

AN OPTIMAL STRATEGY FOR SCHEDULING SEED POPULATIONS

Bruno S. Miranda

FEEC/UNICAMP

Av. Albert Einstein, 400, 13083-852, Campinas, Brasil

miranda.bsm@live.com

Akebo Yamakami

FEEC/UNICAMP

Av. Albert Einstein, 400, 13083-852, Campinas, Brasil

akebo@dt.fee.unicamp.br

Priscila C. Berbert Rampazzo

FCA/UNICAMP

R. Pedro Zaccaria, 1300, 13484-350, Limeira, Brasil

priscila.rampazzo@fca.unicamp.br

RESUMO

Alocação de terras cultiváveis é um ponto de início de muitas aplicações em Pesquisa Operacional. Trata-se de um problema desafiador, cujas aspirações e cenários são diversos. Neste artigo, dois modelos de otimização são propostos para uma derivação do problema de alocação, que consiste na distribuição de populações de semente ao longo de um intervalo determinado sem ultrapassar a capacidade de processamento da produção semanal. O Modelo A atua com base em graus-dia de crescimento para determinar o período de colheita. O Modelo B produz o planejamento do plantio das populações de semente, com o objetivo de atingir o melhor nível de aproveitamento da capacidade de processamento de sementes. Soluções ótimas foram encontradas para um número grande de populações de sementes. O tempo necessário para encontrar soluções foi bastante razoável, comprovando a próspera performance.

PALAVRAS CHAVE. Produção, Alocação de terras cultiváveis, Otimização.

Tópicos: AG&MA - Agricultura e Meio-ambiente; OC - Otimização Combinatória; PM-Programação Matemática

ABSTRACT

Agrarian land use allocation is the starting point for many applications in Operational Research. It is a challenge problem, which its aspirations and scenarios are diverse. In this paper, we proposed two optimization models of a related allocation problem, which is based on the distribution of seed population along a certain interval without crossing the maximum weekly processing capacity. The first one deals with the Growing Degree Days (GDD) to establish the harvesting period. The second generates the seeding schedule of the seed populations, aiming to do the best with the limited processing capacity. We are able to find optimal solutions within a large number of seed populations. Time spent to find solutions was very compelling, leading to prosperous performance.

KEYWORDS. Production. Agrarian Land Use Allocation. Optimization.

Paper topics: AG&MA - Agricultura e Meio-ambiente; OC - Otimização Combinatória; PM-Programação Matemática

1. Introduction

Allocating activities and resources to some areas in a goal-oriented context is not bounded by agrarian applications. It could be also interesting in a wide range of situations, such as urban development, transport and logistic expansion [Memmah et al., 2015]. From the viewpoint of Operational Research, planning a seed population schedule and assigning tasks have a profound bond in line with the techniques of problem-solving.

Optimization problems are grouped based on their kind of constraint equations and decision variables. One of the most compelling problem classes is Scheduling, which gathers the optimization problems of assigning and arranging personnel, resources and tasks in any general production setting. Scheduling problems are highly regarded among decision-making ones. Service systems usually deal with task attribution, where poorly planed assignments may cause remarkable losses. Making use of optimization techniques to reach powerful schedules can boost management and help organizations to reach their targets, meeting deadlines and improving resource utilization [Miranda et al., 2019b].

Section 2 contains a literature review about agrarian land use allocations. We describe the proposed models and solving techniques in Section 3. The gathered results are presented and talked over in Section 4. Finally, Section 5 contains some final remarks and future research directions.

2. Previous works

An agrarian land use problem comprises intricate tasks, which often gather attributes, discordant goals and many spatial factors. Surveying agrarian and optimization literature, we noticed that the crop rotation problem, which is the problem of allocating different crops in some areas under particular conditions, shares strong similarities with the current challenge. Therefore, in this subsection, we have discussed some fundamental researches about the crop rotation problem. This group of excellent works is our guideline and inspiration trend, inquiring modeling concepts and the state-of-the-art problem-solving methods.

Researching the crop rotation problem, Aliano et al. [2014] proposed a mixed-integer optimization model and hybrid meta-heuristics alternatives. In this research, hybrid algorithms combine local search and Simulating Annealing. Avoiding an unfavorable initialization, Aliano et al. [2014] developed an heuristic procedure to generate a suitable initial population.

Aliano et al. [2018] presented a bi-objective optimization model for the crop rotation problem. Allocating several distinct crops increases diversity, but it could compromise profits as the most lucrative crops weight equally in the sequence. Searching for a solution in the opposite direction might reduce diversity severely. Deterministic approaches for solving multi-objective problems were discussed in this work, exploring conflicting objectives in the proposed model.

Santos et al. [2011] proposed a crop rotation model based on organic farms. Adjacency and spatial limitations have a very clever design in the optimization model. Forrester e Rodriguez, 2018 also developed an integer binary linear model to crop rotation planning on organic vegetable crop production. In this research, modeling covers many characteristics, such as pest control, irrigation, fallow periods, field capacity. The crop rotation planning aims to meet market demand.

Miranda et al. [2019b] presented a mixed-integer optimization model for the crop rotation problem. Although the objective function is to maximize profits, an innovative constraint set relates fertilization parameters that optimize the soil nutrient storage and balance the chemical compound inputs. The outcome is a profitable and more sustainable allocation scheme. They also developed a genetic algorithm with special encoding techniques, providing good results.

Miranda et al. [2019a] developed a multi-objective optimization model. This research discussed the relationship between grown crops and soil fertility. The crop allocation is based on

crop cultural traits. Besides the profit maximization, the decision maker is able to balance input parameters to reach an erosion fight crop sequence or a weed killer solution.

3. The optimization strategy

In the opposite direction of standard farms, which their cultivable land defines their business potential, seed multiplication enterprises overcome the availability of cultivable land with partnerships. They provide financial aid and expertise to associate farmers and they pay more for their harvest production. Hence, in general, seed multiplication businesses can grow all their seed population, unbounded by how many acres they actually own.

Although we have ruled out the space problem of growing extensive field of seed populations, commercial seeds ought to be evaluated with the regulatory standard germination tests, which provide an accurate perspective about their seed quality. After sorting out the harvesting from the fields, packing and storing are critical steps to ensure seed viability until reaching their final consumers. Many farmers usually take part on the seed multiplication step, but processing seeds after harvesting is very limited to a small number of facilities, and not rarely, this job is up to just one site and its fully capacity.

Seed enterprises dedicated much effort in providing high-quality commercial seeds. As we have already suggested, seed production conditions and post-harvest processing treatments have a large effect on seed quality. Genetic developments in seed multiplication also play their essential part in this process.

At this point, we have already portrayed the problem components. The main objective is to schedule the planting date for each seed population to ensure the capacity constraints are attained. In other words, minimize the difference between the weekly harvest quantity and the processing capacity for each week.

At first, we could get an impression that this is small optimization problem with a simple objective. Looking at the huge seed market, we have an idea about how many genetic variations are present, which implies that many seed populations should grow up annually. Each seed enterprise deals with thousands of seed populations. Each population group has its own seeding interval and harvesting criteria. Seeding populations outside their proper interval increases the risk of crop diseases. Seed quality could be damaged by postponing harvest too much.

Figure 1: Achieved harvesting weekly quantity from site 0 in Scenario 1.

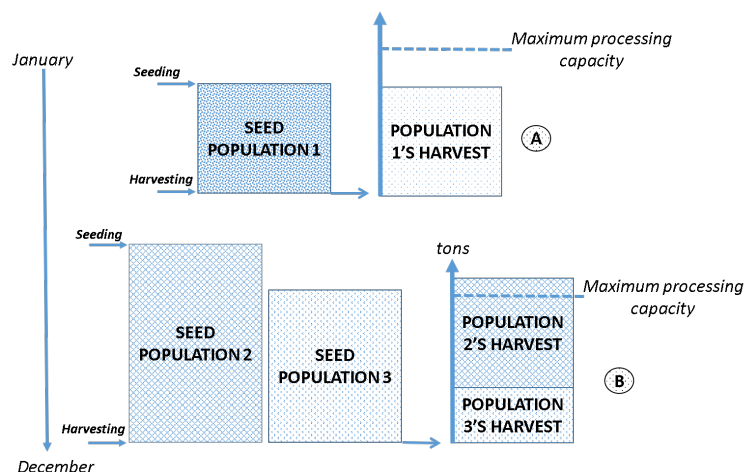
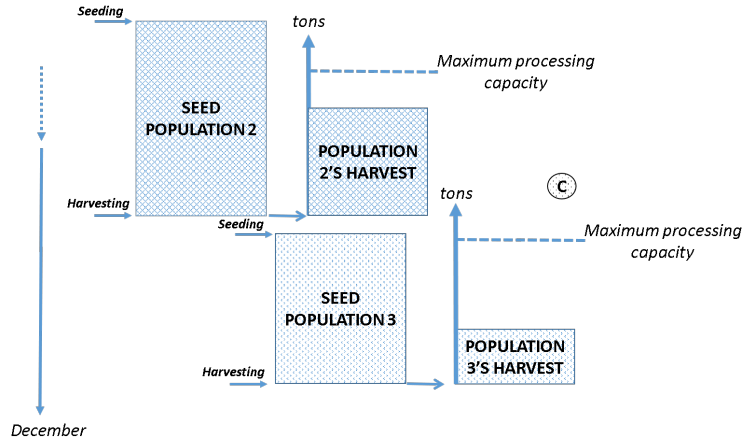


Figure 2: Achieved harvesting weekly quantity from site 0 in Scenario 1.



Eventually, with so many input parameters, sustaining reasonable harvesting quantities per week becomes a challenge. Some kind of predication about each population harvest can help to evaluate whether the site processing capacity would be overflowed or underused. Figure 1 shows the basic problem quest. In this example, processing population 1's production is successfully done without overflowing the maximum processing capacity.

When we observe the harvest of population 2 and 3, we notice that it is too much and exceeds the site capacity at the period. Reorganizing the seeding schedule would prevent overwhelming production at a certain period, as presented in Figure 2.

About harvesting, we based our knowledge on Growing Degree Days (GDD), which are used to estimate the growth and development of many crops during the growing season. The GDD's idea is that crop growth will happen if the temperature exceeds some minimum development threshold or base temperature. The base temperatures are gathered from experiments for each population.

Analyzing the problem context and desirable outcomes, we noticed that GDDs information and scheduling the planting date for each population could be taken apart, which reduces the complexity of handling an all-in-one optimization. Also, GDDs evaluation technique is straight forward, for any given input combination there will be a similar outcome. Scheduling optimization are far away from patterns.

Processing all GDDs at first before proceeding to the real optimization problem represents an expressive downsizing and some steps toward the main goal. Then, our propose consists in two optimization model: the first one, presented in Subsection 3.1, combines all planting dates of seed populations and their needed GDDs for harvest to estimate their harvesting periods; and the second one, described in Subsection 3.2, evaluates the schedule itself.

3.1. Model A: finding harvesting periods from each seed population

The solution of the proposed model is a set of harvesting periods indexed by seed population i and planting date j . For any given combination of seed i and j , $t[i][j]$ holds the harvesting date. Although the matrix $t[i][j]$ could be a long table, splitting the problem in partial optimization models increases solver's performance by far. The objective function in Equation (1) aims to reach the minimum harvesting for each i, j combination. The Model A's solutions will be used in Model B's optimization. Constraint in Equation (2) ensures that the harvesting period is greater the needed GDDs for each population i . Equations (3) and (4) hold the solution domain $1 \leq t[i][j] \leq J$.

The input indexes and parameters are as follows:

- N : number of populations;
- J : late harvesting period;
- Y : late planting period;
- $GDD[j]$: a cumulative growing degree days (GDDs) vector;
- $H[i]$: the number of GDDs in Celsius needed for population i 's harvest.

The decision variable set is $t[i][j]$. As each solution is an index, it should be integer values only.

- $t[i][j]$: harvesting period for seed population i , which has been planted in period j .

Harvesting periods are based on site environment and its GDDs records. They need to be assessed individually. There will be a single solution for each site. Once we completed the optimization process, we can pull together all site solutions.

$$\text{minimize} \quad \sum_{i=1}^N \sum_{j=1}^Y t[i][j] \quad (1)$$

$$\text{subject to} \quad GDD[t[i][j]] - GDD[j] \geq H[i], \quad i = 1, \dots, N, \quad j = 1, \dots, Y \quad (2)$$

$$t[i][j] \geq 1, \quad i = 1, \dots, N, \quad j = 1, \dots, Y \quad (3)$$

$$t[i][j] \leq J, \quad i = 1, \dots, N, \quad j = 1, \dots, Y \quad (4)$$

Considering the following example, which has just two seed populations ($N = 2$) and the scheduling span is just six periods away (late planting period $Y = 3$ and late harvesting period $J = 6$). Vector H holds the needed GDD for harvesting. $S2$ matrix represents the planting interval for seed populations 1 and 2. GDD is a cumulative vector with size J .

- $N = 2$;
- $J = 6$;
- $H = [20, 35]$;
- $Y = 3$;
- $S2 = [[1, 3], [2, 3]]$;
- $GDD = [10 \ 25 \ 40 \ 70 \ 85 \ 100]$.

The outcome presents in Table 1. For each combination of crop i and planting period j , we have a respective harvesting period j . Each site has its own GDD records, therefore, if there are more than one site, there will be solutions similar to the one in Table 1 for each site k . Thinking about many sites, we could gather the outcomes in single multidimensional matrix $T[i][j][k]$.

Table 1: Outcome solution: harvesting combinations.

$t[i][j]$	$j = 1$	$j = 2$	$j = 3$
$i = 1$	3	4	4
$j = 2$	4	4	5

3.2. Model B: scheduling seed populations

In this subsection, we describe the proposed optimization model for scheduling seed populations. Model parameters introduced ahead:

- $W[k]$: storage capacity limitation of site k ;

- $Ib[i], Ie[i]$: early and late seeding periods of seed population i , it begins in $Ib[i]$ and ends in $Ie[i]$;
- $S2 = \{[Ib[1], Ie[1]], \dots, [Ib[n], Ie[n]], \dots, [Ib[N], Ie[N]], \}$;
- $P[i]$: harvesting quantity of seed population i ;
- $T[i][f][k]$: it returns the harvesting period from planting population i in period j on site k ;
- $S[i]$: it returns the site of population i ;
- N : number of populations;
- Q : number of sites;
- J : late harvesting period;
- Y : late planting period;
- α : number of weeks;
- β : week size;
- M : penalty parameter.

Decision variables are discussed in the following list:

- $x[i][j][k]$: population i seeded on period j at site k ;
- $p[j][k]$: harvesting amount in period j on site k ;
- $a[n][k]$: adjust parameter;
- $wp[n][k]$: harvesting total in interval n on site k .

Constraint in Equation 6 evaluates production from each period j . Harvesting quantity $wp[n][k]$ from a interval β is modeled by constraint in Equation 7. Capacity constraint presents in Equation 8. In the model, constraint in Equation 9 ensures the proper planting date for all seed populations. Equation 10 guarantees that all the seed populations are planted once. Constraint in Equation 11 ensures that seed population i would be planted only in its predefined site $S[i]$.

If storage capacities $W[k]$ were really tight, infeasible solutions would appear often. Taking into account that any solution is better than no solution at all, we introduced the adjust parameter $a[n][k]$. If there are solutions without compromising storage capacity $W[k]$, $a[n][k]$ will be null, which is held certain by constraint in Equation 13.

$$\text{minimize} \quad \sum_{k=1}^Q \sum_{n=1}^{\alpha} |W[k] - wp[n][k]| + M \cdot a[n][k] \quad (5)$$

$$\text{subject to} \quad p[j][k] = \sum_{i=1}^N \sum_{f=1}^Y x[i][j][k] \cdot P_i \cdot (T[i][f][k] == j) \quad (6)$$

$$j = 1, \dots, J, \quad k = 1, \dots, Q$$

$$\sum_{s=1}^{\beta} p[(n-1) \cdot \beta + s][k] = wp[n][k], \quad k = 1, \dots, Q, \quad n = 1, \dots, \alpha \quad (7)$$

$$wp[n][k] \leq W[k] + a[n][k], \quad k = 1, \dots, Q, \quad n = 1, \dots, \alpha \quad (8)$$

$$\sum_{j=1}^Y \sum_{k=1}^Q x[i][j][k] - \sum_{j=Ib[i]}^{Ie[i]} \sum_{k=1}^Q x[i][j][k] \cdot (Ib[i] \leq Ie[i]) \quad (9)$$

$$- \sum_{j=Ib[i]}^Y \sum_{k=1}^Q x[i][j][k] \cdot (Ib[i] > Ie[i]) - \sum_{j=1}^{Ie[i]} \sum_{k=1}^Q x[i][j][k] \cdot (Ib[i] > Ie[i]),$$

$$i = 1, \dots, N$$

$$\sum_{j=1}^Y \sum_{k=1}^Q x[i][j][k] = 1, \quad i = 1, \dots, N \quad (10)$$

$$\sum_{i=1}^N \sum_{j=1}^Y x[i][j][k] \cdot (k \neq S[i]) = 0, \quad k = 1, \dots, Q \quad (11)$$

$$x[i][j][k] \in \{0, 1\} \quad (12)$$

$$a[n][k] \leq 0, \quad k = 1, \dots, Q, \quad n = 1, \dots, \alpha \quad (13)$$

$$\alpha = \{\alpha \in \mathcal{I} | \alpha = J/\beta\} \quad (14)$$

Considering that β represents a week interval ($\beta = 7$) and the late harvesting period J is a two-year-long interval ($J = 730$), by the set definition in Equation 14, the number of weeks α will be equal to 104 ($\alpha = 104$).

Some changes in the model are required to reach minimum processing capacity. We need to replace the objective function in Equation 5 with Equation 15. The processing capacity of site k is represented by $w[k]$:

$$\text{minimize} \quad \sum_{k=1}^Q w[k] \quad (15)$$

We also have to replace the capacity constraint in Equation 8 with Equation 16:

$$wp[n][k] \leq w[k], \quad k = 1, \dots, Q, \quad n = 1, \dots, \alpha \quad (16)$$

For illustration purpose only, we have replaced Equation 7 with Equation 17. By this modification, a period j accounts by a whole week. By that, we also need to introduce a slightly change in the objective function, which is described by Equation 18.

$$p[j][k] \leq W[k], \quad j = 1, \dots, J, \quad k = 1, \dots, Q \quad (17)$$

$$\text{minimize} \quad \sum_{k=1}^Q \sum_{j=1}^J |W[k] - p[j][k]| \quad (18)$$

Input parameters are described as follows:

- $N = 2$;
- $Q = 2$;
- $S2 = [[1, 3], [2, 3]]$;
- $Y = 3$;
- $P = [354, 342]$;
- $W = [400, 400]$;
- $J = 6$;
- $SITE = [1 \quad 2]$;

Optimizing multiple seed processing sites ($Q > 1$), S vector would indicate a corresponding site to manage each seed production, which enables or disables each seed population from site to site as we could observe in the example below:

- $S = [1, 1, \dots, 0, 1, 1]$ (the dimension of S vector is N);

Harvesting periods in Table 2 are based on results from Subsection 3.1.

Table 2: GDDs input.

T[i][j][k]						
	k=1			k=2		
	j = 1	j = 2	j = 3	j = 1	j = 2	j = 3
i = 1	3	4	4	3	4	4
i = 2	4	4	5	4	4	5

Results are presented in Table 3.

Table 3: Minimum example from scheduling optimization model.

x[i][j][k]				p[j][k]		
i	j	k	value	j	k	p
1	1	1	1	3	1	354
2	2	2	1	4	2	342

4. Results and observations

We selected IBM ILOG CPLEX Studio IDE Studio 12.8.0.0, that supports Optimization Programming Language (OPL), to build our proposed optimization models and generate solutions. CPLEX is one of the most popular large-scale solver worldwide due to its exceptional performance and friendly interface. We evaluated the problem in a three-thousand seed populations database ($N = 3000$). About the test workbench in this paper, we developed a database based on real data and applying a probability density function to produce many populations.

The seeding schedule is a year-long plan ($Y = 365$). Seeding at the end of the planning interval implies that harvest will happen in the following year, and so, we need a few more periods to account all the seed production ($J = 730$).

The first step is to generate all the combinations of seed population i and period j that lead to the outcome harvesting period $t[i][j]$, which we get from finding Model A's solutions. Due to the large number of inputs, we split the seed population set in 500-seed-population groups. As the total number of seed populations is 3000, this step would repeat six times to get all the correspondent harvesting combinations. Without further changes in the GDD data, there will be no need to solve Model A again. A brief description of the parameters is ahead:

- $S2 = [[309, 364], [180, 297], \dots, [137, 170], \dots, [87, 133]]$ (the dimension of $S2$ is $N \times 2$);

For each site, there is a GDD vector:

- $GDU[j] = [9, 22, 31, \dots, 7211, 7221, 7230]$ (the dimension of GDU is J);

The list ahead provides a fair description of the parameters required to evaluate Model B:

- $N = 3000$;
- $Y = 365$;
- $J = 730$;
- $H = [1060, 1060, 1060, 965, \dots, 1159, 1259]$ (the dimension of H vector is N);
- $P = [354, 342, 300, \dots, 346, 268]$ (the dimension of P vector is N);
- $Q = 1$;
- $W = 23500$;
- $T = [[109 \ 110 \ 111 \ \dots; \vdots; \dots \ 472 \ 473 \ 473] \dots]$ (the dimension of T is $N \times Y \times Q$).

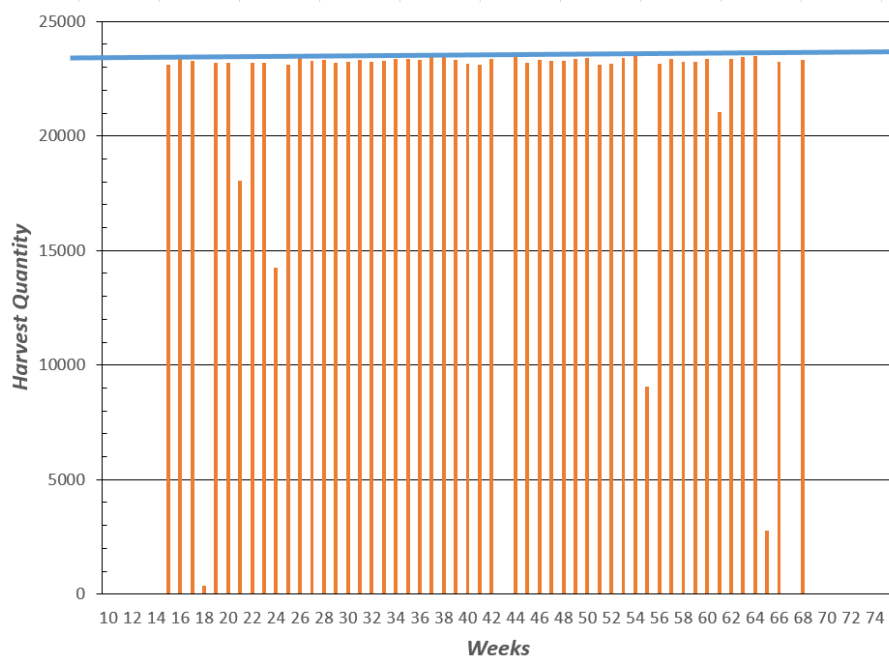
Solving the problem and generating a seed schedule solution took 500 seconds. Table 4 shows other solver parameters related to the optimization of Model B.

Table 4: Model B: solver's parameters.

Statistics		
Solver	Cplex	
Constraints	7146	
Variables	1096250	
	Binary	1095000
	Integer	938
	Other	312
Non-zero coefficients	2995779	
Iterations	3130	

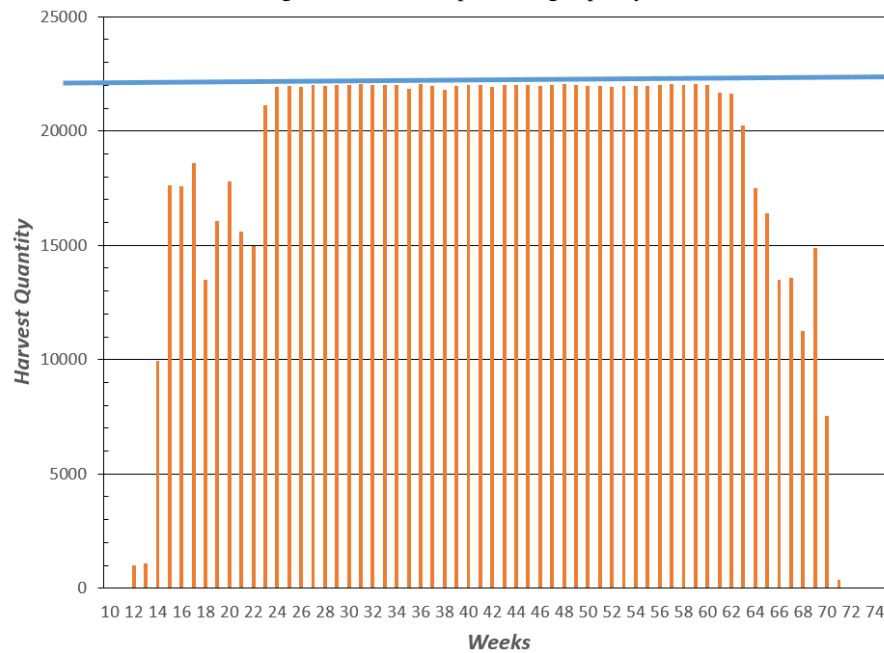
From the optimal solution found, Figure 3 exhibits weekly harvesting quantities of Site 1. Processing capacity in this test is $W = 23500$. We got the minimum deviation from site processing capacity, increasing the efficiency of the whole seed processing facility.

Figure 3: Efficient seed distribution: minimum deviation.



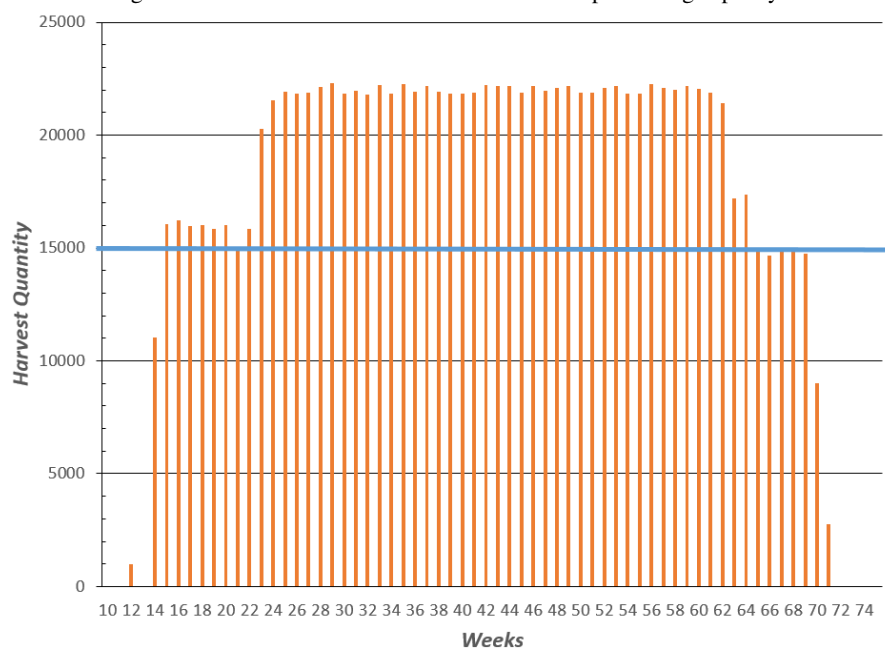
Looking for minimum capacity, we found the results in Figure 4. It is also acclaimed because there is a consistency among every weekly harvest amount. The minimum capacity value was 22047. We are neglecting defined units as the optimization models could handle any kind of seed multiplication.

Figure 4: Minimum processing capacity.



Model B deals with unfeasible problems using the *adjust* variables, which loose the processing capacity. In that case, when capacity is too short, a solution schedule would produce peaks as shown in Figure 5.

Figure 5: Undesirable solution: harvest overflows processing capacity.



5. Conclusion

Planning large-size operations without applying optimization can be an exhaustive task and counterproductive. In this research, we have presented alternatives to manage the distribution of seed populations subject to processing capacity constraints. We proposed mathematical models and tested the optimization overall performance in a testing workbench with plenty of seed populations. As we evaluate the optimization models using deterministic methods, we managed to achieve excellent solutions. Although modeling agrarian problem requires abstraction to navigate through the complexity of living things, creating connections between Agriculture and Operational Research is a delightful experiment. Developing problem-solving strategies in the agrarian field is always ongoing as our understanding about the problem universe evolves.

References

- Aliano, A., Florentino, H., e Pato, M. (2014). Metaheuristics for a crop rotation problem. *Int. J. Metaheuristics*, 3(3).
- Aliano, A., de Oliveira Florentino, H., e Pato, M. V. (2018). Metodologias de escalarizações para o problema de rotação de culturas biobjetivo. *Proceeding Series of the Brazilian Society of Applied and Computational Mathematics*, 6(1).
- Forrester, R. J. e Rodriguez, M. (2018). An integer programming approach to crop rotation planning at an organic farm. *The UMAP Journal*, p. 5–23.
- Memmah, M.-M., Lescourret, F., Yao, X., e Lavigne, C. (2015). Metaheuristics for agricultural land use optimization. a review. *Agronomy for Sustainable Development*, 35(3):975–998. ISSN 1773-0155. URL <https://doi.org/10.1007/s13593-015-0303-4>.
- Miranda, B. S., Yamakami, A., e Rampazzo, P. C. B. (2019a). A multiobjective approach for crop rotation planning. In GALOÁ, editor, *Technological Innovation for Industry and Service Systems*, p. 1–12, Limeira. LI SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL.
- Miranda, B. S., Yamakami, A., e Rampazzo, P. C. B. (2019b). A new approach for crop rotation problem in farming 4.0. In Camarinha-Matos, L. M., Almeida, R., e Oliveira, J., editors, *Technological Innovation for Industry and Service Systems*, p. 99–111, Cham. Springer International Publishing.
- Santos, L., Michelon, P., Arenales, M., e Santos, H. (2011). Crop rotation scheduling with adjacency constraints. *Ann Oper Res*, 190:165–180.