Comparison Tables: BBOB 2015 Testbed in 3-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: 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
BSifeg	**1.7**(2)	**1.5**(1)	**2.1**(0.3)	**2.1**(0.3)	**2.1**(0.3)	15/15
BSif	**1.7**(2)	**1.5**(2)	**2.1**(0.3)	**2.1**(0.3)	**2.1**(0.2)	15/15
BSqi	**1.7**(1)	**1.5**(1)	**2.1**(0.3)	**2.1**(0.3)	**2.1**(0.3)	15/15
BSrr	**1.7**(1)	**1.5**(1)	**2.1**(0.2)	**2.1**(0.2)	**2.1**(0.3)	15/15
CMA-CSA	5.4(2)	5.6(5)	52(3)	52(4)	52(4)	15/15
CMA-MSR	**2.1**(2)	**2.4**(3)	83(8)	83(9)	83(12)	15/15
CMA-TPA	**2.9**(2)	3.4(3)	55(20)	55(14)	55(18)	15/15
GP1-CMAES	**2.4**(2)	3.2(2)	31(3)	31(5)	31(5)	15/15
GP5-CMAES	**2.2**(2)	**2.8**(2)	46(36)	46(29)	46(36)	15/15
IPOPCMAv3p	**2.7**(4)	**2.8**(2)	52(7)	52(8)	52(6)	15/15
LHD-10xDef	**2.1**(2)	3.6(4)	∞	∞	∞ *150*	0/15
LHD-2xDefa	**2.1**(2)	**2.2**(2)	∞	∞	∞ *150*	0/15
RAND-2xDef	**1.8**(2)	**2.3**(2)	∞	∞	∞ *150*	0/15
RF1-CMAES	**2.4**(2)	**2.3**(1)	133(108)	133(106)	133(74)	9/15
RF5-CMAES	12(41)	11(2)	∞	∞	∞ *753*	0/15
Sifeg	**1.7**(2)	**1.5**(1)	7.5(0.7)	7.5(1)	7.5(1)	15/15
Sif	**1.7**(2)	**1.5**(1.0)	7.5(1)	7.5(0.7)	7.5(0.8)	15/15
Srr	**1.7**(1)	**1.5**(1)	7.2(0.6)	7.2(0.8)	7.2(0.9)	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
BSifeg	3.7(4)	**2.8**(0.4)	**1.4**(0.2)	**0.66**(0.3)	**1.2**(0.3)	15/15
BSif	3.7(4)	**2.8**(0.9)	**1.4**(0.5)	**0.68**(0.2)	**1.2**(0.1)	15/15
BSqi	3.7(4)	**2.8**(0.4)	**1.4**(0.2)	**0.62**(0.1)	**1.1**(0.3)	15/15
BSrr	3.7(4)	**2.8**(1)	**1.4**(0.5)	**0.66**(0.0)	**1.3**(0.2)	15/15
CMA-CSA	**1.5**(0.3)	**1.4**(0.7)	3.5(4)	7.3(2)	18(2)	15/15
CMA-MSR	**2.3**(1.0)	**1.3**(0.5)	3.1(5)	6.9(2)	22(3)	15/15
CMA-TPA	**2.2**(1.0)	**2.0**(3)	3.3(5)	7.2(2)	19(3)	15/15
GP1-CMAES	3.1(3)	**2.5**(2)	**2.4**(1)	5.4(1)	45(38)	5/15
GP5-CMAES	3.1(3)	**1.9**(0.9)	**2.3**(1)	**2.5**(1.0)	21(28)	8/15
IPOPCMAv3p	**3.0**(4)	**2.3**(2)	**2.4**(3)	8.9(6)	∞ *751*	0/15
LHD-10xDef	**1.0**(0.3)	**1.2**(1)	**2.7**(3)	70(92)	∞ *150*	0/15
LHD-2xDefa	**1.2**(1.0)	**1.2**(1)	**1.3**(1)	70(89)	∞ *150*	0/15
RAND-2xDef	**1.2**(1.0)	**1.0**(1)	**2.3**(1)	16(16)	∞ *150*	0/15
RF1-CMAES	**2.2**(2)	**1.3**(1)	15(2)	60(69)	∞ *751*	0/15
RF5-CMAES	**1.9**(3)	**1.4**(1)	7.0(1)	167(184)	∞ *760*	0/15
Sifeg	3.7(4)	**2.9**(2)	**1.9**(0.8)	**1.1**(0.1)	**1.7**(0.3)	15/15
Sif	3.7(4)	**2.9**(0.4)	**1.9**(0.8)	**1.2**(0.4)	**1.6**(0.3)	15/15
Srr	3.7(4)	**2.9**(0.2)	**1.9**(1)	**1.0**(0.1)	**1.7**(0.1)	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
BSifeg	3.1(3)	**1.6**(1.0)	**1.3**(0.2)	**0.64**(0.3)	**0.18**(0.1)	15/15
BSif	3.1(3)	**1.6**(1.0)	**1.3**(0.2)	**0.64**(0.4)	**0.19**(0.1)	15/15
BSqi	3.1(2)	**1.6**(0.9)	**1.3**(0.2)	**0.62**(0.2)	**0.19**(0.1)	15/15
BSrr	3.1(2)	**1.6**(0.9)	**1.3**(0.2)	**0.60**(0.2)	**0.19**(0.1)	15/15
CMA-CSA	3.5(4)	**2.1**(2)	**2.5**(2)	3.2(2)	3.3(3)	15/15
CMA-MSR	3.0(3)	**1.8**(2)	**2.5**(2)	5.2(2)	**2.7**(3)	15/15
CMA-TPA	**3.0**(0.9)	**1.8**(2)	**2.3**(2)	**2.3**(2)	**2.6**(2)	15/15
GP1-CMAES	**2.0**(3)	**1.6**(2)	**1.9**(2)	3.7(2)	**1.1**(0.9)	13/15
GP5-CMAES	**1.9**(2)	**1.4**(1)	**1.5**(0.9)	**1.4**(0.6)	**1.6**(2)	11/15
IPOPCMAv3p	**2.8**(3)	**2.3**(3)	3.5(3)	**2.9**(3)	**1.9**(3)	10/15
LHD-10xDef	**2.9**(2)	**1.7**(1)	4.1(2)	4.0(0.3)	3.5(4)	2/15
LHD-2xDefa	**1.6**(0.9)	**1.3**(1)	**1.4**(1)	**1.7**(0.5)	**1.2**(1)	5/15
RAND-2xDef	**1.7**(1)	**1.1**(1)	**1.6**(1.0)	**1.6**(1.0)	**1.2**(1)	5/15
RF1-CMAES	**2.7**(2)	**2.6**(2)	**2.3**(0.6)	5.9(12)	6.8(13)	4/15
RF5-CMAES	**2.4**(1)	8.2(1)	9.2(28)	14(26)	11(18)	3/15
Sifeg	3.1(3)	**1.6**(0.6)	**1.4**(0.3)	**0.72**(0.1)	**0.21**(0.0)	15/15
Sif	3.1(2)	**1.6**(0.2)	**1.4**(0.3)	**0.72**(0.3)	**0.22**(0.1)	15/15
Srr	3.1(3)	**1.6**(1)	**1.4**(0.3)	**0.69**(0.5)	**0.20**(0.0)	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
BSifeg	**2.3**(1)	**2.0**(2)	**2.0**(2)	**1.1**(0.6)	**0.16**(0.1)	15/15
BSif	**2.2**(3)	**2.1**(2)	**2.1**(2)	**1.1**(0.6)	**0.16**(0.1)	15/15
BSqi	**2.2**(3)	**2.0**(2)	**2.0**(1)	**1.1**(0.5)	**0.17**(0.0)	15/15
BSrr	**2.3**(3)	**2.0**(1)	**2.0**(2)	**1.1**(0.6)	**0.16**(0.1)	15/15
CMA-CSA	**1.9**(3)	**1.7**(2)	**1.7**(0.8)	**1.7**(1)	3.9(5)	15/15
CMA-MSR	**3.0**(0.6)	**2.1**(2)	**2.1**(1)	**2.4**(0.9)	7.0(7)	15/15
CMA-TPA	**1.9**(2)	**2.1**(2)	**2.1**(2)	**2.7**(2)	4.0(4)	15/15
GP1-CMAES	**2.5**(3)	**2.3**(1)	**2.3**(2)	**2.0**(0.6)	8.3(10)	2/15
GP5-CMAES	3.1(0.9)	6.0(2)	6.0(5)	4.6(11)	5.5(5)	3/15
IPOPCMAv3p	**1.9**(1)	**2.5**(2)	**2.5**(2)	3.5(1.0)	8.3(13)	2/15
LHD-10xDef	**1.0**(0.9)	**2.1**(1)	**2.1**(4)	**2.9**(2)	∞ *150*	0/15
LHD-2xDefa	**1.2**(2)	**1.4**(2)	**1.4**(2)	**2.5**(3)	3.6(5)	1/15
RAND-2xDef	**1.2**(0.7)	**2.1**(2)	**2.1**(1)	**2.1**(3)	∞ *150*	0/15
RF1-CMAES	**1.4**(2)	3.2(3)	3.2(3)	33(16)	∞ *751*	0/15
RF5-CMAES	19(0.3)	16(15)	16(28)	37(26)	∞ *753*	0/15
Sifeg	**2.1**(2)	**1.7**(0.9)	**1.7**(0.5)	**0.81**(0.5)	**0.14**(0.0)	15/15
Sif	**2.1**(2)	**1.8**(0.5)	**1.8**(1)	**0.85**(0.2)	**0.14**(0.1)	15/15
Srr	**2.1**(2)	**1.7**(1)	**1.7**(1)	**0.80**(0.3)	**0.14**(0.1)	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
BSifeg	3.5(1)	**1.9**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
BSif	3.5(0.5)	**1.9**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
BSqi	3.5(0.9)	**1.9**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
BSrr	3.5(0.7)	**1.9**(0.2)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
CMA-CSA	3.1(3)	**2.4**(5)	5.5(2)	5.5(4)	5.5(3)	15/15
CMA-MSR	**3.0**(2)	**2.0**(1)	4.9(2)	4.9(3)	4.9(3)	15/15
CMA-TPA	**1.8**(2)	**1.9**(1)	3.9(2)	3.9(2)	3.9(2)	15/15
GP1-CMAES	3.7(3)	**2.2**(1)	24(56)	24(6)	24(28)	14/15
GP5-CMAES	3.2(2)	**2.4**(1)	4.7(3)	4.7(3)	4.7(4)	15/15
IPOPCMAv3p	4.3(4)	3.1(4)	10(13)	10(16)	10(13)	15/15
LHD-10xDef	**1.5**(2)	**2.2**(2)	13(0.4)	13(0.4)	13(0.4)	15/15
LHD-2xDefa	**1.8**(1)	**2.0**(2)	3.1(0.8)	3.1(0.4)	3.1(0.4)	15/15
RAND-2xDef	**2.2**(0.9)	**2.2**(1)	3.1(0.4)	3.1(0.4)	3.1(0.4)	15/15
RF1-CMAES	3.7(5)	**2.2**(2)	19(41)	19(22)	19(40)	15/15
RF5-CMAES	3.8(2)	11(32)	150(114)	150(263)	150(97)	7/15
Sifeg	3.5(0.6)	**1.9**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.1)	15/15
Sif	3.5(0.7)	**1.9**(0.2)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	15/15
Srr	3.5(0.5)	**1.9**(0.3)	**1.4**(0.2)	**1.4**(0.2)	**1.4**(0.2)	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
BSifeg	**1.9**(3)	**1.2**(1)	79(273)	146(132)	1254(864)	2/15
BSif	**1.9**(2)	**1.2**(2)	99(184)	209(699)	∞ *3e4*	0/15
BSqi	**1.9**(2)	**1.2**(0.7)	30(19)	231(9)	853(961)	3/15
BSrr	**1.9**(2)	**1.2**(1)	119(211)	114(178)	1324(805)	2/15
CMA-CSA	**1.7**(3)	**1.3**(0.3)	**1.6**(0.5)	**1.5**(2)	**2.6**(0.7)	15/15
CMA-MSR	**2.3**(6)	**2.1**(10)	**2.1**(2)	**2.8**(1)	3.6(0.5)	15/15
CMA-TPA	**2.4**(4)	**2.2**(2)	3.0(2)	3.1(0.7)	**3.0**(0.9)	15/15
GP1-CMAES	3.7(6)	**2.9**(3)	**1.6**(1)	**2.7**(3)	∞ *751*	0/15
GP5-CMAES	3.7(6)	3.3(4)	3.4(4)	**2.5**(2)	∞ *760*	0/15
IPOPCMAv3p	4.0(2)	3.8(3)	**2.8**(4)	**2.8**(2)	**3.2**(0.7)	15/15
LHD-10xDef	**0.81**(0.8)	**1.6**(0.5)	**2.0**(1)	4.2(6)	∞ *150*	0/15
LHD-2xDefa	**0.85**(0.3)	**0.85**(0.5)	**1.4**(2)	4.8(6)	∞ *150*	0/15
RAND-2xDef	**0.96**(0.8)	**1.1**(0.5)	**2.6**(3)	3.1(3)	∞ *150*	0/15
RF1-CMAES	**2.7**(4)	**2.3**(2)	**2.7**(1)	13(33)	∞ *751*	0/15
RF5-CMAES	3.2(2)	34(103)	23(48)	42(54)	∞ *760*	0/15
Sifeg	**1.9**(2)	**1.2**(1)	9.4(11)	81(244)	798(891)	3/15
Sif	**1.9**(3)	**1.2**(1)	25(75)	106(430)	∞ *3e4*	0/15
Srr	**1.9**(2)	**1.2**(1)	35(25)	75(104)	543(577)	4/15

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
BSifeg	**1.7**(1.0)	34(241)	49(133)	142(494)	119(208)	10/15
BSif	**1.7**(2)	34(241)	61(45)	235(400)	233(252)	7/15
BSqi	**1.7**(3)	34(1)	167(178)	215(402)	131(196)	10/15
BSrr	**1.7**(2)	34(121)	131(443)	325(814)	158(233)	9/15
CMA-CSA	**1.9**(1)	**2.4**(2)	**2.8**(2)	**1.9**(1)	**0.89**(1)	15/15
CMA-MSR	**2.2**(2)	**1.8**(2)	3.5(3)	**2.1**(2)	**1.3**(1)	15/15
CMA-TPA	**3.0**(2)	**1.8**(2)	3.8(6)	**2.7**(1)	**1.4**(1)	15/15
GP1-CMAES	**1.4**(0.2)	**1.3**(0.7)	**2.0**(2)	**1.5**(2)	**0.56**(0.6)	15/15
GP5-CMAES	**2.1**(3)	**2.3**(2)	**2.0**(1)	**1.1**(0.8)	**0.63**(1)	15/15
IPOPCMAv3p	**2.8**(2)	3.9(3)	4.9(5)	**2.8**(3)	**1.2**(0.7)	15/15
LHD-10xDef	**1.2**(0.7)	**1.8**(2)	3.7(4)	**2.1**(1)	**1.9**(2)	6/15
LHD-2xDefa	**1.8**(1.0)	**1.8**(2)	**1.9**(1)	**1.5**(2)	**0.98**(0.8)	9/15
RAND-2xDef	**1.7**(1)	**1.2**(1)	**2.4**(2)	**1.6**(2)	**0.90**(0.9)	10/15
RF1-CMAES	**2.0**(2)	3.3(1.0)	7.5(7)	3.8(6)	**2.1**(5)	12/15
RF5-CMAES	**2.5**(3)	**2.7**(2)	10(24)	20(11)	31(36)	2/15
Sifeg	**1.7**(0.5)	**1.9**(2)	37(45)	205(380)	122(76)	11/15
Sif	**1.7**(2)	**2.0**(2)	61(0.6)	234(188)	131(277)	10/15
Srr	**1.7**(3)	**1.9**(2)	97(0.7)	297(543)	154(280)	9/15

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
BSifeg	**1.7**(2)	**2.4**(1)	6.6(14)	7.3(14)	398(343)	6/15
BSif	**1.7**(1)	**2.4**(1)	6.0(10)	6.1(11)	266(223)	7/15
BSqi	**1.7**(1)	**2.4**(1)	4.8(10)	6.7(11)	789(1355)	3/15
BSrr	**1.7**(3)	**2.4**(2)	4.2(12)	4.5(5)	595(363)	4/15
CMA-CSA	4.8(7)	4.1(5)	3.2(2)	3.1(1)	**3.3**(1)	15/15
CMA-MSR	3.4(2)	3.2(4)	3.4(2)	4.0(3)	4.5(1)	15/15
CMA-TPA	**2.6**(3)	**2.5**(4)	**2.9**(2)	3.5(2)	**3.7**(1.0)	15/15
GP1-CMAES	3.2(4)	**2.6**(3)	**2.7**(2)	**2.9**(2)	13(8)	5/15
GP5-CMAES	**2.9**(2)	**2.7**(3)	**2.1**(1.0)	3.0(7)	**4.4**(4)	10/15
IPOPCMAv3p	3.9(3)	3.3(2)	3.2(1)	3.2(1)	5.5(3)	10/15
LHD-10xDef	**2.3**(1)	**2.9**(2)	5.0(3)	10(8)	∞ *150*	0/15
LHD-2xDefa	**1.5**(0.8)	**2.0**(2)	**1.8**(0.5)	3.3(3)	∞ *150*	0/15
RAND-2xDef	**2.0**(1)	**2.2**(2)	**1.9**(1)	3.0(2)	∞ *150*	0/15
RF1-CMAES	**2.0**(0.8)	**2.7**(3)	4.0(9)	13(20)	∞ *751*	0/15
RF5-CMAES	**2.6**(4)	**1.9**(4)	19(22)	53(52)	∞ *753*	0/15
Sifeg	**1.7**(2)	**2.4**(0.8)	**1.4**(1)	**1.8**(2)	123(135)	11/15
Sif	**1.7**(1)	**2.4**(0.8)	**1.3**(0.2)	**2.1**(3)	261(107)	7/15
Srr	**1.7**(1)	**2.4**(1)	**1.3**(2)	**1.7**(2)	211(359)	8/15

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
BSifeg	14(30)	12(25)	48(152)	107(222)	∞ *3e4*	0/15
BSif	12(21)	11(6)	39(15)	179(220)	∞ *3e4*	0/15
BSqi	8.5(14)	7.7(12)	50(13)	88(15)	1267(1585)	2/15
BSrr	8.0(15)	8.2(15)	41(9)	75(17)	∞ *3e4*	0/15
CMA-CSA	3.8(0.9)	3.7(1)	3.2(2)	**3.2**(4)	**4.3**(3)	15/15
CMA-MSR	5.8(3)	6.9(13)	7.0(7)	7.3(7)	**5.9**(3)	15/15
CMA-TPA	4.3(2)	4.6(2)	4.9(1)	6.0(6)	**5.4**(2)	15/15
GP1-CMAES	3.5(1)	4.0(6)	5.5(1)	7.7(8)	35(40)	2/15
GP5-CMAES	**2.4**(1.0)	**2.5**(1)	**3.2**(1)	4.0(0.9)	12(22)	5/15
IPOPCMAv3p	3.5(3)	3.5(2)	**3.2**(2)	**3.3**(4)	8.2(6)	8/15
LHD-10xDef	10(12)	16(14)	16(11)	22(28)	∞ *150*	0/15
LHD-2xDefa	**2.5**(2)	**2.9**(2)	3.3(2)	5.8(3)	∞ *150*	0/15
RAND-2xDef	**2.6**(1)	3.3(3)	**3.0**(1)	**4.0**(4)	∞ *150*	0/15
RF1-CMAES	8.1(20)	7.8(16)	9.1(24)	12(17)	∞ *751*	0/15
RF5-CMAES	37(38)	34(52)	33(34)	32(31)	∞ *753*	0/15
Sifeg	**1.9**(0.8)	**1.8**(2)	63(76)	73(209)	∞ *3e4*	0/15
Sif	**1.8**(1)	**1.7**(2)	40(0.8)	87(211)	∞ *3e4*	0/15
Srr	**1.5**(0.4)	**1.4**(1)	101(184)	135(249)	∞ *2e4*	0/15

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
BSifeg	**1.0**(0.3)	**2.5**(1)	4.7(13)	411(417)	1211(992)	1/15
BSif	**1.0**(0.3)	**2.5**(1)	5.7(0.2)	578(334)	1253(1696)	1/15
BSqi	**1.0**(0.3)	**2.5**(1)	6.9(0.2)	304(351)	725(915)	2/15
BSrr	**1.0**(0.3)	**2.5**(1)	5.4(0.3)	691(454)	1155(1570)	1/15
CMA-CSA	**1.4**(2)	**2.3**(2)	**2.8**(2)	4.8(2)	**3.0**(1.0)	15/15
CMA-MSR	**1.4**(2)	**3.0**(1)	**2.0**(2)	**4.0**(2)	3.4(1)	15/15
CMA-TPA	**1.2**(0.9)	3.2(3)	**1.9**(1)	**3.2**(2)	3.1(0.9)	15/15
GP1-CMAES	**2.6**(0.6)	**2.7**(3)	**2.4**(2)	4.3(2)	**2.8**(1)	15/15
GP5-CMAES	**1.3**(0.4)	**1.8**(3)	**1.6**(1)	**2.3**(1)	**1.1**(0.4)	15/15
IPOPCMAv3p	**2.2**(3)	**2.7**(4)	**2.0**(1)	5.8(5)	4.4(6)	13/15
LHD-10xDef	**1.4**(1)	**2.4**(2)	**2.9**(3)	14(16)	∞ *150*	0/15
LHD-2xDefa	**1.2**(0.6)	3.0(2)	**1.8**(1)	11(11)	∞ *150*	0/15
RAND-2xDef	**1.3**(2)	**3.0**(3)	**2.1**(1)	8.4(12)	∞ *150*	0/15
RF1-CMAES	**1.5**(2)	**2.4**(2)	7.8(32)	34(55)	∞ *751*	0/15
RF5-CMAES	**2.0**(2)	3.2(3)	**1.5**(1)	49(79)	∞ *753*	0/15
Sifeg	**1.0**(1)	**2.7**(1)	**1.5**(0.3)	173(211)	467(452)	1/15
Sif	**1.0**(1)	**2.7**(1)	**1.5**(0.4)	109(121)	482(552)	1/15
Srr	**1.0**(0.3)	**2.7**(0.8)	**1.5**(0.4)	105(181)	194(326)	2/15

Table 12: 03-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	*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
BSifeg	**1.6**(1)	**2.0**(1)	**1.2**(0.7)	138(224)	1140(884)	1/15
BSif	**1.6**(1)	**2.0**(1.0)	**1.2**(0.7)	168(241)	365(243)	3/15
BSqi	**1.6**(3)	**2.0**(1)	**1.2**(0.2)	141(245)	∞ *2e4*	0/15
BSrr	**1.6**(1)	**2.0**(1)	**1.2**(0.2)	126(274)	1061(1784)	1/15
CMA-CSA	**1.0**(0.8)	**1.1**(1)	**1.3**(2)	**6.2**(4)	3.0(0.8)	15/15
CMA-MSR	**2.0**(2)	**2.6**(2)	**1.8**(2)	10(5)	**3.0**(0.5)	15/15
CMA-TPA	**1.5**(2)	**1.5**(1)	**1.5**(1)	7.8(5)	3.1(0.9)	15/15
GP1-CMAES	**2.5**(3)	**2.1**(2)	**1.7**(1)	**6.7**(6)	**2.8**(0.6)	15/15
GP5-CMAES	**2.3**(2)	**1.9**(1)	**1.9**(0.6)	**2.7**(1)	**1.2**(0.2)	15/15
IPOPCMAv3p	**1.8**(2)	**2.3**(2)	**1.9**(2)	11(8)	13(9)	5/15
LHD-10xDef	**1.9**(1)	**2.4**(4)	**2.4**(3)	30(44)	∞ *150*	0/15
LHD-2xDefa	**1.8**(2)	**2.1**(2)	**1.5**(1)	11(16)	∞ *150*	0/15
RAND-2xDef	**1.4**(1)	**1.4**(1)	**1.4**(0.9)	11(12)	∞ *150*	0/15
RF1-CMAES	**1.6**(2)	**1.3**(1)	**1.3**(0.9)	34(48)	61(90)	1/15
RF5-CMAES	**2.3**(2)	**1.9**(1)	**1.9**(1)	15(25)	∞ *753*	0/15
Sifeg	**1.7**(2)	**2.0**(1)	**1.4**(0.9)	50(104)	∞ *7533*	0/15
Sif	**1.7**(1)	**2.0**(1)	**1.4**(0.3)	66(123)	∞ *7579*	0/15
Srr	**1.7**(4)	**2.0**(1)	**1.4**(0.5)	60(177)	189(350)	3/15

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
BSifeg	**1.1**(0.7)	**1.3**(1.0)	**1.1**(0.2)	10(7)	48(52)	11/15
BSif	**1.1**(1.0)	**1.3**(1)	**1.1**(0.3)	8.6(17)	80(72)	9/15
BSqi	**1.1**(2)	**1.3**(2)	**1.1**(0.4)	43(0.1)	71(83)	11/15
BSrr	**1.1**(0.7)	**1.3**(0.6)	**1.1**(0.4)	27(8)	46(128)	10/15
CMA-CSA	**1.2**(2)	**2.2**(1)	**1.6**(1)	5.1(2)	5.1(3)	15/15
CMA-MSR	**1.7**(2)	**1.9**(2)	3.2(2)	7.1(2)	6.7(14)	15/15
CMA-TPA	**1.7**(2)	**2.5**(2)	3.1(2)	5.6(2)	**5.0**(7)	15/15
GP1-CMAES	**1.0**(1)	**1.7**(2)	**2.3**(2)	**4.2**(3)	**4.3**(3)	10/15
GP5-CMAES	**1.2**(0.3)	**2.5**(2)	**2.3**(1)	5.6(1)	**4.9**(10)	8/15
IPOPCMAv3p	**1.4**(1)	**2.6**(3)	**2.9**(3)	5.3(2)	5.9(6)	8/15
LHD-10xDef	**0.96**(1.0)	**1.6**(2)	4.2(3)	∞	∞ *150*	0/15
LHD-2xDefa	**1.1**(0.7)	**1.8**(2)	**1.6**(0.8)	4.7(3)	13(12)	1/15
RAND-2xDef	**1**(0.3)	**1.5**(0.9)	**1.6**(0.9)	4.2(5)	∞ *150*	0/15
RF1-CMAES	**1.0**(3)	**1.6**(2)	**2.4**(2)	5.1(2)	20(20)	3/15
RF5-CMAES	**1.2**(0.5)	10(34)	18(26)	65(104)	∞ *753*	0/15
Sifeg	**1.1**(0)	**1.3**(2)	**1.2**(0.4)	**2.9**(0.6)	20(18)	9/15
Sif	**1.1**(1)	**1.3**(0.9)	**1.4**(0.3)	5.8(13)	20(26)	9/15
Srr	**1.1**(0)	**1.3**(1.0)	**1.3**(0.3)	**3.4**(11)	19(33)	10/15

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
BSifeg	**1.7**(1)	**1.6**(0.7)	**1.4**(0.9)	89(513)	∞ *3e4*	0/15
BSif	**1.7**(2)	**1.6**(1)	**1.4**(0.5)	59(19)	∞ *3e4*	0/15
BSqi	**1.7**(2)	**1.6**(0.7)	**1.3**(0.3)	26(68)	∞ *3e4*	0/15
BSrr	**1.7**(3)	**1.6**(0.8)	**1.4**(0.7)	50(165)	∞ *2e4*	0/15
CMA-CSA	**3.0**(4)	**2.2**(3)	**1.8**(2)	3.7(2)	**3.4**(0.5)	15/15
CMA-MSR	**2.4**(3)	**1.8**(2)	**2.0**(1)	4.2(2)	3.9(0.6)	15/15
CMA-TPA	**1.8**(0.9)	**1.8**(2)	**2.6**(5)	3.5(2)	**3.9**(0.7)	15/15
GP1-CMAES	**0.75**(0.3)	**1.7**(1)	**1.5**(0.8)	**2.0**(0.9)	19(16)	3/15
GP5-CMAES	**2.0**(2)	**1.4**(2)	**1.2**(0.9)	**1.4**(0.4)	**3.2**(3)	11/15
IPOPCMAv3p	**1.6**(3)	**1.7**(2)	**1.9**(0.8)	3.4(2)	12(7)	5/15
LHD-10xDef	**1.6**(4)	**1.5**(0.6)	**2.9**(3)	3.1(0.7)	∞ *150*	0/15
LHD-2xDefa	**1.2**(0.8)	**1.1**(0.8)	**1.3**(0.8)	**1.3**(0.6)	∞ *150*	0/15
RAND-2xDef	**1.4**(0.9)	**0.99**(1)	**1.1**(0.9)	**1.5**(2)	∞ *150*	0/15
RF1-CMAES	**1.9**(2)	**1.6**(1)	**1.8**(2)	4.2(0.7)	∞ *751*	0/15
RF5-CMAES	**1.9**(2)	**2.0**(2)	**2.5**(5)	21(11)	∞ *760*	0/15
Sifeg	**1.7**(3)	**1.5**(1)	**1.3**(0.3)	169(270)	∞ *2e4*	0/15
Sif	**1.7**(1)	**1.5**(1)	**1.3**(0.4)	76(0.9)	∞ *2e4*	0/15
Srr	**1.7**(1)	**1.5**(0.9)	**1.3**(0.2)	58(45)	∞ *2e4*	0/15

Table 15: 03-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_{14} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f14	*1.0e+1*:2.2	*6.3e+0*:4.2	*2.5e+0*:10	*6.3e-2*:31	*2.5e-6*:160	15/15
BSifeg	**1.8**(1)	**2.1**(3)	**2.7**(1)	5.0(8)	∞ *3e4*	0/15
BSif	**1.8**(2)	**2.2**(1)	4.3(10)	5.5(7)	∞ *3e4*	0/15
BSqi	**1.8**(2)	**2.3**(2)	**2.4**(0.9)	4.2(6)	∞ *3e4*	0/15
BSrr	**1.8**(2)	**1.9**(2)	**2.5**(1)	5.6(9)	∞ *3e4*	0/15
CMA-CSA	3.8(3)	3.1(1)	**2.3**(2)	4.1(0.7)	**4.0**(0.5)	15/15
CMA-MSR	**2.5**(2)	**1.8**(2)	**2.4**(2)	5.3(2)	**4.4**(0.9)	15/15
CMA-TPA	4.4(5)	3.7(3)	3.4(2)	4.5(2)	**3.8**(0.5)	15/15
GP1-CMAES	3.9(5)	**2.7**(3)	**2.2**(1)	**2.7**(0.8)	∞ *751*	0/15
GP5-CMAES	3.3(4)	**2.3**(2)	**2.0**(1)	**1.9**(0.6)	68(40)	1/15
IPOPCMAv3p	**2.2**(3)	**2.3**(4)	**2.7**(2)	4.0(1)	23(38)	3/15
LHD-10xDef	**1.5**(2)	**1.4**(3)	**2.0**(2)	4.0(0.4)	∞ *150*	0/15
LHD-2xDefa	**1.9**(1)	**1.6**(1)	**1.6**(0.9)	**1.8**(0.9)	∞ *150*	0/15
RAND-2xDef	**2.3**(2)	**1.5**(0.7)	**1.5**(1)	**2.3**(1.0)	∞ *150*	0/15
RF1-CMAES	**2.9**(5)	**2.7**(4)	4.7(8)	16(30)	∞ *751*	0/15
RF5-CMAES	**1.9**(2)	**1.1**(1.0)	20(57)	47(43)	∞ *753*	0/15
Sifeg	**1.8**(2)	**1.7**(1)	**2.0**(1)	**2.6**(2)	∞ *3e4*	0/15
Sif	**1.8**(2)	**1.7**(2)	**2.0**(2)	**3.0**(3)	∞ *3e4*	0/15
Srr	**1.8**(2)	**1.7**(1)	**1.9**(1)	**1.9**(1)	∞ *3e4*	0/15

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
BSifeg	**1.7**(2)	**1.7**(0.9)	**2.6**(0.4)	43(44)	72(50)	11/15
BSif	**1.7**(1)	**1.7**(2)	**1.9**(4)	44(53)	59(104)	13/15
BSqi	**1.7**(2)	**1.7**(1)	3.4(9)	51(119)	64(117)	12/15
BSrr	**1.7**(2)	**1.7**(1)	3.3(10)	31(5)	56(96)	13/15
CMA-CSA	**3.0**(4)	**2.7**(3)	**2.1**(1)	**1.3**(0.5)	**0.85**(0.5)	15/15
CMA-MSR	**1.7**(1)	**1.6**(0.8)	**1.4**(1)	**1.4**(1)	3.7(3)	15/15
CMA-TPA	**1.9**(1)	**1.5**(2)	**2.2**(2)	**1.4**(0.8)	**0.91**(0.5)	15/15
GP1-CMAES	4.4(3)	**2.9**(2)	**2.3**(1)	**1.0**(0.2)	**1.9**(2)	11/15
GP5-CMAES	**2.8**(3)	**2.0**(1)	**1.7**(0.7)	**0.67**(0.8)	**1.6**(2)	13/15
IPOPCMAv3p	4.2(11)	**2.5**(3)	**1.8**(2)	**1.4**(0.4)	**1.1**(0.5)	14/15
LHD-10xDef	**1.9**(1)	**2.2**(2)	3.3(2)	**2.0**(1)	3.3(4)	3/15
LHD-2xDefa	**2.2**(2)	**1.3**(1)	**1.3**(1)	**0.68**(0.4)	**1.1**(0.7)	7/15
RAND-2xDef	**1.4**(0.9)	**0.99**(0.8)	**0.98**(0.5)	**0.74**(0.5)	**0.57**(0.3)	12/15
RF1-CMAES	**2.4**(3)	**2.3**(3)	**1.7**(0.7)	**0.79**(0.7)	**2.0**(6)	11/15
RF5-CMAES	**2.4**(2)	18(3)	13(2)	5.5(9)	11(14)	4/15
Sifeg	**1.7**(2)	**1.6**(1)	**0.97**(0.6)	17(0.3)	36(78)	13/15
Sif	**1.7**(1)	**1.6**(1)	**0.97**(0.4)	23(83)	45(26)	13/15
Srr	**1.7**(2)	**1.6**(1.0)	**0.95**(0.6)	11(0.3)	32(50)	15/15

Table 17: 03-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
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
BSifeg	**1.6**(1)	**1.5**(1)	**1.9**(1.0)	**1.6**(2)	24(7)	14/15
BSif	**1.6**(1)	**1.5**(1)	**2.7**(7)	**2.0**(1)	20(41)	15/15
BSqi	**1.6**(1)	**1.5**(1)	**1.8**(1)	**2.0**(2)	13(23)	15/15
BSrr	**1.6**(1)	**1.5**(1)	**1.9**(1)	**1.5**(1)	10(2)	15/15
CMA-CSA	**2.8**(3)	**1.5**(1)	4.4(5)	**1.7**(1)	**3.0**(5)	15/15
CMA-MSR	**1.7**(1)	3.3(3)	7.7(4)	6.7(11)	4.1(5)	15/15
CMA-TPA	**1.7**(2)	**1.8**(2)	3.4(7)	3.2(4)	**2.7**(4)	15/15
GP1-CMAES	**2.0**(2)	**1.6**(2)	**2.9**(2)	**1.3**(1)	**2.7**(6)	10/15
GP5-CMAES	**2.1**(2)	**1.2**(2)	**1.4**(2)	**0.78**(0.9)	**2.1**(4)	12/15
IPOPCMAv3p	**1.5**(1)	**1.7**(1)	**2.2**(1)	**1.6**(2)	**1.9**(2)	12/15
LHD-10xDef	**1.7**(1)	**0.76**(0.8)	**2.1**(4)	**0.99**(0.9)	**1.0**(1.0)	8/15
LHD-2xDefa	**1.3**(0)	**1.1**(0.7)	**2.6**(2)	**1.0**(1)	**0.53**(0.5)	11/15
RAND-2xDef	**1.9**(2)	**1.6**(1)	**1.8**(2)	**1.4**(0.7)	**1.0**(0.4)	8/15
RF1-CMAES	**2.5**(4)	**1.8**(2)	**2.8**(3)	**1.1**(1)	**2.6**(3)	10/15
RF5-CMAES	**2.6**(2)	**1.4**(2)	6.4(35)	**2.8**(0.4)	**2.6**(4)	10/15
Sifeg	**1.6**(1)	**1.9**(2)	**2.7**(2)	**1.1**(0.6)	6.3(21)	15/15
Sif	**1.6**(1)	**1.8**(2)	**2.6**(1)	**1.0**(0.7)	3.6(4)	15/15
Srr	**1.6**(1)	**1.9**(2)	3.0(1)	**1.2**(0.7)	4.2(8)	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
BSifeg	4.3(11)	3.6(1)	**1.4**(3)	30(105)	107(81)	10/15
BSif	3.3(11)	**2.9**(2)	**1.2**(0.2)	18(65)	114(177)	10/15
BSqi	3.1(2)	**3.0**(2)	**1.2**(1)	13(42)	82(80)	11/15
BSrr	4.1(2)	5.2(24)	**2.9**(0.6)	**2.7**(5)	112(70)	10/15
CMA-CSA	3.3(2)	**2.3**(0.4)	**1.3**(0.4)	**1.6**(1.0)	**0.88**(0.3)	15/15
CMA-MSR	3.7(3)	3.1(5)	**1.5**(1)	**1.9**(1)	**2.4**(3)	15/15
CMA-TPA	3.8(5)	4.5(2)	**1.8**(2)	**2.1**(1)	**1.0**(0.3)	15/15
GP1-CMAES	**1.8**(2)	**2.3**(2)	**1.2**(1)	**2.9**(6)	**1.3**(2)	13/15
GP5-CMAES	**2.0**(2)	3.2(5)	**2.7**(7)	3.9(3)	3.1(5)	11/15
IPOPCMAv3p	4.9(4)	5.4(4)	**2.8**(4)	**2.5**(1)	**1.1**(0.5)	15/15
LHD-10xDef	**2.9**(2)	**2.8**(4)	**1.6**(2)	**2.4**(1)	**2.2**(2)	5/15
LHD-2xDefa	**2.1**(2)	**2.4**(1)	**1.1**(0.7)	**1.2**(0.4)	5.7(13)	2/15
RAND-2xDef	3.3(3)	**2.6**(1)	**1.1**(0.8)	**1.0**(0.4)	3.8(4)	3/15
RF1-CMAES	3.3(4)	4.1(4)	**1.7**(2)	7.5(16)	5.7(9)	7/15
RF5-CMAES	22(0.6)	22(37)	11(18)	13(11)	13(24)	4/15
Sifeg	4.5(3)	3.8(7)	**1.9**(3)	23(3)	33(27)	14/15
Sif	4.4(19)	4.7(11)	**1.9**(3)	**2.5**(1)	55(76)	13/15
Srr	4.0(2)	3.7(2)	**1.6**(2)	**1.6**(2)	31(26)	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
BSifeg	**2.7**(2)	**1.9**(2)	**1.5**(2)	**1.6**(0.3)	334(420)	5/15
BSif	**2.7**(3)	**1.9**(1)	**1.5**(3)	**1.3**(2)	218(273)	7/15
BSqi	**2.7**(2)	**1.9**(2)	**1.5**(2)	**1.1**(3)	158(197)	9/15
BSrr	**2.7**(2)	**1.9**(2)	**1.3**(0.8)	17(60)	187(356)	7/15
CMA-CSA	3.9(4)	**2.7**(2)	**1.7**(2)	**1.4**(1)	**3.6**(1)	15/15
CMA-MSR	4.6(5)	**2.7**(4)	**1.8**(2)	**1.4**(1.0)	4.7(17)	15/15
CMA-TPA	6.1(9)	3.7(5)	**2.4**(1)	**1.7**(1)	**3.8**(2)	15/15
GP1-CMAES	4.2(4)	**2.2**(0.8)	**1.7**(2)	**1.3**(0.5)	3.8(7)	9/15
GP5-CMAES	5.8(4)	14(78)	14(21)	5.6(12)	**2.7**(3)	11/15
IPOPCMAv3p	3.3(3)	4.2(4)	**2.6**(1)	**1.7**(3)	5.4(4)	8/15
LHD-10xDef	**2.9**(4)	**1.6**(2)	**1.4**(2)	**2.1**(0.2)	12(15)	1/15
LHD-2xDefa	**2.6**(0.8)	**1.8**(1)	**1.2**(0.7)	**0.93**(0.6)	3.9(5)	3/15
RAND-2xDef	3.4(3)	**2.0**(2)	**1.3**(1)	**0.97**(1)	6.0(7)	2/15
RF1-CMAES	4.4(9)	3.3(4)	5.6(2)	7.7(19)	17(16)	3/15
RF5-CMAES	3.5(3)	**2.1**(0.8)	**1.1**(1)	5.1(5)	58(77)	1/15
Sifeg	**2.7**(3)	**2.1**(1)	**1.8**(3)	**1.5**(1)	139(166)	10/15
Sif	**2.7**(3)	**2.1**(1)	4.8(15)	3.7(0.9)	153(143)	9/15
Srr	**2.7**(3)	**2.1**(1)	**1.7**(1)	**1.2**(1)	94(119)	11/15

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
BSifeg	**12**(18)	**17**(46)	**35**(50)	**62**(61)	54(36)	5/15
BSif	**19**(18)	36(58)	66(179)	139(103)	42(51)	6/15
BSqi	19(10)	45(15)	62(56)	99(193)	**37**(64)	7/15
BSrr	23(24)	33(51)	78(127)	149(256)	115(55)	3/15
CMA-CSA	30(31)	39(48)	59(54)	**59**(45)	**12**(14)	15/15
CMA-MSR	69(83)	96(148)	274(546)	677(1164)	131(193)	12/15
CMA-TPA	28(15)	41(49)	**46**(68)	**66**(55)	**7.5**(4)	15/15
GP1-CMAES	43(31)	48(31)	**48**(21)	92(129)	∞ *753*	0/15
GP5-CMAES	68(48)	104(86)	104(117)	∞	∞ *762*	0/15
IPOPCMAv3p	28(47)	∞	∞	∞	∞ *751*	0/15
LHD-10xDef	∞	∞	∞	∞	∞ *150*	0/15
LHD-2xDefa	∞	∞	∞	∞	∞ *150*	0/15
RAND-2xDef	∞	∞	∞	∞	∞ *150*	0/15
RF1-CMAES	23(22)	24(13)	100(105)	93(87)	∞ *751*	0/15
RF5-CMAES	44(61)	∞	∞	∞	∞ *755*	0/15
Sifeg	22(25)	**19**(33)	70(135)	176(205)	170(346)	2/15
Sif	20(51)	36(61)	54(96)	169(173)	48(76)	6/15
Srr	**15**(23)	**21**(35)	62(95)	133(119)	48(14)	6/15

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
BSifeg	**2.4**(2)	**2.2**(2)	4.4(0.2)	5.3(6)	13(24)	15/15
BSif	**2.4**(2)	**2.2**(2)	4.0(9)	6.4(9)	17(13)	15/15
BSqi	**2.4**(2)	**2.2**(2)	**2.8**(5)	5.4(6)	11(16)	15/15
BSrr	**2.4**(2)	**2.2**(2)	**2.0**(1)	4.7(6)	14(27)	14/15
CMA-CSA	**2.0**(2)	**1.6**(2)	**2.2**(1.0)	11(13)	**4.5**(2)	15/15
CMA-MSR	**1.7**(0.5)	**1.7**(2)	**2.6**(2)	20(35)	13(16)	15/15
CMA-TPA	**1.7**(1)	**2.4**(2)	**2.9**(1)	10(10)	7.6(10)	15/15
GP1-CMAES	**2.0**(3)	**1.8**(0.6)	**2.4**(2)	8.6(15)	**3.7**(2)	6/15
GP5-CMAES	**1.2**(0.6)	**1.2**(2)	**2.0**(1)	5.4(5)	**2.0**(3)	9/15
IPOPCMAv3p	**1.4**(1)	**1.5**(1)	**2.8**(2)	12(18)	4.5(9)	5/15
LHD-10xDef	**0.79**(0.5)	**1.1**(2)	4.3(3)	∞	∞ *150*	0/15
LHD-2xDefa	**1.4**(1)	**1.3**(0.9)	**1.8**(1)	17(17)	∞ *150*	0/15
RAND-2xDef	**1.2**(2)	**1.4**(2)	**1.9**(1)	13(15)	∞ *150*	0/15
RF1-CMAES	**1.6**(2)	**1.7**(0.9)	6.4(13)	17(9)	5.9(6)	4/15
RF5-CMAES	**1.8**(2)	**2.0**(2)	33(24)	63(42)	∞ *760*	0/15
Sifeg	**2.5**(2)	**2.3**(2)	**2.0**(1.0)	**3.3**(4)	7.3(9)	15/15
Sif	**2.5**(2)	**2.3**(2)	**2.2**(3)	**3.8**(3)	10(25)	15/15
Srr	**2.5**(2)	**2.3**(1.0)	**1.9**(0.7)	**2.7**(0.9)	9.0(26)	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
BSifeg	**2.8**(3)	**1.6**(1)	152(524)	117(70)	84(90)	12/15
BSif	**2.8**(3)	**1.6**(1)	152(7)	249(439)	114(154)	10/15
BSqi	**2.8**(2)	**1.6**(1)	152(521)	167(259)	150(139)	9/15
BSrr	**2.8**(3)	**1.6**(1)	150(0.9)	113(46)	130(135)	10/15
CMA-CSA	**2.2**(3)	**1.3**(1)	5.8(17)	5.1(1)	6.9(3)	15/15
CMA-MSR	**2.9**(5)	**2.1**(2)	**2.8**(5)	6.7(1.0)	5.9(8)	15/15
CMA-TPA	**2.7**(3)	**1.6**(2)	**2.3**(4)	**1.6**(2)	**1.9**(3)	15/15
GP1-CMAES	**1.1**(2)	**0.93**(0.8)	4.9(28)	5.3(19)	7.1(8)	6/15
GP5-CMAES	**2.2**(2)	**1.4**(2)	**1.3**(1)	**1.2**(0.5)	**2.4**(2)	11/15
IPOPCMAv3p	**2.5**(3)	**1.9**(1)	**1.4**(2)	**1.6**(1)	**2.5**(2)	11/15
LHD-10xDef	**1.2**(0.8)	**1.7**(2)	**1.3**(2)	**1.4**(0.9)	**0.66**(0.2)	14/15
LHD-2xDefa	**2.2**(2)	**1.3**(0.8)	**1.6**(2)	**1.1**(0.7)	**0.89**(1)	10/15
RAND-2xDef	**1.5**(0.6)	**1.2**(2)	**1.1**(0.8)	**1.1**(1)	**0.80**(0.8)	10/15
RF1-CMAES	**2.3**(2)	**2.1**(2)	**1.9**(1)	**2.5**(9)	4.4(8)	8/15
RF5-CMAES	**1.9**(2)	**1.9**(1)	5.8(0.7)	5.2(6)	7.3(11)	7/15
Sifeg	**2.8**(3)	**1.9**(2)	151(522)	92(561)	95(145)	11/15
Sif	**2.8**(3)	**2.1**(2)	152(523)	101(1)	104(204)	10/15
Srr	**2.8**(3)	**1.9**(2)	151(522)	62(183)	142(147)	9/15

Table 23: 03-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_{22} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f22	*4.0e+1*:2.9	*2.5e+1*:5.2	*1.0e+1*:18	*6.3e+0*:33	*1.0e+0*:170	5/5
BSifeg	**3.0**(2)	**2.1**(3)	7.6(2)	79(214)	87(138)	11/15
BSif	**2.6**(2)	**2.0**(2)	10(30)	171(237)	219(410)	7/15
BSqi	3.1(3)	**2.2**(3)	4.8(6)	85(265)	130(261)	10/15
BSrr	3.2(6)	**2.2**(5)	3.9(9)	87(180)	146(210)	9/15
CMA-CSA	**1.8**(2)	**1.7**(2)	**1.4**(1)	3.0(7)	11(20)	15/15
CMA-MSR	**2.0**(1)	**2.3**(3)	**2.0**(2)	6.8(1)	5.9(25)	15/15
CMA-TPA	**1.8**(2)	**2.0**(4)	**1.8**(3)	**1.5**(2)	19(50)	15/15
GP1-CMAES	**1.4**(0.9)	**2.0**(1)	**1.7**(2)	5.2(12)	3.8(9)	9/15
GP5-CMAES	**1.5**(1)	**1.5**(3)	4.2(3)	6.4(6)	10(15)	5/15
IPOPCMAv3p	**2.7**(3)	**2.6**(3)	**2.1**(1)	**1.8**(0.9)	10(24)	5/15
LHD-10xDef	**2.8**(2)	**2.1**(1)	**1.7**(1)	**1.9**(1)	**0.88**(0.9)	11/15
LHD-2xDefa	**1.4**(1)	**1.7**(0.9)	**1.4**(0.7)	**1.4**(1.0)	**2.2**(3)	5/15
RAND-2xDef	**1.6**(0.6)	**2.2**(1)	**0.97**(0.6)	**1.3**(1)	**1.1**(1.0)	9/15
RF1-CMAES	**1.1**(1)	**0.96**(0.5)	**1.7**(1)	7.1(24)	5.5(8)	7/15
RF5-CMAES	12(1)	14(0.8)	6.9(19)	10(27)	11(11)	5/15
Sifeg	**2.3**(3)	**1.8**(2)	**1.8**(3)	31(3)	61(62)	13/15
Sif	**2.3**(3)	**1.8**(3)	3.2(8)	45(12)	67(162)	12/15
Srr	**2.3**(2)	**1.8**(1)	**1.5**(4)	**1.6**(2)	66(172)	12/15

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
BSifeg	3.8(2)	**1.4**(1)	**1.2**(1)	**2.2**(4)	**1.5**(2)	15/15
BSif	3.8(2)	**1.4**(1)	**1.6**(2)	**2.7**(3)	**2.0**(2)	15/15
BSqi	3.7(3)	**1.2**(1)	**1.1**(2)	**1.8**(3)	**1.4**(1)	15/15
BSrr	3.9(3)	**1.4**(1)	**1.4**(1)	**1.6**(1)	**1.3**(1)	15/15
CMA-CSA	**3.3**(6)	**1.5**(2)	**1.4**(2)	**2.8**(2)	8.5(9)	15/15
CMA-MSR	**2.6**(2)	**2.2**(1)	3.4(2)	5.6(4)	4.8(7)	15/15
CMA-TPA	4.2(6)	3.3(6)	7.6(3)	6.5(3)	4.2(2)	15/15
GP1-CMAES	**3.3**(4)	**1.7**(0.7)	**2.3**(2)	3.5(4)	4.3(6)	10/15
GP5-CMAES	5.7(7)	**1.7**(2)	**2.1**(4)	**2.1**(2)	**1.3**(1)	14/15
IPOPCMAv3p	4.3(7)	**1.5**(2)	**1.7**(2)	3.1(4)	4.7(4)	9/15
LHD-10xDef	6.4(3)	**2.0**(2)	**2.3**(1)	8.5(9)	11(20)	1/15
LHD-2xDefa	4.0(4)	**1.3**(0.8)	3.6(5)	3.5(6)	∞ *150*	0/15
RAND-2xDef	4.5(4)	**2.7**(4)	**2.9**(3)	7.7(8)	11(14)	1/15
RF1-CMAES	5.1(3)	**1.7**(1)	**1.9**(1)	**2.5**(4)	6.9(5)	6/15
RF5-CMAES	3.5(2)	**1.4**(1)	**1.6**(1)	5.4(7)	7.9(6)	6/15
Sifeg	3.7(4)	**1.9**(2)	**2.3**(2)	3.9(2)	**2.9**(2)	15/15
Sif	3.7(2)	**1.7**(3)	**2.3**(2)	3.8(3)	**2.8**(1)	15/15
Srr	3.7(4)	**1.7**(3)	**2.2**(2)	3.9(2)	**2.3**(1)	15/15

Table 25: 03-D, running time excess ERT/$\mathrm{ERT}_{\mathrm{best}\ 2009}$ on f_{24} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\mathrm{best}\ 2009}$ (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
f24	*4.0e+1*:4.6	*2.5e+1*:13	*1.6e+1*:47	*1.6e+1*:47	*6.3e+0*:382	15/15
BSifeg	**1.5**(1)	**2.2**(0.5)	**2.5**(4)	**2.5**(4)	17(35)	14/15
BSif	**1.6**(0.5)	**2.4**(1)	**2.3**(1)	**2.3**(2)	24(49)	12/15
BSqi	**1.6**(1)	**1.7**(1)	**2.6**(3)	**2.6**(1)	13(29)	14/15
BSrr	**1.5**(1)	**1.7**(0.7)	**2.5**(5)	**2.5**(4)	8.1(21)	15/15
CMA-CSA	**1.8**(2)	**2.1**(2)	**1.5**(0.5)	**1.5**(0.9)	**2.0**(3)	15/15
CMA-MSR	**1.7**(0.8)	**2.6**(2)	**1.6**(0.9)	**1.6**(2)	3.4(4)	15/15
CMA-TPA	**1.4**(1)	**2.0**(1)	**1.3**(2)	**1.3**(0.3)	**2.0**(0.7)	15/15
GP1-CMAES	**2.2**(2)	**2.2**(1)	**1.7**(0.8)	**1.7**(0.9)	**2.0**(3)	9/15
GP5-CMAES	**1.5**(2)	**2.4**(2)	**1.2**(1)	**1.2**(2)	**2.3**(4)	8/15
IPOPCMAv3p	**2.0**(2)	**2.3**(3)	**1.8**(1)	**1.8**(1)	**1.5**(0.9)	12/15
LHD-10xDef	**2.1**(2)	3.3(3)	3.0(4)	3.0(4)	5.8(5)	1/15
LHD-2xDefa	**1.2**(2)	**2.2**(3)	**1.7**(0.7)	**1.7**(2)	5.6(9)	1/15
RAND-2xDef	**0.93**(1)	**2.3**(1)	**2.2**(3)	**2.2**(0.9)	5.7(8)	1/15
RF1-CMAES	**1.9**(4)	**2.4**(2)	**1.3**(0.8)	**1.3**(0.7)	**1.6**(1)	10/15
RF5-CMAES	**1.2**(1)	**1.3**(0.8)	5.2(5)	5.2(11)	**2.3**(3)	9/15
Sifeg	**1.4**(2)	**1.7**(0.6)	**1.3**(0.9)	**1.3**(1)	6.1(34)	14/15
Sif	**1.4**(1)	**1.6**(0.6)	**1.4**(1)	**1.4**(1)	6.2(3)	15/15
Srr	**1.4**(2)	**1.8**(1.0)	**1.3**(0.6)	**1.3**(0.6)	11(34)	14/15

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.