Comparison Tables: BBOB 2015 Testbed in 20-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: 20-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	*6.3e+1*:24	*4.0e+1*:42	*1.0e-8*:43	*1.0e-8*:43	*1.0e-8*:43	15/15
BSifeg	**2.2**(1.0)	**1.7**(0.1)	**2.5**(0.2)	**2.5**(0.2)	**2.5**(0.3)	15/15
BSif	**2.2**(0.9)	**1.7**(0.1)	**2.5**(0.3)	**2.5**(0.2)	**2.5**(0.3)	15/15
BSqi	**2.2**(0.9)	**1.7**(0.1)	**2.5**(0.2)	**2.5**(0.3)	**2.5**(0.2)	15/15
BSrr	**2.2**(0.4)	**1.7**(0.1)	**2.5**(0.2)	**2.5**(0.2)	**2.5**(0.2)	15/15
CMA-CSA	4.7(2)	4.4(1)	64(4)	64(3)	64(5)	15/15
CMA-MSR	6.0(2)	5.1(2)	75(4)	75(3)	75(3)	15/15
CMA-TPA	5.4(2)	4.0(0.9)	46(3)	46(4)	46(1)	15/15
GP1-CMAES	3.9(2)	3.1(0.6)	58(6)	58(10)	58(7)	15/15
GP5-CMAES	**2.9**(0.9)	**2.0**(0.2)	∞	∞	∞ *5034*	0/15
IPOPCMAv3p	4.8(2)	4.6(1)	64(2)	64(2)	64(2)	15/15
LHD-10xDef	17(0.1)	10(0.1)	∞	∞	∞ *1000*	0/15
LHD-2xDefa	3.9(0.2)	**2.5**(0.2)	∞	∞	∞ *1000*	0/15
RAND-2xDef	4.0(0.5)	**2.8**(0.4)	∞	∞	∞ *1000*	0/15
RF1-CMAES	3.9(1)	3.5(0.8)	73(20)	73(18)	73(18)	15/15
RF5-CMAES	3.7(2)	3.0(0.9)	∞	∞	∞ *5006*	0/15
Sifeg	**2.2**(0.6)	**1.7**(0.2)	16(0.9)	16(0.9)	16(1)	15/15
Sif	**2.2**(0.7)	**1.7**(0.1)	16(1)	16(1)	16(2)	15/15
Srr	**2.2**(0.8)	**1.7**(0.1)	16(1)	16(1)	16(0.7)	15/15

Table 3: 20-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	*4.0e+6*:29	*2.5e+6*:42	*1.0e+5*:65	*1.0e+4*:207	*1.0e-8*:412	15/15
BSifeg	**2.3**(1.0)	**1.8**(0.2)	**1.5**(0.2)	**0.54**(0.1)	**1.2**(0.1)	15/15
BSif	**2.3**(1.0)	**1.8**(0.2)	**1.5**(0.0)	**0.55**(0.1)	**1.2**(0.1)	15/15
BSqi	**2.3**(0.0)	**1.8**(0.2)	**1.5**(0.1)	**0.53**(0.1)	**1.3**(0.2)	15/15
BSrr	**2.3**(1)	**1.8**(0.3)	**1.5**(0.1)	**0.54**(0.1)	**1.3**(0.2)	15/15
CMA-CSA	**1.2**(0.6)	**1.1**(0.8)	14(3)	11(3)	33(1)	15/15
CMA-MSR	**1.0**(0.4)	**1.6**(2)	11(5)	7.9(3)	38(3)	15/15
CMA-TPA	**1.4**(1)	**1.5**(0.7)	11(4)	9.5(3)	36(2)	15/15
GP1-CMAES	**1.5**(0.9)	**1.6**(1)	8.9(5)	8.0(3)	∞ *5006*	0/15
GP5-CMAES	**0.90**(0.9)	**1.3**(1)	5.4(2)	3.9(0.8)	∞ *5006*	0/15
IPOPCMAv3p	**0.93**(0.6)	**1.4**(1)	14(4)	11(2)	∞ *5006*	0/15
LHD-10xDef	**1.6**(1)	4.1(4)	30(19)	∞	∞ *1000*	0/15
LHD-2xDefa	**1.1**(0.8)	**1.9**(0.8)	8.5(5)	72(67)	∞ *1000*	0/15
RAND-2xDef	**1.3**(2)	**1.6**(2)	7.8(4)	70(116)	∞ *1000*	0/15
RF1-CMAES	**1.1**(0.8)	**1.2**(1)	11(3)	29(27)	∞ *5006*	0/15
RF5-CMAES	**1.2**(1)	**1.3**(0.9)	184(149)	∞	∞ *5006*	0/15
Sifeg	**2.3**(0.9)	**1.9**(0.3)	**1.9**(0.2)	**0.77**(0.1)	**2.2**(0.2)	15/15
Sif	**2.3**(0.5)	**1.9**(0.5)	**1.9**(0.2)	**0.82**(0.2)	**2.2**(0.3)	15/15
Srr	**2.3**(1)	**1.9**(0.5)	**1.9**(0.3)	**0.76**(0.1)	**2.2**(0.3)	15/15

Table 4: 20-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	*6.3e+2*:33	*4.0e+2*:44	*1.6e+2*:109	*1.0e+2*:255	*2.5e+1*:3277	15/15
BSifeg	**1.7**(0.2)	**1.6**(0.2)	**0.92**(0.1)	**0.52**(0.0)	**0.12**(0.0)	15/15
BSif	**1.7**(0.2)	**1.6**(0.2)	**0.92**(0.0)	**0.52**(0.0)	**0.12**(0.0)	15/15
BSqi	**1.7**(0.2)	**1.6**(0.2)	**0.92**(0.1)	**0.52**(0.1)	**0.11**(0.0)	15/15
BSrr	**1.7**(0.2)	**1.6**(0.1)	**0.92**(0.1)	**0.52**(0.1)	**0.11**(0.0)	15/15
CMA-CSA	**1.9**(2)	4.0(0.5)	7.4(2)	7.0(3)	3.5(1.0)	15/15
CMA-MSR	**2.9**(2)	4.5(0.8)	5.2(1.0)	**3.0**(0.3)	3.5(1)	15/15
CMA-TPA	3.1(1)	4.0(1.0)	6.0(4)	4.4(1)	**2.5**(1)	15/15
GP1-CMAES	**2.3**(1)	3.2(1.0)	5.9(1)	4.2(2)	22(34)	1/15
GP5-CMAES	**1.9**(0.9)	**2.9**(1)	15(23)	43(50)	∞ *5034*	0/15
IPOPCMAv3p	**1.9**(1)	3.6(2)	7.6(4)	7.0(2)	22(30)	1/15
LHD-10xDef	9.1(4)	10(0.3)	8.7(0.7)	9.0(6)	∞ *1000*	0/15
LHD-2xDefa	**2.5**(0.5)	3.0(0.5)	8.6(7)	18(40)	∞ *1000*	0/15
RAND-2xDef	**2.7**(0.3)	**2.8**(0.4)	5.7(5)	10(24)	∞ *1000*	0/15
RF1-CMAES	**2.0**(1)	3.7(0.9)	6.4(2)	4.0(0.8)	22(41)	1/15
RF5-CMAES	**1.7**(0.8)	**2.6**(0.5)	18(17)	82(128)	∞ *5006*	0/15
Sifeg	**1.7**(0.2)	**1.6**(0.1)	**0.97**(0.1)	**0.54**(0.1)	**0.14**(0.0)	15/15
Sif	**1.7**(0.4)	**1.6**(0.2)	**0.97**(0.1)	**0.55**(0.1)	**0.15**(0.0)	15/15
Srr	**1.7**(0.2)	**1.6**(0.2)	**0.97**(0.2)	**0.54**(0.1)	**0.13**(0.0)	15/15

Table 5: 20-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_4 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f4	*6.3e+2*:22	*4.0e+2*:91	*2.5e+2*:250	*1.6e+2*:332	*6.3e+1*:1927	15/15
BSifeg	**2.8**(1)	**0.98**(0.1)	**0.58**(0.3)	**0.66**(0.3)	**0.20**(0.0)	15/15
BSif	**2.8**(0.9)	**0.98**(0.1)	**0.59**(0.3)	**0.68**(0.4)	**0.21**(0.0)	15/15
BSqi	**2.8**(1)	**0.98**(0.2)	**0.62**(0.5)	**0.70**(0.3)	**0.24**(0.0)	15/15
BSrr	**2.8**(0.5)	**0.98**(0.2)	**0.60**(0.3)	**0.69**(0.3)	**0.21**(0.0)	15/15
CMA-CSA	7.1(3)	3.1(0.4)	**2.1**(0.4)	3.7(1)	**2.2**(1)	15/15
CMA-MSR	8.5(3)	3.3(0.8)	**2.0**(0.8)	3.6(5)	5.4(6)	15/15
CMA-TPA	8.2(2)	**2.9**(0.8)	**1.9**(0.5)	**2.9**(0.6)	3.5(2)	15/15
GP1-CMAES	9.2(8)	4.5(2)	11(4)	42(88)	39(22)	1/15
GP5-CMAES	7.5(2)	5.5(9)	13(45)	215(159)	∞ *5022*	0/15
IPOPCMAv3p	7.4(3)	3.3(1)	**2.3**(0.6)	4.1(1)	**2.4**(2)	10/15
LHD-10xDef	22(10)	18(12)	∞	∞	∞ *1000*	0/15
LHD-2xDefa	9.1(3)	6.8(1)	58(71)	∞	∞ *1000*	0/15
RAND-2xDef	10(8)	8.4(8)	13(11)	∞	∞ *1000*	0/15
RF1-CMAES	8.1(2)	3.4(1)	**2.5**(0.6)	7.7(13)	∞ *5006*	0/15
RF5-CMAES	8.9(3)	13(2)	57(110)	∞	∞ *5006*	0/15
Sifeg	**2.8**(0.8)	**1.1**(0.3)	**0.56**(0.2)	**0.60**(0.1)	**0.17**(0.0)	15/15
Sif	**2.8**(1)	**1.1**(0.2)	**0.58**(0.1)	**0.61**(0.1)	**0.17**(0.0)	15/15
Srr	**2.8**(1)	**1.1**(0.2)	**0.56**(0.1)	**0.59**(0.1)	**0.16**(9e-3)	15/15

Table 6: 20-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	*2.5e+2*:19	*1.6e+2*:34	*1.0e-8*:41	*1.0e-8*:41	*1.0e-8*:41	15/15
BSifeg	**2.1**(0.3)	**1.5**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
BSif	**2.1**(0.3)	**1.5**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
BSqi	**2.1**(0.2)	**1.5**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
BSrr	**2.1**(0.2)	**1.5**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
CMA-CSA	**1.5**(0.7)	**1.8**(0.5)	6.0(0.8)	6.0(1)	6.0(1)	15/15
CMA-MSR	**1.9**(0.6)	**1.9**(0.7)	5.6(1)	5.6(1)	5.6(1)	15/15
CMA-TPA	**1.3**(1)	**1.4**(0.4)	4.9(2)	4.9(1)	4.9(1)	15/15
GP1-CMAES	**1.6**(0.8)	**1.9**(0.5)	92(115)	92(62)	92(85)	11/15
GP5-CMAES	**1.7**(0.9)	**1.7**(0.4)	4.8(0.7)	4.8(2)	4.8(1)	15/15
IPOPCMAv3p	**2.0**(2)	**2.0**(0.9)	36(13)	36(15)	36(19)	15/15
LHD-10xDef	8.1(9)	12(0.0)	11(0.4)	11(0.2)	11(0.3)	15/15
LHD-2xDefa	3.7(2)	**2.5**(0.0)	**3.0**(0.1)	**3.0**(0.2)	**3.0**(0.2)	15/15
RAND-2xDef	3.4(2)	**2.6**(0.1)	3.4(2)	3.4(3)	3.4(3)	15/15
RF1-CMAES	**1.8**(1)	**2.3**(0.8)	50(22)	50(26)	50(24)	15/15
RF5-CMAES	**2.0**(0.9)	**2.0**(0.6)	265(451)	265(361)	265(330)	6/15
Sifeg	**2.1**(0.2)	**1.5**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
Sif	**2.1**(0.3)	**1.5**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15
Srr	**2.1**(0.2)	**1.5**(0.1)	**1.5**(0.0)	**1.5**(0.0)	**1.5**(0.0)	15/15

Table 7: 20-D, running time excess ERT/$\mathrm{ERT}_{\mathrm{best\ 2009}}$ 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 $\mathrm{ERT}_{\mathrm{best\ 2009}}$ (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
f6	*2.5e+5*:16	*6.3e+4*:43	*1.6e+4*:62	*1.6e+2*:353	*1.6e+1*:1078	15/15
BSifeg	**2.3**(1)	**1.3**(0.8)	**1.2**(0.4)	39(42)	713(1132)	3/15
BSif	**2.3**(2)	**1.3**(0.3)	**1.2**(0.5)	159(199)	2558(3325)	1/15
BSqi	**1.9**(2)	**1.1**(0.7)	**1.1**(0.3)	34(59)	419(453)	2/7
BSrr	**2.3**(1)	**1.3**(0.7)	**1.2**(0.4)	36(47)	210(81)	8/15
CMA-CSA	3.6(2)	**2.5**(1)	**2.4**(1)	**2.8**(0.8)	**1.8**(0.3)	15/15
CMA-MSR	3.6(2)	**2.4**(0.6)	**2.4**(0.9)	**1.9**(0.8)	**1.5**(0.6)	15/15
CMA-TPA	3.1(2)	**2.2**(2)	**2.0**(0.7)	**2.3**(0.7)	**1.6**(0.5)	15/15
GP1-CMAES	**2.8**(2)	**2.0**(0.9)	**2.0**(0.8)	**1.7**(1)	4.2(4)	11/15
GP5-CMAES	**2.6**(2)	**1.6**(0.5)	**1.5**(0.4)	19(24)	∞ *5024*	0/15
IPOPCMAv3p	**2.9**(2)	**2.3**(1)	**2.3**(2)	**2.0**(0.5)	**1.5**(0.2)	15/15
LHD-10xDef	17(12)	10(1)	7.1(0.2)	4.3(4)	∞ *1000*	0/15
LHD-2xDefa	4.6(2)	**2.3**(0.3)	**1.9**(0.4)	6.8(6)	∞ *1000*	0/15
RAND-2xDef	4.8(0.2)	**2.3**(0.3)	**1.8**(0.5)	5.7(9)	∞ *1000*	0/15
RF1-CMAES	**2.7**(2)	**2.1**(0.7)	**2.3**(1)	**2.9**(1)	66(118)	1/15
RF5-CMAES	**2.1**(2)	**1.6**(0.6)	**1.8**(0.4)	32(28)	∞ *5006*	0/15
Sifeg	**2.3**(2)	**1.2**(0.5)	**1.2**(0.4)	7.4(9)	151(146)	10/15
Sif	**2.3**(2)	**1.2**(0.7)	**1.2**(0.5)	37(71)	294(301)	7/15
Srr	**2.3**(2)	**1.2**(0.7)	**1.2**(0.3)	10(5)	94(67)	12/15

Table 8: 20-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f7	*1.0e+3*:11	*4.0e+2*:39	*2.5e+2*:74	*6.3e+1*:319	*1.0e+1*:1351	15/15
BSifeg	**1.4**(0.7)	**1.5**(0.5)	**1.0**(0.5)	187(162)	∞ *2e5*	0/15
BSif	**1.4**(2)	**1.5**(0.5)	7.0(0.4)	258(349)	∞ *2e5*	0/15
BSqi	**1.4**(2)	**1.5**(0.5)	**1.0**(0.4)	288(335)	∞ *2e5*	0/15
BSrr	**1.4**(1)	**1.5**(0.6)	**1.0**(0.5)	174(69)	∞ *2e5*	0/15
CMA-CSA	**2.7**(2)	3.5(2)	3.1(1)	**1.7**(0.4)	**1.7**(2)	15/15
CMA-MSR	3.5(2)	3.1(0.6)	**2.7**(1.0)	**1.4**(0.4)	**2.1**(1)	15/15
CMA-TPA	3.2(3)	3.7(1)	**2.6**(0.7)	**1.3**(0.3)	**2.1**(1)	15/15
GP1-CMAES	**2.1**(2)	**2.2**(1)	**1.7**(0.7)	**0.92**(0.4)	3.0(3)	10/15
GP5-CMAES	**2.4**(2)	**1.8**(0.6)	**1.3**(0.1)	**0.58**(0.0)*	**1.6**(0.8)	14/15
IPOPCMAv3p	**1.3**(1)	**2.2**(2)	**2.5**(0.8)	**1.6**(0.2)	**1.3**(0.7)	15/15
LHD-10xDef	**2.4**(2)	10(3)	5.7(0.8)	8.1(13)	∞ *1000*	0/15
LHD-2xDefa	**1.5**(1)	**2.6**(0.3)	**1.9**(0.5)	8.1(4)	∞ *1000*	0/15
RAND-2xDef	**2.2**(3)	**2.8**(2)	**2.1**(0.7)	6.4(6)	∞ *1000*	0/15
RF1-CMAES	**2.0**(1)	**2.6**(2)	**2.3**(0.9)	**1.5**(0.7)	54(73)	1/15
RF5-CMAES	**1.8**(2)	**1.9**(0.9)	**1.8**(0.4)	7.0(7)	∞ *5034*	0/15
Sifeg	**1.4**(2)	**1.5**(0.6)	**3.0**(0.4)	48(6)	∞ *2e5*	0/15
Sif	**1.4**(1)	**1.5**(0.5)	**1.1**(0.6)	21(40)	∞ *2e5*	0/15
Srr	**1.4**(2)	**1.5**(0.7)	**1.1**(0.3)	12(4)	∞ *2e5*	0/15

Table 9: 20-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	*4.0e+4*:19	*2.5e+4*:35	*4.0e+3*:67	*2.5e+2*:231	*1.6e+1*:1470	15/15
BSifeg	**3.3**(0.4)	**2.0**(0.3)	**1.2**(0.1)	4.2(6)	65(107)	12/15
BSif	**3.3**(0.7)	**2.0**(0.4)	**1.2**(0.1)	4.9(14)	518(422)	3/15
BSqi	**3.1**(0.5)	**1.9**(0.4)	**1.2**(0.1)	**1.6**(2)	32(16)	7/8
BSrr	**3.3**(1)	**2.0**(0.3)	**1.2**(0.1)	3.7(0.2)	75(77)	10/15
CMA-CSA	7.3(5)	5.8(2)	4.9(1)	**2.9**(0.5)	**2.4**(0.7)	15/15
CMA-MSR	8.7(3)	5.9(1)	5.0(1.0)	**2.9**(0.7)	**2.5**(0.9)	15/15
CMA-TPA	7.2(2)	4.3(1)	3.8(0.8)	**2.2**(0.5)	**1.9**(0.5)	15/15
GP1-CMAES	5.5(2)	3.4(0.7)	3.1(0.8)	**2.3**(0.3)	3.0(2)	12/15
GP5-CMAES	4.4(0.6)	**2.6**(0.2)	**2.2**(0.4)	11(22)	8.2(9)	5/15
IPOPCMAv3p	6.3(3)	4.8(1.0)	4.4(0.3)	**2.8**(1.0)	**2.9**(2)	12/15
LHD-10xDef	23(0.2)	12(0.1)	6.7(0.4)	3.8(2)	∞ *1000*	0/15
LHD-2xDefa	5.5(0.7)	3.2(0.6)	3.1(0.7)	**2.8**(0.7)	∞ *1000*	0/15
RAND-2xDef	5.4(0.5)	3.2(0.4)	**2.8**(0.4)	**2.5**(0.7)	∞ *1000*	0/15
RF1-CMAES	5.8(1)	4.0(0.6)	4.0(1)	**2.7**(0.8)	25(19)	2/15
RF5-CMAES	5.8(2)	3.7(2)	4.2(2)	96(108)	∞ *5006*	0/15
Sifeg	3.3(1)	**2.0**(0.3)	**1.2**(0.1)	**0.98**(0.6)	49(53)	14/15
Sif	3.3(0.6)	**2.0**(0.1)	**1.2**(0.1)	**1.1**(0.8)	75(113)	12/15
Srr	3.3(0.7)	**2.0**(0.2)	**1.2**(0.1)	**0.95**(0.2)	77(55)	10/15

Table 10: 20-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+2*:357	*6.3e+1*:560	*4.0e+1*:684	*2.5e+1*:756	*1.0e+1*:1716	15/15
BSifeg	19(39)	66(100)	59(81)	57(78)	∞ *2e5*	0/15
BSif	73(20)	115(91)	147(295)	181(258)	∞ *2e5*	0/15
BSqi	7.4(3)	26(26)	25(45)	27(24)	∞ *2e5*	0/5
BSrr	16(11)	65(100)	57(4)	55(107)	∞ *2e5*	0/15
CMA-CSA	**2.2**(1)	**2.0**(4)	**1.8**(2)	**1.9**(0.1)	**3.8**(0.6)	15/15
CMA-MSR	**2.0**(0.2)	**2.4**(1)	**2.1**(4)	**2.1**(1)	**3.8**(0.9)	15/15
CMA-TPA	**1.5**(0.3)	**1.0**(0.3)	**0.97**(0.2)	**1.0**(0.4)	**3.8**(0.7)	15/15
GP1-CMAES	**2.1**(0.8)	**1.6**(0.5)	**1.4**(0.4)	**1.5**(0.4)	∞ *5006*	0/15
GP5-CMAES	12(7)	12(14)	13(9)	12(24)	∞ *5020*	0/15
IPOPCMAv3p	3.6(4)	3.6(0.9)	3.1(5)	**3.0**(3)	∞ *5006*	0/15
LHD-10xDef	8.0(4)	27(35)	∞	∞	∞ *1000*	0/15
LHD-2xDefa	5.3(6)	13(8)	22(21)	∞	∞ *1000*	0/15
RAND-2xDef	5.3(6)	13(23)	22(24)	∞	∞ *1000*	0/15
RF1-CMAES	3.8(1)	5.8(8)	5.3(4)	5.8(4)	∞ *5006*	0/15
RF5-CMAES	∞	∞	∞	∞	∞ *5006*	0/15
Sifeg	3.4(5)	19(28)	16(45)	15(6)	∞ *2e5*	0/15
Sif	16(6)	29(76)	26(29)	25(45)	∞ *2e5*	0/15
Srr	**1.9**(3)	16(3)	14(27)	13(19)	∞ *2e5*	0/15

Table 11: 20-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{10} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f10	*1.6e+6*:15	*1.0e+6*:27	*4.0e+5*:70	*6.3e+4*:231	*4.0e+3*:1015	15/15
BSifeg	**2.4**(2)	**2.0**(0.9)	**1.6**(1)	479(636)	∞ *2e5*	0/15
BSif	**2.4**(2)	**2.0**(1)	**1.6**(0.4)	494(497)	∞ *2e5*	0/15
BSqi	**2.4**(2)	**2.0**(1)	**1.5**(0.6)	221(501)	∞ *2e5*	0/15
BSrr	**2.4**(2)	**2.0**(1)	**1.5**(0.6)	204(583)	∞ *9e4*	0/15
CMA-CSA	7.8(4)	8.3(8)	7.5(4)	5.8(1)	**3.0**(0.5)	15/15
CMA-MSR	8.1(4)	5.8(3)	4.0(2)	**3.4**(1)	**2.7**(0.7)	15/15
CMA-TPA	7.5(6)	6.5(2)	4.7(1)	**3.3**(1)	**2.6**(0.6)	15/15
GP1-CMAES	5.4(4)	5.4(3)	4.2(0.8)	3.5(0.7)	**2.3**(0.9)	15/15
GP5-CMAES	4.8(1)	3.6(2)	3.1(0.7)	**2.0**(0.4)*	**1.1**(0.4)	15/15
IPOPCMAv3p	4.8(5)	4.3(4)	5.1(2)	5.0(1.0)	3.2(0.7)	15/15
LHD-10xDef	15(10)	12(8)	7.5(0.9)	21(19)	∞ *1000*	0/15
LHD-2xDefa	6.7(4)	4.8(3)	3.3(2)	5.8(4)	∞ *1000*	0/15
RAND-2xDef	6.4(2)	4.6(3)	3.6(1)	6.9(9)	∞ *1000*	0/15
RF1-CMAES	5.9(3)	5.3(3)	4.1(0.9)	3.9(1)	74(110)	1/15
RF5-CMAES	5.1(4)	5.2(3)	10(19)	145(184)	∞ *5006*	0/15
Sifeg	**2.5**(2)	**2.0**(2)	**1.9**(1)	21(30)	∞ *1e5*	0/15
Sif	**2.5**(1.0)	**2.0**(1)	**1.7**(1)	32(15)	∞ *1e5*	0/15
Srr	**2.5**(2)	**2.0**(2)	**1.6**(1.0)	14(18)	∞ *7e4*	0/15

Table 12: 20-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	*4.0e+4*:11	*2.5e+3*:27	*1.6e+2*:313	*1.0e+2*:481	*1.0e+1*:1002	15/15
BSifeg	**1.2**(1.0)	**1.3**(0.5)	796(1249)	4369(6014)	∞ *1e5*	0/15
BSif	**1.2**(1)	**1.3**(0.6)	1191(1314)	∞	∞ *2e5*	0/15
BSqi	**1.2**(0.5)	**1.3**(1)	462(666)	2574(3002)	∞ *2e5*	0/15
BSrr	**1.2**(0.8)	**1.3**(1)	1822(2282)	2506(3240)	∞ *8e4*	0/15
CMA-CSA	**2.2**(2)	3.3(4)	12(1)	**8.1**(1.0)	**4.6**(0.3)	15/15
CMA-MSR	**2.0**(2)	**2.9**(2)	**9.2**(2)	**6.9**(1)	**4.7**(0.5)	15/15
CMA-TPA	**2.2**(2)	**2.5**(2)	10(1)	**7.2**(0.8)	**4.5**(0.3)	15/15
GP1-CMAES	**1.3**(1.0)	**1.8**(3)	14(14)	15(5)	∞ *5006*	0/15
GP5-CMAES	**1.8**(1)	**2.3**(2)	**5.2**(3)	17(12)	∞ *5008*	0/15
IPOPCMAv3p	**1.5**(0.9)	**2.5**(2)	52(91)	74(55)	∞ *5006*	0/15
LHD-10xDef	**2.6**(2)	4.0(4)	22(30)	∞	∞ *1000*	0/15
LHD-2xDefa	**1.9**(1)	**2.6**(3)	22(10)	∞	∞ *1000*	0/15
RAND-2xDef	3.0(1)	**2.5**(3)	∞	∞	∞ *1000*	0/15
RF1-CMAES	**1.5**(0.6)	3.2(3)	**8.6**(5)	12(7)	∞ *5006*	0/15
RF5-CMAES	**1.9**(2)	**2.8**(2)	105(152)	146(174)	∞ *5008*	0/15
Sifeg	**1.2**(1)	**1.3**(1)	478(717)	3784(2685)	∞ *1e5*	0/15
Sif	**1.2**(1)	**1.3**(0.5)	668(1268)	∞	∞ *1e5*	0/15
Srr	**1.2**(1)	**1.3**(1)	529(1045)	2003(2431)	∞ *7e4*	0/15

Table 13: 20-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*:23	*6.3e+7*:39	*2.5e+7*:76	*4.0e+6*:209	*1.0e+1*:1042	15/15
BSifeg	3.5(0.9)	11(0.6)	16(38)	25(62)	∞ *1e5*	0/15
BSif	**3.4**(2)	9.3(28)	29(130)	39(27)	∞ *1e5*	0/15
BSqi	**3.2**(0.7)	12(36)	47(64)	63(231)	173(106)	9/15
BSrr	**3.3**(2)	11(33)	89(48)	69(211)	677(608)	2/15
CMA-CSA	5.4(2)	4.5(2)	**3.7**(0.7)	**2.3**(0.4)	**3.6**(3)	15/15
CMA-MSR	6.2(2)	4.9(2)	3.8(0.8)	**2.5**(0.2)	3.7(3)	15/15
CMA-TPA	7.2(3)	5.2(1)	3.8(0.6)	**2.0**(0.5)	3.8(4)	15/15
GP1-CMAES	4.7(3)	3.9(3)	3.7(0.9)	**2.8**(0.9)	**2.4**(4)	13/15
GP5-CMAES	18(82)	22(40)	38(31)	74(142)	21(12)	3/15
IPOPCMAv3p	5.1(3)	4.7(0.4)	3.8(1)	**2.5**(0.3)	3.8(2)	11/15
LHD-10xDef	17(7)	12(0.7)	8.9(4)	17(14)	∞ *1000*	0/15
LHD-2xDefa	4.6(2)	4.1(0.7)	3.8(0.8)	3.2(1)	∞ *1000*	0/15
RAND-2xDef	5.0(1)	4.0(2)	**3.7**(1)	3.8(3)	∞ *1000*	0/15
RF1-CMAES	3.9(2)	3.9(1)	**3.1**(0.5)	**1.8**(0.1)	**3.0**(2)	12/15
RF5-CMAES	5.2(2)	4.6(1)	9.4(2)	39(108)	∞ *5006*	0/15
Sifeg	3.7(0.6)	**3.0**(0.3)	8.1(20)	26(19)	∞ *4e4*	0/15
Sif	3.7(3)	**3.2**(4)	12(37)	23(12)	∞ *5e4*	0/15
Srr	4.6(0.6)	**3.8**(3)	28(19)	20(28)	∞ *4e4*	0/15

Table 14: 20-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	*1.6e+3*:28	*1.0e+3*:64	*6.3e+2*:79	*4.0e+1*:211	*2.5e+0*:1724	15/15
BSifeg	**2.2**(0.3)	**1.3**(0.1)	**1.5**(0.1)	125(442)	99(70)	11/15
BSif	**2.2**(0.7)	**1.3**(0.2)	**1.5**(0.1)	1842(1760)	∞ *2e5*	0/15
BSqi	**2.2**(0.6)	**1.3**(0.2)	**1.4**(0.2)	71(9)	106(67)	9/15
BSrr	**2.2**(0.5)	**1.3**(0.1)	**1.4**(0.2)	67(143)	88(62)	11/15
CMA-CSA	4.3(2)	3.6(0.7)	4.7(0.8)	**5.4**(0.4)	**2.8**(2)	15/15
CMA-MSR	5.6(1)	4.3(1)	4.9(0.4)	6.1(0.3)	**2.8**(1)	15/15
CMA-TPA	3.9(1.0)	3.0(0.9)	3.7(0.6)	**5.3**(0.3)	**4.0**(3)	15/15
GP1-CMAES	**2.5**(0.9)	**2.4**(0.4)	**2.9**(0.5)	49(26)	42(52)	1/15
GP5-CMAES	**2.4**(0.4)	**1.5**(0.3)	**1.6**(0.3)	**2.9**(0.8)$^{\star 2}$	4.5(14)	7/15
IPOPCMAv3p	3.3(2)	3.9(1)	5.5(0.8)	8.0(3)	9.1(7)	4/15
LHD-10xDef	15(0.1)	6.9(0.1)	6.3(0.2)	8.5(7)	∞ *1000*	0/15
LHD-2xDefa	3.3(1)	**2.2**(0.2)	**2.8**(2)	9.4(9)	∞ *1000*	0/15
RAND-2xDef	3.5(0.5)	**2.2**(0.3)	**2.9**(0.5)	7.1(6)	∞ *1000*	0/15
RF1-CMAES	3.2(1)	**3.0**(1)	3.9(1)	7.1(1)	7.3(5)	5/15
RF5-CMAES	3.4(2)	**3.0**(0.7)	4.2(1)	343(659)	∞ *5006*	0/15
Sifeg	**2.2**(0.7)	**1.3**(0.1)	**1.4**(0.1)	19(41)	43(29)	15/15
Sif	**2.2**(0.8)	**1.3**(0.1)	**1.4**(0.1)	23(84)	81(59)	11/15
Srr	**2.2**(0.6)	**1.3**(0.1)	**1.4**(0.1)	16(77)	64(111)	13/15

Table 15: 20-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{14} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f14	*2.5e+1*:15	*1.6e+1*:42	*1.0e+1*:75	*1.6e+0*:219	*6.3e-4*:1106	15/15
BSifeg	3.3(2)	**1.7**(0.4)	**1.5**(0.6)	10(9)	∞ *2e5*	0/15
BSif	3.3(2)	**1.7**(0.7)	**1.5**(0.8)	9.1(10)	∞ *2e5*	0/15
BSqi	3.9(1)	**2.0**(0.3)	**1.5**(0.3)	4.7(3)	∞ *2e5*	0/4
BSrr	3.3(2)	**1.7**(1.0)	**1.4**(0.6)	6.4(5)	∞ *2e5*	0/15
CMA-CSA	9.1(7)	4.8(2)	4.2(1)	**2.9**(0.3)	**3.1**(0.4)	15/15
CMA-MSR	8.7(2)	4.7(2)	4.2(0.8)	**2.7**(0.4)	**2.7**(0.2)	15/15
CMA-TPA	8.9(8)	4.8(3)	3.5(2)	**2.4**(0.3)	**2.6**(0.4)	15/15
GP1-CMAES	7.9(5)	3.9(1)	**3.0**(0.7)	**2.3**(0.6)	4.6(2)	13/15
GP5-CMAES	5.7(2)	**2.6**(0.6)	**2.1**(0.3)	**1.7**(0.5)	67(23)	1/15
IPOPCMAv3p	10(3)	4.6(1)	3.7(1)	**2.9**(0.3)	3.9(0.4)	15/15
LHD-10xDef	25(7)	11(0.8)	6.9(0.4)	7.7(6)	∞ *1000*	0/15
LHD-2xDefa	8.1(2)	3.7(1)	3.1(1)	3.8(3)	∞ *1000*	0/15
RAND-2xDef	8.8(2)	4.1(0.8)	3.3(0.9)	7.3(6)	∞ *1000*	0/15
RF1-CMAES	7.1(3)	4.4(2)	3.5(0.9)	3.1(0.9)	33(32)	2/15
RF5-CMAES	6.5(7)	3.9(2)	3.7(1)	153(104)	∞ *5006*	0/15
Sifeg	**3.3**(2)	**1.8**(0.5)	**1.3**(0.4)	**1.4**(0.7)	∞ *2e5*	0/15
Sif	**3.3**(2)	**1.8**(0.8)	**1.3**(0.4)	**1.5**(0.9)	∞ *2e5*	0/15
Srr	**3.3**(2)	**1.8**(0.7)	**1.3**(0.5)	**1.1**(0.7)	∞ *2e5*	0/15

Table 16: 20-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	*6.3e+2*:15	*4.0e+2*:67	*2.5e+2*:292	*1.6e+2*:846	*1.0e+2*:1671	15/15
BSifeg	**2.2**(2)	18(0.6)	274(399)	525(570)	1563(1672)	1/15
BSif	**2.2**(2)	76(563)	296(322)	943(1318)	1581(3030)	1/15
BSqi	**2.1**(1)	15(0.3)	224(291)	346(493)	∞ *2e5*	0/11
BSrr	**2.2**(1)	5.7(33)	272(279)	450(534)	1391(3523)	1/15
CMA-CSA	4.9(2)	**2.5**(0.6)	**1.1**(0.2)	**0.86**(0.2)	**1.5**(0.6)	15/15
CMA-MSR	6.2(4)	**2.8**(1)	**1.1**(0.3)	**0.68**(0.1)	**0.44**(0.1)	15/15
CMA-TPA	6.1(3)	**2.3**(0.4)	**0.87**(0.3)	**0.70**(0.2)	**0.72**(0.1)	15/15
GP1-CMAES	3.6(2)	**1.9**(0.8)	**0.81**(0.2)	**0.78**(0.4)	**0.64**(0.2)	15/15
GP5-CMAES	3.7(3)	**1.5**(0.3)	**0.60**(0.2)	**2.2**(8)	3.9(3)	7/15
IPOPCMAv3p	4.5(3)	**2.6**(1)	**1.2**(0.3)	**1.2**(0.4)	**1.4**(2)	14/15
LHD-10xDef	16(14)	6.4(0.1)	**1.9**(0.4)	**1.1**(0.2)	**1.2**(0.5)	7/15
LHD-2xDefa	5.5(2)	**2.0**(0.9)	**1.1**(0.9)	**1.5**(1)	8.7(6)	1/15
RAND-2xDef	5.1(2)	**1.9**(0.5)	**1.1**(0.5)	**1.6**(0.4)	8.8(6)	1/15
RF1-CMAES	4.5(2)	**2.2**(1)	**0.93**(0.2)	**0.92**(0.2)	**0.70**(0.1)	15/15
RF5-CMAES	3.6(2)	**1.9**(0.4)	**0.95**(0.5)	**1.7**(2)	6.5(6)	5/15
Sifeg	**2.2**(1)	46(168)	49(91)	126(362)	774(620)	2/15
Sif	**2.2**(2)	52(0.5)	74(144)	170(170)	703(1812)	2/15
Srr	**2.2**(2)	52(190)	72(135)	100(124)	352(371)	4/15

Table 17: 20-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{16} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f16	*4.0e+1*:26	*2.5e+1*:127	*1.6e+1*:540	*1.6e+1*:540	*1.0e+1*:1384	15/15
BSifeg	**1.7**(1)	**0.92**(0.2)	44(119)	44(89)	123(354)	9/15
BSif	**1.7**(1)	**0.92**(0.2)	42(89)	42(88)	178(188)	8/15
BSqi	**1.7**(1)	**0.93**(0.2)	21(42)	21(0.7)	98(223)	10/13
BSrr	**1.7**(0.9)	**0.93**(0.2)	40(46)	40(134)	119(89)	10/15
CMA-CSA	7.8(10)	16(6)	4.3(2)	4.3(1)	**1.9**(0.6)	15/15
CMA-MSR	5.2(2)	**2.1**(0.4)	**0.81**(0.3)	**0.81**(0.4)	**0.80**(1)	15/15
CMA-TPA	4.5(8)	8.4(6)	**2.2**(1)	**2.2**(1)	**1.2**(0.7)	15/15
GP1-CMAES	3.3(2)	4.3(2)	**1.4**(0.3)	**1.4**(0.2)	**0.90**(0.1)	14/15
GP5-CMAES	3.4(3)	**1.6**(0.6)	**0.54**(0.2)	**0.54**(0.2)	**0.57**(0.2)	15/15
IPOPCMAv3p	3.4(4)	10(4)	3.2(1)	3.2(1)	**1.4**(0.4)	15/15
LHD-10xDef	6.5(7)	5.4(2)	4.8(3)	4.8(3)	3.4(4)	3/15
LHD-2xDefa	3.2(3)	3.6(2)	**2.6**(3)	**2.6**(3)	5.1(8)	2/15
RAND-2xDef	3.8(3)	6.4(8)	5.0(8)	5.0(8)	11(16)	1/15
RF1-CMAES	**2.8**(4)	4.6(3)	**1.6**(0.5)	**1.6**(0.4)	**0.79**(0.2)	15/15
RF5-CMAES	4.2(4)	**2.2**(2)	**0.83**(0.2)	**0.83**(0.3)	**1.0**(3)	13/15
Sifeg	**1.9**(2)	**1.5**(0.6)	10(9)	10(7)	15(22)	15/15
Sif	**1.9**(2)	**1.6**(1)	6.1(6)	6.1(6)	12(12)	15/15
Srr	**1.9**(1)	**1.8**(3)	3.3(12)	3.3(0.9)	14(10)	15/15

Table 18: 20-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*:11	*1.0e+1*:63	*6.3e+0*:305	*4.0e+0*:468	*1.0e+0*:1030	15/15
BSifeg	22(2)	493(0.8)	1042(852)	∞	∞ *2e5*	0/15
BSif	35(3)	488(791)	2637(4581)	∞	∞ *2e5*	0/15
BSqi	**2.6**(1)	**1.7**(0.6)	346(328)	∞	∞ *2e5*	0/3
BSrr	27(3)	476(1)	1359(2006)	∞	∞ *2e5*	0/15
CMA-CSA	7.5(5)	**3.0**(2)	**1.1**(0.2)	**1.0**(0.2)	**1.0**(0.2)	15/15
CMA-MSR	7.9(3)	**2.7**(0.6)	**0.90**(0.2)	**1.0**(0.2)	6.5(3)	15/15
CMA-TPA	8.8(7)	**2.7**(0.7)	**0.93**(0.1)	**0.95**(0.3)	**1.4**(0.4)	15/15
GP1-CMAES	**2.4**(3)	**1.4**(0.8)	**0.73**(0.3)	**0.79**(0.3)	3.4(10)	10/15
GP5-CMAES	3.3(2)	**1.6**(0.8)	**0.79**(0.4)	**0.87**(0.6)	11(27)	5/15
IPOPCMAv3p	**2.8**(2)	**2.0**(2)	**0.89**(0.3)	**0.95**(0.3)	**0.99**(0.2)	15/15
LHD-10xDef	12(9)	7.3(2)	**2.2**(0.4)	5.7(2)	∞ *1000*	0/15
LHD-2xDefa	3.3(4)	**2.6**(0.8)	**1.4**(2)	10(8)	∞ *1000*	0/15
RAND-2xDef	5.0(4)	**2.7**(2)	**2.1**(2)	31(38)	∞ *1000*	0/15
RF1-CMAES	3.1(3)	**1.9**(0.7)	**0.78**(0.3)	**0.83**(0.3)	4.2(10)	9/15
RF5-CMAES	4.0(2)	**2.7**(0.8)	3.6(5)	17(13)	∞ *5006*	0/15
Sifeg	3.8(6)	230(797)	797(1392)	5983(5449)	∞ *2e5*	0/15
Sif	3.6(5)	7.1(18)	750(822)	5964(6604)	∞ *2e5*	0/15
Srr	3.6(3)	3.8(6)	979(1795)	5963(5002)	∞ *2e5*	0/15

Table 19: 20-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{18} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f18	*4.0e+1*:116	*2.5e+1*:252	*1.6e+1*:430	*1.0e+1*:621	*4.0e+0*:1090	15/15
BSifeg	5.9(13)	468(752)	5872(6811)	∞	∞ *2e5*	0/14
BSif	123(7)	758(760)	∞	∞	∞ *2e5*	0/15
BSqi	**0.80**(0.3)	368(889)	∞	∞	∞ *2e5*	0/4
BSrr	26(0.2)	649(989)	∞	∞	∞ *2e5*	0/15
CMA-CSA	**1.3**(0.3)	**1.0**(0.3)	**0.96**(0.1)	**0.96**(0.3)	**0.96**(0.3)	15/15
CMA-MSR	**1.3**(0.5)	**1.0**(0.3)	**0.97**(0.6)	**2.8**(0.2)	4.8(13)	15/15
CMA-TPA	**1.3**(0.7)	**1.0**(0.7)	**0.92**(0.3)	**1.6**(3)	**1.3**(0.3)	15/15
GP1-CMAES	**0.81**(0.3)	**0.84**(0.3)	**0.85**(0.3)	**0.93**(0.5)	5.2(6)	8/15
GP5-CMAES	**0.95**(0.9)	**0.85**(0.6)	**1.7**(3)	**2.8**(4)	19(22)	3/15
IPOPCMAv3p	**0.94**(0.7)	**1.1**(0.2)	**1.1**(0.4)	**1.1**(0.4)	**1.1**(0.2)	15/15
LHD-10xDef	3.5(0.5)	**2.7**(0.8)	5.2(3)	∞	∞ *1000*	0/15
LHD-2xDefa	**1.00**(0.6)	**1.6**(0.7)	11(6)	∞	∞ *1000*	0/15
RAND-2xDef	**0.99**(0.4)	**1.6**(1)	16(11)	∞	∞ *1000*	0/15
RF1-CMAES	**0.87**(0.4)	**0.82**(0.2)	**0.87**(0.2)	**1.0**(0.2)	10(9)	5/15
RF5-CMAES	**1.3**(0.4)	**2.7**(1)	5.4(3)	53(99)	∞ *5006*	0/15
Sifeg	48(9)	217(303)	1821(1148)	∞	∞ *2e5*	0/15
Sif	91(59)	367(370)	1063(2314)	4564(4318)	∞ *2e5*	0/15
Srr	30(8)	132(7)	801(1299)	∞	∞ *2e5*	0/15

Table 20: 20-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*:2.5e5	*1.0e-1*:3.4e5	*6.3e-2*:3.4e5	*4.0e-2*:3.4e5	*2.5e-2*:3.4e5	3/15
BSifeg	∞	∞	∞	∞	∞ *2e5*	0/15
BSif	∞	∞	∞	∞	∞ *2e5*	0/15
BSqi	∞	∞	∞	∞	∞ *2e5*	0/8
BSrr	∞	∞	∞	∞	∞ *2e5*	0/15
CMA-CSA	**0.89**(0.5)	**0.82**(0.5)	**0.87**(0.5)	**2.9**(2)	**5.6**(4)	11/15
CMA-MSR	**1.2**(0.5)	**1.2**(0.4)	**1.7**(0.9)	**3.0**(5)	**15**(28)	5/15
CMA-TPA	**1.3**(0.6)	**1.6**(0.6)	**2.3**(0.9)	**3.4**(2)	**7.9**(8)	9/15
GP1-CMAES	∞	∞	∞	∞	∞ *5006*	0/15
GP5-CMAES	∞	∞	∞	∞	∞ *5020*	0/15
IPOPCMAv3p	∞	∞	∞	∞	∞ *5008*	0/15
LHD-10xDef	∞	∞	∞	∞	∞ *1000*	0/15
LHD-2xDefa	∞	∞	∞	∞	∞ *1000*	0/15
RAND-2xDef	∞	∞	∞	∞	∞ *1000*	0/15
RF1-CMAES	∞	∞	∞	∞	∞ *5008*	0/15
RF5-CMAES	∞	∞	∞	∞	∞ *5034*	0/15
Sifeg	∞	∞	∞	∞	∞ *2e5*	0/15
Sif	∞	∞	∞	∞	∞ *2e5*	0/15
Srr	∞	∞	∞	∞	∞ *2e5*	0/15

Table 21: 20-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	*1.6e+4*:38	*1.0e+4*:42	*2.5e+2*:62	*2.5e+0*:250	*1.6e+0*:2536	15/15
BSifeg	**1.8**(0.6)	**1.8**(0.6)	**1.6**(0.6)	**2.1**(4)	3.2(0.8)	15/15
BSif	**1.8**(1)	**1.8**(0.3)	**1.6**(0.8)	**1.6**(0.8)	**1.4**(1)	15/15
BSqi	**1.1**(1)	**1.5**(1)	**1.5**(0.5)	**1.0**(0.3)	**1.3**(0.3)	4/4
BSrr	**1.8**(0.5)	**1.8**(1)	**1.6**(0.5)	**1.1**(0.5)	4.8(14)	15/15
CMA-CSA	3.2(1)	3.7(0.6)	5.5(2)	7.6(9)	18(13)	15/15
CMA-MSR	3.8(0.8)	4.2(1)	5.9(1)	3.2(0.5)	4477(3623)	2/15
CMA-TPA	3.4(0.7)	3.8(0.4)	4.3(0.9)	5.8(0.6)	432(576)	10/15
GP1-CMAES	**2.8**(1)	3.2(1)	3.6(0.2)	4.0(2)	∞ *5006*	0/15
GP5-CMAES	**2.3**(0.5)	**2.2**(0.2)	**2.5**(0.4)	284(618)	∞ *5022*	0/15
IPOPCMAv3p	3.5(1)	4.3(2)	6.1(0.9)	5.9(0.5)	6.2(6)	4/15
LHD-10xDef	11(2)	10(0.2)	8.7(0.5)	7.9(10)	∞ *1000*	0/15
LHD-2xDefa	3.3(0.6)	3.3(0.6)	4.4(2)	**2.8**(2)	5.7(4)	1/15
RAND-2xDef	3.1(0.6)	3.1(0.7)	4.8(3)	4.9(2)	∞ *1000*	0/15
RF1-CMAES	3.5(0.7)	3.9(0.6)	5.2(1)	3.4(1)	6.1(7)	4/15
RF5-CMAES	**2.6**(0.8)	**2.9**(1)	6.6(4)	134(150)	∞ *5006*	0/15
Sifeg	**1.8**(0.6)	**1.8**(0.3)	**1.9**(0.3)	**0.75**(0.1)	**0.55**(0.6)	15/15
Sif	**1.8**(1)	**1.8**(0.6)	**1.9**(0.7)	**0.82**(0.4)	**0.64**(0.5)	15/15
Srr	**1.8**(0.6)	**1.8**(0.7)	**1.9**(0.4)	**0.71**(0.1)	**0.51**(0.3)	15/15

Table 22: 20-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	*6.3e+1*:36	*4.0e+1*:77	*4.0e+1*:77	*1.6e+1*:456	*4.0e+0*:1094	15/15
BSifeg	**2.0**(0.1)	**1.2**(0.2)	**1.2**(0.2)	87(219)	98(149)	11/15
BSif	**2.0**(0.3)	**1.2**(0.2)	**1.2**(0.2)	93(158)	146(109)	9/15
BSqi	**2.1**(0.1)	**1.4**(0.5)	**1.4**(0.5)	187(401)	79(117)	3/4
BSrr	**2.0**(0.1)	**1.2**(0.1)	**1.2**(0.6)	80(44)	162(230)	8/15
CMA-CSA	5.6(2)	3.4(0.9)	3.4(0.5)	**2.5**(3)	99(520)	14/15
CMA-MSR	5.0(1)	3.1(0.8)	3.1(0.9)	**1.6**(0.2)	234(2)	13/15
CMA-TPA	4.9(2)	3.3(1)	3.3(1)	17(4)	198(577)	13/15
GP1-CMAES	3.8(2)	6.3(0.4)	6.3(0.6)	**1.8**(5)	3.4(1)	10/15
GP5-CMAES	3.2(1.0)	**2.9**(0.4)	**2.9**(8)	**1.8**(2)	7.7(10)	6/15
IPOPCMAv3p	5.8(5)	5.5(4)	5.5(2)	3.2(4)	7.8(7)	6/15
LHD-10xDef	13(0.8)	6.6(0.6)	6.6(0.9)	**2.0**(2)	**2.4**(4)	5/15
LHD-2xDefa	3.6(0.6)	**2.6**(1)	**2.6**(3)	**0.88**(0.3)	**1.2**(0.8)	9/15
RAND-2xDef	3.4(0.6)	**1.9**(0.4)	**1.9**(0.5)	**0.46**(0.1)	**0.57**(0.5)	12/15
RF1-CMAES	5.2(2)	3.7(1)	3.7(1)	5.5(6)	8.0(10)	6/15
RF5-CMAES	7.6(2)	6.1(2)	6.1(1)	8.7(6)	14(7)	4/15
Sifeg	**2.0**(0.3)	**1.3**(0.3)	**1.3**(0.5)	65(164)	87(267)	12/15
Sif	**2.0**(0.5)	**1.3**(0.5)	**1.3**(0.1)	106(141)	136(259)	9/15
Srr	**2.0**(0.3)	**1.3**(0.6)	**1.3**(0.3)	67(103)	160(92)	8/15

Table 23: 20-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{22} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f22	*6.3e+1*:45	*4.0e+1*:68	*4.0e+1*:68	*1.6e+1*:231	*6.3e+0*:1219	15/15
BSifeg	**1.6**(0.2)	4.4(24)	4.4(12)	166(86)	156(297)	8/15
BSif	**1.6**(0.3)	6.1(0.3)	6.1(0.1)	196(144)	225(246)	7/15
BSqi	**1.6**(0.0)	6.6(14)	6.6(0.0)	228(445)	230(456)	2/5
BSrr	**1.6**(0.3)	3.5(9)	3.5(9)	167(263)	117(123)	10/15
CMA-CSA	4.6(1)	9.1(2)	9.1(17)	38(98)	230(458)	12/15
CMA-MSR	5.8(1)	14(18)	14(35)	7.9(11)	206(329)	13/15
CMA-TPA	4.0(1)	6.8(1)	6.8(1)	327(0.8)	428(431)	10/15
GP1-CMAES	12(29)	17(17)	17(20)	12(11)	**2.7**(1)	10/15
GP5-CMAES	**2.5**(0.3)	3.6(6)	3.6(12)	4.8(8)	**2.0**(7)	11/15
IPOPCMAv3p	4.6(2)	10(2)	10(2)	6.7(6)	4.6(5)	8/15
LHD-10xDef	10(0.4)	7.8(2)	7.8(2)	5.4(5)	**2.1**(1)	5/15
LHD-2xDefa	3.4(0.6)	3.3(3)	3.3(2)	**2.0**(3)	**1.2**(2)	8/15
RAND-2xDef	3.3(0.6)	3.2(2)	3.2(0.3)	**1.9**(2)	**1.1**(1)	8/15
RF1-CMAES	4.6(3)	5.2(3)	5.2(1)	**4.6**(6)	**3.0**(0.4)	10/15
RF5-CMAES	5.8(3)	11(38)	11(22)	11(17)	5.7(4)	7/15
Sifeg	**1.6**(0.4)	**1.7**(4)	**1.7**(0.2)	230(351)	159(287)	9/15
Sif	**1.6**(0.1)	**1.9**(6)	**1.9**(0.2)	118(156)	125(125)	10/15
Srr	**1.6**(0.4)	**1.5**(0.2)	**1.5**(3)	170(497)	133(101)	9/15

Table 24: 20-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	*6.3e+0*:29	*4.0e+0*:118	*2.5e+0*:306	*2.5e+0*:306	*1.0e+0*:1614	15/15
BSifeg	**2.3**(2)	**1.4**(0.4)	**1.9**(2)	**1.9**(2)	60(52)	14/15
BSif	**2.3**(2)	**1.4**(0.4)	**2.8**(10)	**2.8**(5)	37(39)	15/15
BSrr	**2.3**(2)	**1.4**(0.5)	**2.5**(2)	**2.5**(5)	54(25)	14/15
CMA-CSA	6.7(4)	8.1(5)	37(35)	37(45)	93(10)	15/15
CMA-MSR	4.1(3)	3.9(2)	4.3(0.6)	4.3(10)	**2.0**(6)	15/15
CMA-TPA	5.2(3)	12(10)	34(47)	34(50)	23(38)	15/15
GP1-CMAES	**2.1**(2)	6.4(4)	5.8(5)	5.8(2)	**1.6**(0.9)	14/15
GP5-CMAES	**2.2**(3)	**2.9**(2)	**1.9**(4)	**1.9**(0.2)	**0.84**(0.8)	15/15
IPOPCMAv3p	**1.5**(2)	6.7(4)	75(126)	75(72)	∞ *5020*	0/15
LHD-10xDef	**2.3**(3)	8.2(16)	48(43)	48(25)	∞ *1000*	0/15
LHD-2xDefa	**1.6**(2)	5.6(7)	23(20)	23(22)	∞ *1000*	0/15
RAND-2xDef	**2.1**(0.8)	6.2(5)	∞	∞	∞ *1000*	0/15
RF1-CMAES	**1.3**(0.9)	3.4(6)	244(139)	244(275)	∞ *5010*	0/15
RF5-CMAES	**1.8**(3)	4.5(6)	113(313)	113(144)	∞ *5086*	0/15
Sifeg	**2.3**(2)	4.7(3)	4.2(3)	4.2(2)	6.5(5)	15/15
Sif	**2.3**(2)	4.7(2)	4.4(2)	4.4(2)	11(9)	15/15
Srr	**2.3**(1)	4.8(3)	3.9(1)	3.9(2)	5.2(8)	15/15

Table 25: 20-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{24} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f24	*2.5e+2*:208	*1.6e+2*:918	*1.0e+2*:6628	*6.3e+1*:9885	*4.0e+1*:31629	15/15
BSifeg	61(21)	73(139)	56(58)	∞	∞ *2e5*	0/15
BSif	76(70)	187(223)	196(301)	∞	∞ *2e5*	0/15
BSrr	63(152)	78(66)	75(81)	∞	∞ *2e5*	0/15
CMA-CSA	**1.5**(0.2)	**1.4**(0.2)	**1.4**(0.5)	**0.99**(1)	**0.87**(0.8)	15/15
CMA-MSR	**1.3**(0.3)	**0.73**(0.1)	**0.21**(0.3)	**0.61**(0.4)	**0.34**(0.2)	15/15
CMA-TPA	**1.2**(0.3)	**1.7**(1.0)	**0.71**(0.3)	**0.67**(0.3)	**0.92**(0.3)	15/15
GP1-CMAES	**0.83**(0.3)	**0.84**(0.3)	**0.20**(0.0)	**0.24**(0.1)	**0.38**(0.6)	5/15
GP5-CMAES	**0.51**(0.1)$^{\star}_{\downarrow 4}$	**0.86**(0.2)	**0.37**(0.2)	**0.32**(0.3)	**1.1**(0.9)	2/15
IPOPCMAv3p	**1.4**(0.4)	**1.7**(1)	5.5(3)	7.5(9)	∞ *5008*	0/15
LHD-10xDef	3.0(3)	**1.6**(0.8)	**0.74**(0.8)	∞	∞ *1000*	0/15
LHD-2xDefa	7.5(3)	7.8(10)	∞	∞	∞ *1000*	0/15
RAND-2xDef	9.3(12)	∞	∞	∞	∞ *1000*	0/15
RF1-CMAES	**1.4**(0.3)	**2.5**(2)	**1.0**(1)	**2.4**(4)	**2.3**(3)	1/15
RF5-CMAES	**2.2**(1)	4.1(4)	11(21)	∞	∞ *5034*	0/15
Sifeg	16(45)	19(27)	23(9)	258(219)	∞ *2e5*	0/15
Sif	19(16)	20(30)	28(44)	∞	∞ *2e5*	0/15
Srr	17(12)	12(11)	13(17)	∞	∞ *2e5*	0/15

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.