Information Sciences 586 (2022) 192–208

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A diversity metric for population-based metaheuristic algorithms

Valentín Osuna-Enciso [a,*], Erik Cuevas [a], Bernardo Morales Castañeda [a]

[a] Universidad de Guadalajara, Blvd. Marcelino García Barragán 1421, esq. Calzada Olímpica, C.P. 44430 Guadalajara, Jalisco, Mexico

ARTICLE INFO

Article history:
Received 7 January 2021
Received in revised form 17 November 2021
Accepted 23 November 2021
Available online 4 December 2021

Keywords:
Diversity metric
Metaheuristic algorithm
Exploration–exploitation balance

ABSTRACT

In Metaheuristic Algorithms (MA), the balance between exploration and exploitation is a common issue considered an open research problem in the MA community. Besides its particular parameters, another way to control the Exploration–Exploitation Balance in MA is by using a diversity metric (DM) as a guide. However, this procedure has two drawbacks: its computational cost and its effectiveness to represent the actual diversity of the population. This paper proposes a DM for real-coded candidate solutions. The approach utilizes surrogate hypervolumes to calculate the spatial distribution of the individuals in the population. In a comparison against five DM reported in the literature, our proposal achieves comparable results in terms of stability, sensitivity, and robustness in the presence of outliers, without significant increases in the computational cost.

1. Introduction

Metaheuristic Algorithms (MA) are techniques inspired by several sources such as evolution, physics, nature, or even biology [1]. These methods provide robust performance, flexibility, and adaptability to solve a wide range of problems, e.g., optimization problems [2]. Algorithmically, these approaches adopt two search procedures: 1) one that operates over the whole feasible space (exploration), and 2) another that examines a local portion of the search space (exploitation)[3]. Excessive use of exploration usually degrades the performance of the algorithm [4]. In contrast, too much exploitation can produce premature convergence, among many other issues [5]. Therefore, an exploration–exploitation balance (EEB) is a desirable trait of a successful MA [6,7].

According to [8,9], the techniques to achieve EEB are divided into the following approaches: (1) adaptive, (2) hybrid and (3) interactive. In the adaptive paradigm, a MA can control and change its parameters and operators as the iterations advance, either in a deterministic or a dynamic manner[10–13]. For their operation, they utilize information from the position of candidate solutions, their fitness values, or a mixture of both. The main advantage of an adaptive approach is that it is easy to implement its mechanisms to the original algorithms without much effort. Examples of adaptive approaches include those based on rules, co-evolution, entropy, and fuzzy principles[14,15].

Another important way to achieve the exploration–exploitation balance in MA is by using topologies [16]. Topology refers to how a population is spatially organized. Under this perspective, different population structures produce other communication channels among the individuals as the iterations advance, enhancing population diversity during the optimization process. The most common topologies are panmictic, cellular or diffused (fine-grained), distributed and island (coarse-

* Corresponding author.

https://doi.org/10.1016/j.ins.2021.11.073

grained). However, several other topologies exist in the literature, such as hierarchical, small-world, random, irregular, or co-evolution [17].

Another method to control EEB is the hybridization between two or more MA or between metaheuristics and other optimization tools [18]. The main idea is that a single technique presents the tendency toward either intensification or diversification in the optimization process [19]. Therefore, a combination of algorithms reduces the weakness of each other, increasing their potential. In this case, the main benefit of the hybrid paradigm is the control of the EEB by employing the intrinsic characteristics of the original algorithms.

The interactive approach is a recent proposal that incorporates a human–computer interface that employs expert knowledge to increase the capacities of the optimization strategy [20]. Under this approach, a specialist evaluates the fitness value of every candidate solution because of the lack of an objective function in closed form (in fact, this is the principal advantage of the approach).

In each of the three mentioned mechanisms, the goal is to increase the population diversity of a MA while at the same time avoiding premature convergence [21]. Controlling the diversity by the adaptive approach is a common technique to reach an equilibrium between exploration and exploitation. Therefore, the diversity regulation can be obtained by following several methodologies. One of them is the inclusion of a diversity metric for guiding the evolving process [22–24]. In MA with real-coded candidate solutions, the goal is to maintain diversity in the genotype pool to reach an adequate EEB [25]. Under such conditions, a Diversity Measure (DM) calculates the heterogeneity of the population by utilizing the information of every candidate solution [26].

There exist various studies in the area of diversity-guided for MA. For instance, to alleviate the premature convergence in Particle Swarm Optimization, the authors in [27] proposed an improvement that splits the population into a current swarm and a memory swarm. The second sub-population diversity is modified using two measures: spatial population diversity and population fitness spatial diversity. In [7], the authors compare the convergence behavior of a Genetic Algorithm based Artificial Neural Network for classification and forecasting under different crossover operators. Their study suggests that the arithmetic crossover could be the most appropriate operator for this kind of architecture applied to a classification problem. The reason is that it helps the GA to train an architecture capable of avoiding over-fitting. As an alternative to the uniform random selection in canonical Differential Evolution, the authors in [28] proposed a multi-objective optimization algorithm using fitness and diversity information. This approach provides a simple and efficient EEB, with better results than other metaheuristics.

In [29], it was suggested an improvement to the Teaching Learning-based optimization algorithm for the EEB by introducing the concept of the historical population together with mutation and crossover. Authors in [30] also introduced an EEB mechanism for partitioning the evolved population into buckets to maintain a practical and useful diversity of the population. Similar approaches have been applied to other MA: e.g., [31–37].

As mentioned in [26], the numerical evaluation of EEB has been achieved by methods that consider certain features in the evolving population. In that sense, several diversity metrics have been recently introduced in the literature. For instance, Tilahun [38] designed an exploration metric that utilizes the averaged difference among the average and every candidate solution at a particular iteration to guide a hyper-heuristic process around the Predator–Prey metaheuristic. Experimental results show that the approach improves the original algorithm's performance increasing the computational cost. In the same venue, Gabor and Belzner defined a notion of diversity that is domain-independent, called genealogical diversity [39]. This metric considers the similarity among individuals' genes by adding some 'trash' genes to the original individuals. In [25] the authors proposed a new DM for linear Genetic Programming. It computes the symmetric pairwise difference among individuals, which is called Universal Information Distance. The findings suggest that such a measure is strongly correlated with two other phenotypic-based measures providing a lower computational cost. The authors in [6] presented both an extensive taxonomy (based on different factors) and a survey of MAs that promotes population diversity.

In [6,40], a DM is proposed to consider an average distance between individuals as a stepping stone to calculate the diversity. In our approach, we calculate the hypervolume of only two multi-dimensional cubes at each iteration. Then, the ratio of these volumes is used to determine the DM. Although the proposal is simple, its behavior is similar to other well-known DM in terms of robustness in the presence of outliers, stability, and sensitivity. This method has not been proposed in the reviewed literature to the best of the authors' knowledge.

The rest of the paper is organized as follows: Section 2 describes the proposal for diversity calculus based on the volume of multi-dimensional cubes. The used benchmark and a detailed description of the utilized metrics are reviewed in Section 3. In Section 4, we explore both the proposed and state-of-the-art metrics under several scenarios such as multi-modality of the search space, population size, dimensionality, and presence of outliers to test its stability, sensitivity, and robustness. Moreover, we analyze the computational cost in the same section. In Section 5, a discussion of the experimental results is provided. Finally, Section 6 presents the conclusions of the paper, as well as future work.

2. The proposal

Considering as a reference Fig. 1, in a three-dimensional space, a matrix A can be formed by the coordinates of edge vectors $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$:

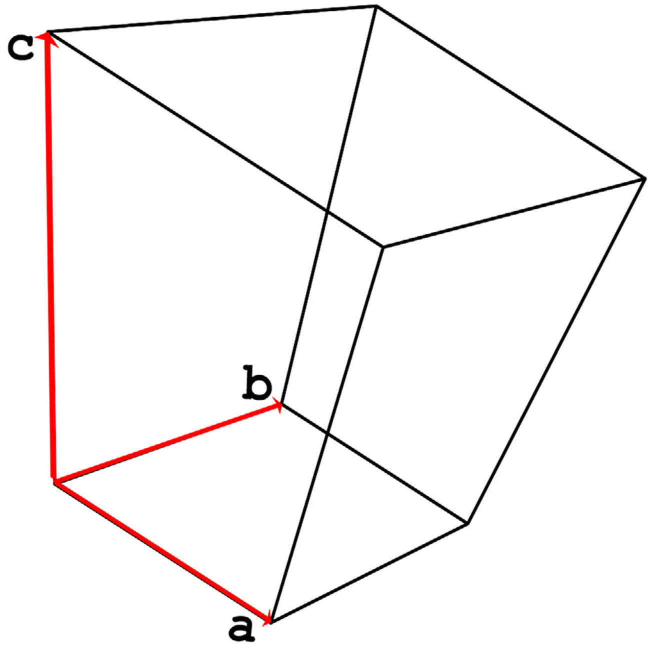


Fig. 1. A non-orthogonal parallelepiped in 3 dimensions.

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \quad (1)$$

whose multiplication by its transpose can be represented as the inner product of vectors:

$$A^T A = \begin{bmatrix} (\mathbf{a}\cdot\mathbf{a}) & (\mathbf{a}\cdot\mathbf{b}) & (\mathbf{a}\cdot\mathbf{c}) \\ (\mathbf{b}\cdot\mathbf{a}) & (\mathbf{b}\cdot\mathbf{b}) & (\mathbf{b}\cdot\mathbf{c}) \\ (\mathbf{c}\cdot\mathbf{a}) & (\mathbf{c}\cdot\mathbf{b}) & (\mathbf{c}\cdot\mathbf{c}) \end{bmatrix} \quad (2)$$

and, substituting $\mathbf{a}, \mathbf{b}$, and $\mathbf{c}$ by $\mathbf{y}_i \in \Re^d, i = 1, \ldots, d$, a general form of Eq. 2 is

$$A^T A = \begin{bmatrix} (\mathbf{y}_1\cdot\mathbf{y}_1) & (\mathbf{y}_1\cdot\mathbf{y}_2) & \ldots & (\mathbf{y}_1\cdot\mathbf{y}_d) \\ (\mathbf{y}_2\cdot\mathbf{y}_1) & (\mathbf{y}_2\cdot\mathbf{y}_2) & \ldots & (\mathbf{y}_2\cdot\mathbf{y}_d) \\ & & \vdots & \\ (\mathbf{y}_d\cdot\mathbf{y}_1) & (\mathbf{y}_d\cdot\mathbf{y}_2) & \ldots & (\mathbf{y}_d\cdot\mathbf{y}_d) \end{bmatrix} \quad (3)$$

By using Eq. 3, the calculus of an arbitrary hypervolume with non-orthogonal edge vectors involves the squared root of its determinant [41]:

$$V_p = \left(det(A^T A)\right)^{1/2} \quad (4)$$

However, by definition of the dot product, if edge vectors are orthogonal with origin in (0,0,...), then Eq. 3 becomes:

$$A^T A = \begin{bmatrix} (\mathbf{y}_1\cdot\mathbf{y}_1) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & (\mathbf{y}_2\cdot\mathbf{y}_2) & \ldots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \ldots & (\mathbf{y}_d\cdot\mathbf{y}_d) \end{bmatrix} \quad (5)$$

and therefore, under such assumptions, an hypervolume can be computed as

$$V_p = \sqrt{det(A^T A)} = \sqrt{\prod_{i=1}^{d} (\mathbf{y}_i\cdot\mathbf{y}_i)} \quad (6)$$

Our proposal considers the computation of two hypervolumes: one corresponding to the limits of the search space and another representing the spatial distribution of the population in the iteration. The volume corresponding to the limits is calculated considering the absolute difference between the inferior and the superior boundaries of the search space:

$$V_{lim} = \sqrt{\prod_{i=1}^{d} |u_i - l_i|} \tag{7}$$

where $\mathbf{l} = \{l_1, l_2, \ldots, l_d\}$ and $\mathbf{u} = \{u_1, u_2, \ldots, u_d\}$ are vectors that represents the lower and upper search space limits, respectively. The volume obtained by Eq. 7 is calculated once at the beginning of the iterations. The second hypervolume represents the evolving population; in this case, to get the edge vectors, we use the inter-quartile range for every dimension of the whole population, being the edge vector coordinates: $\mathbf{y}_1 = (2 * iqr(\mathbf{x}_1), 0, 0, \ldots), \mathbf{y}_2 = (0, 2 * iqr(\mathbf{x}_2), 0, \ldots), \mathbf{y}_3 = (0, 0, 2 * iqr(\mathbf{x}_3), \ldots)$, etc, and its value is computed as follows:

$$V_{pop} = \sqrt{\prod_{i=1}^{d} (\mathbf{y}_i \cdot \mathbf{y}_i)} \tag{8}$$

where $\mathbf{x}_i$ represents the columns in the population of candidate solutions, iqr represents the inter-quartile range, $i = 1, \ldots, d$, and $j = 1, \ldots, N$. Finally, in our proposed approach, the ratio between the hyper-volume of the population of candidate solutions and the hyper-volume of the search space limits is adopted as diversity measure. It is computed as follows:

$$nVOL = \sqrt{V_{pop}/V_{lim}} \tag{9}$$

Authors in [42] proposed a diversity metric that calculates the volume of the union of every cell assigned to every individual in a population. In that case, the measure solves the Klee's measure problem, which has computational complexity $O(N^{d/2} \cdot logN)$ in the best scenario [43]. Our method is $O(d)$ because it only performs a multiplication. Our approach is simple. It adopts the diversity of the population as a proportion between the d-dimensional volume limited by the search space and a similar hypervolume representing the total population. Different from the method suggested in [44], the proposed metric does not use the normalization with the maximum diversity found so far. We call our approach nVOL because it utilizes multidimensional volumes as a base of its procedure (Eq. 9). As a simple visual example of the hypervolumes (areas in this case), let us consider Fig. 2.

Under this condition, ten individuals (o) form the population's distribution, and the search space is located between $[-50, 50]$ per dimension. The triangle markers show the maximum and minimum of the individuals in the population. In Fig. 2, the big square represents the limits of the search space, whose area is the denominator in the ratio of the proposed metric. In contrast, the area of the small square represents the numerator in the mentioned ratio. From the Figure, it is clear

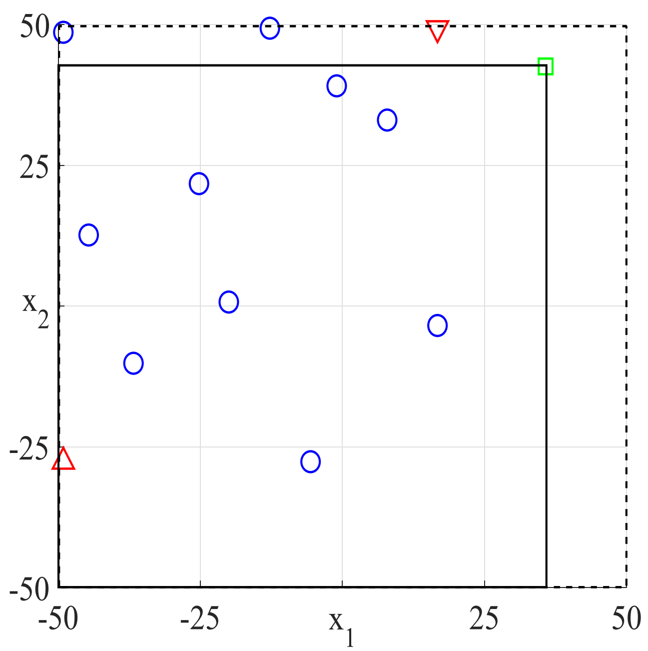


Fig. 2. The areas of a population and a search space in two dimensions.

that two individuals lie outside the representative rectangle; this is because we are using the inter-quartile range of every dimension to avoid a possible bias to the metric due to outliers [45,46]. Since the inter-quartile range only represents half of the data [46], the obtained value is multiplied by 2 to represent the totality of data avoiding the outliers. For the sake of clarification, the proposed metric is explained in Algorithm 1.

Algorithm 1: Calculus of nVOL.

Data: population of candidate solutions ($\mathbf{X}$),
limits of search space ($\mathbf{l}$, $\mathbf{u}$), V_{lim}
Result: Genotype Diversity Measure ($nVOL$)

$V_{pop} = 1$;
for i=1 *to* d **do**
  $V_{pop} = V_{pop} * (2 * iqr(\mathbf{x}_{j,i}))$
$nVOL = \sqrt{V_{pop}/V_{lim}}$;

3. Experimental settings

3.1. Benchmark

The goal of every diversity metric is to accurately reflect the actual diversity, or how different are the candidate solution among them. Therefore, to test new diversity metrics, it is necessary to control the collective behavior of candidate solutions, from a high to a low dispersion, which is in concordance to the convergence behavior of a Metaheuristic Algorithm. In order to evaluate the performance of diversity mechanisms, several benchmarks have been proposed. In our study, we have used the benchmark introduced by [44]. It considers an d-dimensional search space bounded between $[-50, 50]$, with the capability to generate several optimum points. Initially, the individuals are located randomly over the entire search space and reduced 2% at every iteration. Moreover, in [44], the authors suggests that the solutions tend to group around the optima in a good benchmark. An illustrative example of a population formed by $N = 100$ individuals is in Fig. 3. In this case, in the first iteration four optima are randomly placed in the positions $(-27.3, -8.2), (-45.9, 45.6), (-0.9, 12.1), (48.2, 17.2)$, and the remaining 96 individuals are uniformly distributed in the entire search space (Fig. 3(a)). After 25 iterations, the individuals start to proportionally group around every optimum (Fig. 3(c)). At iteration 45, the individuals form four more compact concentrations (Fig. 3(d)), and at the last iteration (50 in the experiments), the candidate solutions are very close to their respective optimum.

3.2. Metrics compared and experimental setup

Because of the nature of the proposed metric index, in this paper we only considered distance-based metrics for real coded representations of candidate solutions, excluding those that use the gene frequency approach. Moreover, from the five original metrics surveyed in [44], we use the expanded version of their standardized versions (except for the measure proposed by the authors in [47]):

$$DTAP = \frac{\frac{1}{N}\sum_{i=1}^{N}\sqrt{\sum_{k=1}^{d}(x_{i,k} - x_{j,k})^2}}{NMDF} \tag{10}$$

$$TD = \frac{\frac{1}{d}\sqrt{\sum_{k=1}^{d}\left(\bar{x}_k^2 - (\bar{x}_k)^2\right)}}{NMDF} \quad where \quad \bar{x}_k^2 = \frac{1}{N}\sum_{i=1}^{N}x_{i,k}^2 \tag{11}$$

$$MI = \frac{\sum_{k=1}^{d}\sum_{i=1}^{N}(x_{i,k} - \bar{x}_k)^2}{NMDF} \tag{12}$$

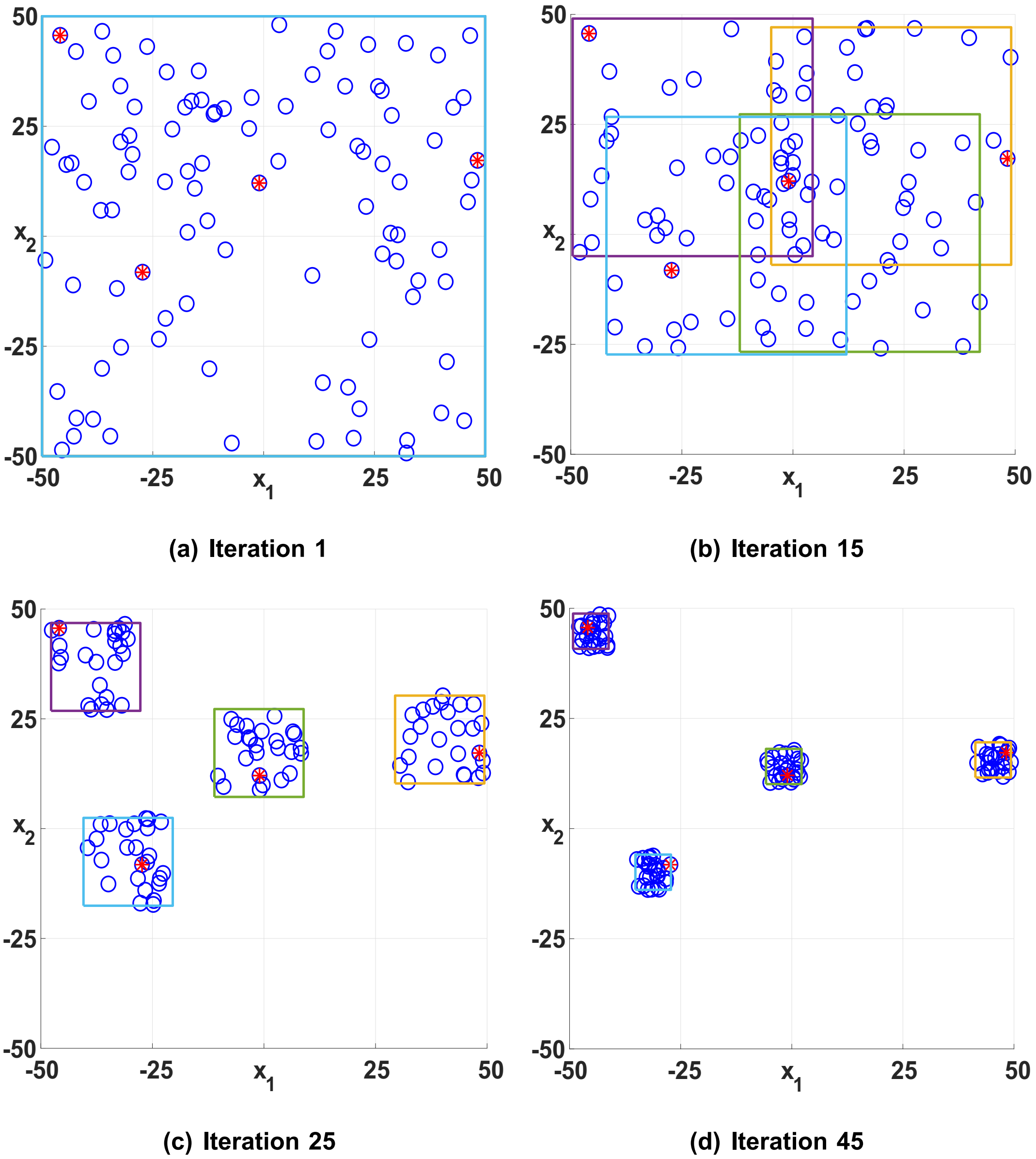


Fig. 3. An evolving population with four optima.

$$PW = \frac{\frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\sqrt{\sum_{k=1}^{d}(x_{i,k}-x_{j,k})^2}}{NMDF} \tag{13}$$

$$VAC = \frac{\frac{1}{N}\sum_{i=1}^{N}(\bar{x}_i-\bar{x})^2}{NMDF} \quad where \quad \bar{x}_i = \frac{1}{d}\sum_{k=1}^{d}x_{i,k} \quad and \quad \bar{x} = \frac{1}{d\cdot N}\sum_{i=1}^{N}\sum_{k=1}^{d}x_{i,k} \tag{14}$$

where *NMDF* stands for normalized in terms of the maximum diversity found so far, and *N* is the population's size of *x* candidate solutions, with a dimensionality denoted by *d*. The name of every metric is, respectively: (10) Distance to average point measure, normalized; (11) True diversity, normalized; (12) Moment of inertia, normalized; (13) Mean of the pairwise distance among individuals in the population; and (14) Variance average of chromosomes, normalized. It is important to notice that the normalization process is because as every measure gives values in different ranges, a direct comparison would be unfeasible. However, this is not the case with our proposal, as its values are always in the range [0,1] avoiding the normalization step.

Concerning the experimental setup, the hardware and software to implement the benchmark and other metrics are Matlab 9.4, 64 bits OS Windows 10, and 16 GB in RAM. In the experiments, 50 runs of 50 iterations have been considered to calculate the average of each metric. Among the complete set of experiments, we experimented with landscapes with 1, 2, 4, 8, and 16 global optima; populations with $N = \{50, 100, 300, 500\}$ individuals; and dimensions from $d = \{2, \ldots, 100\}$ per individual.

4. Experimental study

4.1. Uni-modal landscape

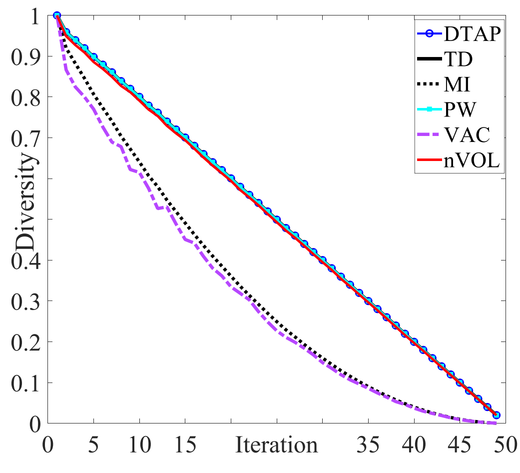
For the first experimental part, we consider a landscape with only one optimum but with a dimensionality of 2 and 100. The goal is to directly compare if the behavior of the different diversity metrics is affected or not by the dimensionality of the search space, for optimization problems with one global optimum. Individuals in the population evolve from completely scattered to a group around the optimum, completing a run. The same process iterates 50 times, after which we compute the average of the results. The arithmetical mean of each of the different measures is shown in Figs. 4(a) and 4(b), for 2 and 100 dimensions, respectively.

Fig. 4 shows the results when the nVOL metric is not normalized against maximum diversity found so far, as a clear difference against the compared metrics. From this figure we can clearly see how some of the algorithms, nVol included, properly follow the expected behavior of a linearly decreasing diversity. This correctly represents the diversity diminishing through the iterations as the whole population begins to group around one global optimum. The MI and VAC measures diverge from this behavior. They show an exponential decreasing of the diversity. After applying the non-parametric Wilcoxon test in pairs, the results provide statistically significant evidence that the behavior of nVOL differs from the behavior of every other metric (Table 1, $p < 0.05$). Fig. 5 shows the results when nVOL employs a normalization step. We can see that its results are not affected in a meaningful manner. Because of this, and to reduce the computational complexity of the proposal, for the remaining experiments the nVOL measure is not going to utilize any normalization.

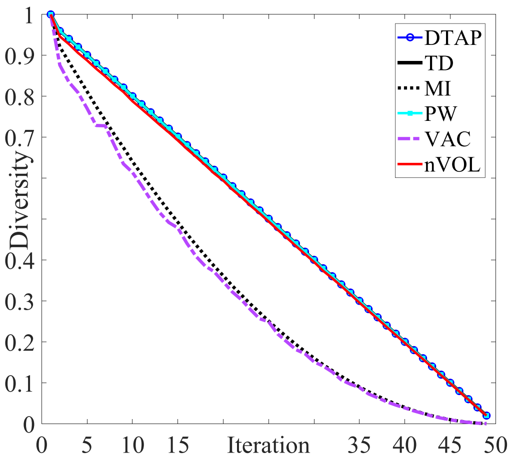
4.2. Multi-modal landscape

Fig. 6 depicts the response of the six presented diversity measures for landscapes with 2, 4, 8, and 16 optima, considering a population of 100 search agents with a dimensionality of 100.

Similar to the previous experiment, although the agents started from an initial random distribution, they finally concentrate around the different optima. This could cause variations in the evaluation of the diversity if the metric employed is not enough robust. As an example, a particular diversity metric could produce a high value if N search agents are grouped too close around the multiple optima of the search space instead of N search agents dispersed through the search space. This is because the distance among the optima (where the search agents are grouped) is longer than the distance between the farthest dispersed search agents. This is easily seen in the graphs, where it is clear that the number of optima affects the ability



(a) 2 dimensions



(b) 100 dimensions

Fig. 4. Averaged diversity for unimodal landscape (nVOL not normalized).

Table 1
Wilcoxon test for the unimodal landscape experiment.

nVOL vs	DTAP	TD	MI	PW	VAC
2	1.18e-77	2.49e-74	0	2.21e-73	0
100	3.78e-12	0	0	1.68e-11	0

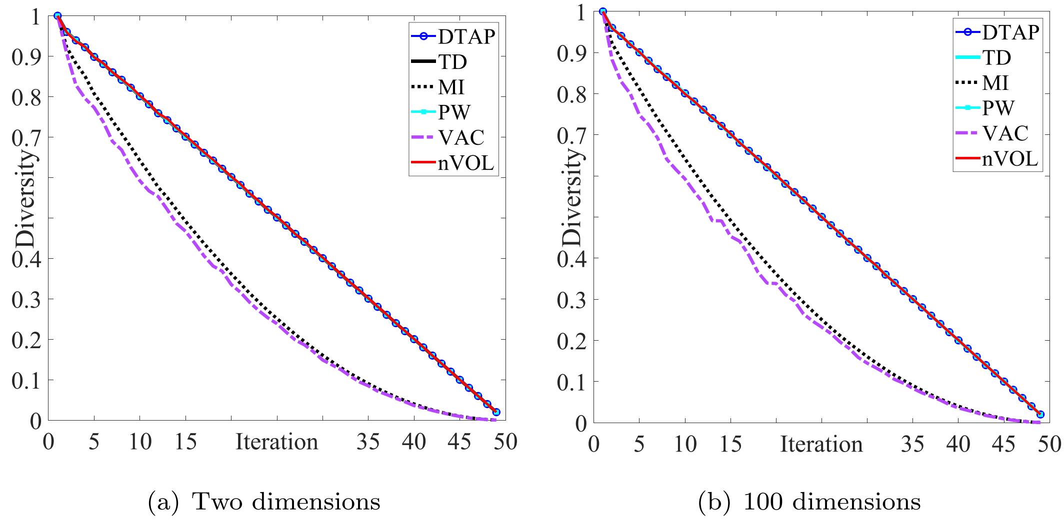


Fig. 5. Averaged diversity for unimodal landscape (nVOL normalized).

of every diversity metric to reflect the true diversity of candidate solutions. For more than two optima, the concentration of the candidate solutions in groups produces a strong bias, privileging the spacing of those groups rather than the difference among the individuals.

Fig. 6(a) shows that PW and nVOL are less affected by this bias when it is considered benchmark functions with 2 modes. This fact can be visualized through the reduction in diversity during the iterations without presenting a random increment due to the bias. In Figs. 6(b), 6(c) and 6(d), it is shown the diversity metrics for 4, 8 and 16 optima. According to these figures, the diversity metrics perform similar as the previous case. At the beginning of the optimization process, the search agents are randomly dispersed in the search space. In this point, every metric presents a more or less linear decrement in diversity as the iterations evolve. Then, several groups begin to consolidate around the distinct optima. In this moment, the diversity measured by every diversity metric tends to be affected by the bias showing an increment in diversity. Under such conditions, we calculate the Euclidean distance between a line representing an ideal diversity and the complete results (not the average) obtained with every diversity metric. Then, we apply the Wilcoxon test to verify the null hypothesis that nVOL median behavior is different from every other metric (Table 3, $p = 0.05$). Moreover, we apply the post hoc Nemenyi test [48] to the same data, with a confidence value of $p = 0.05$ (Fig. 3).

The results from Table 2 demonstrate that every metric's behavior is adversely affected as the number of modes increases. Nevertheless, the magnitude of the effect is different for each metric. Specifically, for more than 2 modes, MI and VAC metrics seem to be the less affected by the bias. However, if it is considered the full set of experiments, these metrics are the least accurate for 1 and 2 modes. If we want to analyze the performance or accuracy of the diversity metrics for the widest range of benchmark functions, these two cannot be considered as the most accurate. In Table 2, the next diversity metric with the less impact from the increment in modes is nVOL. Such data is statistically validated with the Nemenyi test with 95% certainty, as depicted in Fig. 7. Therefore, some interesting conclusions can be formulated if we consider only unimodal functions. Under these conditions, the proposed diversity metric nVOL seems to be the most accurate. Conversely, for multimodal functions MI and VAC both can be considered the most accurate, since no statistical difference has been detected between them. And lastly, when considering both unimodal and multimodal functions, the superior metric is nVOL.

4.3. Stability and sensitivity analysis

Since the experimental results do not follow a normal distribution, instead of using the standard deviation of 50 iterations, we compute the stability of the search diversity metric by considering the dispersion among the 93% of the repeated data. The dispersion value uses the average sum between the fifth high diversity and the second-lowest diversity of one execution. This experiment uses population sizes of 50, 100, 300, and 500 individuals, and dimension $d = 2$. According to the

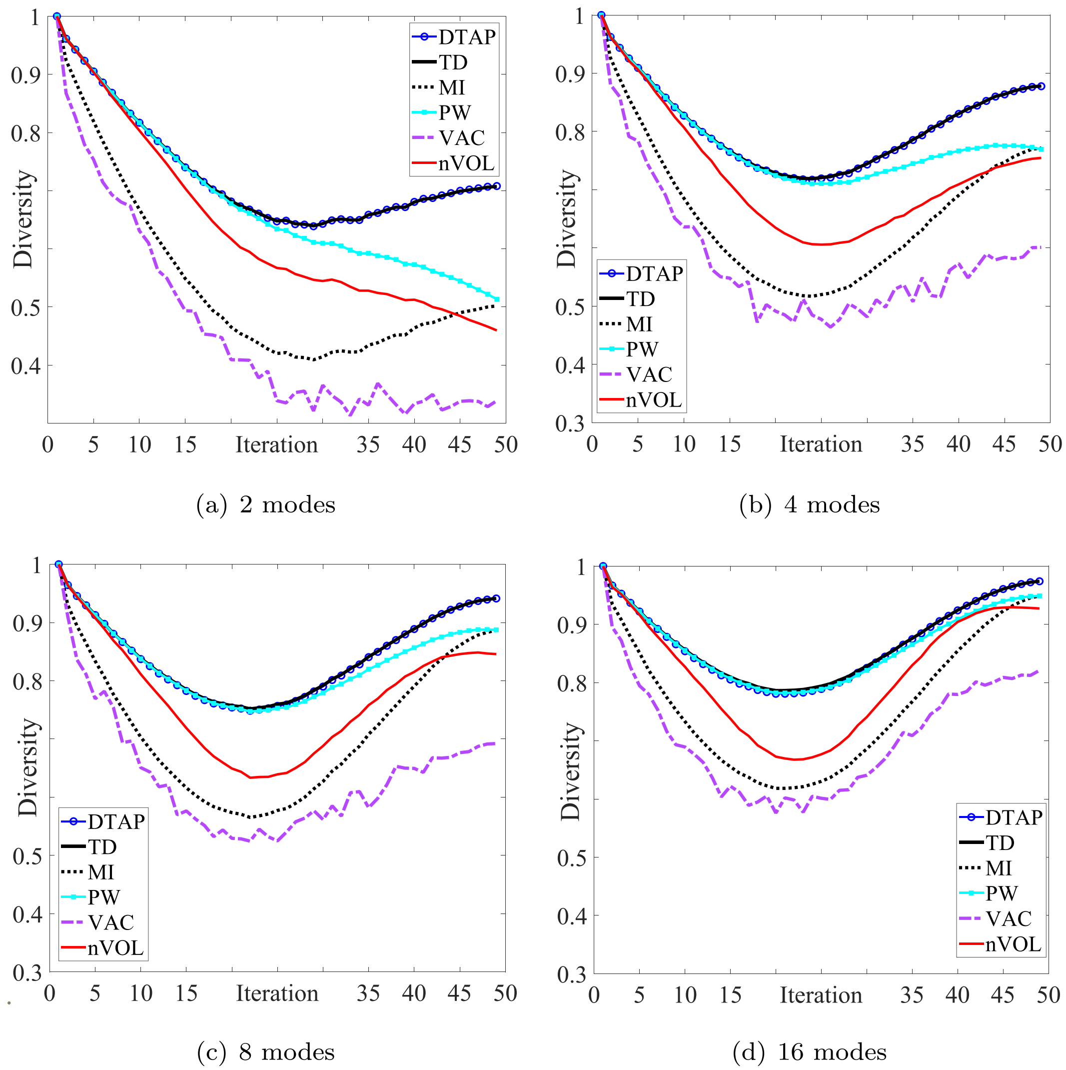


Fig. 6. Averaged diversity for multimodal landscapes.

Table 2
Effect of multimodal search space over each diversity metric.

Modes	Averaged error					
	DTAP	TD	MI	PW	VAC	nVOL
2	2.2146	2.2135	1.4497	1.6951	1.7108	**1.3967**
4	2.8527	2.8557	2.2355	2.5578	**1.8867**	2.2624
8	3.1632	3.1679	2.7003	3.0146	**2.3368**	2.7341
16	3.3378	3.3458	2.9632	3.2706	**2.6374**	3.0815

Table 3
Wilcoxon test for the multimodal landscape experiment.

Modes nVOL vs	DTAP	TD	MI	PW	VAC
2	7.55e-10	7.55e-10	0.0178	7.55e-10	0.03719
4	0	0	0	0	4.04e-249
8	0	0	6.66e-320	0	5.37e-225
16	0	0	0	0	8.13e-258

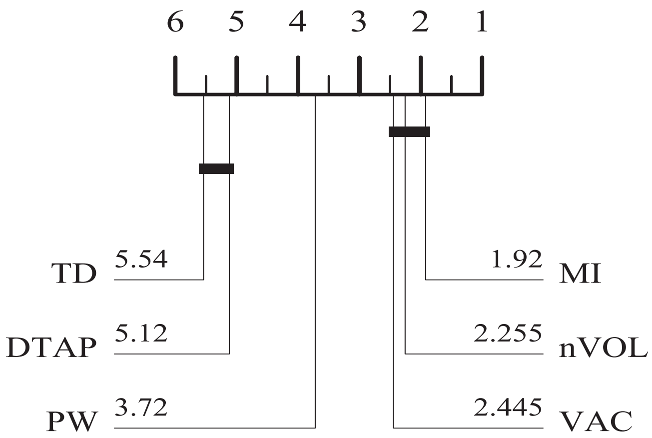


Fig. 7. Comparing the effect of modes.

results shown in Table 4, the stability of every diversity measure improves as the population size increases, which is in agreement with the results presented in [44]. Also, as in the previous experiments, the Wilcoxon test is applied over the same results. According to this analysis, there exist significant evidence that nVOL is different from the other metrics (Table 5, $p = 0.05$). A close analysis at the last column in Table 4 demonstrates that our proposed metric nVOL is the second-worst in terms of stability, only being better than VAC. Applying the Nemenyi test ($p = 0.05$) to the data, the results depicted in Fig. 8 are produced. By considering the experimental results, all the diversity metrics are substantially different from each other. Therefore, a classification in terms of stability, and from the best to the worst, we have the following rank TD, PW, DTAP, MI, nVOL, and VAC. For the sensitivity analysis, we consider as experimental setup the combination of different elements such as diversity measure, population size, the modes of the landscape, and dimensionality. In the first sensitivity experiment, we tested every metric in terms of dimensionality. For example, one case considers DTAP as the metric under scrutiny, one global optimum, a population size of 100 individuals, dimensions of $\{2, 10, 30\}$, and 50 iterations. Once the first three trials are complete, we use the nonparametric statistical Friedman test as a robustness criterion to test the null hypothesis that all samples are equal. As a result, if $p \leqslant \alpha = 5\%$, we accept the alternative hypothesis that the data come from different medians and a counter increases. To get the experiments in percentage form, we complete 100 executions per experimental unit. Table 6 shows the results of the sensitivity analysis as a function of dimensionality. The rejection of the null hypothesis means that the measure is sensitive to changes in dimensionality. Therefore the lower percentages in Table 6 represent the best results, where the best ones are boldfaced for the number of modes. Applying the Nemenyi test over the complete experiment produces Fig. 9. It shows a small difference among the metrics with changes in dimensionality. However, according to the Figure, the ranking of the diversity metrics is, from the best to the worst ranking as follows TD, DTAP, MI, nVOL, PW, and VAC.

The second sensitivity analysis explores the robustness of every metric, with changes in the population size. Table 7 shows the values obtained with this analysis. In this case, it is valid the same criterion as in the previous test: lower percentages mean more robustness to changes in the population size. The calculus of the critical differences with the Nemenyi test, depicted in Fig. 10, complements the data reported in Table 7. From the image, the ranking of the metrics is as follows: DTAP, TD, PW, VAC, MI, and nVOL.

4.4. The effect of outliers

This section analyzes the effect of different outlier levels in the diversity measures, with percentages of 1%, 2%, 5%, and 10% of the population. Under this study, from the whole population, a portion of the individuals will not tend to concentrate around every optimum, and they will be at random positions into the search space. Each experimental unit utilizes a portion of outliers (injected at the 10th iteration), a unimodal landscape, $d = 100, N = 100$, and 50 repetitions per execution. The

Table 4
Results of the stability experiment.

size	Averaged Dispersion Range					
Pop	DTAP	TD	MI	PW	VAC	nVOL
50	0.1066	**0.0964**	0.1140	0.0975	0.1899	0.1720
100	0.0882	**0.0759**	0.0895	0.0764	0.1388	0.1257
300	0.0483	**0.0406**	0.0479	0.0415	0.0858	0.0721
500	0.0377	**0.0338**	0.0401	0.0342	0.0731	0.0546

Table 5
Wilcoxon test for the stability experiment.

nVOL vs DTAP	nVOL vs TD	nVOL vs MI	nVOL vs PW	nVOL vs VAC
1.43e-34	1.43e-34	1.43e-34	1.43e-34	1.64e-34

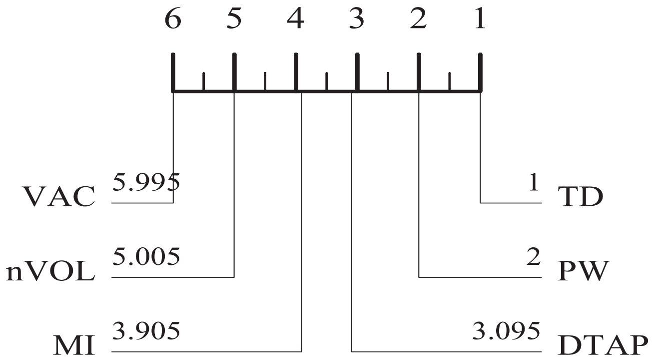


Fig. 8. Comparing the stability of diversity metrics.

Table 6
Results of the sensitivity experiment with $d = \{2, 10, 30\}$ and $N = 100$.

	% p-values $< \alpha$					
Modes	DTAP	TD	MI	PW	VAC	nVOL
1	**37**	42	41	39	46	44
2	42	49	**39**	41	40	40
4	27	**26**	40	49	34	37

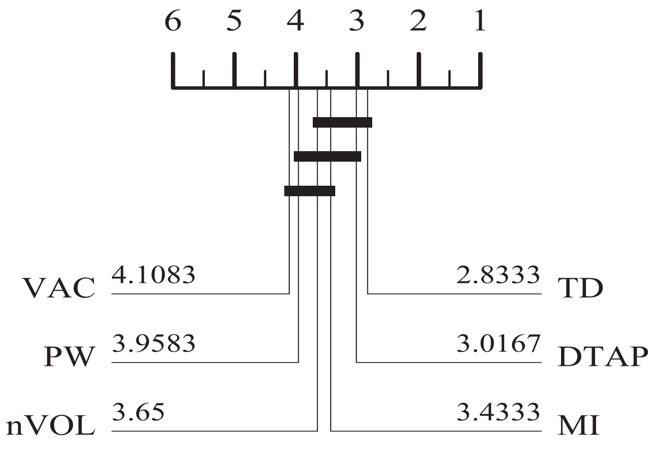


Fig. 9. Comparing the sensitivity of diversity metrics with variable dimensionality.

Table 7
Results of the sensitivity experiment with $N = \{50, 100, 150\}$ and $d = 100$.

	% p-values $< \alpha$					
Modes	DTAP	TD	MI	PW	VAC	nVOL
1	**51**	**51**	67	**51**	57	56
2	48	54	66	**42**	48	54
4	**29**	**29**	33	51	49	57

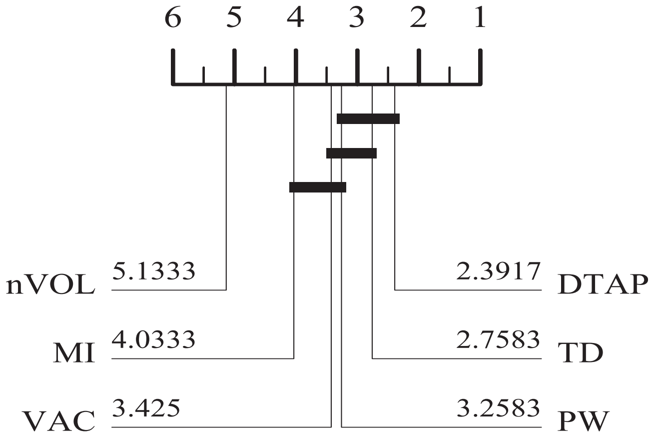


Fig. 10. Comparing the sensitivity of diversity metrics with variable population.

final results averaged over 50 executions are shown in Fig. 12. We also calculate the Euclidean distance between the averaged values and a hypothetical line representing the ideal diversity (Table 8). Without outliers in the population and with a unimodal landscape, almost all measures behave in a quasi-linear way (Fig. 4, and first row of Table 8). However, as the number of outliers increases, it is natural that diversity tends to increase, but the measure must not have much bias toward such outliers. For example, for an outlier level of 1%, almost none measure is biased, but only TD; in fact, the Wilcoxon test gives no significant difference between nVOL and DTAP for this level of outliers (Table 9, $p = 0.05$). Also, TD is the most affected metric by the presence of outliers at 1% (Fig. 12(a)), but it remains closer to a line than MI and VAC. The outlier percentage tends to decrease the robustness of every metric. By reviewing Table 8, it is clear that PW robustness is affected slower than every other measure. In contrast, our proposal presents a better robustness to the outlier factor. The explanation for this behavior is that we are using the not normalized version of nVOL, so there is a slight bias in the metric as the iterations advance (e.g., nVOL at the first iterations in Fig. 4(b)). In the last iterations, our proposal has a degradation in its robustness due to the number of outliers (e.g., the value of nVOL in the final iteration in Fig. 12(a) is 0.01994, whereas the same in Fig. 12 (d) is 0.0218). Therefore, at least for up to 10% of outliers, the previous effects give a robustness improvement in our proposed diversity measure. To further clarify this issue, we apply the Nemenyi test to the complete experimental results from this section (Fig. 11). From the Figure, DTAP and nVOL have no statistical difference ($p = 0.05$), and therefore they share the first rank. Likewise, by taking the accumulated errors from Table 8 and the result of the post hoc test in Fig. 11, the ordering of the metrics for this experiment in terms of robustness is as follows nVOL, DTAP, PW, TD, MI, and VAC.

4.5. Computational cost

Finally, in the last experiment, we calculate the execution time for every diversity metric to compute its value. The settings for this experiment are: $N = \{100, 150, 500\}, d = 100$, one global optimum, 50 iterations, and 50 executions. The averaged times are shown in Table 8, and we apply the post hoc Nemenyi test [48] over the complete data from the 50 executions in order to compare all the metrics against each other, and also we apply the Wilcoxon test over the same results (Table 11, $p = 0.05$). Fig. 13 presents critical differences obtained with the Nemenyi test, with a connection on the metrics that are not statistically different by considering $p = 0.05$. It is important to notice that, in this experiment, we do not apply the Friedman test because of the clear differences among the computational cost of every metric. After reviewing these results, the diversity measures with the lowest computational costs are DTAP, TD, and MI, where the measures almost have no changes as the population's size increases. The result in Fig. 13 suggests that TD and MI have no significant differences, and they are ranked as first, whereas DTAP ranks second place. see Table 10.

Table 8
Effect of outliers over each diversity metric.

%	Averaged error					
Outliers	DTAP	TD	MI	PW	VAC	nVOL
0	0.0800	**0.0798**	1.3549	0.0799	1.5394	0.0895
1	0.0756	0.2564	1.2838	0.0829	1.5260	**0.0748**
2	0.1152	0.4186	1.2194	0.1359	1.4282	**0.0724**
5	0.2971	0.7920	1.0474	0.3383	1.2990	**0.0710**
10	0.6355	1.2557	0.8666	0.6800	1.3424	**0.1147**

Table 9
Wilcoxon test for the outliers experiment.

Outliers (%)nVOL vs	DTAP	TD	MI	PW	VAC
1	0.5147	7.55e-10	7.55e-10	7.55e-10	7.55e-10
2	1.97e-9	7.55e-10	7.55e-10	9.06e-10	7.55e-10
5	7.55e-10	7.55e-10	7.55e-10	7.55e-10	7.55e-10
10	7.55e-10	7.55e-10	7.55e-10	7.55e-10	7.55e-10

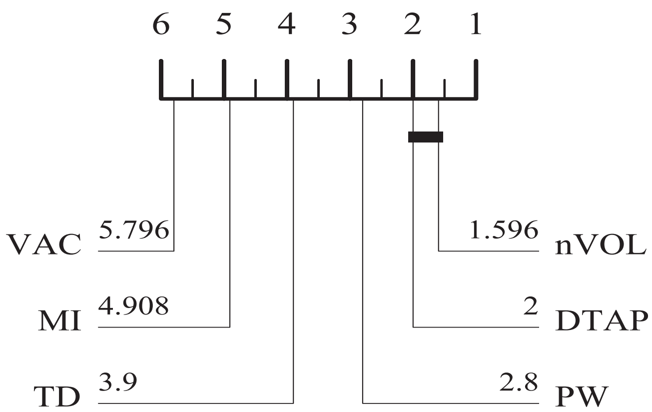


Fig. 11. Comparing the effect of outliers.

On the other hand, VAC, nVOL, and PW are, computationally more expensive than the other metrics if the number of candidate solutions increases. According to Fig. 13, VAC and nVOL are not statistically different, and therefore they can be ranked as third place. The computational cost of PW is almost exponential, behaving more steeply than VAC and nVOL. Therefore, concerning the computational times reported in Table 8 and Fig. 13, we can rank the metrics from the best to the worst as follows TD, MI, DTAP, VAC, nVOL, and PW.

5. Discussion

In this paper, several experiments have been designed so that the population of candidate solutions is distributed along with the search space in the first iteration, then forming compact groups as the iteration advances. The idea is to simulate the behavior of metaheuristics methods when they explore the search space in the first iterations and exploit promissory areas in the last iterations. Also, the experiments consider different scenarios based on the multimodality of the search space, populations of up to 500 individuals, and multidimensionality. We compare the proposed nVOL against the best five diversity metrics found by the study made in [44].

In the experiment with many local optima, the results show that every metric is very sensitive to this feature of the search space, confirming the results of [44]. Nevertheless, in comparing the actual behavior vs. an ideal behavior of every metric for 2, 4, 8, and 16 modes, we found that our approach is the second more linear, being only under MI in such an aspect. The Nemenyi test statistically validated this result, with a $p = 0.05$. It is essential to notice that, despite those results, none of the metrics reflect the unbiased ideal behavior of the metric as the number of modes increases. However, this experiment is important because some existing metaheuristics methods aim to find several local optima at once, such as those designed to solve multimodal problems [49]. This fact could help the diversity metrics that better reflect the actual diversity of the search space. Therefore, the results of the multimodal experiment suggest that our proposal is slightly better than other state-of-the-art diversity metrics.

The stability and sensitivity experiments include the increment of the population and dimensionality. Regarding to the stability, nVOL presents a poor performance, with a rank 5th according to the Nemenyi test. The utilization of the inter-quartile range in the computation of the surrogate hypercube could explain the low stability of our proposal. The sensitivity test includes two parts: changes in dimensionality and changes in population. Our proposal has an average performance in the first case, being ranked 4th according to the Nemenyi test and being only better than PW and VAC. On the other hand, the increment in the population size adversely affects nVOL, ranked the last of all the metrics.

Our approach to measure diversity gives the best result against the other metrics in the presence of outliers. It is ranked as the first place according to the Nemenyi test. Even considering that nVOL stability is affected by the utilization of the inter-quartile range, we believe that its use favors the actual diversity metric in the case of outliers. The computational cost ranks as 5th place to nVOL, according to the statistical post hoc test, which is only below the most expensive metric, PW. However, while our approach is more than 30 times faster than PW for a population of 500 individuals, it is only between

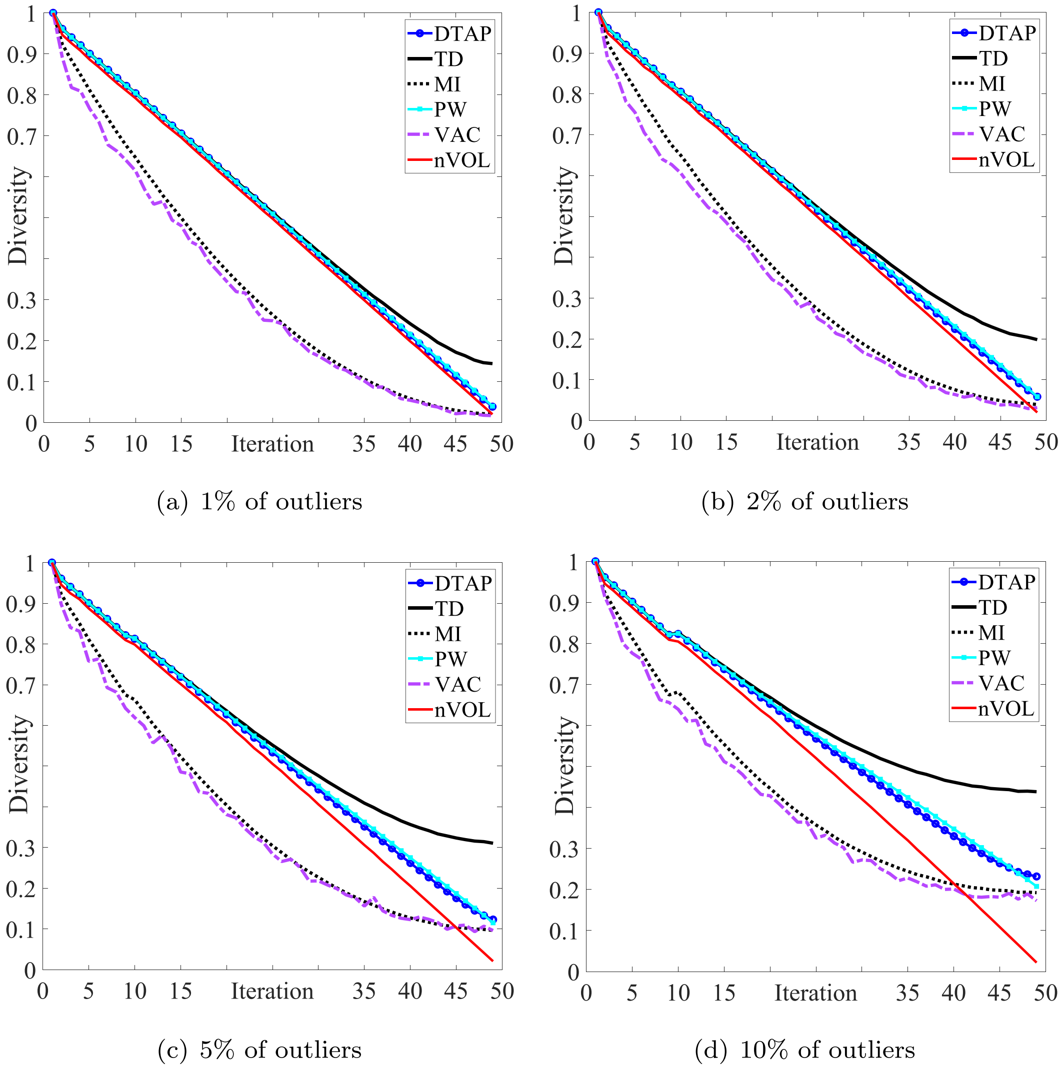


Fig. 12. Averaged diversity for unimodal landscape with outliers.

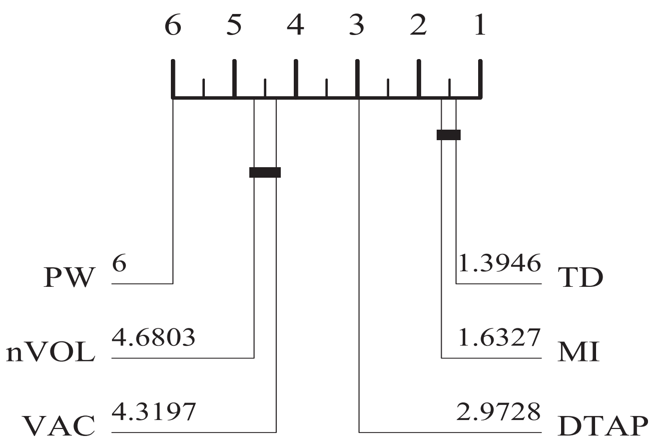


Fig. 13. Comparison of computational times.

Table 10
Averaged computational time of each diversity metric.

Pop	Averaged time in seconds					
size	DTAP	TD	MI	PW	VAC	nVOL
100	0.000241	0.000231	**0.000225**	0.001374	0.000403	0.000576
150	0.000253	**0.000235**	0.000237	0.002981	0.000482	0.000647
500	0.000460	**0.000344**	0.000416	0.041092	0.001420	0.001336

Table 11
Wilcoxon test for the computational cost experiment.

nVOL vs DTAP	nVOL vs TD	nVOL vs MI	nVOL vs PW	nVOL vs VAC
1.37e-24	7.15e-26	7.15e-26	7.15e-26	6.55e-23

4 and 3 slower than the remaining metrics and for the same population size. Under such conditions, we can consider that our proposal is competitive regarding the computational cost in comparison with the five state-of-the-art diversity metrics.

Our approach to measure diversity gives the best result against the other metrics in the presence of outliers, being ranked first according to the Nemenyi test. Even considering that nVOL stability is affected by the utilization of the inter-quartile range, we believe that its use favors the metric in the case of outliers. The computational cost ranks as 5th to nVOL, according to the statistical post hoc test, which is only below the most expensive metric, PW. However, while our approach is more than 30 times faster than PW for a population of 500 individuals, it is only between 4 and 3 slower than the remaining metrics and for the same population size. In that sense, we can consider that our proposal is competitive regarding the computational cost in comparison against five state-of-the-art diversity metrics.

6. Conclusions and future work

This paper has introduced a new diversity metric for populations of real solutions for Metaheuristic Algorithms. The approach considers the numerical computation of two hypervolumes, one belonging to the search space and the other to the individuals in the current iteration. For this reason, we called the diversity metric as nVOL. In order to evaluate the performance of the proposed diversity metrics, we use several tests. The experiments are designed so that the population of candidate solutions is distributed along with the search space in the first iteration, forming compact groups as the iteration advances. Moreover, we consider single and multiple modes in the search space, populations up to 500 individuals, and multi-dimensionality, as proposed by [44]. Also, our method has been compared with five state-of-the-art metrics by considering repeatability, robustness, number of outliers, and computational cost.

The most important advantages of our proposal are its good performance in the presence of outliers and relatively good behavior in multimodal search spaces. Another characteristic of the approach is its competitive computational cost, which is 30 times faster than the worst diversity metric found in the literature, but only three times slower than the quicker metrics. As future work, we intuitively consider that nVOL could be used, with minor changes, to substitute the computation of Pareto dominance in Multi-Objective algorithms. Another possible alternative can be testing our proposal in diversity-guided Metaheuristic methods to improve its convergence speed when the search space is multimodal.

CRediT authorship contribution statement

Valentín Osuna-Enciso: Conceptualization, Methodology, Validation, Writing - original draft, Visualization, Software. **Erik Cuevas:** Writing - review & editing, Funding acquisition. **Bernardo Morales Castañeda:** Writing - review & editing, Methodology, Validation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors express their gratitude to the Mexican National Council for Science and Technology (CONACYT), and to University of Guadalajara. We greatly appreciate the suggestions made by the reviewers of the first drafts of the manuscript.

References

[1] A. Kononova, F. Caraffini, T. Bäck, Differential evolution outside the box, Information Sciences 581 (2021) 587–604.
[2] T. Back, U. Hammel, H. Schwefel, Evolutionary computation: comments on the history and current state, IEEE Transactions on Evolutionary Computation 1 (1) (1997) 3–17.
[3] S. Bansal, A comparative study of nature-inspired metaheuristic algorithms in search of near-to-optimal golomb rulers for the fwm crosstalk elimination in wdm systems, Applied Artificial Intelligence 33 (14) (2019) 1199–1265.
[4] T. Eftimov, P. Korošec, A novel statistical approach for comparing meta-heuristic stochastic optimization algorithms according to the distribution of solutions in the search space, Information Sciences 489 (2019) 255–273.
[5] A. Eiben, C. Schippers, On evolutionary exploration and exploitation, Fundamenta Informaticae 35 (1–4) (1998) 35–50.
[6] G. Squillero, A. Tonda, Divergence of character and premature convergence: A survey of methodologies for promoting diversity in evolutionary optimization, Information Sciences 329 (2016) 782–799.
[7] R.J. Pendharkar P.C., An empirical study of impact of crossover operators on the performance of non-binary genetic algorithm based neural approaches for classification, Computers and Operations Research 31 (2004) 481–498..
[8] J. Xu, J. Zhang, Exploration-exploitation tradeoffs in metaheuristics: Survey and analysis, in: in: Proceedings of the 33rd Chinese Control Conference, 2014, pp. 8633–8638, https://doi.org/10.1109/ChiCC.2014.6896450.
[9] M. Salleh, K. Hussain, S. Cheng, Y. Shi, A. Muhammad, G. Ullah, R. Naseem, Exploration and exploitation measurement in swarm-based metaheuristic algorithms: An empirical analysis, Advances in Intelligent Systems and Computing 700 (2018) 24–32.
[10] T.Y. Liu J., Zheng S., The improvement on controlling exploration and exploitation of firework algorithm, in: Advances in Swarm Intelligence. ICSI 2013. Lecture Notes in Computer Science, Vol. 7928, 2013, pp. 11–23. doi:10.1007/978-3-642-38703-6_2..
[11] M.Y.S. Hussain A., Trade-off between exploration and exploitation with genetic algorithm using a novel selection operator, Complex and Intelligent Systems 6 (2020) 1–14..
[12] L.K. Hansheng L., Balance between exploration and exploitation in genetic search, Wuhan Univ. J. Nat. Sci. 4 (1999) 28–32..
[13] C.S. Noa Vargas Y., Particle swarm optimization with resets - improving the balance between exploration and exploitation, in: Advances in Soft Computing. MICAI 2010. Lecture Notes in Computer Science, Vol. 6438, 2010, pp. 371–381. doi:10.1007/978-3-642-16773-7_32..
[14] S.A., ype-2 fuzzy logic control of trade-off between exploration and exploitation properties of genetic algorithms, in: Swarm and Evolutionary Computation. EC 2012, SIDE 2012. Lecture Notes in Computer Science, Vol. 7269, 2012, pp. 368–376. doi:10.1007/978-3-642-29353-5_43..
[15] G.M. Lin L., Auto-tuning strategy for evolutionary algorithms: balancing between exploration and exploitation, Soft Computing 13 (2009) 157–168..
[16] N. Lynn, M. Ali, P. Suganthan, Population topologies for particle swarm optimization and differential evolution, Swarm and Evolutionary Computation 39 (2018) 24–35.
[17] D. Yousri, D. Allam, M. Eteiba, P. Suganthan, Static and dynamic photovoltaic models' parameters identification using chaotic heterogeneous comprehensive learning particle swarm optimizer variants, Energy Conversion and Management 182 (2019) 546–563.
[18] C.G.M.M. Lozano, Hybrid metaheuristics with evolutionary algorithms specializing in intensification and diversification: Overview and progress report, Computers and Operations Research 37 (2010) 481–497.
[19] C.S. Hussain K., Salleh M.N.M., On the exploration and exploitation in popular swarm-based metaheuristic algorithms, Neural Comput and Applic 31 (2019) 7665–7683..
[20] L. Salas-Morera, L. García-Hernández, C. Carmona-Muñoz, A multi-user interactive coral reef optimization algorithm for considering expert knowledge in the unequal area facility layout problem, Applied Sciences (Switzerland) 11 (15)..
[21] S.H.R. Rezaei F., Guaspso: a new approach to hold a better exploration-exploitation balance in pso algorithm, Soft Computing 24 (2020) 4855–4875..
[22] C. Li, J. Sun, V. Palade, Diversity-guided lamarckian random drift particle swarm optimization for flexible ligand docking, BMC Bioinformatics 21 (1)..
[23] M. Lukovic, Y. Tian, W. Matusik, Diversity-guided multi-objective bayesian optimization with batch evaluations, Vol. 2020-December, 2020..
[24] A. Bartoli, A. De Lorenzo, E. Medvet, G. Squillero, Multi-level diversity promotion strategies for grammar-guided genetic programming, Applied Soft Computing Journal 83..
[25] M. Gaudesi, G. Squillero, A. Tonda, An efficient distance metric for linear genetic programming, in: Proceedings of the 15th Annual Conference on Genetic and Evolutionary Computation, GECCO '13, Association for Computing Machinery, New York, NY, USA, 2013, p. 925–932..
[26] M. Crepinsek, S.-H. Liu, M. Mernik, Exploration and exploitation in evolutionary algorithms: A survey, ACM Computing Surveys 45 (3)..
[27] W. Lim, N. Mat Isa, Two-layer particle swarm optimization with intelligent division of labor, Engineering Applications of Artificial Intelligence 26 (10) (2013) 2327–2348.
[28] J. Wang, J. Liao, Y. Zhou, Y. Cai, Differential evolution enhanced with multiobjective sorting-based mutation operators, IEEE Transactions on Cybernetics 44 (12) (2014) 2792–2805.
[29] X. Ji, H. Ye, J. Zhou, Y. Yin, X. Shen, An improved teaching-learning-based optimization algorithm and its application to a combinatorial optimization problem in foundry industry, Applied Soft Computing Journal 57 (2017) 504–516.
[30] A. Benbassat, Y. Shafet, A simple bucketing based approach to diversity maintenance, in: Proceedings of the 17th Annual Conference on Genetic and Evolutionary Computation, GECCO '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 1559–1564..
[31] G. Leguizamón, C. Coello, An alternative aco r algorithm for continuous optimization problems, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 6234 LNCS (2010) 48–59..
[32] Y. Shuhao, S. Shoubao, H. Li, A simple diversity guided firefly algorithm, Kybernetes 44 (1) (2015) 43–56.
[33] N. Ben Guedria, Improved accelerated pso algorithm for mechanical engineering optimization problems, Applied Soft Computing Journal 40 (2016) 455–467.
[34] R. Sagban, K.R. Ku-Mahamud, M. Shahbani Abu Bakar, Unified strategy for intensification and diversification balance in aco metaheuristic, in: 2017 8th International Conference on Information Technology (ICIT), 2017, pp. 139–143.
[35] L. dos Santos Coelho, C. Richter, V.C. Mariani, A. Askarzadeh, Modified crow search approach applied to electromagnetic optimization, in: 2016 IEEE Conference on Electromagnetic Field Computation (CEFC), 2016, p. 1.
[36] T. Sharma, M. Pant, Shuffled artificial bee colony algorithm, Soft Computing 21 (20) (2017) 6085–6104.
[37] M. Yang, C. Li, Z. Cai, J. Guan, Differential evolution with auto-enhanced population diversity, IEEE Transactions on Cybernetics 45 (2) (2015) 302–315.
[38] S. Tilahun, Prey predator hyperheuristic, Applied Soft Computing Journal 59 (2017) 104–114.
[39] T. Gabor, L. Belzner, Genealogical distance as a diversity estimate in evolutionary algorithms, in: Proceedings of the Genetic and Evolutionary Computation Conference Companion, GECCO '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 1572–1577..
[40] M. Wineberg, F. Oppacher, The underlying similarity of diversity measures used in evolutionary computation, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2724 (2003) 1493–1504.
[41] N. Beume, S-metric calculation by considering dominated hypervolume as klee's measure problem, Evol. Comput. 17 (4) (2009) 477–492.
[42] B. Lacevic, E. Amaldi, Ectropy of diversity measures for populations in euclidean space, Information Sciences 181 (11) (2011) 2316–2339.
[43] M.H. Overmars, C.-K. Yap, New upper bounds in klee's measure problem, SIAM Journal on Computing 20 (6) (1991) 1034–1045.
[44] G. Corriveau, R. Guilbault, A. Tahan, R. Sabourin, Review and study of genotypic diversity measures for real-coded representations, IEEE Transactions on Evolutionary Computation 16 (5) (2012) 695–710.
[45] G. Terrell, The maximal smoothing principle in density estimation, Journal of the American Statistical Association 85 (410) (1990) 470–477.
[46] J.L. Devore, Probability and Statistics for Engineering and the Sciences, 8th Edition, Brooks/Cole, 2011, iSBN-13: 978-0-538-73352-6..
[47] F. Herrera, M. Lozano, Adaptation of genetic algorithm parameters based on fuzzy logic controllers, in, Genetic Algorithms and Soft Computing, Physica-Verlag (1996) 95–125.

[48] J. Demšar, Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research 7 (2006) 1–30.
[49] V.F. Kuyu Y.C., Advanced metaheuristic algorithms on solving multimodal functions: Experimental analyses and performance evaluations, Arch Computat Methods Eng 1 (2021) 1–13..