Comparison Tables: BBOB 2015 Testbed in 10-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: 10-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_1 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f1	*4.0e+1*:8.0	*2.5e+1*:16	*1.0e-8*:23	*1.0e-8*:23	*1.0e-8*:23	15/15
BSifeg	**3.2**(1.0)	**2.0**(0.2)	**2.4**(0.3)	**2.4**(0.3)	**2.4**(0.2)	15/15
BSif	**3.2**(0.9)	**2.0**(0.2)	**2.4**(0.2)	**2.4**(0.2)	**2.4**(0.2)	15/15
BSqi	**3.2**(0.8)	**2.0**(0.2)	**2.4**(0.2)	**2.4**(0.3)	**2.4**(0.2)	15/15
BSrr	**3.2**(0.7)	**2.0**(0.1)	**2.4**(0.2)	**2.4**(0.2)	**2.4**(0.2)	15/15
CMA-CSA	5.2(4)	3.7(1)	63(5)	63(7)	63(5)	15/15
CMA-MSR	4.4(2)	5.2(2)	87(5)	87(10)	87(5)	15/15
CMA-TPA	7.5(6)	5.5(3)	51(5)	51(4)	51(6)	15/15
GP1-CMAES	3.6(3)	**2.6**(0.7)	43(7)	43(8)	43(8)	15/15
GP5-CMAES	**3.0**(2)	**1.9**(1.0)	404(275)	404(741)	404(248)	4/15
IPOPCMAv3p	5.0(4)	3.9(2)	65(4)	65(3)	65(2)	15/15
LHD-10xDef	9.4(9)	12(5)	∞	∞	∞ *500*	0/15
LHD-2xDefa	5.0(2)	**2.9**(1)	∞	∞	∞ *500*	0/15
RAND-2xDef	4.2(2)	**2.9**(1.0)	∞	∞	∞ *500*	0/15
RF1-CMAES	4.3(2)	3.2(2)	522(508)	522(673)	522(761)	3/15
RF5-CMAES	3.9(2)	**2.7**(1)	∞	∞	∞ *2514*	0/15
Sifeg	**3.2**(1)	**2.0**(0.2)	12(2)	12(2)	12(1)	15/15
Sif	**3.2**(1)	**2.0**(0.2)	11(2)	11(2)	11(2)	15/15
Srr	**3.2**(0.9)	**2.0**(0.5)	11(2)	11(2)	11(2)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f2	*2.5e+6*:5.6	*1.0e+6*:17	*1.0e+5*:33	*2.5e+3*:118	*1.0e-8*:196	15/15
BSifeg	5.4(4)	**2.4**(0.9)	**1.5**(0.3)	**0.54**(0.2)	**1.2**(0.2)	15/15
BSif	5.4(2)	**2.4**(1)	**1.5**(0.1)	**0.56**(0.1)	**1.2**(0.3)	15/15
BSqi	5.4(4)	**2.4**(0.9)	**1.5**(0.3)	**0.51**(0.1)	**1.2**(0.2)	15/15
BSrr	5.4(4)	**2.4**(1)	**1.5**(0.2)	**0.54**(0.2)	**1.2**(0.2)	15/15
CMA-CSA	**2.3**(2)	**2.1**(3)	7.4(3)	8.5(4)	23(1.0)	15/15
CMA-MSR	3.2(2)	**3.0**(2)	7.3(2)	8.7(4)	26(2)	15/15
CMA-TPA	**1.8**(2)	**1.8**(1)	5.8(3)	7.4(2)	23(1)	15/15
GP1-CMAES	**1.4**(1)	**1.1**(0.6)	4.8(3)	7.5(2)	∞ *2502*	0/15
GP5-CMAES	**2.8**(3)	**1.9**(2)	3.5(1)	3.2(0.8)	∞ *2502*	0/15
IPOPCMAv3p	**2.5**(3)	**2.0**(3)	6.0(3)	10(3)	∞ *2502*	0/15
LHD-10xDef	**1.5**(0.8)	**1.2**(0.8)	9.0(3)	∞	∞ *500*	0/15
LHD-2xDefa	**1.5**(2)	**1.5**(1)	5.5(2)	∞	∞ *500*	0/15
RAND-2xDef	**1.8**(2)	**1.8**(1)	5.1(4)	∞	∞ *500*	0/15
RF1-CMAES	**2.3**(2)	**2.4**(2)	7.7(7)	58(30)	∞ *2502*	0/15
RF5-CMAES	**2.4**(3)	**2.3**(1)	34(35)	317(661)	∞ *2504*	0/15
Sifeg	5.5(4)	**2.6**(0.5)	**1.8**(0.3)	**0.69**(0.1)	**1.7**(0.2)	15/15
Sif	5.5(4)	**2.6**(0.9)	**1.8**(0.3)	**0.73**(0.1)	**1.7**(0.2)	15/15
Srr	5.5(4)	**2.6**(0.9)	**1.8**(0.2)	**0.68**(0.0)	**1.7**(0.2)	15/15

Table 4: 10-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f3	*4.0e+2*:8.2	*1.6e+2*:37	*1.0e+2*:69	*6.3e+1*:147	*2.5e+1*:1129	15/15
BSifeg	**2.9**(0.5)	**1.1**(0.1)	**0.69**(0.1)	**0.41**(0.1)	**0.11**(0.0)	15/15
BSif	**2.9**(0.9)	**1.1**(0.1)	**0.69**(0.1)	**0.43**(0.2)	**0.12**(0.0)	15/15
BSqi	**2.9**(0.9)	**1.1**(0.1)	**0.69**(0.1)	**0.41**(0.1)	**0.10**(0.0)	15/15
BSrr	**2.9**(0.8)	**1.1**(0.0)	**0.69**(0.1)	**0.42**(0.1)	**0.09**(0.0)	15/15
CMA-CSA	**2.6**(2)	**2.8**(2)	3.1(1)	**2.8**(0.7)	**1.4**(1)	15/15
CMA-MSR	**2.3**(2)	3.3(1)	3.1(1)	**2.8**(0.6)	**1.4**(2)	15/15
CMA-TPA	3.2(4)	3.3(2)	**2.8**(0.7)	**2.4**(1)	**1.1**(1)	15/15
GP1-CMAES	**1.8**(2)	**2.2**(0.8)	**2.2**(0.7)	**2.5**(1)	**1.8**(3)	11/15
GP5-CMAES	**1.9**(2)	**1.7**(0.6)	**1.6**(0.6)	6.6(7)	32(68)	1/15
IPOPCMAv3p	**2.8**(3)	**2.7**(1)	3.1(2)	3.4(1)	**1.3**(0.3)	13/15
LHD-10xDef	**2.3**(3)	5.7(0.1)	3.7(0.4)	**2.4**(0.2)	**2.1**(2)	3/15
LHD-2xDefa	**1.8**(2)	**2.2**(1)	**2.5**(1)	3.4(4)	∞ *500*	0/15
RAND-2xDef	**1.1**(0.7)	**1.8**(0.4)	**1.7**(0.3)	**1.8**(0.7)	**2.0**(2)	3/15
RF1-CMAES	**1.7**(2)	**1.9**(1)	**2.1**(0.7)	**2.1**(0.7)	**1.6**(2)	10/15
RF5-CMAES	**2.5**(1)	4.6(0.7)	5.4(6)	8.5(8)	31(34)	1/15
Sifeg	**2.9**(0.9)	**1.1**(0.1)	**0.74**(0.1)	**0.46**(0.1)	**0.14**(0.0)	15/15
Sif	**2.9**(0.5)	**1.1**(0.1)	**0.74**(0.1)	**0.48**(0.1)	**0.14**(0.0)	15/15
Srr	**2.9**(0.6)	**1.1**(0.1)	**0.74**(0.1)	**0.46**(0.1)	**0.14**(0.0)	15/15

Table 5: 10-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ 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 $\mathrm{ERT}_{\text{best 2009}}$ (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
f4	*2.5e+2*:21	*1.6e+2*:59	*1.6e+2*:59	*6.3e+1*:139	*4.0e+1*:854	15/15
BSifeg	**1.6**(1)	**0.86**(0.4)	**0.86**(0.4)	**0.89**(0.3)	**0.20**(0.1)	15/15
BSif	**1.6**(1)	**0.86**(0.2)	**0.86**(0.2)	**0.97**(0.2)	**0.20**(0.0)	15/15
BSqi	**1.6**(0.9)	**0.86**(0.3)	**0.86**(0.1)	**0.92**(0.5)	**0.23**(0.1)	15/15
BSrr	**1.6**(0.6)	**0.86**(0.4)	**0.86**(0.5)	**0.85**(0.2)	**0.20**(0.1)	15/15
CMA-CSA	4.6(2)	**2.9**(1)	**2.9**(1)	4.3(2)	**1.3**(3)	15/15
CMA-MSR	5.3(3)	3.4(1)	3.4(0.9)	5.2(0.4)	**2.2**(1)	15/15
CMA-TPA	5.3(4)	**3.0**(1)	**3.0**(0.9)	5.4(0.7)	**2.3**(2)	15/15
GP1-CMAES	4.3(2)	3.5(1)	3.5(1)	11(6)	4.2(4)	8/15
GP5-CMAES	3.8(3)	3.1(3)	3.1(3)	∞	∞ *2516*	0/15
IPOPCMAv3p	4.4(4)	3.2(1)	3.2(1)	5.2(3)	**1.6**(0.3)	13/15
LHD-10xDef	12(4)	8.2(7)	8.2(6)	∞	∞ *500*	0/15
LHD-2xDefa	4.4(3)	3.8(2)	3.8(1)	53(72)	∞ *500*	0/15
RAND-2xDef	4.2(2)	3.2(2)	3.2(1)	12(29)	8.7(11)	1/15
RF1-CMAES	4.8(2)	3.2(2)	3.2(2)	13(14)	42(111)	1/15
RF5-CMAES	17(46)	24(17)	24(19)	267(238)	∞ *2504*	0/15
Sifeg	**1.7**(0.8)	**0.94**(0.3)	**0.94**(0.3)	**0.82**(0.2)	**0.17**(0.0)	15/15
Sif	**1.7**(1)	**0.94**(0.2)	**0.94**(0.4)	**0.82**(0.3)	**0.16**(0.0)	15/15
Srr	**1.7**(0.4)	**0.95**(0.2)	**0.95**(0.2)	**0.79**(0.1)	**0.16**(0.0)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f5	*1.0e+2*:16	*6.3e+1*:19	*1.0e-8*:20	*1.0e-8*:20	*1.0e-8*:20	15/15
BSifeg	**1.5**(0.2)	**1.4**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
BSif	**1.5**(0.2)	**1.4**(0.2)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
BSqi	**1.5**(0.2)	**1.4**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
BSrr	**1.5**(0.2)	**1.4**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
CMA-CSA	**1.8**(2)	**2.4**(1)	6.1(1)	6.1(2)	6.1(1)	15/15
CMA-MSR	**1.6**(1)	**2.0**(1)	5.4(2)	5.4(1)	5.4(2)	15/15
CMA-TPA	**1.3**(0.8)	**2.1**(1)	5.0(2)	5.0(2)	5.0(1)	15/15
GP1-CMAES	**1.4**(0.8)	**1.9**(0.3)	42(26)	42(39)	42(30)	15/15
GP5-CMAES	**1.3**(0.9)	**1.6**(0.5)	5.4(4)	5.4(3)	5.4(2)	15/15
IPOPCMAv3p	**1.6**(1)	**2.8**(0.6)	29(11)	29(9)	29(6)	15/15
LHD-10xDef	4.5(6)	11(1.0)	12(0.3)	12(0.4)	12(0.2)	15/15
LHD-2xDefa	**2.4**(0.8)	**2.3**(0.3)	**3.0**(0.1)	**3.0**(0.2)	**3.0**(0.2)	15/15
RAND-2xDef	**2.6**(0.2)	**2.4**(0.1)	3.1(0.3)	3.1(0.2)	3.1(0.2)	15/15
RF1-CMAES	**1.7**(0.9)	**2.4**(1)	35(28)	35(19)	35(17)	15/15
RF5-CMAES	**1.4**(1)	**2.2**(0.9)	120(198)	120(86)	120(81)	10/15
Sifeg	**1.5**(0.3)	**1.4**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
Sif	**1.5**(0.2)	**1.4**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
Srr	**1.5**(0.2)	**1.4**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15

Table 7: 10-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_6 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f6	*1.6e+5*:7.0	*6.3e+4*:16	*4.0e+2*:36	*1.0e+2*:102	*4.0e+0*:504	15/15
BSifeg	**1.6**(2)	**1.4**(1)	**1.2**(0.8)	38(23)	1260(1610)	2/15
BSif	**1.6**(2)	**1.4**(1.0)	**1.4**(0.4)	64(103)	862(1268)	3/15
BSqi	**1.6**(1)	**1.4**(0.6)	**1.2**(0.8)	30(86)	255(280)	7/15
BSrr	**1.6**(1)	**1.4**(0.9)	**1.2**(0.9)	42(24)	694(177)	3/15
CMA-CSA	**1.6**(1)	**1.5**(1)	**1.8**(2)	3.6(1)	**1.8**(0.3)	15/15
CMA-MSR	**2.7**(4)	**1.8**(3)	**2.2**(1)	**2.8**(0.7)	**1.6**(0.4)	15/15
CMA-TPA	**2.4**(3)	**2.2**(1)	4.8(3)	3.8(1)	**1.8**(0.3)	15/15
GP1-CMAES	**2.6**(2)	**1.8**(1)	**2.2**(1)	**2.1**(0.5)	4.8(4)	11/15
GP5-CMAES	**2.4**(3)	**1.5**(1)	**1.3**(0.3)	5.7(8)	∞ *2516*	0/15
IPOPCMAv3p	**2.7**(2)	**1.9**(4)	**2.8**(3)	3.0(2)	**1.7**(0.6)	15/15
LHD-10xDef	**1.9**(3)	4.1(2)	5.8(0.3)	7.2(8)	∞ *500*	0/15
LHD-2xDefa	**2.1**(2)	**1.8**(1)	**1.5**(0.6)	6.3(9)	∞ *500*	0/15
RAND-2xDef	**2.0**(3)	**1.8**(1)	**1.5**(0.2)	6.2(5)	∞ *500*	0/15
RF1-CMAES	**1.7**(2)	**1.4**(1)	**1.9**(0.9)	**2.7**(2)	71(106)	1/15
RF5-CMAES	**2.4**(2)	**1.7**(3)	5.9(3)	39(47)	∞ *2518*	0/15
Sifeg	**1.6**(2)	**1.4**(1)	**1.4**(0.3)	5.9(7)	118(49)	11/15
Sif	**1.6**(2)	**1.4**(0.9)	**1.4**(0.2)	8.6(11)	300(388)	7/15
Srr	**1.6**(1)	**1.4**(1)	**1.3**(0.3)	8.1(5)	110(53)	11/15

Table 8: 10-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ 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	*2.5e+2*:9.2	*1.6e+2*:18	*1.0e+2*:33	*1.0e+1*:172	*4.0e+0*:678	15/15
BSifeg	**2.2**(2)	**1.5**(0.9)	14(32)	1100(1642)	2095(2465)	1/15
BSif	**2.2**(2)	**1.5**(0.9)	14(16)	882(1047)	2193(1494)	1/15
BSqi	**2.2**(2)	**1.5**(1)	14(16)	931(1437)	1037(898)	2/15
BSrr	**2.2**(2)	**1.5**(0.8)	14(31)	894(1247)	∞ *1e5*	0/15
CMA-CSA	4.2(5)	3.3(2)	**2.9**(1)	**2.3**(0.8)	**1.1**(0.8)	15/15
CMA-MSR	3.5(2)	3.1(2)	3.2(1)	**1.9**(0.5)	**0.83**(0.9)	15/15
CMA-TPA	**2.8**(2)	**2.8**(2)	**2.7**(2)	**1.7**(0.6)	**1.2**(1.0)	15/15
GP1-CMAES	3.2(2)	**2.3**(1)	**2.1**(0.9)	**1.6**(0.7)	**1.1**(1)	15/15
GP5-CMAES	**2.3**(2)	**1.7**(1)	**1.4**(0.6)	**1.0**(0.3)*	**0.79**(0.4)	15/15
IPOPCMAv3p	**2.8**(3)	**3.0**(3)	**2.5**(1)	**2.6**(3)	**0.85**(1)	15/15
LHD-10xDef	**2.4**(3)	4.5(4)	5.4(3)	10(5)	∞ *500*	0/15
LHD-2xDefa	**2.3**(3)	**2.4**(2)	**2.3**(0.3)	43(20)	∞ *500*	0/15
RAND-2xDef	3.2(3)	3.0(1)	**2.6**(1)	13(9)	∞ *500*	0/15
RF1-CMAES	4.2(3)	3.6(3)	**3.0**(2)	13(26)	12(12)	4/15
RF5-CMAES	3.3(2)	**2.9**(2)	3.2(5)	31(47)	28(21)	2/15
Sifeg	**2.2**(2)	**1.6**(0.9)	5.5(0.3)	281(354)	363(508)	5/15
Sif	**2.2**(2)	**1.6**(0.9)	5.5(17)	166(150)	444(486)	4/15
Srr	**2.2**(2)	**1.6**(1)	5.5(0.7)	217(540)	220(211)	7/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f8	*1.6e+4*:15	*1.0e+4*:22	*1.6e+3*:34	*2.5e+2*:103	*4.0e+0*:727	15/15
BSifeg	**1.9**(0.1)	**1.5**(0.1)	**1.2**(0.1)	**1.3**(0.4)	51(29)	13/15
BSif	**1.9**(0.9)	**1.5**(0.9)	**1.2**(0.1)	**0.93**(0.1)	123(268)	10/15
BSqi	**1.9**(0.9)	**1.5**(0.5)	**1.2**(0.1)	**0.99**(2)	30(40)	15/15
BSrr	**1.9**(1)	**1.5**(0.4)	**1.2**(0.1)	**0.68**(0.4)	35(44)	13/15
CMA-CSA	4.2(5)	3.6(4)	5.0(3)	3.6(2)	**4.9**(4)	15/15
CMA-MSR	4.4(3)	3.9(3)	6.0(2)	3.3(1)	**5.3**(12)	15/15
CMA-TPA	3.8(4)	3.3(2)	4.0(1)	**2.2**(0.8)	5.4(5)	15/15
GP1-CMAES	**2.9**(2)	**2.5**(1)	3.1(0.4)	**2.3**(1)	12(16)	4/15
GP5-CMAES	**2.3**(0.9)	**1.7**(0.7)	**2.2**(0.7)	**1.8**(0.5)	∞ *2516*	0/15
IPOPCMAv3p	3.2(3)	**3.0**(3)	4.4(2)	**2.5**(0.6)	**4.1**(2)	11/15
LHD-10xDef	8.1(7)	7.1(4)	6.8(0.4)	3.2(0.5)	∞ *500*	0/15
LHD-2xDefa	3.3(0.3)	**2.3**(0.4)	**2.5**(0.4)	**2.5**(0.6)	∞ *500*	0/15
RAND-2xDef	3.3(0.2)	**2.5**(0.5)	**2.6**(0.9)	**2.4**(0.6)	∞ *500*	0/15
RF1-CMAES	4.1(2)	3.4(1)	4.3(2)	3.3(5)	∞ *2502*	0/15
RF5-CMAES	**2.9**(2)	**2.4**(1)	10(1)	15(18)	∞ *2514*	0/15
Sifeg	**1.9**(1)	**1.5**(0.1)	**1.3**(0.1)	**0.66**(0.2)	26(43)	15/15
Sif	**1.9**(1)	**1.5**(0.5)	**1.3**(0.1)	**0.68**(0.3)	31(66)	14/15
Srr	**1.9**(1)	**1.5**(0.4)	**1.3**(0.1)	**0.65**(0.1)	39(125)	13/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f9	*4.0e+1*:125	*2.5e+1*:148	*1.6e+1*:180	*1.0e+1*:200	*1.6e+0*:563	15/15
BSifeg	59(140)	53(131)	46(17)	47(169)	1100(1053)	2/15
BSif	239(371)	267(723)	239(393)	247(362)	2383(4152)	1/15
BSqi	47(120)	43(24)	38(81)	37(140)	1041(1197)	2/15
BSrr	47(38)	43(10)	37(24)	37(19)	1006(1142)	2/15
CMA-CSA	**3.5**(2)	**3.4**(2)	**3.1**(0.4)	**3.3**(0.5)	**5.2**(1)	15/15
CMA-MSR	4.5(2)	4.2(1.0)	3.8(1)	4.0(0.5)	**5.8**(4)	15/15
CMA-TPA	4.6(7)	**4.1**(4)	**3.6**(8)	**3.7**(2)	**5.0**(3)	15/15
GP1-CMAES	**4.3**(4)	4.2(0.6)	3.9(2)	4.3(1.0)	66(29)	1/15
GP5-CMAES	28(11)	71(76)	93(137)	84(139)	∞ *2526*	0/15
IPOPCMAv3p	**3.8**(2)	**3.5**(2)	**3.2**(0.4)	**3.5**(2)	33(101)	2/15
LHD-10xDef	19(16)	50(54)	42(31)	∞	∞ *500*	0/15
LHD-2xDefa	11(9)	12(7)	∞	∞	∞ *500*	0/15
RAND-2xDef	14(15)	25(54)	∞	∞	∞ *500*	0/15
RF1-CMAES	14(13)	15(19)	13(14)	16(16)	∞ *2502*	0/15
RF5-CMAES	∞	∞	∞	∞	∞ *2504*	0/15
Sifeg	69(179)	59(151)	49(126)	46(114)	2334(2673)	1/15
Sif	52(180)	45(151)	38(125)	37(3)	1159(1546)	2/15
Srr	33(53)	28(55)	24(47)	22(4)	2211(2228)	1/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f10	*2.5e+6*:6.0	*1.0e+6*:21	*4.0e+5*:38	*2.5e+4*:104	*6.3e+2*:512	15/15
BSifeg	**1.9**(3)	**1.2**(1)	7.7(50)	481(363)	∞ *7e4*	0/15
BSif	**1.9**(2)	**1.2**(0.9)	7.5(25)	413(1203)	∞ *7e4*	0/15
BSqi	**1.9**(2)	**1.2**(1)	4.5(26)	438(558)	∞ *8e4*	0/15
BSrr	**1.9**(2)	**1.2**(1.0)	3.6(0.7)	334(281)	∞ *5e4*	0/15
CMA-CSA	**2.4**(2)	**1.1**(0.7)	**2.5**(3)	5.0(1)	**2.8**(0.6)	15/15
CMA-MSR	**2.8**(2)	**1.6**(1)	**2.2**(1)	3.5(0.9)	**2.7**(0.3)	15/15
CMA-TPA	**1.2**(2)	**1.4**(1)	**1.9**(1)	**3.5**(1.0)	**2.4**(0.7)	15/15
GP1-CMAES	**2.0**(2)	**1.2**(2)	**1.2**(0.9)	**2.9**(2)	**2.2**(0.6)	15/15
GP5-CMAES	3.2(3)	**1.5**(0.9)	**1.2**(0.9)	**1.8**(1)	**1.0**(0.5)	15/15
IPOPCMAv3p	**1.9**(1)	**1.2**(2)	**2.3**(2)	4.1(2)	3.1(0.7)	15/15
LHD-10xDef	**2.4**(4)	**1.6**(2)	3.7(2)	11(7)	∞ *500*	0/15
LHD-2xDefa	**2.1**(2)	**1.5**(1)	**1.9**(0.9)	6.6(7)	∞ *500*	0/15
RAND-2xDef	**1.5**(1)	**1.1**(1)	**1.4**(1)	12(22)	∞ *500*	0/15
RF1-CMAES	**1.4**(0.9)	**1.3**(1)	**1.7**(2)	7.6(8)	∞ *2502*	0/15
RF5-CMAES	**2.2**(2)	**1.2**(0.6)	**2.3**(2)	61(79)	∞ *2504*	0/15
Sifeg	**1.9**(1)	**1.1**(0.9)	**1.1**(0.4)	117(244)	∞ *4e4*	0/15
Sif	**1.9**(3)	**1.1**(0.8)	**1.1**(0.4)	139(186)	∞ *4e4*	0/15
Srr	**1.9**(3)	**1.1**(0.6)	**1.1**(0.4)	50(13)	∞ *3e4*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f11	*4.0e+4*:6.4	*2.5e+3*:15	*6.3e+1*:217	*4.0e+1*:244	*2.5e+0*:675	15/15
BSifeg	**2.4**(1)	**2.2**(0.9)	1105(1450)	∞	∞ *7e4*	0/15
BSif	**2.4**(2)	**2.2**(1)	339(361)	2153(1833)	∞ *8e4*	0/15
BSqi	**2.4**(1)	**2.2**(1)	363(529)	1107(417)	∞ *8e4*	0/15
BSrr	**2.4**(1)	**2.2**(0.9)	503(409)	1283(718)	∞ *4e4*	0/15
CMA-CSA	**1.5**(1.0)	4.0(3)	7.1(1)	6.6(0.6)	**2.9**(0.2)	15/15
CMA-MSR	3.8(4)	4.2(5)	6.5(0.9)	**6.3**(1)	**3.3**(0.2)	15/15
CMA-TPA	**2.1**(2)	3.1(2)	**5.8**(2)	**5.9**(1)	**3.0**(0.4)	15/15
GP1-CMAES	3.1(3)	3.3(2)	**6.4**(3)	7.8(1)	55(65)	1/15
GP5-CMAES	**2.9**(2)	**2.9**(2)	**3.5**(2)	**5.1**(1)	4.3(1)	11/15
IPOPCMAv3p	3.6(6)	4.1(3)	11(13)	18(10)	∞ *2502*	0/15
LHD-10xDef	5.3(7)	5.3(4)	∞	∞	∞ *500*	0/15
LHD-2xDefa	4.3(3)	3.7(3)	16(13)	∞	∞ *500*	0/15
RAND-2xDef	3.6(4)	3.9(3)	33(49)	∞	∞ *500*	0/15
RF1-CMAES	3.2(4)	4.3(3)	19(12)	153(192)	∞ *2502*	0/15
RF5-CMAES	**2.4**(3)	3.2(2)	55(32)	154(201)	∞ *2514*	0/15
Sifeg	**2.4**(2)	**2.3**(1)	225(495)	1601(2467)	∞ *6e4*	0/15
Sif	**2.4**(2)	**2.3**(2)	244(228)	3570(4531)	∞ *6e4*	0/15
Srr	**2.4**(1)	**2.3**(1)	208(235)	577(785)	∞ *3e4*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f12	*4.0e+7*:15	*2.5e+7*:24	*1.6e+7*:34	*1.0e+6*:103	*1.0e+1*:515	15/15
BSifeg	**2.2**(0.7)	**1.6**(0.3)	**1.3**(0.2)	**3.0**(11)	262(255)	5/15
BSif	**2.2**(0.1)	**1.6**(0.3)	**1.3**(0.3)	3.3(7)	353(240)	4/15
BSqi	**2.2**(0.6)	**1.6**(0.3)	**1.3**(0.1)	6.1(0.2)	42(61)	14/15
BSrr	**2.2**(0.3)	**1.6**(0.3)	**1.3**(0.3)	**2.8**(0.4)	62(108)	11/15
CMA-CSA	3.4(3)	3.3(2)	**2.7**(1)	**2.9**(1)	4.2(5)	15/15
CMA-MSR	3.9(3)	3.8(2)	3.7(1)	3.3(0.6)	4.9(4)	15/15
CMA-TPA	3.7(4)	4.2(3)	3.6(2)	**2.5**(0.5)	**3.6**(3)	15/15
GP1-CMAES	**2.3**(2)	**2.6**(1)	**3.0**(2)	**2.9**(3)	**2.9**(3)	13/15
GP5-CMAES	6.8(4)	13(9)	17(8)	28(34)	21(13)	3/15
IPOPCMAv3p	**3.0**(2)	3.2(2)	3.3(2)	**2.8**(0.7)	**2.8**(4)	13/15
LHD-10xDef	6.7(4)	8.6(2)	6.9(1)	6.2(8)	∞ *500*	0/15
LHD-2xDefa	3.6(2)	**2.8**(0.6)	**2.4**(0.8)	**2.6**(1)	∞ *500*	0/15
RAND-2xDef	**2.4**(2)	**2.3**(0.7)	**2.1**(1)	**2.2**(1)	∞ *500*	0/15
RF1-CMAES	4.4(4)	3.9(2)	3.1(1)	**2.5**(1)	4.2(3)	11/15
RF5-CMAES	3.5(4)	3.1(2)	3.2(3)	21(19)	∞ *2504*	0/15
Sifeg	**2.2**(0.4)	**1.6**(0.1)	**1.2**(0.2)	7.3(17)	74(104)	5/15
Sif	**2.2**(0.3)	**1.6**(0.2)	**1.2**(0.2)	4.1(6)	99(113)	4/15
Srr	**2.2**(0.8)	**1.6**(0.3)	**1.2**(0.0)	3.6(4)	20(33)	10/15

Table 14: 10-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\text{best 2009}}$ on f_{13} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f13	*1.0e+3*:12	*6.3e+2*:32	*4.0e+2*:40	*6.3e+1*:154	*2.5e+0*:521	15/15
BSifeg	**2.9**(0.8)	**1.4**(0.4)	3.4(0.1)	45(68)	163(104)	11/15
BSif	**2.9**(0.5)	**1.4**(0.2)	**1.6**(0.2)	88(7)	∞ *9e4*	0/15
BSqi	**2.9**(0.7)	**1.4**(0.3)	**1.6**(0.2)	34(96)	207(193)	8/15
BSrr	**2.9**(1.0)	**1.4**(0.1)	**1.5**(0.2)	35(52)	208(202)	8/15
CMA-CSA	4.5(2)	3.6(0.5)	4.6(2)	**3.0**(0.3)	**3.3**(2)	15/15
CMA-MSR	5.7(3)	4.3(2)	5.9(1)	3.4(0.8)	**2.7**(0.4)	15/15
CMA-TPA	6.5(3)	4.0(1)	4.7(0.9)	**2.7**(0.5)	3.6(2)	15/15
GP1-CMAES	3.7(1)	**2.5**(0.6)	3.1(0.5)	**1.9**(0.5)	5.9(4)	8/15
GP5-CMAES	**2.6**(1)	**1.6**(0.3)	**1.7**(0.4)	**0.91**(0.1)	3.9(3)	10/15
IPOPCMAv3p	4.5(3)	3.5(3)	5.0(1)	3.2(0.6)	5.3(4)	9/15
LHD-10xDef	14(7)	7.0(0.2)	5.8(0.2)	**2.2**(0.2)	∞ *500*	0/15
LHD-2xDefa	3.7(1)	**2.2**(0.7)	**2.5**(0.8)	**1.6**(0.4)	**3.4**(3)	4/15
RAND-2xDef	3.6(2)	**2.1**(0.5)	**2.2**(0.6)	**1.6**(0.4)	7.2(13)	2/15
RF1-CMAES	3.4(1.0)	3.0(1)	3.5(0.9)	3.1(1)	16(20)	4/15
RF5-CMAES	3.8(2)	3.4(2)	11(16)	245(241)	∞ *2514*	0/15
Sifeg	**2.9**(0.5)	**1.4**(0.2)	**1.4**(0.3)	16(56)	75(153)	13/15
Sif	**2.9**(1)	**1.4**(0.2)	**1.4**(0.1)	17(56)	97(132)	11/15
Srr	**2.9**(1)	**1.4**(0.2)	**1.4**(0.1)	21(0.6)	76(103)	13/15

Table 15: 10-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	*4.0e+1*:7.7	*1.6e+1*:27	*1.0e+1*:37	*6.3e-1*:107	*1.0e-4*:505	15/15
BSifeg	**1.3**(1)	**1.2**(0.2)	**1.4**(1)	10(12)	∞ *1e5*	0/15
BSif	**1.3**(1)	**1.2**(0.4)	**1.4**(0.2)	33(91)	∞ *1e5*	0/15
BSqi	**1.3**(1)	**1.2**(0.4)	**1.3**(0.9)	5.9(3)	∞ *1e5*	0/15
BSrr	**1.3**(2)	**1.2**(0.4)	**1.3**(0.2)	9.0(9)	∞ *1e5*	0/15
CMA-CSA	**2.6**(1)	**2.3**(1)	**2.7**(1)	3.3(0.6)	**3.6**(0.3)	15/15
CMA-MSR	**2.1**(2)	**2.1**(2)	3.1(1)	3.6(0.8)	**3.6**(0.4)	15/15
CMA-TPA	**2.2**(2)	**2.4**(2)	3.0(1)	3.1(0.7)	**3.5**(0.2)	15/15
GP1-CMAES	**1.9**(2)	**1.6**(1.0)	**2.0**(1)	**2.3**(0.6)	25(36)	3/15
GP5-CMAES	**1.7**(1)	**1.5**(0.4)	**1.6**(0.4)	**1.7**(0.3)	∞ *2526*	0/15
IPOPCMAv3p	**1.2**(0.8)	**2.2**(1.0)	**2.4**(3)	3.2(1)	7.1(4)	10/15
LHD-10xDef	**1.8**(2)	3.8(4)	5.8(0.8)	4.0(0.5)	∞ *500*	0/15
LHD-2xDefa	**1.0**(0.5)	**1.8**(0.8)	**2.1**(1.0)	6.4(12)	∞ *500*	0/15
RAND-2xDef	**1.4**(2)	**1.8**(0.5)	**1.8**(0.4)	4.1(4)	∞ *500*	0/15
RF1-CMAES	**1.6**(1)	**1.7**(2)	**2.2**(1)	3.6(2)	∞ *2502*	0/15
RF5-CMAES	**1.5**(1)	**1.6**(1)	5.8(4)	42(41)	∞ *2504*	0/15
Sifeg	**1.3**(2)	**1.2**(0.5)	**1.3**(0.2)	**1.7**(0.7)	∞ *1e5*	0/15
Sif	**1.3**(2)	**1.2**(0.6)	**1.3**(0.3)	**2.1**(3)	∞ *1e5*	0/15
Srr	**1.3**(1)	**1.2**(0.7)	**1.2**(0.5)	**1.5**(0.6)	∞ *1e5*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f15	*2.5e+2*:9.0	*1.6e+2*:72	*1.0e+2*:186	*6.3e+1*:450	*4.0e+1*:872	15/15
BSifeg	3.5(1)	88(161)	78(251)	267(366)	1489(1436)	1/15
BSif	**3.5**(3)	72(0.8)	49(112)	198(356)	722(809)	2/15
BSqi	3.5(0.8)	83(300)	68(179)	277(480)	702(460)	2/15
BSrr	3.5(2)	46(12)	94(122)	163(100)	236(176)	5/15
CMA-CSA	4.9(2)	**1.4**(0.7)	**0.95**(0.5)	**1.1**(0.5)	**1.2**(0.4)	15/15
CMA-MSR	5.6(2)	**1.5**(0.7)	**1.2**(0.2)	**0.81**(0.2)	**0.57**(0.1)	15/15
CMA-TPA	6.9(4)	**1.8**(0.6)	**1.1**(0.3)	**0.94**(0.3)	**1.2**(0.7)	15/15
GP1-CMAES	4.3(5)	**1.1**(0.2)	**0.78**(0.3)	**0.79**(0.5)	**2.2**(2)	11/15
GP5-CMAES	**3.4**(2)	**0.75**(0.3)	**0.70**(0.4)	**3.0**(3)	41(65)	1/15
IPOPCMAv3p	3.9(3)	**1.3**(0.8)	**0.94**(0.4)	**0.92**(0.5)	**1.3**(2)	14/15
LHD-10xDef	10(10)	3.1(0.2)	**1.5**(0.2)	**0.83**(0.2)	**0.98**(0.6)	8/15
LHD-2xDefa	4.0(3)	**0.97**(0.4)	**0.71**(0.2)	**0.75**(1)	**0.97**(1)	7/15
RAND-2xDef	4.7(2)	**0.98**(0.4)	**0.71**(0.3)	**0.91**(1)	**1.9**(3)	4/15
RF1-CMAES	**3.2**(2)	**1.0**(0.5)	**0.83**(0.1)	**0.89**(0.4)	**0.99**(0.3)	14/15
RF5-CMAES	5.6(4)	3.3(0.5)	3.2(6)	4.4(5)	20(22)	2/15
Sifeg	3.8(2)	61(233)	56(43)	51(37)	138(134)	8/15
Sif	3.8(1)	61(114)	59(92)	76(154)	195(181)	6/15
Srr	3.7(3)	59(221)	65(130)	102(134)	153(115)	7/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f16	*4.0e+1*:12	*2.5e+1*:47	*1.6e+1*:88	*1.0e+1*:425	*4.0e+0*:989	15/15
BSifeg	**1.9**(2)	**1.0**(0.4)	13(33)	36(39)	105(92)	9/15
BSif	**1.9**(2)	**1.1**(0.7)	17(32)	38(45)	90(126)	9/15
BSqi	**1.9**(2)	**1.1**(0.6)	42(60)	60(96)	122(95)	8/15
BSrr	**1.9**(2)	**1.1**(0.6)	23(77)	18(27)	109(80)	8/15
CMA-CSA	**1.8**(1)	4.6(3)	6.9(4)	**1.8**(0.6)	**1.5**(3)	15/15
CMA-MSR	**2.5**(3)	**2.7**(2)	6.3(14)	**1.5**(0.4)	**2.5**(5)	15/15
CMA-TPA	4.5(5)	4.8(5)	10(8)	3.1(2)	**2.6**(3)	15/15
GP1-CMAES	**1.1**(0.8)	**2.4**(2)	3.2(3)	**1.1**(0.3)	**1.5**(2)	12/15
GP5-CMAES	**1.4**(2)	**1.6**(1.0)	**1.4**(0.4)	**0.39**(0.1)	**0.89**(1)	13/15
IPOPCMAv3p	**1.1**(1)	3.3(3)	7.7(3)	**2.4**(0.9)	**1.6**(1)	14/15
LHD-10xDef	**1.7**(2)	3.6(2)	3.7(3)	**1.6**(1)	∞ *500*	0/15
LHD-2xDefa	**1.4**(1)	**1.8**(0.7)	6.9(7)	5.4(4)	∞ *500*	0/15
RAND-2xDef	**1.3**(0.9)	**2.6**(3)	3.6(4)	**1.8**(2)	7.3(6)	1/15
RF1-CMAES	**1.6**(1)	3.1(2)	4.5(3)	**1.3**(0.8)	4.4(3)	6/15
RF5-CMAES	**1.5**(2)	**2.0**(2)	**2.4**(1)	3.6(6)	17(21)	2/15
Sifeg	**2.0**(3)	**1.5**(0.8)	**1.4**(0.8)	5.9(18)	28(12)	15/15
Sif	**2.0**(2)	**1.5**(1)	8.9(30)	4.2(4)	39(29)	14/15
Srr	**2.0**(3)	**1.5**(0.9)	12(54)	4.5(6)	21(27)	15/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f17	*1.0e+1*:26	*6.3e+0*:85	*4.0e+0*:155	*2.5e+0*:238	*6.3e-1*:585	15/15
BSifeg	**1.4**(1)	301(1177)	982(1814)	2734(4204)	∞ *1e5*	0/15
BSif	**1.4**(0.9)	237(464)	965(1440)	1182(1806)	∞ *1e5*	0/15
BSqi	**1.4**(0.6)	205(661)	740(647)	1684(1788)	∞ *1e5*	0/15
BSrr	**1.4**(0.7)	296(592)	761(1122)	1812(1492)	∞ *1e5*	0/15
CMA-CSA	3.4(2)	**2.1**(1.0)	**1.7**(0.7)	**1.6**(0.7)	**1.7**(3)	15/15
CMA-MSR	**2.0**(0.9)	**1.4**(0.6)	**1.2**(0.7)	**1.1**(0.3)	**2.8**(2)	15/15
CMA-TPA	**2.4**(2)	**1.4**(0.3)	**1.1**(0.3)	**1.0**(0.4)	**1.2**(0.2)	15/15
GP1-CMAES	**1.7**(0.8)	**1.0**(0.6)	**0.88**(0.3)	**0.80**(0.3)	**1.0**(0.2)	14/15
GP5-CMAES	**1.7**(1)	**1.1**(0.5)	**0.99**(0.8)	**2.4**(6)	18(26)	3/15
IPOPCMAv3p	**2.4**(2)	**1.4**(1)	**1.4**(0.4)	**1.3**(0.6)	**1.0**(0.2)	15/15
LHD-10xDef	3.5(4)	**2.7**(0.4)	**2.1**(0.6)	4.4(6)	∞ *500*	0/15
LHD-2xDefa	**1.6**(0.8)	**0.99**(0.6)	**2.4**(2)	6.9(7)	∞ *500*	0/15
RAND-2xDef	**1.8**(0.7)	**1.2**(0.8)	**1.7**(0.7)	3.7(3)	∞ *500*	0/15
RF1-CMAES	**1.7**(2)	**1.3**(0.6)	**1.2**(0.4)	**3.0**(0.7)	13(17)	4/15
RF5-CMAES	**1.6**(0.2)	**2.6**(0.2)	5.0(6)	10(10)	64(29)	1/15
Sifeg	**1.3**(0.7)	208(202)	402(557)	440(736)	1223(2308)	2/15
Sif	**1.3**(0.8)	194(337)	490(1441)	834(1353)	∞ *1e5*	0/15
Srr	**1.3**(0.4)	182(4)	429(1444)	487(629)	∞ *1e5*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f18	*4.0e+1*:11	*2.5e+1*:56	*1.6e+1*:172	*1.6e+1*:172	*2.5e+0*:561	15/15
BSifeg	3.2(5)	382(511)	354(420)	354(544)	2408(1070)	1/15
BSif	3.5(7)	362(748)	317(415)	317(529)	∞ *1e5*	0/15
BSqi	81(589)	223(397)	281(731)	281(287)	2341(1144)	1/15
BSrr	**3.0**(2)	305(691)	528(814)	528(532)	∞ *9e4*	0/15
CMA-CSA	6.7(6)	**2.5**(0.5)	**1.2**(0.7)	**1.2**(0.4)	**1.0**(0.2)	15/15
CMA-MSR	4.9(4)	**1.8**(0.3)	**0.99**(0.4)	**0.99**(0.3)	**1.9**(4)	15/15
CMA-TPA	5.1(4)	**2.1**(0.9)	**1.1**(0.3)	**1.1**(0.2)	**1.5**(0.3)	15/15
GP1-CMAES	4.4(5)	**1.7**(0.8)	**0.85**(0.5)	**0.85**(0.5)	**1.6**(1)	13/15
GP5-CMAES	3.7(6)	**1.3**(0.5)	**0.71**(0.2)	**0.71**(0.4)	6.0(10)	7/15
IPOPCMAv3p	4.7(3)	**2.1**(1)	**1.2**(0.6)	**1.2**(0.4)	**1.5**(0.3)	14/15
LHD-10xDef	8.2(9)	4.0(0.2)	**2.0**(2)	**2.0**(0.4)	∞ *500*	0/15
LHD-2xDefa	4.2(3)	**1.8**(1)	**1.8**(2)	**1.8**(2)	∞ *500*	0/15
RAND-2xDef	**3.0**(2)	**1.5**(0.6)	**1.6**(2)	**1.6**(3)	∞ *500*	0/15
RF1-CMAES	**2.8**(2)	**1.4**(0.8)	**0.95**(0.4)	**0.95**(0.5)	3.2(2)	10/15
RF5-CMAES	4.0(2)	4.5(22)	3.8(10)	3.8(10)	66(54)	1/15
Sifeg	8.3(2)	83(7)	171(286)	171(236)	1179(1377)	2/15
Sif	320(3)	194(671)	223(324)	223(553)	∞ *1e5*	0/15
Srr	6.7(2)	137(290)	184(432)	184(327)	∞ *9e4*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f19	*1.6e-1*:618	*1.0e-1*:10609	*6.3e-2*:10623	*4.0e-2*:10625	*2.5e-2*:10644	15/15
BSifeg	∞	∞	∞	∞	∞ *1e5*	0/15
BSif	∞	∞	∞	∞	∞ *1e5*	0/15
BSqi	∞	∞	∞	∞	∞ *1e5*	0/15
BSrr	∞	∞	∞	∞	∞ *1e5*	0/15
CMA-CSA	**171**(101)	**12**(5)	**19**(12)	**23**(10)	**26**(25)	15/15
CMA-MSR	**237**(437)	**21**(31)	**35**(55)	**62**(128)	**290**(257)	4/15
CMA-TPA	**156**(83)	**12**(5)	**18**(11)	**24**(13)	**34**(18)	15/15
GP1-CMAES	∞	∞	∞	∞	∞ *2504*	0/15
GP5-CMAES	∞	∞	∞	∞	∞ *2528*	0/15
IPOPCMAv3p	∞	∞	∞	∞	∞ *2504*	0/15
LHD-10xDef	∞	∞	∞	∞	∞ *500*	0/15
LHD-2xDefa	∞	∞	∞	∞	∞ *500*	0/15
RAND-2xDef	∞	∞	∞	∞	∞ *500*	0/15
RF1-CMAES	∞	∞	∞	∞	∞ *2502*	0/15
RF5-CMAES	∞	∞	∞	∞	∞ *2516*	0/15
Sifeg	∞	∞	∞	∞	∞ *1e5*	0/15
Sif	∞	∞	∞	∞	∞ *1e5*	0/15
Srr	∞	∞	∞	∞	∞ *1e5*	0/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f20	*1.0e+4*:17	*6.3e+3*:21	*6.3e+1*:30	*2.5e+0*:122	*1.0e+0*:15426	13/15
BSifeg	**1.9**(0.6)	**1.9**(0.1)	**1.7**(0.8)	**1.9**(2)	3.2(4)	12/15
BSif	**1.9**(0.1)	**1.9**(0.1)	**1.8**(1)	**1.8**(2)	3.5(3)	12/15
BSqi	**1.9**(0.1)	**1.9**(0.1)	**1.7**(0.5)	**1.1**(1)	3.4(7)	12/15
BSrr	**1.9**(1)	**1.9**(0.1)	**1.7**(0.4)	**1.4**(0.3)	**2.8**(3)	13/15
CMA-CSA	**2.3**(1)	**2.4**(2)	5.2(1)	5.5(2)	**1.8**(0.9)	15/15
CMA-MSR	3.1(3)	3.5(2)	5.7(2)	23(2)	∞ *1e6*	0/15
CMA-TPA	3.1(2)	3.2(2)	5.1(0.9)	3.0(1)	18(0.3)	14/15
GP1-CMAES	**1.9**(2)	**2.0**(1)	3.8(1.0)	8.3(9)	∞ *2502*	0/15
GP5-CMAES	**2.1**(1.0)	**1.9**(0.8)	**2.4**(0.6)	25(23)	∞ *2526*	0/15
IPOPCMAv3p	3.2(3)	3.3(2)	5.4(2)	5.1(3)	**2.3**(5)	1/15
LHD-10xDef	3.6(6)	6.0(5)	7.7(0.3)	14(14)	∞ *500*	0/15
LHD-2xDefa	**2.9**(0.7)	**2.9**(0.6)	**2.9**(0.7)	6.2(5)	∞ *500*	0/15
RAND-2xDef	**1.8**(2)	**2.2**(0.5)	**2.9**(2)	4.9(3)	∞ *500*	0/15
RF1-CMAES	**2.6**(1)	3.0(2)	5.4(2)	4.4(2)	∞ *2502*	0/15
RF5-CMAES	**2.4**(2)	**2.5**(2)	21(32)	35(65)	∞ *2514*	0/15
Sifeg	**1.9**(1)	**1.9**(0.2)	**2.0**(0.6)	**0.72**(0.4)	**1.4**(2)	15/15
Sif	**1.9**(0.7)	**1.9**(0.1)	**1.9**(0.7)	**0.76**(0.4)	**0.87**(1)	15/15
Srr	**1.9**(0.7)	**1.9**(0.2)	**1.9**(0.4)	**0.76**(0.2)	**1.9**(3)	14/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f21	*4.0e+1*:30	*2.5e+1*:46	*1.6e+1*:56	*1.0e+1*:130	*6.3e+0*:639	15/15
BSifeg	**2.3**(0.5)	263(547)	374(0.5)	171(317)	116(123)	10/15
BSif	5.7(0.7)	354(54)	520(656)	331(768)	226(279)	7/15
BSqi	**2.5**(8)	217(557)	333(33)	151(29)	92(121)	11/15
BSrr	4.1(0.7)	222(762)	338(25)	159(282)	95(147)	11/15
CMA-CSA	4.7(3)	7.6(16)	11(25)	7.7(11)	5.0(10)	15/15
CMA-MSR	3.5(2)	3.7(3)	15(30)	10(13)	8.4(3)	15/15
CMA-TPA	3.3(1)	3.0(0.8)	8.4(12)	4.8(5)	**1.4**(2)	15/15
GP1-CMAES	**2.1**(1)	**2.1**(0.9)	**2.0**(0.6)	**2.4**(10)	**1.2**(3)	12/15
GP5-CMAES	**1.8**(0.5)	**2.6**(4)	7.1(8)	3.7(13)	**1.3**(2)	13/15
IPOPCMAv3p	3.5(3)	8.1(15)	17(37)	10(10)	3.2(5)	9/15
LHD-10xDef	6.5(3)	4.7(0.2)	4.6(0.3)	**2.7**(2)	**1.0**(2)	9/15
LHD-2xDefa	**2.1**(2)	**1.8**(0.3)	**2.1**(0.8)	**1.3**(1)	**0.52**(0.2)	11/15
RAND-2xDef	**2.0**(0.3)	**1.7**(0.4)	**2.2**(0.3)	**1.2**(0.2)	**0.39**(0.3)	12/15
RF1-CMAES	9.1(2)	7.5(15)	11(2)	9.1(11)	4.0(9)	8/15
RF5-CMAES	**2.9**(2)	6.2(0.8)	6.9(12)	13(20)	12(14)	4/15
Sifeg	**1.6**(0.9)	216(1297)	177(612)	79(385)	88(100)	11/15
Sif	**1.6**(0.6)	256(1092)	210(284)	110(384)	109(150)	10/15
Srr	**1.6**(0.7)	197(0.8)	162(4)	78(78)	65(59)	12/15

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

#FEs/D	0.5	1.2	3	10	50	#succ
f22	*6.3e+1*:18	*4.0e+1*:30	*4.0e+1*:30	*6.3e+0*:155	*4.0e+0*:631	14/15
BSifeg	404(1)	737(829)	737(1656)	530(487)	135(475)	9/15
BSif	404(1405)	1042(2485)	1042(2927)	813(798)	251(673)	6/15
BSqi	404(2810)	713(1339)	713(828)	434(647)	109(119)	10/15
BSrr	404(0.8)	802(3787)	802(1302)	469(416)	116(116)	10/15
CMA-CSA	3.3(4)	3.9(1)	3.9(1)	303(900)	183(249)	13/15
CMA-MSR	5.1(3)	11(25)	11(2)	428(2713)	349(1667)	12/15
CMA-TPA	4.7(3)	4.1(3)	4.1(3)	604(1968)	148(2)	13/15
GP1-CMAES	**3.2**(2)	**3.1**(2)	**3.1**(0.7)	**3.9**(13)	**1.3**(2)	12/15
GP5-CMAES	**2.9**(1)	4.0(1)	4.0(12)	12(14)	3.4(8)	9/15
IPOPCMAv3p	4.5(3)	4.7(2)	4.7(5)	17(21)	5.6(3)	7/15
LHD-10xDef	6.2(6)	7.7(0.1)	7.7(0.3)	5.7(4)	**1.4**(2)	7/15
LHD-2xDefa	3.4(1.0)	**3.1**(2)	**3.1**(2)	**4.4**(6)	**1.3**(1)	7/15
RAND-2xDef	3.4(2)	**2.8**(2)	**2.8**(0.7)	**4.6**(3)	**1.4**(4)	7/15
RF1-CMAES	**3.2**(3)	10(23)	10(65)	16(24)	8.3(8)	5/15
RF5-CMAES	3.9(3)	4.4(2)	4.4(2)	36(52)	9.0(12)	5/15
Sifeg	8.0(22)	442(1261)	442(699)	280(380)	83(125)	11/15
Sif	404(0.5)	772(1364)	772(1968)	611(835)	151(88)	8/15
Srr	404(1406)	672(1655)	672(1655)	389(317)	109(195)	10/15

Table 24: 10-D, running time excess ERT/$\mathrm{ERT}_{\text{best 2009}}$ on f_{23} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding $\mathrm{ERT}_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f23	*6.3e+0*:10	*4.0e+0*:62	*2.5e+0*:162	*2.5e+0*:162	*1.0e+0*:915	15/15
BSifeg	**2.9**(4)	**1.5**(0.7)	**1.7**(2)	**1.7**(2)	13(11)	15/15
BSif	**2.9**(2)	**1.2**(0.9)	**1.6**(2)	**1.6**(2)	12(24)	15/15
BSqi	**2.9**(4)	**1.2**(0.5)	**1.1**(2)	**1.1**(1)	13(10)	15/15
BSrr	**2.9**(4)	**1.2**(1)	**1.5**(2)	**1.5**(2)	11(16)	15/15
CMA-CSA	6.3(5)	6.0(6)	18(14)	18(14)	23(16)	15/15
CMA-MSR	4.8(3)	3.2(3)	**2.8**(2)	**2.8**(0.8)	**2.9**(4)	15/15
CMA-TPA	3.4(4)	3.9(2)	16(23)	16(19)	12(5)	15/15
GP1-CMAES	**2.0**(3)	**1.7**(1)	5.3(4)	5.3(4)	**2.7**(2)	10/15
GP5-CMAES	3.6(4)	**1.7**(2)	**1.6**(0.3)	**1.6**(4)	**0.92**(0.4)	13/15
IPOPCMAv3p	**2.3**(0.8)	**2.8**(2)	4.4(3)	4.4(2)	∞ *2514*	0/15
LHD-10xDef	**2.1**(2)	**2.3**(2)	5.9(4)	5.9(2)	∞ *500*	0/15
LHD-2xDefa	**2.0**(1)	**1.6**(2)	8.3(11)	8.3(17)	∞ *500*	0/15
RAND-2xDef	**2.0**(2)	3.8(9)	10(13)	10(5)	∞ *500*	0/15
RF1-CMAES	**1.7**(1)	3.5(4)	7.1(5)	7.1(10)	∞ *2506*	0/15
RF5-CMAES	**1.6**(3)	3.1(3)	10(8)	10(7)	∞ *2548*	0/15
Sifeg	**2.5**(3)	**1.9**(2)	3.5(3)	3.5(4)	4.6(9)	15/15
Sif	**2.5**(3)	**1.9**(2)	3.7(4)	3.7(4)	6.4(11)	15/15
Srr	**2.5**(3)	**1.9**(2)	3.7(2)	3.7(2)	5.7(4)	15/15

Table 25: 10-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f24	*1.0e+2*:66	*6.3e+1*:596	*4.0e+1*:3181	*2.5e+1*:7668	*1.6e+1*:14353	15/15
BSifeg	22(16)	50(73)	34(31)	87(130)	∞ *9e4*	0/15
BSif	37(13)	37(13)	72(62)	∞	∞ *9e4*	0/15
BSqi	18(10)	71(104)	55(63)	174(352)	∞ *9e4*	0/15
BSrr	19(18)	39(56)	41(40)	162(215)	∞ *8e4*	0/15
CMA-CSA	**2.6**(0.9)	**1.1**(1)	**1.2**(1)	**0.62**(0.5)	**1.2**(0.9)	15/15
CMA-MSR	**2.6**(1)	**0.84**(0.3)	**0.45**(0.6)	**0.50**(0.4)	**0.83**(0.6)	15/15
CMA-TPA	**2.6**(1)	**1.0**(0.5)	**0.67**(0.2)	**0.70**(0.8)	**0.94**(0.6)	15/15
GP1-CMAES	**1.7**(0.8)	**0.90**(0.5)	**2.6**(3)	**1.1**(0.9)	∞ *2514*	0/15
GP5-CMAES	**1.2**(0.3)	**0.81**(0.9)	∞	∞	∞ *2528*	0/15
IPOPCMAv3p	**2.7**(1)	**1.4**(0.9)	**2.7**(1)	4.8(4)	∞ *2504*	0/15
LHD-10xDef	4.8(1)	**2.9**(3)	**2.4**(2)	∞	∞ *500*	0/15
LHD-2xDefa	8.3(8)	∞	∞	∞	∞ *500*	0/15
RAND-2xDef	4.8(2)	∞	∞	∞	∞ *500*	0/15
RF1-CMAES	**2.2**(0.9)	**1.1**(0.9)	**1.2**(2)	**2.3**(2)	∞ *2502*	0/15
RF5-CMAES	**2.4**(1)	3.4(3)	∞	∞	∞ *2514*	0/15
Sifeg	3.0(3)	5.9(13)	9.0(7)	26(33)	∞ *9e4*	0/15
Sif	3.0(2)	10(16)	13(19)	35(33)	44(21)	2/15
Srr	**2.8**(2)	**1.5**(2)	8.8(5)	22(27)	46(55)	2/15

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.