Swarm and Evolutionary Computation 38 (2018) 79–93

Contents lists available at ScienceDirect

Swarm and Evolutionary Computation

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

Weibull-based scaled-differences schema for Differential Evolution

Miguel Cárdenas-Montes

CIEMAT, Department of Fundamental Research, Avda. Complutense 40, 28040 Madrid, Spain

ARTICLE INFO

Keywords:
Differential Evolution
Performance
Weibull probability distribution

ABSTRACT

Differential Evolution is one of the most efficient real-parameter optimization algorithm. It is based on the application of the scaled difference of a pair of population members to another population member, all of them distinct. Diverse variants have been proposed within this schema. In this work, the statistical distribution of these differences of high-performance variants of Differential Evolution is modelled through a Weibull probability distribution. From the application of this model to diverse Differential Evolution variants and benchmark functions, a pattern for the most efficient variants can be drawn. As a consequence, a variant where the scaled differences are replaced by random numbers generated from a Weibull distribution is proposed and evaluated.

1. Introduction

Differential Evolution (DE) is one of the most popular and efficient real-parameter optimization algorithm [1,2]. It is based on perturbing the population members of each generation with other population member, randomly selected, and the scaled difference of a pair of members, being all these members distinct. The alteration of the population is achieved through two operators: the Mutation operator (Eq. (1)) and the Crossover operator (Eq. (2)).

The Mutation operator generates an intermediate population of vectors or members based on the addition of a scaled difference of two randomly-selected vectors, $F\cdot(\overrightarrow{x_2} - \overrightarrow{x_3})$, over a third randomly-selected vector, $\overrightarrow{x_1}$ (base vector), in such a way that $\overrightarrow{x_1}$, $\overrightarrow{x_2}$, and $\overrightarrow{x_3}$ are mutually exclusive, being F a scalar. The mutation factor, F, quantifies the amount of alteration supplied to the base vector. The vectors generated by the mutation operators are termed *donor* or *mutant vectors*.

$$\overrightarrow{v_i} = \overrightarrow{x_1} + F\cdot(\overrightarrow{x_2} - \overrightarrow{x_3}) \tag{1}$$

Next, mutant vectors ($\overrightarrow{v}$) are discretely crossed with the original population members in the current generation (*target vector*) to generate the *trial vectors* ($\overrightarrow{u}$). Two crossover operators are the most frequently used: Binomial Crossover operator (Eq. (2)) and Exponential Crossover operator (Eq. (3)). The Crossover Operator is characterized by a parameter, termed crossover rate, which governs how many components from the mutant vector are inherited by the trial vector.

The Binomial Crossover operator (Eq. (2)) generates random numbers for every element and performs crossover for values lower than C_r.

$$u_i(j) = \begin{cases} v_i(j) & \text{if } rand \le C_r; \\ x_i(j) & \text{otherwise.} \end{cases} \tag{2}$$

Alternatively, the Exponential Crossover operator selects for replacement a random contiguous block of elements (Eq. (3)). For the length of this block, firstly a starting point, randomly picked from a uniform probability distribution, is generated. Next, random numbers are generated, and the number of tries counted, while they are lower than C_r. When exiting from the previous conditions, the number of tries points corresponds to the length of the replacement. The brackets in Eq. (3) indicate the modulus function with modulus n.

$$u_i(j) = \begin{cases} v_i(j), & \text{for } j = \langle n\rangle_D \langle n+1\rangle_D, \ldots, \langle n+L-1\rangle_D \\ x_i(j), & \forall\, j \in [1, D] \end{cases} \tag{3}$$

Following the schema of the mutation operator (Eq. (1)), numerous variants have been proposed (Section 2.1). These variants include the selection as *base vector* a randomly selected vector, the best of the current generation, or the target vector. Other variants vary the number of scaled differences with randomly selected vectors, or other criteria for the selection of the vectors which participate in the scaled differences. The purpose of these combinations is to balance exploration and exploitation behaviours in the algorithm, and as a consequence to produce high-quality solutions.

In this work, the performance of the schemas with base vector the best vector of the current generation is analysed. In these schemas the different performance is originated by the different nature of the scaled differences. From this study, the behaviour exhibited by the top-performer variants is modelled based on Weibull probability distribution. This allows proposing a new DE-variant where the scaled

E-mail address: miguel.cardenas@ciemat.es.

http://dx.doi.org/10.1016/j.swevo.2017.06.004
Received 5 December 2016; Received in revised form 25 May 2017; Accepted 27 June 2017
Available online 01 July 2017
2210-6502/

Table 1
Benchmark functions used in this work.

Name	Function	Search range	Optimum
Ackley	$f(\vec{x}) = -20e^{(-0.2\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2})} - e^{(\frac{1}{n}\sum_{i=1}^{n}cos(2\pi x_i))} + 20 + e$	$[-32.0, 32.0]$	$0\ at\ \vec{0}$
Griewank	$f(\vec{x}) = 1 + \frac{1}{4000}\sum_{i=1}^{D}x_i^2 - \prod_{i=1}^{D}cos(\frac{x_i}{\sqrt{i}})$	$[-600.0, 600.0]$	$0\ at\ \vec{0}$
Hyperellipsoid	$f(\vec{x}) = \sum_{i=1}^{D} i\cdot x_i^2$	$[-5.12, 5.12]$	$0\ at\ \vec{0}$
Rastrigin	$f(\vec{x}) = 10\cdot D + \sum_{i=1}^{D}(x_i^2 - 10\cdot cos(2\pi x_i))$	$[-5.12, 5.12]$	$0\ at\ \vec{0}$
Rosenbrock	$f(\vec{x}) = \sum_{i=1}^{D-1}(x_i - 1)^2 + 100\cdot(x_{i+1} - x_i^2)^2$	$[-30, 30]$	$0\ at\ \vec{0}$
Schaffer's F6	$f(\vec{x}) = 0.5 + \frac{sin^2(\sqrt{\sum_{j=1}^{D}x_i^2}) - 0.5}{[1 + 0.001\cdot(\sum_{j=1}^{D}x_i^2)]^2}$	$[-100.0, 100.0]$	$0\ at\ \vec{0}$
Schaffer's F7	$f(\vec{x}) = \frac{1}{n-1}\sum_{i=1}^{D-1}[\sqrt{s_i}\cdot(sin^2(50.0s_i^{\frac{1}{5}}) + 1)]^2$, $s_i = \sqrt{x_i^2 + x_{i+1}^2}$	$[-100.0, 100.0]$	$0\ at\ \vec{0}$
Schwefel	$f(\vec{x}) = \sum_{i=1}^{D}(-x_i\cdot sin\sqrt{\lvert x_i\rvert}) + 418.982887\cdot D$	$[-500.0, 500.0]$	$0\ at\ \overrightarrow{420.968746}$
Schwefel Problem 1.2	$f(\vec{x}) = \sum_{i=1}^{D}(\sum_{j=1}^{i}x_j)^2$	$[-100.0, 100.0]$	$0\ at\ \vec{0}$
Schwefel Problem 2.22	$f(\vec{x}) = \sum_{i=0}^{D}\lvert x_i\rvert + \prod_{i=0}^{D}\lvert x_i\rvert$	$[-10.0, 10.0]$	$0\ at\ \vec{0}$
Schwefel Problem 2.21	$f(\vec{x}) = \max\lvert x_i\rvert,\ 1 \le i \le D$	$[-100.0, 100.0]$	$0\ at\ \vec{0}$
Sphere	$f(\vec{x}) = \sum_{i=1}^{D}x_i^2$	$[-5.12, 5.12]$	$0\ at\ \vec{0}$
Step	$f(\vec{x}) = \sum_{j=1}^{D}(\lfloor x_i + 0.5\rfloor)^2$	$[-1000, 1000]$	$0\ at\ \vec{0}$
Styblinski-Tang	$f(\vec{x}) = 0.5\cdot\sum_{i=1}^{D}x_i^4 - 16x_i^2 + 5x_i$	$[-5, 5]$	$-39.16599\cdot D\ at\ \overrightarrow{-2.903534}$
Whitley	$f(\vec{x}) = \sum_{i=1}^{D}\sum_{j=1}^{D}(\frac{(s_{ij})^2}{4000} - cos(s_{ij}) + 1)$, $s_{ij} = 100(x_i^2 - x_j)^2 + (1 - x_j)^2$	$[-10.24, 10.24]$	$0\ at\ \vec{1}$
Zakharov	$f(\vec{x}) = \sum_{i=1}^{D}x_i^2 + (\sum_{i=1}^{D}0.5\cdot i\cdot x_i)^2 + (\sum_{i=1}^{D}0.5\cdot i\cdot x_i)^4$	$[-5, 10]$	$0\ at\ \vec{0}$

Table 2
Mean fitness and standard deviation from 25 independent runs for DE parameters $F = Cr = 0.5$ and two configurations for the number of cycles: 100 and 1000, and Binomial Crossover operator. The best DE variant for each configuration and fitness function appears in boldface type.

Function	Cycles	rand/1	rand/2	best/1	best/2	current-to-best/1	rand-to-best/1	rand/2/dir
Ackley	100	8.25 ± 1.48	15.91 ± 1.28	12.12 ± 2.08	**6.83 ± 1.46**	11.52 ± 1.39	10.65 ± 1.96	10.89 ± 2.00
	1000	1.98 ± 1.35	$(1.0 \pm 0.4)10^{-2}$	10.73 ± 1.83	**$(5.1 \pm 9.7)10^{-5}$**	10.74 ± 1.10	10.75 ± 1.86	3.61 ± 2.47
Griewank	100	19.64 ± 6.78	156.60 ± 48.94	56.34 ± 19.20	**17.19 ± 9.88**	49.27 ± 22.23	36.00 ± 16.88	59.51 ± 33.97
	1000	1.16 ± 1.61	$(2.5 \pm 1.4)10^{-5}$	42.18 ± 21.32	**$(0.7 \pm 1.3)10^{-9}$**	41.55 ± 13.86	44.47 ± 16.00	4.02 ± 6.28
Hyperellipsoid	100	83.29 ± 36.11	641.81 ± 186.28	208.18 ± 67.64	**48.88 ± 22.25**	182.96 ± 82.82	130.97 ± 61.91	188.31 ± 80.51
	1000	9.31 ± 12.01	$(9.0 \pm 6.6)10^{-5}$	147.86 ± 82.14	**$(0.8 \pm 3.1)10^{-8}$**	168.87 ± 80.62	124.16 ± 49.05	12.37 ± 23.87
Rastrigin	100	239.05 ± 17.84	308.77 ± 30.11	**115.47 ± 35.91**	236.70 ± 21.74	143.43 ± 25.56	129.46 ± 34.98	273.98 ± 37.23
	1000	96.61 ± 19.53	167.81 ± 19.80	86.45 ± 15.81	96.17 ± 36.72	69.17 ± 18.00	58.58 ± 14.52	**52.29 ± 19.77**
Rosenbrock	100	$(117 \pm 45)10^3$	$(477 \pm 107)10^3$	$(69 \pm 32)10^3$	**$(48 \pm 20)10^3$**	$(49 \pm 21)10^3$	$(50 \pm 20)10^3$	$(179 \pm 127)10^3$
	1000	$(5.7 \pm 4.3)10^3$	$(16 \pm 14)10^3$	$(55 \pm 19)10^3$	**$(1.3 \pm 0.8)10^3$**	$(45 \pm 19)10^3$	$(51 \pm 25)10^3$	$(14 \pm 8)10^3$
Schaffer's F6	100	$(4996 \pm 1)10^{-4}$	$(4999 \pm 1)10^{-4}$	$(4974 \pm 24)10^{-4}$	$(4991 \pm 6)10^{-4}$	$(4944 \pm 48)10^{-4}$	**$(4925 \pm 58)10^{-4}$**	$(4998 \pm 2)10^{-4}$
	1000	**0.27 ± 0.11**	0.39 ± 0.04	0.49 ± 0.01	0.31 ± 0.10	0.49 ± 0.01	0.49 ± 0.01	0.42 ± 0.08
Schaffer's F7	100	46.73 ± 7.63	94.98 ± 12.96	40.86 ± 10.15	47.27 ± 13.08	32.77 ± 9.76	**25.61 ± 7.80**	72.77 ± 16.39
	1000	**1.22 ± 1.46**	6.36 ± 2.48	26.96 ± 5.64	3.17 ± 2.47	19.30 ± 4.68	16.57 ± 5.35	7.76 ± 3.80
Schwefel	100	7554 ± 357	7879 ± 373	**5130 ± 734**	7790 ± 389	7014 ± 610	6835 ± 623	7163 ± 626
	1000	**1282 ± 977**	6194 ± 333	4442 ± 361	3506 ± 1641	4837 ± 5147	4492 ± 615	1695 ± 497
Schwefel Problem 1.2	100	$72{,}131 \pm 17{,}082$	$121{,}520 \pm 32{,}543$	$24{,}326 \pm 6497$	$63{,}632 \pm 19{,}322$	**12766 ± 4593**	$16{,}940 \pm 3563$	$94{,}925 \pm 31{,}018$
	1000	$15{,}470 \pm 5420$	$71{,}092 \pm 15{,}653$	$13{,}615 \pm 6002$	**4621 ± 1931**	7407 ± 3058	9143 ± 3589	$21{,}721 \pm 16{,}698$
Schwefel Problem 2.21	100	63.49 ± 8.89	84.17 ± 4.58	44.25 ± 6.37	50.88 ± 9.67	**34.07 ± 6.51**	34.64 ± 7.97	81.49 ± 6.87
	1000	35.27 ± 9.81	**22.94 ± 6.96**	42.28 ± 5.54	23.01 ± 4.72	35.19 ± 6.82	35.33 ± 5.46	47.54 ± 6.68
Schwefel Problem 2.22	100	22.78 ± 5.91	$32{,}876 \pm 143{,}271$	26.62 ± 8.90	22.24 ± 5.71	22.11 ± 7.68	**18.07 ± 7.00**	27.78 ± 9.76
	1000	0.35 ± 0.72	0.03 ± 0.02	24.36 ± 9.78	**$(0.3 \pm 1.8)10^{-3}$**	20.64 ± 5.01	20.27 ± 7.79	2.11 ± 2.80
Sphere	100	5.82 ± 2.94	54.96 ± 15.67	14.50 ± 7.19	4.01 ± 2.07	14.25 ± 5.21	11.11 ± 4.77	**1.99 ± 2.28**
	1000	0.37 ± 0.47	$(8.6 \pm 8.0)10^{-6}$	12.08 ± 4.90	**$(1.7 \pm 3.5)10^{-10}$**	11.01 ± 5.21	10.03 ± 3.95	10.70 ± 5.74
Step	100	$(0.22 \pm 0.09)10^6$	$(1.82 \pm 0.37)10^6$	$(0.65 \pm 0.27)10^6$	**$(0.16 \pm 0.07)10^6$**	$(0.54 \pm 0.18)10^6$	$(0.43 \pm 0.14)10^6$	$(0.59 \pm 0.27)10^6$
	1000	$(0.2 \pm 0.4)10^5$	**1.12 ± 1.63**	$(5.0 \pm 2.0)10^5$	529 ± 887	$(5.8 \pm 2.0)10^5$	$(5.0 \pm 2.1)10^5$	$(0.5 \pm 0.5)10^5$
Styblinski Tang	100	-750 ± 37	-559 ± 70	-927 ± 36	-800 ± 64	-912 ± 41	**-947 ± 41**	-650 ± 116
	1000	**-1111 ± 26**	-1008 ± 95	-963 ± 40	-1058 ± 34	-944 ± 42	-955 ± 37	-1040 ± 38
Whitley	100	$(23 \pm 27)10^6$	$(3 \pm 2)10^9$	$(31 \pm 38)10^6$	**$(5 \pm 4)10^6$**	$(12 \pm 14)10^6$	$(10 \pm 12)10^6$	$(1.8 \pm 1.9)10^9$
	1000	$(1.9 \pm 4.9)10^6$	**1.25 ± 1.78**	$(44 \pm 102)10^6$	4.72 ± 13.43	$(10 \pm 16)10^6$	$(12 \pm 20)10^6$	$(19 \pm 43)10^6$
Zakharov	100	399.2 ± 95.5	671.7 ± 129.2	167.8 ± 74.3	335.4 ± 94.5	135.0 ± 50.2	**93.7 ± 38.3**	557.0 ± 150.3
	1000	1.19 ± 1.11	284.89 ± 117.54	17.42 ± 17.92	**0.41 ± 1.14**	28.18 ± 25.93	40.48 ± 23.63	5.30 ± 7.44

differences is replaced by random numbers generated from a Weibull probability distribution. The performance of this new DE-variant is tested by confronting it with other variants (Section 2.1) over a wide set of benchmark functions (Section 2.2).

The rest of the paper is organized as follows: the DE-variants considered in this work are described in Section 2.1, whereas the benchmark functions necessary to evaluate their efficiency are described in Section 2.2. In Section 2.3, the Weibull probability distribution is briefly described. The statistical tests used for analysing the results are described in Section 2.4. In Section 3, the experimental study is presented and a new variant based on this analysis is proposed and evaluated. Finally, Section 4 contains the conclusions of this work.

2. Methods and materials

2.1. Differential Evolution variants

From the initial publication [1] in the 90', numerous variants of DE have been proposed to improve the efficiency of the initial version (Algorithm 1). Early, the original authors of DE propose some variants [2]. A detailed review of these variants and other ground-breaking modifications for DE can be found in [3–5]. These surveys include the portfolio of DE-variants, performance studies, and applications to real-world optimization.

Algorithm 1. Algorithm pseudocode for the DE/rand/1/bin variant.

```
Initialize a random population x⃗;
Establish the parameters of DE: mutation factor F and crossover
probability Cr;
Evaluate the current population with the fitness function;
while stopping criterion is not met do
    forall vector in population do
        Randomly choose 3 distinct vectors ;
        Generate a random index and a random number, and created
        the mutant vectors v⃗ (Eq.1);
    forall vector in population do
        Apply the Binomial Crossover Operator (Eq. 2) and generate
        the trial vectors u⃗;
    Evaluate the trial population;
    forall vector in population do
        Apply Selection Operator and generate the next generation
        population, x⃗_{G+1};
```

For mentioning the DE-variants, a nomenclature based on the pattern DE/X/Y/Z is frequently used. The two initial letters (DE) correspond to the name of the algorithm, whereas the following ones consecutively correspond to the mechanism to select the *base vector* (X), the number of scaled differences involved (Y), and the crossover operator (Z). Following this nomenclature, the original DE is DE/rand/1/bin. When using the Exponential Crossover operator instead of Binomial one, then the nomenclature is DE/rand/1/exp.

In this work the following DE-variants have been analysed.[1] In all tests, the Binomial Crossover operator is used.

DE/rand/1 Original proposal of DE.$\vec{v_i} = \vec{x_1} + F\cdot(\vec{x_2} - \vec{x_3})$

DE/rand/2 In this case, the scaled differences are generated with four vectors.$\vec{v_i} = \vec{x_1} + F\cdot(\vec{x_2} + \vec{x_3} - \vec{x_4} - \vec{x_5})$

DE/best/1 In this variant, the base vector is not randomly selected, but the best vector of the current generation is used.- $\vec{v_i} = \vec{x}_{best} + F\cdot(\vec{x_2} - \vec{x_3})$

DE/best/2 Similar to DE/best/1, but the number of vectors involved in the scaled difference is doubled.- $\vec{v_i} = \vec{x}_{best} + F\cdot(\vec{x_1} + \vec{x_2} - \vec{x_3} - \vec{x_4})$For comparison purpose, DE/best/3 is also used in this work.- $DE/best/3 \equiv \vec{v_i} = \vec{x}_{best} + F\cdot(\vec{x_1} + \vec{x_2} + \vec{x_3} - \vec{x_4} - \vec{x_5} - \vec{x_6})$

DE/current-to-best/1 In this variant, a difference including the best vector in the current generation and the target vector is included, at the same time that the target vector is used as base vector.$\vec{v_i} = \vec{x_i} + F\cdot(\vec{x}_{best} - \vec{x_i}) + F\cdot(\vec{x_2} - \vec{x_3})$This variant and the previous ones were proposed by Price and Storn in [2].

DE/rand-to-best/1 In the scaled differences, a difference including the best vector in the current generation and the base vector is present in the variant.$\vec{v_i} = \vec{x_1} + F\cdot(\vec{x}_{best} - \vec{x_1}) + F\cdot(\vec{x_2} - \vec{x_3})$

DE/rand/2/dir This variant incorporates information of the fitness function. The objective is to guide the evolution towards favourable regions [6].$\vec{v_i} = \vec{x_1} + \frac{\mu}{2}\cdot(\vec{x_1} - \vec{x_2} + \vec{x_3} - \vec{x_4})$where $f(\vec{x_1}) < f(\vec{x_2})$ and $f(\vec{x_3}) < f(\vec{x_4})$.

All the previous variants have steady values of parameters F and C_r, and therefore they can be considered as static variants. However, these parameters can vary along the cycles. In [2] two variants, termed *jitter* and *dither*, are proposed with random values for the scaling factor, F. *Dither* is used when generating anew random F value for each difference vector; whereas if the new random F value is generated for each dimension of each difference vector, then the variant is termed *jitter*. In [7] a dither schema with F randomly varying between 0.5 and 1 for each vector is proposed.

For comparison purpose, in this work two dither variants are implemented: the first one for which the scaling parameter is generated from a uniform probability distribution in the range $(0.5, 1.0)$, and the second one for which F is generated from a Gaussian probability distribution with parameters $N(0.5, 0.25)$ [7]. Also in [7], a schema where F is reduced from 1.0 to 0.5 is proposed. This schema, termed DETVSF (DE with time varying scale factor), aims at increasing the importance of the exploration in the initial cycles, and reinforcing the exploitation in the final ones.

In addition to randomly generated scaling factors, this parameter can be modified taking into account the best suited vectors from the previous generations, *self-adapting* variants [8–11]. Similarly to randomly generated, this mechanism aims at improving the overall

[1] For the sake of brevity, the symbol corresponding to the crossover operator has been omitted.

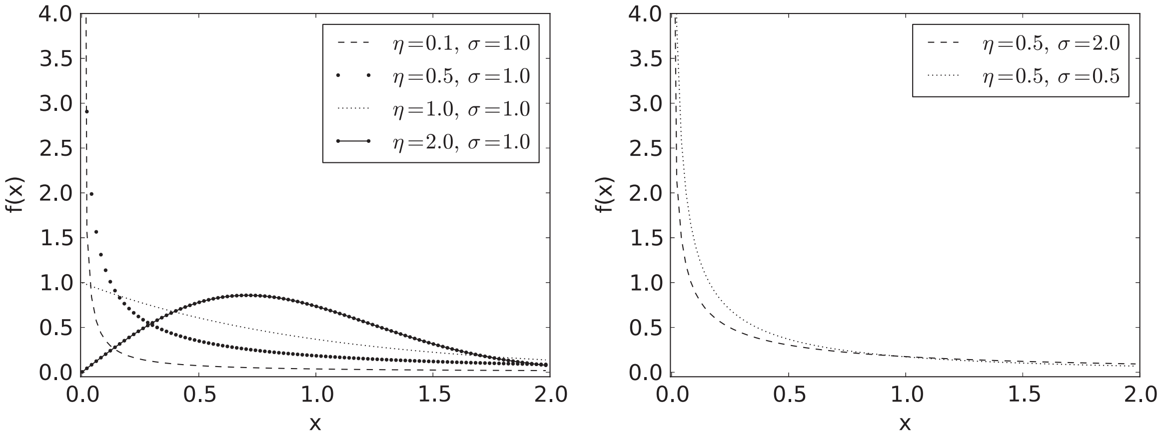


Fig. 1. Examples of the Weibull probability distribution for diverse values of shape and scale (Eq. (4)).

performance. Comparisons with self-adapting variants are not considered in this work, and are proposed as Future Work.

The efforts in self-adapting variants arise from the initial studies about the importance of the control parameters for the final performance of the variants [12], and specially due to the dispersion of the values for these parameters. In [1] a value for the mutation factor of F=0.5 is proposed. In [13], authors proposed mutation factor in the range $0.5 < F < 0.95$ with F=0.9 as initial choice.

Finally, two examples of the hybridization of probability distributions and DE variants are cited. An example of incorporation of Gaussian probability distribution to an operator of DE is presented in [14]. This work aims at modifying the Compact Differential Evolution (cDE) [15] without losing performance, but reducing the computational intensity. cDE belongs to the category of Estimation Distribution Algorithms. In this proposal the mutation operation is performed by sampling the vectors from a truncated Gaussian probability distribution. On the other hand, in [16] an analysis of the probability distribution of mutant vectors under the schema DE/rand/1 is performed. From this work, a new DE variant is proposed. It samples the mutant vectors from a beta distribution.

In the previous works, multiple DE-variants have been presented. From the original ones with a defined strategy for generating the trial vectors and static values for the parameters of DE, to the DE-variants with non-static-parameters, all of them use a unique strategy for generating the trial vectors.

Some works have explored the DE performance when combining several trial-vector generation strategies, namely multi-strategy variants. An example of this kind of strategy can be found in CoDE, composite DE [17]. This method uses three trial-vector generation strategies: DE/rand/1/bin, DE/rand/2/bin, and DE/current-to-rand/1, and three control parameter settings: $F = 1.0$, $C_r = 0.1$; $F = 1.0$, $C_r = 0.9$; and $F = 0.8$, $Cr = 0.2$. In the DE/current-to-rand/1 strategy, $\vec{u_i} = \vec{x_i} + F\cdot(\vec{x_1} - \vec{x_i}) + F\cdot(\vec{x_1} - \vec{x_2})$, the binominal crossover operator is not applied. In each generation and for each target vector, the tree strategies are used for generating three trial vectors. Each strategy uses a randomly-selected parameter setting from the pool. After the evaluation of the three trial vectors, the best one is retained and confronted to the target vectors. As a result of the multiple search strategies and parameters, the search ability of these algorithms is improved.

In Ensemble of Mutation and Crossover Strategies and Parameters (EPSDE), a pool of mutation, crossover strategies, and parameter settings are used along the whole evaluation to produce at each cycle high-quality offspring [18]. Differently to CoDE, EPSDE learns good combinations from evolution. This variant learns from the combination which performs better in the previous cycles, for promoting it in the next cycles. Whereas in CoDE three trial vectors are generated per cycle and target vector, in EPSDE only one trial vector is generated per cycle and target vector. Finally, in CoDE the control parameter settings are static, whereas in EPSDE F is chosen in the range from 0.4 to 0.9 with step-size 0.1 and C_r is chosen from 0.1 to 0.9 with step-size 0.1. For increasing the efficiency of the parameter setting, an improved version of EPSDE via the optimization of these parameters with Harmony Search is proposed [19].

In [20] an adaptive DE with a pool of 4 mutation strategies is proposed. This pool includes DE/current-to-pbest/1 and DE/rand-to-pbest/1 strategies. These strategies select the best candidate between a subset randomly selected from the whole population. For each vector, a particular mutation strategy is selected from the pool. Besides, the values of the parameters are adapted via the JADE algorithm [21].

In [22], the authors incorporate the distribution of the information provided by the population in the optimization process of DE. In the paper, the cumulative population distribution information of DE is used for establishing an Eigen coordinate system by making use of covariance matrix adaptation. In practice, this mechanism creates two trial vectors: the first one on the original coordinate system, and the second one on the Eigen coordinate system. Later, the target vector is compared with these two trial vectors, promoting the best one in the next generation.

2.2. Benchmark functions

In order to evaluate the efficiency to produce high-quality solutions, a wide set of fitness functions are used (Table 1). This set includes both multimodal and monomodal, separable and non-separable functions. These fitness functions have been used in CEC contests [23,24] and also in works where presenting cutting-edge DE-variants [25,26,8,28,27].

In all the tests performed in this work, two dimensionality configurations: 30 and 100, and a population size of 10 members have been used. Two configurations are used for the number of cycles: 100 and 1000. Generally, a larger number of cycles leads to higher-quality solutions. However, it is specially interesting when these high-quality solutions are produced with few cycles. For this reason, in this work the executions are performed with a maximum of 1000 cycles. As pseudorandom number generator, a subroutine based on Mersenne Twister [29] has been used.

In Table 2, the efficiency of the DE-variants is shown. For this initial numerical experiments, 25 independent runs are executed per DE-variant and fitness function, with mutation factor and crossover probability, $F = Cr = 0.5$. Since, this work aims at modelling the relative efficiency of the variants, these values are appropriate for this purpose. Higher values of F are more appropriate for non-separable function [13], but not for separable ones.

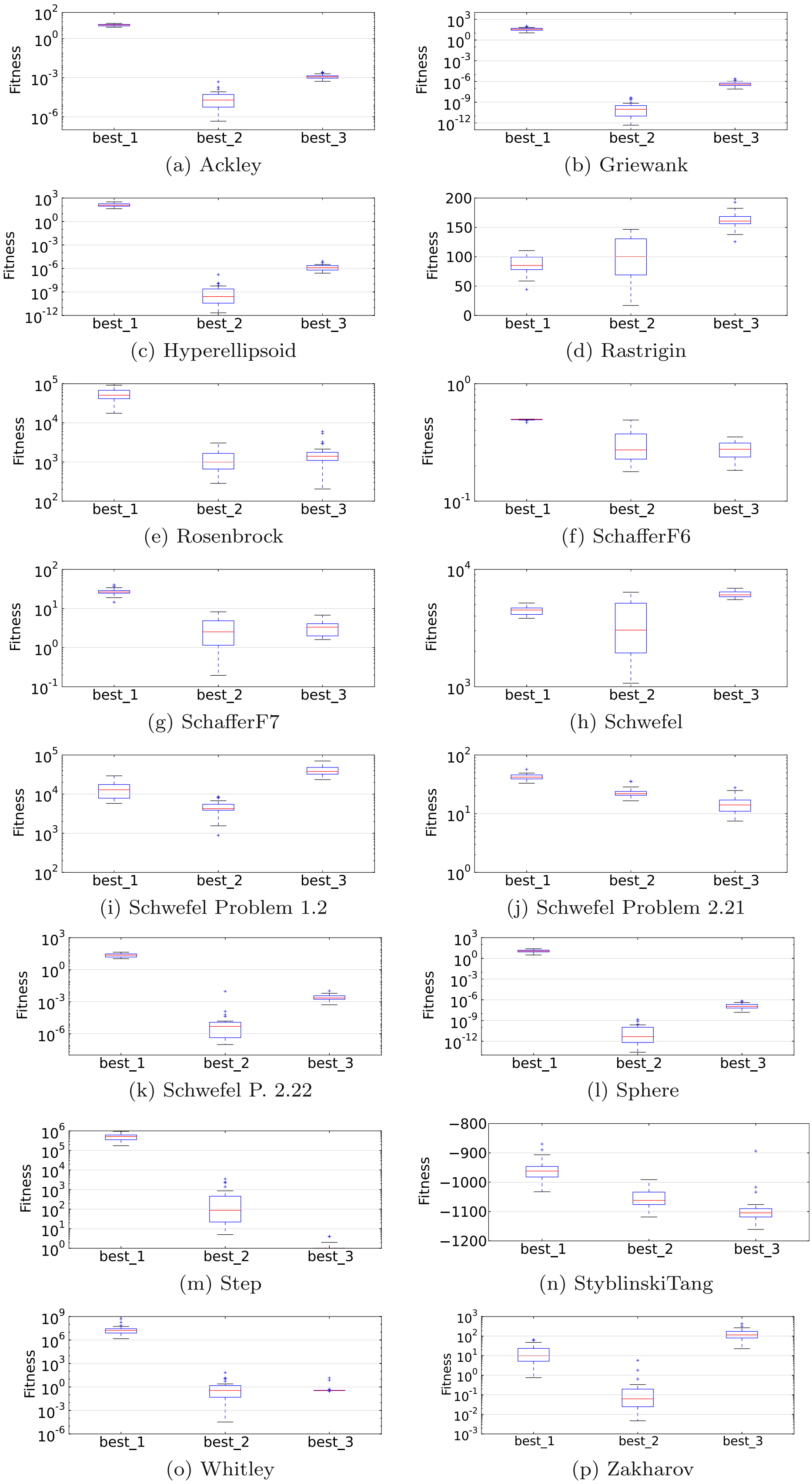


Fig. 2. Fitness comparison for the variants DE/best/1, DE/best/2, and DE/best/3. The box plots have been generated with 25 independent runs with configuration 1000 cycles and $F = Cr = 0.5$.

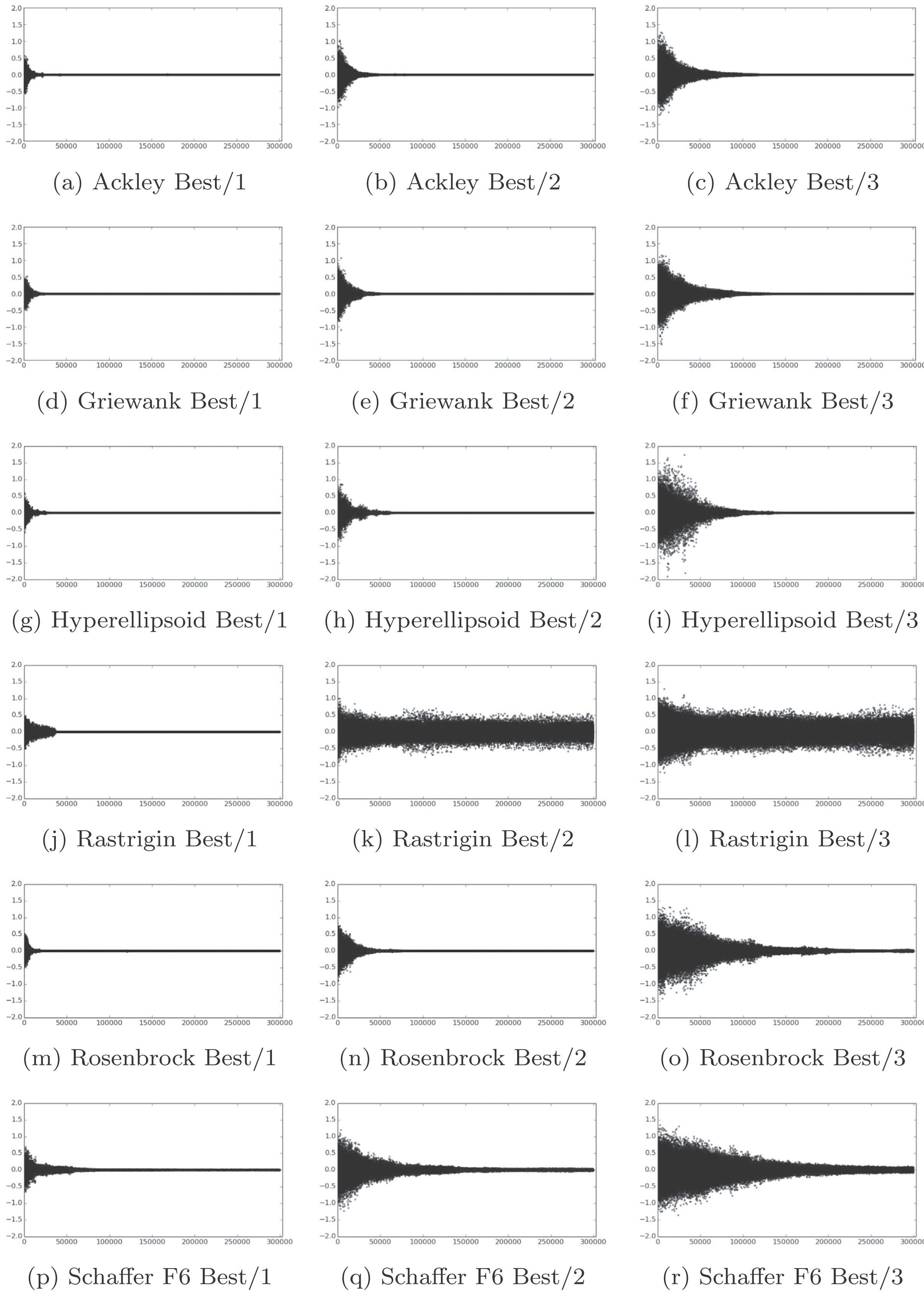
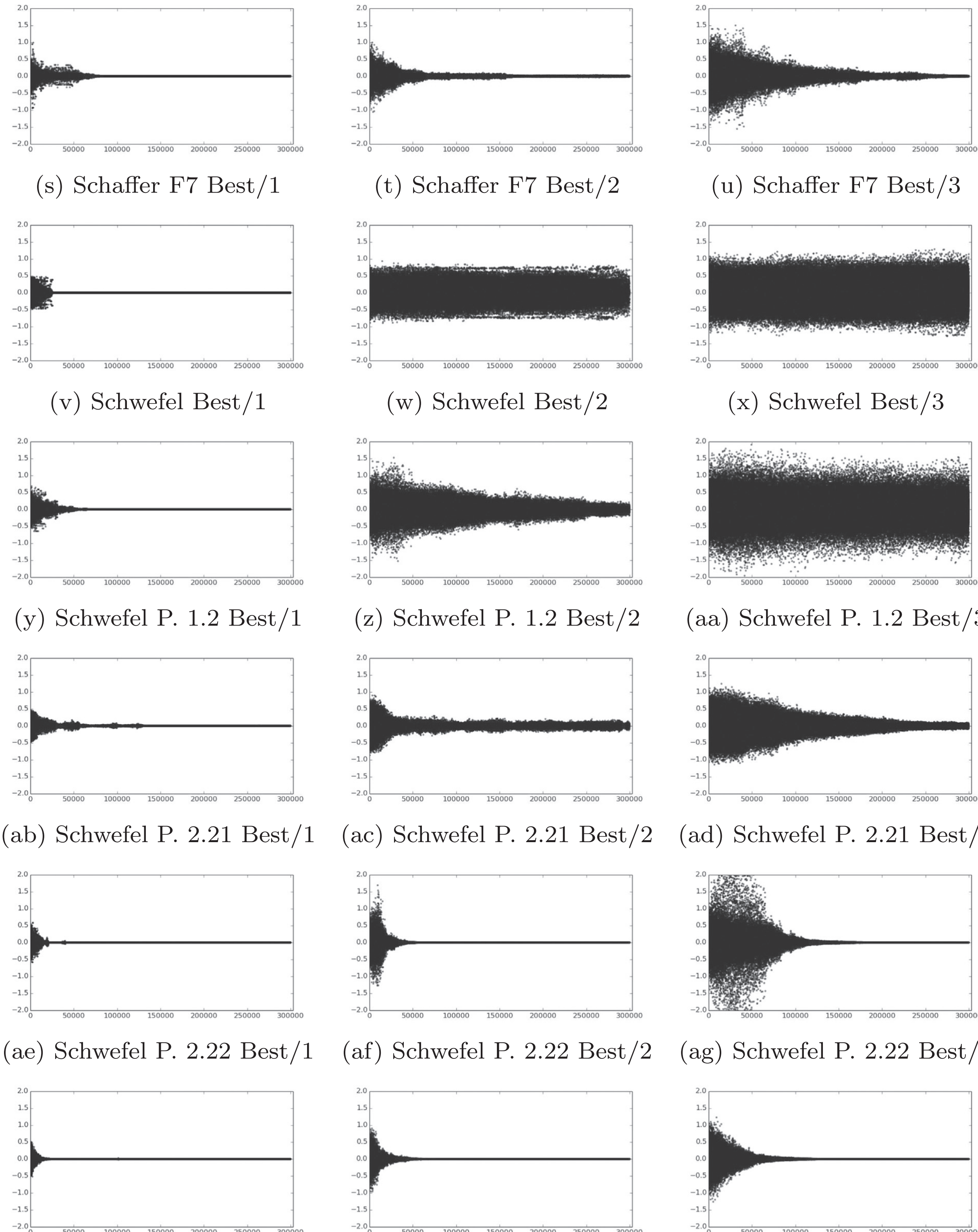


Fig. 3. Scaled differences for DE variants DE/best/1, DE/best/2, and DE/best/3 for 1000 cycles and a population of 10 vectors in a search space of 30.



(s) Schaffer F7 Best/1 (t) Schaffer F7 Best/2 (u) Schaffer F7 Best/3

(v) Schwefel Best/1 (w) Schwefel Best/2 (x) Schwefel Best/3

(y) Schwefel P. 1.2 Best/1 (z) Schwefel P. 1.2 Best/2 (aa) Schwefel P. 1.2 Best/3

(ab) Schwefel P. 2.21 Best/1 (ac) Schwefel P. 2.21 Best/2 (ad) Schwefel P. 2.21 Best/3

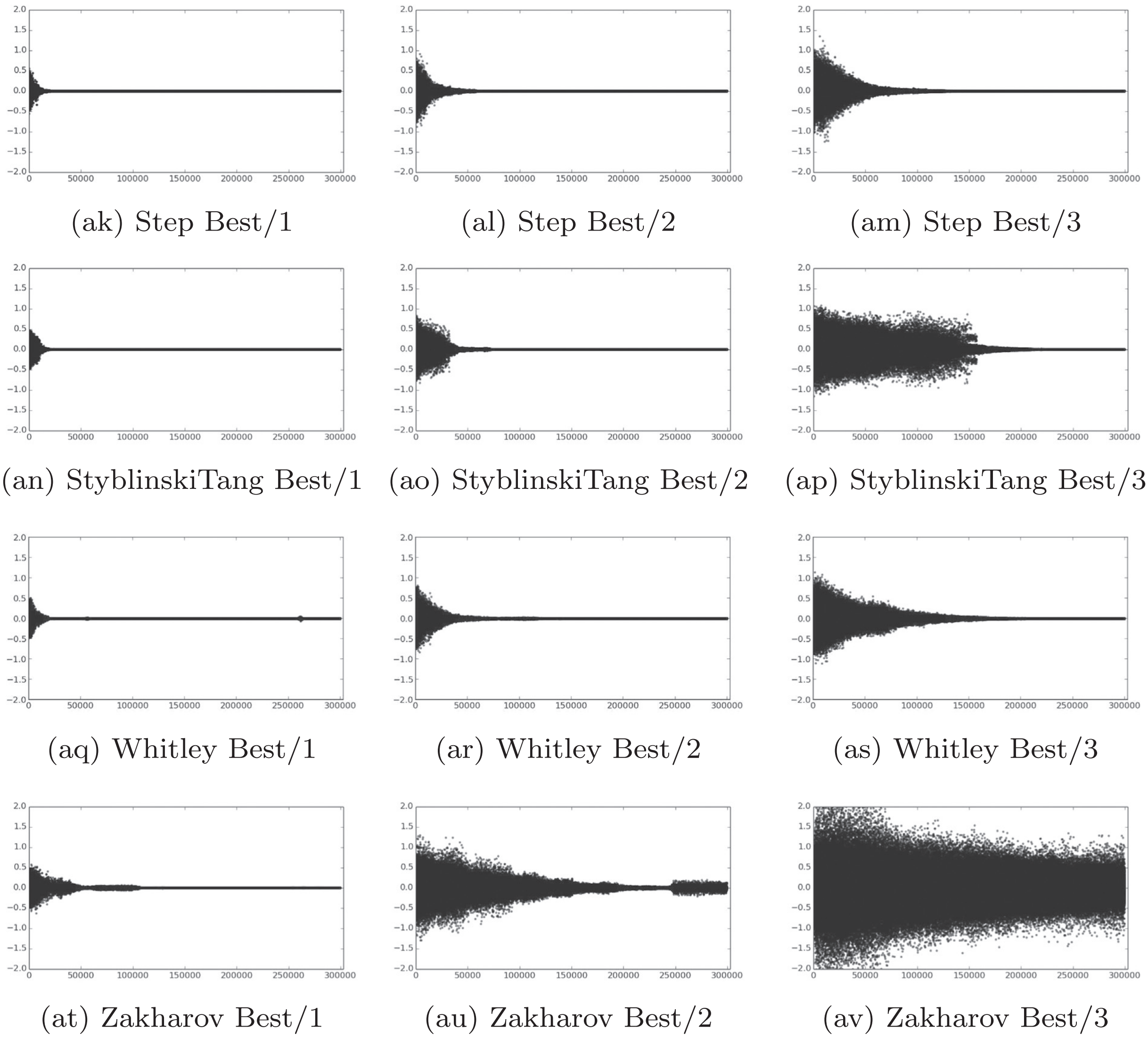
(ae) Schwefel P. 2.22 Best/1 (af) Schwefel P. 2.22 Best/2 (ag) Schwefel P. 2.22 Best/3

(ah) Sphere Best/1 (ai) Sphere Best/2 (aj) Sphere Best/3

Fig. 3. Continued.

2.3. Weibull probability distribution

The Weibull probability distribution is a continuous distribution based on two parameters: shape η and scale σ (Eq. (4)) [30]. The Weibull distribution is identical to the Exponential probability distribution when $\eta = 1$, and the Rayleigh probability distribution for $\eta = 2$. A representation of the Weibull distribution for diverse values of shape and scale parameters is shown in Fig. 1. The modification of the



(ak) Step Best/1 (al) Step Best/2 (am) Step Best/3

(an) StyblinskiTang Best/1 (ao) StyblinskiTang Best/2 (ap) StyblinskiTang Best/3

(aq) Whitley Best/1 (ar) Whitley Best/2 (as) Whitley Best/3

(at) Zakharov Best/1 (au) Zakharov Best/2 (av) Zakharov Best/3

Fig. 3. Continued.

scale factor varies the size of the figure, but not its shape.

$$f(x;\eta,\sigma) = \frac{\eta}{\sigma}\cdot\left(\frac{x}{\sigma}\right)^{\eta-1}\cdot e^{-(\frac{x}{\sigma})^{\eta}} \tag{4}$$

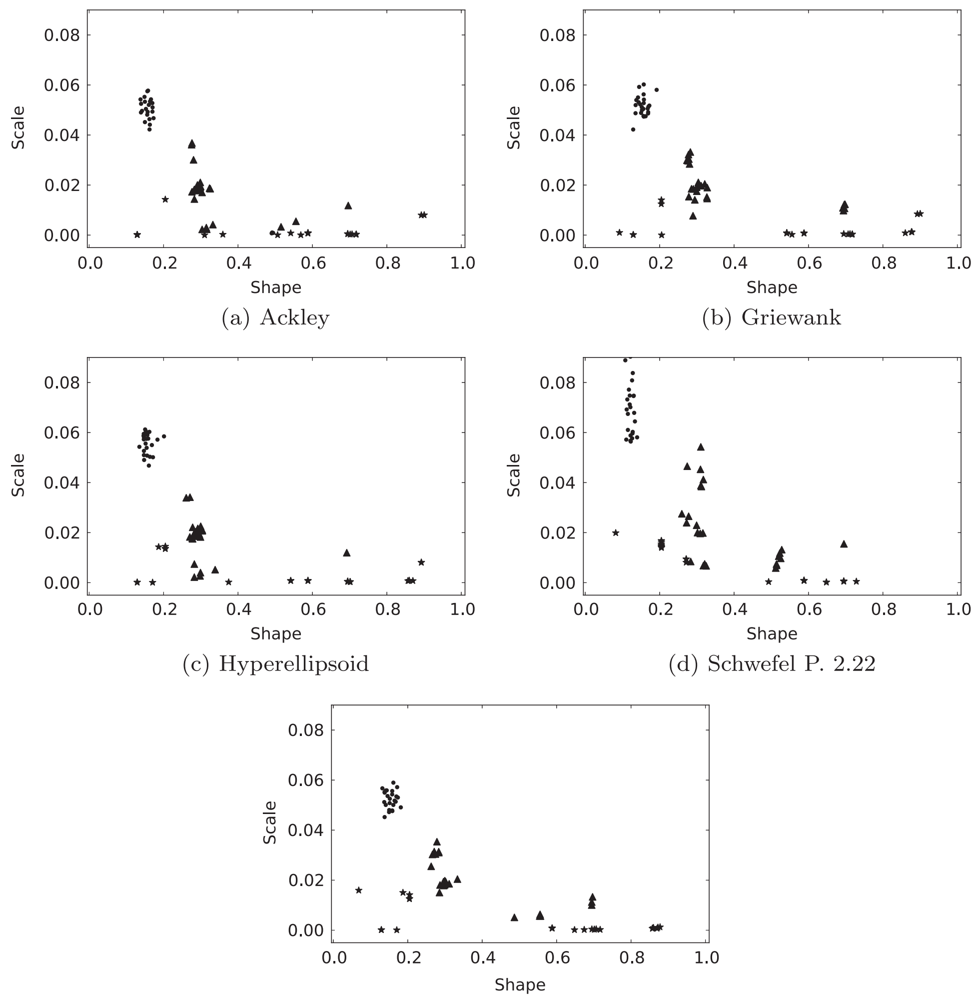
A typical application of the Weibull distribution is to model the lifetime of components. However, in this work it is used to model the evolution, along the cycles of the optimization process, of the scaled differences, $F\cdot(\overrightarrow{x_2} - \overrightarrow{x_3})$, of DE algorithm. The election of this probability distribution holds on its flexibility to reproduce very different curves in function of its parameters: shape and scale.

2.4. Statistics

Statistical inference is used in this work to infer which strategy produces better results, and if the differences are significant or not. Two types of tests can be applied to these cases: parametric and non-parametric. The differences between both rely on the assumption that data is normally distributed for parametric tests, whereas non-explicit conditions are assumed in non-parametric tests. For this reason, the latter is recommended when the statistical model of the data is unknown [32,31].

The Kruskal-Wallis test is a non-parametric test used to compare three or more groups of sample data [33]. For this test, the null hypothesis assumes that the samples are from identical populations. The procedure when using multiple comparison to test whether the null hypothesis is rejected implies the use of a post-hoc test to determine which sample makes the difference. The most typical post-hoc test is the Wilcoxon signed-rank test with the Bonferroni or Holm correction [31].

The Wilcoxon signed-rank test belongs to the non-parametric category. It is a pairwise test that aims to detect significant differences between two sample means, that is, the behaviour (fitness values achieved by the DE-variants in our study) of two DE-variants. If necessary, the Bonferroni or Holm correction can be applied to control the Family-Wise Error Rate (FWER). In our case, the Bonferroni correction for this purpose. FWER is the cumulative error when more than one pairwise comparison (e.g. more than one Wilcoxon signed-rank test) is performed. Therefore, when multiple pairwise comparisons are performed, Bonferroni correction allows maintaining the control over the FWER.



(a) Ackley (b) Griewank (c) Hyperellipsoid (d) Schwefel P. 2.22 (e) Sphere

Fig. 4. Weibull parameters when modelling the scaled differences of DE-best variants DE/best/1 (star), DE/best/2 (dot), and DE/best/3 (triangle). A total of 25 independent runs per case has been executed.

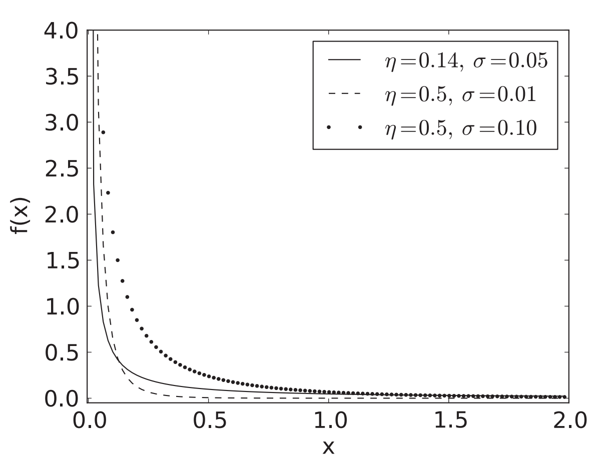


Fig. 5. Examples of Weibull probability distribution for diverse values of parameters shape and scale.

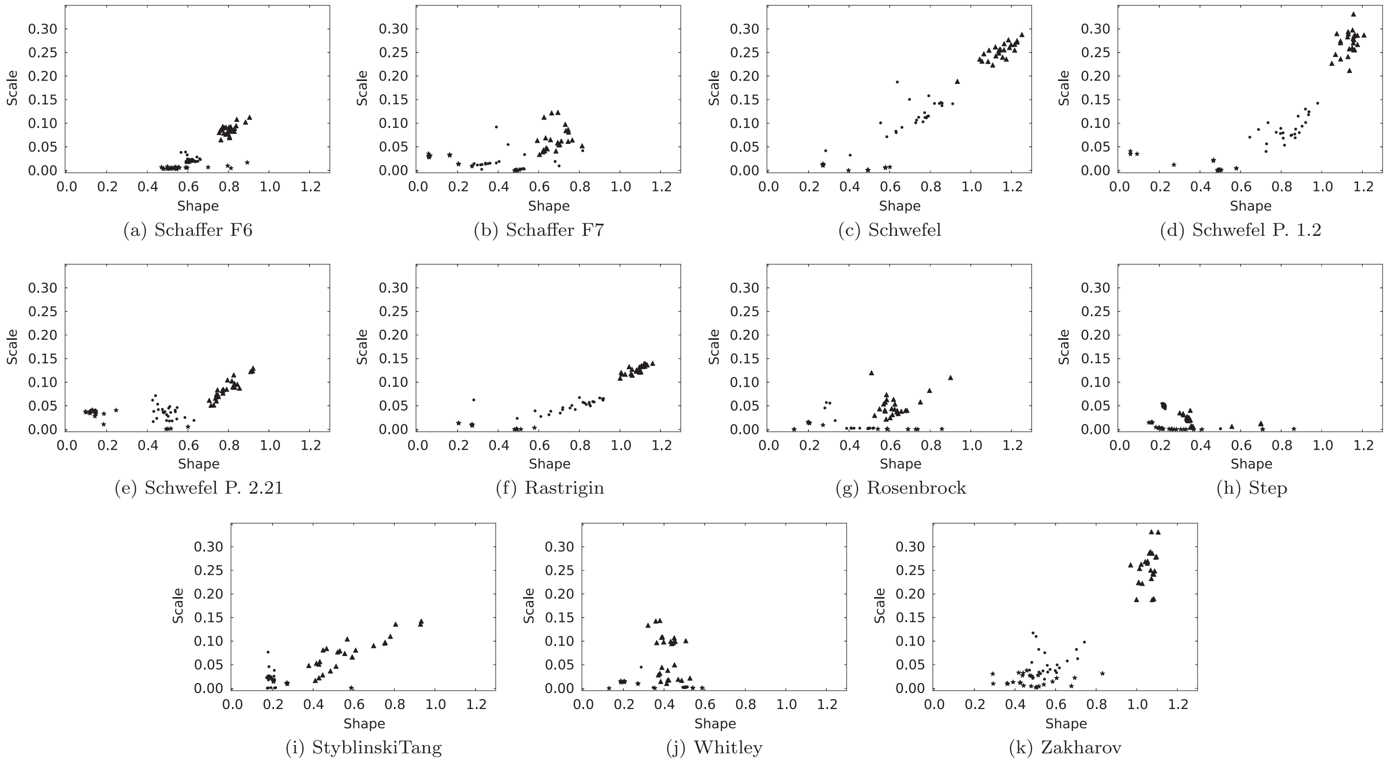
3. Proposed variant and experimental study

In this section, the scaled differences of DE-variants are modelled through a Weibull probability distribution. It is observed that high-performance DE-variants exhibit similar values of shape and scale. The values of these parameters are useful for separating by their performance the DE-variants analysed.

3.1. Comparison among DE/best/Y variants

Numerous DE-variants have been developed (see Section 2.1) for improving the performance of the DE/rand/1/bin. In Table 2, a review of the performance of a set of the most common variants has been presented. As can be observed DE/best/2 produces the largest number of best results. Specially relevant is the number of cases where DE/best/2 outperforms DE/best/1. In DE/best/1, the scaled difference of two vectors is added to the best vector in the current generation, whereas DE/best/2 simply uses four vectors for the scaled differences. Since in both variants the base vector is identical, the best vector in the current generation, the capacity to produce better solutions of DE/best/2 should be relied on the scaled differences. To complete this study, the results of DE/best/3 have been included (Fig. 2).

The box plots in Fig. 2 present the fitness of the variants DE/best/1, DE/best/2, and DE/best/3 after 25 independent runs with a configuration for the mutation factor and the crossover probability, $F = Cr = 0.5$, and 1000 cycles. As can be observed, in some cases DE/best/2 critically outperforms the two other variants: Ackely (Fig. 2a), Griewank (Fig. 2b), Hyperellipsoid (Fig. 2c), Schwefel Problem 2.22 (Fig. 2k), and Sphere (Fig. 2l). In these cases, the



(a) Schaffer F6 (b) Schaffer F7 (c) Schwefel (d) Schwefel P. 1.2 (e) Schwefel P. 2.21 (f) Rastrigin (g) Rosenbrock (h) Step (i) StyblinskiTang (j) Whitley (k) Zakharov

Fig. 6. Weibull parameters when modelling the scaled differences in DE variants DE/best/1 (star), DE/best/2 (dot), and DE/best/3 (triangle). A total of 25 independent runs per case has been executed.

Table 3
Mean value and standard deviation for 25 independent runs of the parameters shape and scale when modelling the scaled differences with a Weibull probability distribution. The functions are separately presented for those which DE/best/2 outperforms the other variants.

Function	DE/Best/1		DE/Best/2		DE/Best/3	
	Shape	Scale	Shape	Scale	Shape	Scale
Ackley	0.530 ± 0.231	0.002 ± 0.003	0.197 ± 0.109	0.045 ± 0.017	0.332 ± 0.100	0.017 ± 0.009
Griewank	0.558 ± 0.253	0.002 ± 0.004	0.155 ± 0.014	0.052 ± 0.004	0.376 ± 0.160	0.019 ± 0.007
Hyperellipsoid	0.548 ± 0.261	0.003 ± 0.005	0.157 ± 0.013	0.056 ± 0.004	0.306 ± 0.080	0.018 ± 0.008
Schwefel P. 2.22	0.446 ± 0.216	0.006 ± 0.007	0.122 ± 0.008	0.070 ± 0.012	0.378 ± 0.118	0.021 ± 0.015
Sphere	0.548 ± 0.281	0.003 ± 0.005	0.154 ± 0.013	0.052 ± 0.004	0.379 ± 0.148	0.019 ± 0.009
Schaffer F. 6	0.571 ± 0.109	0.006 ± 0.003	0.615 ± 0.024	0.022 ± 0.006	0.805 ± 0.035	0.087 ± 0.011
Schaffer F. 7	0.250 ± 0.177	0.017 ± 0.014	0.447 ± 0.138	0.017 ± 0.020	0.684 ± 0.056	0.067 ± 0.025
Schwefel	0.397 ± 0.127	0.007 ± 0.005	0.716 ± 0.142	0.114 ± 0.035	1.142 ± 0.070	0.252 ± 0.020
Schwefel P. 1.2	0.438 ± 0.146	0.007 ± 0.012	0.838 ± 0.085	0.089 ± 0.025	1.136 ± 0.037	0.270 ± 0.025
Schwefel P. 2.21	0.203 ± 0.146	0.030 ± 0.013	0.498 ± 0.049	0.036 ± 0.014	0.795 ± 0.061	0.086 ± 0.021
Ratrigin	0.349 ± 0.117	0.007 ± 0.005	0.757 ± 0.147	0.050 ± 0.012	1.085 ± 0.042	0.127 ± 0.008
Rosenbrock	0.410 ± 0.238	0.006 ± 0.006	0.449 ± 0.075	0.009 ± 0.017	0.629 ± 0.083	0.051 ± 0.024
Step	0.325 ± 0.197	0.003 ± 0.005	0.243 ± 0.076	0.046 ± 0.014	0.379 ± 0.104	0.019 ± 0.010
StyblinskiTang	0.273 ± 0.098	0.011 ± 0.004	0.189 ± 0.013	0.022 ± 0.016	0.579 ± 0.161	0.075 ± 0.034
Whitley	0.294 ± 0.153	0.010 ± 0.006	0.488 ± 0.052	0.004 ± 0.008	0.418 ± 0.048	0.070 ± 0.045
Zakharov	0.487 ± 0.121	0.017 ± 0.011	0.562 ± 0.075	0.052 ± 0.027	1.055 ± 0.034	0.259 ± 0.042

improvement achieves several orders of magnitude. However, in other cases, the improvement is not relevant, e.g. Schaffer Function 7 (Fig. 2g), and Zakharov (Fig. 2p); or simply DE/best/2 is outperformed by the other DE-variants, Schwefel Problem 2.21 (Fig. 2j).

In order to understand this differential behaviour, the differences for all the coordinates of the vectors, vectors of the population and cycles for the three variants under study are plotted (Fig. 3). These plots do not exactly correspond with the temporal evolution of the difference, though larger values in x-axis roughly indicate more advanced cycle. In all plots, x-axis ranges from 0 to 300,000, corresponding to the product of 1000 cycles, times 10 vectors, times 30 dimensions; while y-axis ranges from − 2 to 2.

Regarding the cases where DE/best/2 outperforms DE/best/1 and DE/best/3, some visual similarities are observed in the evolution of the scaled differences. For Ackley function (Fig. 3a–c), the initial dispersion of differences of DE/best/2 is larger than DE/best/1, but shorter than DE/best/3. Beside, its convergence to low values for the scaled differences is slower than DE/best/1, but faster than DE/best/3. Apparently, the combination of an intermediated initial dispersion for exploring the search space, together with a fast convergence to exploit the most promising solution leads to a higher efficiency. This combined behaviour is also observed for the other functions where DE/best/2 critically outperforms the other DE/best variants: Ackley function (Fig. 3a–c), Griewank (Fig. 3d–f,), Hyperellipsoid (Fig. 3g–i), Schwefel Problem 2.22 (Fig. 3ae–ag), and Sphere (Fig. 3ah–aj).

Oppositely, for functions where DE/best/2 does not take any

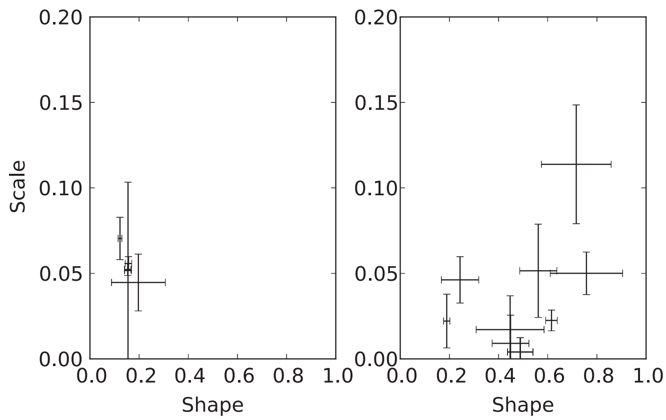


Fig. 7. Mean values and standard deviations of the shape and scale when modelling the scaled difference of DE/best/2 with a Weibull probability distribution. In left panel the values for the functions Ackely, Griewank, Hyperellipsoid, Schwefel Problem 2.22, and Sphere; whereas the remaining ones are plotted in right panel.

Table 4
Comparison of DE/best/binweibull/bin variant with best results of the previous DE-variants. 25 independent runs per case with $F = Cr = 0.5$, for two configurations for the number of cycles: 100 and 1000, 30 dimensions, a population size of 10 vectors, and Binomial Crossover operator.

Function	Cycles	Best previous		DE/Best/ binweibull
Ackley	100	DE/best/2	6.83 ± 1.46	$\mathbf{0.37 \pm 0.02}$
	1000	DE/best/2	$\mathbf{(5.1 \pm 9.7)10^{-5}}$	0.30 ± 0.02
Griewank	100	DE/best/2	17.19 ± 9.88	$\mathbf{(6.8 \pm 0.6)10^{-5}}$
	1000	DE/best/2	$\mathbf{(0.7 \pm 1.3)10^{-9}}$	$(5.0 \pm 0.2)10^{-5}$
Hyperellipsoid	100	DE/best/2	48.88 ± 22.25	$\mathbf{3.65 \pm 0.29}$
	1000	DE/best/2	$\mathbf{(0.8 \pm 3.1)10^{-8}}$	2.56 ± 0.19
Rastrigin	100	DE/best/1	115.47 ± 35.91	$\mathbf{45.39 \pm 3.52}$
	1000	DE/rand/2/dir	52.29 ± 19.77	$\mathbf{33.56 \pm 1.84}$
Rosenbrock	100	DE/best/2	$(48 \pm 20)10^{3}$	$\mathbf{66.4 \pm 7.6}$
	1000	DE/best/2	$(1.3 \pm 0.8)10^{3}$	$\mathbf{51.4 \pm 3.1}$
Schaffer's F6	100	DE/rand-to-best/1	$(4925 \pm 58)10^{-4}$	$\mathbf{0.0097 \pm 10^{-6}}$
	1000	DE/rand/1	0.27 ± 0.11	$\mathbf{1 \cdot 10^{-2} \pm 8 \cdot 10^{-9}}$
Schaffer's F7	100	DE/rand-to-best/1	25.61 ± 0.80	$\mathbf{0.25 \pm 0.02}$
	1000	DE/rand/1	1.22 ± 1.46	$\mathbf{0.19 \pm 0.02}$
Schwefel	100	DE/best/1	$\mathbf{5130 \pm 734}$	9903 ± 615
	1000	DE/rand/1	$\mathbf{1282 \pm 977}$	7804 ± 863
Schwefel P. 1.2	100	DE/current-to-best/1	$12,766 \pm 4593$	$\mathbf{0.56 \pm 0.14}$
	1000	DE/best/2	4621 ± 1931	$\mathbf{0.30 \pm 0.05}$
Schwefel P. 2.21	100	DE/current-to-best/1	34.07 ± 6.51	$\mathbf{0.18 \pm 0.02}$
	1000	DE/rand/2	22.94 ± 6.96	$\mathbf{0.129 \pm 0.006}$
Schwefel P. 2.22	100	DE/rand-to-best/1	18.07 ± 7.00	$\mathbf{2.47 \pm 0.12}$
	1000	DE/best/2	$\mathbf{(0.3 \pm 1.8)10^{-3}}$	2.11 ± 0.14
Sphere	100	DE/rand/2/dir	1.99 ± 2.28	$\mathbf{0.26 \pm 0.03}$
	1000	DE/best/2	$\mathbf{(1.7 \pm 3.5)10^{-10}}$	0.18 ± 0.02
Step	100	DE/best/2	$(0.16 \pm 0.07)10^{6}$	$\mathbf{0.0 \pm 0.0}$
	1000	DE/rand/2	1.12 ± 1.63	$\mathbf{0.0 \pm 0.0}$
StyblinskiTang	100	DE/rand-to-best/1	$\mathbf{-947 \pm 41}$	-771 ± 57
	1000	DE/rand/1	$\mathbf{-1111 \pm 26}$	-1001 ± 43
Whitley	100	DE/best/2	$(5 \pm 4)10^{6}$	$\mathbf{1.57 \pm 0.22}$
	1000	DE/rand/2	1.25 ± 1.78	$\mathbf{1.01 \pm 0.05}$
Zakharov	100	DE/rand-to-best/1	93.7 ± 38.3	$\mathbf{0.30 \pm 0.04}$
	1000	DE/best/2	0.41 ± 1.14	$\mathbf{0.20 \pm 0.014}$

advantage over the other variants, the patter observed is clearly different. For example, for Rastrigin function, the scaled differences of DE/best/2 exhibit a lack of convergence as the cycles progress (Fig. 3k). This behaviour impedes an efficient exploitation of the most promising solutions. A similar lack of convergence toward low scaled differences is observed for Schwefel (Fig. 3w) function and the Schwefel Problem 1.2 (Fig. 3z).

Other cases, for example Whitley function (Fig. 2o) and the Step function (Fig. 2m), the visual analysis induces to expect a higher performance than the finally obtained one. In these functions, an in-depth analysis is necessary to ascertain the differential behaviour in relation to the cases of high performance. In order to undertake this analysis, the scaled differences are modelled as Weibull probability distribution.

3.2. A statistical model based on Weibull probability distribution

Until this point, it has been observed how the differential performance of the DE/best-variants is reflected in the scaled differences generated as far as the cycles of DE progress. In order to analyse the scaled differences, the absolute value of the differences is modelled as a Weibull probability distribution.

In Fig. 4, the values of the shape and scale for the benchmark functions where DE/best/2 outperforms the other DE/best-variants are represented. The scaled differences have been modelled as a Weibull probability distribution, and the parameters of the distribution extracted: shape and scale. The DE/best/1 distribution parameters are labelled as star, DE/best/2 as dot, and DE/best/3 as triangle.

As illustrated by Fig. 4, the values of shape and scale obtained of DE/best/2 extend in a reduced area, $0.12 < shape < 0.20$ and $0.05 < scale < 0.07$. Oppositely, the values of these parameters for the other variants extend in much larger areas. In general, the values of the shape are larger than the ones of DE/best/2, at the same time the scale parameter is lower. This combination for the shape and scale parameters leads to premature convergence: excessive low values for the generated scaled differences lead to a poor exploration of the search space (Fig. 5). Otherwise, larger values of the scale for a similar shape value lead to slower convergence: too large scaled differences are generated, so that they make difficult an intensive exploitation of the candidate solutions. Both scenarios conclude in a low-performance for these variants.

Regarding the values of shape and scale for the functions where the performance of DE/best/2 is not critically better than the other DE/best-variants (Fig. 6), completely different patterns to those previously analysed are observed. As illustrated by Fig. 6, and conversely to Fig. 4, the shape and scale values for DE/best/2 are not longer constrained to a reduced area, but they scatter in a larger area and mix with the parameters of the other DE/best-variants. They are not longer separable from the others variants.

The numerical results of shape and scale are shown in Table 3. The values of the shape for the functions Ackely, Griewank, Hyperellipsoid, Schwefel Problem 2.22, and Sphere are in the range $0.12 < shape < 0.20$; whereas the values for the scale are in the range $0.05 < scale < 0.07$. These functions are separately presented since they correspond to the fitness functions where DE/best/2 critically outperforms the other variants.

In the previous section the case of the Step function was mentioned. Visually, the relative performance of the DE/best-variants (Fig. 2m) is not similar to others where DE/best/2 outperforms the two other DE/best-variants, but the scaled differences seem similar to those cases (Fig. 3al). In spite of a numerical value of shape in the range of high-performance cases, its value of the scale is much lower, 0.022 ± 0.016, than the range of the high-performance case $0.05 < scale < 0.07$. For this reason, it is not considered as part of the high-performance cases of DE/best/2.

Table 5
Comparison of DE/best/binweibull/bin variant with best results of the previous DE-variants. 25 independent runs per case with $F = Cr = 0.5$, for two configurations for the number of cycles: 100 and 1000, 100 dimensions, a population size of 10 vectors, and Binomial Crossover operator.

Function	Cycles	Best previous		DE/Best/binweibull
Ackley	100	DE/rand-to-best/1	15.853 ± 0.592	**0.568 ± 0.027**
	1000	DE/best/2	4.820 ± 1.614	**0.473 ± 0.014**
Griewank	100	DE/rand-to-best/1	456.84 ± 58.406	**$(52.6 \pm 5.3)10^{-5}$**
	1000	DE/best/2	20.12 ± 12.58	**$(36.4 \pm 2.5)10^{-5}$**
Hyperellipsoid	100	DE/current-to-best/1	5973 ± 1114	**94 ± 11**
	1000	DE/best/2	253 ± 254	**64 ± 7**
Rastrigin	100	DE/best/1	701.23 ± 87.60	**256.83 ± 12.56**
	1000	DE/best/2	342.29 ± 78.96	**215.81 ± 9.40**
Rosenbrock	100	DE/current-to-best/1	$(668.31 \pm 103.21)10^{3}$	**456.68 ± 46.46**
	1000	DE/best/2	$(59.71 \pm 26.58)10^{3}$	**369.51 ± 21.41**
Schaffer's F6	100	DE/current-to-best/1	$0.49992 \pm 3{\cdot}10^{-5}$	**$0.00972.1 \pm 1.1{\cdot}10^{-5}$**
	1000	DE/rand/1	0.4996 ± 0.0002	**$0.00971593 \pm 3{\cdot}10^{-8}$**
Schaffer's F7	100	DE/rand-to-best/1	60.2107 ± 6.314	**0.385 ± 0.0212**
	1000	DE/rand/1	21.776 ± 6.206	**0.331 ± 0.011**
Schwefel	100	DE/best/1	**28572 ± 2732**	36,786 ± 1022
	1000	DE/best/1	**23290 ± 1361**	30,621 ± 2113
Schwefel P. 1.2	100	DE/current-to-best/1	199,852 ± 43,829	**12.5 ± 4**
	1000	DE/current-to-best/1	92,434 ± 27,566	**7.4 ± 1.2**
Schwefel P. 2.21	100	DE/current-to-best/1	55.28 ± 5.41	**0.48 ± 0.05**
	1000	DE/current-to-best/1	54.02 ± 4.06	**0.32 ± 0.03**
Schwefel P. 2.22	100	DE/current-to-best/1	175.40 ± 21.62	**11.37 ± 0.56**
	1000	DE/rand/1	19.90 ± 6.54	**10.23 ± 0.29**
Sphere	100	DE/rand-to-best/1	129.8 ± 15.7	**2.0 ± 0.2**
	1000	DE/best/2	4.39 ± 3.04	**1.41 ± 0.12**
Step	100	DE/current-to-best/1	$(5.2 \pm 0.8)10^{6}$	**0.08 ± 0.27**
	1000	DE/best/3	$(0.3 \pm 0.2)10^{6}$	**0.0 ± 0.0**
StyblinskiTang	100	DE/current-to-best/1	**−2495 ± 101**	−1653 ± 141
	1000	DE/best/2	**−3324 ± 84**	−2803 ± 79
Whitley	100	DE/current-to-best/1	$(1.7 \pm 1.1)10^{9}$	**96 ± 22**
	1000	DE/rand-to-best/1	$(1.4 \pm 0.6)10^{9}$	**42.2 ± 7.4**
Zakharov	100	DE/rand-to-best/1	1111.94 ± 158.31	**2.52 ± 0.43**
	1000	DE/current-to-best/1	546.833 ± 125.627	**1.68 ± 0.16**

Graphically the values of parameters shape and scale of DE/best/2 for the two performance cases separate them in two disjoint groups. In Fig. 7, each point is elaborated with the mean values of shape and scale, and the uncertainties with the standard deviations. In left panel of Fig. 7, the high-performance cases are plotted, whereas the remaining ones are plotted in right panel. From Fig. 7, it can be seen that the high-performance cases behave differently that the remaining cases. They form a compact cluster, whereas the other cases spread in a much larger area.

In order to calculate the centre of the cluster formed by the mean values of the shape and scale of the cases where DE/best/2 variant outperforms the other variants (left panel of Fig. 7), the cost function of K-means algorithm (Eq. (5)) [34] appropriated modified to deal with the uncertainty of the values, can be used (Eq. (6)).

$$J(\vec{x_1}, \vec{x_2}, \ldots, \vec{x_n}, \vec{\mu_1}, \vec{\mu_2}, \ldots, \vec{x_m}) = \sum_k \sum_{\vec{x} \in \omega_k} (\vec{x} - \vec{\mu}(\omega_k))^2 \tag{5}$$

where k represents each cluster, and $\vec{x} \in \omega_k$ are the points belonging to the cluster k.

$$J(\vec{x_1}, \vec{x_2}, \ldots, \vec{x_n}, \vec{\mu_1}, \vec{\mu_2}, \ldots, \vec{x_m}) = \sum_k \sum_{\vec{x} \in \omega_k} \left(\frac{\vec{x} - \vec{\mu}(\omega_k)}{\sigma_{\vec{x}}} \right)^2 \tag{6}$$

The centre of the cluster (centroid) is obtained when minimizing the previous cost function of K-means (Eq. (7)).

$$\begin{aligned}
\frac{\partial J}{\partial \mu_{\vec{x}}} &= -2 \sum_{\vec{x} \in \omega_k} \left(\frac{\vec{x} - \vec{\mu}(\omega_k)}{\sigma^2_{\vec{x}}} \right) = 0 \Longrightarrow \\
\sum_{\vec{x} \in \omega_k} \frac{\vec{x}}{\sigma^2_{\vec{x}}} &= \sum_{\vec{x} \in \omega_k} \frac{\vec{\mu}(\omega_k)}{\sigma^2_{\vec{x}}} \Longrightarrow \\
\sum_{\vec{x} \in \omega_k} \frac{\vec{x}}{\sigma^2_{\vec{x}}} &= \vec{\mu}(\omega_k) \sum_{\vec{x} \in \omega_k} \frac{1}{\sigma^2_{\vec{x}}} \Longrightarrow \\
\vec{\mu}(\omega_k) &= \frac{1}{\sum_{\vec{x} \in \omega_k} \frac{1}{\sigma^2_{\vec{x}}}} \sum_{\vec{x} \in \omega_k} \frac{\vec{x}}{\sigma^2_{\vec{x}}}
\end{aligned} \tag{7}$$

The centroid coordinates, arisen from applying the Eq. (7) to the values of shape and scale, and their uncertainties, are $\eta = 0.14$ and $\sigma = 0.05$.

3.3. DE/best/binweibull variant

Based on the acquired knowledge, a new variant is proposed. In this variant, the base vector is the best vector in the current generation and the scaled differences are generated by the product of a binomial distribution for the sign of the difference, $(-1)^{B(0.5)}$, and a random number from a Weibull probability distribution of parameters $f(x; \eta = 0.14, \sigma = 0.05)$. The new variant is termed DE/best/binweibull. Since the Binomial Crossover operator is used, then the full descriptor of the variant is DE/best/binweibull/bin.

Random numbers based on Weibull probability distribution can be

Table 6
Comparison of DE/best/binweibull/bin variant with DE dither and jitter variants for 25 independent runs, and two configurations for the number of cycles: 100 and 1000, 30 dimensions, a population size of 10 vectors, and Binomial Crossover operator.

Function	Cycles	Dither uniform	Dither Gaussian	Jitter	DE/best/binweibull
Ackley	100	16.07 ± 1.08	8.49 ± 1.62	8.27 ± 1.22	**0.37 ± 0.02**
	1000	**0.015 ± 0.009**	4.65 ± 3.18	2.80 ± 2.28	0.30 ± 0.02
Griewank	100	190.70 ± 57.76	24.65 ± 10.45	25.27 ± 9.79	**(6.8 ± 0.6)10^{-5}**
	1000	**(4.3 ± 3.3)10^{-5}**	3.06 ± 2.60	2.55 ± 6.057	(5.0 ± 0.2)10^{-5}
Hyperellipsoid	100	559.09 ± 209.25	108.23 ± 41.46	89.37 ± 35.55	**3.65 ± 0.29**
	1000	**(14 ± 10)10^{-5}**	22.65 ± 26.50	3.87 ± 7.84	2.56 ± 0.19
Rastrigin	100	314.74 ± 26.72	207.32 ± 22.22	235.46 ± 33.70	**45.39 ± 3.52**
	1000	165.80 ± 19.95	**30.03 ± 8.20**	103.11 ± 22.95	33.56 ± 1.84
Rosenbrock	100	441,043 ± 112,688	80,855 ± 39,906	145,208 ± 61,860	**66.4 ± 7.6**
	1000	8621 ± 5030	8013 ± 7775	3438 ± 2235	**51.4 ± 3.1**
Schaffer's F6	100	0.49987 ± $8 \cdot 10^{-5}$	0.4994 ± 0.0003	0.4996 ± 0.0003	**0.0097 ± 10^{-6}**
	1000	0.38 ± 0.05	0.36 ± 0.10	0.26 ± 0.10	**0.0097 ± $8 \cdot 10^{-9}$**
Schaffer's F7	100	100.98 ± 10.47	43.62 ± 10.91	49.09 ± 9.36	**0.25 ± 0.02**
	1000	4.79 ± 2.01	4.25 ± 2.62	0.79 ± 0.59	**0.19 ± 0.02**
Schwefel	100	10,184 ± 492	**7949 ± 670**	9383 ± 495	9903 ± 615
	1000	7862 ± 453	**2459 ± 586**	4517 ± 1915	7804 ± 863
Schwefel Problem 1.2	100	126,024 ± 34,941	64,622 ± 12,833	76,447 ± 20,686	**0.56 ± 0.14**
	1000	67,049 ± 18,226	10,405 ± 5685	22,894 ± 10,063	**0.30 ± 0.05**
Schwefel Problem 2.21	100	81.11 ± 5.18	56.90 ± 8.84	68.72 ± 9.59	**0.18 ± 0.02**
	1000	23.41 ± 8.88	45.18 ± 9.13	37.63 ± 7.82	**0.129 ± 0.006**
Schwefel Problem 2.22	100	848.70 ± 1776.78	16.22 ± 3.58	29.26 ± 8.34	**2.47 ± 0.12**
	1000	**0.015 ± 0.007**	1.73 ± 2.35	0.36 ± 0.80	2.11 ± 0.14
Sphere	100	52.46 ± 18.05	7.80 ± 4.74	6.38 ± 2.04	**0.26 ± 0.03**
	1000	**(0.11e ± 9.14)10^{-6}**	0.98 ± 1.60	0.56 ± 1.01	0.18 ± 0.02
Step	100	2,029,877 ± 583,464	263,838 ± 120,059	235,486 ± 93,078	**0.0 ± 0.0**
	1000	1.36 ± 2.33	83,705 ± 128,632	16,079 ± 25,020	**0.0 ± 0.0**
Styblinski Tang	100	−570 ± 48	**−905 ± 48**	−744.9 ± 60.618	−771 ± 57
	1000	−1101 ± 73	−1072 ± 25	**−1104 ± 25**	−1001 ± 43
Whitley	100	(4.0 ± 3.0)10^{9}	(36.7 ± 67.9)10^{6}	(28.8 ± 22.7)10^{6}	**1.57 ± 0.22**
	1000	2.07 ± 3.43	(8.5 ± 20.2)10^{6}	0.24 ± 0.50)10^{6}	**1.01 ± 0.05**
Zakharov	100	683.84 ± 126.39	437.43 ± 148.26	427.25 ± 93.55	**0.30 ± 0.04**
	1000	290.00 ± 113.28	4.53 ± 7.60	5.14 ± 6.45	**0.202 ± 0.014**

produced by using Eq. (8), where ξ are numbers from a uniform probability distribution in the range (0, 1).

$$x = \sigma(-\ln\xi)^{\frac{1}{\eta}} \tag{8}$$

where ξ is a random number extracted from a uniform probability distribution in the range (0, 1), the parameters of the Weibull probability distribution are: $\eta = 0.14$ and $\sigma = 0.05$ and x is the scaled difference. Therefore the DE/best/binweibull variant for mutant vectors generation can be expressed as Eq. (9).

$$\vec{v_i} = \vec{x}_{best} + \sigma(-\ln\xi)^{\frac{1}{\eta}} \tag{9}$$

The pseudocode for the scaled difference under the schema DE/best/binweibull is shown in Algorithm 2.

Algorithm 2. Algorithm pseudocode for the scaled differences for the DE/best/binweibull variant.

foreach *Vector* **do**
  foreach *Dimension* **do**
    best vector component $+ (-1)^{B(0.5)} \times 0.05(-ln\xi)^{\frac{1}{0.14}}$

In Table 4, the results are compared with the previous best results of the static variants (Section 2.2). The configuration of this study is: 25 independent runs per case with $F = Cr = 0.5$, for two configurations for the number of cycles: 100 and 1000, two dimensionality configurations: 30 and 100, a population size of 10 vectors, and Binomial Crossover operator. As can be appreciated, DE/best/binweibull outperforms the other statics variants in 23 of 32 cases. Most of the confrontations, 14 cases, are versus DE/best/2. For these particular confrontations, DE/best/binweibull outperforms DE/best/2 in 10 cases. The second most frequent confrontation is versus DE/rand-to-best/2, 5 confrontations.

An in-depth analysis demonstrates that DE/best/binweibull specifically produces better results, 14 cases of 23, when using a short number of cycles in comparison other variants. This indicates that this variant greedily behaves, but falling in premature convergence.

For the functions Schwefel and StyblinskiTang, and independently of the number of cycles, DE/best/binweibull is outperformed by the other variants. Conversely, for 8 functions (Rastrigin, Rosenbrock, Schaffer's F6 and F7, Schwefel Problem 1.2 and 2.21, Whitley and Zakharov), DE/best/binweibull is the most suitable variant for both configurations of the number of cycles.

The statistical analysis using the Wilcoxon signed-rank test of the best result obtained with the independent runs (Table 4) indicates that in most of the cases the differences between the DE/best/binweibull variant and the other best-performing variant of the table are significant for a confidence level of 95% (p-value under 0.05). This means that the differences are unlikely to have occurred by chance with a probability of 95%. In only two cases the differences are not significant: for the functions Whitley ($p - value = 0.183$) and Zakharov

Table 7
Comparison of DE/best/binweibull/bin variant with DE dither and jitter variants for 25 independent runs, and two configurations for the number of cycles: 100 and 1000, 100 dimensions, a population size of 10 vectors, and Binomial Crossover operator.

Function	Cycles	Dither uniform	Dither Gaussian	Jitter	DE/best/binweibull
Ackley	100	20.56 ± 0.18	17.21 ± 0.66	19.47 ± 0.41	$\mathbf{0.57 \pm 0.03}$
	1000	8.15 ± 1.16	9.78 ± 1.13	4.36 ± 1.20	$\mathbf{0.473 \pm 0.014}$
Griewank	100	2488 ± 156	710 ± 122	1764 ± 370	$\mathbf{(52.6 \pm 5.3)10^{-5}}$
	1000	105 ± 35	113 ± 41	18 ± 12	$\mathbf{(36.4 \pm 2.5)10^{-5}}$
Hyperellipsoid	100	$32{,}204 \pm 3830$	9263 ± 1293	$17{,}162 \pm 2829$	$\mathbf{94 \pm 11}$
	1000	975 ± 342	$13{,}823 \pm 554$	267 ± 201	$\mathbf{64 \pm 7}$
Rastrigin	100	1640 ± 104	1171 ± 64.72	1442.24 ± 96.39	$\mathbf{257 \pm 13}$
	1000	1041 ± 52	292 ± 45	826 ± 44	$\mathbf{216 \pm 9}$
Rosenbrock	100	$4{,}182{,}787 \pm 558475$	$1{,}336{,}657 \pm 335463$	$3{,}015{,}338 \pm 674{,}769$	$\mathbf{457 \pm 46}$
	1000	$467{,}931 \pm 292{,}116$	$151{,}687 \pm 55{,}791$	$58{,}399 \pm 20{,}080$	$\mathbf{370 \pm 21}$
Schaffer's F6	100	$0.4999952 \pm 7{\cdot}10^{-7}$	$0.499991 \pm 3{\cdot}10^{-6}$	$0.4999949 \pm 1.3{\cdot}10^{-6}$	$\mathbf{(972.1 \pm 1.1)10^{-5}}$
	1000	$0.499986 \pm 4{\cdot}10^{-6}$	$0.49982 \pm 11{\cdot}10^{-5}$	$0.49985 \pm 7{\cdot}10^{-5}$	$\mathbf{(971593 \pm 3)10^{-8}}$
Schaffer's F7	100	147.51 ± 7.21	99.96 ± 10.38	126.27 ± 10.11	$\mathbf{0.39 \pm 0.02}$
	1000	73.69 ± 13.29	25.60 ± 3.67	31.01 ± 7.24	$\mathbf{0.331 \pm 0.011}$
Schwefel	100	$39{,}048 \pm 1315$	$37{,}077 \pm 1218$	$38{,}725 \pm 1093$	$\mathbf{36786 \pm 1022}$
	1000	$38{,}974 \pm 1131$	$\mathbf{17277 \pm 1281}$	$38{,}622 \pm 1014$	$30{,}621 \pm 2113$
Schwefel Problem 1.2	100	$(1.36 \pm 0.51)10^{6}$	$(0.86 \pm 0.21)10^{6}$	$(1.10 \pm 0.18)10^{6}$	$\mathbf{12.5 \pm 4.0}$
	1000	$(1.19 \pm 0.28)10^{6}$	$(0.46 \pm 0.09)10^{6}$	$(0.81 \pm 0.16)10^{6}$	$\mathbf{7.4 \pm 1.2}$
Schwefel Problem 2.21	100	97.27 ± 0.89	94.50 ± 2.19	97.21 ± 1.12	$\mathbf{0.48 \pm 0.05}$
	1000	97.01 ± 1.24	75.59 ± 6.33	97.19 ± 0.98	$\mathbf{0.32 \pm 0.03}$
Schwefel Problem 2.22	100	$(1.1 \pm 5.1)10^{41}$	$(0.4 \pm 1.9)10^{16}$	$(1.8 \pm 8.7)10^{31}$	$\mathbf{11.37 \pm 0.56}$
	1000	224.92 ± 117.18	41.89 ± 12.01	12.74 ± 4.33	$\mathbf{10.23 \pm 0.29}$
Sphere	100	734 ± 53	205 ± 29	508 ± 98	$\mathbf{2.0 \pm 0.2}$
	1000	28 ± 10	29 ± 11	4.8 ± 2.5	$\mathbf{1.41 \pm 0.12}$
Step	100	$(27.2 \pm 2.4)10^{6}$	$(7.8 \pm 1.2)10^{6}$	$(19.0 \pm 4.0)10^{6}$	$\mathbf{0.08 \pm 0.27}$
	1000	$(1.1 \pm 0.4)10^{6}$	$(1.3 \pm 0.4)10^{6}$	$(0.2 \pm 0.2)10^{6}$	$\mathbf{0.0 \pm 0.0}$
Styblinski Tang	100	-930 ± 207	$\mathbf{-1791 \pm 205}$	-906 ± 181	-1653 ± 141
	1000	-943 ± 207	$\mathbf{-3237 \pm 75}$	-1442 ± 709	-2803 ± 79
Whitley	100	$(244 \pm 38)10^{9}$	$(63 \pm 40)10^{9}$	$(247 \pm 36)10^{9}$	$\mathbf{96 \pm 22}$
	1000	$(253 \pm 37)10^{9}$	$(2.8 \pm 2.1)10^{9}$	$(160 \pm 113)10^{9}$	$\mathbf{42 \pm 7}$
Zakharov	100	2526.95 ± 358.85	2303.81 ± 237.40	2439.10 ± 239.72	$\mathbf{2.52 \pm 0.43}$
	1000	2544.78 ± 259.99	1579.88 ± 420.34	2323.45 ± 237.24	$\mathbf{1.68 \pm 0.16}$

($p - value = 0.051$) when using 1000 cycles. When analysing the results corresponding to the configuration of 100 dimensions (Table 5), in all the cases the differences are significant for a confidence level of 95% (p-value under 0.05).

When using as benchmark configuration a dimensionality of 100 dimensions without modifying of the remaining configuration (Table 5), the DE/best/binweibull variant outperforms the other statics variants in 28 of 32 cases. Similarly to the 30 dimensions configuration, the DE/best/binweibull variant is outperformed by other variants for the functions Schwefel and StyblinskiTang. Most of the confrontations are versus DE/current-to-best/1 with 12 cases, and DE/best/2 with 7 cases. The proposal of DE/best/binweibull has been derived from the differential analysis of the performance of the DE/best/X variants, being the parameters of the Weibull probability distribution inferred from the results for 30 dimensions. In the spite of this, the proposed variant keeps a high performance for 100 dimensions too.

Beyond the static DE-variants, further comparisons are made with other non-static DE-variants: dither and jitter variants (Tables 6 and 7). In comparison with static variants, DE-variants with random variation of F parameter have demonstrated better performance.

When comparing with *dither* and *jitter* variants for functions with dimensionality 30, DE/best/binweibull outperforms these variants in 22 of 32 cases. Similarly to the previous study, for the Schwefel and StyblinskiTang, DE/best/binweibull is outperformed for the two configuration of number of cycles. From the previous 22 cases, in 14 cases, the best performances is obtained for 100 cycles. For functions with dimensionality 100 (Table 7), DE/best/binweibull outperforms the other variants in 29 of 32 cases.

The application of the Kruskal-Wallis test (Tables 6, 7) indicates that the differences between the medians of the best result obtained in the independent runs are significant for a confidence level of 95%.[2]

From the comparison performed (Tables 4–7), it can be concluded that the application of scaled differences generated from a Weibull probability distribution with parameters, shape and scale (0.14, 0.05) and as base vector the best vector in the current generation produces excellent results in comparison with other DE-variants. However, from these results it is also derived that this implementation slightly reduces its performance when large number of cycles is used.

4. Conclusions

In this paper, a new variant of Differential Evolution is proposed. This new variant has its origin in the study of the performance when analysing the scaled differences in the DE-variants using the best vector of each generation as base vector. This statistical study demonstrated that high performance is reached with the scaled differences modelled with Weibull probability distribution with the parameters of the distribution in a narrow range. Outside this range, the performance critically falls down.

The results of the high-performance cases permit modelling the scaled differences with a Weibull probability distribution with shape and scale (0.14, 0.05). Based on this model, a new variant is proposed: DE/best/binweibull. In this variant the best vector from the previous generation is used as base vector and the differences are picked from a Weibull distribution with the previously mentioned parameters and a Binomial probability distribution of the sign:

[2] A confidence level of 95% (p-value under 0.05) is used in this analysis. This means that the differences are unlikely to have occurred by chance with a probability of 95%.

$w_i = best + (-1)^{binomial(0.5)} \cdot f_{Weibull}(x; 0.14, 0.05)$. This variant is confronted to a wide set of non-adaptive DE-variants, including *dither* and *jitter* ones. In general, DE/best/binweibull outperforms the other variants tested.

Beyond the creation a new DE-variant, further works have to explore the capacity of the statistical modelling technique for predicting the performance of a particular DE-variant. Additional refinements in DE/best/binweibull, mainly the degradation of the performance for very large number of cycles or higher dimensionality configurations, are also considered as future work. Furthermore, comparisons and hybridizations with self-adaptive variants are open questions for future work. More benchmarking comparisons including, Black-Box Optimization Benchmarking and rotated fitness functions, are also proposed.

Acknowledgements

The research leading to these results has received funding by the Spanish Ministry of Economy and Competitiveness (MINECO) for funding support through the Grants FPA2013-47804-C2-1-R, FPA2016-80994-C2-1-R, and “Unidad de Excelencia María de Maeztu”: CIEMAT - FÍSICA DE PARTÍCULAS through the Grant MDM-2015-0509.

References

[1] R. Storn, K.V. Price, J. Glob. Optim. 11 (4) (1997) 341–359.
[2] K.V. Price, R.M. Storn, J.A. Lampinen, Differential Evolution – A Practical Approach to Global Optimization, Natural Computing Series, Springer-Verlag, Berlin, Germany, 2005.
[3] S. Das, P.N. Suganthan, IEEE Trans. Evol. Comput. 15 (1) (2011) 4–31.
[4] S. Das, S.S. Mullick, P.N. Suganthan, Swarm Evol. Comput. 27 (2016) 1–30.
[5] F. Neri, V. Tirronen, Artif. Intell. Rev. 33 (1) (2010) 61–106.
[6] V. Feoktistov, S. Janaqi, GeneralizAtion of the Strategies in Differential Evolution, in: Proceedings of the 18th IPDPS, Apr. 2004, p. 165a.
[7] S. Das, A. Konar, U.K. Chakraborty, Two improved differential evolution schemes for faster global search, Genet. Evolut. Comput. Conf. GECCO (2005) 991–998.
[8] A. Zamuda, J. Brest, Swarm Evol. Comput. 25 (2015) 72–99.
[9] X. Lu, K. Tang, B. Sendhoff, X. Yao, Neucom 146 (2014) 2–16.
[10] A.K. Qin, P.N. Suganthan, Self-adaptive differential evolution algorithm for numerical optimization, in: Proceedings of the 2005 IEEE Congress on Evolutionary Computation, 2005, pp. 1785–1791.
[11] A.K. Qin, V.L. Huang, P.N. Suganthan, IEEE Trans. Evol. Comput. 13 (2) (2009) 398–417.
[12] R. Gämperle, S.D. Müller, P. Koumoutsakos, A parameter study for differential evolution, Adv. Intell. Syst. Fuzzy Syst. Evol. Comput. 10 (2002) 293–298.
[13] J. Rönkkönen, S. Kukkonen, K.V. Price, Real-parameter optimization with differential evolution, Congr. Evolut. Comput. (2005) 506–513.
[14] G. Iacca, F. Caraffini, F. Neri, J. Comput. Sci. Technol. 27 (5) (2009) 1056–1076.
[15] E. Mininno, F. Neri, F. Cupertino, D. Naso, vol. 15(1), 2011, pp. 32–54.
[16] M.M. Ali, L.P. Fatti, J. Glob. Optim. 35 (4) (2006) 551–572.
[17] Y. Wang, Z. Cai, Q. Zhang, Trans. Evol. Comp. 15 (1) (2011) 55–66.
[18] R. Mallipeddi, P.N. Suganthan, Q.K. Pan, M.F. Tasgetiren, Appl. Soft Comput. 11 (2) (2011) 1679–1696.
[19] R. Mallipeddi, J. Appl. Math. 2013 (2013) 750819:1–750819:12.
[20] W. Gong, Z. Cai, X.L. Charles, L. Hui, IEEE Trans. Syst. Man Cybern. Part B 41 (2) (2011) 397–413.
[21] J. Zhang, A. Sanderson, IEEE Trans. Evol. Comput. 13 (5) (2009) 945–958.
[22] Y. Wang, Z.Z. Liu, J. Li, H.X. Li, G.G. Yen, Appl. Soft Comput. 48 (2016) 329–346.
[23] K. Tang, X. Yao, P.N. Suganthan, Y. Zhenyu, T. Weise, Benchmark Functions for the CEC'2010 Special Session and Competition on Large-Scale Global Optimization, Tech. Report, 2009.
[24] K. Tang, X. Yao, P.N. Suganthan, C. MacNish, Y.P. Chen, C.M. Chen, Z. Yang, Benchmark Functions for the C 2008 Special Session and Competition on Large Scale Global Optimization, Tech. Report, 2007.
[25] F.R. Peñuñuri-Anguiano, C.A. Cab-Cauich, O. Carvente-Muñoz, M.A. Zambrano-Arjona, J.A. Tapia-González, Swarm Evol. Comput. 26 (2016) 86–96.
[26] X. Lu, K. Tang, B. Sendhoff, X. Yao, Neurocomput 146 (2014) 2–16.
[27] E. Mezura-Montes, J. Velázquez-Reyes, C.A.C. Coello, A comparative study of differential evolution variants for global optimization, in: Proceedings of the GECCO, 2006, pp. 485–492.
[28] Y. Chen, W. Xie, X. Zou, Neucom 149 (2015) 1038–1047.
[29] M. Matsumoto, T. Nishimura, Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator, ACM Trans. Model. Comput. Simul. 8 (1) (1998) 3–30.
[30] C. Walck, Hand-book on Statistical Distributions for Experimentalists, SUF-PFY/96-01, Particle Physics Group, Fysikum, University of Stockholm, 1996.
[31] S. García, A. Fernández, J. Luengo, F. Herrera, Soft Comput. 13 (10) (2009) 959–977.
[32] S. García, D. Molina, M. Lozano, F. Herrera, J. Heuristics 15 (6) (2009) 617–644.
[33] D. Sheskin, Handbook of Parametric and Non-parametric Statistical Procedures, CRC Press, 2004.
[34] J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2000.