Comparison Tables: CEC BBOB 2015 Testbed in 5-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: 05-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	*2.5e+1*:4.8	*1.6e+1*:7.6	*1.0e-8*:12	*1.0e-8*:12	*1.0e-8*:12	15/15
MATSUMOTO-	**1.8**(1)	**1.7**(0.8)	∞	∞	∞ *250*	0/15
R-DE-10e2-	**2.5**(4)	**2.6**(2)	∞	∞	∞ *500*	0/15
R-DE-10e5-	**2.5**(2)	**2.8**(3)	**67**(13)	**67**(8)	**67**(14)	15/15
RL-SHADE-1	**1.7**(1)	**2.7**(1)	614(553)	614(430)	614(932)	1/15
RL-SHADE-1	**2.8**(11)	3.8(5)	632(17)	632(20)	632(36)	15/15
R-SHADE-10	**2.8**(2)	**2.6**(1)	∞	∞	∞ *500*	0/15
R-SHADE-10	**2.2**(2)	3.4(2)	**110**(9)	**110**(9)	**110**(11)	15/15
SOO-Derbel	**0.99**(1)	**1.3**(0.8)	194(12)	194(14)	194(18)	15/15

Table 3: 05-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	*1.6e+6*:2.9	*4.0e+5*:11	*4.0e+4*:15	*6.3e+2*:58	*1.0e-8*:95	15/15
MATSUMOTO-	**1.6**(1.0)	**0.68**(0.4)	**4.0**(5)	∞	∞ *250*	0/15
R-DE-10e2-	**2.0**(4)	**1.2**(1)	**4.0**(2)	**2.8**(0.5)	∞ *500*	0/15
R-DE-10e5-	**2.1**(1)	**1.6**(2)	4.4(4)	**2.9**(0.9)	**10**(0.6)	15/15
RL-SHADE-1	**2.5**(3)	**1.4**(0.6)	5.2(3)	**2.9**(0.7)	∞ *500*	0/15
RL-SHADE-1	**1.8**(2)	**0.92**(3)	10(10)	22(5)	113(4)	15/15
R-SHADE-10	**2.9**(8)	**2.2**(3)	4.6(4)	3.7(0.6)	∞ *500*	0/15
R-SHADE-10	**1.7**(0.7)	**0.85**(1)	4.2(6)	5.4(1)	**22**(2)	15/15
SOO-Derbel	6.6(11)	**2.5**(3)	6.7(5)	6.6(2)	872(2643)	13/15

Table 4: 05-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	*1.6e+2*:4.1	*1.0e+2*:15	*6.3e+1*:23	*2.5e+1*:73	*1.0e+1*:716	15/15
MATSUMOTO-	**1.8**(2)	**1.2**(1)	**1.8**(1)	**2.7**(2)	**1.1**(1)	4/15
R-DE-10e2-	**2.5**(3)	**1.1**(1)	**1.5**(0.9)	**1.8**(1)	**0.41**(0.1)	15/15
R-DE-10e5-	**1.9**(3)	**1.5**(1)	**1.7**(1)	**1.5**(0.8)	**0.36**(0.1)	15/15
RL-SHADE-1	**2.9**(2)	**2.5**(1)	**2.5**(1)	**1.7**(0.4)	**0.33**(0.4)	14/15
RL-SHADE-1	**1.3**(1.0)	**1.9**(4)	7.8(2)	11(3)	3.1(1)	15/15
R-SHADE-10	4.6(3)	**2.3**(1)	**2.3**(1.0)	**1.6**(0.5)	**0.33**(0.0)	15/15
R-SHADE-10	**1.5**(2)	**1.1**(1)	**2.4**(0.9)	3.0(0.6)	**1.1**(0.5)	15/15
SOO-Derbel	**1.9**(1)	**1.2**(0.5)	**1.9**(1.0)	**2.4**(1)	**1.2**(2)	15/15

Table 5: 05-D, running time excess ERT/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 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	*2.5e+2*:2.6	*1.6e+2*:10	*1.0e+2*:19	*4.0e+1*:65	*1.6e+1*:434	15/15
MATSUMOTO-	**2.6**(4)	**1.1**(0.7)	**2.9**(2)	3.0(1)	9.0(12)	1/15
R-DE-10e2-	**2.1**(2)	**1.5**(1)	**1.8**(0.7)	**1.9**(0.5)	**0.92**(0.4)	12/15
R-DE-10e5-	4.3(4)	**2.5**(2)	**2.7**(2)	**2.2**(1)	**1.3**(2)	15/15
RL-SHADE-1	**2.7**(0.8)	**2.0**(2)	**2.5**(2)	**1.9**(0.7)	**0.49**(0.2)	15/15
RL-SHADE-1	3.0(1)	**2.2**(1)	4.2(5)	10(3)	4.1(1)	15/15
R-SHADE-10	3.3(3)	**2.0**(2)	**2.0**(1)	**1.6**(0.6)	**0.54**(0.2)	15/15
R-SHADE-10	**2.4**(2)	**1.5**(0.7)	**2.1**(1)	**2.5**(0.6)	**1.4**(0.5)	15/15
SOO-Derbel	**0.69**(0)	**0.68**(0.7)	**1.3**(1)	**2.3**(1)	3.7(19)	15/15

Table 6: 05-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	*6.3e+1*:4.0	*4.0e+1*:10	*1.0e-8*:10	*1.0e-8*:10	*1.0e-8*:10	15/15
MATSUMOTO-	**1.6**(0.8)	**1.2**(0.6)	**1.9**(0.3)$^{\star 4}$	**1.9**(0.3)$^{\star 4}$	**1.9**(0.3)$^{\star 4}$	15/15
R-DE-10e2-	**2.4**(2)	**2.7**(1)	∞	∞	∞ *500*	0/15
R-DE-10e5-	**2.3**(3)	**2.4**(1)	**184**(115)	**184**(44)	**184**(93)	15/15
RL-SHADE-1	**2.0**(2)	3.1(3)	372(325)	372(187)	372(538)	2/15
RL-SHADE-1	3.4(3)	4.4(5)	613(20)	613(21)	613(27)	15/15
R-SHADE-10	3.9(5)	3.9(2)	∞	∞	∞ *500*	0/15
R-SHADE-10	**2.6**(4)	5.6(3)	219(30)	219(20)	219(15)	15/15
SOO-Derbel	3.6(0.1)	**1.4**(0.1)	1054(0.1)	1054(0.1)	1054(0.1)	15/15

Table 7: 05-D, running time excess ERT/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 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	*1.0e+5*:3.0	*2.5e+4*:8.4	*1.0e+2*:16	*2.5e+1*:54	*2.5e-1*:254	15/15
MATSUMOTO-	**1.3**(0.5)	**0.90**(0.7)	**1.7**(1)	12(14)	∞ *250*	0/15
R-DE-10e2-	**1.9**(2)	**2.0**(2)	3.5(4)	**2.0**(2)	**6.9**(10)	4/15
R-DE-10e5-	**2.7**(2)	**1.9**(1)	3.5(3)	**2.4**(2)	16(40)	15/15
RL-SHADE-1	3.0(3)	**1.9**(5)	3.8(3)	**2.4**(2)	14(10)	2/15
RL-SHADE-1	**1.9**(3)	**2.6**(3)	11(10)	6.7(5)	13(2)	15/15
R-SHADE-10	**1.9**(2)	**1.6**(1)	3.3(4)	**2.3**(2)	29(35)	1/15
R-SHADE-10	**2.4**(1)	**2.5**(4)	4.2(4)	**2.7**(2)	**2.7**(0.9)	15/15
SOO-Derbel	**1.7**(4)	**1.2**(2)	**1.8**(2)	**2.0**(2)	6105(3636)	4/15

Table 8: 05-D, running time excess ERT/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 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	*1.6e+2*:4.2	*1.0e+2*:6.2	*2.5e+1*:20	*4.0e+0*:54	*1.0e+0*:324	15/15
MATSUMOTO-	**1.3**(2)	**1.8**(1)	**1.5**(0.7)	7.6(9)	5.4(4)	2/15
R-DE-10e2-	**2.3**(2)	**2.3**(5)	**2.5**(1)	5.4(5)	5.3(5)	4/15
R-DE-10e5-	**2.0**(0.9)	**2.1**(1)	**2.8**(1)	4.0(3)	**2.7**(1.0)	15/15
RL-SHADE-1	**1.5**(2)	**2.3**(4)	3.5(2)	5.9(5)	**2.8**(1)	7/15
RL-SHADE-1	**2.2**(0.9)	**2.1**(1)	8.2(9)	12(7)	5.5(2)	15/15
R-SHADE-10	**2.9**(3)	3.1(2)	4.0(3)	**3.9**(2)	**1.9**(3)	10/15
R-SHADE-10	**2.3**(2)	**2.6**(5)	**2.4**(2)	**2.6**(2)	**1.3**(2)	15/15
SOO-Derbel	**1.1**(0.9)	**2.2**(1)	**2.1**(2)	4.1(4)	**2.1**(2)	15/15

Table 9: 05-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	*1.0e+4*:4.6	*6.3e+3*:6.8	*1.0e+3*:18	*6.3e+1*:54	*1.6e+0*:258	15/15
MATSUMOTO-	**1.7**(1)	**1.9**(1)	**1.4**(0.6)	**2.6**(2)	∞ *250*	0/15
R-DE-10e2-	3.3(4)	**2.9**(2)	**2.9**(2)	**2.4**(1)	**9.0**(19)	3/15
R-DE-10e5-	**2.1**(3)	**2.0**(1)	**2.0**(1)	**2.6**(0.6)	22(57)	15/15
RL-SHADE-1	3.3(2)	3.8(4)	3.9(0.5)	3.6(5)	14(12)	2/15
RL-SHADE-1	**2.5**(2)	3.5(7)	4.9(4)	13(6)	17(3)	15/15
R-SHADE-10	**2.4**(3)	**1.8**(2)	**2.5**(1)	3.4(2)	∞ *500*	0/15
R-SHADE-10	3.6(5)	3.0(3)	3.2(4)	3.9(3)	**4.3**(2)	15/15
SOO-Derbel	**1.4**(0.5)	**1.3**(2)	**1.4**(1.0)	**2.1**(1)	16(42)	15/15

Table 10: 05-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	*2.5e+1*:20	*1.6e+1*:26	*1.0e+1*:35	*4.0e+0*:62	*1.6e-2*:256	15/15
MATSUMOTO-	18(20)	34(13)	35(44)	64(65)	∞ *250*	0/15
R-DE-10e2-	**11**(7)	10(5)	**8.1**(3)	11(13)	∞ *500*	0/15
R-DE-10e5-	12(4)	13(12)	11(6)	10(5)	487(633)	15/15
RL-SHADE-1	13(10)	11(3)	10(4)	15(17)	∞ *500*	0/15
RL-SHADE-1	54(20)	51(10)	46(11)	45(4)	**28**(3)	15/15
R-SHADE-10	14(7)	13(8)	10(4)	28(20)	∞ *500*	0/15
R-SHADE-10	11(4)	**10**(1)	8.1(3)	**8.5**(1)	**9.1**(8)	15/15
SOO-Derbel	**6.5**(2)	**6.3**(2)	**5.7**(1)	**5.1**(2)	202(348)	15/15

Table 11: 05-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	*2.5e+6*:2.9	*6.3e+5*:7.0	*2.5e+5*:17	*6.3e+3*:54	*2.5e+1*:297	15/15
MATSUMOTO-	**1.5**(2)	**1.9**(1)	**1.4**(0.9)	6.4(9)	∞ *250*	0/15
R-DE-10e2-	**1.4**(2)	**1.2**(0.7)	**1.2**(1)	**2.6**(0.7)	∞ *500*	0/15
R-DE-10e5-	**1.3**(3)	**1.9**(2)	**1.2**(0.4)	7.1(4)	169(117)	15/15
RL-SHADE-1	**1.6**(3)	**1.6**(0.9)	**1.5**(2)	4.5(3)	**12**(15)	2/15
RL-SHADE-1	**1.6**(1)	**2.3**(2)	**1.4**(0.5)	14(6)	15(1)	15/15
R-SHADE-10	**1.7**(1)	**1.7**(3)	**1.2**(2)	6.7(8)	∞ *500*	0/15
R-SHADE-10	**1.8**(3)	**1.2**(1)	**1.3**(2)	**3.5**(1)	**2.2**(0.6)	15/15
SOO-Derbel	**1.4**(0.7)	**0.74**(0.8)	**0.56**(0.6)	4.6(2)	74(254)	15/15

Table 12: 05-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+6*:3.0	*6.3e+4*:6.2	*6.3e+2*:16	*6.3e+1*:74	*6.3e-1*:298	15/15
MATSUMOTO-	**1.4**(2)	**2.5**(2)	4.7(1)	8.9(6)	∞ *250*	0/15
R-DE-10e2-	**2.3**(4)	**2.0**(2)	5.1(5)	3.8(3)	∞ *500*	0/15
R-DE-10e5-	**1.9**(2)	3.3(1)	4.6(2)	36(121)	4079(4104)	5/15
RL-SHADE-1	**1.5**(0.8)	**2.2**(5)	**3.7**(2)	**3.1**(3)	∞ *500*	0/15
RL-SHADE-1	**3.0**(6)	3.4(5)	6.6(5)	6.7(3)	**15**(3)	15/15
R-SHADE-10	**2.0**(3)	**1.8**(2)	4.0(5)	4.0(4)	∞ *500*	0/15
R-SHADE-10	**1.2**(1)	4.2(3)	**3.9**(3)	**2.2**(0.9)	**2.7**(1)	15/15
SOO-Derbel	**0.98**(0.3)	**1.6**(3)	5.0(2)	7.5(20)	2683(2985)	6/15

Table 13: 05-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	*4.0e+7*:3.6	*1.6e+7*:7.6	*4.0e+6*:19	*1.6e+4*:52	*1.0e+0*:268	15/15
MATSUMOTO-	**1.2**(0.6)	**1.4**(1.0)	**1.7**(0.5)	**3.2**(1)	∞ *250*	0/15
R-DE-10e2-	**1.8**(2)	**2.6**(4)	3.1(1)	**4.5**(2)	∞ *500*	0/15
R-DE-10e5-	**1.6**(2)	**2.3**(2)	**2.4**(2)	5.3(3)	112(66)	15/15
RL-SHADE-1	**1.4**(1)	**2.5**(1.0)	3.3(1)	4.8(0.9)	∞ *500*	0/15
RL-SHADE-1	**2.1**(2)	4.0(1)	6.9(10)	39(11)	29(4)	15/15
R-SHADE-10	**2.1**(0.6)	**2.7**(1)	**2.9**(1)	5.6(2)	∞ *500*	0/15
R-SHADE-10	**0.96**(0.8)	**1.8**(1)	**2.3**(2)	6.0(2)	**7.1**(8)	15/15
SOO-Derbel	**0.50**(0.1)	**0.66**(2)	**1.2**(0.7)	5.4(2)	**6.8**(1)	15/15

Table 14: 05-D, running time excess ERT/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 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	*1.0e+3*:2.8	*6.3e+2*:8.4	*4.0e+2*:17	*6.3e+1*:52	*6.3e-2*:264	15/15
MATSUMOTO-	**2.0**(1.0)	**1.8**(1)	**1.5**(0.5)	**1.7**(0.3)$^{\star 2}$	∞ *250*	0/15
R-DE-10e2-	**2.5**(2)	**2.7**(0.7)	**2.4**(2)	4.5(1)	∞ *500*	0/15
R-DE-10e5-	**2.0**(2)	**3.0**(2)	3.3(2)	4.6(3)	739(735)	13/15
RL-SHADE-1	**1.7**(0.8)	**2.6**(3)	3.4(2)	3.9(0.9)	∞ *500*	0/15
RL-SHADE-1	3.4(8)	5.0(5)	6.7(5)	23(6)	**28**(2)	15/15
R-SHADE-10	**2.5**(5)	3.4(2)	**2.8**(1)	5.3(5)	∞ *500*	0/15
R-SHADE-10	**1.2**(0.8)	**1.9**(5)	**2.8**(3)	4.2(2)	**4.8**(2)	15/15
SOO-Derbel	**1.0**(2)	**0.96**(0.7)	**1.2**(0.4)	**3.9**(0.3)	29(26)	15/15

Table 15: 05-D, running time excess ERT/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 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*:3.0	*1.0e+1*:10	*6.3e+0*:15	*2.5e-1*:53	*1.0e-5*:251	15/15
MATSUMOTO-	**2.1**(1)	**1.4**(0.8)	**1.3**(0.5)	**2.3**(1)	∞ *250*	0/15
R-DE-10e2-	3.4(6)	**1.7**(2)	**2.3**(2)	**3.4**(2)	∞ *500*	0/15
R-DE-10e5-	3.5(2)	**2.1**(2)	**1.8**(2)	3.7(2)	579(1050)	14/15
RL-SHADE-1	**2.3**(2)	**2.2**(3)	**2.6**(3)	3.6(1)	∞ *500*	0/15
RL-SHADE-1	5.3(3)	3.1(4)	3.5(3)	20(7)	**29**(2)	15/15
R-SHADE-10	**2.7**(4)	**1.6**(2)	**2.1**(2)	3.9(0.5)	∞ *500*	0/15
R-SHADE-10	3.2(2)	**1.3**(0.7)	**1.4**(1)	4.4(0.6)	**5.3**(0.9)	15/15
SOO-Derbel	**1.0**(1)	**0.59**(0.8)	**0.74**(0.9)	3.6(1)	1342(1364)	12/15

Table 16: 05-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*:3.0	*1.0e+2*:13	*6.3e+1*:24	*4.0e+1*:55	*1.6e+1*:289	5/5
MATSUMOTO-	**2.4**(2)	**1.3**(0.1)	**1.6**(2)	**1.7**(1.0)	**0.97**(0.8)	10/15
R-DE-10e2-	**2.6**(2)	**1.6**(0.8)	**1.8**(0.8)	**1.9**(2)	**2.1**(0.8)	10/15
R-DE-10e5-	**2.9**(2)	**1.6**(1)	**1.8**(1)	**1.6**(0.4)	3.4(7)	15/15
RL-SHADE-1	**2.5**(2)	**2.8**(2)	3.1(0.8)	**2.0**(0.4)	3.0(3)	7/15
RL-SHADE-1	**1.9**(3)	**2.6**(4)	6.1(5)	7.3(5)	7.6(3)	15/15
R-SHADE-10	4.8(5)	**2.3**(2)	**2.4**(1)	**2.0**(0.4)	**2.8**(3)	8/15
R-SHADE-10	4.2(6)	**2.0**(1)	**2.4**(1)	**2.4**(3)	**2.0**(0.6)	15/15
SOO-Derbel	**1.1**(2)	**0.64**(0.5)	**1.3**(0.7)	**1.8**(2)	**1.6**(0.6)	15/15

Table 17: 05-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	*4.0e+1*:4.8	*2.5e+1*:16	*1.6e+1*:46	*1.0e+1*:120	*4.0e+0*:334	15/15
MATSUMOTO-	**1.4**(2)	**0.86**(0.7)	**1.2**(2)	**1.2**(0.9)	3.9(4)	3/15
R-DE-10e2-	**1.4**(0.8)	**1.9**(2)	**1.1**(1)	**0.95**(1)	5.3(4)	4/15
R-DE-10e5-	**1.0**(0.9)	**1.2**(1.0)	**1.7**(2)	**1.7**(2)	4.5(5)	15/15
RL-SHADE-1	**1.8**(0.8)	**1.4**(2)	**1.3**(0.5)	**1.2**(0.9)	**1.7**(1.0)	10/15
RL-SHADE-1	**1.1**(1)	**1.3**(0.6)	**1.6**(3)	**1.6**(2)	7.5(5)	15/15
R-SHADE-10	**1.9**(2)	**1.6**(2)	**1.6**(1)	**1.4**(0.7)	**2.2**(4)	8/15
R-SHADE-10	**1.7**(2)	**0.85**(0.5)	**1.5**(2)	**1.1**(0.8)	**2.7**(1)	15/15
SOO-Derbel	**1.7**(2)	**1.6**(1)	**1.9**(1)	**1.2**(0.7)	**0.85**(0.5)	15/15

Table 18: 05-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	*1.0e+1*:5.2	*6.3e+0*:26	*4.0e+0*:57	*2.5e+0*:110	*6.3e-1*:412	15/15
MATSUMOTO-	3.1(2)	**1.2**(1)	**1.1**(0.5)	**1.9**(1)	9.2(8)	1/15
R-DE-10e2-	4.1(4)	**1.8**(1)	**1.5**(0.8)	**1.2**(0.7)	**0.94**(0.2)	12/15
R-DE-10e5-	4.2(2)	**1.8**(2)	**1.7**(1)	5.0(7)	7.4(11)	15/15
RL-SHADE-1	**2.4**(2)	**1.0**(1)	**1.1**(0.5)	**1.3**(0.8)	**1.8**(2)	8/15
RL-SHADE-1	**2.8**(2)	**1.7**(1)	3.1(3)	4.3(2)	4.6(1)	15/15
R-SHADE-10	3.3(4)	**2.1**(2)	**1.9**(1)	**1.8**(2)	**2.7**(2)	6/15
R-SHADE-10	**2.8**(2)	**1.3**(0.4)	**1.2**(0.9)	**1.3**(1)	**1.0**(0.4)	15/15
SOO-Derbel	**1.4**(2)	**0.64**(0.6)	**0.86**(0.7)	**0.93**(0.5)	**0.92**(0.6)	15/15

Table 19: 05-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	*6.3e+1*:3.4	*4.0e+1*:7.2	*2.5e+1*:20	*1.6e+1*:58	*1.6e+0*:318	15/15
MATSUMOTO-	**1.9**(2)	**1.4**(2)	**1.1**(0.9)	**0.85**(1)	∞ *250*	0/15
R-DE-10e2-	**2.1**(2)	**2.5**(2)	**1.9**(2)	**1.3**(0.3)	4.2(2)	5/15
R-DE-10e5-	**2.6**(1)	**2.7**(6)	**2.1**(2)	**1.3**(0.5)	**2.9**(3)	15/15
RL-SHADE-1	**1.5**(3)	**1.7**(1)	**2.2**(2)	**1.4**(1)	4.4(3)	5/15
RL-SHADE-1	**1.1**(0.8)	**2.1**(1)	3.1(2)	3.9(3)	8.2(3)	15/15
R-SHADE-10	**1.5**(1)	**2.8**(5)	**1.8**(2)	**1.7**(0.8)	12(13)	2/15
R-SHADE-10	**1.4**(1)	**1.5**(2)	**1.8**(2)	**1.5**(0.9)	**1.5**(1)	15/15
SOO-Derbel	**0.76**(0.8)	**1.0**(1)	**0.76**(0.4)	**0.79**(0.5)	**1.6**(0.7)	15/15

Table 20: 05-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*:172	*1.0e-1*:242	*6.3e-2*:675	*4.0e-2*:3078	*2.5e-2*:4946	15/15
MATSUMOTO-	∞	∞	∞	∞	∞ *250*	0/15
R-DE-10e2-	∞	∞	∞	∞	∞ *500*	0/15
R-DE-10e5-	1109(1356)	1951(2356)	3319(4999)	1163(1634)	1470(1238)	1/15
RL-SHADE-1	∞	∞	∞	∞	∞ *500*	0/15
RL-SHADE-1	273(149)	269(150)	134(116)	70(52)	116(136)	9/15
R-SHADE-10	∞	∞	∞	∞	∞ *500*	0/15
R-SHADE-10	**90**(115)	**111**(153)	**85**(75)	**68**(108)	**61**(82)	11/15
SOO-Derbel	**3.7**(1)	**10**(28)	**6.3**(0.3)	**1.4**(0.1)	**0.90**(3)	15/15

Table 21: 05-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	*6.3e+3*:5.1	*4.0e+3*:8.4	*4.0e+1*:15	*2.5e+0*:69	*1.0e+0*:851	15/15
MATSUMOTO-	**1.7**(1)	**1.4**(1)	**1.9**(1)	4.1(3)	∞ *250*	0/15
R-DE-10e2-	**2.1**(3)	**1.8**(2)	3.5(3)	**2.0**(0.7)	**1.1**(1)	7/15
R-DE-10e5-	**2.4**(2)	**1.8**(0.8)	**3.3**(2)	**2.2**(0.8)	**0.95**(1)	15/15
RL-SHADE-1	**1.4**(1)	**1.3**(1)	4.3(3)	**2.0**(0.6)	**1.3**(1)	6/15
RL-SHADE-1	**2.0**(2)	**1.8**(2)	12(4)	11(6)	5.6(2)	15/15
R-SHADE-10	**2.5**(3)	**2.2**(3)	4.0(3)	**2.5**(1)	**1.4**(2)	6/15
R-SHADE-10	**2.3**(4)	**1.9**(2)	3.9(2)	3.2(3)	**1.9**(0.8)	15/15
SOO-Derbel	**0.88**(0.1)	**1.5**(0.1)	11(0.0)	3.7(7e-3)	**1.2**(6e-4)	15/15

Table 22: 05-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	*4.0e+1*:3.9	*2.5e+1*:11	*1.6e+1*:31	*6.3e+0*:73	*1.6e+0*:347	5/5
MATSUMOTO-	**1.3**(0.6)	**1.3**(2)	**0.73**(0.8)	**1.2**(0.7)	**1.5**(1)	7/15
R-DE-10e2-	**1.8**(1)	**1.2**(0.8)	**0.82**(0.8)	**1.6**(1)	4.7(4)	4/15
R-DE-10e5-	**2.2**(3)	**1.6**(0.7)	**1.1**(0.7)	3.6(6)	33(23)	15/15
RL-SHADE-1	**1**(0.7)	**1.4**(1)	**1.9**(2)	**2.1**(1.0)	**2.3**(3)	7/15
RL-SHADE-1	**1.5**(2)	**1.5**(2)	**1.6**(1)	5.0(3)	4.5(2)	15/15
R-SHADE-10	**1.9**(2)	**1.5**(3)	**1.3**(1)	**2.0**(0.7)	**2.1**(4)	8/15
R-SHADE-10	**2.1**(0.8)	**2.2**(2)	**1.6**(1)	4.5(9)	5.3(10)	15/15
SOO-Derbel	**1.3**(1.0)	**1.1**(2)	**0.88**(0.7)	**0.99**(0.6)	**0.90**(0.8)	15/15

Table 23: 05-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	*6.3e+1*:3.6	*4.0e+1*:15	*2.5e+1*:32	*1.0e+1*:71	*1.6e+0*:341	5/5
MATSUMOTO-	**1.6**(1)	**1.1**(0.9)	**1.1**(0.6)	**1.1**(0.7)	**2.3**(3)	5/15
R-DE-10e2-	**2.8**(3)	**1.5**(2)	**1.2**(0.6)	**1.3**(0.9)	**2.3**(3)	7/15
R-DE-10e5-	**1.4**(0.7)	**1.3**(1)	**1.5**(2)	**2.1**(1)	17(29)	15/15
RL-SHADE-1	**1.4**(2)	**1.9**(2)	**1.3**(0.7)	**1.6**(0.7)	**2.9**(4)	6/15
RL-SHADE-1	**2.5**(3)	**2.0**(1.0)	**1.9**(3)	3.7(5)	4.2(2)	15/15
R-SHADE-10	**2.7**(5)	**1.5**(1)	**1.1**(0.4)	**1.6**(2)	6.4(11)	3/15
R-SHADE-10	3.0(2)	**1.8**(1)	**1.2**(1)	**1.5**(0.8)	3.3(7)	15/15
SOO-Derbel	**1.5**(2)	**0.94**(0.6)	**0.77**(0.7)	**1.0**(0.5)	**1.00**(1.0)	15/15

Table 24: 05-D, running time excess ERT/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 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	*1.0e+1*:3.0	*6.3e+0*:9.0	*4.0e+0*:33	*2.5e+0*:84	*1.0e+0*:518	15/15
MATSUMOTO-	**1.8**(2)	**1.7**(2)	**2.4**(3)	**3.9**(6)	∞ *250*	0/15
R-DE-10e2-	**1.6**(2)	**1.1**(1)	**2.1**(2)	4.5(3)	14(10)	1/15
R-DE-10e5-	**2.0**(1)	**2.4**(3)	3.8(4)	6.0(7)	34(34)	15/15
RL-SHADE-1	**2.8**(3)	**2.8**(1)	3.1(3)	7.3(9)	∞ *500*	0/15
RL-SHADE-1	**2.2**(2)	**2.3**(0.9)	**2.5**(1)	5.1(6)	18(9)	15/15
R-SHADE-10	3.1(4)	**2.6**(1)	3.5(4)	5.2(6)	∞ *500*	0/15
R-SHADE-10	**2.9**(3)	**2.1**(2)	3.2(3)	6.3(7)	**6.2**(3)	15/15
SOO-Derbel	**1.7**(2)	**2.9**(5)	**2.5**(3)	**3.8**(2)	**1.4**(0.2)	15/15

Table 25: 05-D, running time excess ERT/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 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	*6.3e+1*:15	*4.0e+1*:37	*4.0e+1*:37	*2.5e+1*:118	*1.6e+1*:692	15/15
MATSUMOTO-	**1.9**(3)	3.1(2)	3.1(4)	10(8)	5.3(4)	1/15
R-DE-10e2-	**1.7**(0.9)	**2.7**(1)	**2.7**(1)	3.9(4)	11(18)	1/15
R-DE-10e5-	**1.9**(2)	**2.0**(2)	**2.0**(2)	**2.0**(3)	**2.0**(1)	15/15
RL-SHADE-1	**2.4**(3)	3.7(1)	3.7(2)	5.3(4)	11(14)	1/15
RL-SHADE-1	**2.8**(3)	8.0(6)	8.0(7)	10(4)	5.0(3)	15/15
R-SHADE-10	**2.8**(1)	3.3(2)	3.3(2)	5.7(4)	11(16)	1/15
R-SHADE-10	**2.4**(3)	3.1(2)	3.1(2)	**3.0**(2)	**1.8**(1)	15/15
SOO-Derbel	**1.1**(2)	5.1(4)	5.1(3)	5.2(2)	**1.8**(1)	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.