Applied Soft Computing 13 (2013) 9–26

Contents lists available at SciVerse ScienceDirect

Applied Soft Computing

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

Review of phenotypic diversity formulations for diagnostic tool

Guillaume Corriveau[a,*], Raynald Guilbault[a], Antoine Tahan[a], Robert Sabourin[b]

[a] Department of Mechanical Engineering, Ecole de technologie superieure, 1100 rue Notre-Dame Ouest, Montreal, Canada H3C 1K3
[b] Department of Automated Manufacturing Engineering, Ecole de technologie superieure, 1100 rue Notre-Dame Ouest, Montreal, Canada H3C 1K3

ARTICLE INFO

Article history:
Received 13 May 2012
Received in revised form 7 August 2012
Accepted 17 August 2012
Available online 31 August 2012

Keywords:
Diversity measures
Evolutionary algorithms
Exploration/exploitation balance
Premature convergence

ABSTRACT

Practitioners often rely on search results to learn about the performance of a particular optimizer as applied to a real-world problem. However, even the best fitness measure is often not precise enough to reveal the behavior of the optimizer's added features or the nature of the interactions among its parameters. This makes customization of an efficient search method a rather difficult task.

The aim of this paper is to propose a diagnostic tool to help determine the impact of parameter setting by monitoring the exploration/exploitation balance (EEB) of the search process, as this constitutes a key characteristic of any population-based optimizer. It is common practice to evaluate the EEB through a diversity measure. For any diagnostic tool developed to perform this function, it will be critical to be able to certify its reliability. To achieve this, the performance of the selected measure needs to be assessed, and the EEB framework must be able to accommodate any landscape structure. We show that to devise a diagnostic tool, the EEB must be viewed from an orthogonal perspective, which means that two diversity measures need to be involved: one for the exploration axis, and one for the exploitation axis. Exploration is best described by a genotypic diversity measure (GDM), while exploitation is better represented by a phenotypic convergence measure (PCM). Our paper includes a complete review of PCM formulations, and compares nearly all the published PCMs over a validation framework involving six test cases that offer controlled fitness distribution. This simple framework makes it possible to portray the underlying behavior of phenotypic formulations based on three established requirements: monotonicity in fitness varieties, twinning, and monotonicity in distance. We prove that these requirements are sufficient to identify phenotypic formulation weaknesses, and, from this conclusion, we propose a new PCM, which, once validated, is shown to comply with all the above-mentioned requirements. We then compare these phenotypic formulations over three specially designed fitness landscapes, and, finally, the new phenotypic formulation is combined with a genotypic formulation to form the foundation of the EEB diagnostic tool. The value of such a tool is substantiated through a comparison of the behaviors of various genetic operators and parameters.

1. Introduction

To estimate the performance of a particular optimizer, practitioners commonly rely on search results, such as the best fitness. However, this information alone may not reveal the underlying behavior of a customized search strategy. Furthermore, theories in the field of metaheuristics are generally difficult to translate into the realities of real-world problems. In fact, these theories are usually either restricted to specific landscape problems or derived for an isolated component of the search process [1]. For these reasons, and considering the No Free Lunch theorem, which stipulates that no one optimizer can dominate in all situations [2], designing an efficient search strategy may be difficult. In this paper, efficiency refers to the ability to find a valuable solution, or solutions, in the shortest possible time.

Diagnostic tools for optimizers may help practitioners determine the impact of different strategies implemented during the search process. More importantly, the information gathered can serve to devise a better search strategy, customized for the problem at hand.

Monitoring the search exploration/exploitation balance (EEB) offers a valuable description of the *working of an algorithm* [3]. In other words, as it is responsible for the specific search path pursued, the EEB may be regarded as a basic efficiency characteristic for any population-based optimizer. The EEB summarizes the way in which resources are allocated. Samples directed toward exploration help in the gathering of knowledge on infrequently visited landscape areas, while exploitation relates to resources dedicated to digging in promising regions. Clearly, excessive exploration can lead to random searching and a waste of computational resources.

* Corresponding author.
E-mail address: guillaume.corriveau@etsmtl.ca (G. Corriveau).

1568-4946/$ – see front matter.
http://dx.doi.org/10.1016/j.asoc.2012.08.046

At the same time, excessive exploitation can lead to local searching and convergence to a suboptimal solution. In fact, what is needed for conducting a search over unknown landscapes with limited resources is a precise EEB, and tools that capitalize on EEB information can be a powerful means for diagnosing the impact of a search strategy and for selecting the best combinations of search parameters.

In a similar line of thought, Bassett and De Jong [4] have provided an evolutionary algorithm (EA) customization tool for monitoring the EEB, with the aim of diagnosing customized reproductive operators. They use multivariate quantitative genetics theory to develop two indicators, perturbation and heritability. The former describes exploration capacity, and the latter estimates exploitation capability. However, this customization tool does not support search component interaction. Turkey and Poli [5] considered a different approach to describe the emergent collective behavior of population-based search process. They used a self-organizing map (SOM), which is a kind of artificial neural network, for tracking the population dynamics. With this system, they extracted multiple properties for characterizing the EEB. Nevertheless, the impact of the SOM parameters, such as grid size and training approach, on the quality of the retrieved EEB features remains unclear.

Our objective here is to develop a diagnostic tool based on population diversity formulations for indicating the optimizer EEB. Two kinds of diversity descriptors can be used to define this framework: genotypic diversity measures (GDMs), and phenotypic diversity measures (PDMs). GDMs characterize the spatial distribution of the population, whereas PDMs depict its fitness distribution, and so refer to the quantity and the quality of the population diversity respectively [6].

To develop the diagnostic tool, two underlying objectives must be achieved. First, the role of both diversity measures must be established. Moreover, since numerous diversity measures have been proposed in the literature over the years, the efficiency and reliability of these formulations must be established. Some studies compare the similarities and differences of GDMs [7–9]. However, to the best of the authors' knowledge, no such study involving PDMs has been conducted. This leads us to state the second objective, which is to review and assess the performance of phenotypic formulations.

The paper is organized as follows: in the next section, we show how the EEB can be represented through diversity measures; in Section 3, we review the phenotypic formulations proposed in the literature; in Section 4, we propose a validation framework and analyze some phenotypic formulations; in Section 5, we develop and validate a new phenotypic formulation; in Section 6, we compare all these formulations over specially designed landscapes; in Section 7, we establish and assess the desirable qualities of a formulation; in Section 8, we present the proposed diagnostic tool and describe it through an application in a genetic algorithm (GA) parameter setting context; finally, in Section 9, we conclude the paper.

2. EEB concept

The EEB can be viewed in terms of one of two paradigms [10]: (1) exploration and exploitation act as opposing forces, where increasing one reduces the other; or (2) they can be considered as orthogonal forces. This second perspective offers the possibility of increasing both exploration and exploitation simultaneously. In fact, it has been shown that the opposing forces paradigm is a special case of the orthogonal forces paradigm [9].

Consequently, monitoring the EEB must involve two metrics: one for the exploration axis, and one for the exploitation axis. Exploration is best described by the genotypic formulation, as it summarizes the distribution of the individuals over the search space, while exploitation is best characterized by phenotypic formulations, as promising regions are targeted based on fitness information. This orthogonal EEB framework is illustrated in Fig. 1. With normalized evaluation, unitary genotypic and phenotypic values relate to maximum exploration and exploitation respectively.

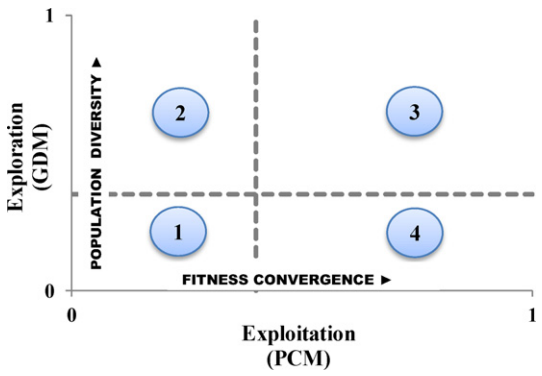

Fig. 1. Orthogonal EEB framework with differentiation into four search zones.

According to this framework, exploration increases with a rise in genotypic diversity. In contrast, exploitation corresponds to the intensification of phenotypic convergence. To avoid confusion, we will refer to the phenotypic convergence measure (PCM) instead of the phenotypic diversity measure (PDM) when dealing with the EEB framework.

Since a mode can be generated from neighboring fitness values belonging to highly scattered individuals, phenotypic convergence should not be employed to terminate a search process. Generally, for population-based optimizers, advancing toward convergence indicates that individuals are becoming increasingly similar. Therefore, phenotypic convergence without genotypic convergence indicates that multiple solutions perform equally well. However, such a condition does not necessarily correspond to a multimodal landscape. In reality, it could refer to a "ring" formation produced by the individuals around a particular optimum. Despite this condition, representation of the orthogonal EEB framework through a GDM–PCM combination provides a way to depict the concept of *useful diversity* introduced by Goldberg and Richardson [11]. As they point out, preserving diversity by itself is not the ultimate goal; it is maintaining diversity that can lead to the identification of good individuals.

Finally, to further illustrate the value of the orthogonal EEB framework, four zones are proposed in Fig. 1. Zone 1 characterizes a population with spatially similar individuals (low genotypic diversity) and heterogeneous fitness (low phenotypic convergence). Zone 2 is characterized by high exploration capability and low exploitation strength. A random search, for instance, would be located in this zone. Zone 3 is the useful diversity area discussed previously, in which exploration and exploitation are maximized simultaneously. Finally, in Zone 4, a searching process is directed toward converges to a single solution. The orthogonal EEB framework appears, therefore, to be more descriptive of the search process than the concept of opposing forces, where, to some extent, only the second and the fourth zones are distinguished.

3. Review of phenotypic formulations

For any landscape structure, the orthogonal EEB framework portrays the way resources are allocated, and, consequently, optimizer performance. In fact, the use of a phenotypic formulation is only justifiable from this perspective. To reduce computational effort, some researchers only consider phenotypic diversity (the EEB concept of opposing forces), on the assumption that fitness differences

Table 1
Variables defined for this study.

Variable	Definition
$d(i,j)$	Distance function used to monitor diversity between i and j
$D(F)$	Diversity of the fitness distribution
i,j	Individual number $\in \{1,\ldots,N\}$
f_i	Fitness of individual i
f_{best}	Fitness of the best individual in the population
f_{avg}	Average fitness of the population
f_{worst}	Fitness of the worst individual in the population
$\hat{\sigma}_f$	Standard deviation of the fitness distribution
F	Fitness distribution of a population
m	Interval number
M	Total number of intervals
N	Population size
p_m	Fraction of N that belongs to interval m
P	Population
$U(F)$	Fitness distribution F based on a uniform distribution
NMDF	Normalization with maximum diversity so far
VMD	Virtual maximum diversity

reflect genotypic space diversity [12]. This is a limitation, however, and few researchers using this approach take it into account [13]-[16]. The following scenario illustrates the problem: a population of N individuals located on N different peaks of the same magnitude would be considered to be in a state of convergence from the phenotypic point of view, whereas from a genotypic perspective, the diversity would be clearly visible. Therefore, in the presence of an unknown landscape structure, relying solely on phenotypic measurement could be misleading in the search performance analysis.

Phenotypic formulations have frequently been involved in the heuristic formulations used to adapt EA parameters to control the EEB. However, modifying the EEB during a search considerably increases process complexity. Burke, Gustafson, and Kendall [17] summarized the problem as follows: "*The type and amount of diversity required at different evolutionary times remains rather unclear*."

The first objective of the paper having been met with the above detailed discussion on phenotypic measures, we present below nine formulations retrieved from the literature, as well as some variants adapted to the present context. This results in a total of 19 different PCMs.

3.1. General concept

In this study, minimization problems are considered, which requires the adaptation of some PCM formulations. Table 1 summarizes the variables used in this study.

Phenotypic formulations can be evaluated from two perspectives: (1) based on distance-based measurement, where the measurements are estimated by the best fitness (f_{best}), the average fitness (f_{avg}), the worst fitness (f_{worst}), or the standard deviation of the fitness distribution ($\hat{\sigma}_f$) (evaluations based on the distances between solution responses are also possible, and may be evaluated according to a particular descriptive fitness (f_{best}, f_{avg}), or between each individual fitness measure); or (2) by scanning the fitness frequency of a population. However, two points have restricted the latter perspective from being adopted for phenotypic formulation evaluation. First, since the fitness distribution is generally continuous, the fitness space has to be partitioned. Also, the maximum range of the fitness values is unknown, unless the search space is completely enumerated, and so the partitioning process needs to be adaptive, to account for the extension of the fitness range.

3.2. Normalization

In this study, all PCMs are limited to unitary ranges. Full exploitation in the EEB framework is associated with a PCM value of 1, as the phenotypic convergence state is achieved, while a 0 value represents the maximum phenotypic diversity state. Aside from the advantage of normalized measures for comparison purposes, this makes the PCM equal to 1–PDM.

Some existing PCMs are normalized in their original formulation, while others rely on the maximum diversity obtained so far in the optimization process (*NMDF*). However, this normalization is not suitable in a phenotypic context, as it could distort the measurement. In fact, *NMDF* assumes that the starting population is drawn from a uniform distribution. Since the fitness distribution is a function of the landscape relief, this assumption cannot be made. Consequently, this approach would then consider the initial phenotypic distribution as the most diversified state, regardless of its real level. We therefore propose the virtual maximum diversity (*VMD*) as a normalization alternative. For a given population size, this approach considers that the most diverse population state is achieved when the fitness distribution is uniformly distributed between the worst and the best values obtained up to that point in the optimization process. This means that the diversity is computed for a virtual population in which the fitness of the individuals is uniformly distributed over the absolute fitness range respecting a predefined distance ($|f_{worst} - f_{best}|/(N-1)$). Since *VMD* is established for N individuals, it must be recalculated when the population size and/or the absolute fitness range are modified.

For fitness frequency measurements, the maximum diversity value is obtained when the fitness distribution is partitioned uniformly over the total number of intervals (M). The maximum value is calculated by setting $p_m = 1/M$ in the formulation. However, for $N < M$, the maximum value is achieved when $p_m = 1/N$.

Finally, the presence of phenotypic outliers could lead to an overestimation of the convergence state, due to the widening of the absolute fitness range. However, in real-world problems, identifying phenotypic outliers is difficult, since they can represent unvisited regions, instead of a single extreme value.

3.3. PCM formulation

Table 2 presents the PCMs considered in our comparison, some of which were developed specifically for phenotypic distribution, while others were proposed in multivariate distribution contexts and so are reformulated here. The latter are marked with an asterisk in the reference column of the table.

PCM_1 and PCM_2 are simple ratio indicators, whereas PCM_3 could be considered as an extreme ratio. Lee and Takagi [18] used PCM_1, PCM_2, and the change in best fitness as inputs of fuzzy logic controllers for adapting GA parameters. Subbu et al. [19] later proposed a similar adaptation scheme, in which they promote PCM_1 and a GDM based on the Hamming distance as inputs. Herrera and Lozano [6] also used PCM_1 and a GDM based on Euclidian distance as inputs to their fuzzy logic controller. Finally, Vasconcelos et al. [20] and Pellerin et al. [21] promoted the use of a PCM with the same meaning as PCM_1 to adapt GA parameters following heuristic rules.

PCM_4 represents a family of PCMs based on the difference between the average and the best fitness. This difference could serve as a phenotypic convergence detector [22]. $PCM_{4.1}$ is normalized by the fitness range [23], while $PCM_{4.2}$ is the absolute version of $PCM_{4.1}$ proposed by Neri et al. [24,25] to adapt parameters and activate local searchers with heuristic rules in a memetic algorithm (MA) context. Caponio et al. [26] proposed $PCM_{4.3}$, which is an *NMDF* normalized version. They use this indicator with a hybrid algorithm to detect super-fit individuals, and thus activate different local searchers following heuristic rules. $PCM_{4.4}$ was proposed by Caponio et al. [14]. Again, it is used to adapt EA parameters and activate local search procedures following heuristic rules. $PCM_{4.4}$ was later used for other applications with similar adaptation rules [13,27,28]. $PCM_{4.5}$ is the *VMD* normalized version proposed here.

Table 2
PCM formulations used for the comparative study.

No.	PCM formulation	Ref.	No.	PCM formulation	Ref.
1.	$PCM_1 = \frac{f_{best}}{f_{avg}}$	[18]	13.	$PCM_6 = 1 - \frac{(1/N)\sum_{i=1}^{N} \lvert f_i - f_{avg}\rvert}{VMD}$	[31]*, [32]*
2.	$PCM_2 = \frac{f_{avg}}{f_{worst}}$	[18]	14.	$PCM_7 = 1 - \frac{\sum_{i=1}^{N} (f_i - f_{avg})^2}{VMD}$	[33]*
3.	$PCM_3 = \frac{f_{best}}{f_{worst}}$				
4.	$PCM_{4.1} = 1 - \frac{f_{avg} - f_{best}}{f_{worst} - f_{best}}$		15.	$PCM_8 = 1 - \frac{2/(N(N-1))\sum_{i=2}^{N}\sum_{j=1}^{i-1} \lvert f_i - f_j\rvert}{VMD}$	[34]*
5.	$PCM_{4.2} = 1 - \left\lvert \frac{f_{avg} - f_{best}}{f_{worst} - f_{best}} \right\rvert$	[24]	16.	$PCM_9 = 1 + \frac{1}{\log(u)} \sum_{m=1}^{M} p_m \log(p_m)$	[36]
6.	$PCM_{4.3} = 1 - \frac{\lvert f_{avg} - f_{best}\rvert}{NMDF}$	[26]	17.	$PCM_{10} = 1 - \frac{1 - \sum_{m=1}^{M} p_m^{\alpha}}{1 - u^{1-\alpha}}$	[39]
7.	$PCM_{4.4} = 1 - \min\left\{ \left\lvert \frac{f_{avg} - f_{best}}{f_{best}} \right\rvert, 1 \right\}$	[14]	18.	$PCM_{11} = 1 - \frac{\log\left(\sum_{m=1}^{M} p_m^{\alpha}\right)}{(1-\alpha)\log(u)}$	[40]
8.	$PCM_{4.5} = 1 - \frac{\lvert f_{avg} - f_{best}\rvert}{VMD}$		19.	$PCM_{12} = 1 - \beta \cdot \sum_{m=1}^{M} p_m(1 - p_m)$	[7]*
9.	$PCM_{5.1} = 1 - \frac{\hat{\sigma}_f}{\lvert f_{worst} - f_{best}\rvert}$	[15]		where,	
10.	$PCM_{5.2} = 1 - \frac{\hat{\sigma}_f}{\sqrt{((f_{worst} - f_{avg})^2 + (f_{best} - f_{avg})^2)/2}}$	[29]		$\beta = \begin{cases} \frac{M}{(M-1) - (r(M-r)/N^2)} & \text{if } M < N \\ \frac{N}{N-1} & \text{otherwise} \end{cases}$	
11.	$PCM_{5.3} = 1 - \min\left\{ \frac{\hat{\sigma}_f}{\lvert f_{avg}\rvert}, 1 \right\}$	[30]			
12.	$PCM_{5.4} = 1 - \frac{\hat{\sigma}_f}{VMD}$				

The PCM_5 family is based on the standard deviation, or dispersion, of the fitness values. We consider the unbiased standard deviation in this study. $PCM_{5.1}$ was proposed by Tirronen and Neri [15] to adapt differential evolution parameters following heuristic rules. $PCM_{5.2}$ is known as the degree of dispersion, and was proposed by Miao et al. [29] to adapt a particle swarm optimizer parameter. $PCM_{5.3}$ was promoted by Tirronen et al. [30] to activate local searchers in MA following heuristic rules. In this PCM formulation, as for $PCM_{4.4}$, the minimum operator suggests normalization issues, since the unitary range is not guaranteed. $PCM_{5.4}$ is the *VMD* normalized version proposed in this paper.

PCM_6 to PCM_8 are reformulations of multivariate diversity measurements [31–34]. In the phenotypic context, PCM_6 describes the mean location of the fitness values with respect to the average fitness of the distribution. PCM_7 is based on the underlying idea of allocating more importance to fitness values away from the mean of the distribution. Finally, PCM_8 corresponds to the mean pairwise distance from all fitness values. The idea behind this PCM was used by Hutter and Legg [35] to motivate the development of the fitness uniform selection scheme (FUSS).

PCM_9 to PCM_{12} belong to the fitness frequency category. This category involves the entropy concept which, at first sight, could be well suited to being a phenotypic descriptor, since it describes the level of disorder of a distribution. PCM_9 represents the Shannon entropy [36]. Rosca [37] uses this formulation to correlate genetic programming (GP) statistical measures to the phenotypic state with the aim of controlling the EEB, whereas Darwen [38] uses it to compare problem-specific learning strategies involved in a GA optimizer. PCM_{10} and PCM_{11} are two other entropy families ($\alpha > 0$ and $\alpha \neq 1$) [39,40]. By letting $\alpha \to 1$, PCM_{10} and PCM_{11} tend toward PCM_9. In contrast, PCM_{12} is an approximation of PCM_9 [7]. The variable u, shared by PCM_9 to PCM_{11}, stands for the normalization part, as $u = \min\{M, N\}$ (Section 3.2). For PCM_{12}, there is a similar normalization. Nevertheless, in the original formulation, a correction term ($r = N \bmod M$) was considered for cases where M is not a common divisor of N.

As we have shown, most of the phenotypic indicators formulated in the literature have been used alone to describe the population's EEB state, and no performance analysis was conducted to assess the suitability of these various formulations.

4. Validation of phenotypic formulations

Validation of the phenotypic formulation selected is mandatory, in order to ensure the reliability of the EEB diagnostic tool. Since no framework is available in the literature, three diversity requirements are proposed to determine the relevance of phenotypic formulations. These requirements are validated by means of a deterministic frozen diversity case framework, which is a simple framework that can represent them efficiently. In order to avoid potential issues arising from normalization approaches, phenotypic formulations are considered here solely by studying their characteristics at the family level, which reflect their computed diversity.

4.1. Requirements for a suitable diversity measure

In pioneering research, Weitzman [41] listed 14 principal characteristics of reliable diversity measures. Weitzman acknowledged that these properties are not equally important. Later, Solow and Polasky [42] identified three of them as natural requirements:

1. *Monotonicity in species*: adding a species (or individuals, in the current context) should not decrease diversity or $D(P') \leq D(P)$, if P' is a subset of population P.
2. *Twinning:* the addition of an individual or a species already in the population should not increase the diversity or $D(P \cup i) = D(P)$, if $d(i, j) = 0$, where $j \in P$ and $i \notin P$.
3. *Monotonicity in distance:* an unambiguous increase in distance between individuals should be reflected in the measurement or $D(P') \leq D(P)$, if $d(i', j') \leq d(i, j)$.

The ideas governing these requirements apply to phenotypic measurement. In reality, the diversity measurement should be understood as a description of the coverage of the search space. This concept is completely and rigorously expressed by those diversity

Table 3
Requirements of the phenotypic formulation.

#	Requirement	Brief description
1	Monotonicity in fitness varieties	Adding a fitness value should not decrease diversity Uniformly distributed fitness provides upper bound diversity
2	Twinning	Duplicate fitness values should reduce diversity, as the distribution moves away from the uniform distribution case (diversity upper bound)
3	Monotonicity in distance or shuffling dependence	Diversity should decrease as fitness values move closer together

requirements. Nevertheless, the three requirements must first be reformulated in terms of fitness distribution.

Species monotonicity will be referred to here as *monotonicity in fitness varieties*. This first quality specifies that diversity will increase with the addition of new fitness values. This implies that the maximum phenotypic diversity is produced by a uniform distribution ($U(F)$) over the fitness range. Therefore, the mathematical formulation is: $D(F') \leq D(F) \leq D(U(F))$, where F' is a subset of the fitness distribution F.

The initial definition of the *twinning* requirement is directly transferable to the present context. However, for fixed population sizes, the presence of duplicate individuals inevitably reduces the diversity of a population. The mathematical form then becomes $D((F \setminus f_k) \cup f_i) < D(F)$, if $d(f_i, f_j) = 0$, where $f_j \in F, f_i \notin F$. Here, f_k is a non-duplicated individual removed from the population F.

The *monotonicity in distance* requirement also corresponds to the *shuffling dependence property* [43]. This requirement states that permutation of fitness values impacts the phenotypic measurement directly. In this context, the mathematical formulation becomes: $D(F') \leq D(F)$, if $d(f_i', f_j') \leq d(f_i, f_j)$.

Table 3 lists and describes the final phenotypic formulation requirements, which will be shown to be sufficient for evaluating their relevance.

4.2. Validation framework for the requirements analysis

Six deterministic cases of frozen fitness diversity are proposed to evaluate and illustrate the phenotypic formulation responses, as follows:

Case 1: All the individuals are located at f_{best}.
Case 2: 50% of the population is located at the mid-point between f_{worst} and f_{best}, while the remaining portion is at f_{best}.
Case 3: $N-1$ of the population is located at the mid-point between f_{worst} and f_{best}, while the remaining individual is at f_{best}.
Case 4: 50% of the population is located at f_{worst}, while the remaining portion is at f_{best}.
Case 5: $N-1$ of the population is located at f_{worst}, while the remaining individual is at f_{best}.
Case 6: The individuals are uniformly distributed over a predefined fitness range (*VMD* case).

The first case corresponds to a converged situation. Cases 2 and 3, and Cases 4 and 5 present equivalent phenotypic diversities. However, Cases 4 and 5 present higher diversities than Cases 2 and 3. Furthermore, Cases 2–5 have low phenotypic diversity (two fitness values). In contrast, Case 6 corresponds to the highest phenotypic diversity state.

During the tests, a population size (N) of 100 and a total number of intervals (M) of 100 are used. In addition, the fitness range is defined between 150 for f_{worst} and 50 for f_{best}.

4.3. Relevance of the phenotypic formulations

Table 4 presents the diversity levels obtained from each phenotypic family in Table 2 and applied to the validation framework.

Results indicate that all the phenotypic formulations identify the converged distribution (Case 1 = 0). However, none of them conforms to the diversity requirements.

In fact, families 1–8 violate the monotonicity in fitness varieties; Case 6 is not identified as the highest diversity level. All the descriptors found Case 4 or Case 5 to represent the highest diversity condition, even though they each involve only two fitness values.

In addition, all but the third phenotypic family violate the twinning requirement; their state evaluations are dependent on the number of individuals located at a given fitness position (Case 2 $\neq$ Case 3, and Case 4 $\neq$ Case 5). In fact, the third family is able to fulfill this requirement only because its extreme ratio takes advantage of the proposed fitness value distribution.

Finally, none of the fitness frequency families (9–12) is capable of adequately describing the monotonicity in distance requirement. This is because they all show identical phenotypic measurement, since they do not take into account the location of the intervals over the fitness distribution [9] in cases where the individuals are different distances apart (Cases 2 $\not\prec$ 4, or Cases 3 $\not\prec$ 5).

Table 5 summarizes the behavior of the phenotypic families over the diversity requirements identified.

5. New phenotypic formulation proposed

To control the EEB within the orthogonal framework, a reliable PCM is required. The previous section revealed that no phenotypic formulation available in the literature offers a perfect description of the scattering of the fitness distribution. Therefore, our aim in this section is to present a new formulation that meets the requirements listed in Table 3.

This new formulation is based on multiplication of the phenotypic value differences established between neighbors. Once the fitness distribution has been sorted, the computation can start from any side of the sorted distribution. The formulation ensures that the state of maximum phenotypic diversity occurs when all the individuals are uniformly spread out within the fitness range, which leads to the *VMD* case and fulfillment of the monotonicity in fitness varieties requirement.

To demonstrate, Fig. 2(a) depicts the diversity level of a phenotypic distribution with three individuals located within a 10-unit fitness range. One individual is located at $f_{best}(0)$, another at $f_{worst}(10)$, and the third between the boundaries of this range. This example shows that the behavior of the proposal is generally good. The maximum diversity state appears when the third individual is located at 5, and the performance deteriorates as the third individual approaches the fitness range boundaries. This response degradation comes about as a result of the multiplication effect, and so it appears when the third individual is located closer than 1 unit from any other individual. Furthermore, the multiplication of the fitness difference between neighbors could rapidly lead to very high numbers, as the population and fitness range increase. The multiplication is therefore replaced by the addition of the logarithms of the neighbor differences. Moreover, the addition of 1 in the logarithmic operator automatically eliminates duplicate fitness values, which ensures that the twinning requirement is met.

Table 4
Behavior of phenotypic formulations over the six frozen case framework.

Family	Diversity formulation	Phenotypic distribution					
		Case 1 N_{best}	Case 2 $50_{middle}/50_{best}$	Case 3 $(N-1)_{middle}/1_{best}$	Case 4 $50_{worst}/50_{best}$	Case 5 $(N-1)_{worst}/1_{best}$	Case 6 VMD
1	$1-(f_{best}/f_{avg})$	0.00	0.33	0.50	0.50	0.66	0.50
2	$1-(f_{avg}/f_{worst})$	0.00	0.25	0.005	0.33	0.007	0.33
3	$1-(f_{best}/f_{worst})$	0.00	0.50	0.50	0.67	0.67	0.67
4	$\lvert f_{avg}-f_{best}\rvert$	0.00	25.00	49.50	50.00	99.00	50.00
5	$\hat{\sigma}_f$	0.00	25.13	5.00	50.25	10.00	29.30
6	$\frac{1}{N}\sum_{i=1}^{N}\lvert f_i-f_{avg}\rvert$	0.00	25.00	0.99	50.00	1.98	25.25
7	$\sum_{i=1}^{N}(f_i-f_{avg})^2$	0.00	6.25E+04	2.48E+03	2.50E+05	9.90E+03	8.50E+04
8	$\frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\lvert f_i-f_j\rvert$	0.00	25.25	1.00	50.51	2.00	34.01
9	$-\sum_{m=1}^{M}p_m\log(p_m)$	0.00	0.69	0.06	0.69	0.06	4.59
10	$\frac{1}{\alpha-1}\left(1-\sum_{m=1}^{M}p_m^{\alpha}\right)$	0.00	0.96	0.70	0.96	0.70	68.34
11	$\frac{1}{1-\alpha}\log\left(\sum_{m=1}^{M}p_m^{\alpha}\right)$	0.00	0.69	0.54	0.69	0.54	4.59
12	$\frac{N^2}{2}\sum_{m=1}^{M}p_m(1-p_m)$	0.00	2500.00	99.00	2500.00	99.00	4949.00

Table 5
Summary of the diversity requirement fulfillment by the phenotypic formulations.

Family	Diversity formulation	Requirement (cases):	#1 (6 > all others)	#2 (2 = 3 and 4 = 5)	#3 (4 > 2 and 5 > 3)
1	$1-(f_{best}/f_{avg})$		No	No	Yes
2	$1-(f_{avg}/f_{worst})$		No	No	Yes
3	$1-(f_{best}/f_{worst})$		No	(Yes)	Yes
4	$\lvert f_{avg}-f_{best}\rvert$		No	No	Yes
5	$\hat{\sigma}_f$		No	No	Yes
6	$\frac{1}{N}\sum_{i=1}^{N}\lvert f_i-f_{avg}\rvert$		No	No	Yes
7	$\sum_{i=1}^{N}(f_i-f_{avg})^2$		No	No	Yes
8	$\frac{2}{N(N-1)}\sum_{i=2}^{N}\sum_{j=1}^{i-1}\lvert f_i-f_j\rvert$		No	No	Yes
9	$-\sum_{m=1}^{M}p_m\log(p_m)$		Yes	No	No
10	$\frac{1}{\alpha-1}\left(1-\sum_{m=1}^{M}p_m^{\alpha}\right)$		Yes	No	No
11	$\frac{1}{1-\alpha}\log\left(\sum_{m=1}^{M}p_m^{\alpha}\right)$		Yes	No	No
12	$\frac{N^2}{2}\sum_{m=1}^{M}p_m(1-p_m)$		Yes	No	No

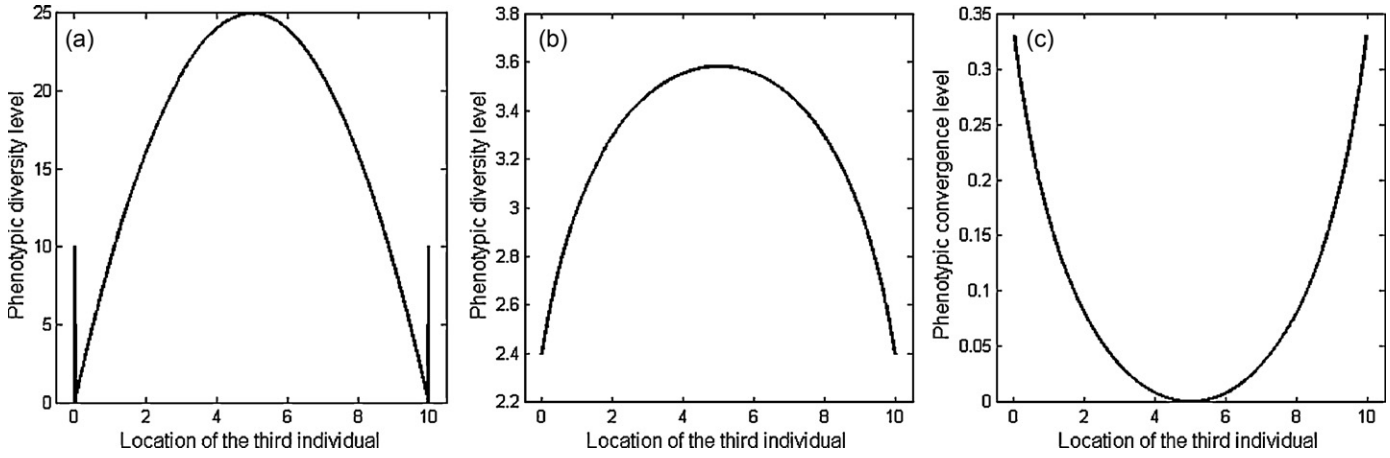

Fig. 2. Phenotypic diversity level of the new proposal registered over a population of three individuals. (a) Multiplicative formulation. (b) Logarithmic formulation. (c) Phenotypic convergence level of PCM_{13}.

Equation 1 gives the descriptor formulation. The previous example is repeated in Fig. 2(b) with this enhanced formulation.

$$\sum_{i=1}^{N-1} \ln(1 + |f_i - f_{i+1}|) \tag{1}$$

Finally, (1) can be rewritten as a PCM formulation, as follows:

$$PCM_{13} = 1 - \frac{\sum_{i=1}^{N-1} \ln(1 + |f_i - f_{i+1}|)}{VMD} \tag{2}$$

Fig. 2(c) presents the behavior of PCM_{13} over the previous example. It confirms that the lowest convergence state (or highest diversity state) is achieved when the population is uniformly distributed (with the third individual located at 5). It can also be observed that the maximum achieved convergence level is 0.33, since at least two individuals are always differentiated by the maximum distance allowed from the fitness range.

Analysis of the new phenotypic formulation over the diversity requirements

The formulation can be evaluated by means of the validation framework introduced in the previous section, the results of which are provided in Table 6. As with the preceding descriptors, the new proposal in its non-normalized version (1) detects the converged fitness distribution (Case 1 = 0). Moreover, Table 6 reveals that the new formulation conforms to the three diversity requirements: Monotonicity in fitness varieties is respected, as Case 6 presents the highest diversity level. Twinning is followed, since a different distribution of fitness values has no impact on the diversity level (Case 2 = Case 3, and Case 4 = Case 5). Finally, the monotonicity in distance requirement is met, as the distance between individuals is accurately taken into account (Case 2 < Case 4, and Case 3 < Case 5).

6. Analysis of PCMs over specifically designed landscapes

Now that the new phenotypic formulation has been proved to perform in accordance with the diversity requirements, this section examines its behavior over the course of a search process. PCM_1 to PCM_{12} are also included in the investigation. However, this analysis requires that the phenotypic state be known quantitatively, or at least qualitatively, throughout the optimization process. This would become a serious issue if the search were based on an EA, since the sampled fitness distribution depends on the search path followed, which is a stochastic process. The result would be to hide the phenotypic distribution structure of a chosen benchmark. Furthermore, replications of the simulations, which are essential for validating the reliability of a PCM, would be useless.

In order to circumvent this problem, a generic benchmark is proposed to ensure uniform fitness distribution sampling, as well as control of the phenotypic states by the landscape definition and the search dynamic. Furthermore, with this benchmark, no genetic operator is involved in the evolution of the population. Instead, at each iteration, a new fitness distribution is sampled over the landscape. Phenotypic convergence is simulated by reducing the sampling boundary as the process evolves. Consequently, the process begins in a full phenotypic diversity state (PCM = 0) and proceeds to a convergence state (PCM = 1) following a predefined schedule. As a result, the phenotypic distribution is known throughout the evolution process.

We propose three landscapes here; a linear landscape, a double-slope landscape, and a saw tooth landscape, as depicted in Fig. 3. The analysis is conducted with a population size (N) of 100, while an interval number (M) of 100 is assigned and applied for PCM_9 to PCM_{12} inclusive. All the results are averaged over 50 repetitions.

6.1. Linear function

6.1.1. Landscape definition

The first landscape includes a linear function (Fig. 3(a)). A good *PCM* has to reflect the intended linear convergence pattern. The fitness function is given by:

$$f(x) = -x + b \tag{3}$$

Table 6
Diversity levels of the new proposal over the validation framework.

Family	Diversity formula	Phenotypic distribution					
		Case 1 N_{best}	Case 2 $50_{middle}/50_{best}$	Case 3 $(N-1)_{\text{middle}}/1_{\text{best}}$	Case 4 $50_{worst}/50_{best}$	Case 5 $(N-1)_{worst}/1_{best}$	Case 6 VMD
13	$\sum_{i=1}^{N-1} \ln(1 + \|f_i - f_{i+1}\|)$	0.00	3.93	3.93	4.62	4.62	69.12

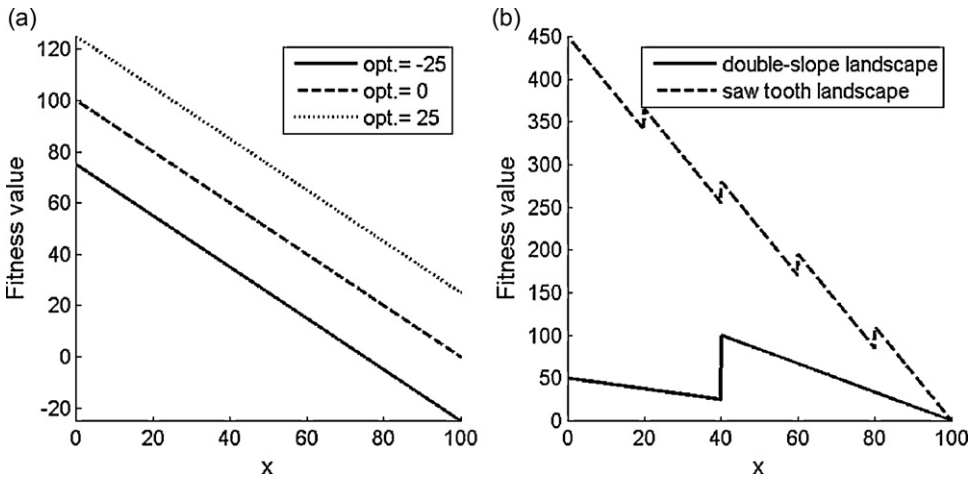

Fig. 3. Fitness functions of the generic benchmark. (a) Three translating cases of the linear landscape. (b) Double-slope landscape and the saw tooth landscape.

The variable x denotes the genotypic position of the individuals, and ranges from 0 to 100. Since the study considers only the fitness of the individuals, a univariate genotype is sufficient. Three translations of (3) are considered: b equal to {75, 100, 125}. These cases allow the assessment of PCM coherence. Indeed, an accurate PCM should provide a constant evaluation, regardless of the fitness value sign. This is important, as the fitness range is often unknown for real-world problems.

The population is uniformly generated within the genotypic range. At each iteration, the convergence of the lower genotypic boundary is increased by 2% toward the optimum, and the process goes from PCM = 0 to PCM = 1 in 51 iterations. In addition, in order to ensure that f_{best} always represents the optimum value $f(x^* = 100)$, an elite individual is inserted at the optimum position. This landscape simulates the dynamics of a search process over a unimodal landscape.

6.1.2. Behavioral results of the PCMs

As demonstrated by the characteristic response curves in Fig. 4, PCM_1 to PCM_3, $PCM_{4.4}$, and $PCM_{5.3}$ appear to be unreliable. The first three PCMs fall outside the unitary range for the negative optimum case. However, $PCM_{4.1}$, $PCM_{4.2}$, $PCM_{5.1}$, and $PCM_{5.2}$ (Fig. 5(a)) are unable to describe the convergence progression, as their value remains constant throughout the entire process. This behavior is generated by the numerator and the denominator decreasing at the same rate over the process.

The fitness frequency measures PCM_9 to PCM_{12} present a similar trend, and do not monitor phenotypic diversity well, as demonstrated in Figs. 5(b)–6(a). Since the descriptors do not meet the monotonicity in distance requirement, their convergence values remain quite low for a significant part of the process. PCM_9 and PCM_{12} are not explicitly presented, as their behaviors are similar to those of PCM_{10} ($\alpha = 1.1$) and PCM_{10} ($\alpha = 2.0$) respectively.

Fig. 6(b) presents the evolution of the state of convergence of PCM_{13}. The curves reveal good coherence and show a generally good trend. Nevertheless, the linear pattern is not perfectly represented, as the process does not start in a state of full diversity. This may be a result of the sampling error of the population. We investigate this in a section below, as part of a discussion on a sensitivity analysis procedure. The phenotypic state estimation error at the beginning of the process could also be linked to the random number generator (RNG) imprecision.

Finally, the remaining PCMs ($PCM_{4.3}$, $PCM_{4.5}$, $PCM_{5.4}$, and PCM_6 to PCM_8), although they are not included in a figure, provide excellent descriptions of the linear function convergence pattern.

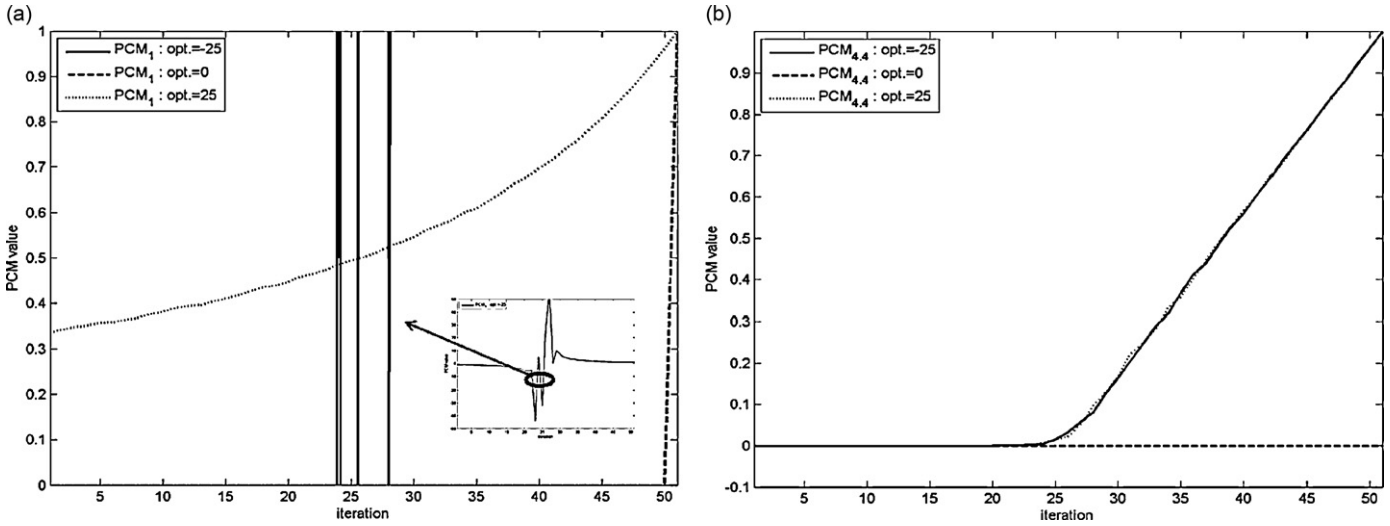

Fig. 4. PCM behavior observed over the linear landscape. (a) PCM_1. (b) $PCM_{4.4}$.

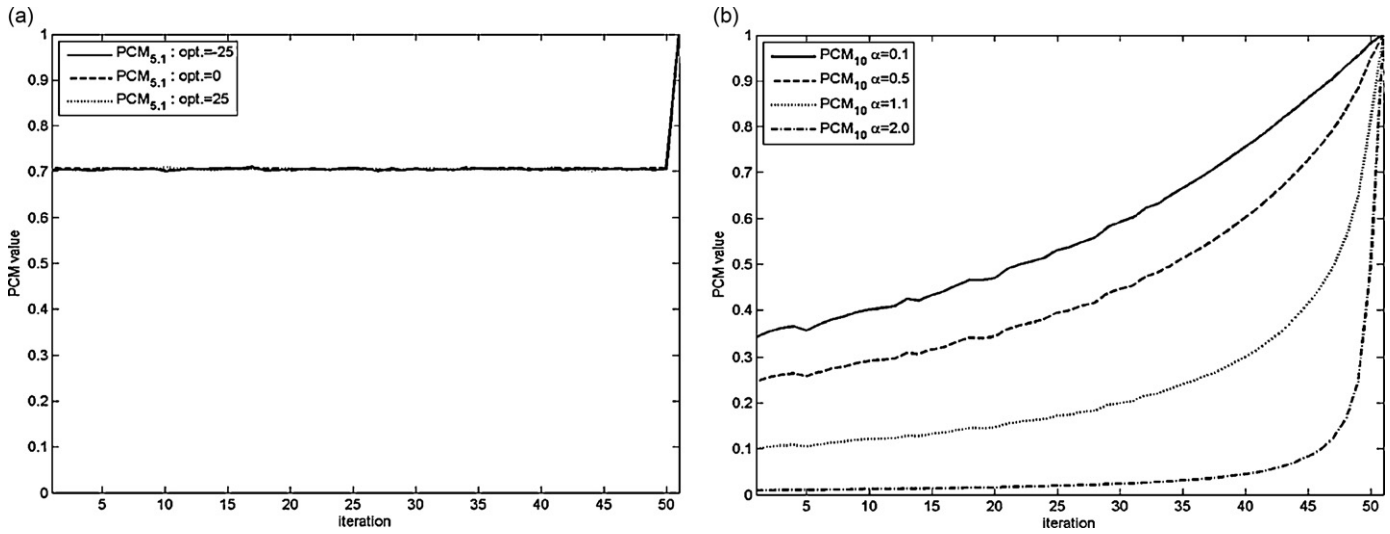

Fig. 5. PCM behavior observed over the linear landscape. (a) $PCM_{5.1}$. (b) PCM_{10}.

6.2. Double-slope landscape

6.2.1. Landscape definition

The second landscape (Fig. 3(b)) is proposed in order to study the impact of an increasing fitness range. The fitness function is given by:

$$f(x) = \begin{cases} -25/40x + 50 & \text{if } x \leq 40 \\ -5/3x + 166.67 & \text{otherwise} \end{cases}. \quad (4)$$

Again, the genotypic position (x) ranges from 0 to 100. An elite individual, located at $f(x^* = 100) = 0$, is kept in the population, and the remaining individuals are uniformly generated within the genotypic boundaries.

The following convergence schedule is adopted for the analysis: during the first 20 iterations, individuals (except the elite one) are located only on the first slope ($0 \leq x \leq 40$). Thereafter, they are located solely on the second slope ($40 < x \leq 100$) (31 remaining iterations). The lower genotypic bound is brought closer to the global optimum location by 2% of the total range at each iteration. Therefore, the jump to the second slope is implicitly controlled by the lower boundary, while the upper genotypic bound is constrained as previously defined.

Intuitively, reliable PCMs should reflect the two linear patterns. Locating the elite individual at the global optimum ensures that the first convergence pattern does not cover the total fitness range. So, the first pattern should start with a phenotypic convergence level slightly below 0.5, due to the fact that half the fitness range is covered (fitness $\in$ [25,50]) with the presence of an elite individual at the global optimum. The same pattern should end with a phenotypic convergence at around 1, since there are only two fitness values in the population at iteration 20. The second pattern should start in a full phenotypic diversity state, as the fitness range grows and the population is distributed over the entire area. This pattern should end with a PCM value of 1, due to the full convergence of the population at the global optimum position ($x^* = 100$).

The double-slope landscape emulates the dynamics of a search process, which might follow a tuned restart and move the population into a second region.

6.2.2. Behavioral results of the PCMs

Figs. 7 and 8 draw the responses over the double-slope landscape of the PCMs that presented a good description of the linear landscape ($PCM_{4.3}$, $PCM_{4.5}$, $PCM_{5.4}$, PCM_6 to PCM_8, PCM_{13}).

The curves in Fig. 7(a) reveal incorrect descriptions resulting from the *NMDF* normalization used by $PCM_{4.3}$, where the first

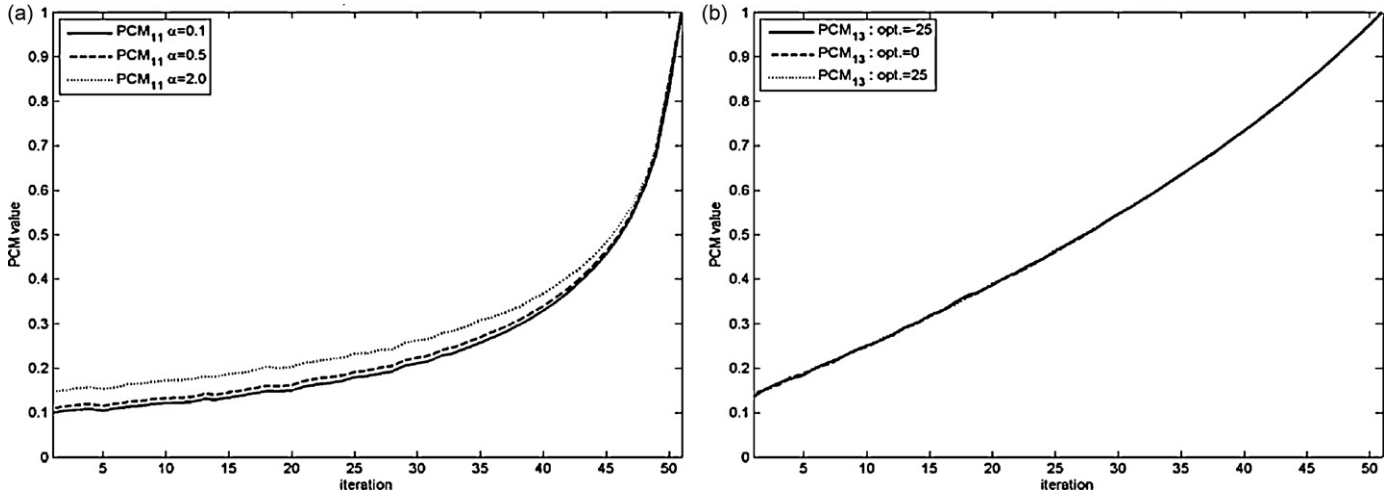

Fig. 6. PCM behavior observed over the linear landscape. (a) PCM_{11}. (b) PCM_{13}.

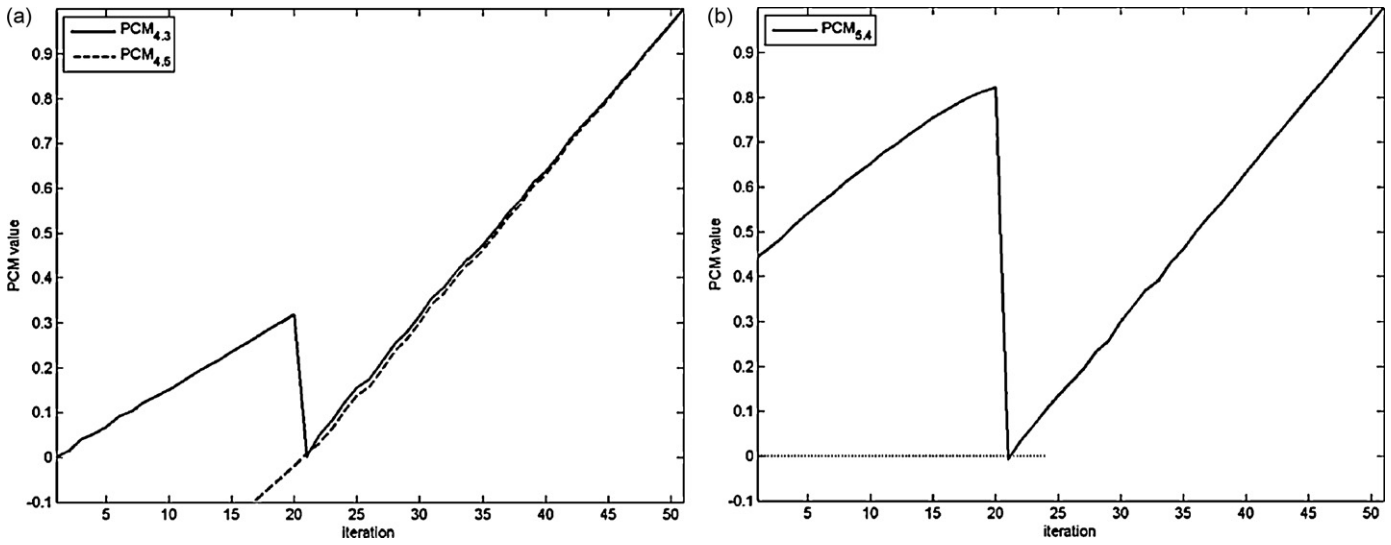

Fig. 7. PCM behavior observed over the double-slope landscape. (a) $PCM_{4.3}$ and $PCM_{4.5}$. (b) $PCM_{5.4}$.

iteration is assumed to offer the highest diversities. Moreover, the first convergence pattern ends far from the convergence state. $PCM_{4.5}$ also appears to be imprecise, as it does not demonstrate the two intended patterns. More fundamentally, the response of $PCM_{4.5}$ does not remain in the unitary range. Since $PCM_{4.5}$ is based on *VMD* normalization, the negative diversity estimations indicate that $|f_{avg} - f_{best}|$ (the PCM_4 family) violates the requirement of monotonicity in fitness varieties.

$PCM_{5.4}$ (Fig. 7(b)) shows a relatively good convergence pattern. Nevertheless, when the fitness range is increased (iteration 21), $PCM_{5.4}$ generates negative values. Even though the error remains small, the underlying problem is similar to the one described for $PCM_{4.5}$. Therefore, phenotypic formulations based on standard deviation (the PCM_5 family) are not recommended, since they contravene the monotonicity in fitness varieties requirement. The same conclusion applies to PCM_6 to PCM_8 (Fig. 8(a)).

PCM_{13} shows a good trend over the two patterns (Fig. 8(b)). Nonetheless, the descriptor cannot perfectly predict a null convergence state when the fitness range is increased (iteration 21). Again, the deviation can be attributed to the population sampling error and RNG inaccuracy.

6.3. Saw tooth landscape

6.3.1. Landscape definition

The third landscape (Fig. 3(b)) reproduces a multimodal fitness distribution. The fitness function is given by:

$$f(x) = \begin{cases} (-11/2)x + 450 & \text{if } x \leq 20 \\ (-11/2)x + 475 & \text{else if } x \leq 40 \\ (-11/2)x + 500 & \text{else if } x \leq 60 \\ (-11/2)x + 525 & \text{else if } x \leq 80 \\ (-11/2)x + 550 & \text{otherwise} \end{cases} \quad (5)$$

Once again, an elite individual is located at the global optimum, and the genotypic position (x) ranges from 0 to 100.

The population is divided into five equivalent groups, and each group is located over a different tooth. This means that the genotypic boundaries are relative to a group. Moreover, the process is divided into five phases, each with 11 iterations. The dynamics of the landscape demands that every group converge toward its local tooth optimum. After this first convergence, the phase is considered completed, and the groups jump to the next tooth, where the

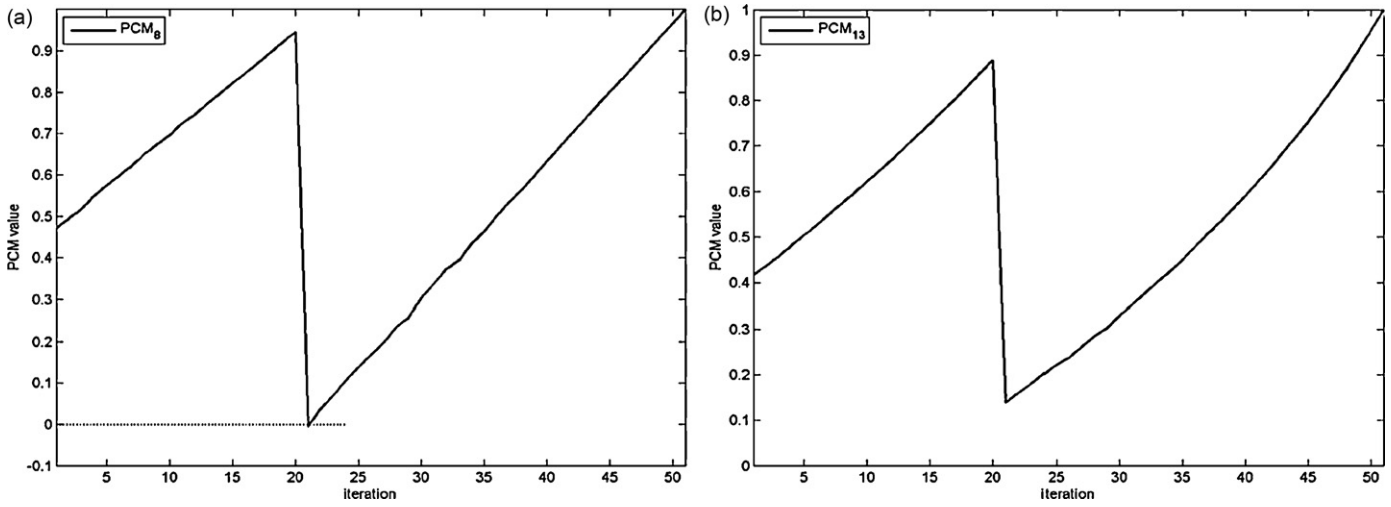

Fig. 8. PCM behavior observed over the double-slope landscape. (a) PCM_8. (b) PCM_{13}.

Table 7
Convergence schedules for Group 1 (G1) to Group 5 (G5) over the saw tooth landscape.

Phase	Iterations	Saw tooth number					
		1 $x \in [0\ 20]$	2 $x \in [2040]$	3 $x \in [4060]$	4 $x \in [6080]$	5 $x \in [80100]$	Frozen $x = 100$
Phase 1	1–11	G1	G2	G3	G4	G5	
Phase 2	12–22		G1	G2	G3	G4	G5
Phase 3	23–33			G1	G2	G3	G4 to G5
Phase 4	34–44				G1	G2	G3 to G5
Phase 5	45–55					G1	G2 to G5

process is repeated. After completion of the last tooth, the groups remain frozen at the global optimum position ($f_{best} = f(x^* = 100) = 0$) until the end of the process. This structure is presented in Table 7.

To summarize, at the end of phase 1 (iteration 11), the fitness distribution is evaluated over five modes. At the end of phase 2 (iteration 22), the fitness distribution is evaluated over four modes, and so on, up to the end of phase 5 (iteration 55), where all the individuals are located at the global optimum. The evolution of the process proceeds from a 10% increase in the lower genotypic bound at each iteration, while the upper bound corresponds to the tooth's local optimum. The groups are generated following a uniform distribution between their boundaries.

Evaluation of the PCMs should present the following: the first convergence phase should start at around 0 (the complete fitness range is covered); the second phase should start at around 0.2 (the fitness distribution covers 4/5 of the fitness range), and so on. All phases should end near convergence, with a value increasing toward 1 as the process evolves.

The search process simulates an algorithm that clusters its resources or individuals over different regions of the landscape.

6.3.2. Behavioral results of the PCMs

None of the PCMs introduced in Table 2 was able to adequately describe the intended phenotypic convergence pattern of the saw tooth landscape. However, some characteristic behaviors of these PCMs are depicted in Fig. 9. PCM_1, PCM_3, and $PCM_{4.4}$ show a constant state of full diversity throughout the process, due to the influence of f_{best}, which is fixed here at 0. PCM_2 presents a completely misleading trend.

Regarding the PCM_4 family, $PCM_{4.1}$ and $PCM_{4.2}$ show the same behavior, and their formulations (similar to those of PCM_2) are not able to describe the convergence progression within each phase. In contrast, $PCM_{4.3}$ and $PCM_{4.5}$ show a convergence progression, but are unable to describe the intended convergence peak at the end of each phase. $PCM_{4.5}$ starts below 0 (−0.004), as it does for the double-slope landscape.

The PCM_5 family provides very diversified patterns. This family appears to be incapable of characterizing the convergence of each phase, since no converged peak is observable over the five phases. A similar conclusion can be drawn from the results of PCM_6 to PCM_8.

The PCM_9 to PCM_{12} estimations show clear evidence of the convergence peaks. As demonstrated in Fig. 9(b), PCM_{10} with $\alpha = 0.1$ is the best option for estimating diversity at the end of each phase, whereas PCM_9 is best for estimating the diversity at the beginning of each phase.

Finally, PCM_{13} accurately represents the convergence pattern over the five phases (Fig. 9(b)). The converged peaks of each phase are well established. However, as for the two preceding landscapes, the convergence state at the beginning of the first phases is slightly overestimated.

In summary, the proposed landscapes ensure a detailed description of the PCM_1 to PCM_{12} response. PCM13 showed the best overall description of the phenotypic distributions.

7. Assessment of desirable PCM qualities

Following the very good performance demonstrated by PCM_{13} over the validation framework and the proposed landscapes, it is now relevant to establish the quality criteria for a reliable PCM. These qualities are applicable to any PCM formulation.

7.1. Definition of desirable PCM qualities

The three following characteristics are proposed as desirable qualities:

(1) *Reliability*: A PCM should be reliable over similarly scattered phenotypic distributions.

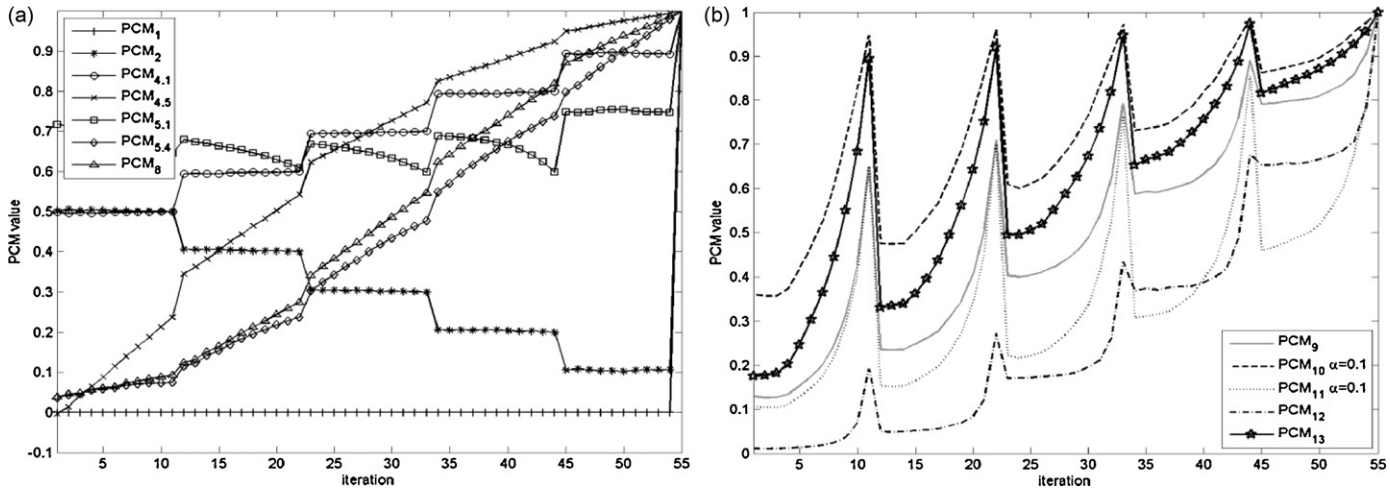

Fig. 9. PCM behavior observed over the saw tooth landscape. (a) PCM_1 to PCM_8. (b) PCM_9 to PCM_{13}.

Table 8
Average range among 96% of the repetition data for PCM_{13} over the three landscape designs.

Landscape	Population size (N)			
	50	100	300	500
Linear	0.060	0.035	0.014	0.008
Double-slope	0.050	0.028	0.010	0.006
Saw tooth	0.058	0.038	0.020	0.013

(2) *Sensitivity*: A PCM should be as insensitive as possible to the simulation parameters.
(3) *Outlier influence*: A PCM should adequately consider the impact of phenotypic outliers.

The performance of PCM_{13} is evaluated in the following section with respect to the desirable qualities. The assessment makes use of the landscapes introduced in the previous section.

7.2. PCM_{13} reliability analysis

The reliability of PCM_{13} is studied through a dispersion analysis, based on 50 repetitions per iteration. In a preliminary step, normality tests were carried out in this study using the Kolmogorov-Smirnov test (0.05 significance level), which indicated that, for each iteration, the 50-repetition sample does not follow a normal distribution. Therefore, to picture PCM reliability, it is preferable to compute the dispersion for 96% of the repetition data. For the sake of clarity, the dispersion values are averaged over the whole process. Table 8 illustrates the stability analysis for four common EA populations ($N \in \{50, 100, 300, 500\}$).

The results clearly indicate that PCM_{13} gives a stable phenotypic state description. On average, for a relatively small population, 96% of all repetition data are stacked with a range smaller than 0.06. Moreover, the analysis reveals that the stability rapidly increases as the population increases, which validates the influence of the sampling error associated with the population size mentioned above.

7.3. PCM_{13} sensitivity analysis

The second experiment constitutes an analysis of the sensitivity of a PCM to the simulation parameters. Considering the PCM_{13} formulation, and since the fitness distribution is univariate, the population size (N) is the only parameter involved. The high reliability of PCM_{13} shown during the stability analysis suggests that the mean curves of the 50 repetitions are representative of the convergence process. Fig. 10 presents the mean curve results obtained for the previous population size samples ($N \in \{50, 100, 300, 500\}$) over the three landscapes described in Section 6.

The trends observed over the three different landscapes are very similar. The population size parameter shows only a slight influence, no matter what the landscape. In all cases, the maximum difference appears between population sizes 50 and 500. The maximum discrepancy values are 0.089, 0.095, and 0.019 for the linear, double-slope, and saw tooth landscapes respectively. More importantly, this analysis verifies that the linearity and coverage of the phenotypic responses increase as the population increases, which validates the explanation given in Section 6 of the behavior of PCM_{13}.

7.4. PCM_{13} analysis with outliers

The final experiment assesses the performance of PCM_{13} in the presence of outliers within the fitness distribution. In reality, the impact of outliers on phenotypic distribution remains unclear. While their presence should normally increase the diversity, they could also increase the fitness range, leading to an over-converged state of the remaining population, as compared to a population without outliers. This analysis will therefore help shed light on the effect of outliers.

To conduct the analysis, the landscapes introduced in Section 6 are adjusted as follows: from the 10th iteration, a given percentage of the population is randomly generated between the lower genotypic bounds of the 1st and 10th iterations. For the saw tooth landscape, this group is restricted to the first tooth. So, as the process evolves, these individuals act as outliers. Herein, the percentage of outliers is set to 1%, 2%, 5%, and 10%.

Fig. 11 presents the impact of outliers over the linear, double-slope, and saw tooth landscapes. The simulations are repeated 50 times, with the population size kept at 100. PCM_{13} adequately considers the presence of outliers, since the diversity level increases (or the convergence level decreases) as the number of outliers increases. With the fitness scaling factor set at 1, there is no sign of convergence value overestimation.

In real-world problems, however, outliers may modify the fitness range. This condition is simulated through a scaling factor applied to the fitness of the outliers. Figs. 12–14 present the results with 1% and 10% of outliers. Scaling factors of 10 and 100 are

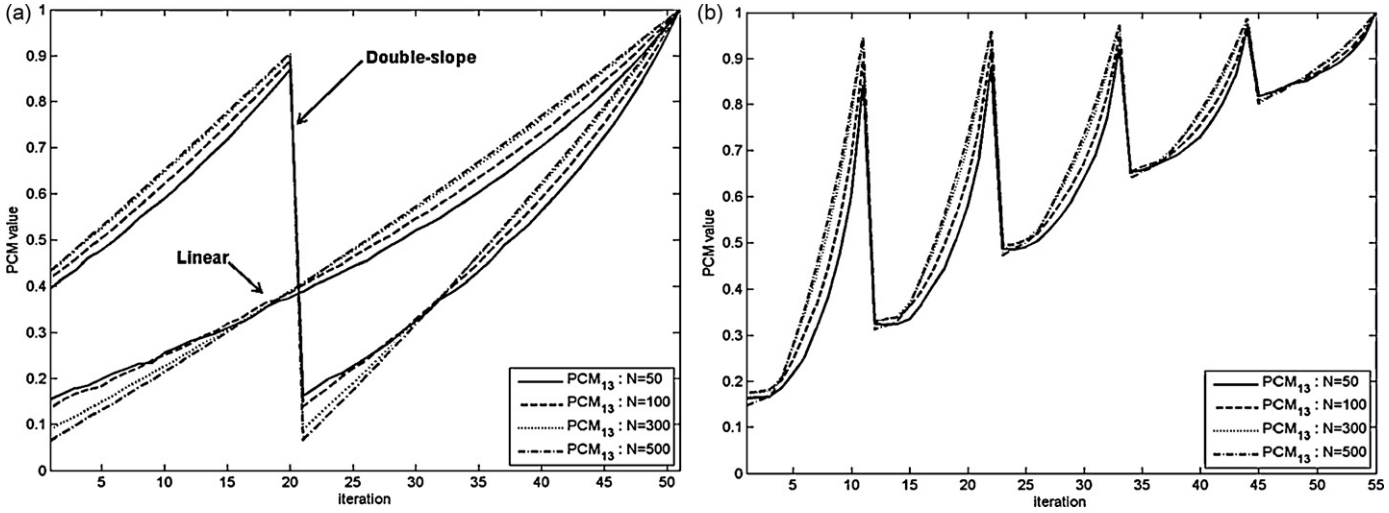

Fig. 10. Sensitivity of PCM_{13} with respect to the population size (N) observed over: (a) linear and double-slope landscapes, (b) saw tooth landscape.

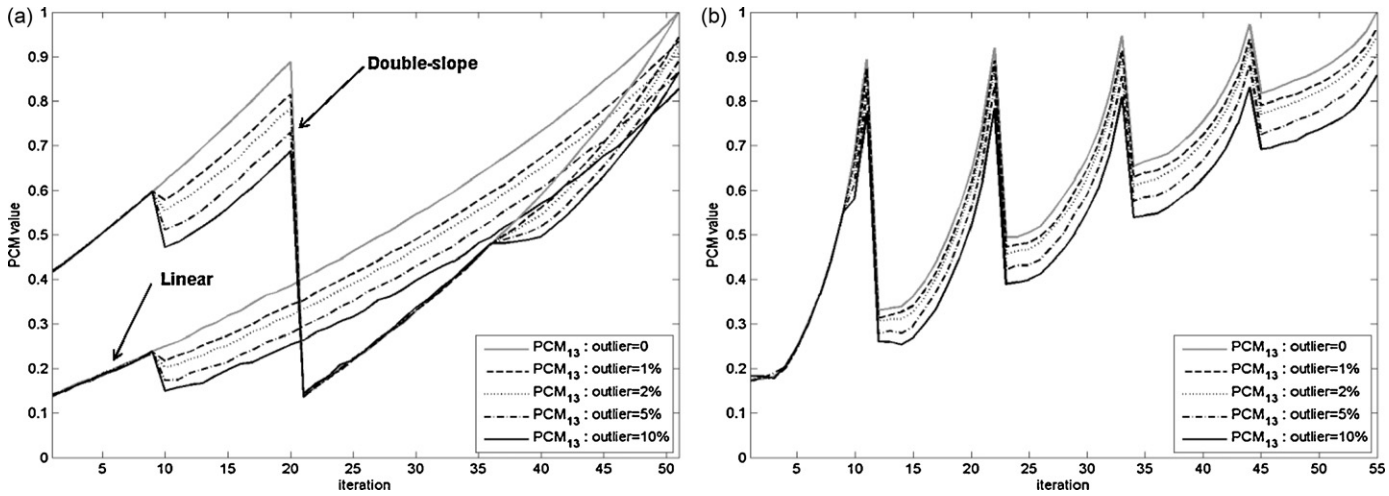

Fig. 11. Impact of outliers on PCM_{13} observed over: (a) linear and double-slope landscape, (b) saw tooth landscape.

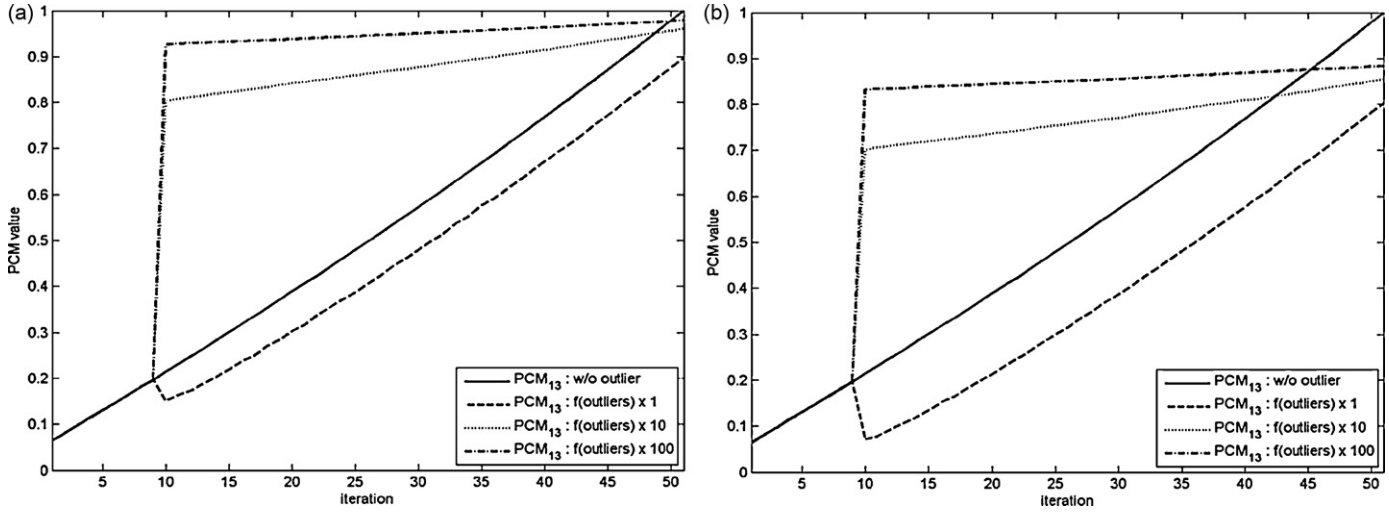

Fig. 12. Impact on PCM_{13} of outliers that are far away, observed over the linear landscape. (a) 1% outliers. (b) 10% outliers.

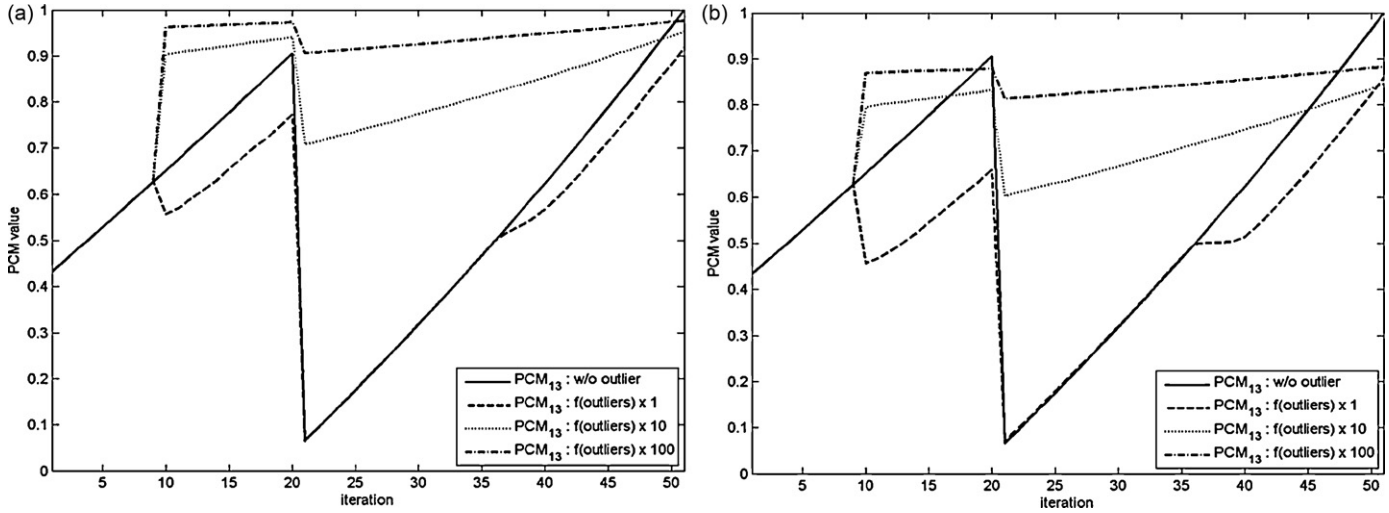

Fig. 13. Impact on PCM_{13} of outliers that are far away, observed over the double-slope landscape. (a) 1% outliers. (b) 10% outliers.

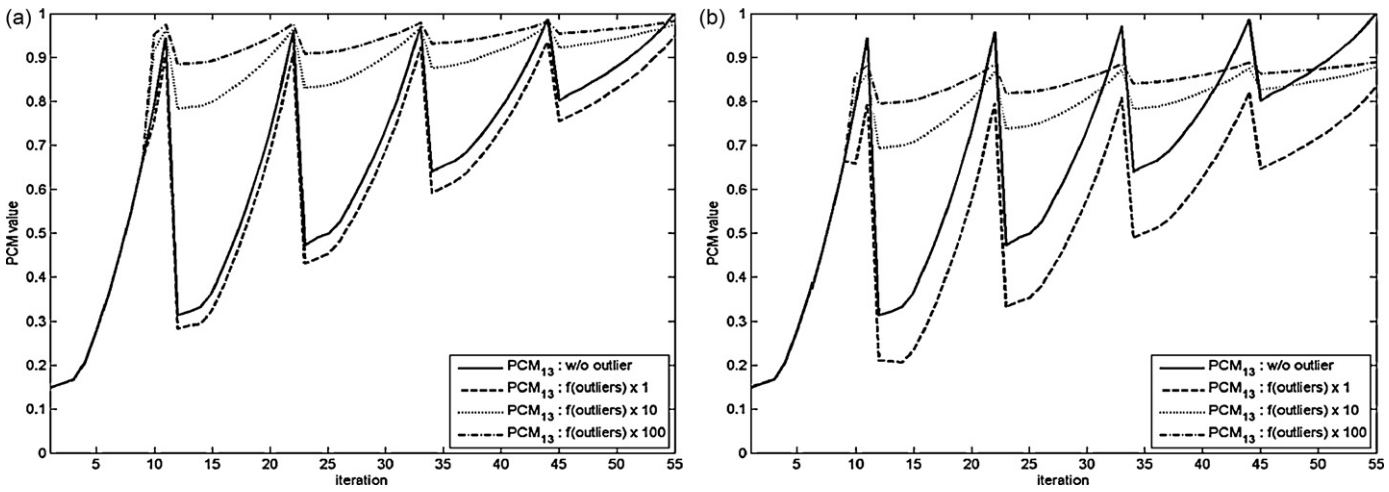

Fig. 14. Impact on PCM_{13} of outliers that are far away, observed over the saw tooth landscape. (a) 1% outliers. (b) 10% outliers.

compared to both the unit scaling factor and the cases without outliers. Even though the results are obtained for a population of 100, the trend is similar with other population sizes.

The simulations explicitly demonstrate that the enlargement of the fitness range by outliers leads to an overestimation of the convergence levels. More interestingly, increasing the number of outliers that disrupt the fitness range decreases the convergence level overestimation as they increasingly become part of the population. Finally, for all the landscapes studied, even with a scaling factor of 100, the overestimation of the phenotypic convergence level does not lead to an early full convergence state. In other words, even if outliers disturb the convergence pattern, the remaining fitness values continue to play an active role in the computed phenotypic state.

These observations suggest thus that PCM_{13} is capable of properly managing the presence of outliers within the fitness distribution.

8. Application of the EEB diagnostic tool

A good way to increase the efficiency of an optimizer over a given problem is to adjust its internal parameters [44]. Parameter setting can be considered in the broadest sense of the term, so that the number of populations involved, the type of evolution model, the diversity promoting features, and the restart strategies are all viewed as adjustable parameters. Since PCM_{13} reliability for phenotypic convergence description has been demonstrated, this section proposes an efficient diagnostic tool developed based on EEB information to help evaluate the impact of any particular parameter setting procedure. The information acquired from PCM_{13} is completed along the exploration axis of the EEB with D^N_{LN} (6). D^N_{LN} represents a generalized multivariate genotypic descriptor based on PCM_{13}. This measure acts on individual genotypic materials. The difference between individuals is defined by the minimum distance with respect to their neighbors. In this formulation, $x_{i,k}$ and $x_{j,k}$ stand for the value of gene k ($k \in \{1, \ldots, n\}$) of the individuals i and j respectively.

$$D^N_{LN} = \frac{\sum_{i=1}^{N-1} \ln\left(1 + \min_{j\in[i+1,N]} \frac{1}{n}\sqrt{\sum_{k=1}^{n}(x_{i,k} - x_{j,k})^2}\right)}{NMDF} \quad (6)$$

To illustrate and demonstrate the efficacy of the EEB diagnostic tool, the following experiments integrate and examine various selection plans, genetic operators, replacement plans, and population sizes. The impact of the evaluation of these parameters is measured by means of a real-coded steady-state GA (SSGA). SSGA, which allows smooth transitions between generations, can be summarized as follows: two offspring are created at each generation; two individuals are removed from the population, following a selected replacement plan, to make room for the new individuals; thereafter, the best individuals from this temporary pool are inserted back into the population.

Five selection plans are considered here: (1) random selection of the parents; (2–4) a tournament scheme with 2, 5, and 10 competitors; and (5) the recently proposed FUSS approach [35], while 4 genetic operators are integrated: (1) parent-centric crossover (PCX) [45,46]; (2) unimodal normal distribution crossover (UNDX) [47,48], which requires three parents; (3) uniform crossover (UX) [49]; and (4) the blended crossover (BLX-α) [50], where only two parents are involved. Five values of α (0.1, 0.3, .0.5, 0.7, and 0.9) are considered within the BLX operator, leading to a comparison of eight genetic operators. Note that no mutation is considered in the search process, as all these crossover operators (except UX) have the ability to incorporate new genetic material into the population. Two methods are compared for the replacement plan: (1) randomly removing individuals, and (2) removing the worst individuals from the population. Finally, four common population sizes are analyzed: 50, 100, 300, and 500. The comparisons involve a default setting with random selection and replacement, a PCX crossover, and a population size of 300. In other words, the impact of each choice is evaluated, one choice at a time.

The following figures (Figs. 15–18) present the results of our comparative studies. Continuous curves correspond to PCM information, and dashed curves identify GDM information. The curves present the median run of 25 repetitions. Since each run has a unique convergence history, averaging is difficult. Nonetheless, the curves are completed by a shaded area indicating the range of values obtained throughout the repetitions. Therefore, narrow shaded areas indicate that the median run is representative. For the sake of brevity, only the CEC'05 benchmark 10-D F2 and 10-D F10 functions from [51] are illustrated in Figs. 15–17, while the genetic operator study (Fig. 18) utilizes 10-D F2 and 10-D F21. This latter choice provides a better demonstration of the relevance of the orthogonal EEB framework (Section 2). In fact, Fig. 18(b) indicates that exploitation (PCM) and exploration (GDM) are not complementary measures, but are complementary concepts portraying the EEB history. This observation is corroborated by the difference between the shaded area patterns of the PCM and the GDM.

In order to control the simulation duration, and since the purpose of the experiments is to monitor the impact of the EA

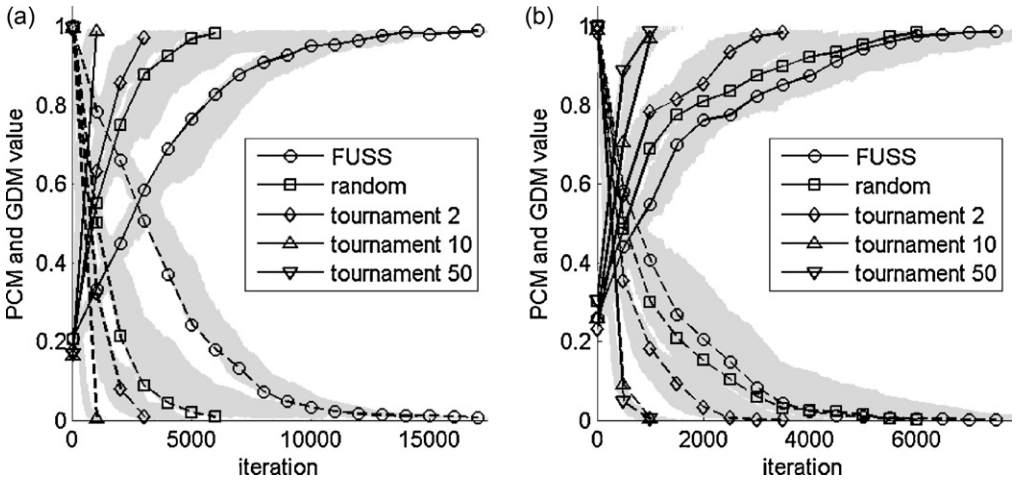

Fig. 15. Impact of various selection plans over the EEB. (a) 10D-F2. (b) 10D-F10.

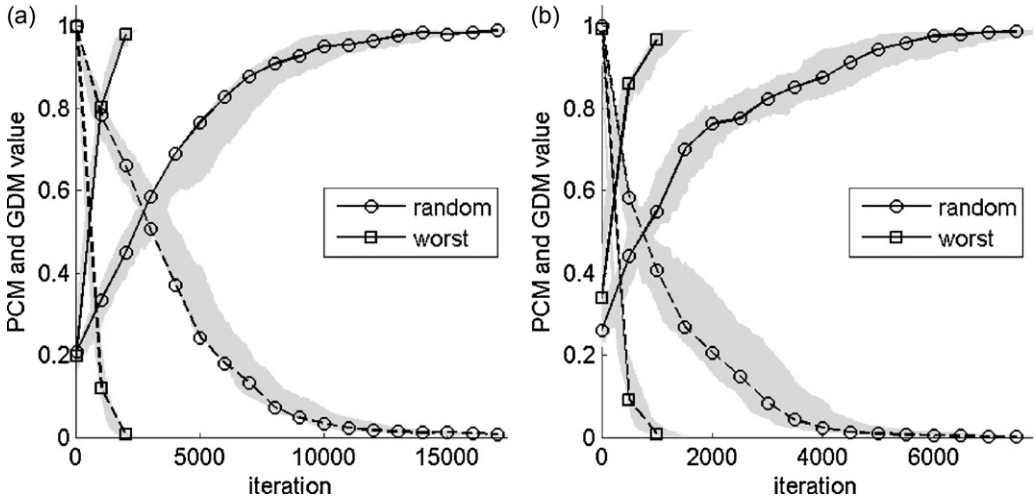

Fig. 16. Impact of various replacement plans combined with FUSS selection over the EEB. (a) 10D-F2. (b) 10D-F10.

parameters over the EEB (but not performance, in terms of end of solution quality), termination of the process was based on the CEC'05 criterion (100,000 evaluation cutoff) and a threshold applied over the PCM value (>0.99). Consequently, none of the simulated configurations found the global optimum. For instance, Fig. 18(a) clearly demonstrates that, in some cases, median runs ended due to phenotypic convergence, even though genotypic convergence had not been reached. This condition undoubtedly confirms that the PCM and the GDM have their own role to play. More importantly, it underscores the fact that premature

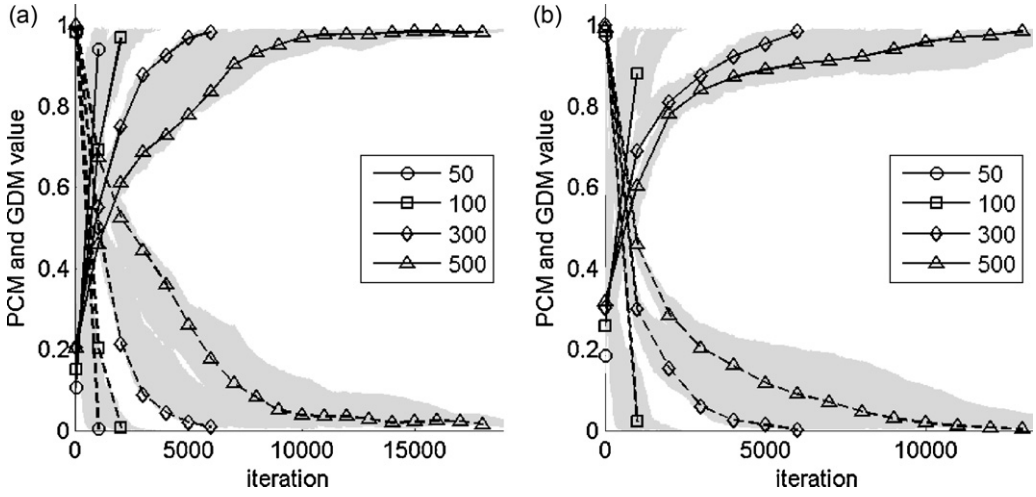

Fig. 17. Impact of various population sizes over the EEB. (a) 10D-F2. (b) 10D-F10.

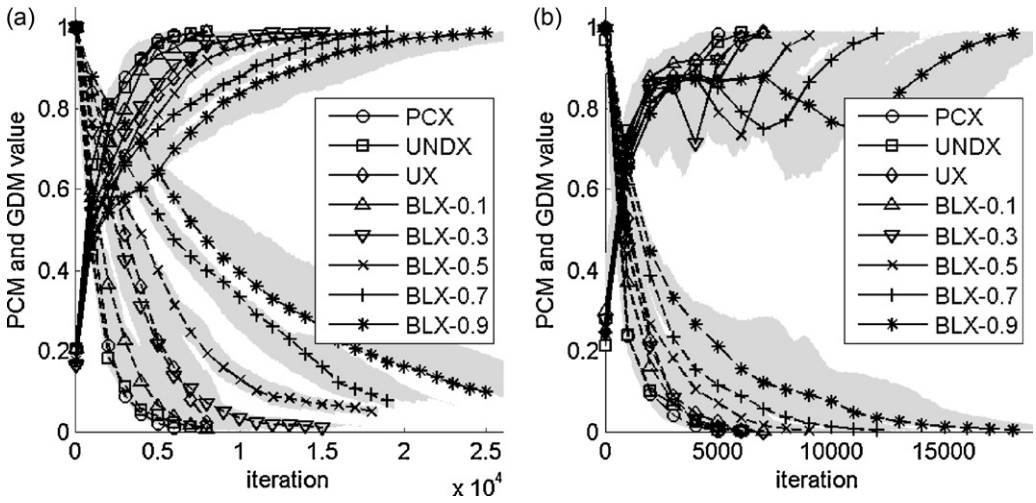

Fig. 18. Impact of various crossover types over the EEB. (a) 10D-F2. (b) 10D-F21.

convergence must be evaluated through the GDM instead of the PCM, as identical fitness values came from different locations.

Our experimental results support the conviction of many in the community that the impact of parameter choice is critically important. Regarding selection plans (Fig. 15), FUSS is shown to be the best option for delaying convergence, even better than the random search. In fact, FUSS was designed to avoid convergence [35]. However, in the presence of other evolutionary mechanisms, specifically the update plan, which promotes the best individuals, convergence is the inevitable result for any selection scheme. In contrast, compared to the previous schemes, the rate of convergence is higher for tournament selection. Obviously, this is accentuated as the tournament size increases [52].

Fig. 16 shows the impact of the replacement plan when tested with FUSS selection. As expected, replacing the worst individuals increases the convergence speed. Considering other simulations not included in this paper, this conclusion may be extended to all other selection methods. As a matter of fact, FUSS appears to be the most reactive selection method, while random selection appears to be the least responsive.

Fig. 17 describes the impact of population size on the EEB. It turns out, as expected, that increasing the population helps maintain both genotypic and phenotypic diversity.

Crossover types are examined in Fig. 18. It is observed that over F2, PCX and UNDX converge the fastest, followed by BLX-0.1, UX, and BLX-0.3 to BLX-0.9. The trend is similar for F21. However, no marked difference appears among PCX, UNDX, UX, and BLX-0.1 to BLX-0.3.

Finally, the search paths look very similar, regardless of the problem considered or the parameters selected. This may be surprising at first glance. However, since no mechanism that explicitly promotes diversity is incorporated, all the processes lead to similar search paths directed toward the best individual.

Even though they are constrained to a limited number of problems, the experiments presented here provide valuable insights into the true behavior of particular parameter choices. Globally, the results suggest that monitoring the EEB could serve as a powerful tool for characterizing EA differences and parameter influence, and may ultimately help in the design of better hybrid or improvement strategies.

9. Conclusion

Designing an efficient optimizer for a given problem is an issue that practitioners commonly encounter. It is quite a difficult task, as no single parameter setting procedure can drive the search path toward its optimal course over any landscape. This issue has led us to develop a diagnostic tool designed to help identify the best optimizer options for the problem at hand. This tool records the EEB of the optimizer throughout the search process. As the EEB is responsible for the way resources are committed, monitoring it is a basic requirement for achieving efficiency for any population-based search method.

The study pursued two main objectives. The first objective was to investigate the conceptualization of the EEB framework. Our results show that considering exploration and exploitation as two orthogonal axes offers an effective description of EEB. Subsequently, we identified the genotypic diversity measure (GDM) as the best description of exploration, and the phenotypic convergence measure (PCM) as an accurate portrayal of exploitation.

The second objective was to assess the performance of the diversity measures. Numerous formulations have been proposed in the literature over the years for doing so. However, to the best of the authors' knowledge, performance assessments of the various diversity formulations have only been conducted for GDM, although this evaluation is an important one. Consequently, we conducted a complete review of the PCM formulations, and compared nearly all the published PCMs over a validation framework involving six test cases with a controlled fitness distribution. With this simple framework, the underlying behavior of phenotypic formulations can be represented based on three requirements that we propose: monotonicity in fitness varieties, twinning, and monotonicity in distance. We proved that these requirements are sufficient for identifying phenotypic formulation weaknesses.

In summary, all the distance-based formulations fail to meet the monotonicity in fitness varieties and twinning requirements, and the fitness frequency formulations fail to meet to the twinning and monotonicity in distance requirements.

To improve the existing descriptive capacities of the formulations, we developed a new formulation (PCM_{13}), based on the neighbor fitness difference. Validation of PCM_{13} proves that it complies with all three requirements. In addition, we compared the phenotypic formulations over three specifically designed fitness landscapes. The same landscapes also served as a platform for assessing the desirable qualities of PCMs. In fact, PCM_{13} proved to be reliable over similarly scattered fitness distributions, and showed slight sensitivity to population size. However, the observed sensitivity level remained irrelevant for proper convergence evaluation. Furthermore, the influence of outliers was investigated, the results suggesting that PCM_{13} reliably takes into account the influence of outliers, even when they greatly disturb the fitness range.

The new phenotypic formulation, combined with a genotypic formulation based on a generalized version of PCM_{13}, is therefore proposed as the foundation for an EEB diagnostic tool. Its usefulness has been shown by comparing behavior of various genetic operators and parameters over a real-coded SSGA. With this EEB diagnostic tool, it is now possible to compare the underlying mechanisms of various maintaining/promoting diversity approaches, and to better understand them [53]. Finally, the next step would be to leverage the EEB diagnostic tool to develop an EEB management tool, to enable the search process to adapt its own evolutionary path as required, based on the PCM and GDM knowledge gathered.

References

[1] A.E. Eiben, J.E. Smith, Introduction to Evolutionary Computing, Berlin, Springer-Verlag, 2003.
[2] D.H. Wolpert, W.G. Macready, No free lunch theorems for optimization, IEEE Transaction on Evolutionary Computation 1 (1997) 67–82.
[3] A.E. Eiben, C.A. Schippers, On evolutionary exploration and exploitation, Fundamenta Informaticae 35 (1–4) (1998) 35–50.
[4] J.K. Bassett, K.A. De Jong, Using multivariate quantitative genetics theory to assist in EA customization, in: Proceedings of the 11th International Workshop on Foundations of Genetic Algorithms – FOGA 2011, ACM, Schwarzenberg, Austria, 2011, pp. 219–229.
[5] M. Turkey, R. Poli, An empirical tool for analysing the collective behaviour of population-based algorithms, in: Proceedings of the Application of Evolutionary Computation – EvoApplications 2012, vol. 7248 of Lecture Notes in Computer Science, Springer-Verlag, 2012, pp. 103–113.
[6] F. Herrera, M. Lozano, Adaptation of genetic algorithm parameters based on fuzzy logic controllers Genetic Algorithms and Soft Computing, 8, 1st ed., Physica-Verlag, 1996, pp. 95–125.
[7] M. Wineberg, F. Oppacher, The underlying similarity of diversity measures used in evolutionary computation, in: Proceedings of Genetic and Evolutionary Computation Conference – GECCO 2003, vol. 2724 of Lecture Notes in Computer Science, Springer, Chicago, IL, 2003, pp. 1493–1504.
[8] O. Olorunda, A.P. Engelbrecht, Measuring exploration/exploitation in particle swarms using swarm diversity, in: Proceedings of the IEEE Congress on Evolutionary Computation – CEC 2008, Hong Kong, 2008, pp. 1128–1134.
[9] G. Corriveau, R., Guilbault, A. Tahan, and R. Sabourin, Review and study of genotypic diversity measures for real-coded representations, IEEE Transaction on Evolutionary Computation, in press
[10] A.K. Gupta, K.G. Smith, C.E. Shalley, The interplay between exploration and exploitation, Academy of Management Journal 49 (4) (2006) 693–706.
[11] D.E. Goldberg, J. Richardson, Genetic algorithms with sharing for multimodal function optimization, in: Proceedings of the 2nd International Conference on Genetic Algorithms – ICGA 1987, Lawrence Erlbaum Associates, Cambridge, MA, 1987, pp. 41–49.
[12] T. Motoki, Calculating the expected loss of diversity of selection schemes, Evolutionary Computation 10 (4) (2002) 397–422.
[13] F. Neri, J. Toivanen, R.A.E. Mäkinen, An adaptive evolutionary algorithm with intelligent mutation local searchers for designing multidrug therapies for HIV, Applied Intelligence 27 (3) (2007) 219–235.
[14] A. Caponio, G.L. Cascella, F. Neri, N. Salvatore, M. Summer, A fast adaptive memetic algorithm for online and offline control design of PMSM drives, IEEE Transaction on Systems, Man and Cybernetics Part B: Cybernetics 37 (1) (2007) 28–41.
[15] V. Tirronen, F. Neri, Differential evolution with fitness diversity self-adaptation, in: Nature-Inspired Algorithms for Optimisation, vol. 193 of Studies in Computational Intelligence, Springer-Verlag, 2009, pp. 199–234.
[16] T. Friedrich, N. Hebbinghaus, F. Neumann, Comparison of simple diversity mechanisms on plateau functions, Theoretical Computer Science 410 (26) (2009) 2455–2462.
[17] E.K. Burke, S. Gustafson, G. Kendall, Diversity in genetic programming: an analysis of measures and correlation with fitness, IEEE Transaction on Evolutionary Computation 8 (1) (2004) 47–62.
[18] A. Lee, H. Takagi, Dynamic control of genetic algorithms using fuzzy logic techniques, in: Proceedings of the 5th International Conference on Genetic Algorithms – ICGA 1993, Morgan Kaufmann, Urbana-Champaign, IL, 1993, pp. 76–83.
[19] R. Subbu, A.C. Sanderson, P.P. Bonissone, Fuzzy logic controlled genetic algorithms versus tuned genetic algorithms: an agile manufacturing application, in: Proceedings 1998 IEEE ISIC/CIRA/ISAS Joint Conference, IEEE, Gaithersburg, MD, 1998, pp. 434–440.
[20] J.A. Vasconcelos, J.A. Ramirez, R.H.C. Takahashi, R.R. Saldanha, Improvements in genetic algorithms, IEEE Transactions on Magnetics 37 (5) (2001) 3414–3417.
[21] E. Pellerin, L. Pigeon, S. Delisle, Self-adaptive parameters in genetic algorithms, in: Proceedings of the SPIE – Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI, vol. 5433, Bellingham, WA, 2004, pp. 53–64.
[22] M. Srinivas, L.M. Patnaik, Adaptive probabilities of crossover and mutation in genetic algorithms, IEEE Transaction on Systems, Man and Cybernetics 24 (4) (1994) 656–667.
[23] S. Arnone, M. Dell'Orto, A. Tettamanzi, Toward a fuzzy government of genetic populations, in: Proceedings of the 6th International Conference on Tools Artificial Intelligence – ICTAI 1994, IEEE, New Orleans, LA, 1994, pp. 585–591.
[24] F. Neri, J. Toivanen, G.L. Cascella, Y.-S. Ong, An adaptive multimeme algorithm for designing HIV multidrug therapies, IEEE/ACM Transaction on Computational Biology and Bioinformatics 4 (2) (2007) 264–278.
[25] F. Neri, N. Kotilainen, M. Vapa, A memetic-neural approach to discover resources in P2P networks, in: Recent Advances in Evolutionary Computation for Combinatorial Optimization, vol. 153 of Studies in Computational Intelligence, Springer-Verlag, 2008, pp. 113–129.
[26] A. Caponio, F. Neri, V. Tirronen, Super-fit control adaptation in memetic differential evolution frameworks, Soft Computing – A Fusion of Foundations, Methodologies, and Applications 13 (8–9) (2009) 811–831.
[27] F. Neri, G.L. Cascella, N. Salvatore, A.V. Kononova, G. Acciani, Prudent-daring vs. tolerant survivor selection schemes in control design of electric drives, in: Proceedings of the Application of Evolutionary Computation–EvoWorkshops, vol. 3907 of Lecture Notes in Computer Science, Springer-Verlag, 2006, pp. 805–809.
[28] F. Neri, R.A.E. Mäkinen, Hierarchical evolutionary algorithms and noise compensation via adaptation, in: Evolutionary Computation in Dynamic and Uncertain Environments, vol. 51 of Studies in Computational Intelligence, Springer-Verlag, 2007, pp. 345–369.
[29] A.-M. Miao, X.-L. Shi, J.-H. Zhang, E.-Y. Wang, S.-Q. Peng, A modified particle swarm optimizer with dynamical inertia weight, in: Fuzzy Information and Engineering, vol. 62 of Advances in Soft Computing, Springer-Verlag, 2009, pp. 767–776.
[30] V. Tirronen, F. Neri, T. Karkkainen, K. Majava, T. Rossi, A memetic differential evolution in filter design for defect detection in paper production, in: Proceedings of the Application of Evolutionary Computation – EvoWorkshops, vol. 4448 of Lecture Notes in Computer Science, Springer-Verlag, 2007, pp. 320–329.
[31] R.K. Ursem, Diversity-guided evolutionary algorithms, in: Proceedings of the 7th International Conference on Parallel Problem Solving from Nature – PPSN VII, vol. 2439 of Lecture Notes in Computer Science, Springer, Granada, Spain, 2002, pp. 462–471.
[32] H.A. Abbass, K. Deb, Searching under multi-evolutionary pressures, in: Proceedings of the 2nd International Conference on Evolutionary Multi-Criterion Optimization – EMO 2003, vol. 2632 of Lecture Notes in Computer Science, Springer, Faro, Portugal, 2003, pp. 391–404.
[33] R.W. Morrison, K.A. De Jong, Measurement of population diversity, in: Selected papers from the 5th International Conference on Artificial Evolution – AE 2001, vol. 2310 of Lecture Notes in Computer Science, Springer-Verlag, Le Creusot, France, 2002, pp. 31–41.
[34] A.L. Barker, W.N. Martin, Dynamics of a distance-based population diversity measure, in: Proceedings of the IEEE Congress on Evolutionary Computation – CEC 2000, La Jolla, CA, 2000, pp. 1002–1009.
[35] M. Hutter, S. Legg, Fitness uniform optimization, IEEE Transaction on Evolutionary Computation 10 (5) (2006) 568–589.
[36] C.E. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (1948) 623–656, 379-423.
[37] J.P. Rosca, Entropy-driven adaptive representation, in: Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications, Rochester, NY, 1995, pp. 23–32.
[38] P.J. Darwen, Black magic: interdependence prevents principled parameter setting, self-adapting costs too much computation, in: Applied Complexity: from Neural Nets to Managed Landscapes, NZ Crop & Food Research, Dunedin, New Zealand, 2000, pp. 227–237.
[39] J. Havrda, F. Charvát, Quantification method of classification processes. Concept of structural a-entropy, Kybernetika 3 (1) (1967) 30–35.
[40] A. Rényi, On measures of entropy and information, in: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, 1961, pp. 547–561.
[41] M.L. Weitzman, On diversity, The Quarterly Journal of Economics 107 (2) (1992) 363–405.
[42] A.R. Solow, S. Polasky, Measuring biological diversity, Environmental and Ecological Statistics 1 (2) (1994) 95–107.
[43] S.-H. Cha, S.N. Srihari, On measuring the distance between histograms, Pattern Recognition 35 (6) (2002) 1355–1370.
[44] A.E. Eiben, R. Hinterding, Z. Michalewicz, Parameter control in evolutionary algorithms, IEEE Transaction on Evolutionary Computation 3 (2) (1999) 124–141.
[45] K. Deb, D. Joshi, A. Anand, Real-coded evolutionary algorithms with parent-centric recombination, in: Proceedings of the IEEE 2002 World Congress on Computational Intelligence, Honolulu, HI, 2002, pp. 61–66.
[46] K. Deb, A. Anand, D. Joshi, A computationally efficient evolutionary algorithm for real-parameter optimization, Evolutionary Computation 10 (4) (2002) 371–395.
[47] I. Ono, S. Kobayashi, A real-coded genetic algorithm for function optimization using unimodal normal distribution crossover, in: Proceedings of the 7th International Conference on Genetic Algorithms – ICGA 1997, Morgan Kaufmann, East Lansing, MI, 1997, pp. 246–253.
[48] H. Kita, I. Ono, S. Kobayashi, Theoretical analysis of the unimodal normal distribution crossover for real-coded genetic algorithms, in: Proceedings of the IEEE Congress on Evolutionary Computation – CEC 1998, Anchorage, AK, 1998, pp. 529–534.

[49] G. Sywerda, Uniform crossover in genetic algorithms, in: Proceedings of the 3rd International Conference on Genetic Algorithms – ICGA 1989, Morgan Kaufmann, Fairfax, VA, 1989, pp. 2–9.
[50] L.J. Eshelman, J.D. Schaffer, Real-coded genetic algorithms and interval-schemata, in: Proceedings of the 2nd International Workshop on Foundations of Genetic Algorithms – FOGA 1992, Morgan Kaufmann, Vail, CO, 1992, pp. 187–202.
[51] P.N. Suganthan, et al., Problem definitions and evaluation criteria for the CEC 2005 special session on real parameter optimization, Nanyang Technical University, Singapore, Technical Report #2005005, May 2005.
[52] D.E. Goldberg, K. Deb, A comparative analysis of selection schemes used in genetic algorithms, in: Proceedings of the 1st International Workshop on Foundations of Genetic Algorithms – FOGA 1990, Bloomington, IN, 1990, pp. 69–93.
[53] S. Das, S. Maity, B.-Y. Qu, P.N. Suganthan, Real-parameter evolutionary multimodal optimization – a survey of the state-of-the-art, Swarm and Evolutionary Computation 1 (2) (2011) 71–88.