Comparison Tables: CEC BBOB 2015 Testbed in 3-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: 03-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*:3.0	*1.0e+1*:3.6	*1.0e-8*:8.0	*1.0e-8*:8.0	*1.0e-8*:8.0	15/15
MATSUMOTO-	**1.6**(2)	**1.9**(1)	∞	∞	∞ *150*	0/15
R-DE-10e2-	**1.6**(0.9)	**2.6**(3)	∞	∞	∞ *300*	0/15
R-DE-10e5-	**2.4**(3)	**2.5**(3)	**72**(33)	**72**(68)	**72**(43)	15/15
RL-SHADE-1	**1.4**(1)	**1.5**(1)	279(150)	279(234)	279(385)	2/15
RL-SHADE-1	**1.4**(0.7)	**2.1**(2)	387(38)	387(29)	387(26)	15/15
R-SHADE-10	**1.8**(2)	**2.9**(3)	∞	∞	∞ *300*	0/15
R-SHADE-10	**2.3**(2)	3.0(2)	**74**(5)	**74**(11)	**74**(8)	15/15
SOO-Derbel	**0.69**(0.5)	**1.1**(1.0)	99(4)	99(7)	99(10)	15/15

Table 3: 03-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	*6.3e+6*:1.5	*6.3e+5*:4.3	*4.0e+4*:10	*1.0e+2*:32	*1.0e-8*:49	15/15
MATSUMOTO-	**1.3**(1.0)	**0.89**(0.8)	**1.8**(1)	22(30)	∞ *150*	0/15
R-DE-10e2-	**1.9**(0.8)	**1.8**(2)	**2.1**(3)	**3.3**(1)	∞ *300*	0/15
R-DE-10e5-	**1.8**(6)	**1.6**(3)	**2.1**(2)	**3.6**(3)	**26**(8)	15/15
RL-SHADE-1	**2.1**(2)	**1.7**(0.7)	4.0(2)	3.9(1)	∞ *300*	0/15
RL-SHADE-1	**1.1**(1.0)	**1.3**(0.8)	**2.8**(3)	18(8)	91(6)	15/15
R-SHADE-10	**1.4**(1.0)	**1.5**(0.8)	4.0(3)	4.8(2)	∞ *300*	0/15
R-SHADE-10	**1.3**(0)	**1.2**(1)	**2.7**(1)	4.7(1)	**18**(3)	15/15
SOO-Derbel	**1.5**(1)	**2.5**(2)	3.7(3)	6.0(2)	39(5)	15/15

Table 4: 03-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*:2.2	*6.3e+1*:6.1	*4.0e+1*:10	*1.6e+1*:32	*4.0e+0*:319	15/15
MATSUMOTO-	**1.9**(1)	**1.1**(1.0)	**1.4**(0.8)	**1.6**(0.8)	**0.95**(1)	6/15
R-DE-10e2-	**2.2**(1)	**2.2**(2)	**2.7**(1.0)	**2.0**(0.8)	**1.1**(1)	10/15
R-DE-10e5-	**2.0**(2)	**1.6**(0.9)	**2.3**(1.0)	**2.9**(0.6)	**1.0**(0.8)	15/15
RL-SHADE-1	**2.1**(2)	**1.8**(2)	**2.5**(1)	**2.2**(1)	**0.44**(0.2)	15/15
RL-SHADE-1	**2.2**(3)	**1.6**(1)	**2.6**(3)	6.5(3)	**2.0**(0.9)	15/15
R-SHADE-10	**2.2**(0.9)	**2.3**(3)	**2.1**(0.8)	**2.3**(0.9)	**0.57**(0.1)	14/15
R-SHADE-10	**2.3**(1)	**1.6**(2)	3.1(1)	**2.1**(2)	**1.7**(2)	15/15
SOO-Derbel	**0.91**(0)	**0.80**(0.5)	**2.2**(1)	**2.2**(1)	**1.4**(1)	15/15

Table 5: 03-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	*1.0e+2*:5.4	*6.3e+1*:10	*6.3e+1*:10	*2.5e+1*:36	*4.0e+0*:617	15/15
MATSUMOTO-	**1.6**(2)	**1.8**(1)	**1.8**(1)	**1.4**(1.0)	**1.2**(2)	3/15
R-DE-10e2-	**2.4**(4)	**2.0**(2)	**2.0**(2)	**1.6**(1)	**0.60**(0.4)	10/15
R-DE-10e5-	**1.2**(1)	**2.0**(1)	**2.0**(1)	**1.5**(0.5)	**1.3**(1)	15/15
RL-SHADE-1	**1.0**(1)	**2.2**(1)	**2.2**(2)	**1.5**(1)	**0.58**(0.5)	9/15
RL-SHADE-1	**2.4**(3)	**2.5**(2)	**2.5**(5)	5.4(1)	**1.7**(0.8)	15/15
R-SHADE-10	**1.5**(2)	**1.8**(3)	**1.8**(1)	**1.8**(0.6)	**0.49**(0.4)	11/15
R-SHADE-10	**1.1**(0.9)	**1.2**(1)	**1.2**(1)	**2.0**(3)	**1.3**(1)	15/15
SOO-Derbel	**0.69**(2)	**0.99**(1.0)	**0.99**(1)	**2.1**(1)	**0.94**(0.5)	15/15

Table 6: 03-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*:2.2	*2.5e+1*:4.8	*1.0e-8*:6.6	*1.0e-8*:6.6	*1.0e-8*:6.6	15/15
MATSUMOTO-	**1.5**(1)	**1.0**(0.8)	**1.9**(0.6)$^{\star 4}$	**1.9**(0.5)$^{\star 4}$	**1.9**(0.5)$^{\star 4}$	15/15
R-DE-10e2-	**2.2**(1)	**3.0**(3)	∞	∞	∞ *300*	0/15
R-DE-10e5-	**2.8**(2)	**3.0**(2)	256(292)	256(170)	256(299)	15/15
RL-SHADE-1	**1.8**(3)	**1.6**(3)	∞	∞	∞ *300*	0/15
RL-SHADE-1	3.8(2)	3.9(4)	431(21)	431(8)	431(15)	15/15
R-SHADE-10	4.7(4)	4.4(3)	∞	∞	∞ *300*	0/15
R-SHADE-10	**1.5**(0.9)	3.6(4)	**150**(17)	**150**(26)	**150**(26)	15/15
SOO-Derbel	**2.0**(0.2)	**1.8**(0.1)	531(0.1)	531(0.1)	531(0.1)	15/15

Table 7: 03-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	*6.3e+4*:1.8	*6.3e+3*:3.7	*4.0e+1*:13	*1.0e+1*:34	*6.3e-4*:159	15/15
MATSUMOTO-	**1.2**(2)	**1.1**(1)	3.2(3)	6.3(5)	∞ *150*	0/15
R-DE-10e2-	**2.9**(3)	**2.1**(3)	3.2(1)	**1.9**(3)	∞ *300*	0/15
R-DE-10e5-	**2.8**(6)	3.4(3)	**2.0**(1)	**2.2**(0.5)	26(50)	15/15
RL-SHADE-1	**1.9**(3)	**2.6**(4)	**2.9**(1)	**2.7**(0.4)	28(31)	1/15
RL-SHADE-1	**1.6**(0.6)	4.8(4)	6.3(7)	4.9(5)	**16**(3)	15/15
R-SHADE-10	**2.0**(1)	3.9(8)	**3.0**(2)	**2.6**(1)	∞ *300*	0/15
R-SHADE-10	**1.3**(1)	3.7(8)	3.8(8)	**2.3**(2)	**3.4**(0.3)	15/15
SOO-Derbel	**1.4**(2)	**1.7**(1.0)	**2.0**(2)	**1.9**(2)	1.3e4(8958)	2/15

Table 8: 03-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	*2.5e+2*:1.5	*6.3e+1*:4.2	*1.0e+1*:11	*2.5e+0*:38	*4.0e-1*:174	15/15
MATSUMOTO-	**1.3**(0.8)	**1.4**(1)	**2.3**(2)	**1.6**(0.7)	**1.7**(1.0)	7/15
R-DE-10e2-	**1.0**(1.0)	**1.4**(2)	**2.2**(1)	**1.4**(1)	**1.0**(1)	13/15
R-DE-10e5-	**1.2**(1)	**0.94**(1.0)	**1.7**(2)	**1.4**(2)	**2.4**(1)	15/15
RL-SHADE-1	**1.7**(2)	**2.1**(3)	3.5(4)	**2.3**(1)	**1.2**(0.6)	13/15
RL-SHADE-1	**2.0**(0.7)	3.8(2)	6.6(7)	5.0(3)	**2.5**(0.4)	15/15
R-SHADE-10	**1.3**(1.0)	**1.3**(2)	3.4(3)	**2.2**(0.7)	**2.5**(3)	8/15
R-SHADE-10	**2.2**(2)	3.6(2)	3.7(6)	**2.2**(2)	**1.4**(0.7)	15/15
SOO-Derbel	**1.1**(0.7)	**1.1**(2)	**2.3**(0.7)	**1.5**(0.7)	**2.6**(8)	15/15

Table 9: 03-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*:1.8	*1.6e+3*:4.0	*1.0e+2*:15	*6.3e+0*:31	*1.0e-1*:152	15/15
MATSUMOTO-	**1.6**(2)	**1.1**(0.5)	**1.5**(0.8)	4.1(5)	15(15)	1/15
R-DE-10e2-	**1.7**(1)	3.5(5)	3.2(2)	3.7(1)	5.6(6)	5/15
R-DE-10e5-	**1.3**(1.0)	**1.6**(2)	**2.2**(2)	3.5(1)	14(19)	15/15
RL-SHADE-1	**1.4**(0.8)	3.8(2)	3.2(1)	**3.3**(1)	14(10)	2/15
RL-SHADE-1	**1.1**(0.6)	**2.6**(1)	5.7(3)	11(4)	14(4)	15/15
R-SHADE-10	**1.5**(1)	**2.1**(2)	3.1(2)	5.3(3)	∞ *300*	0/15
R-SHADE-10	**2.0**(2)	**2.0**(2)	**2.2**(2)	3.4(1)	**4.6**(2)	15/15
SOO-Derbel	**1.4**(2)	**1.1**(0.9)	**1.1**(0.8)	**1.8**(0.7)	**4.6**(2)	15/15

Table 10: 03-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	*1.0e+1*:21	*6.3e+0*:25	*4.0e+0*:32	*2.5e+0*:48	*6.3e-3*:152	15/15
MATSUMOTO-	**2.5**(2)	**2.5**(1)	**2.7**(3)	**2.4**(3)	∞ *150*	0/15
R-DE-10e2-	4.5(2)	4.4(2)	4.2(2)	4.4(1)	∞ *300*	0/15
R-DE-10e5-	8.1(17)	8.4(24)	15(30)	26(34)	36(23)	15/15
RL-SHADE-1	5.1(4)	5.2(10)	5.0(2)	5.4(5)	∞ *300*	0/15
RL-SHADE-1	14(12)	15(13)	15(8)	13(6)	19(4)	15/15
R-SHADE-10	6.2(3)	8.3(7)	12(12)	9.3(9)	∞ *300*	0/15
R-SHADE-10	3.7(1)	3.6(2)	3.6(2)	3.1(2)	**4.4**(4)	15/15
SOO-Derbel	**2.2**(1)	**2.1**(0.5)	**2.0**(1.0)	**1.7**(0.8)	**12**(5)	15/15

Table 11: 03-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	*6.3e+6*:1.7	*1.6e+5*:4.4	*4.0e+4*:12	*4.0e+2*:37	*1.0e+0*:152	15/15
MATSUMOTO-	**1.2**(1)	**2.2**(1)	**1.1**(0.5)	10(18)	∞ *150*	0/15
R-DE-10e2-	**0.88**(0.4)	**2.1**(3)	**1.6**(1)	5.5(5)	29(33)	1/15
R-DE-10e5-	**1.3**(4)	**2.3**(2)	**1.6**(1)	14(7)	52(40)	15/15
RL-SHADE-1	**1.0**(0.8)	3.0(3)	**2.0**(2)	8.9(9)	∞ *300*	0/15
RL-SHADE-1	**1.3**(0.6)	3.7(3)	4.5(4)	11(11)	**14**(4)	15/15
R-SHADE-10	**2.0**(0.9)	**2.8**(2)	**2.3**(1)	20(17)	∞ *300*	0/15
R-SHADE-10	**1.2**(0.6)	**2.6**(3)	**1.8**(2)	**3.0**(1)	**2.2**(0.4)	15/15
SOO-Derbel	**1.0**(1)	**1.8**(2)	**1.2**(0.9)	**3.7**(2)	92(178)	15/15

Table 12: 03-D, running time excess ERT/$\mathrm{ERT}_{\mathrm{best\ 2009}}$ 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 $\mathrm{ERT}_{\mathrm{best\ 2009}}$ (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	*2.5e+6*:1.9	*4.0e+5*:4.5	*6.3e+4*:9.4	*2.5e+1*:36	*2.5e-1*:174	15/15
MATSUMOTO-	**1.6**(2)	**1.3**(1)	**1.4**(0.8)	7.8(12)	∞ *150*	0/15
R-DE-10e2-	**1.2**(1)	**1.3**(1)	**1.4**(1)	13(16)	∞ *300*	0/15
R-DE-10e5-	**2.0**(2)	**1.4**(1)	**1.7**(2)	23(39)	100(176)	15/15
RL-SHADE-1	**2.8**(3)	**2.3**(2)	**2.6**(2)	17(9)	∞ *300*	0/15
RL-SHADE-1	**1.6**(0.8)	**1.3**(1)	**1.3**(0.5)	15(16)	**12**(2)	15/15
R-SHADE-10	**1.2**(1)	**0.99**(0.5)	**1.1**(2)	9.2(13)	∞ *300*	0/15
R-SHADE-10	**1.4**(0.8)	**2.3**(2)	**1.6**(1)	**4.6**(2)	**3.3**(4)	15/15
SOO-Derbel	**1.8**(2)	**1.6**(2)	**1.8**(0.8)	**4.6**(3)	183(927)	14/15

Table 13: 03-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	*1.0e+8*:1.5	*1.0e+7*:3.6	*6.3e+5*:13	*6.3e+2*:31	*1.0e+0*:168	15/15
MATSUMOTO-	**0.83**(0)	**1.1**(2)	**1.4**(0.5)	**3.4**(1)	14(19)	1/15
R-DE-10e2-	**0.91**(0.7)	**1.6**(1)	**2.4**(2)	6.7(0.9)	∞ *300*	0/15
R-DE-10e5-	**1.1**(1.0)	**2.5**(4)	**2.5**(1)	10(8)	76(126)	15/15
RL-SHADE-1	**1.2**(1.0)	**1.9**(2)	**2.4**(1)	5.4(2)	∞ *300*	0/15
RL-SHADE-1	**1.0**(0.3)	**2.3**(3)	4.2(2)	25(16)	20(7)	15/15
R-SHADE-10	**1.3**(0.7)	**2.0**(2)	**2.3**(2)	11(12)	∞ *300*	0/15
R-SHADE-10	**0.87**(0.3)	**1.3**(1)	**2.7**(2)	5.5(2)	**8.2**(20)	15/15
SOO-Derbel	**0.87**(0)	**0.89**(0.6)	**1.3**(1)	**4.8**(0.8)	**4.3**(0.8)	15/15

Table 14: 03-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*:1.6	*4.0e+2*:6.8	*2.5e+2*:11	*4.0e+1*:30	*2.5e-3*:182	15/15
MATSUMOTO-	**1.5**(1)	**1.0**(0.9)	**0.86**(0.5)	**1.6**(0.4)	∞ *150*	0/15
R-DE-10e2-	**1.0**(0.9)	**1.2**(1)	**1.1**(0.6)	**2.6**(0.8)	∞ *300*	0/15
R-DE-10e5-	**1.6**(2)	**1.9**(2)	**1.7**(2)	3.9(4)	44(24)	15/15
RL-SHADE-1	**1.2**(1)	**1.6**(2)	**1.7**(2)	3.8(9)	∞ *300*	0/15
RL-SHADE-1	**1.4**(0.3)	**2.4**(2)	3.3(4)	12(3)	21(1)	15/15
R-SHADE-10	**2.0**(1)	**1.5**(2)	**1.8**(2)	3.9(2)	∞ *300*	0/15
R-SHADE-10	**1.3**(0.8)	**2.1**(3)	**2.4**(1)	3.7(1)	**3.8**(1)	15/15
SOO-Derbel	**0.83**(1)	**0.74**(0.7)	**0.87**(0.6)	**2.4**(2)	**20**(17)	15/15

Table 15: 03-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	*1.0e+1*:2.2	*6.3e+0*:4.2	*2.5e+0*:10	*6.3e-2*:31	*2.5e-6*:160	15/15
MATSUMOTO-	**1.5**(1)	**1.5**(2)	**1.6**(1)	**2.8**(4)	∞ *150*	0/15
R-DE-10e2-	**1.8**(1)	**1.7**(1)	**2.2**(1)	3.7(1)	∞ *300*	0/15
R-DE-10e5-	**2.8**(7)	**2.0**(2)	**2.2**(2)	4.9(5)	107(113)	15/15
RL-SHADE-1	**1.7**(2)	**2.0**(2)	**2.5**(2)	3.7(2)	∞ *300*	0/15
RL-SHADE-1	**1.3**(1)	**0.90**(0.8)	**2.7**(5)	19(5)	**22**(2)	15/15
R-SHADE-10	3.5(6)	**2.5**(4)	**2.4**(2)	4.9(2)	∞ *300*	0/15
R-SHADE-10	**2.7**(3)	**2.0**(2)	**2.5**(4)	4.5(1)	**3.8**(0.4)	15/15
SOO-Derbel	**1.3**(1)	**0.92**(0.8)	**0.93**(0.5)	**3.4**(1)	225(75)	14/15

Table 16: 03-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*:1.6	*6.3e+1*:5.6	*4.0e+1*:12	*1.6e+1*:68	*6.3e+0*:221	15/15
MATSUMOTO-	**1.3**(0.9)	**0.96**(0.6)	**1.3**(1)	**0.66**(0.6)	**0.82**(0.6)	9/15
R-DE-10e2-	**2.2**(4)	**1.9**(2)	**2.0**(2)	**0.92**(0.7)	**1.3**(1)	11/15
R-DE-10e5-	**2.1**(4)	**1.9**(2)	**1.5**(1)	**0.97**(0.4)	3.3(3)	15/15
RL-SHADE-1	**1.8**(2)	**2.2**(2)	**2.1**(1.0)	**1.3**(0.4)	**0.91**(0.6)	13/15
RL-SHADE-1	**2.0**(2)	**2.0**(3)	**2.6**(2)	**2.9**(1)	**2.8**(2)	15/15
R-SHADE-10	**0.96**(0.3)	**1.4**(4)	**1.9**(2)	**1.6**(0.9)	**2.1**(2)	8/15
R-SHADE-10	**3.0**(2)	**3.0**(2)	3.1(4)	**1.4**(0.6)	**1.3**(0.6)	15/15
SOO-Derbel	**1.3**(1)	**1.0**(1)	**1**(1)	**1.1**(0.5)	**0.76**(0.3)	15/15

Table 17: 03-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	*6.3e+1*:1.5	*2.5e+1*:8.2	*1.6e+1*:10	*1.0e+1*:41	*2.5e+0*:208	15/15
MATSUMOTO-	**1.5**(1)	**1.8**(3)	**2.7**(4)	**1.3**(1.0)	**0.82**(0.7)	10/15
R-DE-10e2-	**2.1**(2)	**1.6**(1)	3.2(4)	**1.8**(0.8)	**1.6**(2)	10/15
R-DE-10e5-	**2.1**(4)	**1.1**(0.5)	**1.3**(0.8)	**1.7**(2)	**2.7**(5)	15/15
RL-SHADE-1	**2.0**(0.8)	**1.9**(2)	3.1(3)	**1.3**(2)	**1.0**(0.8)	12/15
RL-SHADE-1	**1.5**(0.7)	**1.3**(2)	**2.7**(3)	**1.9**(1)	**2.9**(3)	15/15
R-SHADE-10	**1.8**(2)	**1.4**(0.7)	**2.7**(4)	**0.98**(0.7)	**2.3**(2)	7/15
R-SHADE-10	**1.3**(0.7)	**1.3**(2)	**2.2**(3)	**1.0**(0.9)	**1.6**(3)	15/15
SOO-Derbel	**2.0**(2)	**1.2**(1)	**1.6**(0.8)	**0.91**(0.6)	**0.45**(0.3)	15/15

Table 18: 03-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	*1.6e+1*:1.8	*1.0e+1*:3.6	*6.3e+0*:14	*2.5e+0*:34	*2.5e-1*:189	5/5
MATSUMOTO-	**2.4**(2)	**2.6**(4)	**1.1**(1)	**1.3**(0.4)	**2.8**(4)	4/15
R-DE-10e2-	**2.7**(3)	**2.2**(2)	**1.5**(2)	**1.6**(1)	**1.4**(2)	13/15
R-DE-10e5-	**1.6**(1)	**2.1**(1)	**1.1**(1)	**1.6**(1)	**1.4**(0.7)	15/15
RL-SHADE-1	**2.1**(2)	**2.5**(2)	**1.6**(2)	**1.6**(2)	3.5(5)	6/15
RL-SHADE-1	**2.6**(0.8)	**2.1**(2)	**0.90**(1)	3.6(3)	5.1(2)	15/15
R-SHADE-10	**2.4**(2)	**2.6**(2)	**1.6**(2)	**2.1**(0.8)	4.7(3)	5/15
R-SHADE-10	**1.9**(2)	**1.8**(1)	**0.98**(0.4)	**1.2**(0.8)	**1.1**(0.6)	15/15
SOO-Derbel	**0.67**(0)	**1.2**(1)	**0.72**(0.3)	**0.96**(0.7)	**0.98**(0.1)	15/15

Table 19: 03-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*:1.8	*4.0e+1*:4.8	*2.5e+1*:13	*1.0e+1*:40	*6.3e-1*:184	15/15
MATSUMOTO-	**3.0**(2)	**1.6**(0.9)	**1.2**(1)	**1.1**(0.4)	∞ *150*	0/15
R-DE-10e2-	**1.3**(0.4)	**1.7**(1)	**1.2**(2)	**1.3**(1)	**2.4**(3)	9/15
R-DE-10e5-	**2.1**(3)	**2.1**(0.8)	**1.5**(1)	**1.3**(0.9)	21(0.5)	15/15
RL-SHADE-1	4.0(6)	**2.9**(3)	3.2(3)	**2.2**(1)	24(20)	1/15
RL-SHADE-1	3.1(3)	**2.4**(2)	**2.2**(3)	**2.9**(2)	6.7(0.8)	15/15
R-SHADE-10	3.9(6)	**1.9**(3)	**2.3**(2)	**2.4**(2)	24(12)	1/15
R-SHADE-10	3.1(3)	**1.9**(2)	**1.1**(0.9)	**1.2**(1)	**1.8**(0.4)	15/15
SOO-Derbel	**0.96**(1)	**1.2**(1)	**0.90**(0.5)	**0.97**(0.7)	**1.5**(0.8)	15/15

Table 20: 03-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*:81	*1.0e-1*:109	*6.3e-2*:109	*4.0e-2*:119	*1.6e-2*:1230	15/15
MATSUMOTO-	∞	∞	∞	∞	∞ *150*	0/15
R-DE-10e2-	25(26)	39(75)	∞	∞	∞ *300*	0/15
R-DE-10e5-	**20**(25)	31(19)	54(67)	103(190)	20(14)	15/15
RL-SHADE-1	27(20)	41(49)	∞	∞	∞ *300*	0/15
RL-SHADE-1	25(11)	25(20)	34(29)	64(61)	11(4)	15/15
R-SHADE-10	52(73)	38(55)	38(64)	**35**(31)	∞ *300*	0/15
R-SHADE-10	24(30)	**24**(41)	**30**(19)	50(60)	**10**(28)	15/15
SOO-Derbel	**2.8**(2)	**2.8**(2)	**3.1**(4)	**4.4**(6)	**0.92**(0.7)	15/15

Table 21: 03-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*:3.5	*2.5e+3*:4.3	*4.0e+0*:13	*1.6e+0*:41	*1.0e+0*:385	5/5
MATSUMOTO-	**1.1**(0.6)	**0.86**(0.6)	**1.5**(0.8)	17(30)	**2.9**(3)	2/15
R-DE-10e2-	**1.8**(0.7)	**1.6**(2)	**2.9**(2)	3.4(1)	**0.89**(1.0)	10/15
R-DE-10e5-	**1.3**(0.7)	**1.3**(0.8)	18(2)	8.7(40)	**2.7**(3)	15/15
RL-SHADE-1	**1.6**(1)	**1.5**(0.5)	**2.5**(2)	**3.2**(2)	**0.62**(0.5)	12/15
RL-SHADE-1	**1.2**(0.4)	**1.2**(0.8)	4.9(3)	13(7)	**2.7**(1)	15/15
R-SHADE-10	**1.6**(1)	**1.8**(2)	3.3(2)	3.8(1.0)	**1.3**(1)	8/15
R-SHADE-10	**1.6**(2)	**1.6**(2)	**2.8**(2)	6.0(5)	**1.8**(2)	15/15
SOO-Derbel	**0.75**(0.1)	**0.61**(0.1)	3.8(0.0)	**1.8**(6e-3)	**0.19**(1e-3)$^{\star 2}$	15/15

Table 22: 03-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	*1.6e+1*:2.5	*1.0e+1*:5.9	*6.3e+0*:14	*2.5e+0*:41	*1.6e+0*:167	15/15
MATSUMOTO-	**2.1**(2)	**1.6**(1)	**1.3**(0.8)	**0.70**(0.8)	**0.63**(0.4)	11/15
R-DE-10e2-	**1.4**(2)	**1.3**(0.9)	**1.3**(2)	**1.1**(1)	**1.1**(1)	12/15
R-DE-10e5-	**2.7**(4)	**1.8**(3)	**1.9**(2)	4.2(9)	4.1(8)	15/15
RL-SHADE-1	**2.3**(2)	**1.2**(2)	**1.1**(1)	**0.96**(1)	**1.3**(1)	12/15
RL-SHADE-1	**2.3**(2)	**1.3**(1)	**1.7**(3)	**1.8**(1)	**2.0**(2)	15/15
R-SHADE-10	**2.4**(1)	**2.6**(1)	**2.7**(2)	**2.3**(2)	**2.3**(3)	8/15
R-SHADE-10	**2.5**(3)	**1.4**(1)	**1.5**(1)	**1.5**(2)	**1.4**(3)	15/15
SOO-Derbel	**1.3**(2)	**1.3**(1)	**1.2**(0.4)	**0.90**(0.5)	**0.43**(0.7)	15/15

Table 23: 03-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 Eru $_{\text{best 2009}}$ (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. uhe median number of conduc ted 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*:2.9	*2.5e+1*:5.2	*1.0e+1*:18	*6.3e+0*:33	*1.0e+0*:170	5/5
MATSUMOTO-	**2.1**(3)	**2.2**(2)	**1.4**(0.8)	**1.4**(2)	**1.5**(2)	7/15
R-DE-10e2-	**1.8**(0.9)	**1.7**(2)	**1.7**(3)	**1.9**(1)	**1.5**(0.9)	11/15
R-DE-10e5-	**2.1**(6)	**1.9**(1)	8.2(28)	5.6(30)	5.7(7)	15/15
RL-SHADE-1	**1.2**(0.5)	**1.4**(1)	**2.9**(4)	**2.3**(0.7)	**1.5**(1)	11/15
RL-SHADE-1	**1.3**(2)	**0.78**(0.4)	**2.0**(3)	**2.0**(2)	**2.4**(2)	15/15
R-SHADE-10	**1.8**(2)	**2.4**(3)	**2.2**(3)	**2.9**(4)	**1.6**(1)	10/15
R-SHADE-10	**1.3**(0.9)	**1.5**(0.7)	**0.87**(0.5)	**1.2**(1)	**1.9**(3)	15/15
SOO-Derbel	**1.0**(1)	**0.74**(0.6)	**0.71**(0.7)	**0.78**(0.8)	**0.47**(0.2)	15/15

Table 24: 03-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*:2.6	*6.3e+0*:16	*4.0e+0*:44	*2.5e+0*:79	*1.6e+0*:198	15/15
MATSUMOTO-	4.3(2)	**1.5**(2)	**2.1**(2)	5.0(6)	∞ *150*	0/15
R-DE-10e2-	**3.1**(2)	**2.1**(3)	**1.8**(2)	3.4(3)	7.0(4)	3/15
R-DE-10e5-	4.6(3)	**2.5**(2)	**1.9**(2)	**2.8**(1)	6.8(6)	15/15
RL-SHADE-1	3.4(4)	**0.90**(0.6)	**1.3**(4)	3.6(3)	7.2(8)	3/15
RL-SHADE-1	3.3(2)	**1.5**(2)	**2.0**(2)	**2.9**(2)	7.9(6)	15/15
R-SHADE-10	4.1(4)	3.6(5)	**2.8**(2)	5.3(6)	6.9(9)	3/15
R-SHADE-10	**2.8**(3)	**1.7**(0.8)	**1.5**(2)	3.3(2)	**3.0**(3)	15/15
SOO-Derbel	4.4(6)	**1.8**(3)	**2.2**(2)	**2.8**(2)	**1.8**(1.0)	15/15

Table 25: 03-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	*4.0e+1*:4.6	*2.5e+1*:13	*1.6e+1*:47	*1.6e+1*:47	*6.3e+0*:382	15/15
MATSUMOTO-	**1.6**(1)	**1.9**(2)	**1.7**(1)	**1.7**(4)	5.8(5)	1/15
R-DE-10e2-	**1.9**(0.7)	**2.2**(2)	**1.8**(1)	**1.8**(0.7)	**2.4**(2)	4/15
R-DE-10e5-	**1.1**(1)	**1.6**(0.9)	**2.7**(4)	**2.7**(4)	4.4(7)	15/15
RL-SHADE-1	**1.6**(2)	**2.2**(1)	**1.3**(0.4)	**1.3**(0.5)	**2.6**(3)	4/15
RL-SHADE-1	**1.2**(0.5)	**1.8**(2)	**2.0**(1)	**2.0**(1.0)	3.1(3)	15/15
R-SHADE-10	**1.0**(0.6)	**1.5**(2)	**1.5**(2)	**1.5**(2)	5.4(4)	2/15
R-SHADE-10	**1.7**(1)	**1.9**(2)	**1.5**(0.9)	**1.5**(0.9)	**1.1**(0.9)	15/15
SOO-Derbel	**1.4**(2)	**1.4**(1)	**2.1**(0.7)	**2.1**(3)	**1.4**(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.