6th GECCO Workshop on Blackbox Optimization Benchmarking (BBOB): Wrap-Up

The BBOBies

https://github.com/numbbo/coco

Numerical Blackbox Optimization

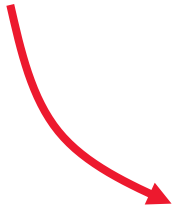
Optimize $f: \Omega \subset \mathbb{R}^n \mapsto \mathbb{R}^k$

derivatives not available or not useful

Need: Benchmarking

- understanding of algorithms
- algorithm selection
- putting algorithms to a standardized test
  - simplify judgement
  - simplify comparison
  - regression test under algorithm changes

that's where COCO and BBOB come into play



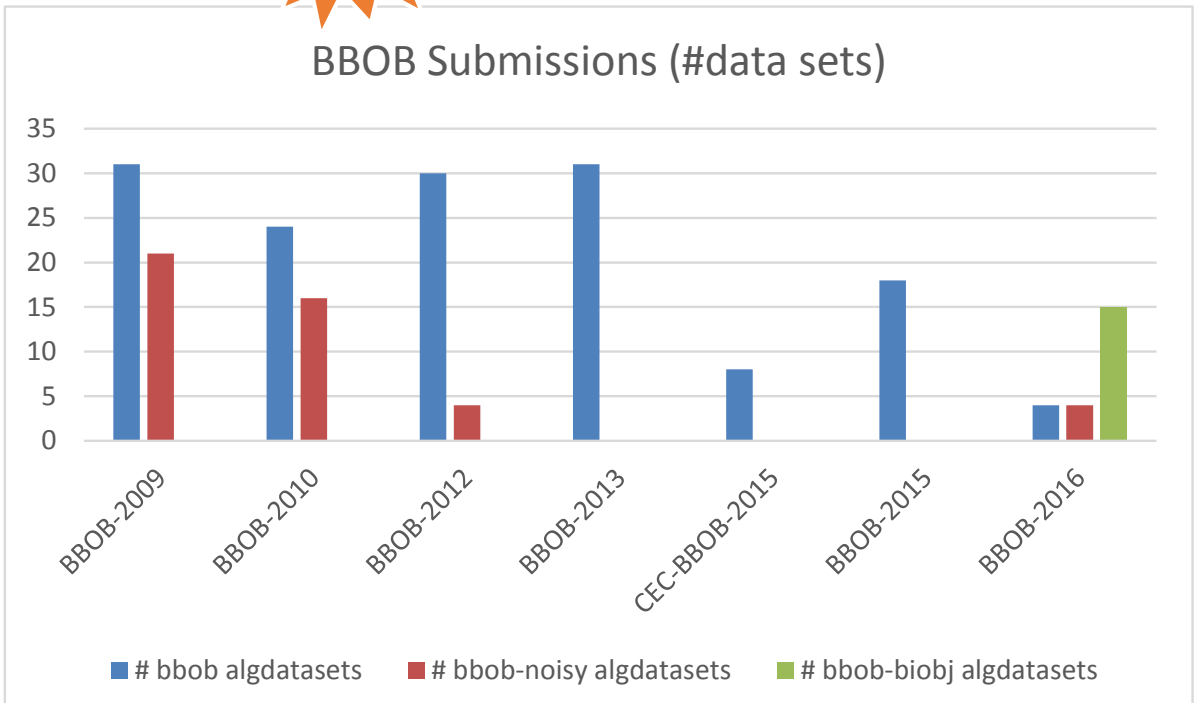
Comparing Continuous Optimizers Platform

`https://github.com/numbbo/coco`

Available Data Sets in COCO

before and after 2016

	Before 2016	In 2016
• bbob	140+ algo data sets	4 data sets
• bbob-noisy	40+ algo data sets	4 data sets
• bbob-biobj	new in 2016	15 data sets



extension of COCO to multi-objective optimization

bbob-biobj Testbed (new in 2016)

- 55 functions, combining **bbob** functions
- 6 dimensions (2..40D)
- no normalization
- ideal/nadir known
- but Pareto set/front not (only refsets)

1 Separable Functions	
f1	Sphere Function ✓
f2	Ellipsoidal Function ✓
f3	Rastrigin Function
f4	Büche-Rastrigin Function
f5	Linear Slope
2 Functions with low or moderate conditioning	
f6	Attractive Sector Function ✓
f7	Step Ellipsoidal Function
f8	Rosenbrock Function, original ✓
f9	Rosenbrock Function, rotated
3 Functions with high conditioning and unimodal	
f10	Ellipsoidal Function
f11	Discus Function
f12	Bent Cigar Function
f13	Sharp Ridge Function ✓
f14	Different Powers Function ✓

4 Multi-modal functions with adequate global structure	
f15	Rastrigin Function ✓
f16	Weierstrass Function
f17	Schaffers F7 Function ✓
f18	Schaffers F7 Functions, moderately ill-conditioned
f19	Composite Griewank-Rosenbrock Function F8F2
5 Multi-modal functions with weak global structure	
f20	Schwefel Function ✓
f21	Gallagher's Gaussian 101-me Peaks Function ✓
f22	Gallagher's Gaussian 21-hi Peaks Function
f23	Katsuura Function

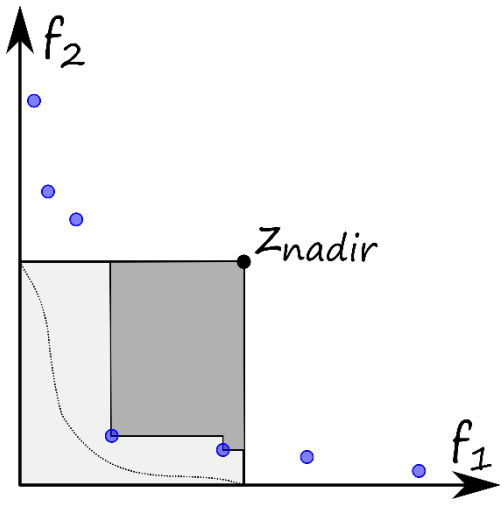
	f_1	f_2	f_6	f_8	f_{13}	f_{14}	f_{15}	f_{17}	f_{20}	f_{21}
f_1	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10
f_2		f11	f12	f13	f14	f15	f16	f17	f18	f19
f_6			f20	f21	f22	f23	f24	f25	f26	f27
f_8				f28	f29	f30	f31	f32	f33	f34
f_{13}					f35	f36	f37	f38	f39	f40
f_{14}						f41	f42	f43	f44	f45
f_{15}							f46	f47	f48	f49
f_{17}								f50	f51	f52
f_{20}									f53	f54
f_{21}										f55

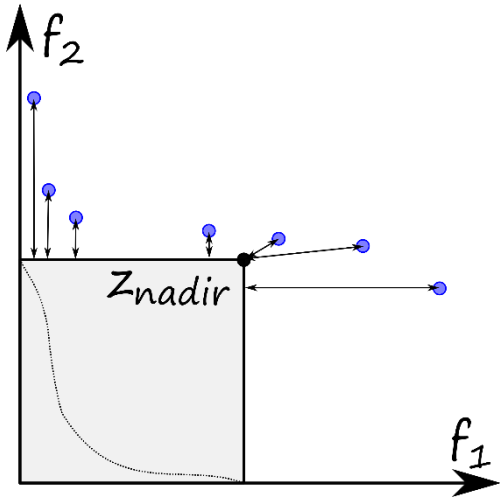
Bi-objective Performance Assessment

algorithm quality =

normalized* hypervolume (HV)
of all non-dominated solutions
if a point dominates nadir

closest normalized* negative distance
to region of interest $[0,1]^2$
if no point dominates nadir





* such that ideal=[0,0] and nadir=[1,1]

Bi-objective Performance Assessment

Again, as in last session's wrap-up:

- results are relative to a reference set, given as the best Pareto front approximation known (since exact Pareto set not known)
  - note: improved reference sets compared to workshop papers
- actual absolute hypervolume targets used are

  HV(refset) – targetprecision

  with 51 fixed targetprecisions between 1 and 10^{-5} (same for all functions, dimensions, and instances) in the displays
- all 10 instances are displayed

let's dig into the data...

all results for the bbob-biobj suites

Bi-objective Performance Assessment

The ECDFs are actually influenced by

❶ the number and set of instances and by

❷ the reference set and the reference hypervolume values

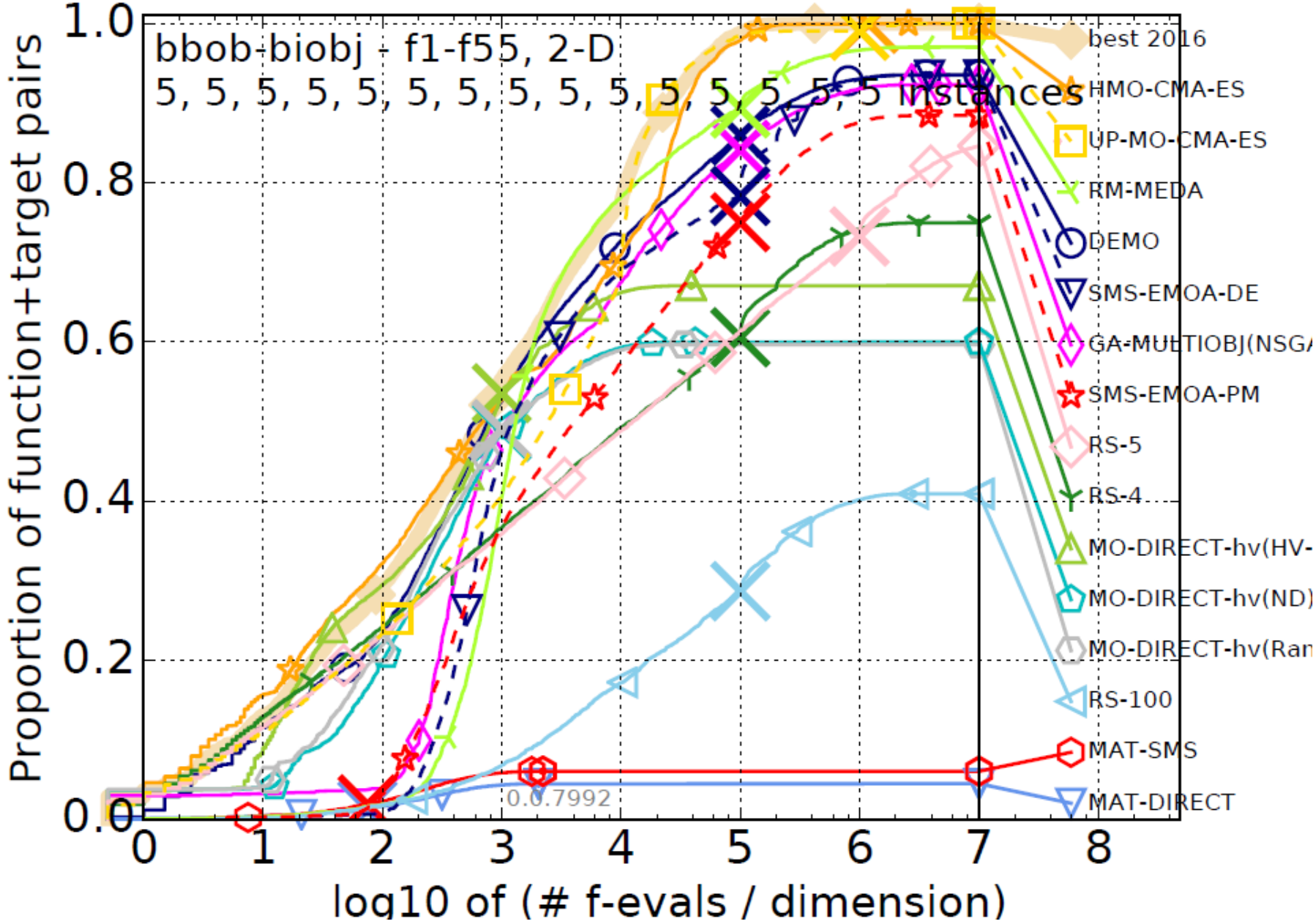
...so let's have a brief look behind the scenes of BBOB

❶ **the influence of the instances**

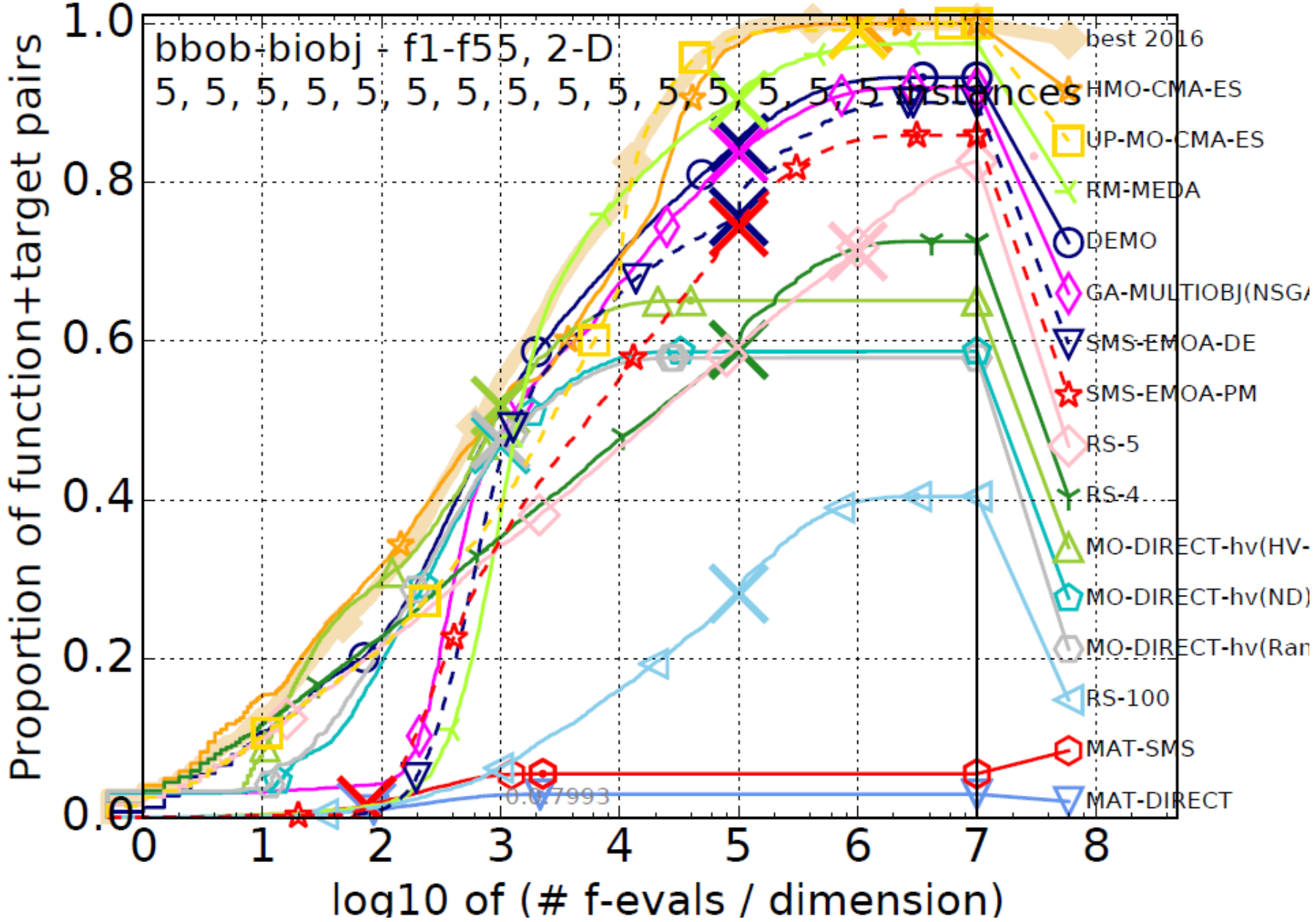
BBOB-2016: Instances

- on the **`bbob-biobj`** test suite, experiments were run on 10 instances
- but all plots were based on the first 5 instances only
  - practical reason 1: we did not have enough data to produce good hypervolume reference values for all instances
  - practical reason 2: setting allows to investigate potential overfitting ("split between training and test")

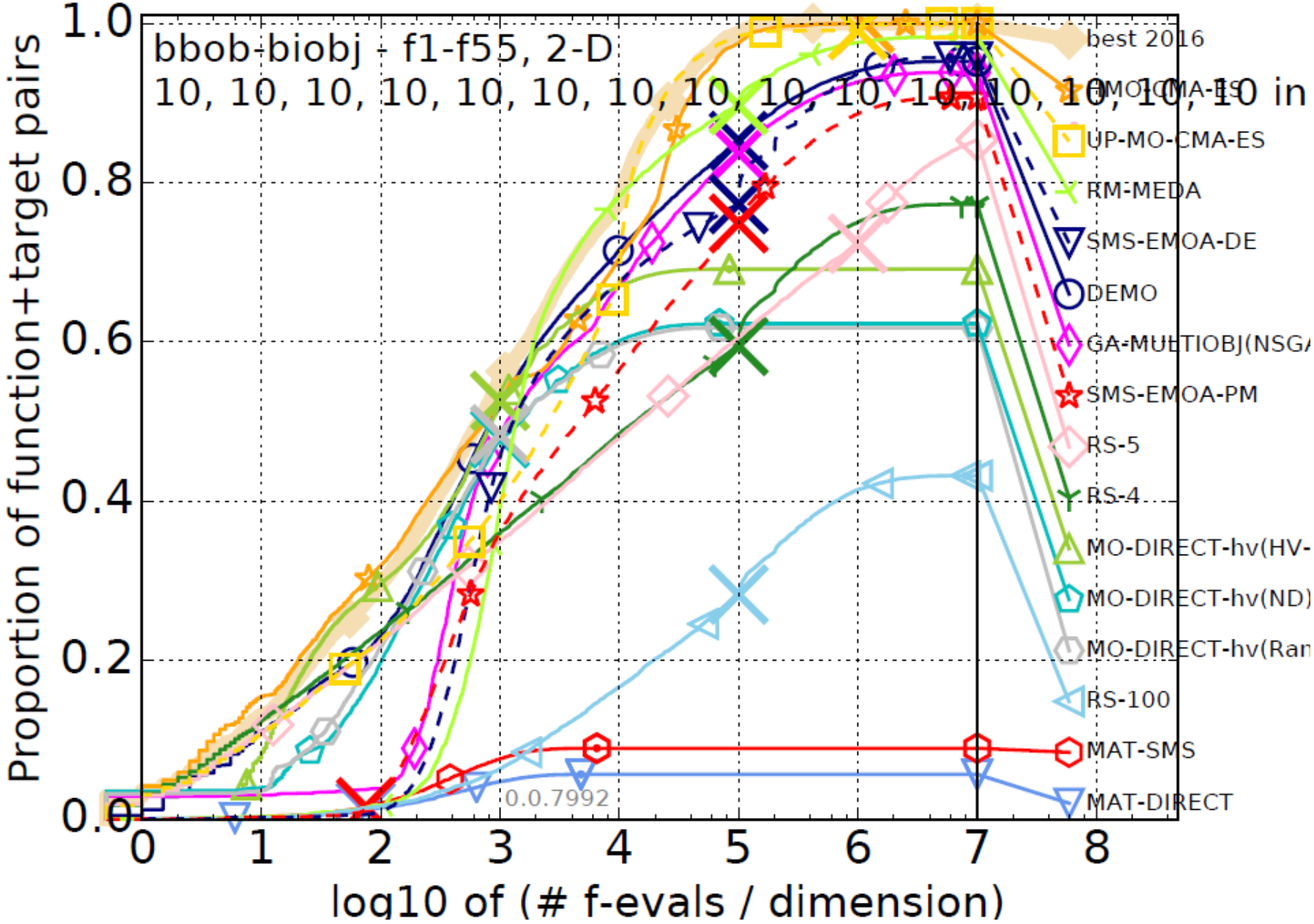
BBOB-2016: Instances 1-5 (2-D)



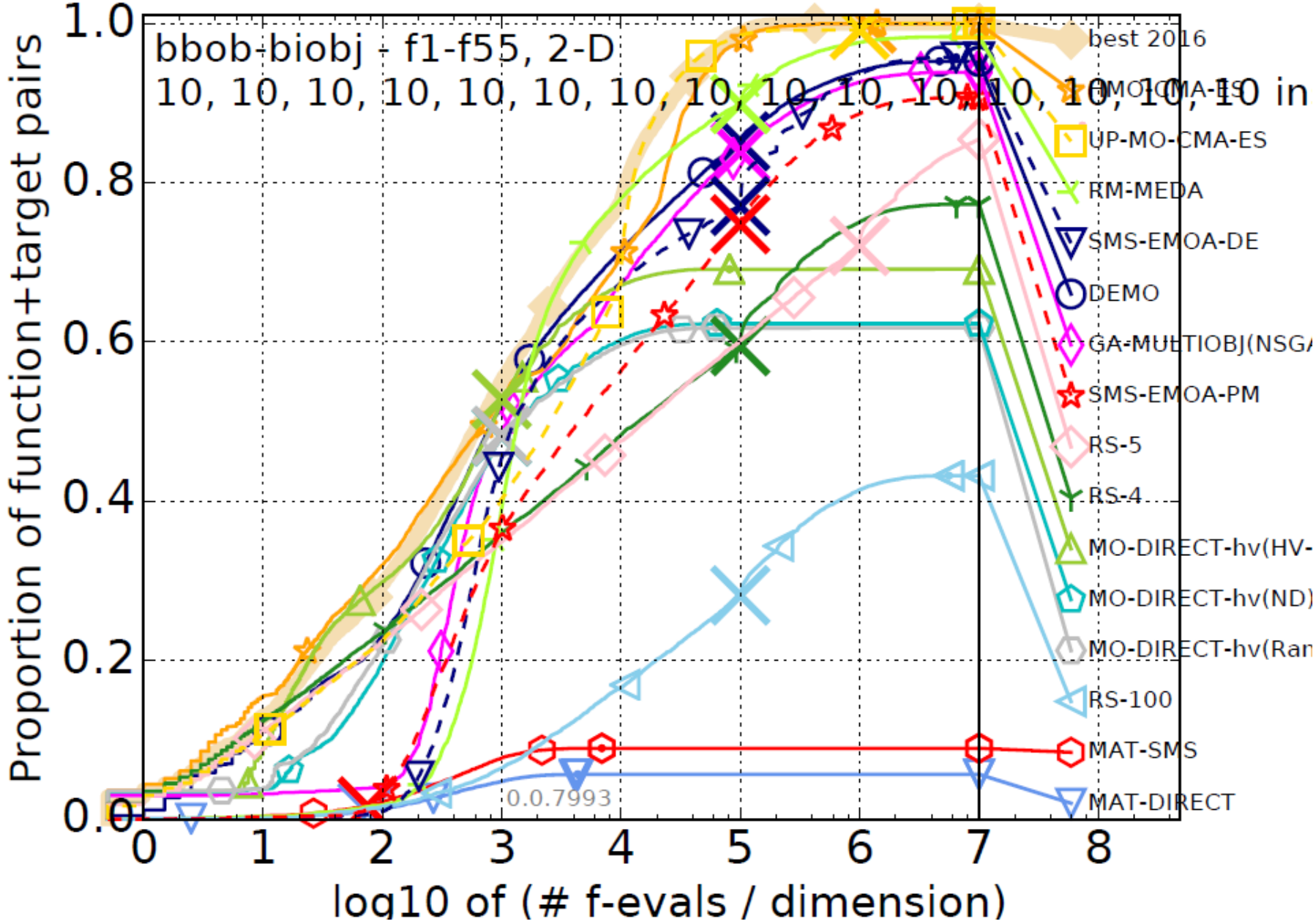
BBOB-2016: Instances 6-10 (2-D)



BBOB-2016: Instances 1-10 (2-D)

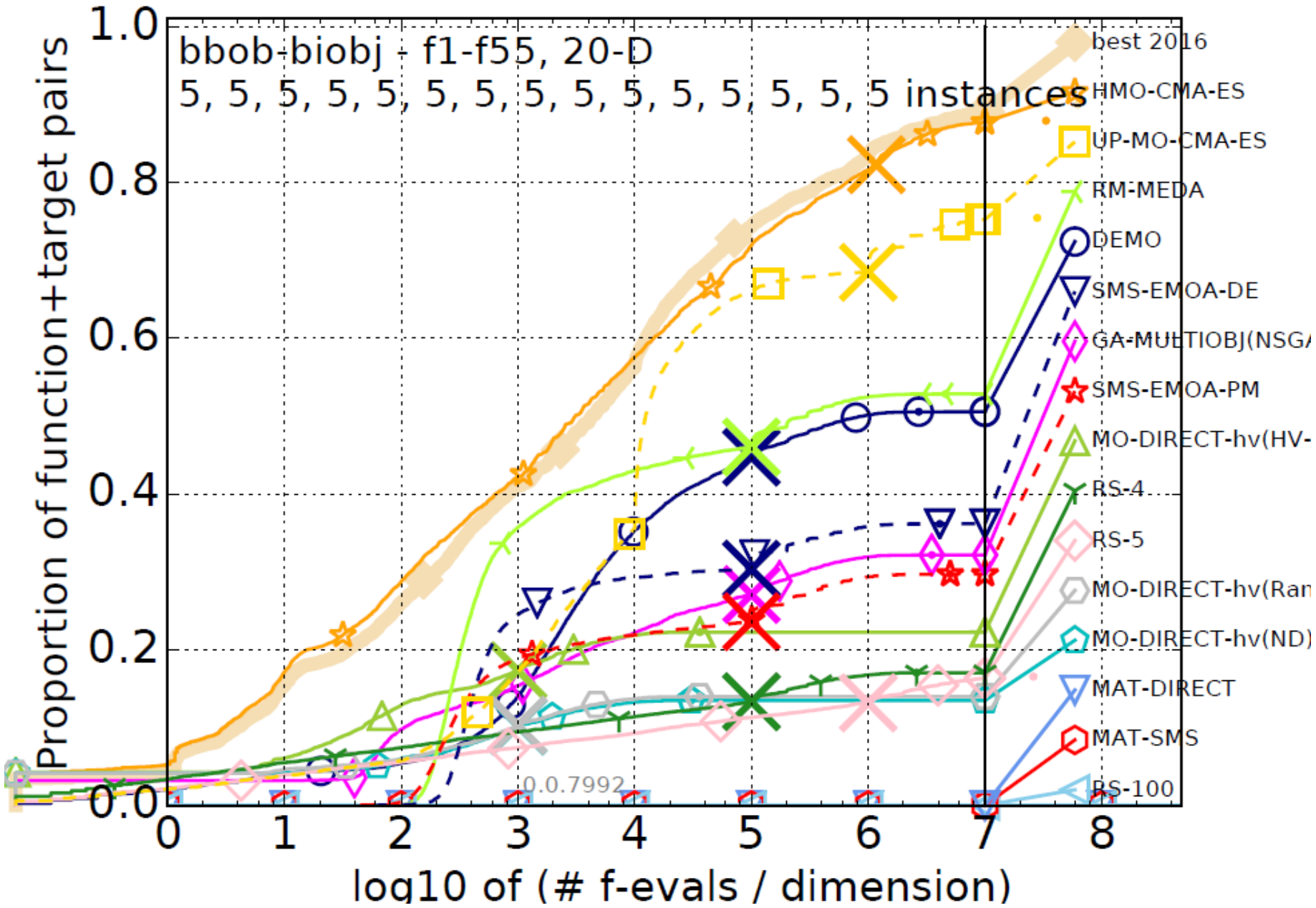


BBOB-2016: Instances 1-10 (same data)

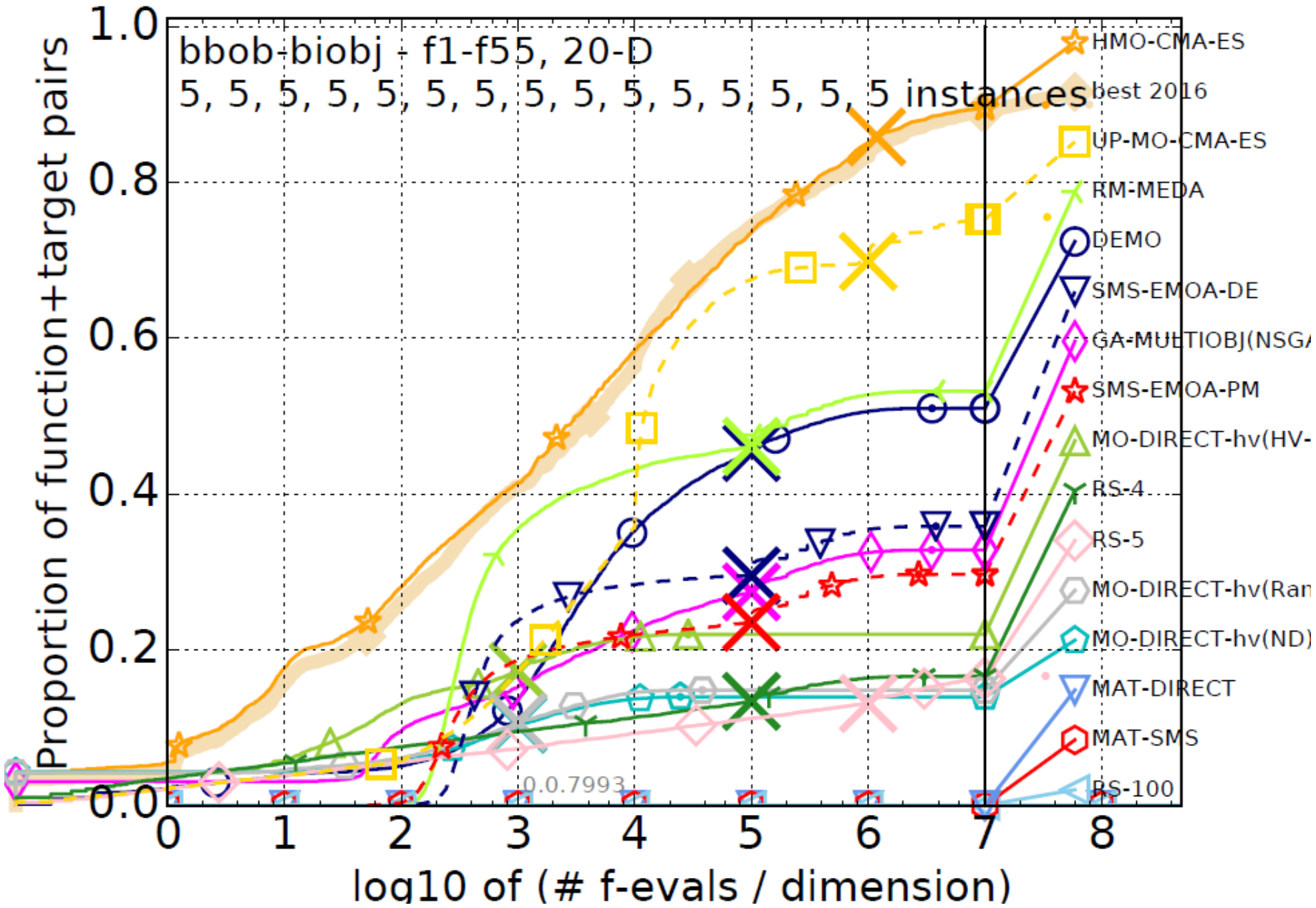


now 20-D

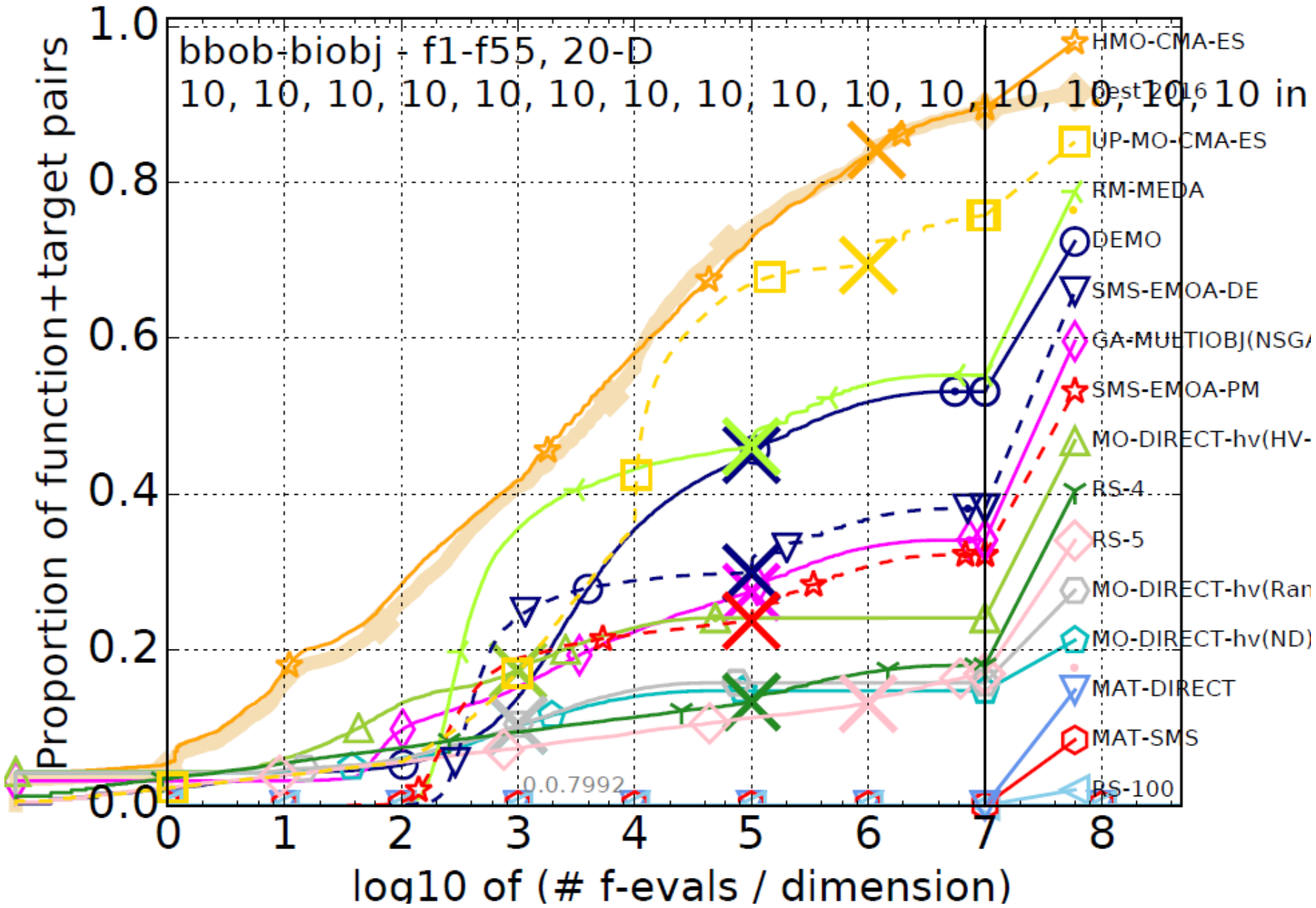
BBOB-2016: Instances 1-5



BBOB-2016: Instances 6-10



BBOB-2016: Instances 1-10

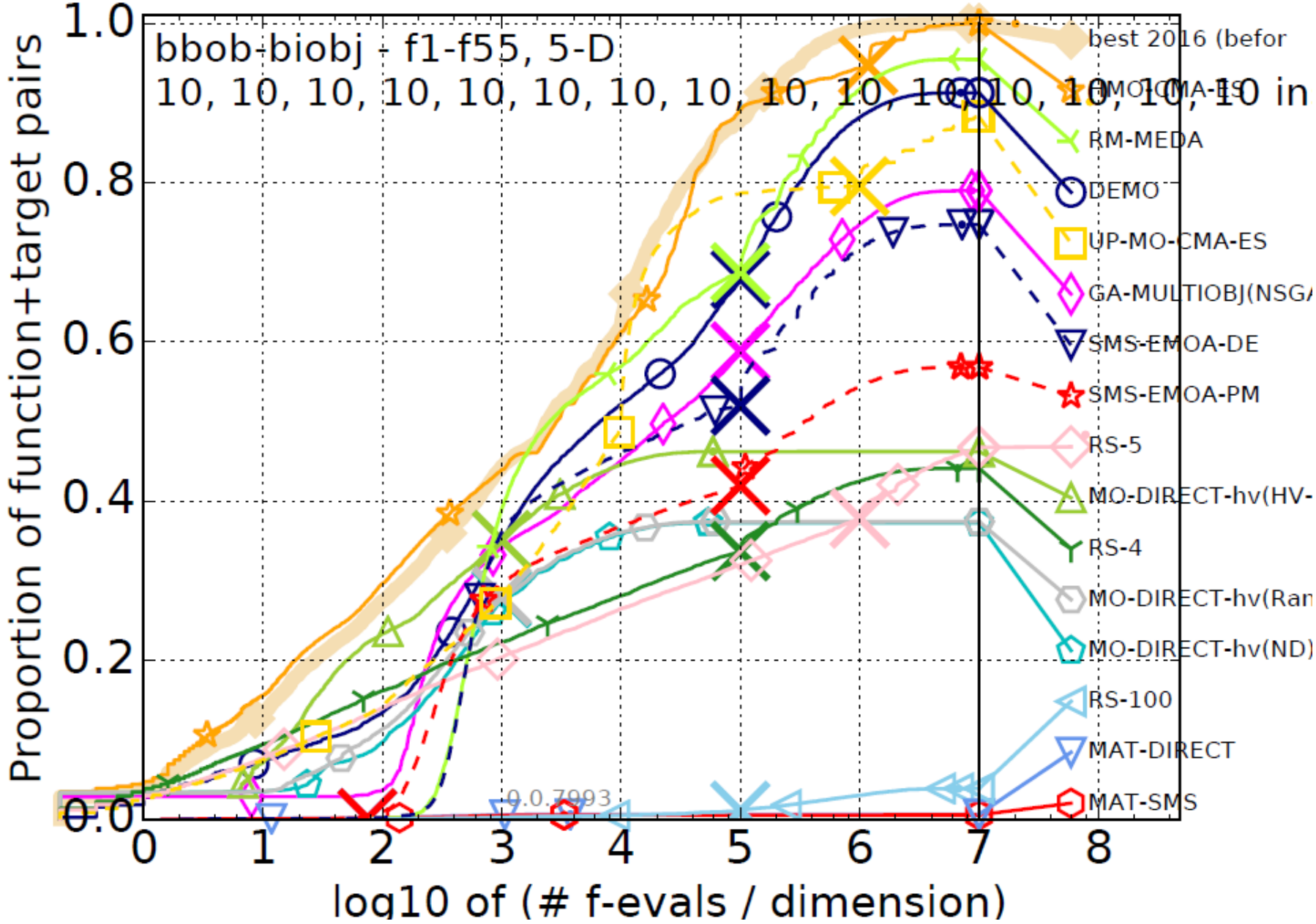


Influence of the Instance Set

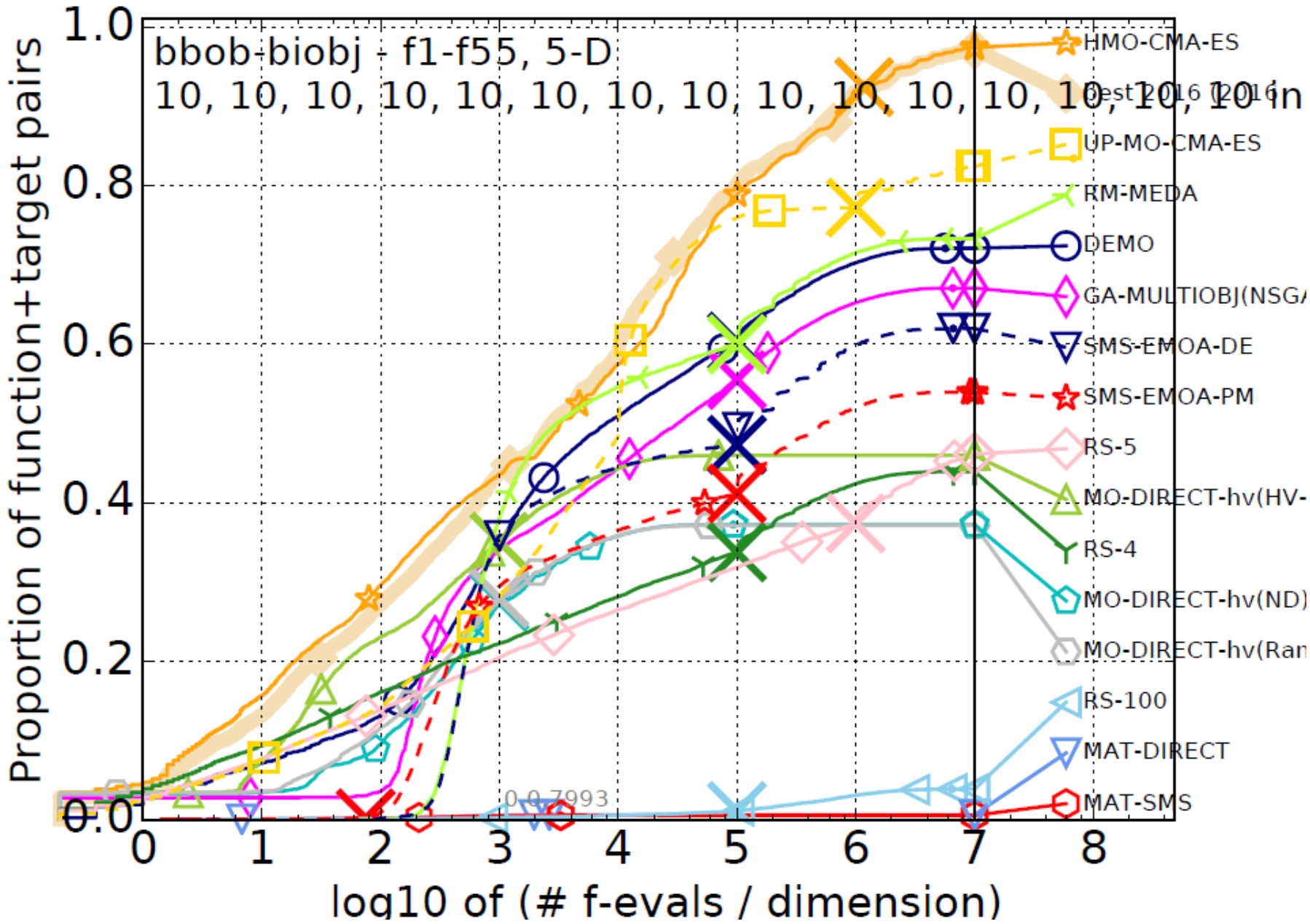
- is relatively small
- sometimes, last 5 instances harder, sometimes first 5 (depending on dimension)
- no indication of overfitting to the first 5 instances

❷ the influence of the reference set

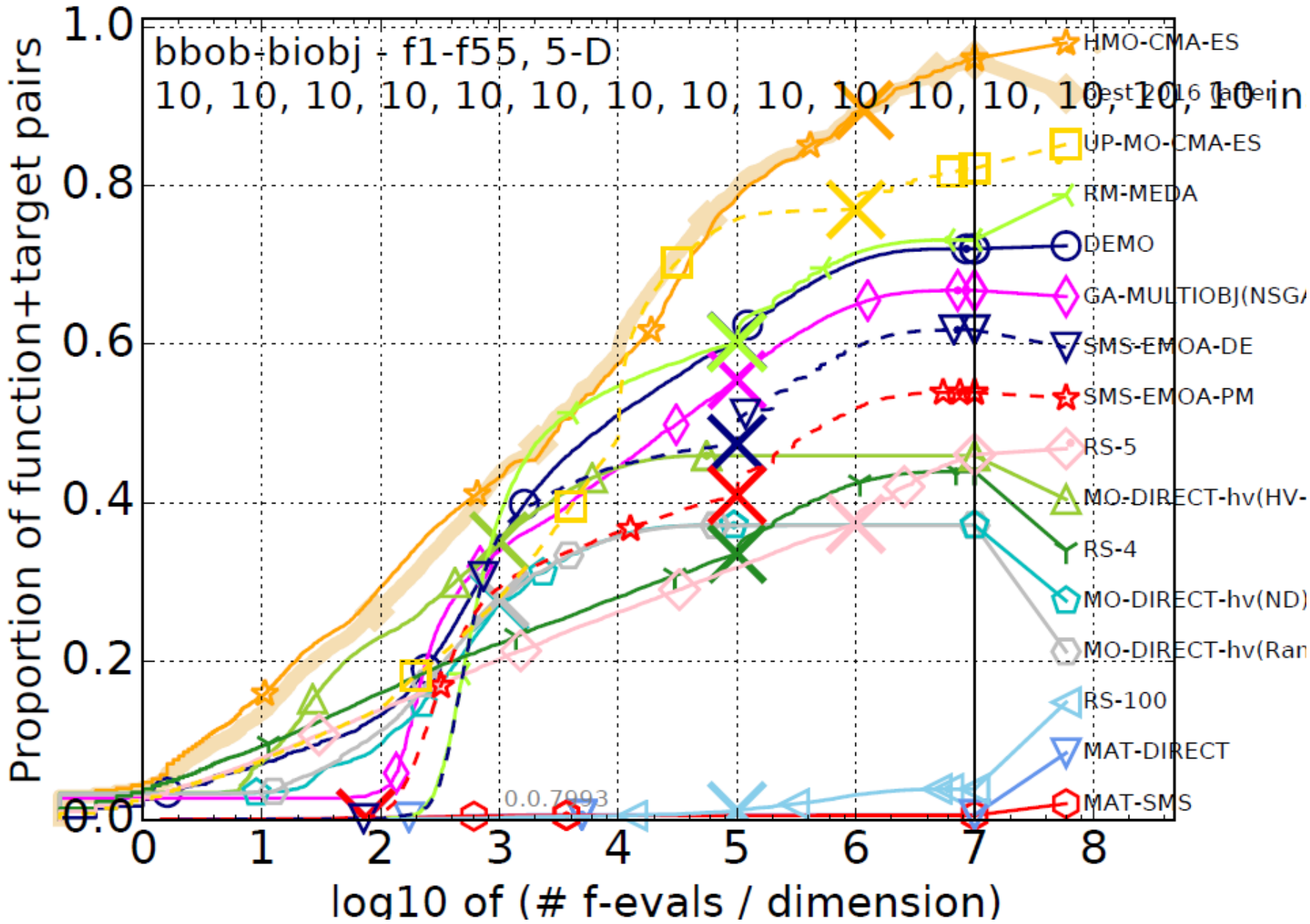
BBOB-2016: Reference Set Before



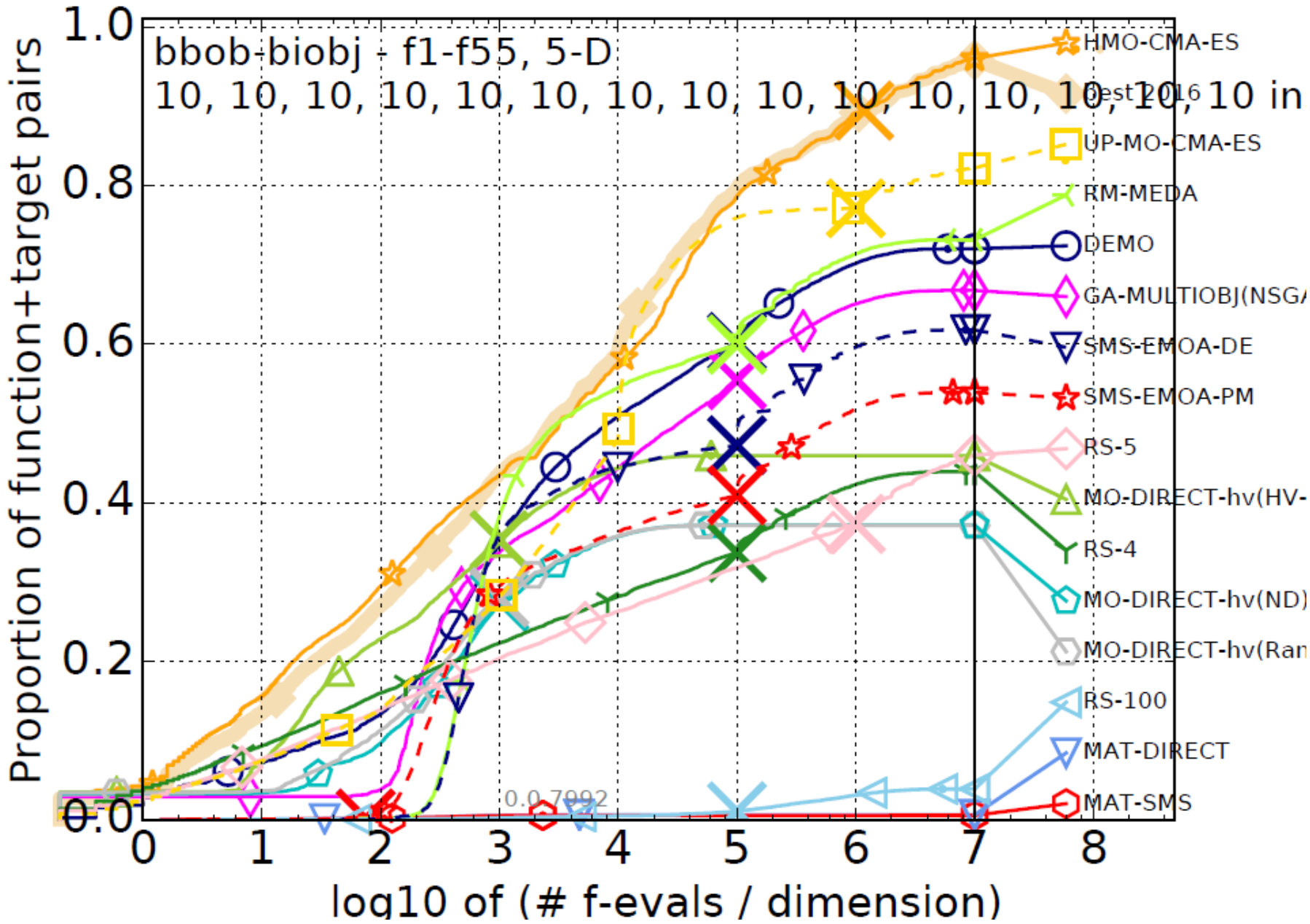
BBOB-2016: Reference Set BBOB-2016



BBOB-2016: Reference Set After

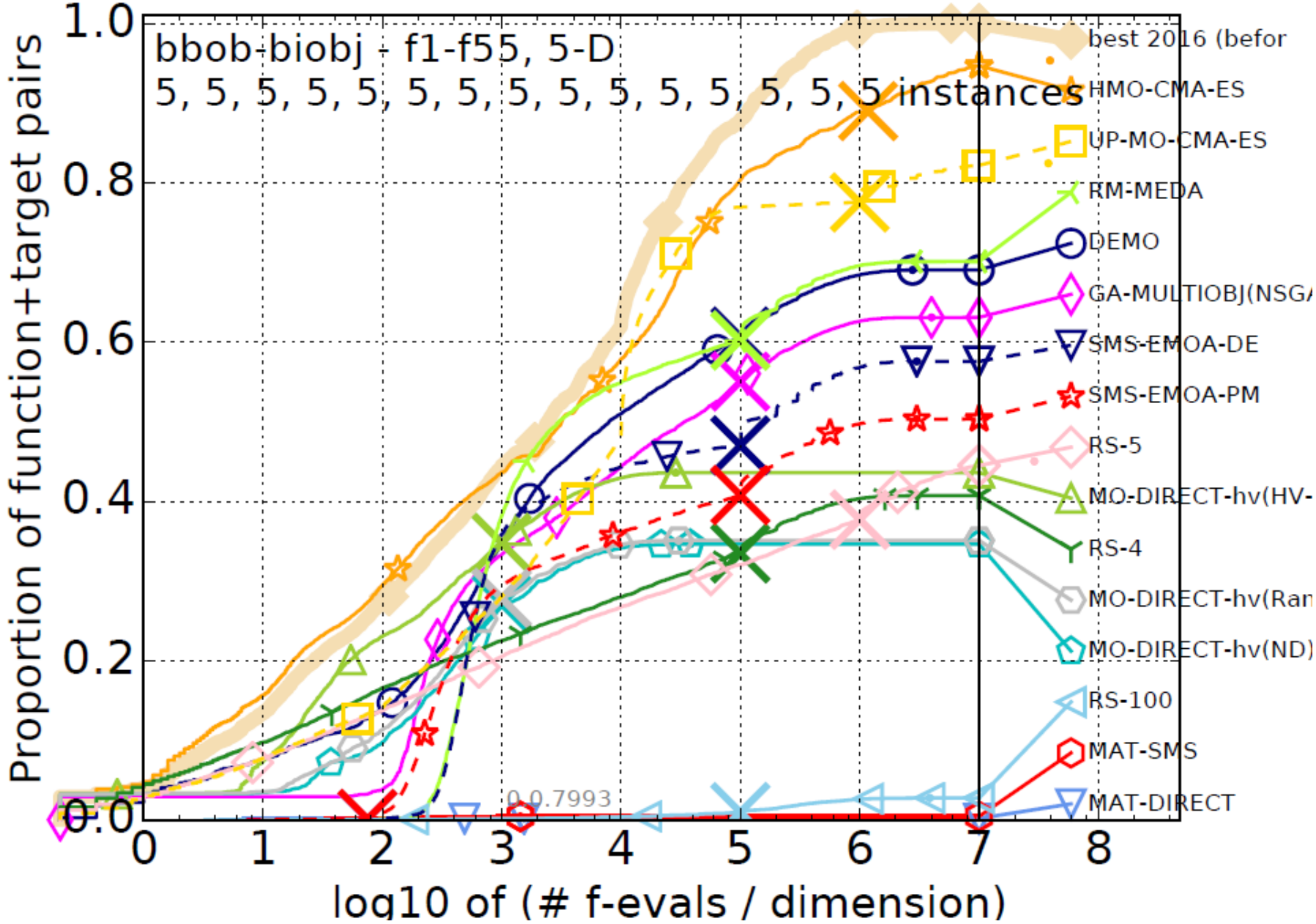


BBOB-2016: Reference Set Current_Best

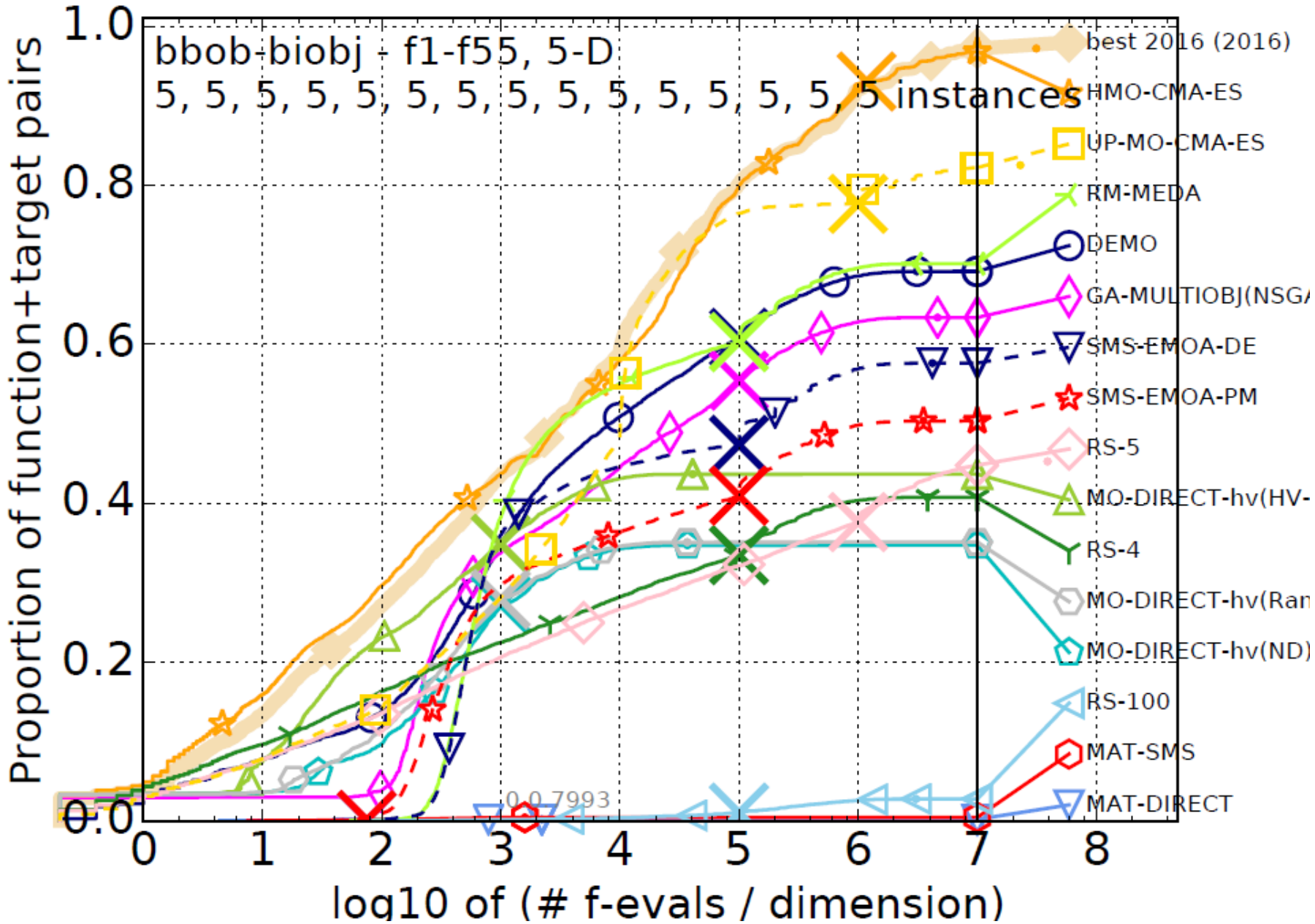


quick check: first 5 instances

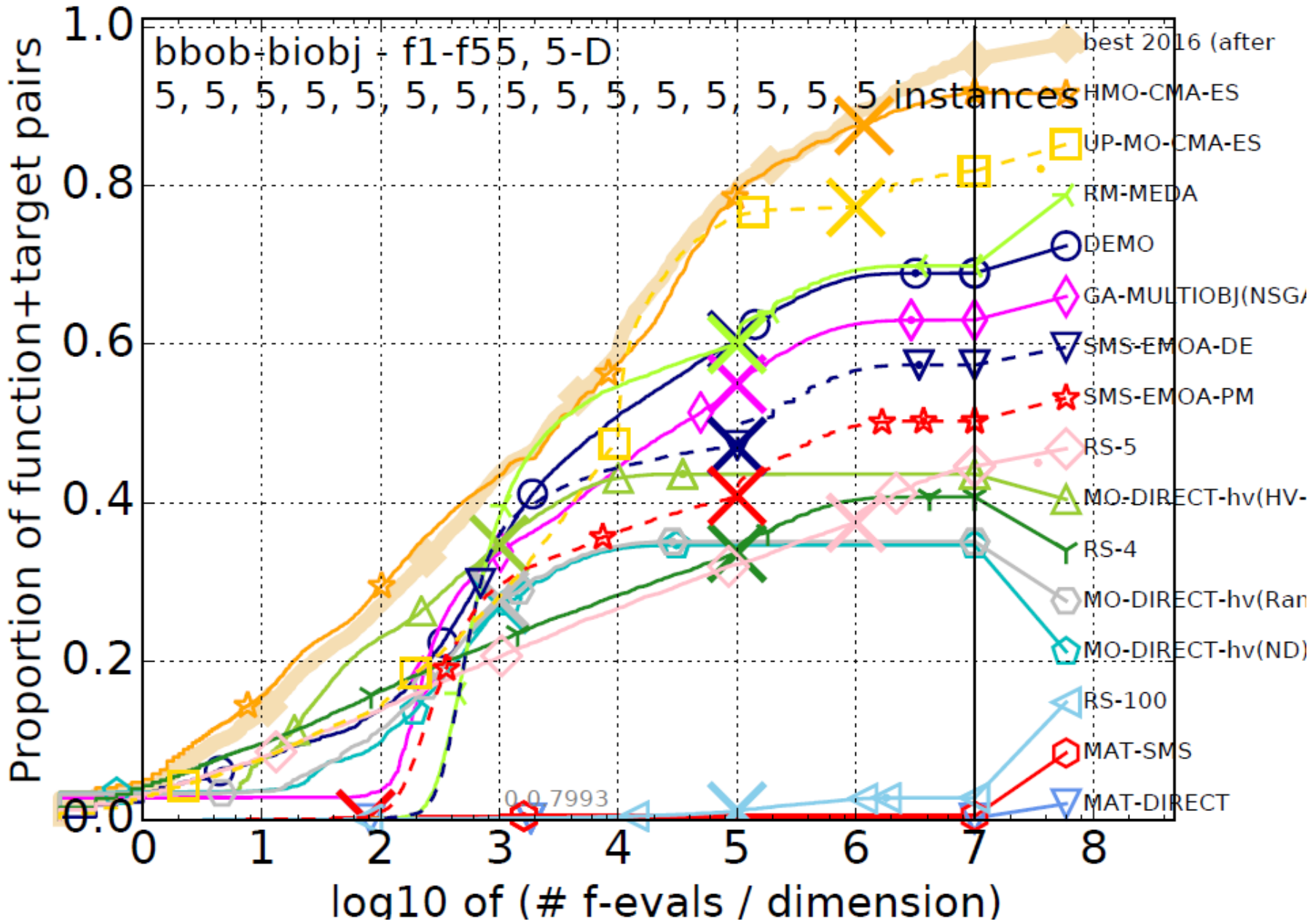
BBOB-2016: Reference Set Before



BBOB-2016: Reference Set BBOB-2016



BBOB-2016: Reference Set After



Influence of the Reference Set

- impact by the workshop algorithms the largest
- mainly on second five instances
  - which means the provided and displayed reference sets were okay
- continue with the current best in the future
  - updated reference hypervolume values will be provided in one of the next releases (this summer for sure)
- investigations on the single functions show that for some, we still do not have a good enough reference set yet

The Future of COCO

- bi-objective data will be made available online in the next days
- towards more realistic problems
  - large-scale test suite soon ready for release
  - constraints potentially ready in 2017
  - "almost real-world" problems
- online visualization of data

Your Participation it Welcome...

- ...always ☺

- benchmark your own algorithm and submit next year
- report bugs, issues, and feature requests
  - https://github.com/numbbo/coco
- contribute to the code base on github
  - issue tracker has special flag easy

- or even join us in Paris
  - as an engineer (funding for 1 year available)
  - or as postdoc, PhD student, or intern