Comparison Tables: CEC BBOB 2015 Testbed in 20-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: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*6.3e+1*:24	*4.0e+1*:42	*1.0e-8*:43	*1.0e-8*:43	*1.0e-8*:43	15/15
MATSUMOTO-	**2.5**(0.3)	**2.0**(0.3)$^{\star 2}$	∞	∞	∞ *1000*	0/15
R-DE-10e2-	3.7(2)	**3.2**(1)	∞	∞	∞ *2000*	0/15
R-DE-10e5-	10(3)	10(1)	**230**(6)	**230**(5)	**230**(6)	15/15
RL-SHADE-1	11(0.5)	7.1(0.8)	∞	∞	∞ *2000*	0/15
RL-SHADE-1	53(17)	47(15)	1041(15)	1041(17)	1041(10)	15/15
R-SHADE-10	6.8(3)	5.5(1)	∞	∞	∞ *2000*	0/15
R-SHADE-10	12(4)	12(3)	**294**(18)	**294**(20)	**294**(32)	15/15
SOO-Derbel	**3.6**(2)	4.3(1)	1042(23)	1042(26)	1042(19)	15/15

Table 3: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*4.0e+6*:29	*2.5e+6*:42	*1.0e+5*:65	*1.0e+4*:207	*1.0e-8*:412	15/15
MATSUMOTO-	**0.79**(1)	**1.3**(1)	9.3(6)	71(145)	∞ *1000*	0/15
R-DE-10e2-	**0.87**(0.8)	**0.95**(0.3)	**3.8**(2)$^{\star 2}$	**2.6**(1)	∞ *2000*	0/15
R-DE-10e5-	**1.4**(2)	**2.0**(2)	12(2)	6.9(0.7)	**32**(0.9)	15/15
RL-SHADE-1	**1.3**(1)	**1.3**(2)	**6.1**(1)	4.2(2)	∞ *2000*	0/15
RL-SHADE-1	**1.5**(1)	3.0(1)	85(10)	61(4)	152(2)	15/15
R-SHADE-10	**1.9**(2)	**2.0**(1)	6.2(1)	**3.6**(0.3)	∞ *2000*	0/15
R-SHADE-10	**1.0**(1)	**2.1**(1)	16(3)	10(2)	**44**(2)	15/15
SOO-Derbel	8.8(3)	12(12)	43(9)	24(7)	3.3e4(2e4)	2/15

Table 4: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*6.3e+2*:33	*4.0e+2*:44	*1.6e+2*:109	*1.0e+2*:255	*2.5e+1*:3277	15/15
MATSUMOTO-	**1.9**(0.6)	**2.5**(0.7)	8.0(5)	10(9)	∞ *1000*	0/15
R-DE-10e2-	**1.4**(0.9)	**2.5**(1)	5.4(9)	**4.3**(1)	9.1(13)	1/15
R-DE-10e5-	4.3(2)	6.8(2)	10(2)	8.7(3)	**2.8**(0.6)	15/15
RL-SHADE-1	5.2(3)	6.4(0.6)	**5.1**(3)	5.9(5)	8.9(10)	1/15
RL-SHADE-1	10(7)	36(8)	75(11)	91(22)	28(2)	15/15
R-SHADE-10	**2.8**(2)	4.0(1)	**5.2**(0.4)	**3.9**(0.7)	**3.0**(5)	3/15
R-SHADE-10	3.8(3)	7.7(2)	20(8)	33(6)	7.7(0.6)	15/15
SOO-Derbel	5.0(3)	5.8(2)	12(5)	12(3)	234(655)	12/15

Table 5: 20-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+2*:22	*4.0e+2*:91	*2.5e+2*:250	*1.6e+2*:332	*6.3e+1*:1927	15/15
MATSUMOTO-	7.6(8)	5.1(4)	10(7)	∞	∞ *1000*	0/15
R-DE-10e2-	**4.4**(3)	**2.0**(0.4)	**1.5**(0.4)*	**2.2**(0.7)	∞ *2000*	0/15
R-DE-10e5-	13(4)	6.1(0.8)	4.0(0.4)	5.0(1)	**2.6**(0.3)	15/15
RL-SHADE-1	12(2)	3.6(0.5)	**1.9**(0.4)	4.2(3)	7.4(10)	2/15
RL-SHADE-1	40(19)	30(3)	25(3)	38(6)	29(3)	15/15
R-SHADE-10	7.9(2)	3.2(0.5)	**2.1**(0.3)	**2.4**(0.5)	**0.84**(0.1)	15/15
R-SHADE-10	11(4)	6.6(2)	5.5(0.8)	13(3)	9.0(1)	15/15
SOO-Derbel	**1.1**(2)*	**2.7**(0.8)	3.9(3)	7.0(1.0)	47(1)	15/15

Table 6: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*2.5e+2*:19	*1.6e+2*:34	*1.0e-8*:41	*1.0e-8*:41	*1.0e-8*:41	15/15
MATSUMOTO-	**1.8**(0.6)	**1.3**(0.1)$^{\star 4}$	**2.4**(0.1)$^{\star 4}$	**2.4**(1)$^{\star 4}$	**2.4**(0.1)$^{\star 4}$	15/15
R-DE-10e2-	**2.2**(0.7)	**3.1**(0.8)	∞	∞	∞ *2000*	0/15
R-DE-10e5-	4.3(3)	10(2)	332(7)	332(5)	332(8)	15/15
RL-SHADE-1	6.3(5)	7.9(0.5)	**242**(173)	**242**(333)	**242**(209)	3/15
RL-SHADE-1	8.0(23)	57(13)	1574(45)	1574(34)	1574(44)	15/15
R-SHADE-10	3.5(3)	4.8(0.6)	∞	∞	∞ *2000*	0/15
R-SHADE-10	6.1(5)	19(4)	959(45)	959(33)	959(31)	15/15
SOO-Derbel	10(0.0)	8.7(0.0)	4928(0.0)	4928(0.0)	4928(0.0)	15/15

Table 7: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*2.5e+5*:16	*6.3e+4*:43	*1.6e+4*:62	*1.6e+2*:353	*1.6e+1*:1078	15/15
MATSUMOTO-	**2.4**(0.2)	**1.4**(0.4)	**1.4**(0.7)*	9.5(21)	∞ *1000*	0/15
R-DE-10e2-	**2.9**(2)	**2.2**(1)	**2.4**(1)	**3.4**(2)	27(44)	1/15
R-DE-10e5-	6.8(7)	7.2(5)	8.8(2)	4.9(1.0)	**12**(3)	15/15
RL-SHADE-1	8.0(9)	6.3(1)	5.1(1)	3.9(2)	∞ *2000*	0/15
RL-SHADE-1	16(22)	32(19)	36(23)	19(5)	20(0.7)	15/15
R-SHADE-10	4.0(3)	4.0(2)	4.1(2)	**2.5**(1)	∞ *2000*	0/15
R-SHADE-10	7.7(10)	9.0(7)	11(6)	4.3(1)	**4.3**(0.4)	15/15
SOO-Derbel	**2.9**(3)	4.2(3)	5.3(2)	6.9(3)	1.3e4(2e4)	2/15

Table 8: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*1.0e+3*:11	*4.0e+2*:39	*2.5e+2*:74	*6.3e+1*:319	*1.0e+1*:1351	15/15
MATSUMOTO-	**1.2**(2)	**2.2**(2)	**2.0**(0.5)	5.6(6)	∞ *1000*	0/15
R-DE-10e2-	**1.9**(1)	**2.1**(1)	**2.0**(2)	**3.0**(2)	**11**(6)	2/15
R-DE-10e5-	**1.9**(3)	4.9(3)	5.0(3)	5.0(2)	12(6)	15/15
RL-SHADE-1	**2.0**(4)	5.6(1)	4.1(0.3)	3.5(6)	∞ *2000*	0/15
RL-SHADE-1	**2.5**(2)	22(16)	25(3)	22(6)	11(0.5)	15/15
R-SHADE-10	**2.1**(2)	3.0(1)	3.2(2)	**2.2**(0.7)	11(7)	2/15
R-SHADE-10	3.4(5)	4.5(2)	5.2(1)	3.1(0.4)	**2.0**(0.7)	15/15
SOO-Derbel	**1.3**(0.5)	**2.0**(1)	**1.9**(1)	5.7(5)	59(85)	15/15

Table 9: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*4.0e+4*:19	*2.5e+4*:35	*4.0e+3*:67	*2.5e+2*:231	*1.6e+1*:1470	15/15
MATSUMOTO-	**4.5**(1)	**2.9**(0.8)	**3.3**(1)	**4.0**(7)	∞ *1000*	0/15
R-DE-10e2-	5.1(1)	**3.6**(1)	**4.7**(1)	6.9(2)	∞ *2000*	0/15
R-DE-10e5-	15(5)	11(2)	12(3)	9.1(2)	51(5)	15/15
RL-SHADE-1	14(2)	8.2(0.9)	7.2(2)	22(20)	∞ *2000*	0/15
RL-SHADE-1	67(28)	50(14)	61(11)	48(4)	22(3)	15/15
R-SHADE-10	8.6(2)	5.4(1)	5.8(1)	**3.5**(1)	**20**(25)	1/15
R-SHADE-10	16(5)	9.3(3)	12(3)	8.3(1)	**7.0**(3)	15/15
SOO-Derbel	**3.8**(5)	4.5(5)	8.2(6)	12(4)	2107(3738)	6/15

Table 10: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*1.0e+2*:357	*6.3e+1*:560	*4.0e+1*:684	*2.5e+1*:756	*1.0e+1*:1716	15/15
MATSUMOTO-	7.9(9)	∞	∞	∞	∞ *1000*	0/15
R-DE-10e2-	7.0(3)	12(15)	14(12)	19(19)	∞ *2000*	0/15
R-DE-10e5-	20(7)	18(6)	16(7)	16(3)	3022(3021)	5/15
RL-SHADE-1	20(48)	∞	∞	∞	∞ *2000*	0/15
RL-SHADE-1	37(2)	26(2)	24(1)	25(1)	**33**(2)	15/15
R-SHADE-10	**5.3**(3)	**4.2**(4)	**5.3**(4)	**6.1**(5)	∞ *2000*	0/15
R-SHADE-10	**5.7**(0.5)	**4.3**(0.3)	**4.1**(0.4)	**4.6**(0.3)	**16**(2)	15/15
SOO-Derbel	10(3)	11(3)	15(5)	21(6)	5000(4642)	3/15

Table 11: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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*:15	*1.0e+6*:27	*4.0e+5*:70	*6.3e+4*:231	*4.0e+3*:1015	15/15
MATSUMOTO-	**4.4**(1)	**3.1**(1)	**3.4**(8)	9.1(8)	∞ *1000*	0/15
R-DE-10e2-	4.6(4)	4.2(0.8)	**4.1**(2)	12(13)	∞ *2000*	0/15
R-DE-10e5-	9.2(11)	10(6)	12(9)	70(21)	∞ *2e6*	0/15
RL-SHADE-1	12(4)	8.9(1)	5.9(3)	10(9)	∞ *2000*	0/15
RL-SHADE-1	32(15)	31(15)	38(12)	57(8)	**20**(2)	15/15
R-SHADE-10	6.9(3)	5.8(2)	5.8(3)	**5.6**(2)	29(27)	1/15
R-SHADE-10	10(6)	8.8(4)	8.0(2)	**6.4**(1.0)	**3.8**(2)	15/15
SOO-Derbel	**2.9**(4)	**3.1**(2)	4.8(4)	14(12)	1743(1348)	10/15

Table 12: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*4.0e+4*:11	*2.5e+3*:27	*1.6e+2*:313	*1.0e+2*:481	*1.0e+1*:1002	15/15
MATSUMOTO-	**1.7**(0.7)	**2.5**(2)	21(17)	∞	∞ *1000*	0/15
R-DE-10e2-	**1.7**(1)	**2.0**(1)	13(10)	∞	∞ *2000*	0/15
R-DE-10e5-	**1.7**(1.0)	**2.2**(2)	11(11)	51(83)	∞ *2e6*	0/15
RL-SHADE-1	**2.1**(4)	3.8(2)	11(23)	60(52)	∞ *2000*	0/15
RL-SHADE-1	**1.9**(2)	3.4(2)	14(12)	**22**(17)	**27**(2)	15/15
R-SHADE-10	**1.6**(1)	**1.6**(0.8)	**10**(7)	61(31)	∞ *2000*	0/15
R-SHADE-10	**1.8**(1)	**2.6**(2)	**2.3**(1)	**2.9**(1)	**7.6**(6)	15/15
SOO-Derbel	**1.4**(1.0)	**2.0**(2)	11(0.9)	343(1430)	∞ *2e6*	0/15

Table 13: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*1.0e+8*:23	*6.3e+7*:39	*2.5e+7*:76	*4.0e+6*:209	*1.0e+1*:1042	15/15
MATSUMOTO-	**3.1**(0.7)	**2.9**(0.4)	**3.0**(0.5)	3.6(3)	∞ *1000*	0/15
R-DE-10e2-	**3.1**(2)	**3.1**(2)	**3.3**(1)	**3.0**(2)	∞ *2000*	0/15
R-DE-10e5-	9.4(4)	8.1(2)	9.5(2)	7.1(0.5)	**31**(36)	15/15
RL-SHADE-1	11(1)	7.5(0.8)	5.0(0.4)	6.1(7)	∞ *2000*	0/15
RL-SHADE-1	37(19)	38(16)	44(7)	44(6)	37(9)	15/15
R-SHADE-10	7.8(2)	6.3(2)	5.0(1)	**3.2**(0.5)	∞ *2000*	0/15
R-SHADE-10	11(5)	9.1(3)	10(3)	8.1(1)	**8.5**(0.7)	15/15
SOO-Derbel	3.6(5)	5.0(4)	6.0(3)	11(9)	1110(972)	10/15

Table 14: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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
f13	*1.6e+3*:28	*1.0e+3*:64	*6.3e+2*:79	*4.0e+1*:211	*2.5e+0*:1724	15/15
MATSUMOTO-	**2.6**(0.6)	**1.8**(0.6)$^{\star 2}$	**2.6**(2)$^{\star 2}$	**8.0**(9)	∞ *1000*	0/15
R-DE-10e2-	**2.9**(1.0)	**3.6**(0.9)	**5.1**(1.0)	∞	∞ *2000*	0/15
R-DE-10e5-	7.1(2)	7.8(2)	12(2)	27(3)	43(46)	15/15
RL-SHADE-1	8.8(0.7)	5.2(0.2)	7.4(0.7)	∞	∞ *2000*	0/15
RL-SHADE-1	30(13)	41(5)	65(5)	93(2)	**18**(0.5)	15/15
R-SHADE-10	5.5(2)	4.8(1)	6.2(0.6)	**11**(3)	∞ *2000*	0/15
R-SHADE-10	8.5(3)	9.3(2)	13(1)	21(4)	**7.4**(5)	15/15
SOO-Derbel	**2.4**(3)	4.9(4)	11(4)	1339(1890)	5103(4830)	3/15

Table 15: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*2.5e+1*:15	*1.6e+1*:42	*1.0e+1*:75	*1.6e+0*:219	*6.3e-4*:1106	15/15
MATSUMOTO-	**5.7**(2)	**3.1**(1)	**2.9**(1)	9.5(11)	∞ *1000*	0/15
R-DE-10e2-	7.3(5)	**4.5**(1)	**4.5**(4)	**6.1**(7)	∞ *2000*	0/15
R-DE-10e5-	13(7)	10(4)	9.4(4)	9.3(2)	∞ *2e6*	0/15
RL-SHADE-1	17(6)	7.7(0.7)	5.8(1)	31(43)	∞ *2000*	0/15
RL-SHADE-1	58(27)	45(10)	40(15)	47(4)	**30**(1)	15/15
R-SHADE-10	12(5)	6.0(2)	4.8(1)	**3.7**(0.6)	∞ *2000*	0/15
R-SHADE-10	13(4)	8.4(1)	8.2(2)	9.0(2)	**10**(2)	15/15
SOO-Derbel	**3.9**(6)	4.7(3)	5.7(3)	34(8)	∞ *2e6*	0/15

Table 16: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*6.3e+2*:15	*4.0e+2*:67	*2.5e+2*:292	*1.6e+2*:846	*1.0e+2*:1671	15/15
MATSUMOTO-	**3.5**(1)	**1.7**(0.4)	**1.1**(1.0)	**2.0**(2)	8.8(12)	1/15
R-DE-10e2-	3.7(2)	**2.3**(2)	**2.2**(3)	11(12)	∞ *2000*	0/15
R-DE-10e5-	7.2(3)	5.0(2)	3.0(0.5)	4.5(3)	32(44)	15/15
RL-SHADE-1	12(4)	4.4(0.4)	**1.9**(0.8)	3.2(1)	18(37)	1/15
RL-SHADE-1	16(25)	24(13)	13(3)	13(3)	23(10)	15/15
R-SHADE-10	6.7(4)	3.5(1)	**1.6**(0.6)	**1.3**(0.4)	**2.8**(6)	6/15
R-SHADE-10	7.1(5)	5.2(1)	**2.6**(0.5)	**2.9**(0.7)	6.3(2)	15/15
SOO-Derbel	**1.4**(3)*	**2.6**(3)	**2.0**(0.8)	3.6(2)	**3.2**(0.8)	15/15

Table 17: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*4.0e+1*:26	*2.5e+1*:127	*1.6e+1*:540	*1.6e+1*:540	*1.0e+1*:1384	15/15
MATSUMOTO-	3.4(3)	4.9(4)	3.3(2)	3.3(2)	11(15)	1/15
R-DE-10e2-	**2.1**(4)	39(29)	∞	∞	∞ *2000*	0/15
R-DE-10e5-	3.4(3)	18(24)	50(30)	50(50)	236(346)	15/15
RL-SHADE-1	3.8(3)	3.8(1)	**2.8**(2)	**2.8**(2)	**6.8**(13)	3/15
RL-SHADE-1	3.8(5)	38(47)	92(45)	92(22)	73(22)	15/15
R-SHADE-10	**1.8**(1)	**3.4**(1)	3.2(1)	3.2(3)	22(20)	1/15
R-SHADE-10	4.1(4)	23(16)	23(7)	23(6)	27(21)	15/15
SOO-Derbel	**2.3**(3)	**3.6**(2)	**2.2**(1.0)	**2.2**(1)	**1.6**(0.4)	15/15

Table 18: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*1.6e+1*:11	*1.0e+1*:63	*6.3e+0*:305	*4.0e+0*:468	*1.0e+0*:1030	15/15
MATSUMOTO-	3.3(2)	**2.2**(0.6)	**1.3**(0.5)	32(46)	∞ *1000*	0/15
R-DE-10e2-	**2.3**(2)	**2.1**(2)	**1.7**(0.8)	4.1(2)	∞ *2000*	0/15
R-DE-10e5-	4.4(4)	4.1(1.0)	**2.7**(1)	4.3(1)	**12**(6)	15/15
RL-SHADE-1	5.5(10)	3.9(2)	**2.2**(2)	11(6)	∞ *2000*	0/15
RL-SHADE-1	4.6(13)	13(8)	10(3)	13(2)	18(2)	15/15
R-SHADE-10	6.0(6)	3.7(1)	**1.7**(0.7)	**3.3**(2)	29(39)	1/15
R-SHADE-10	5.3(13)	3.7(2)	**1.9**(0.9)	**2.3**(0.9)	**3.4**(0.8)	15/15
SOO-Derbel	**0.99**(3)$^{\star 2}$	**1.3**(0.9)	**1.6**(1)	3.5(2)	16(6)	15/15

Table 19: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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+1*:116	*2.5e+1*:252	*1.6e+1*:430	*1.0e+1*:621	*4.0e+0*:1090	15/15
MATSUMOTO-	**1.0**(0.4)	**2.4**(5)	8.3(12)	∞	∞ *1000*	0/15
R-DE-10e2-	**0.78**(0.6)	**1.2**(0.7)	**2.1**(1)	8.4(13)	∞ *2000*	0/15
R-DE-10e5-	**1.8**(0.7)	**2.4**(0.6)	4.1(2)	6.2(2)	28(44)	15/15
RL-SHADE-1	**2.0**(1)	**2.2**(3)	4.2(0.7)	11(13)	∞ *2000*	0/15
RL-SHADE-1	5.3(2)	10(4)	13(2)	17(2)	18(1)	15/15
R-SHADE-10	**1.4**(0.6)	**2.0**(1)	3.0(1)	5.6(7)	28(44)	1/15
R-SHADE-10	**1.8**(0.6)	**2.0**(0.8)	**2.7**(0.8)	**3.1**(0.7)	**3.4**(0.8)	15/15
SOO-Derbel	**0.54**(0.5)	**1.6**(2)	3.2(2)	**5.0**(2)	**11**(5)	15/15

Table 20: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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*:2.5e5	*1.0e-1*:3.4e5	*6.3e-2*:3.4e5	*4.0e-2*:3.4e5	*2.5e-2*:3.4e5	3/15
MATSUMOTO-	∞	∞	∞	∞	∞ *1000*	0/15
R-DE-10e2-	∞	∞	∞	∞	∞ *2000*	0/15
R-DE-10e5-	∞	∞	∞	∞	∞ *2e6*	0/15
RL-SHADE-1	∞	∞	∞	∞	∞ *2000*	0/15
RL-SHADE-1	115(99)	∞	∞	∞	∞ *2e6*	0/15
R-SHADE-10	∞	∞	∞	∞	∞ *2000*	0/15
R-SHADE-10	**55**(59)	∞	∞	∞	∞ *2e6*	0/15
SOO-Derbel	**0.55**(2)	**3.2**(3)	**26**(24)	**83**(95)	∞ *2e6*	0/15

Table 21: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*1.6e+4*:38	*1.0e+4*:42	*2.5e+2*:62	*2.5e+0*:250	*1.6e+0*:2536	15/15
MATSUMOTO-	**2.2**(0.5)	**2.4**(0.9)	**4.6**(2)	**4.1**(5)	∞ *1000*	0/15
R-DE-10e2-	**1.9**(0.9)	**2.3**(1)	**4.9**(7)	**2.9**(0.6)	**1.0**(0.6)	9/15
R-DE-10e5-	6.4(4)	8.4(2)	15(2)	13(2)	4.5(2)	15/15
RL-SHADE-1	7.1(0.5)	6.9(1.0)	8.8(2)	7.4(6)	12(16)	1/15
RL-SHADE-1	22(11)	30(11)	68(15)	133(61)	35(5)	15/15
R-SHADE-10	4.4(1)	4.8(2)	7.7(0.5)	4.3(0.4)	**1.8**(2)	6/15
R-SHADE-10	5.0(2)	5.7(2)	10(3)	44(22)	13(3)	15/15
SOO-Derbel	4.8(0.0)	5.6(0.0)	51(8e-3)	13(2e-3)	3.6(2e-4)	15/15

Table 22: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*6.3e+1*:36	*4.0e+1*:77	*4.0e+1*:77	*1.6e+1*:456	*4.0e+0*:1094	15/15
MATSUMOTO-	**2.7**(0.5)	**1.9**(1)*	**1.9**(0.4)*	**0.56**(0.5)	**0.97**(2)	9/15
R-DE-10e2-	**3.7**(2)	**3.1**(1)	**3.1**(0.9)	**1.2**(0.4)	**2.2**(3)	8/15
R-DE-10e5-	14(7)	12(5)	12(6)	7.3(3)	41(57)	15/15
RL-SHADE-1	9.1(2)	6.6(2)	6.6(3)	4.5(5)	6.1(2)	4/15
RL-SHADE-1	56(16)	52(14)	52(14)	15(3)	169(642)	14/15
R-SHADE-10	8.4(4)	7.0(2)	7.0(2)	3.5(3)	5.0(4)	5/15
R-SHADE-10	13(6)	11(3)	11(3)	**3.0**(0.9)	17(8)	15/15
SOO-Derbel	3.8(1)	4.6(5)	4.6(5)	**2.4**(2)	8.7(40)	15/15

Table 23: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*6.3e+1*:45	*4.0e+1*:68	*4.0e+1*:68	*1.6e+1*:231	*6.3e+0*:1219	15/15
MATSUMOTO-	**2.2**(0.6)	**2.1**(0.6)$^{\star 2}$	**2.1**(2)$^{\star 2}$	**1.5**(2)*	**0.58**(0.7)*	11/15
R-DE-10e2-	**4.4**(2)	**4.7**(2)	**4.7**(2)	6.9(9)	5.2(3)	4/15
R-DE-10e5-	12(8)	13(5)	13(9)	81(66)	25(47)	15/15
RL-SHADE-1	7.7(2)	7.2(4)	7.2(2)	7.2(13)	**2.5**(2)	8/15
RL-SHADE-1	49(17)	59(12)	59(9)	493(1742)	221(2)	14/15
R-SHADE-10	8.1(5)	10(3)	10(3)	**5.1**(4)	**2.1**(2)	9/15
R-SHADE-10	12(7)	12(3)	12(10)	28(161)	14(29)	15/15
SOO-Derbel	4.7(2)	6.8(4)	6.8(5)	180(656)	69(247)	15/15

Table 24: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*6.3e+0*:29	*4.0e+0*:118	*2.5e+0*:306	*2.5e+0*:306	*1.0e+0*:1614	15/15
MATSUMOTO-	**2.0**(2)	10(15)	∞	∞	∞ *1000*	0/15
R-DE-10e2-	**2.1**(2)	6.2(6)	95(126)	95(110)	∞ *2000*	0/15
R-DE-10e5-	**1.1**(2)	6.5(5)	81(17)	81(56)	∞ *2e6*	0/15
RL-SHADE-1	**2.7**(3)	5.4(5)	48(61)	48(39)	∞ *2000*	0/15
RL-SHADE-1	**1.0**(1)	**4.0**(3)	57(11)	57(79)	116(25)	15/15
R-SHADE-10	**2.4**(2)	6.7(6)	**45**(81)	**45**(46)	∞ *2000*	0/15
R-SHADE-10	**1.7**(1)	**4.8**(5)	68(100)	68(88)	**95**(97)	15/15
SOO-Derbel	**1.7**(4)	10(11)	**12**(10)	**12**(3)	**3.9**(2)	15/15

Table 25: 20-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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	*2.5e+2*:208	*1.6e+2*:918	*1.0e+2*:6628	*6.3e+1*:9885	*4.0e+1*:31629	15/15
MATSUMOTO-	15(10)	∞	∞	∞	∞ *1000*	0/15
R-DE-10e2-	**2.6**(3)	15(13)	∞	∞	∞ *2000*	0/15
R-DE-10e5-	4.8(1)	16(5)	592(369)	∞	∞ *2e6*	0/15
RL-SHADE-1	3.0(1)	**5.7**(6)	∞	∞	∞ *2000*	0/15
RL-SHADE-1	16(5)	16(2)	15(2)	20(3)	**9.4**(1.0)	15/15
R-SHADE-10	3.0(0.7)	32(39)	∞	∞	∞ *2000*	0/15
R-SHADE-10	**2.7**(0.5)	**2.7**(2)	**3.2**(1)	**5.4**(0.6)	**6.1**(3)	15/15
SOO-Derbel	6.9(4)	7.3(5)	**5.6**(3)	**13**(17)	20(9)	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.