MaxSAT Evaluation 2023

Ruben Martins
Carnegie Mellon University

Jeremias Berg
University of Helsinki

Matti Järvisalo
University of Helsinki

Andreas Niskanen
University of Helsinki

https://maxsat-evaluations.github.io/

SAT 2023, July 7, 2023

What is Maximum Satisfiability?

- Maximum Satisfiability (MaxSAT):
  - Set of (hard) clauses.
    - must be satisfied.
  - Linear objective with positive coefficients
    - Alternatively, a set of (weighted) soft clauses

- **Goal**: Minimize a linear objective function w.r.t. set of hard clauses.
  - Equivalent to maximizing (minimizing) the sum of the weights of satisfied (unsatisfied) soft clauses

Setup

Same structure as the one used in MaxSAT Evaluations 2017-2022:

- Source disclosure requirement:
  - Increase the dissemination of solver development
- Solver description using IEEE Proceedings style:
  - Better understanding of the techniques used by each solver
- Benchmark description using IEEE Proceedings style
  - Better understanding of the nature of each benchmark
- **Descriptions collected in proceeding, published at MSE website.**

News for this year

- New track names
  - Complete track $\rightarrow$ Exact track
  - Incomplete track $\rightarrow$ Anytime track
    - Anytime track moved away from StarExec
- Incremental track
  - No solvers submitted.

News for this year

- New track names
  - Complete track $\rightarrow$ Exact track
  - Incomplete track $\rightarrow$ Anytime track
    - Anytime track moved away from StarExec
- Incremental track
  - No solvers submitted.

Evaluation tracks

Evaluation tracks:

- Exact:
  - Weighted
  - Unweighted
- Anytime:
  - Weighted
  - Unweighted
- No distinction between industrial and crafted benchmarks

Execution environment

The exact track of MSE 2023 was run on the **new** nodes at StarExec:

- `https://www.starexec.org/`
- Intel(R) Xeon(R) Gold 6334 CPU @ 3.60GHz
- 18432 KB Cache, 264 GB Memory
- Two solvers per node

Anytime track run on a computing cluster at the University of Helsinki, part of the Finnish Computing Competence Infrastructure (FCCI).

- `https://www2.helsinki.fi/en/infrastructures/fcci`
- Intel(R) Xeon(R) CPU E5-2670 0 @ 2.60GHz
- 20 MB Cache, 57 GB Memory

Execution environment

Execution environment:

- Exact:
  - Time limit: 3600 seconds
  - Memory limit: 32 GB

- Anytime track:
  - Two time limits: 60 seconds and 300 seconds
  - Memory limit: 32 GB

Benchmark Selection

- Exact track:
  - Benchmarks selected **randomly** with a limit on the number of instances per family

- Anytime track:
  - Hard instances of the exact track.
    - Instances that **cannot be solved optimally in 60 seconds** by any participants of the exact tracks.

New benchmarks

- Judgment Aggregation (90 instances)
  - Ari Conati, Andreas Niskanen, Matti Jarvisalo
- Inconsistency Measurement(184 instances)
  - Andreas Niskanen, Isabelle Kuhlmann, Matthias Thimm, Matti Jarvisalo
- Optimizing BDDs (120 instances)
  - Hao Hu, Marie-José Huguet, Mohamed Siala
- Pareto-Optimal Interpretations for Black-Box Models (352 instances)
  - Hazem Torfah, Shetal Shah, Supratik Chakraborty, S. Akshay, Sanjit A. Seshia

Thank you to everyone who submitted benchmarks!

MSE 2023 benchmarks

Exact track:

- Unweighted (572 instances)
- Weighted (558 instances)

Anytime track:

- Unweighted (179 instances)
- Weighted (160 instances)

Exact Track

Solvers

MaxSAT approaches in MSE 2023:

Solver	Hitting Set	Unsat-based	Sat-Unsat	Other
CASHWMaxSAT		✓		ILP
MaxCDCL	✓	✓		B&B, ILP
EvalMaxSAT		✓		ILP
Open-WBO		✓		
CGSS		✓		ILP
Pacose			✓	Pre

- Running a sequence of solvers is becoming popular
- ILP is widely adopted by MaxSAT solvers as a preprocessing step
- Each solver has multiple versions with different features
  - 14 submissions for unweighted
  - 12 submissions for weighted

Results*

Unweighted: 572 instances

Solver	#Solved	Time (Avg)

Weighted: 558 instances

Solver	#Solved	Time (Avg)

* Only one version of each solver is shown in top-3

Results*

Unweighted: 572 instances

Solver	#Solved	Time (Avg)
EvalMaxSAT (SCIP)	433	326.11
MaxCDCL (SCIP600, MaxHS900)	430	244.12
CASHWMAXSAT-CorePlus (SCIP)	428	401.84

Weighted: 558 instances

Solver	#Solved	Time (Avg)

* Only one version of each solver is shown in top-3

Results*

Unweighted: 572 instances

Solver	#Solved	Time (Avg)
EvalMaxSAT (SCIP)	433	326.11
MaxCDCL (SCIP600, MaxHS900)	430	244.12
CASHWMAXSAT-CorePlus (SCIP)	428	401.84

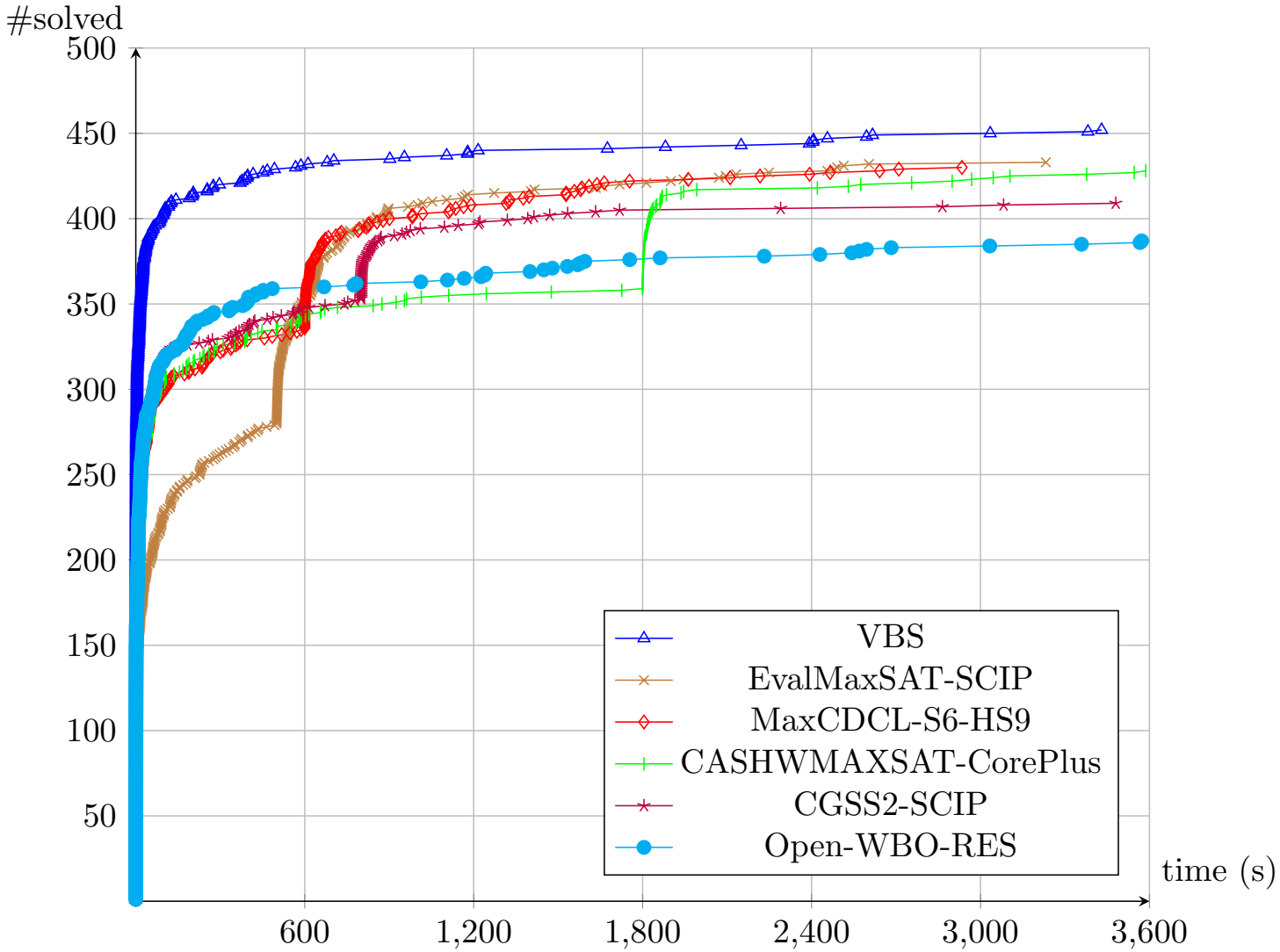
Weighted: 558 instances

Solver	#Solved	Time (Avg)
MaxCDCL (SCIP600, MaxHS1200)	446	276.74
EvalMaxSAT (SCIP)	442	405.04
CASHWMAXSAT-CorePlus	432	128.04

* Only one version of each solver is shown in top-3

Detailed Results

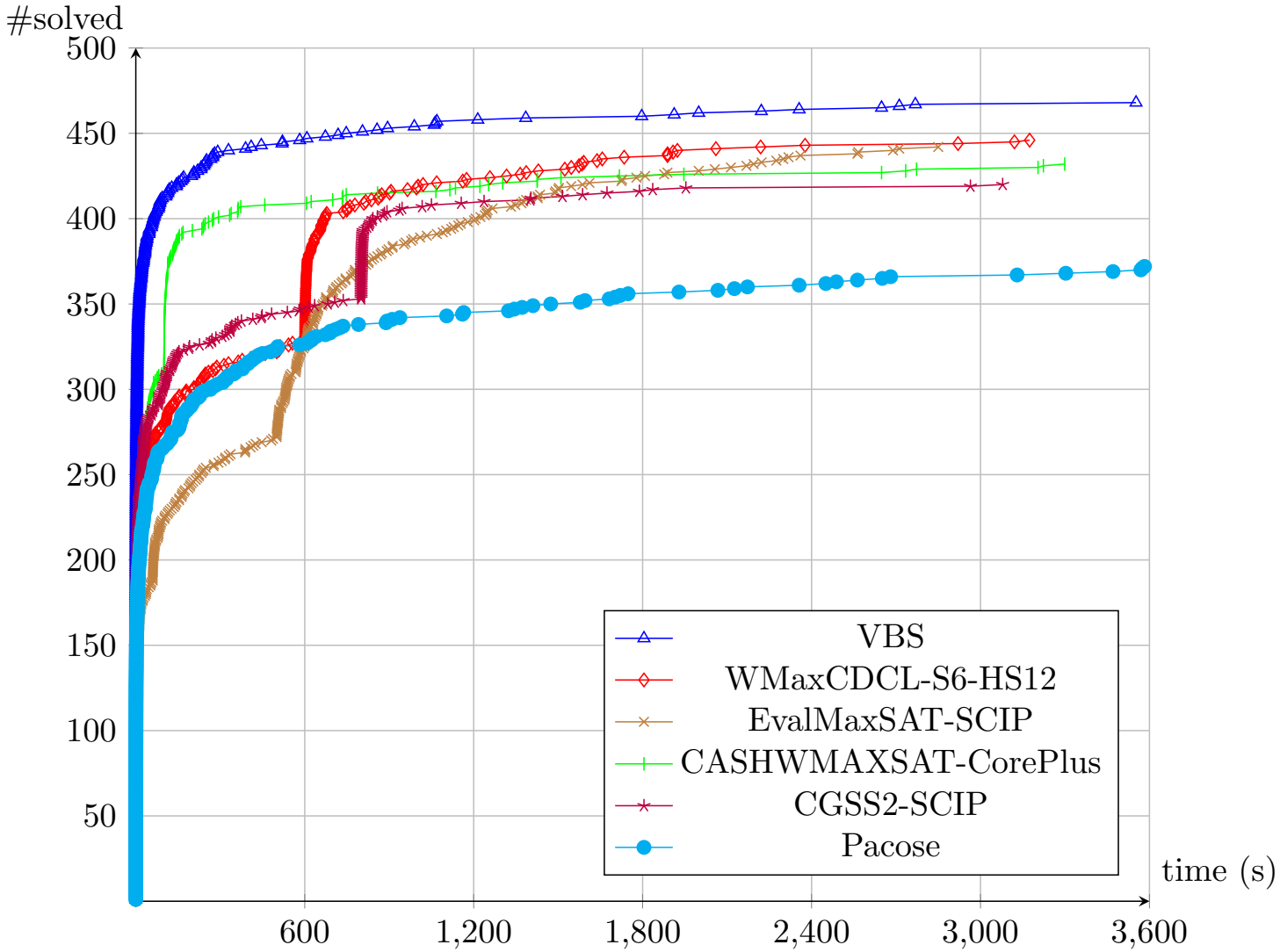
Unweighted



- Only one version of each solver is shown in the cactus
- Gap between best solver and VBS is small

Detailed Results

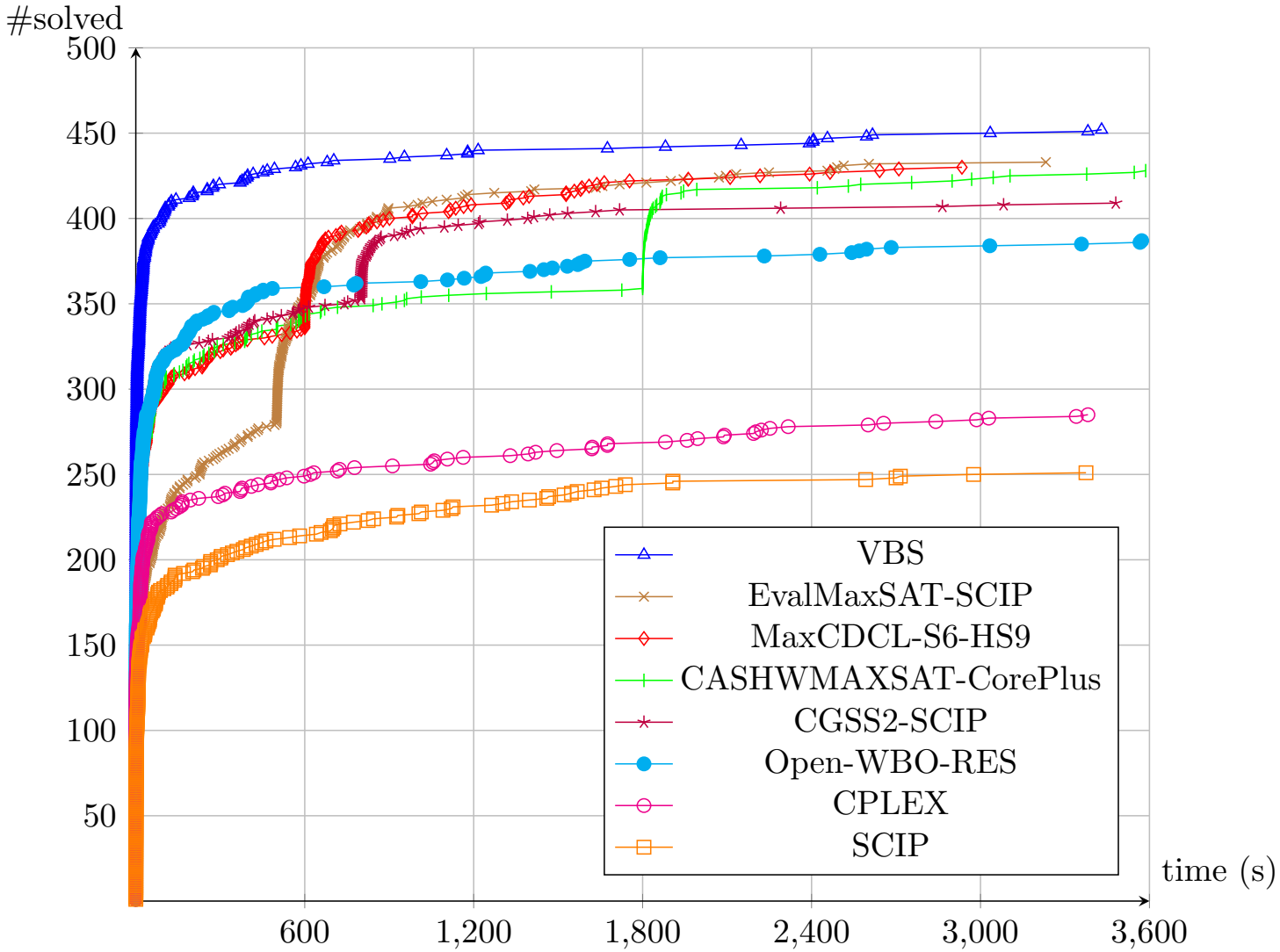
Weighted



- Only one version of each solver is shown in the cactus
- Gap between best solver and VBS is small

How well do ILP solvers perform by themselves?

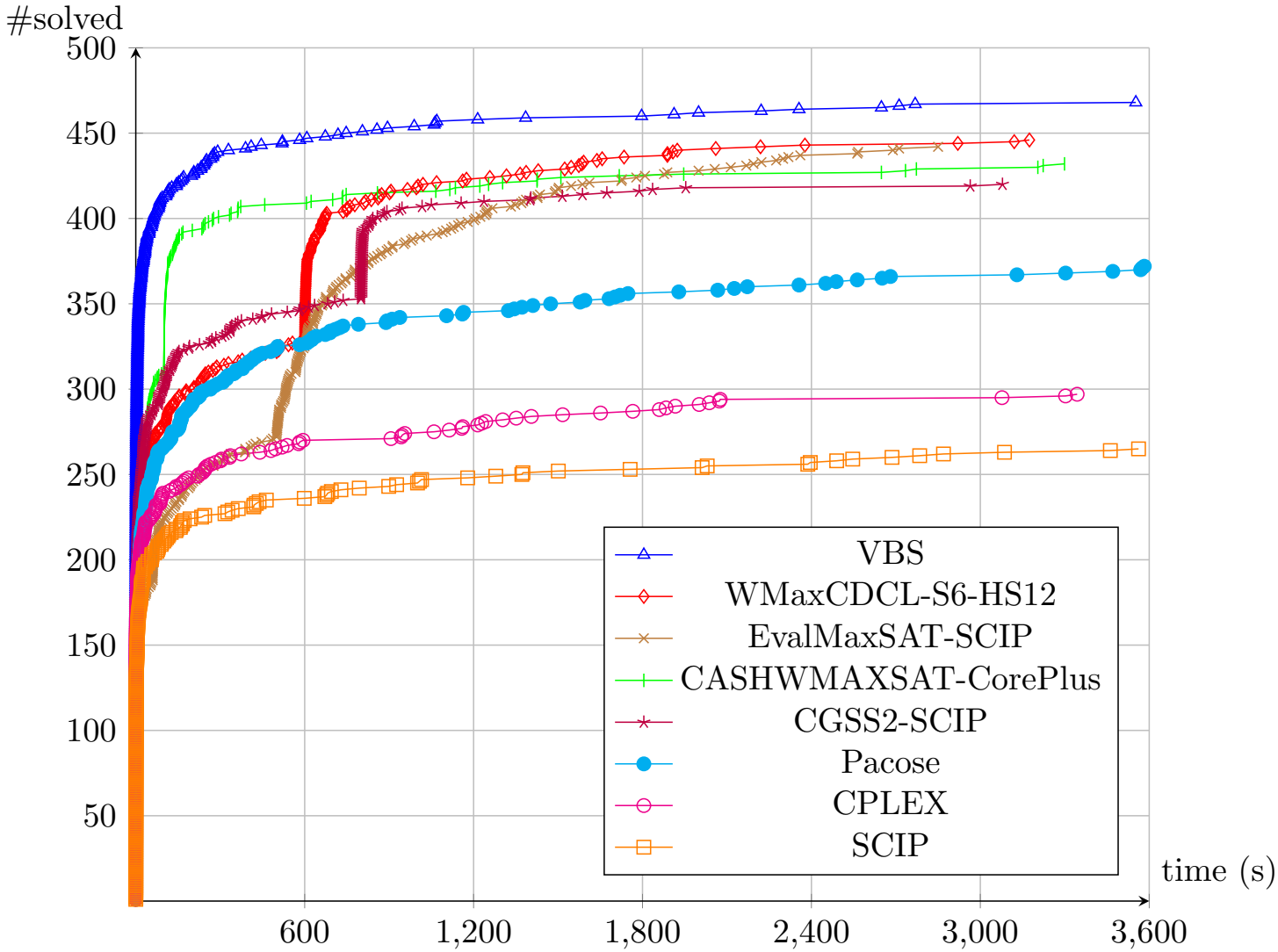
Unweighted



- ILP solvers by themselves are not competitive with MaxSAT solvers

How well do ILP solvers perform by themselves?

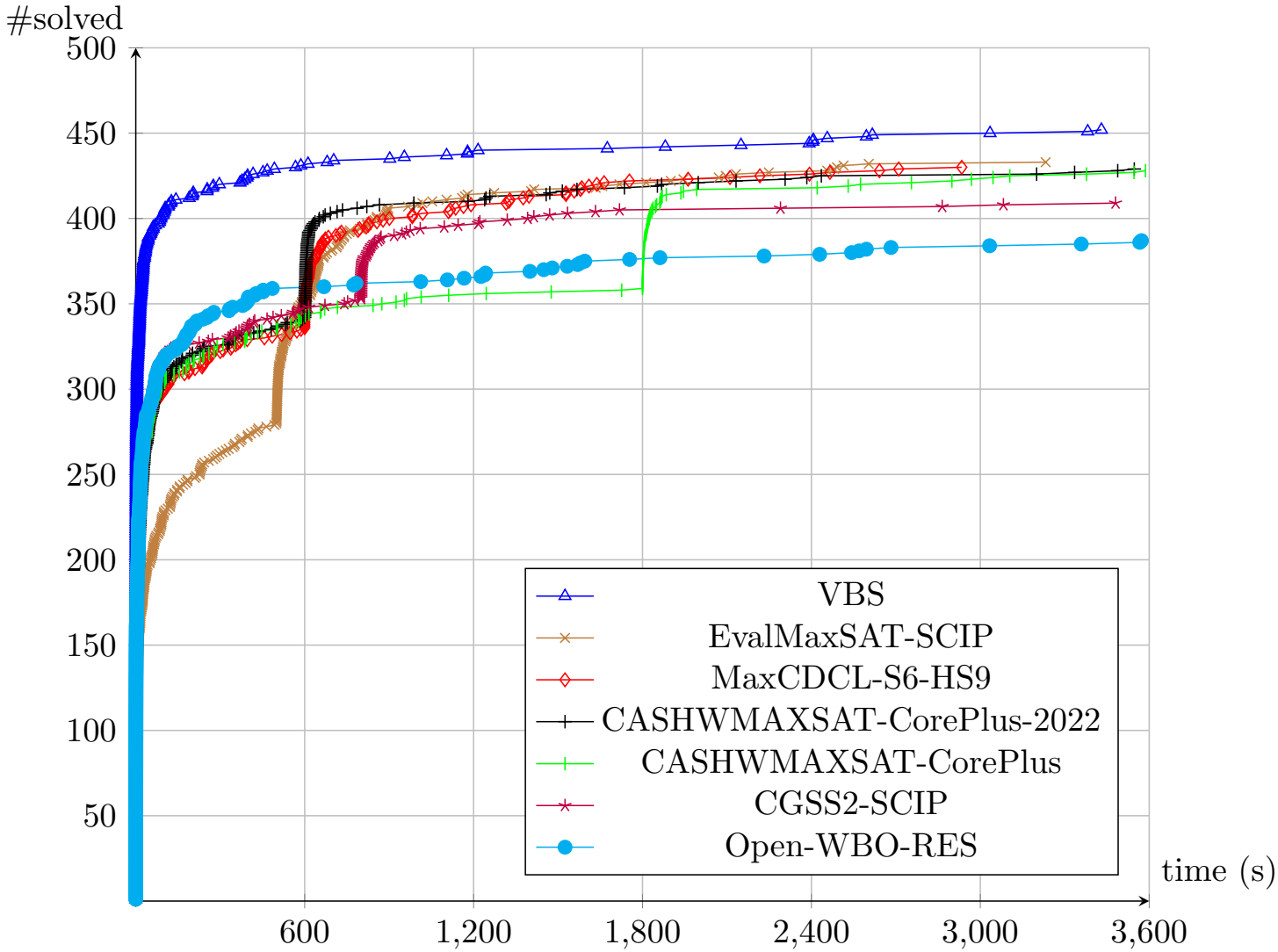
Weighted



- ILP solvers by themselves are not competitive with MaxSAT solvers

What is the improvement from MSE 2022?

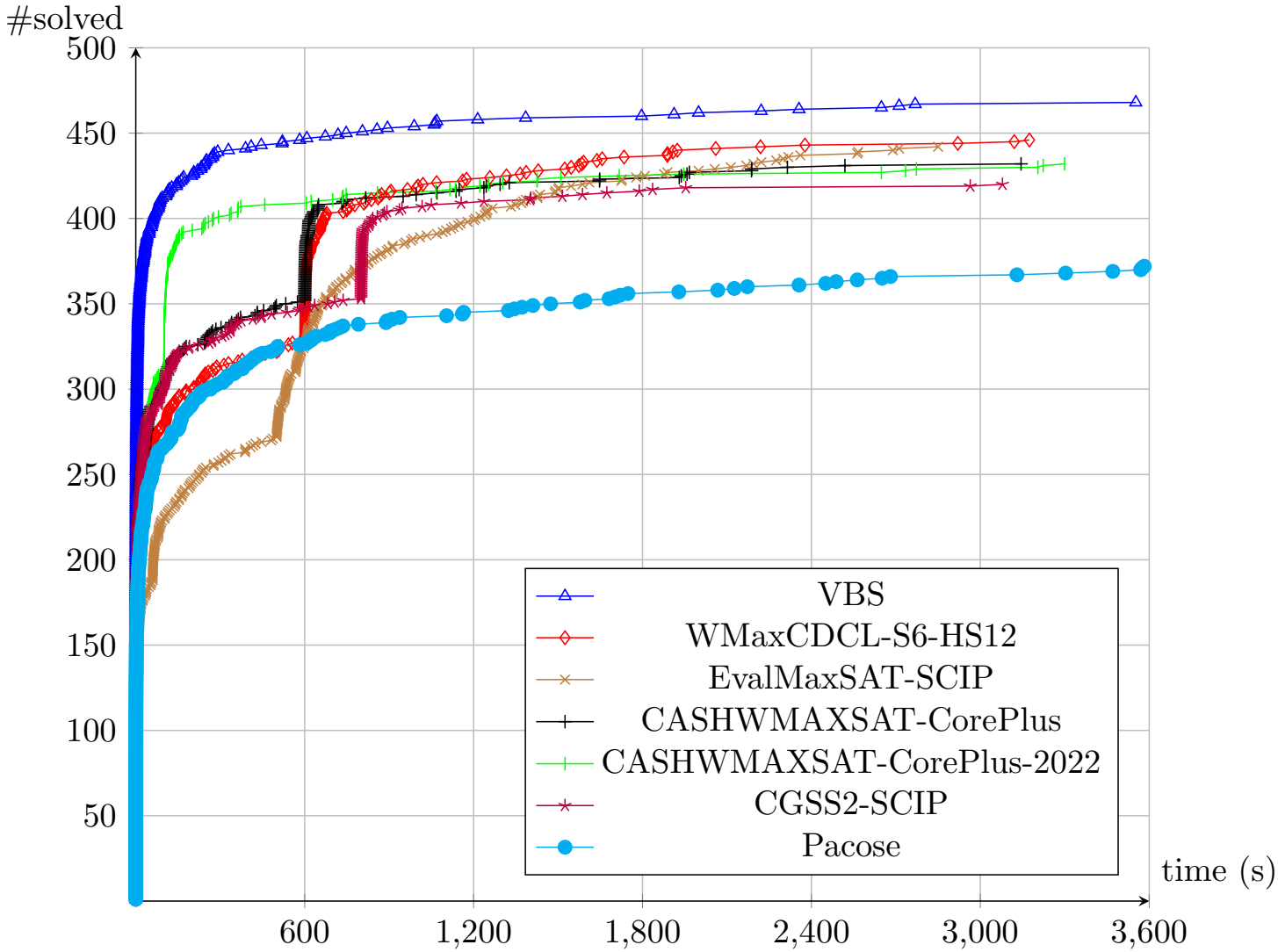
Unweighted



- Small improvement from MSE 2022

What is the improvement from MSE 2022?

Weighted



- Small improvement from MSE 2022

Anytime Track

Ranking for anytime tracks

- Anytime score: computed by the sum of the ratios between the best solution found by a given solver and the best known solution:
  - $\text{SCORE}(\textit{solver}, i) = \frac{(\text{cost of best known solution for i} + 1)}{(\text{cost of solution for i found by solver} + 1)}$
  - For an instance i score is 0 if no solution was found by that solver
  - For each instance the anytime score is a value in $[0, 1]$
- Ranking based on average of all scores.

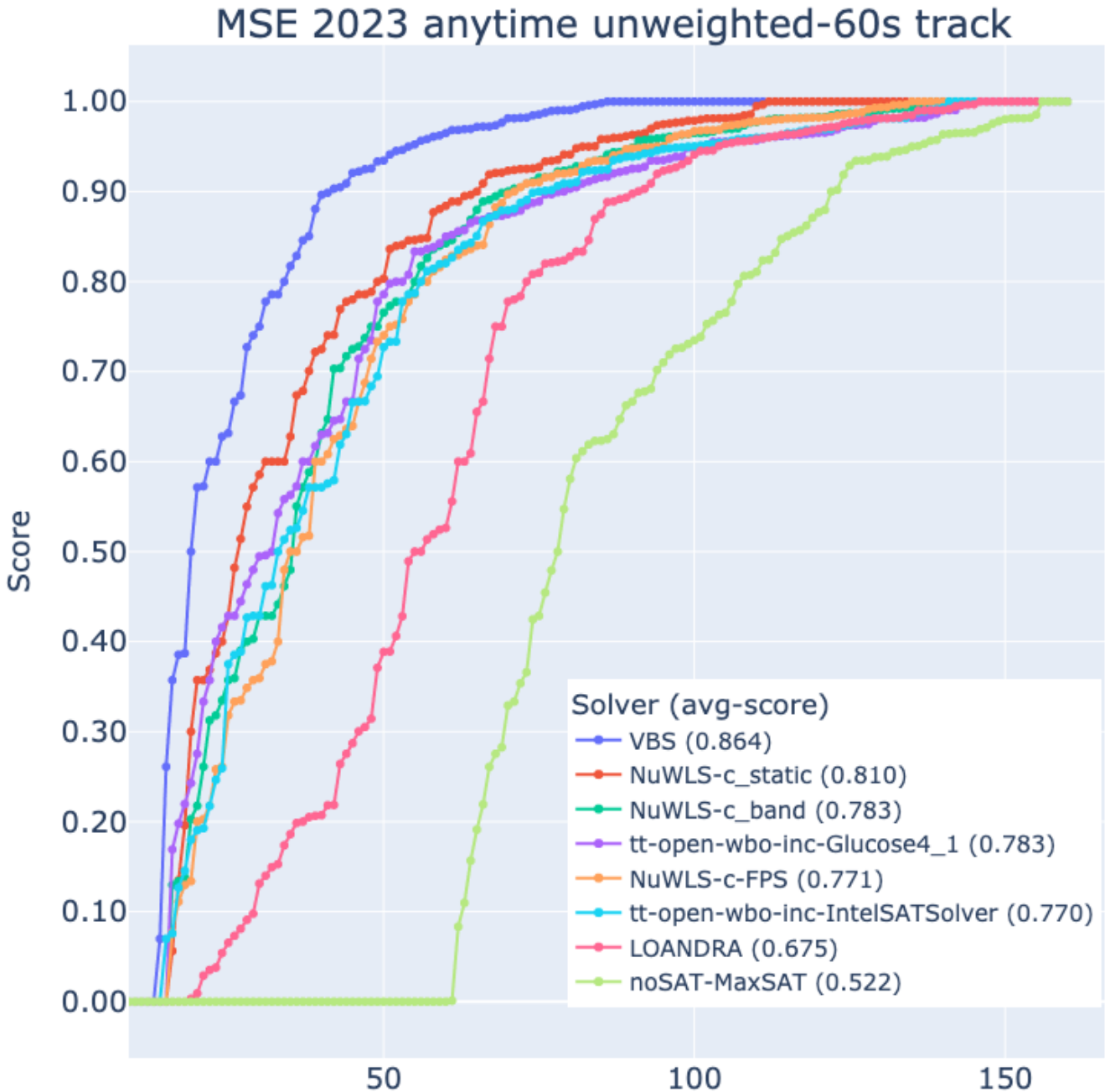
Solvers

- noSAT-MaxSAT
  - Ole Lübke and Sibylle Schupp
  - Local search without a SAT solver.
- NuWLS-c 2023 (last year's winner)
  - Yi Chu, Shaowei Cai, Chuan Luo
  - Local search combined with TT-Open-WBO-Inc
- NuWLS-c-Band, and -FPS (two variants)
  - Jiongzhi Zhen, Kun He, Mingming Jin, Zhuo Chen and Jinghui Xue
  - Local search (BandMaxSAT, farsighted sampling) combined with TT-Open-WBO-Inc
- Loandra (2022 version)
  - Jeremias Berg
  - Preprocessing, core-guided and SAT/UNSAT search combined.
- TT-Open-WBO-Inc (two variants)
  - Alexander Nadel,
  - One variant with Glucose 4.1 and another with IntelSAT.

Anytime track: Unweighted (60 seconds)

Solver	Score (avg)
NuWLS-c 2023	0.810
TT-Open-WBO-inc (Glucose 4.1)	0.783
NuWLS-c-Band	0.783
NuWLS-c-FPS	0.771
TT-Open-WBO-inc (IntelSAT)	0.770
Loandra	0.675
noSAT-MaxSAT	0.522

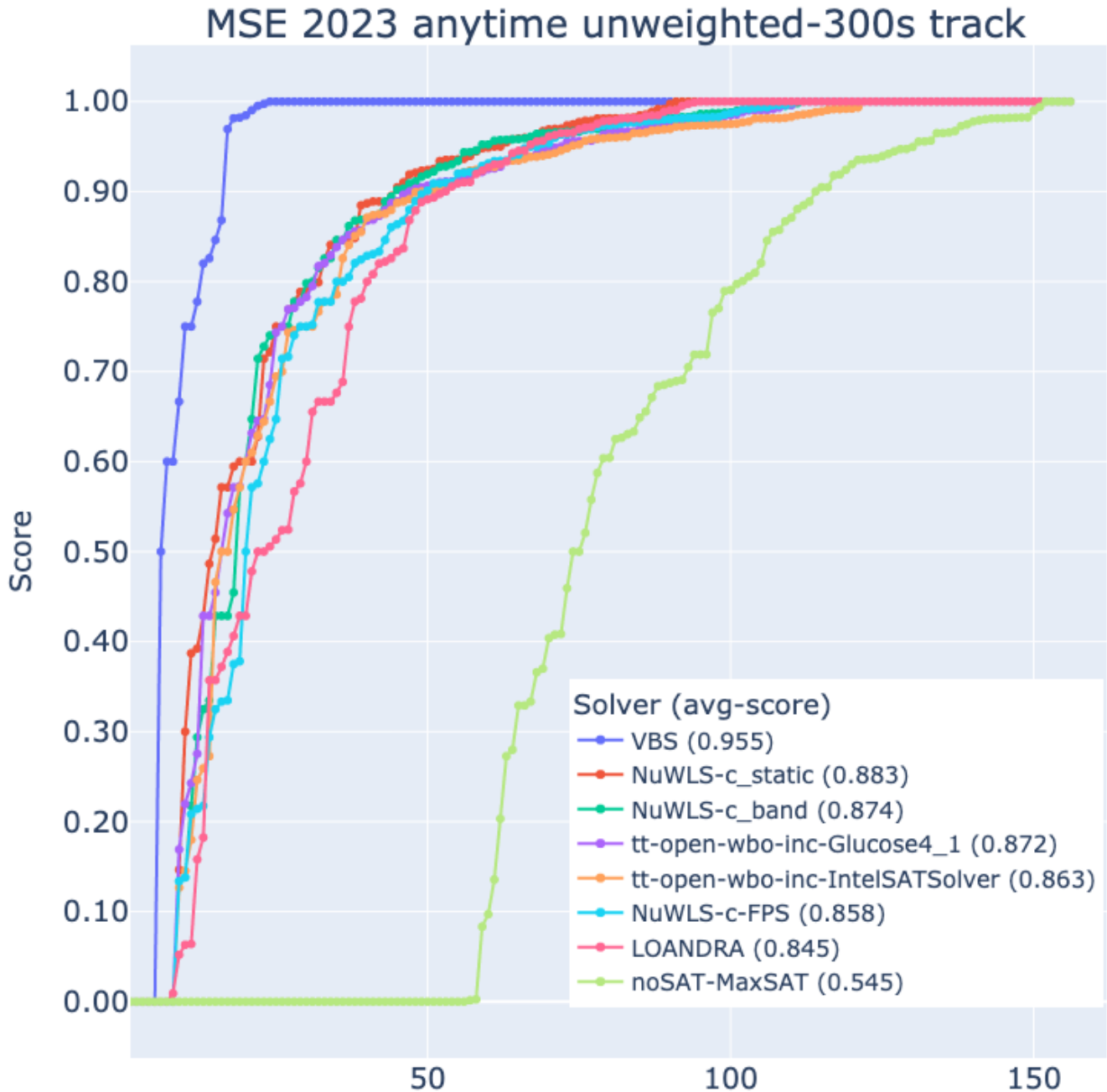
Unweighted 60s



Anytime track: Unweighted (300 seconds)

Solver	Score (avg)
NuWLS-c 2023	0.883
NuWLS-c-Band	0.874
TT-Open-WBO-inc (Glucose 4.1)	0.872
TT-Open-WBO-inc (IntelSATSolver)	0.863
NuWLS-c-FPS	0.858
Loandra	0.845
noSAT-MaxSAT	0.545

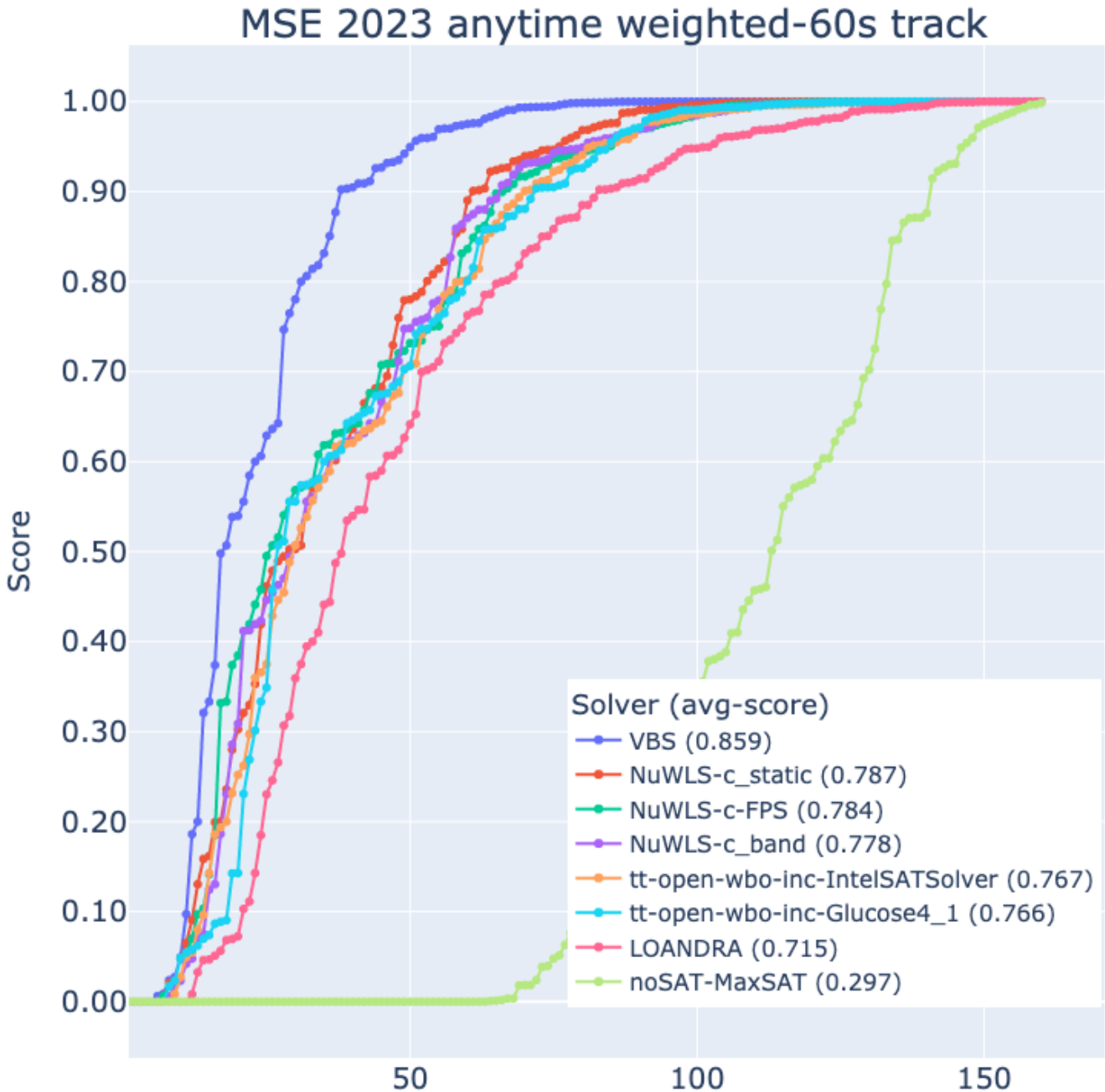
Unweighted 300s



Anytime track: Weighted (60 seconds)

Solver	Score (avg)
NuWLS-c 2023	0.787
NuWLS-c-FPS	0.784
NuWLS-c-Band	0.778
TT-Open-WBO-inc (IntelSAT)	0.767
TT-Open-WBO-inc (Glucose 4.1)	0.766
Loandra	0.715
noSAT-MaxSAT	0.297

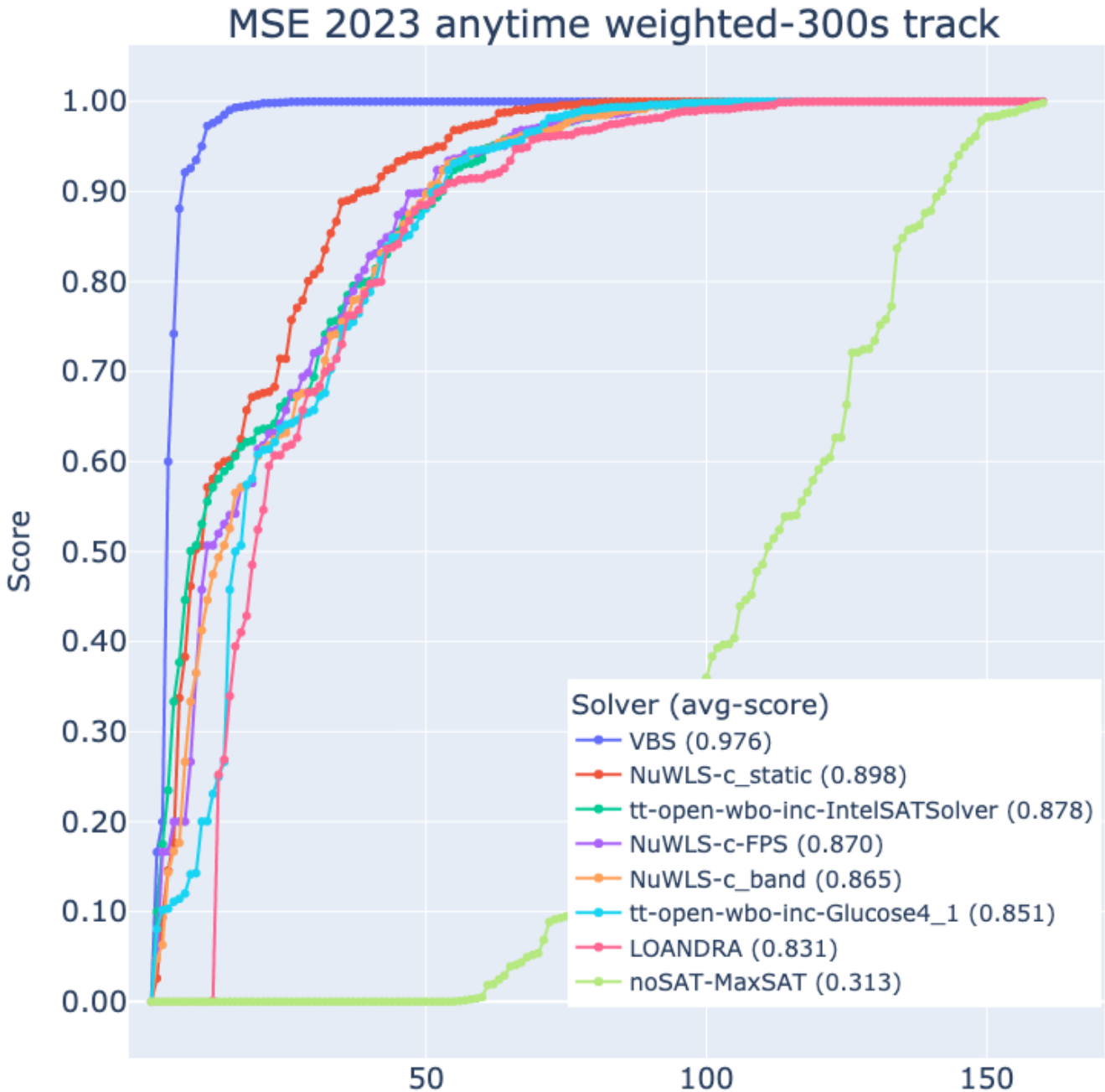
Weighted 60s



Anytime track: Weighted (300 seconds)

Solver	Score (avg)
NuWLS-c 2023	0.898
TT-Open-WBO-inc (IntelSAT)	0.878
NuWLS-c-FPS	0.870
NuWLS-c-Band	0.865
TT-Open-WBO-inc (Glucose 4.1)	0.851
Loandra	0.831
noSAT-MaxSAT	0.313

Weighted 300s



Webpages

MaxSAT Evaluation 2023 webpage

`https://maxsat-evaluations.github.io/2023/`

- Detailed results for each instance
- Description of the solvers and benchmarks
- Source code of the solvers
- Benchmarks and log files.

Thanks

Thanks to everyone that contributed solvers and benchmarks!
Without you this evaluation would not be possible!

Thanks to StarExec for allowing us to use their cluster for the exact track:

https://www.starexec.org/

Thanks to FCCI for supporting the anytime track with computational and data storage resources:

https://www2.helsinki.fi/en/infrastructures/fcci