Comparison Tables: BBOB 2015 Testbed in 40-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: 40-D, running time excess $\mathrm{ERT}/\mathrm{ERT}_{\mathrm{best\ 2009}}$ 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 $\mathrm{ERT}_{\mathrm{best\ 2009}}$ (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
f1	*2.5e+2*:48	*1.6e+2*:82	*1.0e-8*:83	*1.0e-8*:83	*1.0e-8*:83	15/15
CMA-CSA	**2.0**(0.5)	2.3(0.9)	64(4)	64(2)	64(4)	15/15
CMA-MSR	2.6(0.9)	2.8(0.7)	66(3)	66(3)	66(2)	15/15
CMA-TPA	2.5(0.7)	**2.3**(0.4)	**43**(1)$^{\star 4}$	**43**(2)$^{\star 4}$	**43**(2)$^{\star 4}$	15/15

Table 3: 40-D, running time excess ERT/$\mathrm{ERT}_{\mathrm{best\ 2009}}$ 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 $\mathrm{ERT}_{\mathrm{best\ 2009}}$ (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
f2	*1.0e+7*:39	*6.3e+6*:71	*4.0e+5*:121	*2.5e+4*:499	*1.0e-8*:1188	15/15
CMA-CSA	**1.3**(0.9)	1.9(0.9)	17(8)	12(1)	**40**(1)$^{\star 4}$	15/15
CMA-MSR	2.0(1)	**1.8**(1)	**8.0**(2)	**8.2**(2)	47(0.9)	15/15
CMA-TPA	2.1(2)	2.2(0.8)	8.5(2)	8.9(3)	46(0.9)	15/15

Table 4: 40-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	*1.6e+3*:68	*1.0e+3*:222	*6.3e+2*:471	*4.0e+2*:662	*6.3e+1*:6332	15/15
CMA-CSA	**1.8**(1)	1.4(0.4)	1.1(0.2)	1.8(0.3)	3.6(1)	15/15
CMA-MSR	2.3(1)	1.3(0.5)	1.1(0.2)	**1.3**(0.2)	4.0(1)	15/15
CMA-TPA	2.3(1)	**1.1**(0.3)	**0.83**(0.1)*	1.4(0.5)	**2.7**(2)	15/15

Table 5: 40-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f4	*1.0e+3*:439	*6.3e+2*:670	*4.0e+2*:707	*2.5e+2*:735	*1.0e+2*:5369	15/15
CMA-CSA	1.2(0.4)	1.5(0.2)	3.3(0.8)	4.5(0.8)	**3.6**(3)	15/15
CMA-MSR	1.1(0.2)	1.2(0.3)	**1.8**(0.4)	12(8)	8.1(3)	15/15
CMA-TPA	**0.94**(0.2)	**1.1**(0.1)	2.0(0.4)	**2.5**(0.4)	4.5(2)	15/15

Table 6: 40-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f5	*4.0e+2*:51	*2.5e+2*:81	*1.0e-1*:120	*1.0e-8*:121	*1.0e-8*:121	15/15
CMA-CSA	1.7(0.5)	1.8(0.4)	4.4(0.5)	4.4(0.3)	4.4(0.4)	15/15
CMA-MSR	1.7(0.9)	1.8(0.3)	3.6(0.3)	**3.6**(0.7)	**3.6**(0.5)	15/15
CMA-TPA	**1.2**(0.5)*	**1.2**(0.4)$^{\star 2}$	**3.5**(0.5)	3.6(0.4)	3.6(0.3)	15/15

Table 7: 40-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_6 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f6	*6.3e+5*:50	*4.0e+5*:82	*4.0e+4*:127	*4.0e+2*:734	*6.3e+1*:2121	15/15
CMA-CSA	**1.6**(0.5)	**1.4**(0.7)	2.1(0.5)	2.9(0.2)	1.8(0.2)	15/15
CMA-MSR	1.8(0.8)	1.6(0.7)	**2.0**(0.5)	**1.7**(0.2)	1.7(0.9)	15/15
CMA-TPA	1.6(0.4)	1.9(0.5)	2.3(0.7)	1.9(0.3)	**1.5**(0.5)	15/15

Table 8: 40-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_7 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f7	*1.6e+3*:35	*1.0e+3*:106	*6.3e+2*:165	*2.5e+2*:489	*2.5e+1*:2987	15/15
CMA-CSA	**2.9**(1)	2.4(0.8)	2.7(1)	1.7(0.3)	**1.1**(0.8)	15/15
CMA-MSR	3.5(2)	2.2(0.2)	2.1(0.5)	1.2(0.3)	6.7(4)	15/15
CMA-TPA	3.4(1)	**2.0**(0.2)	**1.8**(0.4)	**1.1**(0.2)	10(30)	15/15

Table 9: 40-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_8 for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f8	*1.0e+5*:85	*6.3e+4*:111	*4.0e+4*:125	*2.5e+3*:430	*6.3e+1*:2106	15/15
CMA-CSA	3.4(1)	3.2(0.3)	3.4(0.7)	1.9(0.2)	3.3(8)	15/15
CMA-MSR	3.4(0.9)	3.1(0.8)	3.1(0.5)	1.7(0.2)	3.3(4)	15/15
CMA-TPA	**2.6**(0.6)	**2.2**(0.5)$^{\star 2}$	**2.3**(0.3)$^{\star 2}$	**1.3**(0.1)$^{\star 3}$	**1.4**(3)	15/15

Table 10: 40-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f9	*2.5e+2*:676	*1.6e+2*:865	*1.0e+2*:1397	*6.3e+1*:1896	*4.0e+1*:2180	15/15
CMA-CSA	2.1(0.3)	1.9(0.3)	1.6(0.4)	2.2(8)	2.2(4)	15/15
CMA-MSR	2.1(0.5)	1.9(0.5)	1.7(1)	2.5(5)	2.4(8)	15/15
CMA-TPA	**1.5**(0.6)*	**1.5**(1)	**1.1**(0.6)	**0.91**(0.5)*	**0.98**(0.3)*	15/15

Table 11: 40-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	*1.0e+7*:44	*6.3e+6*:80	*2.5e+6*:126	*2.5e+5*:408	*6.3e+3*:2376	15/15
CMA-CSA	1.5(0.3)	1.7(0.9)	5.3(2)	6.1(1)	4.2(0.7)	15/15
CMA-MSR	1.8(2)	**1.6**(0.7)	**2.3**(0.9)	**3.2**(1)	**3.2**(0.6)	15/15
CMA-TPA	**1.4**(1)	2.0(1)	2.7(0.4)	3.3(0.9)	3.3(0.7)	15/15

Table 12: 40-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{11} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f11	*1.0e+4*:22	*2.5e+3*:52	*2.5e+2*:432	*1.6e+2*:887	*1.6e+1*:2204	15/15
CMA-CSA	1.8(2)	1.8(3)	23(1)	12(0.8)	5.3(0.2)	15/15
CMA-MSR	**1.0**(1)	1.5(1.0)	**15**(2)$^{\star 2}$	**8.5**(0.4)$^{\star 2}$	5.0(0.3)	15/15
CMA-TPA	1.6(1)	**1.3**(0.9)	18(2)	9.5(0.7)	**4.8**(0.2)*	15/15

Table 13: 40-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f12	*2.5e+8*:54	*1.6e+8*:218	*1.0e+8*:284	*1.0e+7*:424	*4.0e+1*:2479	15/15
CMA-CSA	**4.7**(2)	1.7(0.4)	1.7(0.2)	2.2(0.1)	**1.5**(0.1)	15/15
CMA-MSR	5.6(2)	1.8(0.5)	1.7(0.5)	2.2(0.3)	2.2(2)	15/15
CMA-TPA	4.8(2)	**1.5**(0.5)	**1.4**(0.3)	**1.7**(0.4)$^{\star 2}$	1.9(2)	15/15

Table 14: 40-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f13	*2.5e+3*:85	*1.6e+3*:121	*1.6e+3*:121	*6.3e+1*:429	*1.0e+1*:2029	15/15
CMA-CSA	2.3(0.6)	3.7(0.6)	3.7(0.3)	5.0(0.6)	2.5(2)	15/15
CMA-MSR	2.4(0.8)	3.4(0.5)	3.4(0.5)	4.5(0.3)	2.8(3)	15/15
CMA-TPA	**2.2**(0.5)	**2.9**(0.5)	**2.9**(0.6)	**3.8**(0.4)$^{\star 3}$	**2.3**(3)	15/15

Table 15: 40-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	*6.3e+1*:34	*4.0e+1*:137	*2.5e+1*:176	*4.0e+0*:438	*1.0e-3*:2207	15/15
CMA-CSA	**4.1**(1)	2.2(0.6)	3.0(0.4)	2.7(0.4)	3.6(0.2)	15/15
CMA-MSR	5.7(2)	2.3(0.5)	2.5(0.7)	2.2(0.4)	**2.5**(0.3)	15/15
CMA-TPA	6.0(1)	**2.1**(0.6)	**2.3**(0.6)	**2.1**(0.2)	2.6(0.1)	15/15

Table 16: 40-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{15} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f15	*1.0e+3*:192	*6.3e+2*:458	*4.0e+2*:1170	*2.5e+2*:3875	*2.5e+2*:3875	15/15
CMA-CSA	1.5(0.4)	1.1(0.3)	1.1(0.2)	1.0(0.2)	1.0(0.2)	15/15
CMA-MSR	1.4(0.6)	0.98(0.3)	**0.73**(0.1)	**0.30**(0.0)$^{\star 4}$	**0.30**(0.0)$^{\star 4}$	15/15
CMA-TPA	**1.2**(0.3)	**0.88**(0.2)	0.92(0.7)	0.57(0.1)	0.57(0.2)	15/15

Table 17: 40-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*:117	*2.5e+1*:297	*1.6e+1*:4010	*1.6e+1*:4010	*1.0e+1*:5244	15/15
CMA-CSA	25(17)	14(3)	1.2(0.3)	1.2(0.1)	1.0(0.5)	15/15
CMA-MSR	**2.4**(0.4)$^{\star 4}$	**1.6**(0.4)$^{\star 4}$	**0.17**(0.0)$^{\star 4}$	**0.17**(0.0)$^{\star 4}$	**0.43**(0.1)$^{\star 2}$	15/15
CMA-TPA	11(4)	5.3(1)	0.48(0.1)	0.48(0.1)	1.3(1)	15/15

Table 18: 40-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*:54	*1.0e+1*:399	*6.3e+0*:688	*4.0e+0*:1115	*1.0e+0*:4220	15/15
CMA-CSA	5.4(3)	1.3(0.6)	1.1(0.3)	1.0(0.3)	**0.56**(0.1)	15/15
CMA-MSR	3.9(2)	0.96(0.3)	1.1(0.5)	9.1(4)	7.3(7)	15/15
CMA-TPA	**3.8**(1)	**0.93**(0.3)	**0.99**(0.3)	**1.0**(0.4)	5.0(9)	15/15

Table 19: 40-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{18} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f18	*6.3e+1*:55	*4.0e+1*:329	*4.0e+1*:329	*2.5e+1*:579	*6.3e+0*:2006	15/15
CMA-CSA	3.4(2)	1.2(0.4)	1.2(0.2)	1.2(0.3)	**0.97**(0.4)	15/15
CMA-MSR	3.2(0.9)	0.99(0.4)	0.99(0.2)	1.00(0.6)	10(15)	15/15
CMA-TPA	**2.7**(1)	**0.89**(0.2)	**0.89**(0.2)	**0.86**(0.4)	1.6(3)	15/15

Table 20: 40-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*:8.6e5	*1.0e-1*:1.4e6	*6.3e-2*:3.1e6	*4.0e-2*:5.2e6	*2.5e-2*:8.7e6	15/15
CMA-CSA	**1.2**(0.7)	**1.0**(0.6)	**0.56**(0.4)	**0.54**(0.2)	**0.66**(0.7)	9/15
CMA-MSR	1.8(0.9)	1.4(0.7)	0.76(0.3)	0.75(0.9)	1.3(2)	5/15
CMA-TPA	1.2(0.6)	1.0(0.2)	0.60(0.2)	0.62(0.4)	0.74(0.4)	8/15

Table 21: 40-D, running time excess ERT/$\text{ERT}_{\text{best 2009}}$ 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 $\text{ERT}_{\text{best 2009}}$ (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
f20	*2.5e+4*:83	*1.6e+4*:86	*1.0e+3*:125	*2.5e+0*:515	*1.6e+0*:5582	15/15
CMA-CSA	4.4(0.8)	4.9(0.7)	5.6(0.9)	5.2(0.7)	**57**(24)$^{\star 2}$	15/15
CMA-MSR	3.8(0.7)	4.2(0.7)	4.6(0.8)	2.6(0.2)	∞ *3e6*	0/15
CMA-TPA	**3.1**(0.4)*	**3.4**(0.5)$^{\star 2}$	**3.7**(0.5)$^{\star 2}$	**2.3**(0.2)	9480(1e4)	1/15

Table 22: 40-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*:160	*4.0e+1*:305	*2.5e+1*:380	*1.6e+1*:784	*6.3e+0*:2510	30/30
CMA-CSA	2.8(0.5)	1.9(0.4)	1.8(0.5)	**1.1**(0.6)	2.4(3)	15/15
CMA-MSR	2.2(0.7)	1.5(0.4)	**1.6**(0.5)	2.9(0.1)	96(4)	14/15
CMA-TPA	**2.0**(0.2)	**1.4**(0.2)	2.9(13)	2.2(3)	**2.0**(4)	15/15

Table 23: 40-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*:160	*4.0e+1*:231	*2.5e+1*:687	*1.6e+1*:1392	*1.0e+1*:3090	15/15
CMA-CSA	2.9(1)	9.1(25)	4.2(9)	57(205)	87(211)	14/15
CMA-MSR	5.8(12)	**4.5**(0.9)	347(2415)	172(2)	77(0.8)	14/15
CMA-TPA	**2.4**(0.3)	7.2(0.3)	**3.2**(3)	**3.0**(7)	**1.6**(1)	15/15

Table 24: 40-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*:68	*4.0e+0*:292	*2.5e+0*:603	*2.5e+0*:603	*1.6e+0*:2487	15/15
CMA-CSA	12(6)	72(51)	108(79)	108(121)	27(11)	15/15
CMA-MSR	**7.0**(3)	**2.5**(0.2)$^{\star 4}$	**1.4**(0.2)$^{\star 4}$	**1.4**(0.2)$^{\star 4}$	**0.41**(0.1)$^{\star 4}$	15/15
CMA-TPA	12(6)	85(213)	86(205)	86(114)	27(51)	15/15

Table 25: 40-D, running time excess ERT/ERT$_{\text{best 2009}}$ on f_{24} for given run-length based budgets (0.5D, 1.2D, 3D, 10D, and 50D function evaluations). The ERT and in braces, as dispersion measure, the half difference between 90 and 10%-tile of bootstrapped run lengths appear for each algorithm and run-length based target, the corresponding ERT$_{\text{best 2009}}$ (preceded by the target Δf-value in *italics*) in the first row. #succ is the number of trials that reached the target value of the last column. The median number of conducted function evaluations is additionally given in *italics*, if the target in the last column was never reached. Entries with succeeding star are statistically significantly better (according to the rank-sum test) compared to all other algorithms in the table, with $p = 0.05$ or $p = 10^{-k}$ when the number k following the star is larger than 1, with Bonferroni correction by the number of instances.

#FEs/D	0.5	1.2	3	10	50	#succ
f24	*4.0e+2*:1404	*2.5e+2*:17825	*1.6e+2*:18980	*1.0e+2*:38677	*6.3e+1*:1.6e5	15/15
CMA-CSA	0.98(0.4)	1.4(0.9)	1.4(2)	1.2(0.5)	0.80(0.6)	15/15
CMA-MSR	**0.63**(0.1)	**0.07**(4e-3)$^{\star 4}$	**0.43**(0.4)	**0.47**(0.2)	**0.31**(0.1)	15/15
CMA-TPA	0.76(0.4)	0.48(0.3)	0.48(0.3)	0.68(0.4)	0.60(0.5)	15/15

References

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

[2] Anne Auger, Steffen Finck, Nikolaus Hansen, and Raymond Ros. BBOB 2009: Comparison tables of all algorithms on all noiseless functions. Technical Report RT-0383, INRIA, April 2010.

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

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

[5] S. Finck, N. Hansen, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009. Updated February 2010.

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

[7] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009. Updated February 2010.

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

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