# Compilation of Results on the 2005 CEC Benchmark Function Set

Nikolaus Hansen

Computational Laboratory (CoLab)
Institute of Computational Science
ETH Zurich

May 4, 2006

# A Note on Evaluation Criteria

- *Quantitative* performance measurements, the Success Performance, abbreviated `SP` or `FEs`, is used as a measure for the expected number of function evaluations to reach a target function value.
- Invariance is a non-empirical statement on the ability to generalize performance results. Invariance guarantees identical performance on a class of functions. Possible invariances are
  - invariance against translation, scaling, or even order preserving transformations of the objective function value
  - invariance against angle preserving (rigid) transformations of the search space (translation, rotation)
- Meta-Parameters
  - how many parameters of the algorithm need to be adjusted to the object function?
  - how many different setups were tested?
  - how many different setups were finally used?

# Methods

- task: black-box optimization of 25 benchmark functions
- 25 runs on each benchmark function for each dimension $n = 10, 30$
- a run is successful if the global optimum is reached with the given precision, before the
- maximum number of function evaluations

$$\mathrm{FE}_{\max} = \begin{cases} 10^5 & \text{for } n = 10 \\ 3 \times 10^5 & \text{for } n = 30 \end{cases}$$ is reached

Remark

the setting of $\mathrm{FE}_{\max}$ has a remarkable influence on the results, if the target function value can be reached only for a (slightly) larger number of function evaluations with a high probability. Where $\mathrm{FEs} \geq \mathrm{FE}_{\max}$ the result must be taken with great care.

Reference

Suganthan, Hansen, Liang, Deb, Chen, Auger, and Tiwari (2005). *Problem Definitions and Evaluation Criteria for the CEC 2005 Special Session on Real-Parameter Optimization*, Technical report, Nanyang Technological University, Singapore, May 2005, http://www.ntu.edu.sg/home/EPNSugan

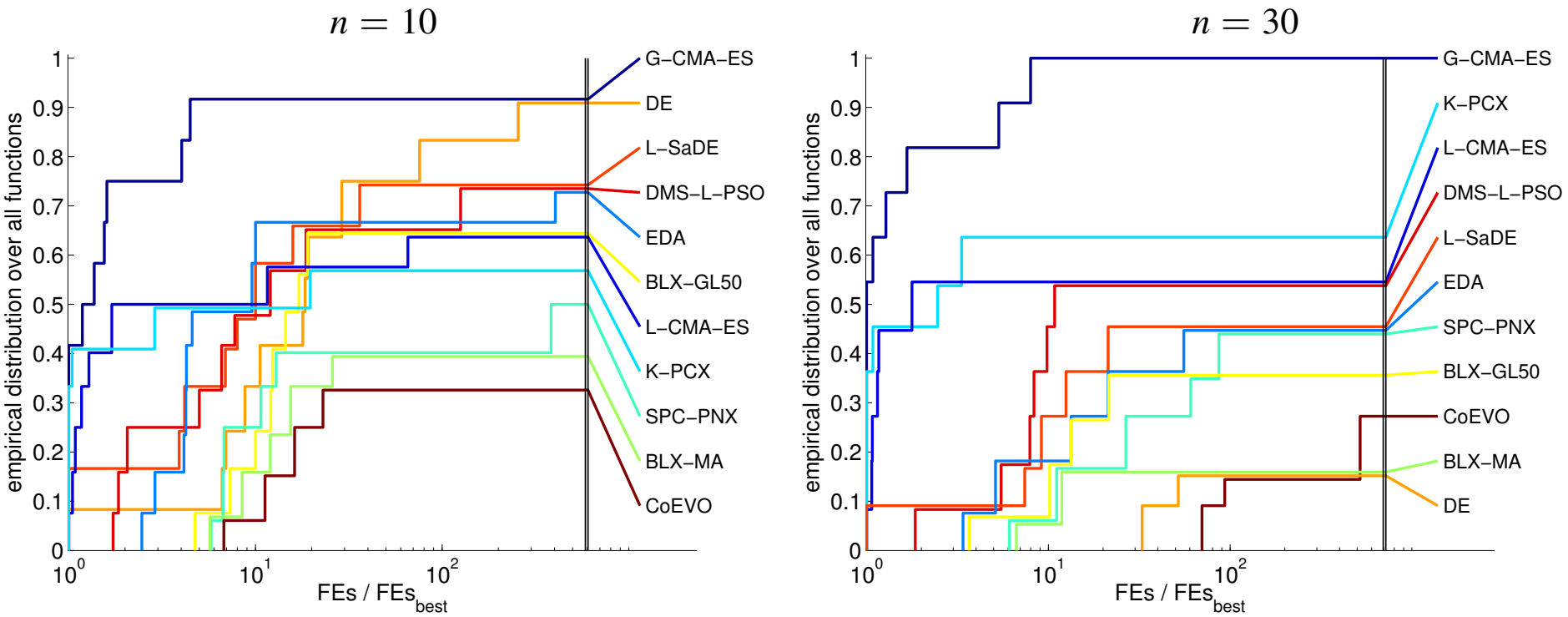
# References to Algorithms

| | |
|---|---|
| BLX-GL50 | García-Martínez and Lozano (Hybrid Real-Coded...) |
| BLX-MA | Molina et al. (Adaptive Local Search...) |
| CoEVO | Pošík (Real-Parameter Optimization...) |
| DE | Rönkkönen et al. (Real-Parameter Optimization...) |
| DMS-L-PSO | Liang and Suganthan (Dynamic Multi-Swarm...) |
| EDA | Yuan and Gallagher (Experimental Results...) |
| G-CMA-ES | Auger and Hansen (A Restart CMA...) |
| K-PCX | Sinha et al. (A Population-Based,...) |
| L-CMA-ES | Auger and Hansen (Performance Evaluation...) |
| L-SaDE | Qin and Suganthan (Self-Adaptive Differential...) |
| SPC-PNX | Ballester et al. (Real-Parameter Optimization...) |

In: CEC 2005 IEEE Congress on Evolutionary Computation, Proceedings

# Summarized Results

## Empirical Distribution of Normalized Success Performance




$\texttt{FEs} = \texttt{mean}(\#\mathit{fevals}) \times \frac{\#\text{all runs (25)}}{\#\text{successful runs}}$, where $\#\mathit{fevals}$ includes only successful runs.

Shown: **empirical distribution function** of the Success Performance $\texttt{FEs}$ divided by $\texttt{FEs}$ of the best algorithm on the respective function.

Results of all functions are used where at least one algorithm was successful at least once, i.e. where the target function value was reached in at least one experiment (out of $11 \times 25$ experiments).

Small values for $\texttt{FEs}$ and therefore large (cumulative frequency) values in the graphs are preferable.

# Function Sets

We split the function set into three subsets

- unimodal functions
- solved multimodal functions

  at least one algorithm conducted at least one successful run
- unsolved multimodal functions

  no single run was successful for any algorithm

# Unimodal Functions

## Normalized Success Performance, Tabulated

$n = 10$, $\mathrm{FE}_{\max} = 100000$

| | solved functions | success rate | 1 Sphere 1000 | 2 Schwefel 1.2 2400 | 3 Ellipsoid Condition $10^6$ 6500 | 4 Schwefel 1.2 with Noise 2900 | 5 Schwefel 2.6 on Bounds 5900 | 6 Rosenbrock 7100 |
|---|---|---|---|---|---|---|---|---|
| G-CMA-ES | 6 | 100% | 1.6(25) | 1.0(25) | 1.0(25) | 1.0(25) | 1.0(25) | 1.5(25) |
| EDA | 6 | 97% | 10.0(25) | 4.6(25) | 2.5(23) | 4.1(25) | 4.2(25) | 9.6(22) |
| DE | 6 | 96% | 29.0(25) | 19.2(25) | 18.5(20) | 17.9(25) | 6.9(25) | 6.6(24) |
| L-CMA-ES | 6 | 88% | 1.7(25) | 1.1(25) | 1.0(25) | 65.5( 7) | 1.0(25) | 1.3(25) |
| BLX-GL50 | 5 | 83% | 19.0(25) | 17.1(25) | [9] | 14.5(25) | 4.7(25) | 7.3(25) |
| DMS-L-PSO | 5 | 80% | 12.0(25) | 5.0(25) | 1.8(25) | [11] | 18.6(20) | 7.7(25) |
| L-SaDE | 5 | 77% | 10.0(25) | 4.2(25) | 8.0(16) | 15.9(24) | [9] | 6.9(25) |
| SPC-PNX | 4 | 67% | 6.7(25) | 12.9(25) | [11] | 10.7(25) | 6.8(25) | [10] |
| CoEVO | 4 | 67% | 23.0(25) | 11.3(25) | 6.8(25) | 16.2(25) | [10] | [11] |
| K-PCX | 4 | 62% | 1.0(25) | 1.0(25) | [8] | 19.7(21) | [11] | 1.0(22) |
| BLX-MA | 3 | 49% | 12.0(25) | 15.4(25) | [10] | 25.9(24) | [8] | [9] |

- **First row**: Success Performance $\mathtt{FEs} = \mathtt{mean}(\#\mathit{fevals}) \times \frac{\#\text{all runs (25)}}{\#\text{successful runs}}$ of the best algorithm, where $\#\mathit{fevals}$ includes only successful runs
- **Table entries**: Success Performance $\mathtt{FEs}$ divided by $\mathtt{FEs}$ of the best algorithm (first row), (number of successful runs in round brackets), [rank of median final function value, where $\mathtt{FEs}$ is not available, in square brackets]

# Unimodal Functions

## Normalized Success Performance, Tabulated

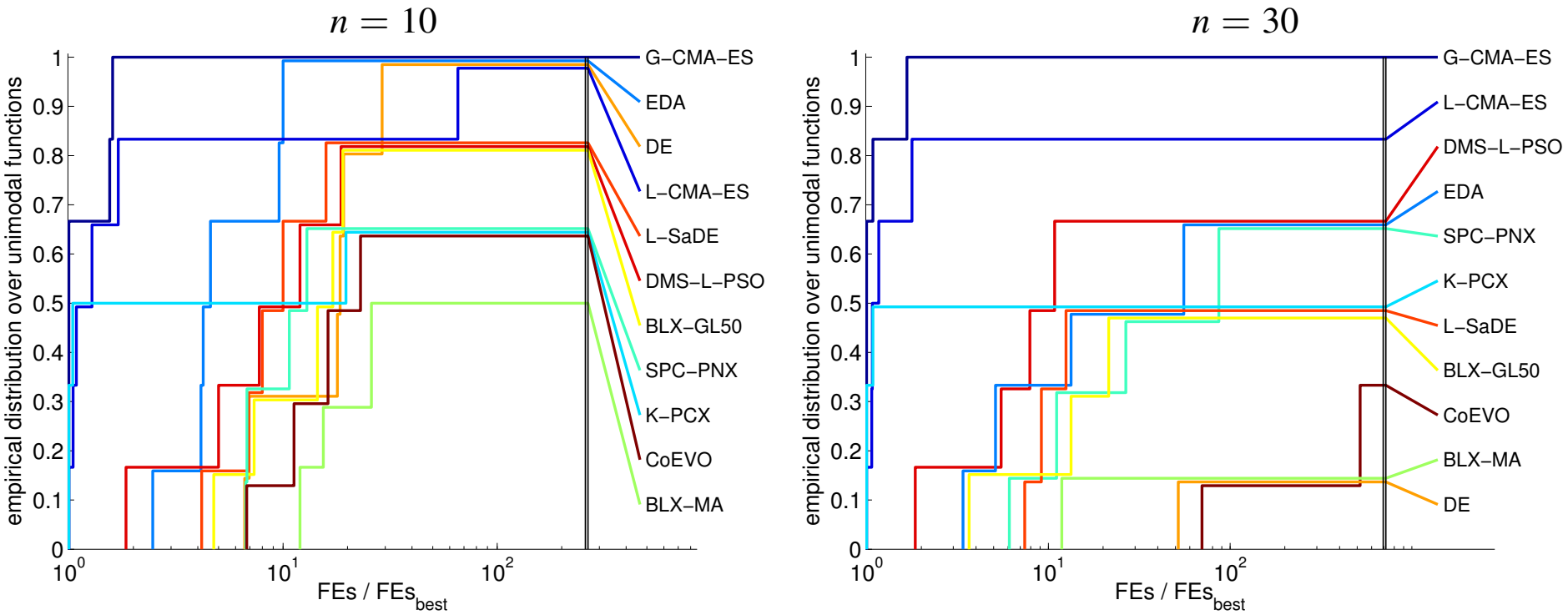
$n = 30$, $\mathrm{FE}_{\max} = 300000$

| | solved functions | success rate | 1 Sphere 2700 | 2 Schwefel 1.2 12000 | 3 Ellipsoid Condition $10^6$ 43000 | 4 Schwefel 1.2 with Noise 59000 | 5 Schwefel 2.6 on Bounds 66000 | 6 Rosenbrock 60000 |
|---|---|---|---|---|---|---|---|---|
| G-CMA-ES | 6 | 90% | 1.7(25) | 1.1(25) | 1.0(25) | 1.0(10) | 1.0(25) | 1.0(25) |
| L-CMA-ES | 5 | 83% | 1.8(25) | 1.2(25) | 1.0(25) | [11] | 1.1(25) | 1.1(25) |
| EDA | 4 | 67% | 55.6(25) | 13.3(25) | 5.1(25) | 3.4(25) | [3] | [10] |
| DMS-L-PSO | 4 | 63% | 1.9(25) | 10.8(25) | 7.9(21) | [9] | [9] | 5.5(24) |
| BLX-GL50 | 3 | 50% | 21.5(25) | 13.3(25) | [7] | [6] | [6] | 3.7(25) |
| SPC-PNX | 4 | 45% | 11.1(25) | 26.7(22) | [11] | 6.1(19) | [10] | 86.7( 1) |
| K-PCX | 3 | 43% | 1.0(25) | 1.0(25) | [6] | [8] | [7] | 1.1(14) |
| L-SaDE | 3 | 41% | 7.4(25) | 12.5(24) | [8] | 9.2(13) | [4] | [8] |
| BLX-MA | 1 | 17% | 11.9(25) | [10] | [10] | [7] | [7] | [8] |
| DE | 1 | 17% | 51.9(25) | [11] | [9] | [5] | [5] | [7] |
| CoEVO | 2 | 7% | 519 ( 3) | 70.0( 8) | [5] | [10] | [11] | [11] |

- **First row**: Success Performance $\mathtt{FEs} = \mathtt{mean}(\#\mathit{fevals}) \times \frac{\#\text{all runs (25)}}{\#\text{successful runs}}$ of the best algorithm, where $\#\mathit{fevals}$ includes only successful runs
- **Table entries**: Success Performance $\mathtt{FEs}$ divided by $\mathtt{FEs}$ of the best algorithm (first row), (number of successful runs in round brackets), [rank of median final function value, where $\mathtt{FEs}$ is not available, in square brackets]

# Unimodal Functions

## Empirical Distribution of Normalized Success Performance




Empirical distribution function of the Success Performance $\texttt{FEs}$ divided by $\texttt{FEs}$ of the best algorithm (table entries of last slides).

$\texttt{FEs} = \texttt{mean}(\#\mathit{fevals}) \times \frac{\#\text{all runs (25)}}{\#\text{successful runs}}$, where $\#\mathit{fevals}$ includes only successful runs.

Small values of $\texttt{FEs}$ and therefore large values in the empirical distribution graphs are preferable.

# Solved Multimodal Functions

## Normalized Success Performance, Tabulated

$n = 10$, $\mathrm{FE}_{\max} = 100000$

| | solved functions | success rate | 7 Griewank out Bounds | 9 Rastrigin Separable | 10 Rastrigin Rotated | 11 Weierstrass | 12 Schwefel 2.13 | 15 Hybrid Separable |
|---|---|---|---|---|---|---|---|---|
| | | | 4700 | 17000 | 55000 | 190000 | 8200 | 33000 |
| G-CMA-ES | 5 | 63% | 1.0(25) | 4.5(19) | 1.2(23) | 1.4( 6) | 4.0(22) | [6] |
| L-SaDE | 4 | 53% | 36.2( 6) | 1.0(25) | [5] | [8] | 3.9(25) | 1.0(23) |
| DMS-L-PSO | 4 | 47% | 126 ( 4) | 2.1(25) | [3] | [7] | 6.6(19) | 1.7(22) |
| K-PCX | 3 | 40% | [10] | 2.9(24) | 1.0(22) | [10] | 1.0(14) | [11] |
| DE | 5 | 30% | 255 ( 2) | 10.6(11) | [9] | 1.0(12) | 8.8(19) | 75.8( 1) |
| L-CMA-ES | 2 | 25% | 1.2(25) | [11] | [10] | [6] | 11.6(12) | [6] |
| BLX-GL50 | 3 | 17% | 12.3( 9) | 10.0( 3) | [5] | [5] | 12.1(13) | [9] |
| BLX-MA | 2 | 15% | [11] | 5.7(18) | [7] | [9] | [9] | 8.5( 5) |
| EDA | 3 | 9% | 404 ( 1) | [9] | [4] | 2.9( 3) | 4.3(10) | [9] |
| SPC-PNX | 2 | 1% | 383 ( 1) | [8] | [8] | 5.8( 1) | [10] | [6] |
| CoEVO | 0 | 0% | [9] | [10] | [11] | [11] | [11] | [8] |

- **First row**: Success Performance $\mathrm{FEs} = \frac{\#fevals}{p_{\mathrm{succ}}}$ of the best algorithm
- **Table entries**: Success Performance FEs normalized (divided) by the first row, (number of successful runs in round brackets), [rank of median final function value, where FEs is not available, in square brackets]

# Solved Multimodal Functions

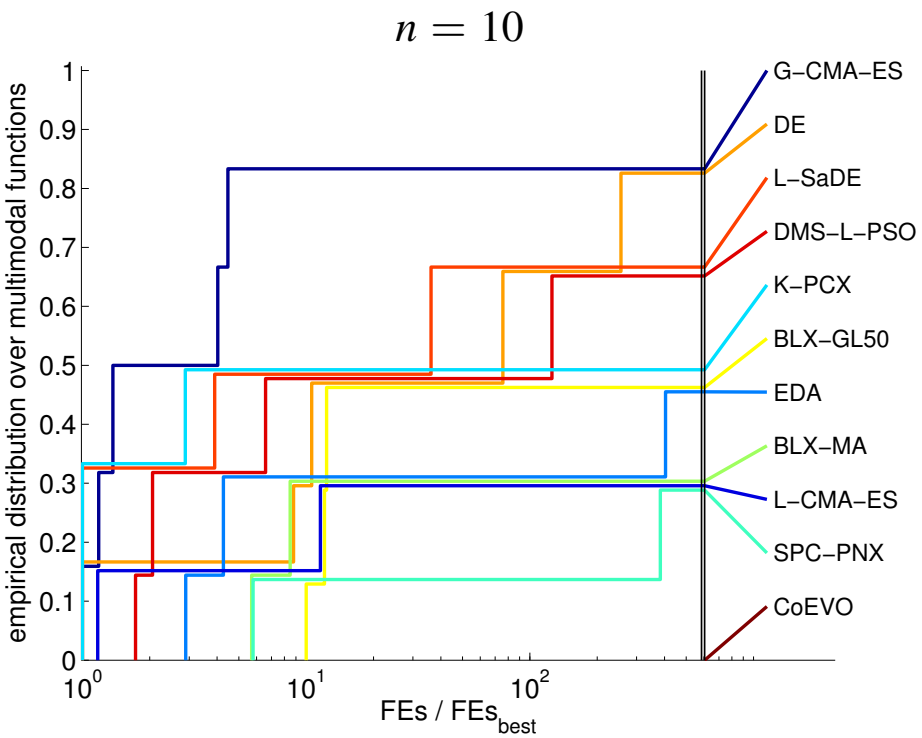
## Normalized Success Performance, Tabulated

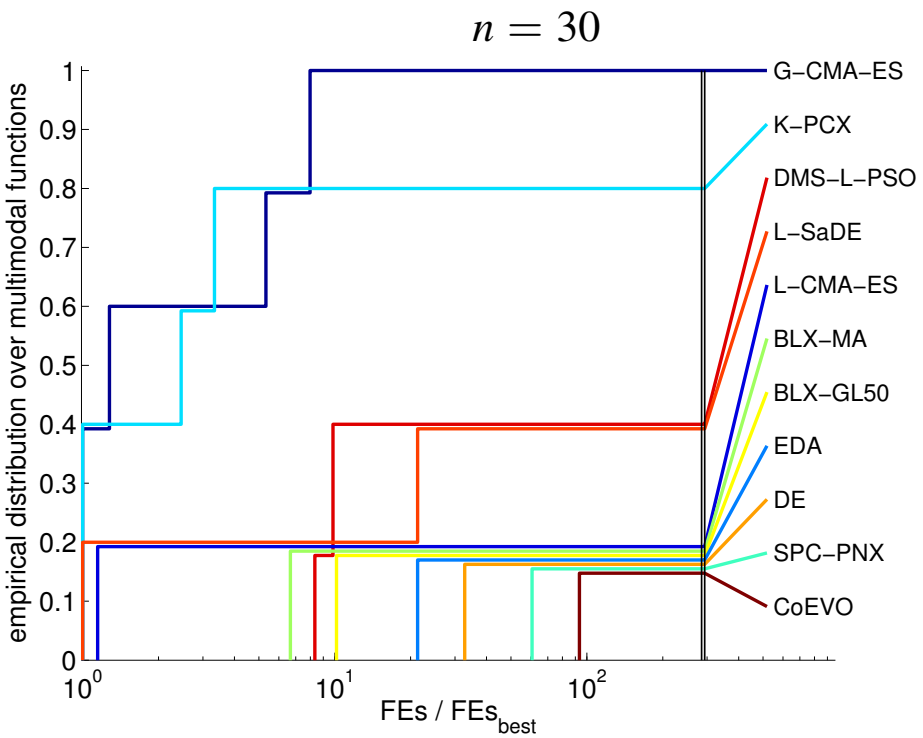
$n = 30$, $\mathrm{FE}_{\max} = 300000$

| | solved functions | success rate | 7 Griewank out Bounds 6100 | 9 Rastrigin Separable 99000 | 10 Rastrigin Rotated 450000 | 11 Weierstrass 5000000 | 12 Schwefel 2.13 180000 | 15 Hybrid Separable |
|---|---|---|---|---|---|---|---|---|
| K-PCX | 4 | 38% | 2.5(10) | 3.3(18) | 1.0(14) | [7] | 1.0( 5) | [11] |
| G-CMA-ES | 5 | 37% | 1.0(25) | 8.0( 9) | 5.3( 3) | 1.0( 1) | 1.3( 8) | [1] |
| L-SaDE | 2 | 36% | 21.3(20) | 1.0(25) | [4] | [5] | [4] | [2] |
| DMS-L-PSO | 2 | 22% | 9.8(24) | [6] | [5] | [5] | 8.3( 4) | [4] |
| EDA | 1 | 20% | 21.3(25) | [10] | [9] | [10] | [7] | [4] |
| BLX-GL50 | 1 | 20% | 10.2(25) | [5] | [3] | [4] | [8] | [4] |
| DE | 1 | 20% | 32.8(22) | [6] | [6] | [10] | [5] | [8] |
| L-CMA-ES | 1 | 20% | 1.1(25) | [11] | [11] | [3] | [9] | [8] |
| SPC-PNX | 1 | 13% | 60.7(16) | [8] | [7] | [2] | [10] | [8] |
| CoEVO | 1 | 9% | 93.4(11) | [9] | [10] | [9] | [11] | [10] |
| BLX-MA | 1 | 7% | [11] | 6.7( 9) | [8] | [8] | [6] | [4] |

- **First row**: Success Performance $\mathtt{FEs} = \mathtt{mean}(\#\mathit{fevals}) \times \frac{\#\text{all runs (25)}}{\#\text{successful runs}}$ of the best algorithm, where $\#\mathit{fevals}$ includes only successful runs
- **Table entries**: Success Performance $\mathtt{FEs}$ divided by $\mathtt{FEs}$ of the best algorithm (first row), (number of successful runs in round brackets), [rank of median final function value, where $\mathtt{FEs}$ is not available, in square brackets]

# Multimodal Functions
## Empirical Distribution of Normalized Success Performance




Empirical distribution function of the Success Performance $\texttt{FEs}$ divided by $\texttt{FEs}$ of the best algorithm (table entries of last slides).

$\texttt{FEs} = \texttt{mean}(\#\mathit{fevals}) \times \frac{\#\text{all runs (25)}}{\#\text{successful runs}}$, where $\#\mathit{fevals}$ includes only successful runs.

Small values of $\texttt{FEs}$ and therefore large values in the empirical distribution graphs are preferable.

# Never Solved Multimodal Functions

## Rank of Final Function Value

$n = 10$, $\mathrm{FE}_{\max} = 10^5$

| | mean | 8 Ackley Condition $10^2$ | 13 Expanded 6&7 | 14 Expanded Schaffer F6 | 16 Hybrid Rotated | 17 Hybrid with Noise | 18 Hybrid F18 | 19 Hybrid Narrow | 20 Hybrid on Bounds | 21 Hybrid F21 | 22 Hybrid High Condition | 23 Hybrid Non-Continuous | 24 Hybrid F24 | 25 Hybrid out Bo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G-CMA-ES | 3.8 | 5.5 | 4 | 7 | 1 | 5.5 | 3 | 2.5 | 2.5 | 5.5 | 1.5 | 4 | 5 | 2.5 |
| BLX-GL50 | 4.2 | 5.5 | 6 | 1.5 | 2.5 | 3 | 3 | 2.5 | 2.5 | 5.5 | 5.5 | 4 | 5 | 8.5 |
| L-SaDE | 5.3 | 5.5 | 1 | 6 | 5.5 | 3 | 8 | 7.5 | 7.5 | 5.5 | 5.5 | 4 | 5 | 4.5 |
| DMS-L-PSO | 5.7 | 5.5 | 2 | 3.5 | 4 | 3 | 8 | 7.5 | 7.5 | 5.5 | 5.5 | 8 | 5 | 8.5 |
| L-CMA-ES | 6.0 | 5.5 | 3 | 11 | 7.5 | 11 | 6 | 5 | 5 | 1 | 3 | 4 | 11 | 4.5 |
| SPC-PNX | 6.2 | 11 | 8 | 8 | 7.5 | 5.5 | 3 | 2.5 | 2.5 | 5.5 | 9 | 4 | 5 | 8.5 |
| EDA | 6.2 | 5.5 | 10 | 5 | 9 | 8 | 3 | 7.5 | 7.5 | 5.5 | 8 | 4 | 5 | 2.5 |
| BLX-MA | 6.6 | 5.5 | 7 | 1.5 | 5.5 | 7 | 8 | 7.5 | 7.5 | 10 | 5.5 | 10 | 5 | 6 |
| K-PCX | 7.0 | 5.5 | 5 | 3.5 | 2.5 | 1 | 10 | 11 | 10 | 11 | 1.5 | 11 | 10 | 8.5 |
| DE | 7.0 | 5.5 | 11 | 10 | 11 | 10 | 3 | 2.5 | 2.5 | 5.5 | 10 | 4 | 5 | 11 |
| CoEVO | 8.2 | 5.5 | 9 | 9 | 10 | 9 | 11 | 10 | 11 | 5.5 | 11 | 9 | 5 | 1 |

- **Table entries**: Rank of the median of the final best function values from 25 runs measured with two digits of precision

# Never Solved Multimodal Functions

## Rank of Final Function Value

$n = 30$, $\mathrm{FE}_{\max} = 3 \times 10^5$

| | mean | 8 Ackley Condition $10^2$ | 13 Expanded 6&7 | 14 Expanded Schaffer F6 | 15 Hybrid Separable | 16 Hybrid Rotated | 17 Hybrid with Noise | 18 Hybrid F18 | 19 Hybrid Narrow | 20 Hybrid on Bounds | 21 Hybrid F21 | 22 Hybrid High Condition | 23 Hybrid Non-Continuo | 24 Hybrid F24 | 25 Hyb |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| G-CMA-ES | 4.1 | 2.5 | 5 | 6.5 | 1 | 1 | 6 | 5.5 | 5.5 | 5.5 | 4.5 | 1 | 3.5 | 6 | 4 |
| EDA | 4.8 | 7.5 | 11 | 6.5 | 8 | 7 | 7 | 1 | 1 | 1 | 4.5 | 3 | 3.5 | 2.5 | 4 |
| BLX-MA | 4.9 | 7.5 | 4 | 6.5 | 4.5 | 10 | 5 | 3 | 3 | 3 | 4.5 | 7 | 3.5 | 2.5 | 4 |
| SPC-PNX | 5.0 | 7.5 | 8 | 6.5 | 8 | 4.5 | 2 | 5.5 | 5.5 | 5.5 | 4.5 | 3 | 3.5 | 2.5 | 4 |
| BLX-GL50 | 5.1 | 7.5 | 6.5 | 2 | 4.5 | 4.5 | 3 | 5.5 | 5.5 | 5.5 | 4.5 | 5 | 7 | 7 | 4 |
| L-CMA-ES | 5.6 | 2.5 | 2 | 10 | 2 | 3 | 10 | 8.5 | 8.5 | 8.5 | 4.5 | 3 | 3.5 | 9 | 4 |
| DE | 6.1 | 7.5 | 6.5 | 6.5 | 8 | 8 | 8 | 5.5 | 5.5 | 5.5 | 4.5 | 6 | 3.5 | 2.5 | 8 |
| K-PCX | 6.2 | 11 | 10 | 10 | 11 | 2 | 1 | 2 | 2 | 2 | 9 | 9 | 9 | 5 | 4 |
| CoEVO | 8.5 | 7.5 | 9 | 6.5 | 10 | 9 | 9 | 10 | 10 | 10 | 4.5 | 8 | 8 | 8 | 9 |
| L-SaDE | – | 2.5 | 1 | 2 | 4.5 | – | – | – | – | – | – | – | – | – | – |
| DMS-L-PSO | – | 2.5 | 3 | 2 | 4.5 | 6 | 4 | 8.5 | 8.5 | 8.5 | – | – | – | – | – |

- **Table entries**: Rank of the median of the final best function values from 25 runs measured with two digits of precision