Comparison Tables: CEC BBOB 2015 Testbed in 2-D (Expensive Setting)

The BBOBies

May 27, 2015

Abstract

This document provides tabular results of the special session on Black-Box Optimization Benchmarking at CEC 2015 with a focus on benchmarking black-box algorithms for small function evaluation budgets ("expensive setting"), see `http://coco.gforge.inria.fr/doku.php?id=cec-bbob-2015`. Overall, eight algorithms have been tested on 24 benchmark functions in dimensions between 2 and 20. A description of the used objective functions can be found in [6, 4]. The experimental set-up is described in [5].

The performance measure provided in the following tables is the expected number of objective function evaluations to reach a given target function value (ERT, expected running time), divided by the respective value for the best algorithm in BBOB-2009 (see [1]) if an algorithm from BBOB-2009 reached the given target function value. The ERT value is given otherwise ($\mathrm{ERT}_{\mathrm{best}}$ is noted as infinite). See [5] for details on how ERT is obtained. Bold entries in the table correspond to values below 3 or the top-three best values. Table 1 gives an overview on all algorithms submitted to the noise-free testbed at CEC 2015.

Table 1: Names and references of all algorithms submitted for the noise-free testbed

algorithm short name	paper	reference
MATSuMoTo	Comparison of the MATSuMoTo Library for Expensive Optimization on the Noiseless Black-Box Optimization Benchmarking Testbed	[2]
R-DE-10e2	Parameter Tuning for Differential Evolution for Cheap, Medium, and Expensive Computational Budgets	[7]
R-DE-10e5	Parameter Tuning for Differential Evolution for Cheap, Medium, and Expensive Computational Budgets	[7]
R-SHADE-10e2	Parameter Tuning for Differential Evolution for Cheap, Medium, and Expensive Computational Budgets	[7]
R-SHADE-10e5	Parameter Tuning for Differential Evolution for Cheap, Medium, and Expensive Computational Budgets	[7]
RL-SHADE-10e2	Parameter Tuning for Differential Evolution for Cheap, Medium, and Expensive Computational Budgets	[7]
RL-SHADE-10e5	Parameter Tuning for Differential Evolution for Cheap, Medium, and Expensive Computational Budgets	[7]
SOO	Simultaneous Optimistic Optimization on the Noiseless BBOB Testbed	[3]

Table 2: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_1 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f1	*1.6e+1*:1.2	*4.0e+0*:2.6	*2.5e-2*:6.2	*1.0e-8*:6.2	*1.0e-8*:6.2	15/15
MATSUMOTO-	**1.7**(2)	**1.8**(1)	**2.8**(0.8)$^{\star 2}$	∞	∞ *100*	0/15
R-DE-10e2-	**2.7**(1)	**2.3**(3)	10(5)	∞	∞ *200*	0/15
R-DE-10e5-	**2.3**(1)	**3.0**(4)	66(222)	102(6)	102(150)	15/15
RL-SHADE-1	**1.8**(4)	**2.1**(0.8)	11(4)	240(210)	240(330)	2/15
RL-SHADE-1	**2.1**(2)	3.3(4)	30(17)	225(20)	225(26)	15/15
R-SHADE-10	**2.7**(2)	**2.8**(2)	13(5)	∞	∞ *200*	0/15
R-SHADE-10	**1.9**(1)	**2.9**(1)	11(4)	**45**(6)	**45**(7)	15/15
SOO-Derbel	**1.1**(0.2)	**1.2**(0.8)	**5.6**(2)	**52**(7)	**52**(3)	15/15

Table 3: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_2 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f2	*1.0e+7*:1.4	*1.6e+6*:2.7	*1.0e+5*:6.1	*6.3e-1*:20	*1.0e-8*:30	15/15
MATSUMOTO-	**1.5**(0.7)	**1.6**(0.5)	**1.3**(1)	∞	∞ *100*	0/15
R-DE-10e2-	**1.6**(1)	**1.9**(2)	**2.2**(2)	**7.3**(3)	∞ *200*	0/15
R-DE-10e5-	**1**(0.5)	**1.2**(2)	**2.1**(2)	13(46)	**24**(3)	15/15
RL-SHADE-1	**1.1**(1)	**1.0**(0.5)	3.2(2)	**6.4**(2)	∞ *200*	0/15
RL-SHADE-1	**0.90**(0.4)	**1.5**(0.9)	**2.8**(3)	27(13)	73(4)	15/15
R-SHADE-10	**1.1**(0.9)	**1.1**(0.7)	5.0(3)	12(6)	∞ *200*	0/15
R-SHADE-10	**1.1**(1)	**2.3**(2)	**2.2**(3)	8.7(1)	**15**(2)	15/15
SOO-Derbel	**1.1**(0.4)	**1.2**(1)	**2.6**(1)	8.9(2)	27(4)	15/15

Table 4: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_3 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f3	*1.0e+2*:1.4	*4.0e+1*:4.1	*2.5e+1*:6.6	*6.3e+0*:26	*2.5e+0*:112	15/15
MATSUMOTO-	**1.6**(0.7)	**1.4**(1)	**1.5**(0.9)	**2.7**(2)	**1.4**(0.5)	8/15
R-DE-10e2-	**1.3**(1)	**1.3**(1)	**1.7**(1)	**2.2**(2)	**0.87**(0.6)	15/15
R-DE-10e5-	**1.0**(0.7)	**2.1**(2)	**2.3**(1)	**2.1**(2)	**2.2**(2)	15/15
RL-SHADE-1	**1.4**(1)	**1.5**(1)	**2.3**(2)	**1.8**(0.7)	**1.0**(0.2)	13/15
RL-SHADE-1	**1.8**(2)	**2.3**(2)	3.8(3)	4.9(4)	**2.8**(1.0)	15/15
R-SHADE-10	**1.2**(0.5)	**2.0**(2)	**2.0**(2)	**1.9**(1)	**0.80**(0.6)	15/15
R-SHADE-10	**1.5**(1)	**2.3**(2)	**2.8**(3)	5.4(9)	**2.0**(2)	15/15
SOO-Derbel	**0.86**(0.2)	**1.3**(0.9)	**1.7**(2)	**1.9**(1)	**0.90**(0.5)	15/15

Table 5: 02-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_4 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f4	*6.3e+1*:2.4	*4.0e+1*:5.2	*2.5e+1*:8.5	*1.0e+1*:22	*2.5e+0*:120	5/5
MATSUMOTO-	**1.9**(1)	**1.5**(0.9)	**1.8**(1)	**1.9**(1)	**2.2**(2)	5/15
R-DE-10e2-	**1.6**(0.9)	**1.7**(1)	**2.0**(2)	**1.9**(2)	**1.2**(0.4)	13/15
R-DE-10e5-	**2.5**(2)	**2.8**(2)	5.5(2)	5.1(11)	**1.7**(0.3)	15/15
RL-SHADE-1	**1.4**(1)	**1.6**(1)	**1.5**(2)	**1.8**(0.9)	**0.81**(0.2)	14/15
RL-SHADE-1	**2.1**(2)	**1.5**(0.9)	4.1(3)	5.2(1)	**2.9**(1)	15/15
R-SHADE-10	**1.8**(1)	**2.4**(2)	**1.9**(1)	**2.0**(2)	**1.1**(0.5)	13/15
R-SHADE-10	**1.9**(2)	**1.6**(2)	**1.8**(1)	**2.4**(0.6)	**3.0**(6)	15/15
SOO-Derbel	**0.83**(0.4)	**1.2**(0.6)	**1.5**(1)	**1.5**(0.8)	**1.2**(0.4)	15/15

Table 6: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_5 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f5	*4.0e+1*:1.4	*1.6e+1*:3.5	*1.0e-8*:4.4	*1.0e-8*:4.4	*1.0e-8*:4.4	15/15
MATSUMOTO-	**1.5**(1)	**1.2**(0.9)	**2.5**(0.9)$^{\star 4}$	**2.5**(1)$^{\star 4}$	**2.5**(0.6)$^{\star 4}$	15/15
R-DE-10e2-	**1.9**(0.7)	3.1(3)	∞	∞	∞ *200*	0/15
R-DE-10e5-	**1.5**(1)	**2.7**(2)	319(272)	319(367)	319(295)	15/15
RL-SHADE-1	**2.0**(1)	3.0(2)	320(474)	320(253)	320(276)	2/15
RL-SHADE-1	**1.3**(0.4)	**1.7**(3)	349(24)	349(33)	349(14)	15/15
R-SHADE-10	**2.0**(1)	**2.9**(3)	∞	∞	∞ *200*	0/15
R-SHADE-10	**1.9**(1)	**2.0**(0.9)	**133**(49)	**133**(52)	**133**(22)	15/15
SOO-Derbel	**1.8**(0.4)	**2.5**(0.1)	338(0.1)	338(0.1)	338(0.1)	15/15

Table 7: 02-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_6 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f6	*6.3e+4*:1.4	*1.0e+2*:2.8	*1.6e+1*:10	*1.0e+0*:23	*2.5e-6*:103	15/15
MATSUMOTO-	**1.3**(0.9)	**1.7**(1)	**1.3**(0.9)	14(9)	∞ *100*	0/15
R-DE-10e2-	**1.4**(0.9)	**2.1**(2)	**1.4**(2)	**2.3**(0.8)	∞ *200*	0/15
R-DE-10e5-	**1.5**(0.9)	**1.8**(2)	12(4)	20(66)	18(21)	15/15
RL-SHADE-1	**1.2**(0.4)	**2.0**(2)	**2.0**(2)	4.2(2)	∞ *200*	0/15
RL-SHADE-1	**1.4**(2)	**2.0**(2)	**2.4**(2)	7.1(5)	**16**(4)	15/15
R-SHADE-10	**1.0**(0.5)	**1.3**(1)	**1.2**(0.8)	4.2(2)	∞ *200*	0/15
R-SHADE-10	**1.4**(2)	**2.3**(3)	**1.7**(3)	**3.2**(2)	**4.9**(3)	15/15
SOO-Derbel	**1**(1)	**0.95**(0.4)	**1.2**(2)	7.6(7)	1.4e4(2e4)	2/15

Table 8: 02-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_7 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f7	*4.0e+2*:1.6	*1.0e+1*:3.2	*2.5e+0*:14	*1.6e+0*:21	*1.6e-2*:188	15/15
MATSUMOTO-	**1.4**(1)	**2.8**(2)	**1.2**(0.9)	**1.1**(0.3)	**0.55**(0.4)	10/15
R-DE-10e2-	**1.5**(0.8)	3.1(2)	**1.9**(1)	**2.0**(2)	**1.0**(1)	11/15
R-DE-10e5-	**0.87**(0.6)	**2.2**(2)	**1.5**(2)	**2.6**(5)	**1.6**(0.8)	15/15
RL-SHADE-1	**1**(1)	5.9(4)	**2.5**(2)	**2.8**(4)	**2.4**(2)	6/15
RL-SHADE-1	**0.96**(0.5)	4.9(7)	**2.9**(4)	**3.0**(4)	**2.3**(1)	15/15
R-SHADE-10	**1.0**(0.3)	**2.9**(1)	**1.4**(0.8)	**2.0**(2)	4.9(4)	3/15
R-SHADE-10	**0.75**(0.3)	**2.6**(3)	**1.4**(2)	**1.4**(1)	**0.81**(0.6)	15/15
SOO-Derbel	**1.2**(1)	**1.8**(2)	**1.0**(1.0)	**1.3**(0.7)	**0.76**(0.4)	15/15

Table 9: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_8 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f8	*2.5e+3*:1.2	*1.0e+2*:3.2	*6.3e+0*:7.0	*1.6e-1*:27	*1.6e-6*:100	15/15
MATSUMOTO-	**1.7**(1)	**2.1**(1)	**2.6**(1)	7.7(11)	∞ *100*	0/15
R-DE-10e2-	**1.5**(0.8)	4.2(4)	4.1(4)	16(5)	∞ *200*	0/15
R-DE-10e5-	**2.0**(0.6)	**2.5**(0.8)	**2.8**(3)	16(8)	**12**(6)	15/15
RL-SHADE-1	**1.8**(1)	**2.5**(3)	3.8(4)	6.3(2)	∞ *200*	0/15
RL-SHADE-1	**2.1**(0.4)	**2.6**(1)	7.8(8)	12(5)	23(3)	15/15
R-SHADE-10	**2.3**(1)	**3.0**(3)	7.6(8)	15(34)	∞ *200*	0/15
R-SHADE-10	**2.0**(2)	**2.7**(2)	3.4(2)	**4.6**(5)	**5.3**(5)	15/15
SOO-Derbel	**1.7**(0.4)	**1.4**(1)	**1.9**(1)	**4.4**(3)	21(13)	15/15

Table 10: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_9 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f9	*6.3e+0*:13	*4.0e+0*:15	*2.5e+0*:15	*2.5e-1*:21	*1.0e-8*:94	15/15
MATSUMOTO-	**1.3**(1)	**1.4**(0.9)	**1.6**(1)	**4.3**(2)	∞ *100*	0/15
R-DE-10e2-	**3.0**(6)	3.5(5)	4.9(1)	8.3(9)	∞ *200*	0/15
R-DE-10e5-	3.8(3)	4.6(2)	6.0(5)	27(86)	**20**(13)	15/15
RL-SHADE-1	3.9(6)	4.0(3)	4.4(7)	10(8)	∞ *200*	0/15
RL-SHADE-1	3.2(2)	3.4(4)	3.6(4)	15(8)	28(2)	15/15
R-SHADE-10	3.4(2)	3.7(3)	5.0(4)	16(15)	∞ *200*	0/15
R-SHADE-10	**2.2**(2)	**2.0**(2)	**2.3**(3)	7.3(3)	**8.0**(12)	15/15
SOO-Derbel	**1.4**(0.8)	**1.6**(0.9)	**1.9**(0.8)	**2.8**(2)	25(16)	15/15

Table 11: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{10} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f10	*1.6e+6*:2.0	*4.0e+5*:3.2	*6.3e+2*:8.8	*1.0e+1*:30	*2.5e-8*:101	15/15
MATSUMOTO-	**2.3**(2)	**2.4**(2)	**2.9**(3)	9.1(11)	∞ *100*	0/15
R-DE-10e2-	**2.2**(1)	**2.0**(1)	6.8(6)	7.3(5)	∞ *200*	0/15
R-DE-10e5-	**2.8**(2)	**2.8**(2)	6.9(6)	27(30)	74(51)	15/15
RL-SHADE-1	**1.7**(2)	**1.7**(1)	7.1(6)	18(24)	∞ *200*	0/15
RL-SHADE-1	**2.0**(2)	**2.2**(1)	14(8)	15(8)	**26**(2)	15/15
R-SHADE-10	**2.2**(2)	**1.6**(1)	11(10)	13(11)	∞ *200*	0/15
R-SHADE-10	**2.2**(2)	**2.4**(2)	5.0(3)	**3.2**(0.9)	**4.4**(1)	15/15
SOO-Derbel	**1.3**(0.8)	**1.1**(0.9)	**3.9**(3)	**3.3**(2)	494(196)	15/15

Table 12: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{11} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f11	*1.0e+7*:1.1	*1.6e+6*:3.2	*1.0e+4*:6.6	*4.0e+1*:23	*4.0e-8*:100	15/15
MATSUMOTO-	**1.6**(0.9)	**1.2**(1)	**2.4**(1)	**2.8**(2)	∞ *100*	0/15
R-DE-10e2-	**1.8**(0.5)	**1.1**(0.8)	4.2(4)	5.1(3)	∞ *200*	0/15
R-DE-10e5-	**1.9**(2)	**1.8**(1)	**3.0**(2)	10(20)	94(103)	15/15
RL-SHADE-1	**2.0**(0.7)	**1.6**(2)	4.4(4)	5.7(5)	∞ *200*	0/15
RL-SHADE-1	**2.1**(3)	**1.9**(2)	8.6(12)	12(5)	**25**(3)	15/15
R-SHADE-10	**1.8**(0.9)	**1.5**(1)	4.1(2)	9.0(7)	∞ *200*	0/15
R-SHADE-10	**1.8**(0.5)	**2.2**(1)	4.4(4)	4.1(5)	**5.0**(2)	15/15
SOO-Derbel	**1.2**(0.5)	**0.94**(0.8)	**2.7**(1)	**2.9**(0.9)	481(197)	15/15

Table 13: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{12} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f12	*2.5e+8*:1.3	*6.3e+6*:2.7	*6.3e+5*:6.3	*4.0e+1*:21	*1.6e-3*:101	15/15
MATSUMOTO-	**1.1**(0.8)	**0.93**(0.7)	**0.77**(0.4)	4.6(3)	∞ *100*	0/15
R-DE-10e2-	**0.85**(0.2)	**0.80**(0.5)	**1.0**(1.0)	5.3(5)	29(22)	1/15
R-DE-10e5-	**0.95**(0)	**0.98**(0.5)	**0.98**(0.5)	8.0(1.0)	37(55)	15/15
RL-SHADE-1	**1.6**(2)	**1.5**(2)	**1.1**(0.6)	**3.9**(2)	29(27)	1/15
RL-SHADE-1	**1**(1)	**0.76**(0.7)	**1.4**(1)	13(5)	**26**(9)	15/15
R-SHADE-10	**1.1**(0.6)	**1.7**(1)	**1.4**(1.0)	7.8(3)	∞ *200*	0/15
R-SHADE-10	**0.80**(0.2)	**0.90**(0.7)	**1.6**(0.8)	4.7(4)	**15**(15)	15/15
SOO-Derbel	**0.95**(0.8)	**0.73**(0.7)	**0.59**(0.5)	**4.3**(2)	33(20)	15/15

Table 14: 02-D, running time excess ERT/$ERT_{best\ 2009}$ on f_{13} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $ERT_{best\ 2009}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f13	*4.0e+2*:1.6	*2.5e+2*:3.1	*6.3e+1*:8.7	*1.0e+1*:23	*4.0e-6*:100	15/15
MATSUMOTO-	**1.6**(0.9)	**1.5**(0.6)	**1.4**(0.5)	**1.6**(0.4)	∞ *100*	0/15
R-DE-10e2-	**1.5**(2)	**1.6**(4)	**2.1**(2)	4.0(0.7)	∞ *200*	0/15
R-DE-10e5-	**1.8**(1)	**1.5**(1)	**2.4**(5)	4.5(9)	100(94)	15/15
RL-SHADE-1	**1.4**(2)	**1.5**(0.6)	**2.4**(2)	3.9(3)	∞ *200*	0/15
RL-SHADE-1	**1.2**(0.3)	**0.68**(0.5)	**2.5**(4)	4.7(4)	**25**(2)	15/15
R-SHADE-10	**2.3**(1)	**1.6**(1)	**3.0**(2)	5.8(6)	∞ *200*	0/15
R-SHADE-10	**1.5**(1)	**1.3**(1)	**1.8**(1)	**2.6**(0.8)	**4.9**(0.4)	15/15
SOO-Derbel	**1.4**(3)	**1.1**(1)	**1.8**(1)	**2.8**(4)	153(161)	15/15

Table 15: 02-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{14} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f14	*1.6e+1*:1.4	*2.5e+0*:4.2	*1.0e+0*:7.4	*2.5e-2*:21	*1.0e-8*:101	15/15
MATSUMOTO-	**1.0**(0.7)	**1.4**(1)	**1.5**(1)	**1.7**(0.5)$^{\star 2}$	∞ *100*	0/15
R-DE-10e2-	**1.3**(0.7)	**2.0**(2)	3.1(2)	4.2(1)	∞ *200*	0/15
R-DE-10e5-	**1.0**(0.7)	**1.1**(1)	**1.9**(0.8)	4.2(0.4)	94(129)	15/15
RL-SHADE-1	**1.2**(0.7)	**2.2**(3)	**2.7**(3)	4.1(2)	∞ *200*	0/15
RL-SHADE-1	**1.3**(0.4)	3.2(1)	4.3(4)	13(6)	**24**(2)	15/15
R-SHADE-10	**1.4**(1)	**2.0**(3)	3.3(3)	5.5(3)	∞ *200*	0/15
R-SHADE-10	**1.3**(0.7)	**2.8**(3)	3.8(4)	4.2(1)	**4.1**(0.4)	15/15
SOO-Derbel	**0.81**(0)	**1.6**(1)	**1.8**(1)	**3.2**(1)	43(19)	15/15

Table 16: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{15} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f15	*1.6e+2*:1.2	*4.0e+1*:4.7	*2.5e+1*:10	*1.0e+1*:37	*2.5e+0*:118	5/5
MATSUMOTO-	**1.6**(1)	**1.1**(0.7)	**0.79**(0.8)	**0.85**(0.6)	**1.0**(0.8)	9/15
R-DE-10e2-	**1.7**(0.8)	**1**(0.8)	**1.3**(1)	**1.2**(0.5)	**1.9**(2)	10/15
R-DE-10e5-	**1.8**(1)	**1.4**(0.9)	**1.2**(1)	**0.94**(1)	**2.1**(4)	15/15
RL-SHADE-1	**2.2**(1)	**1.6**(1)	**1.4**(1)	**1.8**(1)	**2.0**(2)	9/15
RL-SHADE-1	**1.8**(1)	**2.0**(4)	**1.4**(2)	**1.5**(2)	**2.7**(2)	15/15
R-SHADE-10	**1.5**(1)	**2.0**(2)	**1.4**(0.8)	**1.7**(1)	4.4(2)	5/15
R-SHADE-10	**1.3**(0.2)	**1.2**(1)	**1.7**(0.8)	**1.0**(0.8)	**1.4**(2)	15/15
SOO-Derbel	**1.1**(0.4)	**0.84**(1)	**0.74**(0.6)	**0.75**(0.4)	**1.1**(0.8)	15/15

Table 17: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{16} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f16	*1.0e+2*:1.1	*2.5e+1*:3.9	*1.6e+1*:6.5	*4.0e+0*:31	*2.5e-1*:127	5/5
MATSUMOTO-	**1.3**(1)	**1.7**(1)	**1.6**(2)	**1.1**(1)	**1.3**(2)	7/15
R-DE-10e2-	**1.4**(0.4)	**2.1**(2)	**1.8**(1)	**1.8**(2)	3.8(8)	5/15
R-DE-10e5-	**1.3**(0.9)	**1.3**(1)	**2.4**(3)	4.3(1)	**2.6**(3)	15/15
RL-SHADE-1	**1.1**(0)	**2.1**(2)	**1.8**(2)	**2.2**(1)	5.0(5)	4/15
RL-SHADE-1	**1.4**(0.9)	**2.4**(2)	3.3(3)	**2.2**(2)	4.5(2)	15/15
R-SHADE-10	**1.4**(0.7)	**2.2**(2)	**2.3**(3)	**1.7**(1)	4.0(5)	5/15
R-SHADE-10	**1.2**(0.4)	**2.3**(2)	**1.8**(2)	**1.9**(3)	**2.2**(1)	15/15
SOO-Derbel	**1.5**(0.4)	**2.3**(2)	**1.9**(1)	**1.0**(0.6)	**0.93**(1)	15/15

Table 18: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{17} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f17	*4.0e+1*:1.2	*1.0e+1*:2.7	*4.0e+0*:10	*2.5e+0*:28	*1.6e-1*:119	5/5
MATSUMOTO-	**1.2**(0.6)	**1.5**(1)	**1.2**(0.3)	**0.52**(0.4)	**2.1**(3)	5/15
R-DE-10e2-	**1.7**(1)	3.2(2)	**1.6**(1)	**1.1**(1)	**1.3**(0.9)	14/15
R-DE-10e5-	**1.2**(0.4)	**1.8**(0.8)	**1.5**(2)	**0.85**(0.3)	**1.7**(2)	15/15
RL-SHADE-1	**1.2**(0.4)	**2.3**(3)	**1.8**(1.0)	**1.2**(2)	**1.6**(1)	11/15
RL-SHADE-1	**1.6**(2)	**1.9**(2)	**2.4**(3)	**1.4**(1)	3.4(1)	15/15
R-SHADE-10	**1.5**(1)	**2.8**(1)	**1.9**(2)	**1.0**(1)	4.6(6)	5/15
R-SHADE-10	**1.5**(1)	**2.3**(3)	**1.8**(2)	**1.1**(0.6)	**1.2**(0.5)	15/15
SOO-Derbel	**0.94**(0)	**1.8**(1)	**0.85**(0.5)	**0.48**(0.3)	**0.95**(0.8)	15/15

Table 19: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{18} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f18	*4.0e+2*:1.2	*1.0e+2*:3.2	*4.0e+1*:7.2	*6.3e+0*:32	*1.6e+0*:104	5/5
MATSUMOTO-	**1.1**(0.4)	**0.83**(0.4)	**0.95**(0.4)	**0.86**(0.8)	**1.0**(0.6)	10/15
R-DE-10e2-	**1.3**(0.4)	**1.2**(0.5)	**1.5**(1)	**1.6**(0.7)	**1.0**(0.9)	14/15
R-DE-10e5-	**1.2**(0.2)	**1.2**(0.9)	**1.3**(1)	3.2(0.8)	3.0(7)	15/15
RL-SHADE-1	**1.2**(0.4)	**1.3**(1)	**2.1**(2)	**1.2**(1)	**1.6**(1)	10/15
RL-SHADE-1	**1.4**(0.6)	**1.6**(2)	**1.9**(2)	**2.6**(1)	**2.3**(1)	15/15
R-SHADE-10	**1.2**(0.4)	**0.98**(0.3)	**0.77**(1)	**1.2**(2)	**1.6**(0.8)	12/15
R-SHADE-10	**1.5**(0.6)	**1.8**(0.6)	**1.4**(1)	**1.4**(0.8)	**2.4**(0.4)	15/15
SOO-Derbel	**1.1**(0.4)	**1.6**(1)	**0.94**(0.9)	**0.73**(0.4)	**0.94**(0.5)	15/15

Table 20: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{19} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f19	*1.6e-1*:23	*1.0e-1*:26	*6.3e-2*:38	*4.0e-2*:40	*1.0e-2*:216	15/15
MATSUMOTO-	6.9(8)	9.3(4)	11(11)	11(12)	6.6(8)	1/15
R-DE-10e2-	4.3(2)	**4.2**(3)	5.5(2)	6.6(5)	6.7(5)	2/15
R-DE-10e5-	5.2(5)	5.6(6)	6.4(4)	12(14)	11(6)	15/15
RL-SHADE-1	5.4(6)	7.1(7)	6.9(7)	13(10)	**6.5**(7)	2/15
RL-SHADE-1	7.0(11)	7.6(8)	7.0(6)	14(14)	9.1(6)	15/15
R-SHADE-10	9.4(7)	8.4(11)	11(18)	24(33)	∞ *200*	0/15
R-SHADE-10	**2.9**(2)	**3.3**(1)	**3.2**(3)	**4.9**(5)	**5.2**(3)	15/15
SOO-Derbel	**4.1**(2)	5.0(4)	**3.6**(2)	**3.7**(2)	17(15)	15/15

Table 21: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{20} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f20	*4.0e+3*:1.9	*2.5e+2*:2.8	*4.0e+0*:6.3	*2.5e+0*:21	*6.3e-1*:139	15/15
MATSUMOTO-	**1.5**(0.8)	**2.0**(2)	**2.1**(1)	**1.8**(1)	3.4(2)	3/15
R-DE-10e2-	**1**(0.5)	**1.7**(1)	**2.6**(3)	**1.8**(1)	**2.0**(3)	8/15
R-DE-10e5-	**1.2**(1)	**2.2**(2)	**1.9**(1)	**1.8**(1.0)	**2.4**(1)	15/15
RL-SHADE-1	**1.4**(0.9)	**2.7**(2)	3.1(2)	**2.1**(1)	**2.7**(3)	7/15
RL-SHADE-1	**1.7**(1)	3.3(3)	4.7(3)	3.1(3)	4.6(2)	15/15
R-SHADE-10	**2.0**(5)	**2.6**(4)	4.0(2)	**2.4**(1)	6.7(5)	3/15
R-SHADE-10	**1.5**(2)	**2.5**(2)	3.2(2)	**2.1**(1)	**2.3**(1)	15/15
SOO-Derbel	**1.4**(0.3)	5.1(0.2)	**3.0**(0.1)	**1.3**(0.0)	**1.2**(4e-3)	15/15

Table 22: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{21} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f21	*1.0e+1*:1.7	*6.3e+0*:2.6	*2.5e+0*:7.9	*1.6e+0*:30	*4.0e-1*:105	15/15
MATSUMOTO-	**1.1**(1)	**1.8**(2)	**1.7**(2)	**0.92**(1)	**0.64**(1)	12/15
R-DE-10e2-	**1.6**(2)	**2.6**(10)	**1.7**(4)	**2.1**(2)	**2.7**(2)	7/15
R-DE-10e5-	**1.3**(0.6)	**1.3**(1)	**1.5**(0.4)	6.9(2)	5.3(7)	15/15
RL-SHADE-1	**1.1**(0.4)	**1.5**(1)	**1.2**(0.8)	**0.92**(0.2)	**1.1**(0.5)	11/15
RL-SHADE-1	**1.6**(1)	**2.2**(2)	**1.3**(2)	**1.1**(1)	**1.2**(2)	15/15
R-SHADE-10	**1.5**(2)	**1.7**(2)	**1.4**(0.9)	**1.6**(1)	**1.1**(1)	12/15
R-SHADE-10	**1.3**(0.3)	**1.9**(3)	**2.0**(2)	**1.2**(0.7)	3.3(6)	15/15
SOO-Derbel	**0.88**(0.3)	**1.5**(0.7)	**1.3**(0.9)	**0.93**(0.8)	**0.53**(0.6)	15/15

Table 23: 02-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{22} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f22	*4.0e+1*:1.3	*1.6e+1*:3.2	*6.3e+0*:9.3	*1.6e+0*:25	*1.0e-1*:168	15/15
MATSUMOTO-	**1.7**(0.8)	**2.0**(2)	**1.2**(1.0)	**1.1**(0.8)	**0.90**(0.9)	8/15
R-DE-10e2-	**1.9**(2)	**1.7**(2)	**1.5**(0.6)	**1.8**(4)	**0.64**(0.4)	13/15
R-DE-10e5-	**1.6**(1)	**1.5**(1)	**1.5**(2)	11(2)	3.0(6)	15/15
RL-SHADE-1	**1.5**(0.4)	**1.4**(0.9)	**1.4**(1.0)	**1.8**(3)	**0.90**(2)	10/15
RL-SHADE-1	**1.3**(0.4)	**0.58**(0.2)	**0.74**(1)	**2.0**(2)	17(1)	15/15
R-SHADE-10	**1.2**(0.6)	**0.77**(0.7)	**1.3**(2)	**2.5**(3)	**1.5**(2)	8/15
R-SHADE-10	**1.3**(0.8)	**1.2**(0.8)	**0.81**(0.5)	4.2(12)	**1.9**(4)	15/15
SOO-Derbel	**1.3**(0.4)	**1.2**(0.9)	**0.78**(0.8)	**0.89**(1)	**0.71**(0.7)	15/15

Table 24: 02-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{23} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f23	*4.0e+1*:1.5	*2.5e+1*:2.6	*1.0e+1*:7.8	*4.0e+0*:55	*2.5e+0*:103	5/5
MATSUMOTO-	**1.2**(0.7)	**1.4**(1)	**1.2**(2)	**1.4**(2)	**1.9**(2)	6/15
R-DE-10e2-	**1.7**(0.8)	**1.9**(1)	**1.8**(1.0)	**1.3**(1.0)	**1.6**(3)	10/15
R-DE-10e5-	**1.6**(2)	**1.8**(1)	**1.7**(2)	**2.1**(7)	4.5(7)	15/15
RL-SHADE-1	**0.96**(0.7)	**1.5**(2)	**1.9**(2)	**2.0**(4)	3.5(5)	6/15
RL-SHADE-1	**1.2**(1.0)	**1.7**(3)	**2.0**(1)	**2.3**(1)	**2.1**(1)	15/15
R-SHADE-10	**1.3**(0.5)	**1.4**(1)	**2.4**(3)	**2.0**(2)	**3.0**(5)	7/15
R-SHADE-10	**1.3**(2)	**1.4**(0.9)	**2.3**(4)	**2.4**(3)	**2.1**(1)	15/15
SOO-Derbel	**1.4**(2)	**1.3**(0.8)	**2.6**(3)	**1.5**(1)	**1.6**(1)	15/15

Table 25: 02-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ on f_{24} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f24	*4.0e+1*:1.1	*2.5e+1*:2.7	*1.6e+1*:7.7	*6.3e+0*:44	*2.5e+0*:275	5/5
MATSUMOTO-	**1.6**(0.9)	**1.7**(2)	**1.1**(0.8)	**1.2**(2)	∞ *100*	0/15
R-DE-10e2-	**1.5**(0.9)	**1.3**(0.9)	**0.77**(0.7)	**0.96**(0.6)	**2.4**(3)	4/15
R-DE-10e5-	**1.5**(0.9)	**1.7**(1)	**1.0**(1)	**1.5**(1)	8.3(7)	15/15
RL-SHADE-1	**1.6**(1)	**1.6**(1)	**1.4**(2)	**1.7**(2)	5.0(6)	2/15
RL-SHADE-1	**1.6**(0.9)	**1.4**(0.9)	**1.4**(2)	**1.3**(1)	**2.9**(2)	15/15
R-SHADE-10	**1.2**(0)	**2.5**(3)	**1.8**(3)	**1.0**(1)	11(13)	1/15
R-SHADE-10	**1.5**(0.9)	**2.0**(2)	**1.2**(1)	**0.75**(0.6)	**1.8**(1)	15/15
SOO-Derbel	**1.1**(0.2)	**1.4**(1)	**1.9**(0.8)	**0.88**(0.7)	**1.6**(3)	15/15

References

[1] Anne Auger, Steffen Finck, Nikolaus Hansen, and Raymond Ros. BBOB 2009: Comparison tables of all algorithms on all noiseless functions. Technical Report RT-0383, INRIA, April 2010.

[2] Dimo Brockhoff. Comparison of the matsumoto library for expensive optimization on the noiseless black-box optimization benchmarking testbed. In *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2015, 25-28 May, Sendai, Japan*, 2015.

[3] Bilel Derbel and Philippe Preux. Simultaneous optimistic optimization on the noiseless bbob testbed. In *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2015, 25-28 May, Sendai, Japan*, 2015.

[4] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009. Updated February 2010.

[5] N. Hansen, A. Auger, S. Finck, and R. Ros. Real-parameter black-box optimization benchmarking 2012: Experimental setup. Technical report, INRIA, 2012.

[6] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009. Updated February 2010.

[7] Ryoji Tanabe and Alex Fukunaga. Parameter tuning for differential evolution for cheap, medium, and expensive computational budgets. In *Proceedings of the IEEE Congress on Evolutionary Computation, CEC 2015, 25-28 May, Sendai, Japan*, 2015.