2023-11

# Multi-Objective Metaheuristic Solution Approach for the Crop Plant Scheduling Problem

Lalić Maksim, Obrenović Nikola, Atac S., Bortolomiol S., Marko Oskar, Brdar Sanja, Crnojević Vladimir, Luković Ivan

# Multi-Objective Metaheuristic Solution Approach for the Crop Plant Scheduling Problem

M. Lalić[1], N. Obrenović[1], S. Ataç[2], S. Bortolomiol[3], M. Oskar[1], S. Brdar[1], V. Crnojević[1], I. Luković[4]

[1] Institute BioSense, University of Novi Sad, Serbia
{maksim.lalic, nikola.obrenovic, oskar.marko, sanja.brdar, crnojevic}@biosense.rs
[2] École Polytechnique Fédérale de Lausanne, Transport and Mobility Laboratory (TRANSP-OR), Switzerland, selin.atac@epfl.ch
[3] Optit Srl, Bologna, Italy, stefano.bortolomiol@optit.net
[4] Faculty of Organizational Sciences, University of Belgrade, Serbia, ivan.lukovic@fon.bg.ac.rs

## 1 Introduction

Most of the farm field operations are scheduled quite imprecisely or, even worse, determined ad hoc. Following intuitionally and empirically handcrafted schedules can lead to suboptimal harvest time. As a consequence, it can affect harvest effectiveness by reducing the total amount of yield and increasing the amount of unnecessary waste, as well as its efficiency by imposing a longer harvest horizon than required or by making harvest intensity and labor force engagement unevenly distributed over the horizon.

Results from our previous study [1] suggest that by proper planting and harvest scheduling, there is room for significant improvement in waste reduction and consistent harvest intensity. Our study was inspired by the Syngenta Crop Challenge in Analytics 2021 (the challenge) [2]. The aim of the challenge was to determine desired planting and harvest schedule for the farm field whose crop is divided into *populations*. Each population represents a management unit that will be treated individually and uniformly in any aspect. For each population, the allowed planting time window and expected yield are known. Afterward, based on the weather forecast, it is possible to derive time-window during which the population should be harvested. Finally, we have defined a limit on the amount of weekly harvested yield that can be adequately processed and stored while any surplus beyond that would be wasted.

Some other papers discuss slightly different solution approaches for a problem proposed at the Challenge [3–5]. However, all proposed decision frameworks have certain specificities. They all incorporate integer-linear problems (ILP) solved by exact off-the-shelf solvers and the multi-objective nature of the problem has been treated either by aggregation of multiple objectives as a weighted sum or by iterative or sequential exploitation of models as part of a broader decision framework.

Here, we simplify the optimization process and approach it with a single run of a multi-objective metaheuristic method. This enables us to produce and analyze Pareto fronts of solutions with various effectiveness and efficiency and to deal with instances unmanageable for exact solvers. The optimization problem has been approached with two, in some sense complementary, metaheuristics. We examined adaptive large neighborhood search (ALNS) [6] as a model-driven single-solution-based metaheuristic and non-dominated sorting genetic algorithm II (NSGA-II) [7] as a population-based data-driven metaheuristic. We provide our preliminary results, which show that ALNS consistently and significantly outperforms NGSA-II.

## 2 Methodology

Given a set of populations $P$ and a set of weeks comprising harvesting horizon $W$, our problem formulation is as follows:

$$F_1 = \sum_{w \in W} |q_w - C| \tag{1}$$

$$F_2 = |\{w \in W : q_w > 0\}| \tag{2}$$

$$q_w = \sum_{p \in P} hq_p \cdot 1_w([t_p]), \qquad \forall w \in W \tag{3}$$

$$e_p \leq t_p \leq l_p, \qquad \forall p \in P \tag{4}$$

The first objective function $F1$ expresses the effectiveness of a schedule and it is calculated as the sum of deviations between weekly harvested quantities $q_w$ and the stated storage capacity $C$. The second objective function $F2$ expresses the schedule efficiency and it is equal to the number of working weeks, i.e. the number of weeks with a positive harvested quantity.

Constraints (3) are used to determine weekly harvested quantities $q_w$, for each week $w \in W$, based on selected harvesting time $t_p$ for a population $p \in P$ and its expected yield $hq_p$. Since $t_p$ takes real values, we consider that a population $p$ would be harvested in a week $w$ if an indicator function $1_w([t_p])$ yields 1, i.e. if a rounded value of $t_p$ equals to $w$. Constraints (4) ensure that each population $p \in P$ is harvested within its feasible time window $[e_p, l_p]$.

We solved the proposed problem using two well-known metaheuristics, ALNS and NSGA-II. For the ALNS application, we designed five domain-specific neighborhood operators, where each of them was tailored to favor a particular desirable characteristic of the schedule. Operator one is a rebalancing operator designed to reschedule populations from highly stressed weeks to least stressed weeks. Operator two is a stability operator designed to equalize the amount of harvested quantities between consecutive weeks. Operator three is an emptying operator which finds two consecutive weeks with the lowest total harvested quantity and empties the less stressed one into another one. Operator four is also an emptying operator. This operator will empty a week which is preceded by an already emptied one. Operator five is a capacity operator rescheduling populations for weeks that are loaded over storage capacity $C$.

As for the second metaheuristic, we used real-coded constrained NSGA-II without any modifications. We experimented with two different ways of initial population generation. By our first approach, called *random*, the initial population is created by uniformly sampling value for each gene $t_p$ inside its allowed time window $[e_p, l_p]$. By our second approach, called *smart*, we select our initial population as a subset of 100 solutions out of 500 initially produced distinct solutions generated by the previous run of the ALNS for the same case.

## 3 Results

We synthesized 20 different case studies, which may have $n \in \{500, 1000, 2000, 3000, 4000\}$ populations and planning horizon $l \in \{26, 52, 78, 104\}$ weeks long. For each case study, we run each algorithm 25 times. For each case, we combined all solutions nondominated individually per optimization run and derive the overall Pareto front of all solutions on the case level.

Figures 1 and 2 show combined Pareto fronts for cases with the highest and the lowest ratio between the number of populations and the duration of the planning horizon. Solutions produced by ALNS are depicted as ▲, or ■ if they are overall nondominated. Individually nondominated solutions produced by NSGA-II are depicted as x for *random* initialization or ★ for *smart* initialization. Bounding boxes of overall nondominated solutions are presented as red rectangles. Figures emphasize a strong dominance of ALNS over the NSGA-II approach. All overall nondominated solutions are produced by ALNS. NSGA-II has never produced any solution with a relatively low $F2$ objective value and a significant portion of individually optimal solutions produced by NSGA-II turns out to be out of the bounding box surrounding overall nondominated solutions. On the other side, individually nondominated solutions produced by ALNS are close to overall nondominated solutions.

The presented results are consistent with all other our results. For each case, all overall nondominated solutions have been produced solely by ALNS. Pareto fronts generated by ALNS are firmly consistent from run to run and they converge relatively fast, in about 10000 to 15000 iterations. On the other side, all individually nondominated solutions produced by NSGA-II are from the first generation, which indicates that populations don't improve themselves by evolution at all. That also explains why *smart* initialization strategy provides better results than *random* strategy since all solutions from its first generation are actually previously produced by the ALNS.

ALNS dominance over the NSGA-II is particularly obvious regarding the $F2$ objective function, where NSGA-II performed poorly all the time. That is in accordance with our most recent experimental results, which are beyond the scope of this paper. Results indicate that offsprings' $F2$ objective values consistently degrade in comparison to parents' until they finally reach a certain poor quality standard where their values retain.

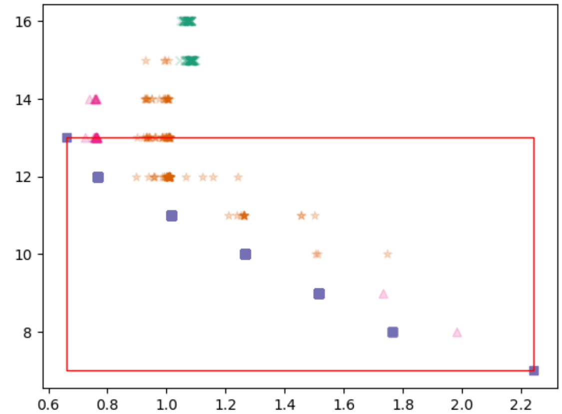


Fig. 1: Pareto front for $n = 4000$ and $l = 26$

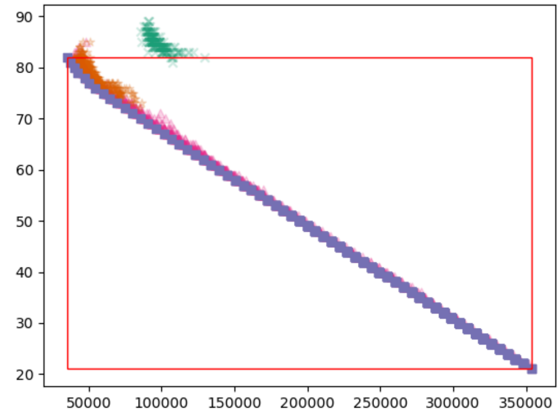


Fig. 2: Pareto front for $n = 500$ and $l = 104$

## 4 Conclusion

In this paper, we outlined a bi-objective formulation of the Crop Plant Scheduling Problem proposed in our previous work [1]. We adapted two metaheuristics to solve the problem. Our experimental results indicate a very strong dominance of ALNS over NSGA-II. Since those results are quite indicative, our future work shall be focused on gaining deeper insights into the reasons for such dominance. More precisely, we shall analyze the structure of the $F2$ objective function in relation to the operators integrated into the ALNS and NSGA-II and derive assumptions on what makes ALNS capable and NSGA-II incapable to provide good Pareto fronts for such a simple objective function.

## 5 Acknowledgements

This work was supported by the Provincial Secretariat for Higher Education and Scientific Research, Autonomous Province of Vojvodina (No. 142-451-2261/2022-01/01).

## References

1. N. Obrenović, S. Ataç, S. Bortolomiol, S. Brdar, M. Oskar, V. Crnojević, "The Crop Plant Scheduling Problem", *International Conference on Optimization and Decision Science, ODS 2022*, Florence, Italy, 2022.
2. Syngenta Crop Challenge in Analytics 2021, url: https://www.ideaconnection.com/syngenta-crop-challenge/challenge.php, [Accessed: April, 2023.]
3. S. S. Sajid, G. Hu, "Optimizing Crop Planting Schedule Considering Planting Window and Storage capacity", *Frontiers in Plant Science*, 13:762446, 2022, doi: https://doi.org/10.3389/fpls.2022.762446
4. Z. Khalilzadeh, L. Wang, "An MILP Model for Corn Planting and Harvest Scheduling Considering Storage Capacity and Growing Degree Units", *bioRxiv*, 2021, doi: https://doi.org/10.1101/2021.02.06.430062
5. J. Ansarifar, F. Akhavizadegan, L. Wang, "Scheduling Planting time Through Developing an Optimization Model and Analysis of Time Series Growing Degree Units", *arXiv*, 2022, doi: https://doi.org/10.48550/ARXIV.2207.00745
6. S. Ropke, D. Pisinger, "An adaptive large neighborhood search heuristic for the pickup and delivery problem with time windows", *Transportation Science*, vol. 40, issue. 4, pp. 455-472, 2006., doi: https://doi.org/10.1287/trsc.1050.0135
7. K. Deb, A. Pratap, S. Agarwal, T. Meyarivan "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II", *IEEE Transactions on Evolutionary Computation*, pp. 182-197, vol 6, no. 2, 2002, doi: https://doi.org/10.1109/4235.996017