Operational Research

Robust Planting Plan

Group 7

Marolicaru Elena Roxana - s287503
Matarazzo Emanuele Kaled - s280143
Spoturno Felipe - s287530
Terzolo Giulia - s281990

September 2022

Contents

Abstract

In the phases of production and sale of vegetables, the scheduling of planting and harvesting is crucial to avoid crop waste and maximize profits. In this project, the operational research approach was used to ensure the optimal planning of the planting and harvesting of Brussels sprouts, taking into account many crop-related variables, including species, size, diseases, and market-related variables. Moreover, the planning of planting and harvesting of Brussels sprouts must respect a considerable number of constraints.

The problem was solved using two different approaches: firstly, the exact problem was solved using Gurobi, thanks to which the optimal solution was obtained. Secondly, a heuristic method based on the ALNS (Adaptive Large Neighbourhood Search) algorithm was implemented. Finally, the performance of the two methods was compared, paying particular attention to the computational time, which is crucial when dealing with complex problems with many variables.

Keywords: Optimization, Heuristics, ALNS

1 Introduction

The problem of planning the planting and harvesting of vegetables has begun to be of increasing interest with the increase in demand for these items on a global scale. Consequently, there is a clear need to optimize the planning of the planting and the harvesting phases of fresh items in order to satisfy the demand on the market, maximizing profits and therefore minimizing waste. This study will address the problem of optimizing the plantation and harvesting of Brussels sprouts with the aim of maximizing the grower's profit. Various parameters and constraints will be considered in the analysis of the problem. Two different approaches will be used and compared: the first will analyze the problem by searching for the exact optimal solution, using Gurobi optimizer, a powerful mathematical solver which solves the problem by using Linear Programming (LP). The second one will be a heuristic approach, based on ALNS (Adaptive Large Neighborhood Search) algorithm which consists in recurrent distructions and repairings of the solution.

2 Literature review

The optimization of planting and harvesting phases of the supply chain have been analysed in several studies with different approaches. Different studies analize the problem at increasing levels of complexity: in the case of Fresh fruit supply chains, the models that deal with them range from planting, to harvesting, to distribution of final products (Masini et al. (2008)[13]). One of the first studies related to the optimization of the plantation was developed by Willis et al. (1976)[18] who used dynamic linear programming to optimize the combination of apple varieties to be planted. Many subsequent studies have focused on the plantation step, considering harvesting as a given. For this phase, attention is for the choice of crops, the choice of the machinery to be used (Annetts and Ambiudsley (2002)[1]), optimization of the environmental benefits (Biswas and Pal (2005)[10], Huh and Lall (2013)[9]). Regarding harvesting modeling, studies address the problem with various objectives (minimization of the cost of harvesting, maximization of profit, minimization of waste). These studies mainly use MILP (Mixed-Integer Linear Programming): Herrera-Cáceres et al. (2017)[8] implemented plans that are flexible for different climatological risk levels; Ferrer et al. (2008)[6] used a MILP model for scheduling the harvesting of wine grapes: the goal was to minimize harvesting costs and the loss of quality of wine grapes due to a wrong harvesting schedule; González-Araya et al. (2015)[7] proposed a MILP model for apple orchard harvest planning. As before, the goal was to minimize harvesting costs, but the objective function also contained a penalty cost for fruit that did not meet the export requirements.

In several cases, the studies carried out were then applied in real companies: a MILP for olive harvest planning found in practice an increase in oil extraction of 4% (Herrera-Cáceres et al. (2017)[8]). Other studies focus on the search for a heuristic solution, which can reduce the time and complexity of computation. Thuankaewsing et al. (2015)[17] implemented a heuristic algorithm to identify the amount of sugarcane harvest with the best schedule that could guarantee equal benefits to all growers. Blanco et al. (2010)[3], Carpente et al. (2010)[4] and Stray et al. (2012)[16] show that models using mathematical programming can be used for small problems, while when the problem becomes complex using heuristic solutions saves time with a good approximation of optimization. Amin Khajepour et al. (2020)[11] used the Adaptive Large Neighborhood Search (ALNS) algorithm for a large instance which consists in removing and inserting heuristics to reach the optimal solution: this approach should have as result a more efficient computation of the solution with respect to traditional heuristics.

3 Instance generation

The problem is composed by a two-stage stochastic model, used to determine optimal planting plans for a vegetable crop. The first stage of the model relates to finding a planting plan which is common to all scenarios and the second stage is concerned with deriving a harvesting schedule for each scenario.

In order to run a simulation of the model, an instance of the problem must be introduced. A crop is defined as a particular variety of Brussels sprouts planted in a given week at a given spacing and the problem is to decide how much of each of a set of possible crops should be planted, where the exact yield profile is not known.

United Kingdom's National Institute of Agricultural Botany [14] provides a list of Brussel sprouts varieties and some experimental parameters, based on harvest trials. This paper will base on it instance generation about varieties growing rate, spacing, cost of land and cost of harvesting.

3.1 Demand

Customers specify a demand as a quantity in tonnes (t) for each harvesting week, and the size band they want.

Harvesting week is considered 0 from the first day of September, increasing every 7 days.

The client chooses a size band based on the market distribution of the vegetable: smaller sizes are perfect to be sold frozen, bigger sizes instead are used for the fresh market. Instance parameters are shown in Tab.1.

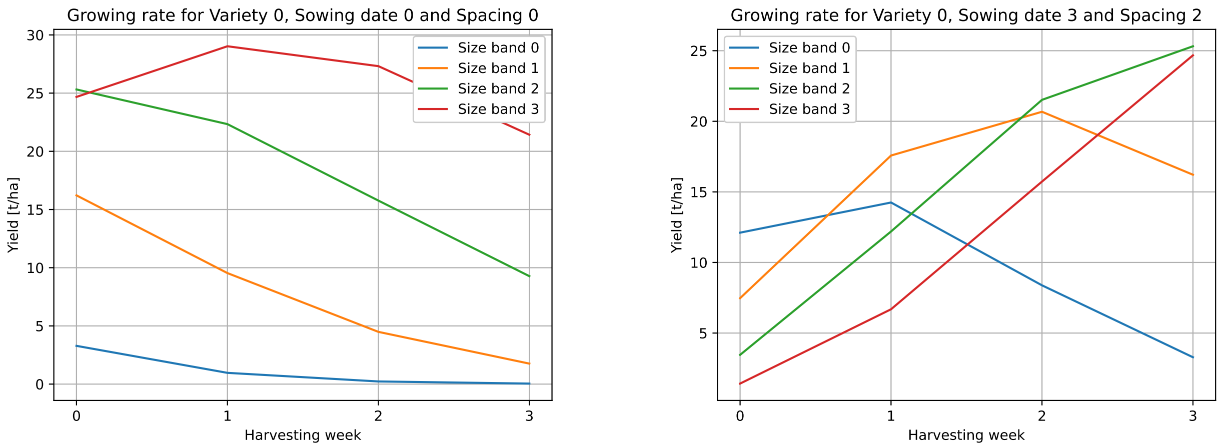
Table 1: Customers demand.

Variable	Unit	Type	Values	Mean	Variance
Quantity	t	continuous	-	100	0.35
Harvesting week	week number	integer	0,1,2,3	-	-
Size band	-	integer	0,1,2,3	-	-

3.2 Crop

A crop is defined as a particular variety of Brussels sprouts planted in a given week at a given spacing. Each crop has a different growing rate, called Yield (t/ha). Fig.1 shows an example for a vegetable with variety 0. If it is sowed in the first possible week (Sub-fig.1a), in the harvesting week 0 the bigger size bands are ripe, while the number of smaller sprouts is lower.

Planting the crop 3 weeks after and with a different spacing, leads to different yield profiles (Sub-fig.1b): a bigger number of small sprouts is present at the first harvesting week, while the bigger ones ripen after some weeks.



(a) (b)

Figure 1: *Yield profile for the same variety but with different sowing dates and spacings.*

The cost of harvesting depends on the crop, the harvesting week and the scenario. It is assumed to be proportional to the quantity.

Susceptibility to diseases is the lost proportion of crop harvested in a certain week due to a disease.

An overview of the instance variable related to the crops is plotted in Tab.2.

Table 2: Crop related variables.

Variable		Unit	Type	Values	Mean	Variance
Crop	Variety	-	integer	[0,8]	-	-
	Sowing week	week number	integer	0,1,2,3	-	-
	Spacing	-	integer	0,1,2,3	-	-
Susceptibility to diseases		-	continuous	[0,0.3]	-	-
Cost of harvesting		£/ha	continuous	-	8700	0.25

3.3 Costs and income

Land cost includes seeds and plants, nutrients, land preparation, planting and machinery [12]. Extra land can be rented to add production capacity. If extra land is bought, land preparation costs are summed to it. Unused land can be rented to others. Costs and income are shown in Tab.3.

Table 3: Costs and income.

Variable	Unit	Type	Values	Mean	Variance
Cost of land	$£/ha$	continuous	-	15000	0.1
Cost of extra land	$£/ha$	continuous	[2000-3500]	-	-
Value of unused land	$£/ha$	continuous	[2000-3500]	-	-
Profit from customers	$£/t$	continuous	-	2000	0.1
Profit from open market	$£/t$	continuous	-	2000	0.5

3.4 Choice of number of scenarios

The analysed problem of planning of the planting and harvesting of Brussels sprouts is a two-stage problem (Fig.2). In particular:

- The first stage consists of the planting plan which is constant for all scenarios of the second stage.
- The second stage consists of the harvesting plan, which is different for each scenario. The second stage constitutes the stochastic part of the problem.

The number of scenarios used for analysing the problem is 31, a choice derived from the fact that the study from which the model was taken (Darby-Dowman et al. [5]) used 31 years of weather data to create 31 yield scenarios. The in-sample stability analysis (7.1) confirmed this to be a reasonable choice.

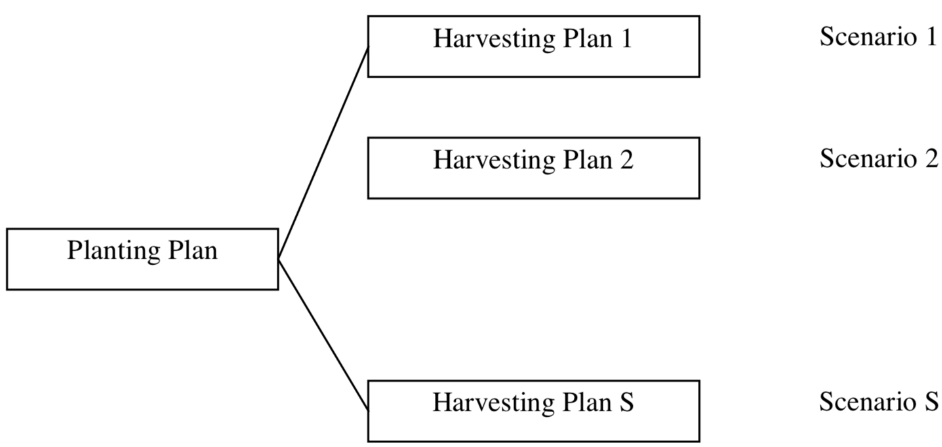


Figure 2: *Stochastic Programming Problem Structure.*

4 Mathematical model

In this section, the mathematical model is presented.

4.1 Sets

All the used sets, that control the size of the model through their cardinality, are displayed in Tab. 4.

Table 4: Sets table.

Notation	Meaning
I	crop
V	variety
J	week
K	size band
M	customer
Q	disease
S	scenario

K_m contains the indices of the size band specified by customer m.

4.2 Data

In Table 5, all the known data referring to the model are reported together with their unit of measurement, notation and meaning.

Table 5: Data table.

Notation	Meaning	Unit
c'	cost of land	$£/ha$
c_{sij}	cost of harvesting crop i in week j under scenario s	$£/ha$
d_{mj}	demand of customer m in week j	t
f_{mj}	profit from satisfying the demand of customer m in week j	$£/ha$

s_{sj}	profit from selling surplus-to-demand sprouts on the open market in week j under scenario s	$£/ha$
y_{sijk}	yield of crop i in week j in size band k under scenario s	t/ha
a	area of grower's land	ha
c^-	cost of extra land required by grower	$£/ha$
c^+	value of land unused by grower	$£/ha$
p_{smj}	penalty for failure to satisfy demand of customer m under scenario s	$£/t$
r_{iq}	1 if crop i is susceptible to disease q, 0 otherwise	$1, 0$
u_q	upper limit on the portion of crop harvested each week that is susceptible to disease q	t
$prob_s$	probability of scenario s	/

4.3 Variables

In Tab. 7, the control variables are displayed. They are needed to keep track of the choices that are performed during the simulation. In particular, the variable $Trash_i$ was introduced to remove the amount of crops discarded due to diseases from the final balance.

Table 7: Variables table.

Notation	Meaning	Unit
F_{sjmk}	weight of sprouts sold in week j to customer m in size band k under scenario s	t
H_{sij}	area of crop i harvested in week j under scenario s	ha
S_{sjk}	weight of surplus-to-demand sprouts of size k sold on the open market in week j under scenario s	t

L^-	area of extra land required by grower	ha
L^+	area of land unused by grower	ha
P_{smj}	shortage in demand of customer m in week j under scenario s	t
A_i	total area of crop i planted	ha
$Trash_i$	Quantity of crop i discarded due to disease	t

4.4 Objective function

The objective function, reported in equation 1 and 2, has to be maximized. The term w, being part of it, represents the risk aversion coefficient and it is user-specified.

$$O.F. = (weighted) * ExpecteProfit - RiskTerm = \tag{1}$$

$$(1-w) * E(Profit_s) - w * E(|Profit_s - E(Profit_s)|)$$

$$E(Profit_s) = \sum_s prob_s * \{\sum_{j,k} s_{sj} * S_{sjk} + \sum_{m,j,k} f_{mj} * F_{sjmk} - \sum_{i,j} c_{sij} * H_{sij} \tag{2}$$
$$+c^+ * L^+ - c^- * L^- - c' * \sum_i A_i - \sum_{m,j} p_{smj} * P_{smj}\}$$

4.4.1 Linearization of the objective function

In order to lienarize the non-linear objective function, a variable z was introduced:

$$z_s - (Profit_s - E(Profit_s)) \geqslant 0 \tag{3}$$

$$z_s + (Profit_s - E(Profit_s)) \geqslant 0 \tag{4}$$

4.5 Constraints

The objective function previously explained in subsec. 4.4 is subject to some constraints that are needed to express the effects of uncertainty on the model.

In the next part of the document they are described in detail.

4.5.1 Marketing Constraint

The marketing constraint, equation n°. 5, expresses the fact that the weekly harvest is sold either to customers or on the open market.

$$\sum_i y_{sijk} * H_{sij} - S_{sjk} - \sum_m F_{sjmk} = 0 \tag{5}$$

$$\forall s \; \forall j \; \forall k$$

4.5.2 Marketing Constraint 2

Marketing constraint 2, equation n°6 was added to the model in order to control the sale in relation to the customer's willingness: if the customer is not interested in purchasing, F_{sjmk} is set to 0. This was introduced because the exact solver sold most of the crop to customers who were not interested, in order to avoid the sell on open market constraint.

$$F_{sjmk} = 0 \tag{6}$$

$$\forall s \; \forall j \; \forall m \qquad k \notin K_m$$

4.5.3 Demand Constraint

The demand constraint, equation n°. 7, states that the quantity of sprouts sold is given by the subtraction of the demand the customer made for them and shortage. The original sign of P_{smj} was changed from + to − as it is the shortage of customer m 's demand in week$j$ for a given scenario s.

$$\sum_k F_{sjmk} = d_{mj} - P_{smj} \tag{7}$$

$$\forall s \; \forall j \; \forall m \qquad k \in K_m$$

4.5.4 Sell on Open Market

The sell on open market constraint, equation n°. 8, imposes a limit on the amount of sprouts that can be sold on free-trade option.

$$\sum_k S_{sjk} \leq 0.25 * \sum_m d_{mj} \tag{8}$$

$$\forall s \; \forall j$$

4.5.5 Land Use Constraints

Two land use constraints have to be set.

The first one, eq. 9, declares that the total area of all crops is equal to the amount of required land, given by the initial quantity plus the one that is needed to fulfill the demand necessity and minus the unused terrain.

$$\sum_i A_i = a + L^- - L^+ \tag{9}$$

The second one, instead, formula n°. 10, makes sure that the area of each crop planted matches the harvested area. Compared to the original equation, the variable $Trash_i$ was also incorporated into the area calculation for crop i since it represents the part of the harvest that cannot be sold due to diseases.

$$A_i = \sum_j H_{sij} + Trash_i \tag{10}$$

$$\forall s \; \forall i$$

4.5.6 Disease Constraint

The disease constraint, eq. 11, sets limits on the amount of demand of crops that can be susceptible to each kind of plant illness. quantità di crop da buttare per disease.

$$Trash_i = \sum_q (r_{iq} * u_q * A_i) \tag{11}$$

$$\forall i \; \forall q$$

This formula imposes that the planting plans produced with this model are less risky than they would otherwise be.

4.5.7 Individual Variety Limit

The individual variety boundary, formula n°. 11, imposes limits on the amount of area that can be assigned to each variety of brussel sprouts.

$$\sum_{i \in I_v} A_i \leq 0.4 * \sum_i A_i \tag{12}$$

$$\forall v$$

In this equation, $I_v$contains the indices of the crops associated with each variety v.

This equation imposes that the planting plans produced with this model are less risky than they would otherwise be.

4.5.8 Individual Crop Limit

The individual crop boundary, equation n°. 12, sets limits on the amount of area assigned to each crop.

$$A_i \leq 0.2 * \sum_i A_i \tag{13}$$

$$\forall i$$

This formula imposes that the planting plans produced with this model are less risky than they would otherwise be.

5 Variation of results with respect to the risk adversion coefficient w

The graph in Fig.3 shows how the profit and standard deviation of profit vary as w varies. The graph was obtained by averaging over 25 instances for each value of w and normalising with respect to profit for $w = 0$. In particular, Fig.3 shows how the standard deviation decreases as w increases, since risk is introduced into the objective function. As result, the profit is lower and the variance of the profit decreases as well.

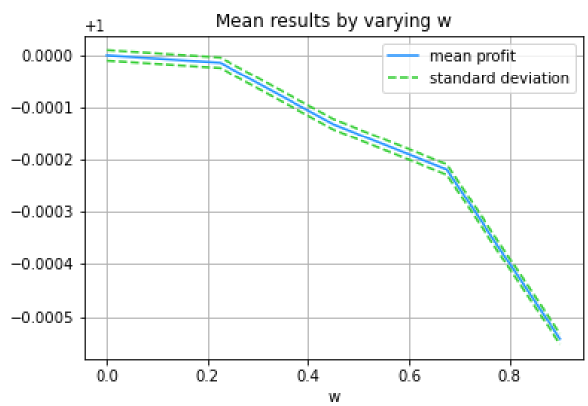


Figure 3: *Profit mean and standard deviation in function of w.*

6 Heuristic method

Heuristics is a problem-solving method that aims at building a practical resolution of the problem that would give a fairly good result and close enough to the exact solution. The advantage over the exact solution is mainly a shorter computational time, which turns out to be a crucial point especially when analyzing a very complex problem with a large number of iterations. In our case, heuristics was implemented using the ALNS algorithm in the first stage of the problem. The ALNS algorithm consists of repeated destructions and reconstructions of the solution (Roozbeh et al.[15]).

In the analyzed problem the constructor builds a solution in a random fashion. The two destroyers act in the following way:

- The former removes 40% of the crops that were planted using more hectares. The reason for this is that it has been assumed that since the customers' demands are similar, planting more hectares for a similar crop means that the yield is low.

- The latter is a random destroyer, which randomly chooses 40% of the crops to discard. This choice allows for a better exploration of the solutions' space.

The initial stage of the algorithm consists of random planting with the aim of satisfying all customers and respecting all the constraints of the mathematical model. It then considers a scenario at the second stage to calculate the exact solution, which then has the objective function as output. At this point, in the first stage, the destruction of part of the solution and its reconstruction takes place . This is iterated for N number of times. Finally, the objective function that turns out to be the best is used to identify the best A_i vector to use.

Two variables were added to the second stage exact solver in order to cope with model's constraints:

- OH_i: Over-harvesting of crop i. Since, according to the model, part of the production cannot be harvested due to disease, this variable stores this quantity to eliminate it from the final balance.

- O_{jk}: Over-production of size band k in week j. This variable was introduced to comply with the Sell on Open Market constraint, according to which no more than 25% of the harvest can be sold on the open market. If the first stage has an output that is too high, which would lead to violating this constraint, that quantity is stored in this variable.

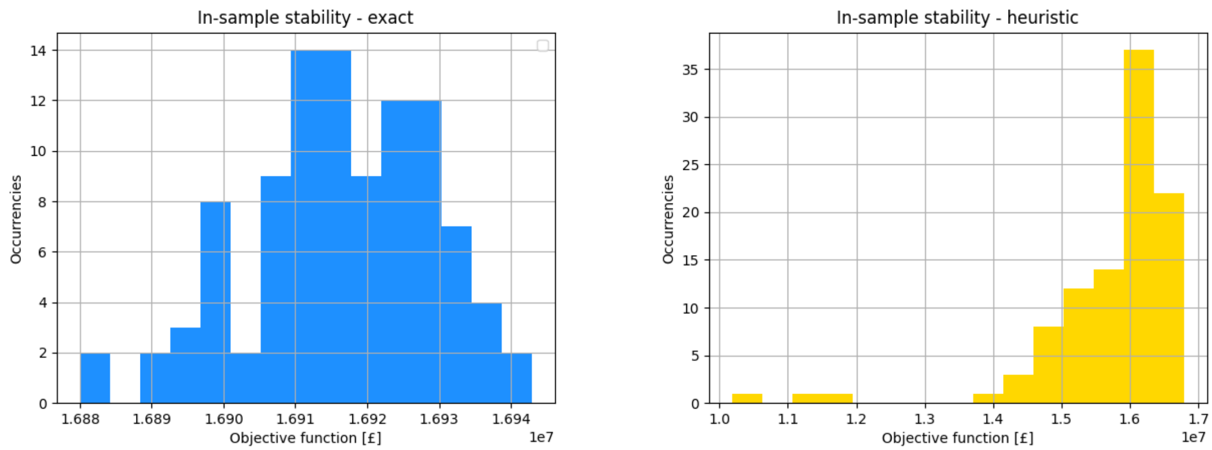
7 Results

In this section, stability analysis and computational time comparison between exact problem solution and heuristic solution will be presneted.

7.1 In-sample stability

For the stability analysis, in-sample stability was studied first. In-sample stability occurs when the optimal value of the objective function remains approximately similar for any scenario tree.

To test the in-sample stability, 100 scenario trees were generated. The problem was solved and an objective function was obtained for each scenario tree. The histogram with the results for the exact problem and for heuristic is shown in Fig.4.



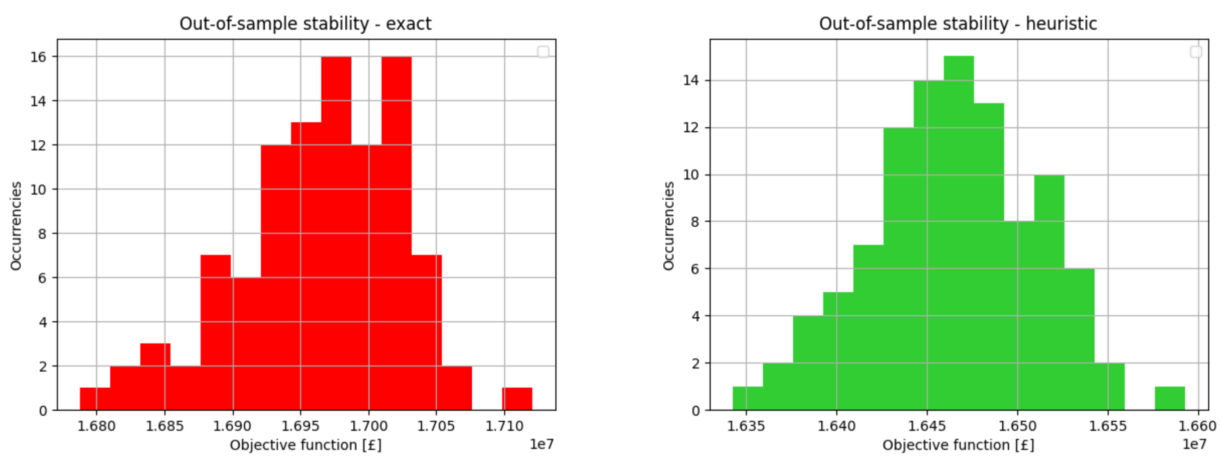
(a) (b)

Figure 4: *In-sample stability histogram.*

7.2 Out-of-sample stability

Subsequently, out-of-sample stability was studied, which occurs when the real performance of the solution does not depend on the scenario tree.

For out-of-sample stability, 10 scenarios were considered and the complete problem was solved for these. Then, starting with the first stage of the exact solution, 1000 new scenarios were generated and the second stage was evaluated for them. The histogram related to out-of-sample stability is shown in Fig.5.



(a) (b)

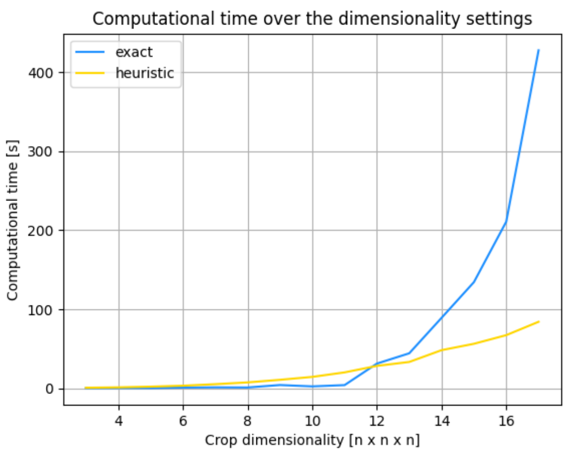
Figure 5: *Out-of--sample stability histogram.*

7.3 Computational time

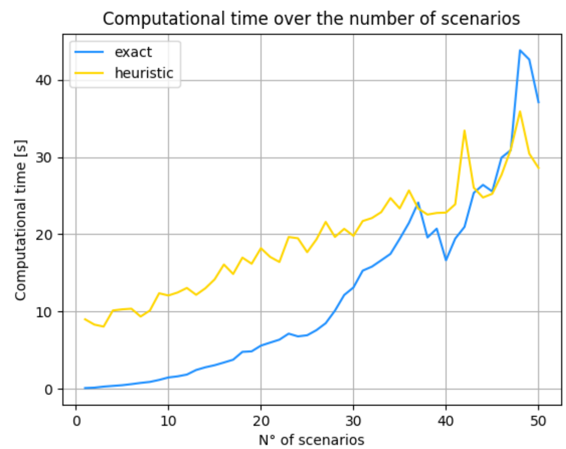
In Fig.6a it is shown the variation of computational time with increasing dimensionality $n \times n \times n$ ($variaties \times spacing \times sowingdates$). Although for few iterations the computational time of the heuristic is longer than the computational time of the exact solver, for a large number of iterations the heuristic performs significantly better than the exact solver, thus confirming that a heuristic is preferable to the exact solver for very complex problems.

In Fig.6b the computational time as the number of scenarios varies is presented. Up to about 40 scenarios, the exact solver performs better than the heuristic, but for a higher number of scenarios, the heuristic has better computational performance.

Finally, the advantage of heuristics over the exact solver is more immediately visible in the variation of the dimensionality of the problem rather than the number of scenarios.



(a)



(b)

Figure 6: *Computational time.*

7.4 Performances comparison

In this subsection, the exact solver and the heuristic are compared in terms of dimensionality ($variaties \times spacing \times sowingdates$) and number of scenarios for different experiments. For the experiments 3 diseases, 5 size bands and 5 harvesting weeks were considered.

The results are presented in Table9.

Table 9: Performance comparison.

Exp n°	Method	Dimensionality $[n \times n \times n]$	Scenarios	Time [s]	OF value [£]	$\triangle$%
1	Exact solution	$3 \times 3 \times 3$	5	0.07	13 252 574	5.75
	Heuristic solution			5.84	12 490 910	
2	Exact solution	$10 \times 10 \times 10$	5	6.86	18 139 068	7.68
	Heuristic solution			14.72	16 745 809	
3	Exact solution	$4 \times 4 \times 4$	31	3.89	17 336 258	2.41
	Heuristic solution			25.86	16 917 778	
4	Exact solution	$16 \times 16 \times 16$	50	3514.52	19 074 688	6.50
	Heuristic solution			1674.75	17 835 327	
5	Exact solution	$100 \times 100 \times 100$	50	-	-	-
	Heuristic solution				17 637 881	

The heuristic algorithm, with the available hardware, is able to handle very big dimensionalities: in case of $100 \times 100 \times 100$, Gurobi ends before reaching the exact value of the objective function.

8 Conclusions

This paper was based on the work of Darby-Dowman et al. [5] to create a robust planting and harvesting plan considering the risk introduced by weather variability. Modifications were made to the model under analysis so that it could be written as a linear optimisation problem, as explained in Sect..4. After having solved the problem in exact form using Gurobi, a heuristic algorithm was designed to reduce the computational time in the case of high dimensionality with an acceptable difference in the solution.

The first stage was decoupled from the second stage. The ALNS algorithm at the first stage, and Gurobi at the second stage, solving the exact problem for one scenario at a time. This allowed to reduce the dimensionality of the problem and consequently also the computational time.

References

[1] Annetts, J. E., Audsley, E., (2002). Multiple objective linear programming for environmental farm planning.

[2] Biswas, A., Pal, B. B., (2005). Application of fuzzy goal programming technique to land use planning in agricultural system.

[3] Blanco, V., Carpente., L., Hinojosa., Y., Puerto, J., (2010). Planning for agricultural forage harvesters and trucks.

[4] Carpente, L., Casas-Mendez, B., Jacome, C., Puerto, J., (2010). A model and two heuristic approaches for a forage harvester planning problem.

[5] Darby-Dowman, K., Barker, S., Audsley, E., Parsons, D., (2017). A two-stage stochastic programming with recourse model for determining robust planting plans in horticulture.

[6] Ferrer, J. C., MacCawley, A., Maturana, S., Toloza, S., Vera, J., (2008) An optimization approach for scheduling wine grape harvest operations.

[7] González-Araya, M.C., Soto-Silva, W.E., Acosta, E., (2015) Harvest planning in apple orchards using an optimization model.

[8] Herrera-Cáceres, C. , Pérez-Galarce, F. , Álvarez-Miranda, E. , Candia-Véjar, A. (2017). Optimization of the harvest planning in the olive oil production: A case study in Chile.

[9] Huh, W.T., Lall, U., (2013). Optimal crop choice, irrigation allocation, and the impact of contract farming.

[10] Kaut, M., Wallace S. W., (2003). Evaluation of scenario-generation methods for stochastic programming.

[11] Khajepour., A., Sheikhmohammady, M., Nikbakhsh, E. (2020). Field path planning using capacitated arc routing problem.

[12] KPU (2022), Southwest British Columbia Small-Scale Farm Enterprise Budget: Brussel Sprout. Kwantlen Polytechnic University. Online resource: https://www.kpu.ca/sites/default/files/ISFS/Brussel%20Sprouts.pdf [opened: 17/06/2022]

[13] Masini, G. L., Blanco, A. B. M., Petracci, N. C., Bandoni, A. (2008). Optimal operational planning in the fruit industry supply chain.

[14] NIAB (1990) Descriptive list of varieties of Brussel sprouts grown in NIAB trials. National Institute of Agricultural Botany, Vegetable Growers Leaflet No. 3, 1989.

[15] Roozbeh, I., Hearne, J., Abbasi B., Ozlen M. (2021). Decision support for wildfire asset protection: A two-stage stochastic programming approach.

[16] Stray, B.J., van Vuuren, J.H., Bezuidenhot, C.N., (2012) An optimisation-based seasonal sugarcane harvest scheduling decision support system for commercial growers in South Africa.

[17] Thuankaewsing, S., Khamjan, S., (2015) Harvest scheduling algorithm to equalize supplier benefits: A case study from the Thai sugar cane industry.

[18] Willis, C., Hanlon, W., (1976) Temporal model for long-run orchard decisions.

[19] https://www.kpu.ca/sites/default/files/ISFS/Brussel%20Sprouts.pdf