A new tool for analysing and reporting solutions for the RCPSP and MMRCPSP

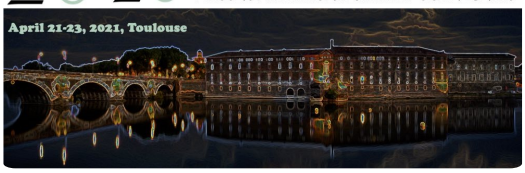
José Coelho[1,3], Mario Vanhoucke[1,2] and Ricardo Amaro[3]

Keywords: RCPSP, datasets, reporting results

[1]Ghent University, Belgium
e-mail: jose.coelho, mario.vanhoucke@ugent.be

[2]Vlerick Business School, Belgium
[3]Universidade Aberta, Portugal
e-mail: jose.coelho@uab.pt

PMS 2020 17TH. INTERNATIONAL WORKSHOP ON PROJECT MANAGEMENT AND SCHEDULING



1

1

PMS 2020

Agenda

- Introduction
- Reporting new results
- Update of tables of BKS on RCPSP and MMRCPSP
- An example of an experiment with NetRes
- Diversity of dataset CV
- Conclusion

If you want to refer to this presentation, please refer to:

- Coelho, J., Vanhoucke, M., & Amaro, R. (2021). A new tool for analysing and reporting solutions for the RCPSP and. Proceeding of the 17th International Workshop on Project Management and Scheduling, 4.
- Or refer to the papers mentioned at the end of this presentation

2

2

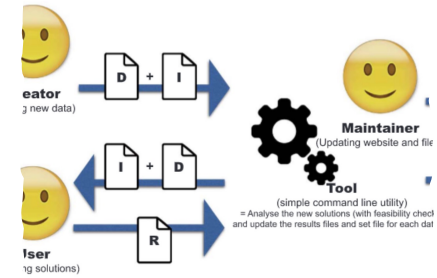
Introduction

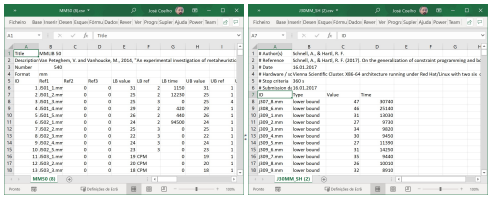
- Method in Vanhoucke and Coelho (2018) for reporting solutions in RCPSP and MMRCPSP, is implemented in the website: http://solutionsupdate.ugent.be/
- Complement other libraries: PSPLIB, MMLIB or OR-LIBRARY
- Record data on a standard format with:
  - LB/UB/OPT values
  - Best known solutions (starting times for each activity and activity-mode)
  - Project indicators calculated
- One result file for each run on a dataset, instead a submission for each instance
- Two new datasets for the RCPSP:
  - **NetRes**, presented in Vanhoucke and Coelho (2018) with a large number of instances that have a wide variety in the project indicators
  - **CV**, presented in Coelho and Vanhouke (2020), contains small but hard instances

3

3

Reporting new results 1/3



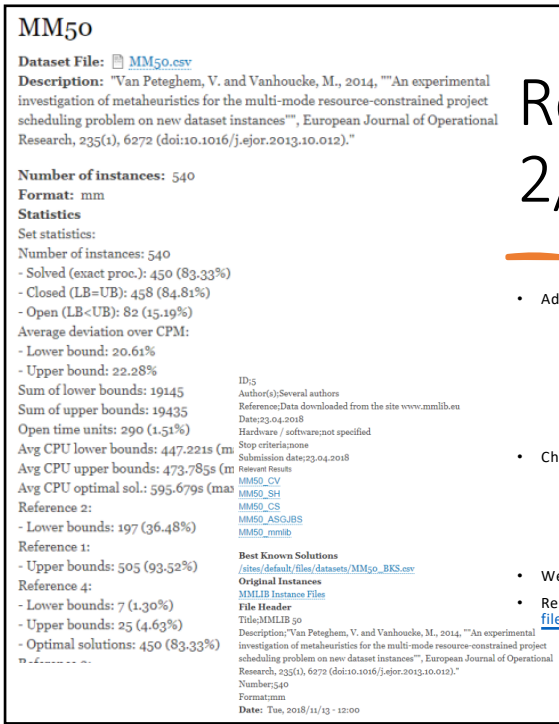


- Roles:
  - Creator – researcher that propose a new dataset
  - User – researcher that upload new results on a dataset
  - Maintainer – researcher that mantains the most recent data
- Two standard files types:
  - Results file (CSV format): single file with all results from a run, on a dataset, one line per result (instance/run)
  - Dataset file (CSV format): specify all instances in the dataset, one line per instance, having updated results and also project indicators;
  - Instance files: one file for each instance in the original format

4

4

Reporting new results 2/3



MM50

Dataset File: MM50.csv
Description: "Van Peteghem, V. and Vanhoucke, M., 2014, ""An experimental investigation of metaheuristics for the multi-mode resource-constrained project scheduling problem on new dataset instances"", European Journal of Operational Research, 235(1), 6272 (doi:10.1016/j.ejor.2013.10.012)."

Number of instances: 540
Format: mm
Statistics
Set statistics:
Number of instances: 540
- Solved (exact proc.): 450 (83.33%)
- Closed (LB=UB): 458 (84.81%)
- Open (LB<UB): 82 (15.19%)
Average deviation over CPM:
- Lower bound: 20.61%
- Upper bound: 22.28%
Sum of lower bounds: 19145
Sum of upper bounds: 19435
Open time units: 290 (1.51%)
Avg CPU lower bounds: 447.2215 (m
Avg CPU upper bounds: 473.7855 (m
Avg CPU optimal sol.: 595.6795 (ma
Reference 2:
- Lower bounds: 197 (36.48%)
Reference 1:
- Upper bounds: 505 (93.52%)
Reference 4:
- Lower bounds: 7 (1.30%)
- Upper bounds: 25 (4.63%)
- Optimal solutions: 450 (83.33%)

ID;5
Author(s);Several authors
Reference;Data downloaded from the site www.mmlib.eu
Date;23.04.2018
Hardware / software;not specified
Stop criteria;none
Submission date;23.04.2018
Relevant Results
MM50_CV
MM50_SH
MM50_CS
MM50_ASGJBS
MM50_mmlib

Best Known Solutions
/sites/default/files/datasets/MM50_BKS.csv
Original Instances
MMLIB Instance Files
File Header
Title;MMLIB 50
Description;"Van Peteghem, V. and Vanhoucke, M., 2014, ""An experimental investigation of metaheuristics for the multi-mode resource-constrained project scheduling problem on new dataset instances"", European Journal of Operational Research, 235(1), 6272 (doi:10.1016/j.ejor.2013.10.012)."
Number;540
Format;mm
Date: Tue, 2018/11/13 - 12:00

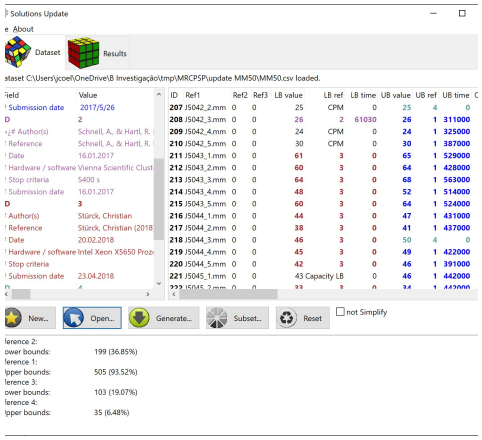
- Advantages:
  - Results can be applied on a dataset by order of publication data
  - A new results file with just the best known solutions of all results files, is automatically created
  - If a researcher want, can download not just the dataset, but also results of other authors
  - All the data updated in the site, can easy replicate by another researcher, after download all dataset and result files
- Checks done on result files:
  - Upper bounds – the solutions are validated against all restrictions.
  - Lower bounds – if a lower bound is incompatible with a upper bound, then the results file of the lower bound became invalid, and should removed from the dataset
  - Optimal solutions – this result can be obtained by an exact procedure, and is reported as both, an upper bound and lower bound, and follows the same checks
- Website: solutionsupdate.ugent.be part of projectmanagement.ugent.be/research/data
- Results confirmation: http://solutionsupdate.ugent.be/index.php/results-file/josecoelho/j30mmsh

5

5

Reporting new results 3/3

- Software client tool allow:
  - all web operations to be done locally;
  - to better see the dataset file;
  - check the solutions to exclude bugs from behing reported in the literature
  - calculate average deviation for each run;
  - create subsets (training set) on existing datasets (just the open instances, or other subset);
  - change names and/or format of instance files;
  - to generate files to export new data to PSPLIB and MMLIB.

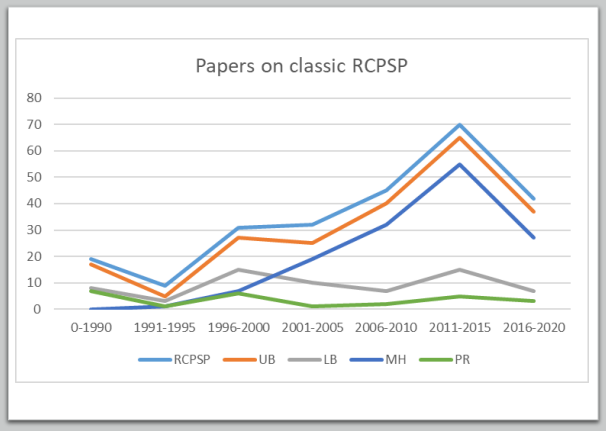


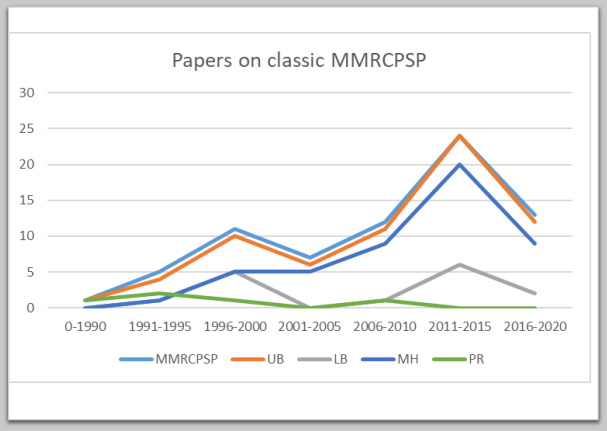
6

6

Update of tables of BKS on RCPSP and MMRCPSP 1/4

- Volume of work in the deterministic RCPSP and MMRCPSP
  - Papers and tags are in https://www.zotero.org/groups/2504748/mrcpsp/library



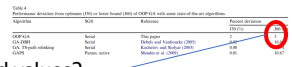


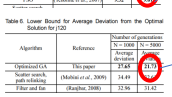
7

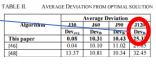
7

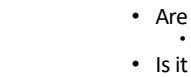
Update of tables of BKS on RCPSP and MMRCPSP 2/4

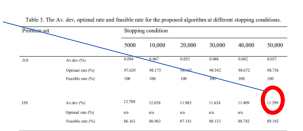
- What can avoid a place with all UB/LB?
  - Can average performance values in a published paper became out of range of valid values?
    - Invalid performance on J60 set (BKS is 10.37% for UB and 9.67% for LB):
      - Ali et al 2016, performance of 0,98 and get worst with more schedules, both below LB
      - Rostami et al 2014, performance of **9.94** - it is possible, but new BKS would need to be reported
      - Montoya-Torres et al 2010, performance of 5, below LB
      - Tchomte et al 2007, performance of 9.01, below LB
    - Invalid performance on J120 set (BKS is 29.01% for UB and 24.98% for LB):
      - Mulin 2018, performance of 50K of **30.25**, it became the best procedure reported, but not in Pellerin 2020 paper
      - Ali et al 2016, performance on 5K schedules of 19.6, below LB
      - Kadan and Kadan 2014, performance of **25.32** - it is possible, but new BKS would need to be reported
      - Rostami et al 2014, performance of 21.73, below LB
    - Invalid performance on J30 mm set (BKS is 12.28% for UB and 10.48% for LB)
      - Soliman et al 2014, performance of **11.30** – it is possible, but new BKS would need to be reported
  - Are wrong instance results appear in international journals?
    - Liess and Michelon 2007, reported invalid LB referenced in Horbach 2010
  - Is it hard to know what instances are closed or open?
    - Araujo et al. 2020 - Table 1 - invalid number of open instances in PSPLIB and MMLIB.
    - Chakrabortty et. al. 2020 – Table 5 – invalid number of closed instances in MMLIB

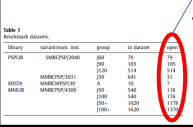


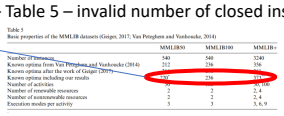










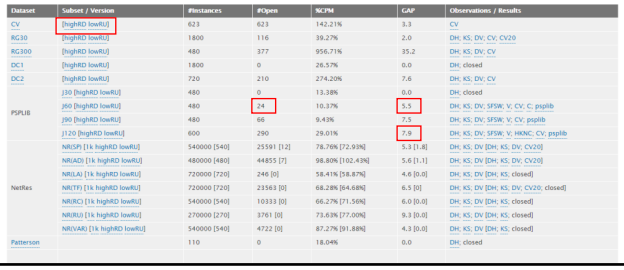


found that the results reported by Liess and Michelon (2008) contain some inconsistencies. The lower bounds indicated for instances j9026-8, j9030-5 and j9030-7 conflict with the upper bounds 82, 83 and 84 listed for them in PSPLIB. These upper bounds are also confirmed

8

Update of tables of BKS on RCPSP and MMRCPSP 3/4

- Table RCPSP:
  - Vanhoucke and Coelho (2018)
  - This presentation
  - http://solutionsupdate.ugent.be/rcpsp
    - highRD/lowRU see Vanhoucke and Coelho (2021)



Dataset		#Instances	#Open	%CPM	Gap
RG30		1800	195	39.33%	7.2
RG300		480	379	956.71%	38.0
DC1		1800	0	26.57%	0.0
DC2	npv25	180	0	124.12%	0.0
	npv50	180	47	225.18%	7.6
	npv75	180	94	331.08%	10.8
	npv100	180	107	416.46%	15.0
PSPLIB	J30	480	0	13.37%	0.0
	J60	480	53	10.37%	5.9
	J90	480	80	9.44%	7.5
	J120	600	308	29.11%	8.5
NetRes	NetRes(SP)	540,000	25,577	78.75%	5.3
	NetRes(AD)	480,000	44,829	98.79%	5.6
	NetRes(LA)	720,000	246	58.41%	4.6
	NetRes(TF)	720,000	23,544	68.27%	6.4
	NetRes(RC)	540,000	10,333	66.27%	6.0
	NetRes(RU)	270,000	3761	73.63%	9.3
	NetRes(VAR)	540,000	4722	87.27%	4.3

Dataset	Subset	#Instances	#Open	%CPM	GAP
CV		623	623	142.21%	3.3
RG30		1,800	116	39.27%	2.0
RG300		480	377	956.71%	35.2
DC1		1,800	0	26.57%	0.0
DC2		720	210	274.20%	7.6
PSPLIB	J30	480	0	13.38%	0.0
	J60	480	37	10.37%	6.3
	J90	480	66	9.43%	7.5
	J120	600	290	29.01%	8.0
NetRes	NR(SP) | 1k	540,000 | 540	25,591 | 12	78.8% | 72.9%	5.3 | 1.8
	NR(AD) | 1k	480,000 | 480	44,855 | 7	98.8% | 102.4%	5.6 | 1.1
	NR(LA) | 1k	720,000 | 720	246 | 0	58.4% | 58.9%	4.6 | 0.0
	NR(TF) | 1k	720,000 | 720	23,544 | 0	68.3% | 64.7%	6.4 | 0.0
	NR(RC) | 1k	540,000 | 540	10,333 | 0	66.3% | 71.6%	6.0 | 0.0
	NR(RU) | 1k	270,000 | 270	3,761 | 0	73.6% | 77.0%	9.3 | 0.0
	NR(VAR) | 1k	540,000 | 540	4,722 | 0	87.3% | 91.9%	4.3 | 0.0
Patterson		110	0	18.04%	0.0

9

Update of tables of BKS on RCPSP and MMRCPSP 4/4

Dataset		#Instances	#Open	%CPM	Gap
PSPLIB	J10	536	0	32.24%	0.0
	J12	547	0	27.57%	0.0
	J14	551	0	23.59%	0.0
	J16	550	0	18.67%	0.0
	J18	552	0	18.32%	0.0
	J20	554	0	17.00%	0.0
	J30	552	245	12.28%	6.5
Boctor	Boct50	120	120	22.74%	52.6
	Boct100	120	120	22.91%	103.6
MMLIB	MMLIB50	540	310	22.33%	9.3
	MMLIB100	540	275	21.56%	10.8
	MMLIB+	3240	2811	79.46%	37.2

- Tables MMRCPSP:
  - Vanhoucke and Coelho (2018)
  - This presentation
  - http://solutionsupdate.ugent.be/mrcpsp

Dataset	Subset	#Instances	#Open	%CPM	GAP
PSPLIB	J10 | J12 | J14 | J16 | J18 | J20	536 | 547 | 551 | 550 | 552 | 554	0 | 0 | 0 | 0 | 0 | 0	32% | 27% | 24% | 19% | 18% | 17%	0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0
Boctor	J30	552	245	12.28%	6.5
	Boct50	120	120	22.74%	52.6
	Boct100	120	120	22.91%	103.6
MMLIB	MMLIB50	540	95	22.29%	9.3
	MMLIB100	540	151	21.35%	10.8
	MMLIB+	3240	2439	78.77%	37.2

Dataset	Subset	#Instances	#Open	%CPM	Gap	Observations / Results
PSPLIB	J10 | J12 | J14 | J16 | J18 | J20	536 | 547 | 551 | 550 | 552 | 554	0 | 0 | 0 | 0 | 0 | 0	32.24% | 27.57% | 23.59% | 18.67% | 18.32% | 17.00%	0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0	psplib | psplib | psplib | psplib | psplib | psplib (all closed)
Boctor	J30	552	31	12.28%	8.7	CV; SH; ASGJBS
	Boct50	120	120	22.74%	52.6	CV
	Boct100	120	120	22.91%	103.6	CV
MMLIB	MM50	540	82	22.28%	3.5	CV; SH; CS; ASGJBS; mmlib
	MM100	540	123	21.35%	4.1	CV; SH; CS; ASGJBS; mmlib
	MMplus	3240	2043	78.64%	15.9	CV; SH; CS; ASGJBS; mmlib

10

10

An example of an experiment with NetRes

- NetRes:
  - High diversity in terms of project indicators;
  - Large number of instances, to allow select subsets needed (3,8 million instances).
- Percentage of hard instances in NetRes depending on each project indicator
  - Hard = not solved by DH92 procedure in 1 second
  - Phase transitions found in all indicators in the table
- The table could not be constructed based in other sets reported in Vanhoucke et. al. (2016) since not enough instances exist for all indicators and values.
- The client tool can be used to select the subset of instances needed and is easy to reference any subset.
- Suggestions:
  - Training set: use instances in steps of 10.000
  - Exact procedures: use instances in steps of 1.000
  - Meta-heuristics: use instances in steps of 100
  - Priority rules / fast lower bounds: use the full dataset

	SP	AD	LA	TF	OS	RC	RS
0-0,1	54%	-	4.9%	0.4%	59%	0%	7.4%
0,1-0,3	13%	1.5%	0.3%	0.5%	31%	9.1%	7.2%
0,3-0,5	0.2%	4.5%	0.1%	2.8%	1.3%	4.9%	1.3%
0,5-0,7	0%	13%	0.1%	8%	0.4%	3.8%	0.5%
0,7-0,9	0%	4.9%	0%	17%	0%	1.8%	0.1%
0,9-1	-	1.9%	-	24%	0%	1.4%	0%

11

11

Diversity of dataset CV 1/2

- CV set:
  - 623 small instances that are open, with 20 to 30 activities and 1 to 4 renewable resources;
  - Build based on existing instances, with a procedure that removes or changes the instance, to allow the instance became smaller and harder;
  - Diversity is taken in account, with a run to get hard instances in zones with few instances generated.
- The instances are not all too parallel, we can see that only LA and RS indicators have most instances concentrated on 0-0,2 range;
- The instances in CV set are open after 20 hours runs, using a diverse set of procedures. Total run time (generating and testing), corresponds to 45 years on a single processor

#Activities	#Resources	CNC	OS	SP	AD
20-21 #4	1 #1	0-1 #405	0-0,1 #85	0-0,1 #232	0-0,2 #8
22-23 #18	2 #39	1-2 #176	0,1-0,2 #416	0,1-0,2 #316	0,2-0,4 #79
24-25 #41	3 #85	2-3 #24	0,2-0,3 #99	0,2-0,3 #62	0,4-0,6 #243
26-27 #95	4 #498	3-4 #6	0,3-0,4 #13	0,3-0,4 #12	0,6-0,8 #229
28-30 #465		4-8 #12	0,4-0,6 #10	0,4-0,5 #1	0,8-1 #64
LA	**TF**	**RC**	**RF**	**RU**	**RS**
0-0,2 #592	0-0,2 #28	0,2-0,3 #57	0,5-0,6 #9	1-2 #34	0-0,1 #533
0,2-0,4 #10	0,2-0,4 #48	0,3-0,4 #258	0,6-0,7 #13	2-3 #84	0,1-0,2 #87
0,4-0,6 #11	0,4-0,6 #138	0,4-0,5 #285	0,7-0,8 #153	3-4 #505	0,2-0,3 #3
0,6-0,8 #7	0,6-0,8 #220	0,5-0,6 #10	0,8-0,9 #245		
0,8-1 #3	0,8-1 #189	0,6-0,8 #13	0,9-1 #203		

12

Diversity of dataset CV 2/2

- Uses of CV set:
  - Can incentivate new radically different procedures. We hope the small size and diversity of the set may help;
  - An in-depth analysis of the instances may reveal what makes these instances are very hard, when with a small change became easy to solve;
  - Since the instances are hard, and have small size, this set is ideal to test heavy lower bounds;
- The results on the time spend to obtain the best UB/LB appear to indicate that is harder to get the best LB, than the best UB, so updates on results are expected from lower bounds and exact procedures

13

Conclusion

- Proposed:
  - SolutionsUpdate to keep results on RCPSP and MMRCPSP updated at all times
  - A new large dataset, NetRes, with diversity in many project indicators and large number of instances available, to fit several research needs
  - A new CV set of small, hard and open instances, mainly to be used in lower bounds and exact procedures
- Website: solutionsupdate.ugent.be part of projectmanagement.ugent.be/research/data
- References:

  - Vanhoucke, M., & Coelho, J. (2018). A tool to test and validate algorithms for the resource-constrained project scheduling problem. *Computers & Industrial Engineering, 118,* 251–265. https://doi.org/10.1016/j.cie.2018.02.001
  - Coelho, J., & Vanhoucke, M. (2020). Going to the core of hard resource-constrained project scheduling instances. Computers & Operations Research, 121, 104976. https://doi.org/10.1016/j.cor.2020.104976
  - Vanhoucke, M., & Coelho, J. (2021). An analysis of network and resource indicators for resource-constrained project scheduling problem instances. Computers & Operations Research, 105260. https://doi.org/10.1016/j.cor.2021.105260
  - Vanhoucke, M., Coelho, J., & Batselier, J. (2016). An overview of project data for integrated project management and control. Journal of Modern Project Management, 3(3), 6–21.

14

14