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

The BBOBies

July 16, 2015

Abstract

This document provides tabular results of the workshop on Black-Box Optimization Benchmarking held at GECCO 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=bbob-2015`. Overall, 18 algorithms have been tested on 24 benchmark functions in dimensions between 2 and 20. Only three of them have been tested on the optional instances in dimension 40. A description of the used objective functions can be found in [7, 5]. The experimental set-up is described in [6].

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 [2]) 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 [6] 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 GECCO 2015.

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

algorithm short name	paper	reference
BSifeg	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
BSif	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
BSqi	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
BSrr	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
CMA-CSA	Benchmarking IPOP-CMA-ES-TPA and IPOP-CMA-ES-MSR on the BBOB Noiseless Testbed	[1]
CMA-MSR	Benchmarking IPOP-CMA-ES-TPA and IPOP-CMA-ES-MSR on the BBOB Noiseless Testbed	[1]
CMA-TPA	Benchmarking IPOP-CMA-ES-TPA and IPOP-CMA-ES-MSR on the BBOB Noiseless Testbed	[1]
GP1-CMAES	SBenchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
GP5-CMAES	Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
IPOPCMAv3p61	Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
LHD-10xDefault-MATSuMoT	The Impact of Initial Designs on the Performance of MATSuMoTo on the Noiseless BBOB-2015 Testbed: A Preliminary Study	[4]
LHD-2xDefault-MATSuMoTo	The Impact of Initial Designs on the Performance of MATSuMoTo on the Noiseless BBOB-2015 Testbed: A Preliminary Study	[4]
RAND-2xDefault-MATSuMoTo	The Impact of Initial Designs on the Performance of MATSuMoTo on the Noiseless BBOB-2015 Testbed: A Preliminary Study	[4]
RF1-CMAES	Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
RF5-CMAES	Benchmarking Gaussian Processes and Random Forests Surrogate Models on the BBOB Noiseless Testbed	[3]
Sifeg	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
Sif	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]
Srr	Dimension Selection in Axis-Parallel Brent-STEP Method for Black-Box Optimization of Separable Continuous Functions	[9]

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

#FEs/D	0.5	1.2	3	10	50	#succ
f1	*1.6e+1*:1.2	*4.0e+0*:2.6	*2.5e-2*:6.2	*1.0e-8*:6.2	*1.0e-8*:6.2	15/15
BSifeg	**1.6**(1)	**2.2**(2)	**1.7**(0.2)	**1.7**(0.2)	**1.7**(0.3)	15/15
BSif	**1.6**(1)	**2.2**(2)	**1.7**(0.2)	**1.7**(0.3)	**1.7**(0.2)	15/15
BSqi	**1.6**(2)	**2.2**(2)	**1.7**(0.3)	**1.7**(0.2)	**1.7**(0.3)	15/15
BSrr	**1.6**(2)	**2.2**(2)	**1.7**(0.3)	**1.7**(0.2)	**1.7**(0.3)	15/15
CMA-CSA	3.1(5)	3.0(3)	11(6)	42(5)	42(6)	15/15
CMA-MSR	**2.7**(1)	4.7(3)	16(11)	69(12)	69(9)	15/15
CMA-TPA	**2.8**(4)	4.5(4)	12(7)	44(8)	44(13)	15/15
GP1-CMAES	**2.4**(5)	**2.5**(2)	6.5(2)	24(4)	24(5)	15/15
GP5-CMAES	3.3(3)	**2.7**(2)	3.4(2)	31(14)	31(16)	15/15
IPOPCMAv3p	4.9(4)	4.6(5)	12(7)	43(9)	43(4)	15/15
LHD-10xDef	**2.6**(2)	3.8(5)	10(0)	∞	∞ *100*	0/15
LHD-2xDefa	**2.4**(1)	**2.8**(2)	3.3(0.8)	∞	∞ *100*	0/15
RAND-2xDef	**2.6**(2)	3.4(2)	3.5(0.9)	∞	∞ *100*	0/15
RF1-CMAES	3.1(4)	4.2(4)	8.0(3)	86(85)	86(111)	10/15
RF5-CMAES	**2.8**(3)	4.7(4)	60(38)	∞	∞ *502*	0/15
Sifeg	**1.6**(1)	**2.2**(2)	**2.3**(0.4)	6.3(0.5)	6.3(0.5)	15/15
Sif	**1.6**(1)	**2.2**(1)	**2.4**(0.3)	6.2(0.6)	6.2(0.6)	15/15
Srr	**1.6**(1)	**2.2**(2)	**2.3**(0.2)	5.9(0.5)	5.9(0.5)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f2	*1.0e+7*:1.4	*1.6e+6*:2.7	*1.0e+5*:6.1	*6.3e-1*:20	*1.0e-8*:30	15/15
BSifeg	**1.7**(1)	3.1(1)	**1.7**(0.2)	**1.2**(0.3)	**1.2**(0.3)	15/15
BSif	**1.7**(1)	3.1(0.4)	**1.7**(0.2)	**1.1**(0.3)	**1.2**(0.1)	15/15
BSqi	**1.7**(1)	3.1(0.9)	**1.7**(0.2)	**0.93**(0.1)	**1.1**(0.2)	15/15
BSrr	**1.7**(1)	3.1(0)	**1.7**(0.3)	**1.1**(0.3)	**1.3**(0.2)	15/15
CMA-CSA	**1.3**(2)	**2.8**(3)	3.2(2)	15(4)	19(2)	15/15
CMA-MSR	**1.8**(0.5)	**2.6**(2)	**2.8**(2)	16(3)	24(5)	15/15
CMA-TPA	**1.5**(1)	**1.2**(0.5)	**1.7**(1)	12(5)	17(2)	15/15
GP1-CMAES	**2.1**(2)	**2.3**(2)	**2.2**(2)	13(6)	19(2)	12/15
GP5-CMAES	**2.2**(2)	**2.2**(2)	**1.5**(1)	5.1(2)	9.2(6)	14/15
IPOPCMAv3p	**2.3**(2)	**3.0**(2)	3.0(2)	17(6)	∞ *506*	0/15
LHD-10xDef	**1.3**(1)	**1.4**(0.8)	**2.6**(1)	∞	∞ *100*	0/15
LHD-2xDefa	**1.0**(0.7)	**1.2**(1)	**1.7**(1)	∞	∞ *100*	0/15
RAND-2xDef	**1.3**(2)	**1.0**(0.7)	**1.3**(1)	∞	∞ *100*	0/15
RF1-CMAES	**2.5**(3)	**2.2**(2)	**2.8**(3)	168(242)	∞ *506*	0/15
RF5-CMAES	**1.8**(1)	**2.8**(2)	**2.8**(2)	167(194)	∞ *502*	0/15
Sifeg	**1.7**(1)	3.2(2)	**2.0**(0.3)	**1.5**(0.3)	**1.8**(0.2)	15/15
Sif	**1.7**(1)	3.2(1)	**2.0**(0.5)	**1.4**(0.3)	**1.7**(0.3)	15/15
Srr	**1.7**(1)	3.2(0.4)	**2.0**(0.5)	**1.4**(0.2)	**1.9**(0.1)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f3	*1.0e+2*:1.4	*4.0e+1*:4.1	*2.5e+1*:6.6	*6.3e+0*:26	*2.5e+0*:112	15/15
BSifeg	**2.2**(1)	**1.8**(1)	**1.4**(0.3)	**0.79**(0.6)	**0.30**(0.2)	15/15
BSif	**2.2**(1)	**1.8**(0.7)	**1.4**(0.7)	**0.77**(0.3)	**0.30**(0.2)	15/15
BSqi	**2.2**(1)	**1.8**(1)	**1.4**(0.5)	**0.76**(0.5)	**0.30**(0.2)	15/15
BSrr	**2.2**(0.7)	**1.8**(1)	**1.4**(0.6)	**0.74**(0.6)	**0.28**(0.2)	15/15
CMA-CSA	**1.8**(1)	**2.4**(2)	**2.4**(2)	**2.2**(1)	**2.9**(4)	15/15
CMA-MSR	3.0(3)	**2.9**(1)	**2.6**(1)	3.3(2)	4.6(9)	15/15
CMA-TPA	**2.0**(3)	**2.9**(3)	**2.9**(2)	6.8(8)	5.5(4)	15/15
GP1-CMAES	**2.8**(3)	**1.7**(2)	**2.6**(2)	**2.2**(1)	**2.9**(4)	11/15
GP5-CMAES	**2.0**(1)	**1.9**(2)	4.0(1)	5.8(7)	**2.3**(2)	13/15
IPOPCMAv3p	**1.6**(0.9)	**1.4**(1.0)	**1.6**(0.8)	**2.6**(1)	3.2(2)	10/15
LHD-10xDef	**1.9**(3)	**1.6**(2)	**1.7**(1)	3.0(2)	**1.5**(0.8)	8/15
LHD-2xDefa	**1.6**(0.4)	**1.2**(1)	**1.4**(1)	**1.0**(0.7)	**0.67**(0.8)	11/15
RAND-2xDef	**1.8**(2)	**1.1**(0.7)	**1.9**(1)	**2.4**(1)	**1.4**(2)	7/15
RF1-CMAES	**1.7**(2)	11(1)	7.2(2)	9.0(6)	5.6(6)	7/15
RF5-CMAES	**2.2**(3)	**2.4**(2)	14(2)	15(15)	6.5(8)	7/15
Sifeg	**2.2**(2)	**1.8**(1)	**1.5**(0.9)	**1.1**(0.5)	**0.30**(0.2)	15/15
Sif	**2.2**(2)	**1.8**(1)	**1.5**(0.7)	**1.1**(0.5)	**0.34**(0.2)	15/15
Srr	**2.2**(1)	**1.8**(1)	**1.5**(1)	**1.1**(0.4)	**0.31**(0.1)	15/15

Table 5: 02-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	*6.3e+1*:2.4	*4.0e+1*:5.2	*2.5e+1*:8.5	*1.0e+1*:22	*2.5e+0*:120	5/5
BSifeg	**1.8**(1)	**1.1**(0.9)	**1.1**(0.3)	**0.91**(0.4)	**0.42**(0.3)	15/15
BSif	**1.8**(2)	**1.1**(0.8)	**1.1**(0.5)	**0.91**(0.6)	**0.42**(0.3)	15/15
BSqi	**1.8**(2)	**1.1**(0.9)	**1.1**(0.5)	**0.98**(0.4)	**0.45**(0.3)	15/15
BSrr	**1.8**(2)	**1.1**(1.0)	**1.1**(0.6)	**0.93**(0.4)	**0.42**(0.3)	15/15
CMA-CSA	**1.9**(2)	**2.3**(2)	**2.4**(1)	**2.3**(2)	4.3(7)	15/15
CMA-MSR	3.2(2)	**2.6**(2)	**2.3**(2)	6.1(20)	11(13)	15/15
CMA-TPA	3.3(4)	**1.9**(0.7)	**2.5**(2)	**2.9**(4)	6.4(9)	15/15
GP1-CMAES	**2.5**(2)	**1.9**(2)	**2.3**(2)	**2.2**(1)	4.6(3)	8/15
GP5-CMAES	3.2(3)	**2.1**(3)	10(15)	6.4(3)	3.8(3)	9/15
IPOPCMAv3p	**2.9**(3)	**2.2**(2)	**2.8**(3)	**2.9**(3)	5.0(10)	8/15
LHD-10xDef	**1.7**(4)	**1.7**(1)	**2.0**(3)	**2.7**(3)	**1.9**(2)	6/15
LHD-2xDefa	**1.5**(1)	**1.9**(1)	**2.0**(1)	**1.8**(1)	**1.7**(1)	6/15
RAND-2xDef	**2.2**(1)	**1.6**(0.8)	**1.2**(0.8)	**1.7**(2)	**2.1**(1.0)	5/15
RF1-CMAES	**2.1**(2)	**2.3**(3)	**2.1**(2)	3.3(3)	28(27)	2/15
RF5-CMAES	17(2)	8.6(1)	12(16)	16(22)	28(26)	2/15
Sifeg	**1.8**(1)	**1.2**(1.0)	**1.3**(0.2)	**0.96**(0.3)	**0.50**(0.3)	15/15
Sif	**1.8**(2)	**1.2**(1)	**1.3**(0.9)	**0.95**(0.1)	**0.52**(0.3)	15/15
Srr	**1.8**(2)	**1.2**(1)	**1.2**(0.3)	**0.96**(0.6)	**0.51**(0.1)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f5	*4.0e+1*:1.4	*1.6e+1*:3.5	*1.0e-8*:4.4	*1.0e-8*:4.4	*1.0e-8*:4.4	15/15
BSifeg	**2.9**(1)	**1.8**(0.1)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
BSif	**2.9**(2)	**1.8**(0.1)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
BSqi	**2.9**(2)	**1.8**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.1)	15/15
BSrr	**2.9**(2)	**1.8**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
CMA-CSA	4.2(4)	**3.0**(3)	4.8(3)	4.8(4)	4.8(3)	15/15
CMA-MSR	4.0(5)	**2.9**(2)	6.2(3)	6.2(2)	6.2(2)	15/15
CMA-TPA	3.3(4)	**2.0**(2)	4.0(1)	4.0(1)	4.0(2)	15/15
GP1-CMAES	3.5(4)	**2.2**(1)	9.3(9)	9.3(14)	9.3(12)	15/15
GP5-CMAES	**2.6**(4)	**1.9**(0.9)	4.4(2)	4.4(3)	4.4(2)	15/15
IPOPCMAv3p	5.0(3)	4.1(4)	13(8)	13(9)	13(16)	15/15
LHD-10xDef	**2.6**(2)	**2.2**(2)	14(0.6)	14(0.3)	14(0.6)	15/15
LHD-2xDefa	**1.4**(0.7)	**2.0**(1)	3.3(0.6)	3.3(0.6)	3.3(0.6)	15/15
RAND-2xDef	**1.4**(0.7)	**1.9**(2)	3.6(0.6)	3.6(0.6)	3.6(0.9)	15/15
RF1-CMAES	**2.4**(3)	**1.9**(2)	22(26)	22(32)	22(26)	15/15
RF5-CMAES	3.7(5)	3.2(3)	61(75)	61(34)	61(76)	12/15
Sifeg	**2.9**(2)	**1.8**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
Sif	**2.9**(2)	**1.8**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
Srr	**2.9**(2)	**1.8**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15

Table 7: 02-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*:1.4	*1.0e+2*:2.8	*1.6e+1*:10	*1.0e+0*:23	*2.5e-6*:103	15/15
BSifeg	**1.3**(1)	**2.0**(3)	245(513)	416(421)	2668(4813)	1/15
BSif	**1.3**(1)	**2.1**(3)	307(617)	442(1192)	∞ *2e4*	0/15
BSqi	**1.3**(1)	**2.0**(1)	355(776)	344(686)	2671(5134)	1/15
BSrr	**1.3**(1)	**2.0**(1)	341(56)	403(547)	1293(1193)	2/15
CMA-CSA	**1.6**(2)	3.6(5)	3.1(2)	**4.3**(4)	**4.2**(0.8)	15/15
CMA-MSR	**1.8**(2)	**2.3**(3)	**2.4**(1)	5.3(3)	5.2(0.8)	15/15
CMA-TPA	**1.3**(0.4)	**1.2**(1)	**1.2**(2)	**3.6**(3)	**3.8**(0.7)	15/15
GP1-CMAES	**2.4**(3)	3.1(5)	**2.8**(2)	4.9(0.7)	∞ *506*	0/15
GP5-CMAES	**1.1**(0.4)	**2.3**(7)	**2.3**(2)	9.2(15)	∞ *506*	0/15
IPOPCMAv3p	**1.5**(2)	**2.7**(2)	**2.6**(5)	**4.6**(2)	**4.8**(4)	13/15
LHD-10xDef	**1.3**(1)	**1.5**(2)	**1.3**(1)	9.0(12)	∞ *100*	0/15
LHD-2xDefa	**1.1**(0)	**1.9**(2)	**1.5**(0.9)	12(16)	∞ *100*	0/15
RAND-2xDef	**1.4**(0.7)	**1.9**(2)	**1.2**(0.7)	6.3(8)	∞ *100*	0/15
RF1-CMAES	**1.8**(0.7)	**2.3**(2)	5.5(14)	67(146)	∞ *506*	0/15
RF5-CMAES	**1.5**(0.9)	3.2(5)	4.8(6)	40(35)	∞ *508*	0/15
Sifeg	**1.3**(1)	**2.3**(4)	333(47)	271(534)	2624(3800)	1/15
Sif	**1.3**(1)	**2.3**(4)	355(1172)	318(750)	2659(2915)	1/15
Srr	**1.3**(1)	**2.1**(3)	319(1229)	242(376)	1251(1158)	2/15

Table 8: 02-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	*4.0e+2*:1.6	*1.0e+1*:3.2	*2.5e+0*:14	*1.6e+0*:21	*1.6e-2*:188	15/15
BSifeg	**1.3**(0.5)	**1.7**(1)	135(216)	376(318)	698(696)	2/15
BSif	**1.3**(0.6)	**1.7**(1)	173(395)	376(1149)	252(216)	5/15
BSqi	**1.3**(1)	**1.7**(1)	206(466)	256(513)	355(565)	4/15
BSrr	**1.3**(1)	**1.7**(1)	268(467)	543(2200)	331(424)	4/15
CMA-CSA	**1.3**(0.6)	4.0(3)	3.6(2)	3.5(10)	**1.1**(1)	15/15
CMA-MSR	**1.7**(2)	4.4(5)	**2.0**(3)	**1.6**(2)	**1.0**(0.3)	15/15
CMA-TPA	**1.1**(0.3)	3.7(3)	**2.5**(2)	**1.9**(2)	**0.80**(0.8)	15/15
GP1-CMAES	**1.4**(1)	4.2(1)	**2.1**(0.9)	**1.9**(2)	**1.0**(0.4)	14/15
GP5-CMAES	**0.67**(0.2)	3.3(3)	**2.0**(3)	**2.0**(3)	**0.98**(0.3)	14/15
IPOPCMAv3p	**1.9**(0.9)	3.9(8)	**2.5**(3)	3.6(4)	**1.7**(1)	13/15
LHD-10xDef	**1.2**(1)	7.6(5)	**3.0**(2)	**2.6**(2)	∞ *100*	0/15
LHD-2xDefa	**1**(0.3)	3.0(2)	**1.2**(0.9)	**1.0**(0.8)	**1.3**(1)	5/15
RAND-2xDef	**1.7**(1)	3.8(1)	**1.2**(0.4)	**1.2**(1)	**1.1**(2)	6/15
RF1-CMAES	**1.5**(1)	4.1(1)	**2.0**(2)	**2.7**(2)	**2.2**(2)	11/15
RF5-CMAES	**1.6**(2)	3.8(4)	**2.5**(2)	3.5(2)	13(15)	3/15
Sifeg	**1.3**(0.3)	**1.8**(2)	68(180)	256(954)	444(405)	3/15
Sif	**1.3**(0.9)	**1.8**(2)	78(36)	204(440)	485(501)	3/15
Srr	**1.3**(0.6)	**1.8**(1)	135(144)	287(367)	473(349)	3/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f8	*2.5e+3*:1.2	*1.0e+2*:3.2	*6.3e+0*:7.0	*1.6e-1*:27	*1.6e-6*:100	15/15
BSifeg	**1.9**(3)	**1.9**(1)	4.0(3)	2208(3277)	∞ *2e4*	0/15
BSif	**1.9**(0.4)	**1.9**(1)	4.2(8)	2303(3150)	∞ *2e4*	0/15
BSqi	**1.9**(2)	**1.9**(1)	4.1(8)	1323(2706)	∞ *2e4*	0/15
BSrr	**1.9**(2)	**1.9**(1)	4.9(5)	2159(3507)	∞ *2e4*	0/15
CMA-CSA	3.5(2)	3.8(3)	8.4(6)	12(11)	**5.9**(3)	15/15
CMA-MSR	**2.1**(0.6)	**1.7**(0.9)	7.3(3)	10(8)	**5.8**(2)	15/15
CMA-TPA	5.2(2)	4.6(3)	4.8(4)	**6.9**(4)	**4.0**(0.9)	15/15
GP1-CMAES	**2.2**(2)	**2.6**(5)	5.0(2)	9.1(6)	10(6)	7/15
GP5-CMAES	**2.6**(2)	**2.6**(2)	**3.8**(6)	11(20)	6.6(6)	9/15
IPOPCMAv3p	3.6(3)	4.0(5)	5.1(4)	**7.8**(12)	8.7(12)	8/15
LHD-10xDef	**2.4**(2)	3.9(8)	6.0(5)	13(12)	∞ *100*	0/15
LHD-2xDefa	**1.6**(2)	**2.4**(2)	**3.0**(2)	**5.1**(2)	∞ *100*	0/15
RAND-2xDef	**2.4**(2)	3.5(1)	**3.2**(1)	10(6)	∞ *100*	0/15
RF1-CMAES	3.2(3)	3.1(2)	9.1(19)	17(10)	∞ *506*	0/15
RF5-CMAES	3.1(3)	22(82)	39(55)	78(92)	∞ *506*	0/15
Sifeg	**1.9**(2)	**2.0**(2)	4.1(6)	933(701)	∞ *2e4*	0/15
Sif	**1.9**(3)	**2.0**(0.6)	4.3(5)	975(1446)	∞ *2e4*	0/15
Srr	**1.9**(2)	**2.0**(1)	3.8(0.4)	1348(1373)	∞ *2e4*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f9	*6.3e+0*:13	*4.0e+0*:15	*2.5e+0*:15	*2.5e-1*:21	*1.0e-8*:94	15/15
BSifeg	15(24)	14(12)	14(17)	210(120)	∞ *2e4*	0/15
BSif	13(10)	12(19)	12(13)	193(218)	∞ *2e4*	0/15
BSqi	13(14)	12(19)	12(13)	207(294)	∞ *2e4*	0/15
BSrr	12(17)	12(20)	14(20)	241(104)	∞ *2e4*	0/15
CMA-CSA	**2.7**(1)	3.5(1)	4.1(5)	8.7(9)	**6.0**(3)	15/15
CMA-MSR	**2.4**(2)	**2.4**(2)	4.6(2)	10(14)	7.0(3)	15/15
CMA-TPA	**2.5**(2)	3.3(4)	4.6(7)	9.3(7)	**6.0**(2)	15/15
GP1-CMAES	**2.8**(2)	4.1(5)	6.8(17)	15(22)	26(44)	3/15
GP5-CMAES	**1.7**(1)	**1.5**(0.2)	**2.0**(2)	**6.0**(3)	**5.8**(6)	10/15
IPOPCMAv3p	**2.3**(2)	**2.2**(2)	**2.5**(2)	6.1(5)	11(10)	7/15
LHD-10xDef	**2.8**(2)	3.4(2)	3.8(3)	16(15)	∞ *100*	0/15
LHD-2xDefa	**1.3**(0.7)	**1.2**(0.8)	**1.9**(1)	**4.6**(3)	∞ *100*	0/15
RAND-2xDef	**1.6**(1)	**1.6**(1)	**2.0**(0.9)	**6.0**(5)	∞ *100*	0/15
RF1-CMAES	4.8(3)	6.6(35)	11(10)	26(40)	∞ *506*	0/15
RF5-CMAES	21(28)	20(19)	21(31)	57(61)	∞ *504*	0/15
Sifeg	**2.6**(3)	**2.8**(2)	4.0(5)	182(348)	∞ *2e4*	0/15
Sif	**2.5**(3)	5.0(3)	5.8(4)	259(407)	∞ *2e4*	0/15
Srr	**2.5**(3)	**2.8**(2)	3.6(3)	232(223)	∞ *2e4*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f10	*1.6e+6*:2.0	*4.0e+5*:3.2	*6.3e+2*:8.8	*1.0e+1*:30	*2.5e-8*:101	15/15
BSifeg	**2.7**(2)	**2.5**(0.2)	**1.5**(0.5)	38(59)	∞ *5883*	0/15
BSif	**2.7**(2)	**2.5**(0.7)	**1.5**(0.5)	71(116)	∞ *8444*	0/15
BSqi	**2.7**(2)	**2.5**(2)	**1.5**(0.5)	135(10)	∞ *1e4*	0/15
BSrr	**2.7**(2)	**2.5**(0.2)	**1.5**(0.5)	46(108)	∞ *6681*	0/15
CMA-CSA	3.7(4)	3.3(4)	7.2(3)	7.5(4)	**5.3**(0.9)	15/15
CMA-MSR	**2.1**(2)	**2.3**(2)	7.0(5)	7.4(4)	6.8(0.6)	15/15
CMA-TPA	3.4(4)	**2.8**(3)	5.9(5)	**6.3**(3)	5.3(1)	15/15
GP1-CMAES	**1.5**(2)	**1.5**(0.5)	4.4(3)	**4.6**(3)	**5.2**(2)	12/15
GP5-CMAES	**2.2**(2)	**2.2**(2)	**2.2**(0.9)	**1.6**(0.6)	**2.5**(0.9)	14/15
IPOPCMAv3p	**2.6**(3)	**2.7**(2)	6.8(3)	8.3(7)	∞ *506*	0/15
LHD-10xDef	**1.7**(1)	**1.6**(2)	7.9(1)	16(9)	∞ *100*	0/15
LHD-2xDefa	**1.7**(2)	**2.2**(1)	3.8(2)	6.9(7)	∞ *100*	0/15
RAND-2xDef	**2.7**(3)	**2.7**(2)	3.9(1)	10(9)	∞ *100*	0/15
RF1-CMAES	**2.3**(3)	**2.2**(2)	4.2(1)	19(25)	∞ *506*	0/15
RF5-CMAES	3.4(2)	**2.9**(2)	8.7(6)	14(18)	∞ *502*	0/15
Sifeg	**2.7**(2)	**2.8**(1)	**1.9**(0.3)	8.6(12)	∞ *2159*	0/15
Sif	**2.7**(2)	**2.8**(2)	**1.9**(0.4)	12(17)	∞ *2178*	0/15
Srr	**2.7**(2)	**2.8**(1)	**1.9**(0.3)	7.9(0.1)	∞ *2193*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f11	*1.0e+7*:1.1	*1.6e+6*:3.2	*1.0e+4*:6.6	*4.0e+1*:23	*4.0e-8*:100	15/15
BSifeg	**1.2**(0)	**1.6**(1)	**1.4**(0.3)	47(49)	∞ *6926*	0/15
BSif	**1.2**(0.5)	**1.6**(1)	**1.4**(0.3)	39(23)	∞ *5430*	0/15
BSqi	**1.2**(0)	**1.6**(1)	**1.4**(0.3)	71(92)	∞ *5998*	0/15
BSrr	**1.2**(0.5)	**1.6**(1)	**1.4**(0.6)	40(109)	∞ *6000*	0/15
CMA-CSA	**1.6**(0.2)	**1.3**(1)	3.5(3)	4.7(5)	**5.2**(0.8)	15/15
CMA-MSR	**1.1**(0.9)	**0.96**(1)	3.8(5)	5.3(0.8)	6.6(0.6)	15/15
CMA-TPA	**1.9**(2)	**1.5**(1)	4.2(3)	6.0(3)	5.3(0.4)	15/15
GP1-CMAES	**2.2**(4)	**1.6**(2)	3.4(3)	**3.6**(3)	**4.7**(0.8)	14/15
GP5-CMAES	3.6(4)	**1.8**(2)	**2.2**(0.9)	**1.3**(0.4)	**2.4**(0.7)	15/15
IPOPCMAv3p	3.1(2)	**1.9**(1)	3.7(1)	4.3(2)	∞ *506*	0/15
LHD-10xDef	**1.5**(0.7)	**1.3**(0.8)	4.9(5)	11(7)	∞ *100*	0/15
LHD-2xDefa	**1.7**(1)	**1.3**(0.9)	**2.5**(1)	5.6(8)	∞ *100*	0/15
RAND-2xDef	**1.8**(1)	**1.2**(1)	**2.7**(2)	**3.0**(0.5)	∞ *100*	0/15
RF1-CMAES	**2.7**(4)	**2.9**(5)	4.9(4)	5.5(7)	∞ *506*	0/15
RF5-CMAES	**1.7**(0.9)	**2.2**(3)	3.8(3)	6.4(12)	∞ *508*	0/15
Sifeg	**1.2**(1)	**1.7**(1)	**1.7**(1.0)	10(6)	∞ *2391*	0/15
Sif	**1.2**(1)	**1.7**(1)	**1.7**(0.6)	10(29)	∞ *2405*	0/15
Srr	**1.2**(0.5)	**1.7**(1)	**1.7**(0.8)	11(11)	∞ *2357*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f12	*2.5e+8*:1.3	*6.3e+6*:2.7	*6.3e+5*:6.3	*4.0e+1*:21	*1.6e-3*:101	15/15
BSifeg	**0.85**(0.4)	**1.2**(1)	**1.3**(0.8)	12(28)	295(517)	5/15
BSif	**0.85**(0.4)	**1.2**(1)	**1.3**(1.0)	29(3)	368(224)	4/15
BSqi	**0.85**(0.2)	**1.2**(1)	**1.3**(0.5)	23(5)	320(351)	4/15
BSrr	**0.85**(0.2)	**1.2**(1)	**1.3**(0.9)	19(0.4)	326(335)	4/15
CMA-CSA	**1.4**(2)	**1.9**(3)	**1.4**(2)	6.5(5)	**8.5**(17)	15/15
CMA-MSR	**1.3**(0.8)	**1.3**(0.7)	**0.87**(0.8)	9.1(7)	9.3(5)	15/15
CMA-TPA	**1.4**(2)	**1.8**(2)	**1.8**(2)	7.0(7)	**6.8**(9)	15/15
GP1-CMAES	**2.0**(6)	**1.5**(2)	**1.2**(1)	3.6(1.0)	10(9)	7/15
GP5-CMAES	**1.2**(2)	**1.6**(1)	**1.3**(2)	**3.5**(6)	**7.8**(5)	7/15
IPOPCMAv3p	**1.7**(3)	**1.8**(1)	**1.7**(2)	4.0(1)	14(26)	5/15
LHD-10xDef	**0.90**(0.4)	**1.2**(0.7)	**0.97**(1)	24(19)	∞ *100*	0/15
LHD-2xDefa	**1**(0.2)	**0.78**(0.6)	**0.97**(0.9)	5.0(4)	∞ *100*	0/15
RAND-2xDef	**1.1**(0.4)	**0.95**(0.7)	**1.1**(0.8)	6.3(11)	∞ *100*	0/15
RF1-CMAES	**1.6**(4)	**1.5**(2)	**1.4**(0.8)	4.8(3)	∞ *506*	0/15
RF5-CMAES	**1.2**(0.2)	**1.9**(3)	**1.9**(2)	11(6)	∞ *504*	0/15
Sifeg	**0.85**(0)	**1.2**(1)	**1.3**(1)	**2.8**(6)	76(138)	5/15
Sif	**0.85**(0)	**1.2**(1)	**1.4**(0.6)	4.9(14)	83(92)	5/15
Srr	**0.85**(0.4)	**1.2**(1)	**1.4**(1)	**3.4**(1)	124(113)	4/15

Table 14: 02-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	*4.0e+2*:1.6	*2.5e+2*:3.1	*6.3e+1*:8.7	*1.0e+1*:23	*4.0e-6*:100	15/15
BSifeg	**2.3**(2)	**1.6**(1)	184(1215)	541(602)	∞ *2e4*	0/15
BSif	**2.3**(2)	**1.6**(1)	224(780)	789(712)	∞ *2e4*	0/15
BSqi	**2.3**(2)	**1.6**(1)	252(879)	430(542)	∞ *2e4*	0/15
BSrr	**2.3**(2)	**1.6**(1)	184(0.7)	491(498)	∞ *2e4*	0/15
CMA-CSA	**2.5**(2)	**2.1**(1)	**2.8**(3)	**2.9**(3)	**5.0**(0.5)	15/15
CMA-MSR	**2.8**(5)	**2.1**(2)	4.1(4)	3.7(1)	**6.2**(1)	15/15
CMA-TPA	**2.8**(3)	**2.5**(2)	3.5(5)	4.5(3)	**5.3**(0.6)	15/15
GP1-CMAES	**1.6**(2)	**1.4**(1)	**1.9**(2)	**2.7**(2)	24(20)	3/15
GP5-CMAES	**2.4**(3)	**1.9**(3)	5.5(1)	3.8(2)	11(4)	6/15
IPOPCMAv3p	**2.2**(3)	**1.8**(2)	**2.6**(2)	3.9(2)	∞ *506*	0/15
LHD-10xDef	**2.0**(1)	**2.2**(2)	**2.8**(2)	**2.9**(1)	∞ *100*	0/15
LHD-2xDefa	**1.7**(2)	**1.5**(0.6)	**1.4**(0.7)	**1.4**(0.8)	∞ *100*	0/15
RAND-2xDef	**1.6**(1)	**1.8**(2)	**1.7**(0.8)	**1.4**(0.8)	∞ *100*	0/15
RF1-CMAES	**2.1**(0.9)	**1.8**(3)	**2.2**(2)	7.9(7)	∞ *506*	0/15
RF5-CMAES	**2.6**(2)	**2.1**(1)	**1.9**(3)	11(33)	∞ *508*	0/15
Sifeg	**2.3**(2)	**1.6**(1)	178(899)	563(826)	∞ *2e4*	0/15
Sif	**2.3**(2)	**1.6**(1)	258(1006)	410(561)	∞ *2e4*	0/15
Srr	**2.3**(2)	**1.6**(1.0)	278(884)	428(499)	∞ *2e4*	0/15

Table 15: 02-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*:1.4	*2.5e+0*:4.2	*1.0e+0*:7.4	*2.5e-2*:21	*1.0e-8*:101	15/15
BSifeg	**2.0**(3)	12(11)	8.5(20)	7.3(22)	∞ *2e4*	0/15
BSif	**2.0**(3)	12(20)	8.7(11)	6.8(12)	∞ *2e4*	0/15
BSqi	**2.0**(3)	7.4(15)	5.7(7)	4.2(4)	∞ *2e4*	0/15
BSrr	**2.0**(2)	11(24)	8.4(13)	6.2(6)	∞ *2e4*	0/15
CMA-CSA	**1.3**(0.7)	**1.7**(1)	**1.9**(2)	4.1(2)	**5.5**(0.9)	15/15
CMA-MSR	**2.1**(2)	**2.4**(1.0)	**2.8**(2)	6.0(2)	**6.5**(0.7)	15/15
CMA-TPA	**2.4**(1)	3.5(4)	3.5(2)	5.8(2)	**5.8**(0.8)	15/15
GP1-CMAES	**0.95**(1)	**2.2**(0.6)	**1.9**(2)	**2.5**(0.7)	∞ *506*	0/15
GP5-CMAES	**1.9**(3)	**2.9**(2)	**2.3**(1)	**1.7**(0.8)	∞ *506*	0/15
IPOPCMAv3p	**1.2**(0.7)	3.3(2)	3.1(2)	3.8(1)	∞ *506*	0/15
LHD-10xDef	**1.4**(0.7)	**1.2**(1)	3.1(3)	8.2(9)	∞ *100*	0/15
LHD-2xDefa	**1.4**(0.9)	**1.6**(2)	**1.4**(1)	**2.0**(1)	∞ *100*	0/15
RAND-2xDef	**1.1**(1)	**2.2**(1)	**1.9**(1)	**2.1**(0.9)	∞ *100*	0/15
RF1-CMAES	**1.4**(2)	5.9(6)	6.3(9)	8.8(9)	∞ *506*	0/15
RF5-CMAES	**1.2**(0.9)	44(46)	48(59)	79(74)	∞ *506*	0/15
Sifeg	**2.0**(1)	**2.8**(1)	**2.3**(1)	3.2(5)	∞ *2e4*	0/15
Sif	**2.0**(2)	**2.8**(3)	**2.4**(0.7)	4.1(4)	∞ *2e4*	0/15
Srr	**2.0**(1)	**2.6**(2)	**2.1**(0.7)	**2.1**(3)	∞ *2e4*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f15	*1.6e+2*:1.2	*4.0e+1*:4.7	*2.5e+1*:10	*1.0e+1*:37	*2.5e+0*:118	5/5
BSifeg	**1.4**(0.4)	**1.8**(1.0)	**2.2**(0.8)	27(5)	111(252)	10/15
BSif	**1.4**(1)	151(561)	70(257)	22(140)	118(120)	10/15
BSqi	**1.4**(2)	3.3(0.9)	4.0(9)	3.3(4)	62(119)	13/15
BSrr	**1.4**(0.4)	6.5(39)	6.4(0.1)	4.2(11)	59(91)	13/15
CMA-CSA	**1.3**(0)	**1.0**(1)	**1.1**(1)	**1.1**(0.5)	**1.6**(3)	15/15
CMA-MSR	**2.1**(0.8)	**1.3**(2)	**1.3**(2)	**0.86**(0.5)	**2.2**(0.6)	15/15
CMA-TPA	**1.8**(2)	**2.2**(5)	**1.5**(0.7)	**1.5**(1)	3.6(5)	15/15
GP1-CMAES	**2.3**(5)	**1.6**(1)	**1.3**(1.0)	**1.3**(1)	3.1(3)	10/15
GP5-CMAES	**2.4**(2)	**2.1**(2)	**1.3**(0.4)	**0.75**(0.3)	**1.6**(3)	14/15
IPOPCMAv3p	**1.4**(0.4)	**2.0**(2)	**1.4**(0.5)	**1.3**(1)	4.8(5)	8/15
LHD-10xDef	**2.5**(2)	**1.8**(2)	**1.6**(2)	**2.2**(1)	**3.0**(2)	4/15
LHD-2xDefa	**1.4**(1)	**0.81**(1)	**0.94**(0.7)	**0.90**(0.8)	**1.4**(2)	7/15
RAND-2xDef	**2.1**(0.8)	**1.6**(2)	**1.1**(0.8)	**1.4**(1.0)	**1.4**(2)	7/15
RF1-CMAES	3.2(6)	**1.8**(0.8)	**1.5**(1)	**1.9**(0.7)	**1.8**(2)	12/15
RF5-CMAES	**2.0**(1)	**1.8**(1)	**1.9**(3)	6.1(1)	8.0(6)	6/15
Sifeg	**1.4**(2)	**1.5**(2)	**1.8**(0.7)	**1.7**(3)	50(103)	12/15
Sif	**1.4**(1)	**1.5**(2)	**1.9**(3)	**1.7**(0.4)	33(42)	13/15
Srr	**1.4**(1)	**1.5**(1)	**1.5**(0.5)	**1.6**(3)	35(46)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f16	*1.0e+2*:1.1	*2.5e+1*:3.9	*1.6e+1*:6.5	*4.0e+0*:31	*2.5e-1*:127	5/5
BSifeg	**1.8**(1)	**2.9**(5)	**2.7**(3)	**1.0**(0.8)	12(29)	15/15
BSif	**1.8**(0.9)	**2.9**(2)	**2.9**(2)	**1.1**(1)	7.0(12)	15/15
BSqi	**1.8**(3)	3.9(1)	**2.9**(5)	**1.1**(0.9)	17(14)	15/15
BSrr	**1.8**(3)	3.4(1)	3.8(4)	**1.2**(1)	20(54)	15/15
CMA-CSA	**1.8**(2)	3.3(3)	4.1(3)	6.1(9)	3.6(4)	15/15
CMA-MSR	**1.5**(0.9)	**1.9**(3)	**2.3**(2)	6.7(0.9)	8.1(7)	15/15
CMA-TPA	**1.8**(2)	**1.6**(2)	**2.1**(2)	3.7(7)	**3.0**(4)	15/15
GP1-CMAES	**1.5**(0.7)	**1.8**(0.6)	**1.5**(1)	**2.3**(1)	4.6(7)	8/15
GP5-CMAES	**1.6**(2)	**2.3**(2)	**2.4**(2)	11(15)	6.6(10)	7/15
IPOPCMAv3p	**1.2**(0.4)	**1.6**(1)	**1.7**(2)	**1.2**(0.9)	**2.5**(4)	11/15
LHD-10xDef	**1.5**(0.9)	**1.8**(0.8)	**1.5**(1)	**1.5**(1)	5.8(6)	2/15
LHD-2xDefa	**1.5**(0.9)	**1.7**(3)	**2.0**(2)	**1.4**(1)	**2.6**(3)	4/15
RAND-2xDef	**1.2**(0.4)	**1.7**(2)	**1.5**(2)	**1.0**(1)	**1.1**(1.0)	8/15
RF1-CMAES	**1.8**(2)	**1.7**(1)	**1.7**(3)	3.6(5)	5.3(4)	7/15
RF5-CMAES	**1.1**(0)	**1.3**(1.0)	**1.7**(2)	6.1(8)	6.4(9)	6/15
Sifeg	**1.8**(3)	**2.4**(2)	**2.1**(0.7)	**0.94**(0.6)	7.4(0.8)	15/15
Sif	**1.8**(3)	**2.4**(2)	**2.0**(1)	**0.89**(0.2)	14(42)	15/15
Srr	**1.8**(3)	**2.3**(2)	**2.0**(1)	**0.91**(0.4)	12(29)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f17	*4.0e+1*:1.2	*1.0e+1*:2.7	*4.0e+0*:10	*2.5e+0*:28	*1.6e-1*:119	5/5
BSifeg	**1.2**(1)	**2.2**(2)	**1.9**(2)	6.6(22)	41(41)	14/15
BSif	**1.2**(1)	**2.2**(2)	**2.1**(3)	11(21)	40(48)	15/15
BSqi	**1.2**(0.8)	**2.2**(2)	**1.9**(2)	32(0.8)	21(28)	15/15
BSrr	**1.2**(0.6)	**2.2**(2)	**1.9**(2)	21(0.7)	40(57)	15/15
CMA-CSA	**1.3**(0.6)	3.1(10)	9.3(14)	3.5(11)	**2.1**(2)	15/15
CMA-MSR	**1.7**(1)	20(59)	12(20)	5.8(1)	3.5(3)	15/15
CMA-TPA	**1.4**(0.8)	3.6(6)	3.3(6)	**1.4**(1)	**2.2**(0.3)	15/15
GP1-CMAES	**2.2**(3)	3.8(9)	6.6(26)	4.1(10)	4.7(8)	8/15
GP5-CMAES	**1.4**(0.6)	7.6(7)	17(29)	8.8(12)	5.7(4)	7/15
IPOPCMAv3p	**1.2**(0.4)	**2.5**(2)	**1.8**(2)	**2.2**(0.4)	**2.3**(3)	12/15
LHD-10xDef	**1.4**(0.6)	**1.8**(2)	**1.8**(2)	**1.6**(1.0)	4.1(5)	3/15
LHD-2xDefa	**1.2**(0.8)	**1.5**(2)	**1.5**(1)	**0.77**(0.6)	3.9(6)	3/15
RAND-2xDef	**1.5**(0.6)	**1.9**(2)	**1.8**(1)	**0.79**(0.3)	**2.6**(4)	4/15
RF1-CMAES	**1.3**(1)	**2.6**(4)	5.6(1)	**2.3**(0.7)	7.2(10)	6/15
RF5-CMAES	**1.4**(0)	50(50)	29(35)	16(15)	61(86)	1/15
Sifeg	**1.2**(0.6)	**2.2**(2)	**1.5**(2)	**1.4**(2)	6.8(5)	15/15
Sif	**1.2**(0)	**2.2**(1)	**1.6**(2)	6.4(1)	8.5(11)	15/15
Srr	**1.2**(0.2)	**2.2**(2)	**1.4**(2)	**1.9**(4)	3.6(5)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f18	*4.0e+2*:1.2	*1.0e+2*:3.2	*4.0e+1*:7.2	*6.3e+0*:32	*1.6e+0*:104	5/5
BSifeg	**1.2**(0)	**1.2**(1)	**1.0**(0.6)	10(2)	133(112)	10/15
BSif	**1.2**(1)	**1.2**(1)	**1.0**(0.7)	12(45)	126(193)	10/15
BSqi	**1.2**(0.6)	**1.2**(0.9)	**1.0**(0.6)	16(37)	119(162)	10/15
BSrr	**1.2**(0.6)	**1.2**(1)	**1.0**(0.5)	13(0.4)	98(172)	12/15
CMA-CSA	**1**(0.4)	**2.0**(4)	**1.8**(2)	**2.7**(6)	3.5(3)	15/15
CMA-MSR	**1.3**(1)	**1.8**(1)	**1.6**(2)	6.1(6)	6.2(6)	15/15
CMA-TPA	**0.94**(0)	**1.5**(0.5)	**1.2**(0.3)	**1.2**(1)	**3.0**(3)	15/15
GP1-CMAES	**2.7**(3)	**2.2**(2)	4.2(9)	**2.0**(5)	4.3(4)	9/15
GP5-CMAES	14(0.6)	26(1)	17(53)	9.5(16)	6.5(10)	8/15
IPOPCMAv3p	**2.4**(8)	**1.9**(0.6)	**1.7**(0.8)	**1.1**(0.9)	**2.7**(10)	11/15
LHD-10xDef	**1.2**(1)	**1.1**(1)	**1.2**(0.8)	**1.8**(1)	**2.3**(2)	6/15
LHD-2xDefa	**1.3**(1)	**1.3**(0.7)	**0.80**(0.5)	**1.0**(0.9)	**1.1**(0.9)	9/15
RAND-2xDef	**1.2**(0.8)	**1.1**(1)	**0.79**(0.8)	**0.96**(0.6)	**1.7**(2)	7/15
RF1-CMAES	**1.2**(0.8)	**1.9**(2)	**1.1**(1)	**1.3**(0.6)	8.0(2)	6/15
RF5-CMAES	**1.6**(2)	3.9(8)	6.4(2)	5.0(12)	9.0(15)	6/15
Sifeg	**1.2**(0.6)	**1.1**(0.8)	**1.0**(0.7)	24(2)	52(98)	14/15
Sif	**1.2**(1)	**1.1**(1)	**1.0**(0.8)	19(1)	119(97)	11/15
Srr	**1.2**(1)	**1.1**(0.8)	**1.0**(0.7)	40(255)	129(97)	10/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f19	*1.6e-1*:23	*1.0e-1*:26	*6.3e-2*:38	*4.0e-2*:40	*1.0e-2*:216	15/15
BSifeg	**4.0**(3)	**4.0**(3)	**2.8**(2)	**3.7**(4)	15(10)	15/15
BSif	4.2(3)	4.2(4)	**3.0**(3)	**3.7**(3)	33(77)	12/15
BSqi	**3.6**(2)	**3.7**(2)	**2.6**(2)	4.2(5)	35(31)	14/15
BSrr	**3.5**(2)	**3.6**(2)	**2.6**(3)	**3.7**(4)	28(58)	13/15
CMA-CSA	13(4)	21(45)	18(20)	26(22)	13(8)	15/15
CMA-MSR	8.0(11)	12(10)	16(23)	26(43)	13(14)	15/15
CMA-TPA	7.0(3)	6.2(6)	14(15)	18(23)	**6.2**(8)	15/15
GP1-CMAES	8.8(9)	11(15)	12(15)	12(17)	16(21)	2/15
GP5-CMAES	12(11)	17(20)	13(13)	15(14)	16(9)	2/15
IPOPCMAv3p	7.9(8)	14(24)	11(16)	12(8)	34(37)	1/15
LHD-10xDef	8.9(12)	18(15)	38(99)	36(28)	**6.9**(11)	1/15
LHD-2xDefa	7.5(7)	8.3(18)	12(18)	37(45)	∞ *100*	0/15
RAND-2xDef	5.6(6)	16(17)	11(13)	11(15)	**6.6**(13)	1/15
RF1-CMAES	17(34)	19(11)	17(17)	28(27)	10(4)	3/15
RF5-CMAES	12(31)	12(10)	15(17)	20(40)	∞ *504*	0/15
Sifeg	7.3(2)	6.8(6)	6.5(6)	7.2(5)	31(22)	14/15
Sif	6.5(5)	6.2(4)	6.7(5)	7.6(4)	25(22)	15/15
Srr	5.0(3)	5.1(4)	5.1(3)	6.3(4)	28(57)	14/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f20	*4.0e+3*:1.9	*2.5e+2*:2.8	*4.0e+0*:6.3	*2.5e+0*:21	*6.3e-1*:139	15/15
BSifeg	**2.1**(2)	14(4)	17(32)	10(22)	28(47)	15/15
BSif	**2.1**(2)	13(3)	19(32)	8.0(3)	62(71)	12/15
BSqi	**2.1**(2)	12(1)	19(24)	8.1(14)	16(40)	15/15
BSrr	**2.1**(2)	14(5)	16(1)	8.3(9)	18(17)	15/15
CMA-CSA	**2.4**(1)	**3.0**(3)	**2.4**(2)	**1.6**(2)	**6.0**(7)	15/15
CMA-MSR	**1.7**(2)	**1.9**(2)	**2.6**(4)	**2.2**(0.9)	13(19)	15/15
CMA-TPA	3.2(3)	4.2(2)	4.0(5)	**2.1**(2)	11(13)	15/15
GP1-CMAES	**1.9**(3)	**2.3**(2)	**1.6**(0.9)	**1.6**(0.8)	**4.3**(4)	8/15
GP5-CMAES	**2.0**(3)	**2.0**(1)	**2.2**(2)	**2.0**(0.3)	8.1(8)	5/15
IPOPCMAv3p	**2.3**(2)	**2.5**(3)	**2.3**(3)	**1.5**(2)	16(21)	3/15
LHD-10xDef	**1.8**(2)	4.0(3)	3.3(2)	**2.1**(2)	**5.1**(4)	2/15
LHD-2xDefa	**1.8**(2)	**2.0**(1)	**2.1**(0.8)	**1.7**(2)	11(14)	1/15
RAND-2xDef	**1.3**(0.7)	**2.8**(2)	**2.5**(1)	**1.9**(0.4)	11(10)	1/15
RF1-CMAES	**2.1**(2)	**2.7**(3)	3.2(4)	3.5(2)	16(11)	3/15
RF5-CMAES	**2.6**(3)	21(8)	27(23)	12(18)	26(33)	2/15
Sifeg	**2.1**(1)	3.6(3)	3.6(0.9)	**1.8**(2)	17(12)	15/15
Sif	**2.1**(2)	3.7(2)	3.7(4)	**1.9**(1)	20(28)	15/15
Srr	**2.1**(2)	3.5(3)	3.0(2)	**1.8**(2)	16(71)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f21	*1.0e+1*:1.7	*6.3e+0*:2.6	*2.5e+0*:7.9	*1.6e+0*:30	*4.0e-1*:105	15/15
BSifeg	**2.1**(1)	**3.0**(2)	3.0(6)	**2.2**(4)	42(163)	13/15
BSif	**2.1**(2)	**2.9**(2)	**3.0**(2)	**2.9**(5)	62(180)	12/15
BSqi	**2.1**(1)	**2.5**(2)	3.4(8)	**2.4**(4)	53(124)	13/15
BSrr	**2.1**(2)	**2.9**(2)	3.4(4)	**2.3**(4)	62(61)	12/15
CMA-CSA	**1.5**(1)	**2.4**(4)	5.2(13)	4.0(14)	4.5(8)	15/15
CMA-MSR	**1.9**(1)	**2.6**(1)	**1.7**(2)	12(35)	156(985)	14/15
CMA-TPA	**1.5**(2)	**2.7**(4)	**2.5**(2)	5.2(12)	37(165)	15/15
GP1-CMAES	**1.8**(2)	**2.1**(1)	**1.6**(2)	10(12)	6.2(14)	7/15
GP5-CMAES	**1.5**(0.6)	**2.5**(2)	3.8(1)	6.4(6)	8.9(8)	6/15
IPOPCMAv3p	**1.6**(0.9)	**2.0**(3)	**1.9**(0.9)	3.4(6)	3.7(9)	9/15
LHD-10xDef	**1.4**(2)	**1.8**(2)	**1.2**(2)	**1.2**(1)	**0.74**(0.7)	12/15
LHD-2xDefa	**1.4**(2)	**1.7**(2)	**1.3**(0.8)	**0.75**(0.5)	**0.45**(0.5)	14/15
RAND-2xDef	**1.5**(0.9)	**2.8**(3)	**1.5**(0.9)	**0.90**(0.6)	**0.49**(0.2)	14/15
RF1-CMAES	**1.2**(0.6)	**3.0**(4)	7.5(32)	13(26)	20(37)	3/15
RF5-CMAES	**1.7**(2)	**2.1**(3)	**1.2**(0.9)	4.4(4)	11(6)	5/15
Sifeg	**2.1**(1)	**2.3**(2)	**1.3**(0.7)	32(2)	88(96)	12/15
Sif	**2.1**(1)	**2.3**(2)	**1.3**(0.7)	49(4)	98(79)	11/15
Srr	**2.1**(1)	**2.3**(2)	**1.3**(0.8)	21(1)	107(79)	11/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f22	*4.0e+1*:1.3	*1.6e+1*:3.2	*6.3e+0*:9.3	*1.6e+0*:25	*1.0e-1*:168	15/15
BSifeg	**1.9**(3)	**1.9**(2)	3.9(2)	18(12)	94(77)	9/15
BSif	**1.9**(3)	**1.9**(2)	3.6(7)	37(107)	80(115)	10/15
BSqi	**1.9**(2)	**1.9**(2)	4.8(5)	14(5)	91(161)	9/15
BSrr	**1.9**(2)	**1.9**(2)	5.5(9)	26(38)	85(90)	10/15
CMA-CSA	**1.8**(2)	**1.9**(3)	5.6(2)	17(34)	14(5)	15/15
CMA-MSR	**1.9**(2)	**2.2**(3)	7.3(22)	22(11)	18(52)	15/15
CMA-TPA	**1.7**(1)	**1.4**(1)	**1.9**(2)	14(24)	7.6(5)	15/15
GP1-CMAES	**1.3**(0.4)	**1.1**(1)	10(14)	6.1(15)	6.8(8)	5/15
GP5-CMAES	**1.6**(2)	**1.4**(1)	3.6(14)	4.1(6)	**2.1**(2)	11/15
IPOPCMAv3p	**1.7**(0.8)	**1.3**(1)	**2.3**(4)	4.8(16)	4.2(6)	7/15
LHD-10xDef	**1.3**(1.0)	**1.2**(0.9)	**1.3**(1)	**1.8**(1)	**0.81**(0.7)	9/15
LHD-2xDefa	**1.3**(1)	**1.9**(2)	**1.3**(2)	**1.5**(1)	**1.9**(2)	4/15
RAND-2xDef	**1.3**(0.6)	**1.3**(1)	**1.2**(0.7)	**1.3**(1)	**0.58**(0.8)	10/15
RF1-CMAES	**1.6**(0.8)	**1.3**(2)	6.3(4)	12(15)	12(8)	3/15
RF5-CMAES	**1.6**(1.0)	**1.6**(0.9)	8.5(26)	13(20)	12(11)	3/15
Sifeg	**2.0**(4)	**2.1**(3)	**2.3**(0.5)	8.9(22)	84(96)	10/15
Sif	**2.0**(0)	**2.3**(1)	**2.3**(2)	10(41)	103(119)	9/15
Srr	**2.0**(2)	**2.0**(2)	**2.4**(4)	27(9)	89(149)	11/15

Table 24: 02-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	*4.0e+1*:1.5	*2.5e+1*:2.6	*1.0e+1*:7.8	*4.0e+0*:55	*2.5e+0*:103	5/5
BSifeg	**2.0**(3)	**1.6**(2)	**1.9**(1)	**1.3**(2)	**0.91**(0.7)	15/15
BSif	**2.0**(3)	**1.6**(2)	**1.9**(2)	**1.3**(2)	**1.0**(0.9)	15/15
BSqi	**2.0**(2)	**1.6**(2)	**1.8**(2)	**1.6**(1)	**1.3**(1)	15/15
BSrr	**2.0**(3)	**1.6**(1)	**2.1**(1)	**1.5**(1.0)	**1.1**(0.8)	15/15
CMA-CSA	**1.3**(2)	**1.1**(0.6)	3.2(2)	**3.0**(2)	**2.8**(6)	15/15
CMA-MSR	**1.7**(1.0)	**2.3**(3)	**2.2**(3)	6.3(6)	4.8(6)	15/15
CMA-TPA	**1.1**(0.2)	**1.2**(1)	**1.9**(3)	3.2(4)	4.8(2)	15/15
GP1-CMAES	**1.2**(0.8)	**1.6**(2)	**2.2**(3)	3.0(3)	4.4(8)	10/15
GP5-CMAES	**1.7**(2)	**1.7**(1.0)	**2.0**(2)	**2.9**(5)	**2.2**(2)	13/15
IPOPCMAv3p	**1.3**(0.7)	**1.0**(0.6)	**0.99**(2)	**1.8**(1)	**2.8**(3)	13/15
LHD-10xDef	**1.6**(1.0)	**1.1**(0.7)	**1.5**(1.0)	**1.7**(2)	**1.8**(1)	6/15
LHD-2xDefa	**1.8**(1)	**2.1**(3)	**2.4**(3)	**1.6**(2)	4.5(7)	3/15
RAND-2xDef	**1.0**(1.0)	**1.4**(1)	**2.0**(2)	**1.5**(0.9)	4.2(5)	3/15
RF1-CMAES	**1.7**(1)	**1.4**(1)	**1.8**(3)	3.5(4)	**2.4**(1)	14/15
RF5-CMAES	**1.8**(1)	**1.4**(1)	**1.5**(2)	**1.8**(3)	**1.9**(2)	14/15
Sifeg	**1.9**(3)	**1.5**(1)	**1.9**(1)	**1.9**(2)	**2.2**(2)	15/15
Sif	**1.9**(3)	**1.5**(2)	**1.8**(2)	**1.8**(2)	**2.2**(2)	15/15
Srr	**1.9**(2)	**1.5**(1)	**1.9**(1)	**1.9**(2)	**2.3**(3)	15/15

Table 25: 02-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*:1.1	*2.5e+1*:2.7	*1.6e+1*:7.7	*6.3e+0*:44	*2.5e+0*:275	5/5
BSifeg	**2.1**(3)	**2.3**(2)	**1.8**(3)	**2.4**(1)	6.3(1.0)	14/15
BSif	**2.1**(4)	**2.3**(2)	**1.6**(2)	**2.3**(2)	8.3(28)	15/15
BSqi	**2.1**(0.9)	**2.3**(0.8)	**2.9**(0.5)	**2.9**(3)	11(29)	14/15
BSrr	**2.1**(4)	**2.3**(2)	**1.9**(6)	**2.6**(2)	6.4(18)	14/15
CMA-CSA	**2.3**(3)	**2.2**(2)	**1.8**(1)	**1.2**(1)	3.3(6)	15/15
CMA-MSR	3.1(4)	**2.4**(2)	**1.6**(2)	3.8(0.6)	5.6(7)	15/15
CMA-TPA	**1.6**(0.5)	**2.2**(3)	**1.3**(1)	**2.3**(5)	4.1(3)	15/15
GP1-CMAES	**1.6**(1)	**1.7**(2)	**1.9**(3)	**1.8**(6)	**3.3**(2)	7/15
GP5-CMAES	**1.8**(0)	**1.9**(2)	**1.4**(3)	3.4(6)	**2.9**(4)	7/15
IPOPCMAv3p	**1.6**(0.9)	**2.1**(4)	**1.5**(1)	**1.2**(1)	3.9(6)	6/15
LHD-10xDef	**1.9**(1)	**1.4**(1)	**1.5**(2)	**1.5**(1)	∞ *100*	0/15
LHD-2xDefa	**1.8**(1)	**1.6**(0.8)	**2.0**(3)	**2.6**(4)	∞ *100*	0/15
RAND-2xDef	**1.2**(0.9)	**2.4**(2)	**1.8**(1)	**1.9**(3)	**2.5**(4)	2/15
RF1-CMAES	**1.4**(0.2)	**1.8**(3)	**1.3**(1)	**0.80**(0.6)	5.3(7)	4/15
RF5-CMAES	**1.8**(2)	**1.2**(0.5)	**1.2**(1)	**2.5**(3)	13(20)	2/15
Sifeg	**2.1**(3)	**2.4**(2)	**1.4**(0.6)	**2.3**(0.5)	4.9(2)	15/15
Sif	**2.1**(3)	**2.4**(3)	**1.4**(0.6)	**2.1**(2)	5.1(13)	15/15
Srr	**2.1**(2)	**2.4**(2)	**1.4**(0.8)	**2.1**(1)	7.4(4)	15/15

References

[1] Asma Atamna. Benchmarking IPOP-CMA-ES-TPA and IPOP-CMA-ES-MSR on the BBOB noiseless testbed. In Laredo et al. [8], pages 1135–1142.

[2] 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.

[3] Lukás Bajer, Zbynek Pitra, and Martin Holena. Benchmarking gaussian processes and random forests surrogate models on the BBOB noiseless testbed. In Laredo et al. [8], pages 1143–1150.

[4] Dimo Brockhoff, Bernd Bischl, and Tobias Wagner. The impact of initial designs on the performance of matsumoto on the noiseless BBOB-2015 testbed: A preliminary study. In Laredo et al. [8], pages 1159–1166.

[5] 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.

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

[7] 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.

[8] Juan Luis Jiménez Laredo, Sara Silva, and Anna Isabel Esparcia-Alcázar, editors. *Genetic and Evolutionary Computation Conference, GECCO 2015, Madrid, Spain, July 11-15, 2015, Companion Material Proceedings*. ACM, 2015.

[9] Petr Posík and Petr Baudis. Dimension selection in axis-parallel brent-step method for black-box optimization of separable continuous functions. In Laredo et al. [8], pages 1151–1158.