Strategic Planning in Citriculture: An Optimization Approach

Cassiano Tavares[a], Pedro Munari[a,*]

[a] *Federal University of São Carlos, Production Engineering Department, Rodovia Washington Luis Km 235, São Carlos, CEP 13565-905, São Paulo, Brazil*

Abstract

The worldwide citrus market has been impacted by various factors in recent years, including population growth, phytosanitary diseases, high costs of agricultural inputs, and diminishing planting areas. As a consequence, producers in this sector have attempted to find tools to support strategic planting decisions, and thus meet international contract demands. This paper proposes an optimization tool for supporting the strategic planning of planting decisions in citriculture, based on mathematical models and algorithms that address real-world requirements. The motivation for this study stems from our collaboration with one of the world's largest orange juice producers. We consider specific characteristics of the citrus business, estimates for productivity and eradication, and desired balance levels for orange varieties and plant age groups. To the best of our knowledge, there are no previous studies proposing optimization approaches that explore these unique characteristics of citrus strategic planting. We validate the effectiveness of the proposed approach through computational experiments using realistic instances based on the company's data. The results show that our approach provides effective support to decision making and can significantly increase fruit box production over a 30-year planning horizon while, most importantly, satisfying all the company's requirements on varietal and age balance as well as planting and eradication control.

Keywords: Optimization, Strategic planning, Planting, Citrus, Mathematical model

---

1. Introduction

Brazil is currently the biggest orange producer in the world, considering both fruit and juice [1]. Overall, agrifood chains represent approximately 25% of Brazil's Gross Domestic Product (GDP), 30% of the country's jobs, and almost 50% of all Brazilian exports [2]. The citriculture generates an annual turnover of around 14 US$ billion for the Brazilian economy [3]. The total shipments of Brazilian orange juice from July to December 2022, which is equivalent to the first six months of the 2022/2023 harvest, reached a total volume of 586,313 tons, which represents an increase of more than 17% compared to the same period of the previous harvest (2021/2022), where 500,323 tons were exported [4]. Regarding revenue, exports of this commodity totalled US$ 1.1 billion in the period, equivalent to an increase of around 37% above the gain of US$ 803.8 million registered between July and December 2021. The largest production of concentrated orange juice takes place in the Citrus Belt region, which includes the state of São Paulo and the south of the state of Minas Gerais [3]. São Paulo is responsible for producing approximately 80% of the national export volume.

---

[*] Corresponding author

Email addresses: `cassiano.tavares@estudante.ufscar.br` (Cassiano Tavares), `munari@dep.ufscar.br` (Pedro Munari)

Orange crops are classified in agribusiness as perennial due to the plant's life cycle, which exceeds 15 years. This extended life cycle allows for multiple harvests from a single planting [5]. This feature introduces several challenges due to variability in production at each harvest, which directly impacts the management process and financial performance of the crop, often relying on external investments [6, 7]. It is not uncommon to observe a sharp drop in fruit productivity at the end of the life cycle of the plants. Therefore, eradicating older plants and planting new seedlings becomes a critical decision [6]. Additionally, the crop is susceptible to phytosanitary diseases, with the most frequent being greening (*huanglongbing* or HLB), citrus variegated *chlorosis* (CVC), and citrus canker [8, 9, 10]. The study of these diseases is not within the scope of this paper, and we refer the interested reader to, e.g., [8, 9, 10]. By the second half of 2023, Brazil was suffering from the lowest stock of orange juice since June 2011, primarily due to climate change and the greening disease, both of which affect orchards and reduce the quality and productivity of the fruit. In August 2023, there were 84,745 tons of juice stored by members of the National Association of Citrus Juice Exporters, which is 40% less than in 2022. This stock balance affects both orange juice and fruit consumers [11].

The above-mentioned situations motivate the need for efficient strategic planning to support planting/eradication decisions in agricultural management. The present study is precisely situated within this context, focusing on using analytics and operations research (OR) techniques to optimize the strategic planning of planting in citriculture. The purpose is to fulfill the essential and desired requirements outlined by producers while simultaneously maximizing orange production. Although the same agro-food chain niche has been addressed in some previous studies [12, 13], the present paper aims to bridge a gap in the state-of-art in this literature by introducing optimization approaches for strategic planning. We are not aware of any study addressing this topic in citriculture thus far.

The contributions of this paper are fourfold. First, we present a systematic literature review (SLR) on quantitative approaches to aid decision-making in planting and harvesting. Second, we present a thorough description of the addressed problem based on the interaction with one of the world's largest orange juice producers. Third, we propose a new optimization model to assist in decision-making in the strategic planning of citrus planting. We validate our model using real-world data provided by the partner company. We considered some specific company characteristics that are common in the planting field, including varietal control, the desired age profile, productivity and eradication curves, and planting and eradication limits. Finally, we solve a large-scale problem instance based on real-world data and present managerial insights for different scenarios.

The remainder of this document is structured as follows. In Section 2, we present the SLR on the research topic. In Section 3, we characterize and define the addressed problem. Section 4 proposes a Mixed-Integer Programming (MIP) model to support decision-making in strategic planning of planting in citriculture. Section 5 presents the computational experiments based on real-world data. Finally, Section 6 presents the conclusions.

2. Literature Review

This section presents an SLR that follows the framework proposed in [14]. The scope, keywords, and inclusion/exclusion criteria definition are based on [15, 16, 17, 18, 19]. The selected keywords were *harvest*, *planting*, *operations research*, and similar expressions, such as *operational research* and *mathematical optimization*. The string used in the search engines of the literature databases was: TITLE-ABS-KEY ((“harvest” OR “planting”) AND (“mathematical optimization” OR “operations research” OR “operational research”)). This search string was used to search the articles' titles, abstracts, and keyword fields. The language of the articles was restricted to Portuguese and English. No time limit regarding the publication date of the articles was imposed. We selected the following literature databases: Scopus, Engineering Village, and Web of Science. For this search, the Scopus base returned 140 papers; the Engineering Village base, 90 papers; and the Web of Science base, six papers. Therefore, 236 articles were identified for analysis in the selection phase.

In the planning phase of our SLR, we adopted the following exclusion criteria: simulation studies and guidelines; not within the scope of the research; and no access to the document. The inclusion criteria were: mathematical models consisting of at least one of the decision variables related to harvest planning and/or production in agro-food chains; mathematical models that include decision variables related to agro-food chains' production planning or logistics; articles that meet the previous rule, but address other perishable agricultural products, such as vegetables instead of fruits; and articles addressing supply chains of food crops that provide food for human consumption.

After analyzing the 236 articles and reading the title, abstract, and keywords fields, we obtained the following result: 18 duplicate articles, 182 rejected articles, and 36 accepted articles. The high number of rejected articles have the following reasons: (i) the keyword *harvest* refers to harvesting; however, shrimp farms use this term for the "harvesting" of fish; (ii) there is a very high number of articles aimed at planting, harvesting, disposal, and routing of trees in forests, especially in countries in Europe and North America (not adhering to the scope); and (iii) due to environmental concerns, several surveys assess water and climate impacts on planting and harvesting operations.

In the SLR extraction phase, we evaluated the 36 articles selected in the previous phase to perform a complete reading of them. After this, 21 papers were rejected based on the following reasons: no access to document; no mathematical model in the article; or not within the scope. The SLR then resulted in 15 accepted papers, which are discussed in the following subsections. They were grouped by strategic, tactical, and operational planning levels, according to the operations management theory [20].

2.1. OR Applications in Agriculture - Operational Level

Higgins et al. [21] and Higgins [22] presented an optimization model and computational experiments for an application in sugarcane harvesting, considering a mill located in the state of Queensland, Australia. The model was based on an extension of the generalized assignment problem, considering the feasibility of transporting sugarcane production by road and rail in five regions of Australia. Caixeta-Filho [12] was a pioneer in the citrus context, introducing the first model for sequencing orange harvests. The author considered two scenarios. In the first, the model aimed to maximize the profit generated by the number of soluble solids in the fruit, while in the second, the model sought to maximize the profit related to each fruit box. He et al. [23] proposed three MIP models to support rice harvesting decisions in one of the leading rice plantations in China. Finally, Escallón-Barrios et al. [24] developed a discrete event simulation model to measure the impact of uncertain events on an oil palm plantation located in the city of Maní in the Colombian Orinoquia.

2.2. OR Applications in Agriculture - Tactical Level

Florentino et al. [25] presented a bi-objective optimization model to support decision-making related to sugarcane planting, considering two objectives, namely (i) minimizing the cost of transferring the straw from the field to the processing center, and (ii) maximizing the energy balance of the residual biomass from the sugarcane harvest. Florentino and Pato [26] addressed the same problem but using a different solution strategy, called GenSugar, which was based on a genetic algorithm (GA) meta-heuristic. In a related application, Poltroniere et al. [27] proposed a MIP model to support decisions in sugarcane harvesting, considering a specific sugarcane variety suitable for energy generation through biomass.

Munhoz and Morabito [13] proposed a robust optimization (RO) model for the tactical planning of concentrated orange juice production, aiming to meet the demand signed in contracts through strategic planning. The model considers two stages: in the first, the planning for processing oranges is carried out, generating intermediate products (juice bases), while in the second stage, these juice bases are mixed to obtain the final products (concentrated orange juice). Osaki and Batalha [28] introduced an optimization model to help the planning process of planting soybeans and corn, considering different sources of uncertainties. The authors considered two objective functions, where the first aimed to maximize the Gross Contribution Margin (GCM), while the second consisted of minimizing the Contribution Margin Risk (CMR). To solve the model, they adopted the so-called Minimization of Absolute Total Deviation (MOTAD) method.

A mixed integer nonlinear programming bi-objective model was introduced by Aliano Filho et al. [29] to support decision-making in the cultivation of several types of crops (vegetables, tomatoes, potatoes, among others). The mathematical model aims to optimize two conflicting objectives, the first given by minimizing the possibility of spreading phytosanitary diseases (pests) between crops, and the second objective given by maximizing the plantation's profit throughout the planning horizon.

2.3. OR Applications in Agriculture - Strategic Level

Darby-Dowman et al. [30] presented a two-stage stochastic programming model to support planting and harvesting decisions at the strategic level for Brussels sprouts in the United Kingdom. Catalá et al. [6] developed a MIP model to aid decision-making related to apple and pear production in the Alto Vale do Rio Negro region in Argentina. The model aimed to maximize the investment project's net present value (NPV) for the harvest. Brulard et al. [31] integrated strategic and tactical decisions in an MIP model focusing on small rural producers that are suppliers of small markets, farms, and restaurants. The model was applied to an experimental garden in France. Rajakal et al. [7] introduced an optimization model in the context of perennial oil palm crops in Malaysia, which was used to determine the ideal maturity level of the plant to fulfill demands. Two approaches with distinct objective functions were evaluated, namely the minimization of the total cost over the planning horizon and the maximization of the discounted carbon value (DCV).

2.4. Summary of the SLR

The presented papers reveal a diversity of crops, solution approaches, performance measures, and objective functions in the applications of OR in Agriculture. Table 1 summarizes the main features of these papers and presents the classification of our study with respect to the state-of-the-art. The columns present the article's reference (Article); the type of the solution approach used (Approach); whether the study considered more than one variety of product (Var); the amplitude of the planning horizon in months; the number of farms; the type of objective function; the crop(s) considered in the study; and, finally, the planning level classification. The table indicates that our study brings a contribution in the context of OR in Agriculture, especially in strategic planning of planting in citriculture. We are unaware of any other study addressing a similar situation.

We observe a balanced distribution of papers among the three planning levels. Specifically, 27, 40, and 33% of papers addressed decisions at strategic, tactical, and operational

Table 1: Summary of our SLR.

Article	Approach	Var	Months	Farms	Objective Function	Crop	Planning level
Higgins et al. [21]	Heuristics	Y	48	—	max Profit	Sugarcane	Operational
Higgins [22]	Heuristics	Y	48	216	max Profit	Sugarcane	Operational
Darby-Dowman et al. [30]	Stochastic	Y	372	1	max Profit	Brussels sprouts	Strategic
Caixeta-Filho [12]	MIP	Y	1	320	max Profit	Orange	Operational
Florentino et al. [25]	MIP	Y	—	1	min EB/ min Costs	Sugarcane	Tatic
Catalá et al. [6]	MIP	Y	240	1	max NPV	Apples/Pears	Strategic
Florentino and Pato [26]	GA	Y	12	1	max Prod. / min Stock	Sugarcane	Tatic
Munhoz and Morabito [13]	RO	Y	12	—	min Costs	Orange	Tatic
Osaki and Batalha [28]	MOTAD	N	60	1	max GCM / min CMR	Soybean/Corn	Tatic
He et al. [23]	MIP	N	—	1	min Harvest time	Rice	Operational
Brulard et al. [31]	MIP	Y	12	1	max Profit	Multiple	Strategic
Rajakal et al. [7]	MIP	N	120	5	min Costs / max DCV	Palm	Strategic
Poltroniere et al. [27]	MIP	Y	12	1	max Prod. / min Stock	Sugarcane	Tatic
Escallón-Barrios et al. [24]	MIP	N	0,65	1	max Profit	Palm	Operational
Aliano Filho et al. [29]	Heuristics	Y	12	1	max Profit	Multiple	Tatic
This study	**MIP**	**Y**	**360**	**≈30**	**max Prod.**	**Orange**	**Strategic**

Method: MIP = Mixed Integer Programming, GA = Genetic Algorithm, RO = Robust Optimization; Objective Function: max = maximize, min = minimize, EB = minimize Energy Balance, DCV = Discounted Carbone Value, Prod. = Productivity, GCM = Gross Contribution Margin, CMR = Contribution Margin Risk; Var: Y = yes, N= no.

levels, respectively. However, different scenarios have been reported in previous studies [16, 19]. Soto-Silva et al. [16] reported the following distribution among the planning areas: 15% - strategic, 50% - tactical, and 35% - operational; while Nguyen et al. [19] found 18% - strategic, 50% - tactical, and 32% - operational. This difference from these previous studies can be explained by our choice of keywords, which is more adherent to the scope of this study. Moreover, we observed an increase in the number of publications related to the tactical level over the last five years.

It is worth mentioning that [6] has a few similarities to our study regarding the solution approach and because it considers varietal control in its MIP model. Nevertheless, the two studies differ in essence according to several points, including objective function, problem dimensions and crop. In [6], the authors considered an objective function based on the investments related to the crop planting while we contemplate the production amount. In our case, this is justified by the choice made by our partner company in not including an intricate financial analysis at this decision stage. While Catalá et al. [6] considered one single farm, our study involves around 30 farms, increasing the computational effort of the solution approach. Additionally, they assumed a planning horizon of 240 months, whereas we consider 360 months (ten more years), as required by our partner company. Finally, [6] addressed the planting of apples and pears while we consider oranges. The peculiarities of each culture led to specificities in the modeling, solution strategies, and analysis.

There are studies related to OR applications in citriculture but not within the scope of our SLR, as they focus on different aspects of the supply chain. For example, Munhoz and Morabito [32, 33, 34] presented optimization approaches for aiding decision-making in tactical production planning of factories that produce concentrated orange juice. As their focus was on industrial aspects but not agricultural, these papers did not adhere to the presented SLR. We refer interested readers to these papers and the references aforementioned by them.

3. Problem description

The strategic planning of planting and production to meet the projection of future revenues is one of the main challenges faced by orange-producing groups [13]. As previously mentioned, the involved decisions are not trivial because the productivity of the farms depends on various characteristics, encompassing orange varieties, the plants' life cycles, fruit maturation, weather conditions, and many others.

The strategic planning process in our partner company takes place annually and considers essential information, including the number of farms and plots in each of their farms; the variety, age, area, and quantity of plants in each plot; estimations on the plant productivity and natural eradication in each plot; and the area available for planting in the farms. Additionally, the decision-makers define their desired percentages of each orange variety type and the desired age profile for the trees throughout the farms. With all this information at hand, they seek to define what should be planted and eradicated in the coming years at each farm to promote maximum orange production in a time horizon of 30 years. This time horizon is chosen to give a broad view of the expected results in the long term based on the planned decisions. Obviously, these decisions are reevaluated every year, and even the decisions for the current year are later revised in tactical and operation planning processes.

The farms are divided into plots containing only trees of the same variety, rootstock, and age, as all the trees in a plot are planted simultaneously. Additionally, one plot has a single irrigation technique and plant density. Thus, each plot has unique attributes that define it: the age of plants (according to the year of planting), variety, rootstock, irrigation technique, and density. According to the company practice, we group plots with the same attributes in a so-called *stratum*. Hence, in our study, we consider a set of strata, where each stratum is specified as a tuple *(variety, rootstock, density, irrigation)* used to represent the group of all plots with the same attributes in a farm.

The main purpose of the planning process is thus to determine *what*, *how much*, *where* and *when* to plant and eradicate at each farm to maximize orange production while satisfying the desired requirements regarding varietal and age balance. The output has to specify what strata and how much of them to plant, referring to all the possible combinations of variety, rootstock, density, and irrigation; where to plant these strata, regarding the available area in each farm, as well as areas that may result from eradication; and when to plant these strata in the upcoming years of the planning horizon. As mentioned, the decisions also involve what areas to eradicate and when. Therefore, given a long-term planning horizon, a set of farms with occupied and available areas, and a set of strata planted as well as available for planting, the decisions can be summarized as follows:

Which stratum to plant/eradicate in what area of each farm at each year to maximize total orange production, considering the varietal and age balance requirements desired by the company?

This is certainly a very complex decision-making process, with too many variables, scenarios, and possibilities. Empirical decisions are likely to yield poor results regarding production and the desired level of varietal and age balance levels. Additionally, as reported by the partner company, the fluctuation of orange production throughout the months and the difference in total production from one year to the next, have been significantly large, which negatively affects resource management and juice production. To overcome these drawbacks, in this paper, we introduce an optimization tool that resorts to mathematical models and algorithms to aid the described decision-making process, making it more effective and efficient.

In the remainder of this section, we detail all the main components involved in the described decision process, also giving the main characteristics and challenges related to orange cultivation, as elucidated together with our partner company. The different types of orange varieties and the motivation for requiring a varietal balance in the production are discussed in Subsection 3.1. The other attributes that define a stratum, namely rootstock, irrigation, and density, are described in Subsections 3.2 to 3.4. In Subsection 3.5, we present the motivation for the desired plant age balance in the farms. Finally, in Subsection 3.6, we describe an important data input used in the decision process, given by the expected plant productivity and eradication curves.

3.1. Fruit Varieties and Varietal Balance

Oranges are classified into several varieties that exhibit a wide variability in terms of fruit characteristics, such as color, taste, yield, maturity date, and many other horticulturally important traits [35]. This variability is the result of field selection, propagation, and diffusion of selected varieties in different cultivation areas over the years. Each variety of orange may have different cues for its ripening stages, influenced by environmental factors and cultivation practices. These stages are heavily dependent on the cultivation environment and the influence of the rootstock. Understanding the ripening stages is crucial for producers to determine the optimal timing for harvesting to maximize flavor, nutritional value, and market appeal. Because oranges do not ripen further once harvested, recognizing their ripening stages while still on the tree is essential for timing the harvest effectively [5, 36].

The last census regarding the Brazilian Citrus Belt [37] showed that 387,169 hectares were used for orange cultivation in 2020. The orchards are made up of three groups of fruit maturation, with late maturation given by 13 to 15 months, medium maturation or mid-season by 10 to 13 months, and early maturation by 8 to 10 months. In recent years, the preference of citrus growers for late-maturing crops has occurred to the detriment of medium-maturing crops, which have lower productivity and multiple flowering, aggravating the control of pests and diseases [38]. The flowering season of citrus varieties is illustrated in Figure 1, which indicates that each variety has a specific time of greater fruit concentration. Hence, the ideal mix of fruit harvest proportion established by our partner company is used to define the varietal balance levels, promoting the uniform production of orange juice over the year by the industry. In our study, the varieties were grouped into five classes according to their maturation cycle and canopy, following the partner company's practice. These classes were named using letters A to E, where the main representatives in each class are Hamlin, Natal, Valencia Americana, Pera, and Valencia, respectively.

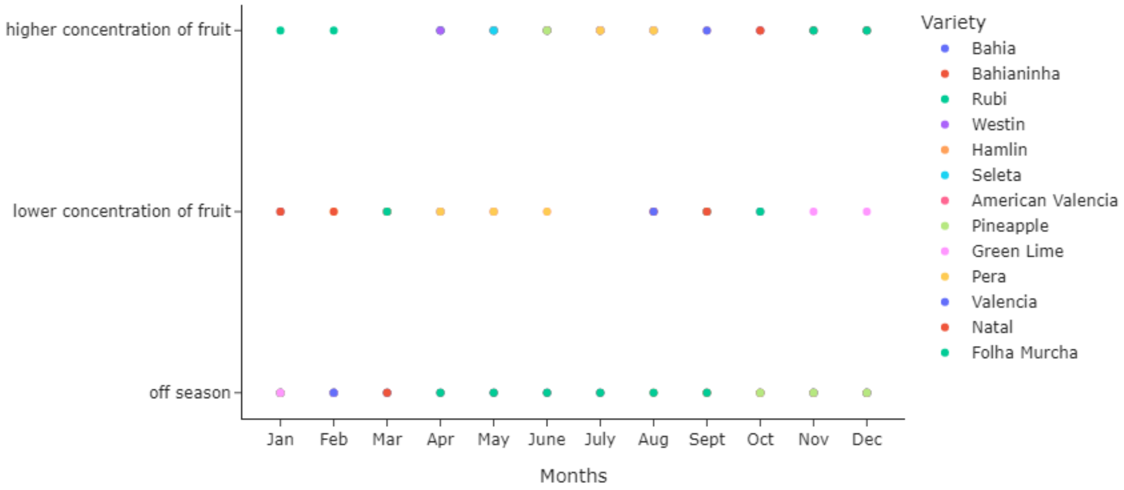

Figure 1: *Flowering season of orange varieties (adapted from [3]).*

3.2. Rootstock

One of the key factors contributing to the success of an orchard is the careful selection of citrus seedlings generated through a grafting process using the T-budding method shown in Figure 2. This method involves combining the graft and the rootstock (matrix plants). The graft forms the crown of the plant, taken from a specific scion variety – it is the visible part comprising leaves and branches that produce fruit. The rootstock develops into the plant's root system under ground. The two plant parts unite and develop as one. Rootstocks have influence on orange production, affecting yield, ripening period, ability of the tree to retain fruit, fruit size and shape, soluble solids, acid concentrations, among many other horticultural and pathological characteristics of the scion cultivar and its fruits. With grafting, the trees mature uniformly and produce fruits earlier than those reproduced by seeds - typically within 3 years compared to around 6 years for seed-produced trees [35, 39]. In our study, we consider that the adequate match between scion and rootstock is already prescribed by the agronomists of the partner company, considering several studies carried out over the years by this company, the Fundação de Defesa Agrícola (FUNDECITROS) and the Luiz de Queiroz School of Agriculture (ESALQ) from the University of São Paulo.

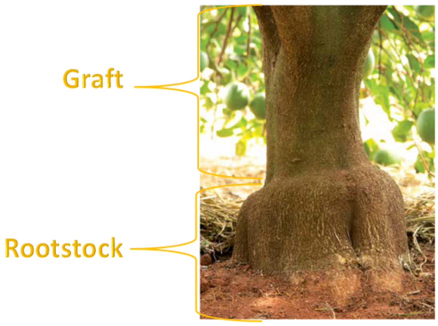

Figure 2: *Illustration of grafting using the T-budding method (adapted from [3]).*

3.3. Irrigation

Due to climatic conditions and long periods of drought in recent years, more than annual rainfall is required to meet all the needs of citrus plants in orchards in the Brazilian Citrus Belt. Drought may promote water stress to the plants, generating significant production breaks. In this context, irrigation becomes fundamental for developing leafy and productive orchards. Therefore, choosing the appropriate irrigation method is necessary to achieve the expected productivity standards in citrus groves [40]. Four irrigation methods are used in citriculture: surface, sprinkler, localized, and the absence of irrigation systems, using only rainwater, known as the rainfed system [40, 41]. According to the partner company's practice, we consider two irrigation systems in our study: localized irrigation and rainfed. The localized irrigation system is carried out by pumping rivers and artesian wells.

3.4. Density

Plots are characterized as portions of rural properties (smaller production units) intended for citrus fruit cultivation, separated by streets, roads and lanes [37]. The planting density refers to the number of plants in a given area, based on the spacing between plants in the same planting row and between planting rows. The spacing in the same planting row cannot be too short, as the shade of the treetop of a plant can prevent the sun's rays from reaching the adjacent plants, harming the photosynthesis process and, consequently, the development of the fruits of the adjacent plants. Another crucial point is that the very close spacing of plants in the same line can lead to competition for soil nutrients in that region – the roots of adjacent plants can consume nutrients intended for other plants, harming their development. The spacing between rows is due to the fertilizing, liming, and harvesting operations. The partner company classifies planting density of their plots according to the following three classes: Low density (hectare with less than 500 plants); Regular density (hectare with 501 to 800 plants) and High density (hectare with more than 801 plants).

3.5. Age balance

It may not be attractive to keep all the plants in the farms at the same age, as many would reach the end of their life cycle at the same time, and fruit production would drop dramatically. Hence, the partner company considers age balance levels for the plants, grouping them by age groups. Due to its significant relevance in productivity, age control is treated as an important requirement in the strategic planning process. The percentages of the age groups are established according to the absolute total number of plants in the entire organization to achieve an age balance with respect to the life cycle of the trees. An example of age groups is as follows: Group 0 (1 to 2 years) - 12%; Group 1 (3 to 5 years) - 12%; Group 2 (6 to 10 years) - 25%; and Group 3 (over 10 years) - 51%.

3.6. Productivity and eradication curves

In agribusiness, crops can be classified into perennial and annual. Perennial crops live for more than two years; therefore, it is possible to carry out several harvests throughout the life cycle of the plant without the need for replanting. Some examples of perennial crops include orange, rubber, oil palm, coconut, sago, coffee, tea, banana, etc. In contrast, annual crops, as the name suggests, have a life cycle of one year or season before harvest. Requiring annual replanting costs. Examples of annual crops include rice, wheat, soybean, corn, etc. However, a point worth noting is that, unlike annual crops, most perennial crops have variable yearly yields. This variation often depends on the age or maturity of culture [7].

The life cycle of perennial crop plants is not linearly associated with their production rate. In the first two years of the plant's life, its productivity is lower and not considered in strategic planning (it is not considered profitable). After that, the plants evolve and reach an intermediate stage in their life cycle where stable and economically viable production begins, which will grow along with the age of the plant until it reaches the highest productivity rate. The plant remains in this highest production stage for a few years. This level begins to decline in the last years of the plant's life cycle until it reaches an economically infeasible level [6, 27, 35].

The partner company relies on statistical methods to estimate productivity and eradication curves according to plant age and stratum. Recall that by stratum, we refer to the grouping of planting plots with the same characteristics regarding variety, rootstock, density, and irrigation. Productivity curves are commonly piecewise linear, with a positive slope from years 3 to 12, on average, and a negative slope after that. They estimate the number of boxes of oranges produced per area unit for a given stratum in a given farm for all the tree ages. Eradication curves are typically linear with positive slopes and estimate the rate of trees that are naturally eradicated throughout the years, according to the tree age. These curves are calculated every year, using specific regression models for each type of curve, and taken as input data in the strategic planning process. The description of these models is not within the scope of this paper, as we assume they are defined in an early decision stage and hence used as input parameters in our optimization model.

4. Mathematical Modeling

We propose an optimization model to aid decision-making in the strategic planning of planting in citriculture, following the characteristics and goals described in the previous section. It consists of a compact MIP formulation that, when solved by general-purpose MIP solvers, effectively provides solutions that recommend how to plant and eradicate at each year of the planning horizon, to maximize production while satisfying technical requirements.

4.1. Sets and parameters

Let $\mathcal{F}$ be the set of farms used for the orange plantation. For each farm $f \in \mathcal{F}$, we define its minimum and maximum planting area as A_f^{min} and A_f^{max}, respectively. Additionally, H_f is the fraction of the farm's area f with localized irrigation, a relevant feature for planting. Farms are grouped into *poles*, according to their proximity and resource sharing. We denote the set of farm poles as $\mathcal{L}$, and the subset of farms in a pole $\ell \in \mathcal{L}$ as $\mathcal{F}(\ell)$.

Recall that we use strata to characterize a plantation in farms according to the variety type, rootstock type, density, and irrigation. Let $\mathcal{E}$ be the set of all strata, which is partitioned into two subsets, namely a subset $\mathcal{E}^b$ of base strata, corresponding to the configuration of the plots at the beginning of the planning horizon (current plantations in the plots), and the subset $\mathcal{E}^n$ of strata that can be used in new plantings, according to technical recommendations. Some strata require localized irrigation, and thus we define $\mathcal{E}^{irri} \in \mathcal{E}$ as the subset of such strata. Additionally, there may be strata that are incompatible with certain farms due to technical reasons. Hence, set $\mathcal{K}(e)$ specifies the subset of farms compatible with stratum $e \in \mathcal{E}$.

Let $\mathcal{V}$ be the set of orange variety types. We denote by $\mathcal{E}(v) \subset \mathcal{E}$ the subset of strata related to oranges of variety type $v \in \mathcal{V}$. Additionally, related to the desired varietal balance levels described in Section 3.1, we define U_v^{min} and U_v^{max} as the minimum and maximum fraction of oranges of variety type $v \in \mathcal{V}$ in each pole at each time period, respectively.

We define $\mathcal{T}$ as the set of time periods in the planning horizon and $\mathcal{I}$ as the set of plant ages. Note that, even though both sets refer to time, $\mathcal{T}$ is used to count the years from the beginning of the planning horizon, while $\mathcal{I}$ is used to specify the ages of orange trees (according to the year they were planted). We assume that both start at 0, such that period 0 represents the beginning of the planning horizon and allows us to impose boundary conditions representing the current planting configuration of the farms. Similarly, age 0 corresponds to the first year of a new planting, as it is only at the end of this first year that the age turns 1. To simplify our notation, we define sets $\mathcal{T}_+ = \mathcal{T} \setminus \{0\}$ and $\mathcal{I}_+ = \mathcal{I} \setminus \{0\}$, which both start at 1. Moreover, we partition set $\mathcal{I}$ into age groups to enforce the plant age balance described in Section 3.5. We denote the set of age groups as $\mathcal{G}$. To represent the age balance levels, we define W_g^{min} and W_g^{max} as the desired minimum and maximum fraction of trees in group age $g \in \mathcal{G}$, respectively, at each period of the planning horizon. We use the notation $i \in g$ to mean that age $i \in \mathcal{I}$ belongs to group age $g \in \mathcal{G}$.

We also need to model the farm's current situation regarding what strata are planted, where they are planted, and what area they occupy on each farm. This is used as a boundary condition and defines the planting configuration at the beginning of the planning horizon. Hence, let parameter X_{eif}^0 represent the number of plants of stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ on farm $f \in \mathcal{F}$. Additionally, we define Y_{eif}^0 as the total area (in hectares) occupied by the plants specified by X_{eif}^0.

We also incorporate the productivity and eradication curves into the model described in Section 3.6. They are represented by parameters P_{eif} and R_{eif}, where P_{eif} represents the estimated number of orange boxes produced by plants of stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ on

Sets	
$\mathcal{F}$	Set of farms
$\mathcal{L}$	Set of farm poles
$\mathcal{E}$	Set of strata
$\mathcal{V}$	Set of orange varieties
$\mathcal{T}$	Set of periods
$\mathcal{T}_+$	Subset of periods greater than 0
$\mathcal{I}$	Set of plant ages
$\mathcal{I}_+$	Subset of plant ages greater than 0
$\mathcal{G}$	Set of plant age groups
$\mathcal{F}(\ell)$	Subset of farms belonging to pole $\ell \in \mathcal{L}$
$\mathcal{K}(e)$	Subset of farms compatible with stratum $e \in \mathcal{E}$
$\mathcal{E}(v)$	Subset of strata that produce the variety $v \in \mathcal{V}$
$\mathcal{E}^b$	Subset of strata that is currently planted at farms (base)
$\mathcal{E}^n$	Subset of strata that can be considered in new plantations
$\mathcal{E}^{irri}$	Subset of strata that require localized irrigation
Parameters	
A_f^{min}	Minimum planting area in farm $f \in \mathcal{F}$
A_f^{max}	Maximum planting area in farm $f \in \mathcal{F}$
H_f	Fraction of the area of farm $f \in \mathcal{F}$ with localized irrigation
U_v^{min}	Desired minimum production percentage of variety $v \in \mathcal{V}$ per pole
U_v^{max}	Desired maximum production percentage of variety $v \in \mathcal{V}$ per pole
W_g^{min}	Desired minimum percentage of plants in age group $g \in \mathcal{G}$
W_g^{max}	Desired maximum percentage of plants in age group $g \in \mathcal{G}$
X_{eif}^0	Number of trees of stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ in farm $f \in \mathcal{F}$ at the beginning of the time horizon
Y_{eif}^0	Area occupied by strata $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ in farm $f \in \mathcal{F}$ at the beginning of the time horizon
P_{eif}	Estimated production for stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ at farm $f \in \mathcal{F}$ (in orange boxes per tree)
R_{eif}	Estimated natural eradication rate for stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ at farm $f \in \mathcal{F}$
N_e	Maximum number of plants per hectare for stratum $e \in \mathcal{E}^n$
P^{min}	Minimum productivity per hectare for each stratum, used for eradication by productivity
I_e^{min}	Minimum age for eradicating stratum $e \in \mathcal{E}$ due to productivity
I_e^{max}	Maximum age allowed for plants in stratum $e \in \mathcal{E}$, used for eradication
M	Maximum number of seedlings planted per year
B	Maximum area allowed for eradication per year

Table 2: Sets and parameters for mathematical modeling.

farm $f \in \mathcal{F}$; and parameter R_{eif} is the estimated rate of natural eradication for plants in the same configuration.

The following parameters represent the technical requirements for planting. The maximum number of seedlings planted per year is defined by parameter M. The maximum number of seedlings per hectare when planting a stratum $e \in \mathcal{E}^n$ is represented as N_e. Parameter I_e^{max} defines the maximum age allowed for plants related to stratum $e \in \mathcal{E}$. After this age, we have to eradicate these plants. Eradication may also be motivated by low productivity, when the total production of a stratum per hectare falls below the threshold defined as P^{min}. Parameter I_e^{min} specifies the minimum age for the eradication decision, as no plant below this age can be eradicated due to low productivity. Moreover, the maximum area allowed for eradication per year is defined as B. Table 2 summarizes all sets and parameters defined in this section.

4.2. Decision variables

We model the decisions involved in the planning process using the following variables:

$x_{eift} \geq 0$	Estimated number of plants of stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ in farm $f \in \mathcal{F}$ in period $t \in \mathcal{T}$;
$y_{eift} \geq 0$	Area occupied by plants of stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ in farm $f \in \mathcal{F}$ in period $t \in \mathcal{T}$;
$z_{eift} \in \{0, 1\}$	1, if there is a plantation of stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ in farm $f \in \mathcal{F}$ in period $t \in \mathcal{T}$; 0, otherwise;
$\theta_{v\ell t} \geq 0$	Total production of orange variety $v \in \mathcal{V}$ at farm pole $l \in \mathcal{L}$ in period $t \in \mathcal{T}$ (in orange boxes).

Variables x_{eift} and y_{eift} define the estimated number of plants and area that should be occupied using stratum e with age i in farm f at period t, respectively. They are defined as continuous variables because their values result from estimations obtained by applying the productivity and eradication curves to the number of plants and area at the initial of the planning horizon or their respective planting year. This becomes clearer in the next subsection when we define the related constraints. The binary decision variable z_{eift} is related to x_{eift} and y_{eift} and indicates if these two variables are positive, i.e., if there are plants of stratum e with age i in farm f at period t. Hence, we should have $z_{eift} = 1$ if, and only if, $x_{eift} > 0$ and $y_{eift} > 0$. As we also clarify in the next subsection, this binary variable is required to ensure eradication when necessary. Finally, $\theta_{v\ell t}$ is an auxiliary variable that calculates the total production of the orange boxes of each variety, farm pole, and time period based on the number of plants specified by x_{eift} and the respective component of the productivity curve, P_{eif}. This variable defines the objective function and constraints related to varietal balance.

4.3. Objective Function and Constraints

The objective function (1) aims to maximize the total production over the planning horizon considering all varieties and farm poles.

$$\max \sum_{v \in \mathcal{V}} \sum_{\ell \in \mathcal{L}} \sum_{t \in \mathcal{T}} \theta_{v\ell t}. \tag{1}$$

Constraints (2) define that the variable $\theta_{v\ell t}$ is given by the total production of each variety $v \in \mathcal{V}$ at each pole $\ell \in \mathcal{L}$ and period $t \in \mathcal{T}$ according to the number of plants of each stratum

$e \in \mathcal{E}(v)$ compatible with the farms of that pole (as defined by $\mathcal{K}(e)$) and considering all plant ages $i \in \mathcal{I}$.

$$\theta_{v\ell t} = \sum_{e \in \mathcal{E}(v)} \sum_{i \in \mathcal{I}} \sum_{f \in \mathcal{F}(\ell) \cap \mathcal{K}(e)} P_{eif} x_{eift}, \qquad \forall v \in \mathcal{V}, \forall \ell \in \mathcal{L}, \forall t \in \mathcal{T}. \tag{2}$$

The varietal distribution balance described in Section 3.1 is ensured by constraints (3). The quantity produced of each variety $v \in \mathcal{V}$ in each pole $l \in \mathcal{L}$ and period $t \in \mathcal{T}$ must respect the desired minimum and maximum percentages U_v^{min} and U_v^{max} considering all orange varieties produced in that pole.

$$U_v^{min} \sum_{v' \in \mathcal{V}} \theta_{v'\ell t} \leq \theta_{v\ell t} \leq U_v^{max} \sum_{v' \in \mathcal{V}} \theta_{v'\ell t}, \qquad \forall v \in \mathcal{V}, \forall \ell \in \mathcal{L}, \forall t \in \mathcal{T}. \tag{3}$$

Constraints (4) and (5) ensure plant age balance as described in Section 3.6. In any period of the planning horizon, the total number of plants in a specific age group $g \in \mathcal{G}$, considering all strata, ages in that group, and farms must respect the desired minimum and maximum percentages W_g^{min} and W_g^{max} of the total number of plants (considering all strata, ages, and farms). Notice that these bounds consider the whole plantation, including all farms.

$$W_g^{min} \sum_{e \in \mathcal{E}} \sum_{i \in \mathcal{I}} \sum_{f \in \mathcal{K}(e)} x_{eifft} \leq \sum_{e \in \mathcal{E}} \sum_{i \in g} \sum_{f \in \mathcal{K}(e)} x_{eifft}, \qquad \forall g \in \mathcal{G}, \forall t \in \mathcal{T}, \tag{4}$$

$$\sum_{e \in \mathcal{E}} \sum_{i \in g} \sum_{f \in \mathcal{K}(e)} x_{eift} \leq W_g^{max} \sum_{e \in \mathcal{E}} \sum_{i \in \mathcal{I}} \sum_{f \in \mathcal{K}(e)} x_{eifft}, \qquad \forall g \in \mathcal{G}, \forall t \in \mathcal{T}. \tag{5}$$

Constraints (6) and (7) enforce the boundary conditions regarding the current planting configuration of farms. Constraints (6) set variable x_{eif0} as the number of plants in stratum $e \in \mathcal{E}$ with age $i \in \mathcal{I}$ in farm $f \in \mathcal{F}$ at the start of the planning horizon (i.e., X^0_{eif}). Similarly, constraints (7) set y_{eif0} equal to the area corresponding to X^0_{eif}, given by Y^0_{eif}.

$$x_{eif0} = X^0_{eif}, \qquad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \in \mathcal{K}(e), \tag{6}$$

$$y_{eif0} = Y^0_{eif}, \qquad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \in \mathcal{K}(e). \tag{7}$$

Constraints (8) forbid the planting of new seedlings (i.e., plants of age $i = 0$) of strata belonging to the base group $\mathcal{E}^b$ in any period $t \geq 1$. These strata are only used to set the boundary conditions related to the current planting configuration of the farms. Only strata in $\mathcal{E}^n$ can be used in new plantations.

$$x_{e0ft} = 0, \qquad \forall e \in \mathcal{E}^b, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}_+. \tag{8}$$

Constraints (9) impose a limit on the annual planting of new seedlings due to operational and financial limitations in crop planning.

$$\sum_{e \in \mathcal{E}} \sum_{f \in \mathcal{K}(e)} x_{e0ft} \leq M, \quad \forall t \in \mathcal{T}_+. \tag{9}$$

Constraints (10)-(12) model the relationship between the variables x_{eift} and y_{eift} using the density of strata in new plantations and the current configuration of the farms. Constraints (10) ensure that new plantations at any period t follow the technical recommendation of density, imposing that the total number of trees (x_{e0ft}) is equal to the number of trees per hectare (N_e) times the total area (y_{e0ft}) used for plantation of stratum e in farm f. In the subsequent years (with age $i \geq 1$), the number of trees may be reduced due to natural eradication. However, the area initially used for planting remains the same, resulting in constraints (11). Constraints (12) ensure a similar logic, but for base strata only, as their density is given by $X^0_{e(i-t)f}/Y^0_{e(i-t)f}$, defined by the boundary conditions that represent the configuration of the farms at the beginning of the planning horizon. Note that we round up this fraction to prevent numerical instability in the solver.

$$
\begin{aligned}
x_{e0ft} &= N_e y_{e0ft}, && \forall e \in \mathcal{E}^n, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}, && (10)\\
x_{eift} &\leq N_e y_{eift}, && \forall e \in \mathcal{E}^n, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}, && (11)\\
x_{eift} &\leq \lceil X^0_{e(i-t)f}/Y^0_{e(i-t)f} \rceil y_{eift}, && \forall e \in \mathcal{E}^b, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T} : t \leq i, && \\
& && Y^0_{e(i-t)f} > 0. && (12)
\end{aligned}
$$

Constraints (13)-(15) relate variables y_{eift} and z_{eift} and impose limits on the planting areas at each farm. Constraints (13) enforce that the area used in a farm f for plants of a stratum e with age i at time period t cannot be less than the minimum plot size A^{min}_f if there is a plantation with this configuration of stratum (i.e., $z_{eift} > 0$). Additionally, they ensure that this area cannot be larger than A^{max}_f. Constraints (14) impose the maximum area utilization limit on each farm, considering all strata with all ages. Recall that H_f is the fraction of the area of farm f with localized irrigation and $\mathcal{E}^{irri}$ is the subset of strata that require localized irrigation. Thus, constraints (15) ensure that the fraction of area with plants that require localized irrigation is not larger than the total available in the farm.

$$
\begin{aligned}
A^{min}_f z_{eift} \leq y_{eift} &\leq A^{max}_f z_{eift}, && \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}, && (13)\\
\sum_{i \in \mathcal{I}} \sum_{e \in \mathcal{E}} y_{eift} &\leq A^{max}_f, && \forall f \in \mathcal{F}, \forall t \in \mathcal{T}_+, && (14)\\
\sum_{i \in \mathcal{I}} \sum_{e \in \mathcal{E} \cap \mathcal{E}^{irri}} y_{eift} &\leq H_f A^{max}_f, && \forall f \in \mathcal{F}, \forall t \in \mathcal{T}_+. && (15)
\end{aligned}
$$

Constraints (16) ensure consistency between variables x_{eift} and z_{eift}. If there is a plantation of stratum e with age i in farm f at time period t ($z_{eift} = 1$), at least one tree has to be in this plantation. Notably, the opposite relationship is guaranteed transitively through constraints (11)-(13).

$$
z_{eift} \leq x_{eift}, \quad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \in \mathcal{F}, \forall t \in \mathcal{T}. \qquad (16)
$$

Constraints (17)–(19) model the evolution of the number of planted trees according to their age. They consider the number of trees in the initial plantation and the natural eradication of trees according to the eradication curve values represented by R_{eif}. Constraints (17) enforce that the number of plants of stratum e with age i in farm f in period t cannot be larger than the number of plants in the same stratum and farm in the previous period $t-1$ (when the plants have age $i-1$) reduced by a factor R_{eif}. Constraints (18)-(19) ensure that the reduction in the number of plants is not larger than stipulated by R_{eif}, imposing thus a lower bound for x_{eift}. Notice that these constraints become inactive if $z_{eift} = 0$, i.e., if there are no plants of stratum e with age i in farm f. Moreover, constraints (18)-(19) work similarly and only differ by the type of stratum considered on them (either those in $\mathcal{E}_b$ or in $\mathcal{E}_n$). This is because the maximum value of x_{eift} used to inactivate these constraints, given by either $X^0_{e(i-t)f}$ or $N_e A^{max}_f$, depends on the stratum type. Note that we round up these values to reduce numerical instability when solving the model. Finally, it is worth mentioning that one may wonder why these three sets of constraints are not modeled using a single set of equality constraints in which x_{eift} is equal to $(1 - R_{eif})x_{e(i-1)f(t-1)}$. The reader should bear in mind, though, that total eradication may be required in a given period (due to reasons such as low productivity or advanced plant age), requiring $x_{eift} = 0$ even if we may have $x_{e(i-1)f(t-1)} > 0$.

$$
\begin{aligned}
x_{eift} &\leq (1 - R_{eif})x_{e(i-1)f(t-1)}, && \forall e \in \mathcal{E}, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \\
& && \forall t \in \mathcal{T}_+, && (17) \\
x_{eift} &\geq (1 - R_{eif})x_{e(i-1)f(t-1)} - \lceil X^0_{e(i-t)f} \rceil (1 - z_{eift}), && \forall e \in \mathcal{E}, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \\
& && \forall t \in \mathcal{T}_+ : i \geq t, X^0_{e(i-t)f} > 0, && (18) \\
x_{eift} &\geq (1 - R_{eif})x_{e(i-1)f(t-1)} - \lceil N_e A^{max}_f \rceil (1 - z_{eift}), && \forall e \in \mathcal{E}^n, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \\
& && \forall t \in \mathcal{T}_+. && (19)
\end{aligned}
$$

Constraints (20) and (21) work similarly to (17)–(19), but for the area variables y_{eift}. They ensure the continuity of using the area initially allocated to a given stratum along the planning horizon. Apart from total eradication (e.g., due to advanced age or low productivity), they would be defined to ensure that $y_{eift} = y_{e(i-1)f(t-1)}$. However, since we need to consider eradication in our planning, we must represent this equality through constraints (20) and (21). Constraints (20) ensure that the area occupied by plants of a given stratum e with age $i \geq 1$ in farm f at period $t \geq 1$ (y_{eift}) is not larger than in period $t-1$ ($y_{e(i-1)f(t-1)}$). Together with constraints (21), they guarantee that this area will remain the same size (as these constraints impose $y_{e(i-1)f(t-1)}$ as a lower bound) unless the stratum is eradicated (i.e., $z_{eift} = 0$) and the corresponding constraints become inactive. Hence, constraints (21) are active only when $z_{eift} = 1$.

$$
\begin{aligned}
y_{eift} &\leq y_{e(i-1)f(t-1)}, && \forall e \in \mathcal{E}, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}_+, && (20) \\
y_{eift} &\geq y_{e(i-1)f(t-1)} - A^{max}_f (1 - z_{eift}), && \forall e \in \mathcal{E}, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}_+. && (21)
\end{aligned}
$$

Recall that the total eradication area per year, considering all farms, cannot exceed the value defined by parameter B. Constraints (22) ensure this requirement by taking the differences of occupied areas between two consecutive periods.

$$
\sum_{e \in \mathcal{E}} \sum_{i \in \mathcal{I}_+} \sum_{f \in \mathcal{K}(e)} y_{e(i-1)f(t-1)} - y_{eift} \leq B, \qquad \forall t \in \mathcal{T}_+. \qquad (22)
$$

Eradication must also respect the minimum and maximum ages of plants (I^{min}_e and I^{max}_e) and has to be applied when the productivity of plants per hectare falls below the threshold P^{min}. Constraints (23) ensure that the plants of stratum e with age below I^{min}_e are never eradicated. Constraints (24) enforce eradication by age, as no plant of stratum e can be older than I^{max}_e. Constraints (25) impose eradication by productivity, ensuring that the estimated production of orange boxes ($P_{eif}x_{eift}$) is never below the threshold considering the respective planting area ($P^{min}y_{eift}$).

$$
\begin{aligned}
z_{eift} &= z_{e(i-1)f(t-1)}, && \forall e \in \mathcal{E}, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}_+ : i < I^{min}_e, && (23) \\
z_{eift} &= 0, && \forall e \in \mathcal{E}, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}_+ : i > I^{max}_e, && (24)
\end{aligned}
$$

$$P_{eif}x_{eift} \geq P^{min}y_{eift}, \quad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}_+, \forall f \in \mathcal{K}(e), \forall t \in \mathcal{T}_+ : i \geq I_e^{min}. \tag{25}$$

Recall that only farms in the subset $\mathcal{K}(e)$ are compatible with a stratum $e \in \mathcal{E}$. Hence, constraints (26)-(28) prohibit the planting of strata that are not compatible with farms.

$$x_{eift} = 0, \quad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \notin \mathcal{K}(e), \forall t \in \mathcal{T}, \tag{26}$$

$$y_{eift} = 0, \quad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \notin \mathcal{K}(e), \forall t \in \mathcal{T}, \tag{27}$$

$$z_{eift} = 0, \quad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \notin \mathcal{K}(e), \forall t \in \mathcal{T}. \tag{28}$$

Finally, constraints (29)-(32) impose the domain of the decision variables.

$$\theta_{v\ell t} \geq 0, \quad \forall v \in \mathcal{V}, \forall \ell \in \mathcal{L}, \forall t \in \mathcal{T}, \tag{29}$$

$$x_{eift} \geq 0, \quad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \in \mathcal{F}, \forall t \in \mathcal{T}, \tag{30}$$

$$y_{eift} \geq 0, \quad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \in \mathcal{F}, \forall t \in \mathcal{T}, \tag{31}$$

$$z_{eift} \in \{0, 1\}, \quad \forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall f \in \mathcal{F}, \forall t \in \mathcal{T}. \tag{32}$$

These constraints conclude the definition of the proposed model to support decision-making in the addressed strategic planning process for planting in citriculture. Model (1)-(32) is hereafter referred to as SPMC, short for *strategic planning model for citriculture*. Note that the proposed model implicitly assumes plot aggregation as it focuses on strategic planning. As mentioned, a plot is a production unit corresponding to a given farm area, delimited by streets, roads, or other means. As the number of plots in the real case that inspired the model is very high (more than 3000 in total), it was observed that the division of farms by plots would be computationally intractable. Furthermore, this level of refinement is beyond the scope of strategic planning, becoming relevant only in tactical or operational-level planning. Thus, plots are implicitly represented in the model by strata in an aggregated way. This simplification was validated by the partner company's team, who also resorted to strata to define productivity and eradication curves. Hence, in the model solution, each decision involving a stratum with a certain age in a given farm and given period represents a planting area that can cover several plots of this farm, all of the same age and plant characteristics. These decisions can be refined later in the tactical and operational planning processes.

4.4. SPMC with soft varietal and age balance

Preliminary computational results using instances based on the data provided by the partner company have shown that the current planting configuration of the farms does not satisfy the varietal and plant age balance levels desired by the company. As mentioned, satisfying these balance levels is challenging because of the many complex decisions and requirements. One of the main purposes behind using an optimization approach such as the one described in this paper was to effectively reduce the violation of these balance levels.

Given that the current planting configuration of the farms is used as input in the SPMC by defining the boundary conditions in constraints (6) and (7), all tested instances were infeasible due to constraints (3)-(5). To overcome this, we developed two strategies based on the use of soft constraints. This approach involves relaxing a subset of constraints that impose goals that are difficult or impossible to achieve, and then penalizing their violation in the objective function to promote their satisfaction as much as possible [42]. Hence, we convert constraints (3)-(5) into soft constraints by introducing artificial variables to them, which are penalized in the objective function. This modeling technique relaxes these constraints at the cost of worsening the objective value when it is impossible to satisfy them, bringing more flexibility to the SPMC, as the satisfaction with balance levels is now not obligatory but strongly enforced by the objective function.

The penalties imposed for violations are weights that represent the relative importance of each soft constraint with respect to the others and the original objective function value. Hence, soft constraints are treated in a weighted goal programming fashion within the objective function [42]. The aim is to obtain solutions that strike a good balance between optimality (concerning the original objective function) and feasibility (with respect to the soft constraints), using a relative simple yet flexible approach with a single objective function [36, 42, 43]. This strategy circumvents the challenges often associated with more sophisticated multi-objective optimization approaches, which may require multiple executions of the model to build Pareto frontiers [44]. Therefore, although the use of soft constraints do not suit every multi-objective situation, in our case, it proves to be an effective and suitable tool for decision-making [42, 45].

We propose two different strategies for converting constraints (3)-(5) into soft, called S1 and S2. In strategy S1, the SPMC is modified by introducing artificial variables related to each single varietal and age balance constraint. Specificaly, we define the non-negative continuous variables $u_{v\ell t}^{min}$ and $u_{v\ell t}^{max}$ for each variety $v \in \mathcal{V}$, pole $\ell \in \mathcal{L}$ and period $t \in \mathcal{T}$; and w_{gt}^{min} and w_{gt}^{max} for each age group $g \in \mathcal{G}$ and period $t \in \mathcal{T}$. Then, constraints (3)-(5) are replaced with:

$$U_v^{min} \sum_{v' \in \mathcal{V}} \theta_{v'\ell t} - u_{v\ell t}^{min} \leq \theta_{v\ell t} \leq U_v^{max} \sum_{v' \in \mathcal{V}} \theta_{v'\ell t} + u_{v\ell t}^{max}, \quad \forall v \in \mathcal{V}, \forall \ell \in \mathcal{L}, \forall t \in \mathcal{T}, \tag{33}$$

$$W_g^{min} \sum_{e \in \mathcal{E}} \sum_{i \in \mathcal{I}} \sum_{f \in \mathcal{K}(e)} x_{eift} - w_{gt}^{min} \leq \sum_{e \in \mathcal{E}} \sum_{i \in g} \sum_{f \in \mathcal{K}(e)} x_{eift}, \quad \forall g \in \mathcal{G}, \forall t \in \mathcal{T}, \tag{34}$$

$$\sum_{e \in \mathcal{E}} \sum_{i \in g} \sum_{f \in \mathcal{K}(e)} x_{eift} \leq W_g^{max} \sum_{e \in \mathcal{E}} \sum_{i \in \mathcal{I}} \sum_{f \in \mathcal{K}(e)} x_{eift} + w_{gt}^{max}, \quad \forall g \in \mathcal{G}, \forall t \in \mathcal{T}. \tag{35}$$

$$u_{v\ell t}^{min}, u_{v\ell t}^{max} \geq 0, \quad \forall v \in \mathcal{V}, \forall \ell \in \mathcal{L}, \forall t \in \mathcal{T}, \tag{36}$$

$$w_{gt}^{min}, w_{gt}^{max} \geq 0, \quad \forall g \in \mathcal{G}, \forall t \in \mathcal{T}. \tag{37}$$

Notably, the artificial variables represent the violations in their respective constraints. They are also inserted into the objective function, multiplied by the positive coefficients ϕ_t^{var} and ϕ_t^{age}, which penalize the violation of varietal and age balance constraints, respectively. Thus, we replace the objective function (1) with:

$$\max \sum_{v \in \mathcal{V}} \sum_{\ell \in \mathcal{L}} \sum_{t \in \mathcal{T}} \theta_{v\ell t} - \sum_{t \in \mathcal{T}} \phi_t^{var} \left(\sum_{v \in \mathcal{V}} \sum_{\ell \in \mathcal{L}} u_{v\ell t}^{min} + u_{v\ell t}^{max} \right) - \sum_{t \in \mathcal{T}} \phi_t^{age} \left(\sum_{g \in \mathcal{G}} w_{gt}^{min} + w_{gt}^{max} \right). \tag{38}$$

Strategy S2 is similar to S1, but it aggregates the violation of several constraints in the artificial variables. We define the non-negative continuous variables $u_v^{min} \geq 0$ and $u_v^{max} \geq 0$, for each variety $v \in \mathcal{V}$; and $w_g^{min} \geq 0$ and $w_g^{max} \geq 0$, for each group $g \in \mathcal{G}$. Constraints (3)-(5) are then replaced with:

$$U_v^{min} \sum_{v' \in \mathcal{V}} \theta_{v'\ell t} - u_v^{min} \leq \theta_{v\ell t} \leq U_v^{max} \sum_{v' \in \mathcal{V}} \theta_{v'\ell t} + u_v^{max}, \quad \forall v \in \mathcal{V}, \forall \ell \in \mathcal{L}, \forall t \in \mathcal{T}, \tag{39}$$

$$W_g^{min} \sum_{e \in \mathcal{E}} \sum_{i \in \mathcal{I}} \sum_{f \in \mathcal{K}(e)} x_{eift} - w_g^{min} \leq \sum_{e \in \mathcal{E}} \sum_{i \in g} \sum_{f \in \mathcal{K}(e)} x_{eift}, \quad \forall g \in \mathcal{G}, \forall t \in \mathcal{T}, \tag{40}$$

$$\sum_{e\in\mathcal{E}}\sum_{i\in g}\sum_{f\in\mathcal{K}(e)} x_{eift} \leq W_g^{max}\sum_{e\in\mathcal{E}}\sum_{i\in\mathcal{I}}\sum_{f\in\mathcal{K}(e)} x_{eift} + w_g^{max}, \quad \forall g\in\mathcal{G}, \forall t\in\mathcal{T}. \tag{41}$$

$$u_v^{min}, u_v^{max} \geq 0, \qquad \forall v\in\mathcal{V}, \tag{42}$$

$$w_g^{min}, w_g^{max} \geq 0, \qquad \forall g\in\mathcal{G}. \tag{43}$$

Hence, each artificial variable represents the total violation over several constraints, grouped by the variety v or age group g. Similarly to S1, we define the penalties ϕ^{age} and ϕ^{var}, which are used to modify the objective function as follows:

$$\max \sum_{v\in\mathcal{V}}\sum_{\ell\in\mathcal{L}}\sum_{t\in\mathcal{T}} \theta_{v\ell t} - \phi^{var}\left(\sum_{v\in\mathcal{V}} u_v^{min} + u_v^{max}\right) - \phi^{age}\left(\sum_{g\in\mathcal{G}} w_g^{min} + w_g^{max}\right). \tag{44}$$

5. Computational Results

We present the results of computational experiments using a realistic instance of the problem, designed to verify the effectiveness and performance of the proposed optimization approach. The SPMC and its variants with soft strategies S1 and S2 were implemented in the language Python, version 3.10, using the libraries Pyomo, Tkinter, Pandas, Numpy, Datetime, and Io. To solve the model, we used the general-purpose MIP solver Gurobi version 9.5.2, with default settings, relative gap tolerance equal to 10^{-4}, and imposing a time limit of 3600 seconds per call. All experiments were executed in a computer with Intel Xeon E5-2680 @ 2.70 GHz x 32 processors, 192 GB of RAM, and Linux Mint operating systems. Detailed computational experiments are presented in Appendix E and summarized in the remainder of this section.

5.1. Data description

We created a realistic problem instance based on the historical data provided by the partner company as well as on their experience in the decisions involved in the model. The data refers to a snapshot of their planting situation at the beginning of the year. All parameters required in the SPMC were set based on this information. Additionally, they were validated based on the company's feedback. The provided data includes around 30 farms grouped into 10 poles and involves around 300 different strata with the full description of the current plantation of the company. There is also an expansion area of around 10,000 hectares, which the company intends to occupy based on the recommendations of this strategic planning process. The desired minimum (U_v^{min}) and maximum (U_v^{max}) production percentages of each variety were defined by specific values in the intervals $[5\%, 25\%]$ and $[15\%, 35\%]$, respectively. The company uses five age groups, hereafter named as G1 to G5. The desired minimum (W_g^{min}) and maximum (W_g^{max}) percentage of plants in each age group assume specific values in the intervals $[15\%, 20\%]$ and $[20\%, 25\%]$, respectively. Each group is defined using levels belonging to these intervals. To protect data confidentiality, we cannot present the mentioned numbers exactly, as requested by the company. All parameter values were defined according to the data provided by the company, but the results we present in this section are multiplied by a given positive scalar and presented as percentage deviations to cope with data privacy.

5.2. Model variants and choice of penalties

To analyze the impact of the varietal and age balance requirements and the different combinations of incorporating soft strategies S1 and S2 into the SPMC, our experiments consider the following SPMC variants:

- No-No: both constraints that impose varietal and age balance are inactive (i.e., they are removed from the SPMC);
- S1-No: requirements related to varietal balance are imposed using the S1 strategy, while the constraints related to age balance are inactive;
- S2-No: requirements related to varietal balance are imposed using the S2 strategy, while the constraints related to age balance are inactive;
- No-S1: inactive varietal balance and active age balance using the S1 strategy;
- No-S2: inactive varietal balance and active age balance using the S2 strategy;
- S1-S1: both varietal and age balance are imposed using the S1 strategy; and
- S2-S2: both varietal and age balance are imposed using the S2 strategy.

It is worth mentioning that experiments with the SPMC in which both varietal and age balance requirements were active as hard constraints resulted in the model's infeasibility because the company's current plantation did not satisfy these constraints. Moreover, model variants No-No, No-S1, No-S2, S1-No, and S2-No are analyzed only to demonstrate the impact of activating the constraints imposed by age and varietal balance.

As mentioned before, the S1 and S2 strategies bring more flexibility to the SPMC. However, no straightforward way exists to define appropriate values for the parameters ϕ_t^{var} and ϕ_t^{age} that penalize the varietal and age balance violations in the objective function. Hence, after running extensive computational experiments with different choices of these parameters in model variants S1-S1 and S2-S2, and based on the feedback of the partner company regarding the results of each particular choice, we found that the best-performing values for ϕ_t^{var} and ϕ_t^{age} in strategy S1 are as follows:

$$\phi_t^{var} = \begin{cases} 0.0, & \text{if } t = 0, \\ 0.1, & \text{if } 1 \leq t \leq \lceil 0.25|\mathcal{T}| \rceil, \\ 1.0, & \text{if } \lceil 0.25|\mathcal{T}| \rceil + 1 \leq t \leq \lceil 0.50|\mathcal{T}| \rceil, \\ 10.0, & \text{otherwise.} \end{cases} \tag{45}$$

$$\phi_t^{age} = \begin{cases} 0.0, & \text{if } t = 0, \\ 0.1, & \text{if } 1 \leq t \leq \lceil 0.25|\mathcal{T}| \rceil, \\ 1.0, & \text{if } \lceil 0.25|\mathcal{T}| \rceil + 1 \leq t \leq \lceil 0.75|\mathcal{T}| \rceil, \\ 10.0, & \text{otherwise.} \end{cases} \tag{46}$$

for $t \in \mathcal{T}$. For variant S2, the best configuration identified for penalty parameters ϕ^{var} and ϕ^{age} are the following:

$$\phi^{var} = \phi^{age} = 1. \tag{47}$$

Note that (45) and (46) define the penalties progressively according to the period. The reason for this is that the solution is strongly influenced by the boundary conditions related to the current planting configuration in the early periods. Hence, it gives more flexibility to the varietal and age balance requirements in the early periods, and gradually reduces this flexibility in the later periods by increasing the penalty for violations. For example, considering a planning horizon of $T = 30$ years, the proposed definition of ϕ_t^{var} uses the relatively low penalty value of 0.1 in years 1 through 8 ($= \lceil 0.25T \rceil$), corresponding to the first 25% of the time horizon. This choice is to avoid significant production loss in the first years, since the planting configuration is mostly determined by the boundary conditions that do not satisfy the varietal balance requirements. In the next 25% of the time horizon, covering years 9 to 15 ($\lceil 0.25T \rceil + 1$ to $\lceil 0.50T \rceil$), the presence of plantations imposed by the boundary conditions is reduced by eradication, and thus we increase the penalty value to 1.0, enforcing a stricter requirement for the varietal balance in these periods (which has influence on the decisions related to new plantations in the early periods). Finally, for the second half of the time horizon, it is highly recommended that the varietal balance be within their required percentages. Hence, we increase the penalty to 10.0 in these periods. The company's team validated the results obtained with these parameter choices as effective and in line with their goals.

5.3. Production Analysis

Figure 3 presents a chart showing the (scaled) total production in orange boxes at each year of the planning horizon, according to the solutions of the different SPMC variants. Each curve in the chart corresponds to one of the SPMC variants defined in the previous subsection. The x-axis represents the periods (years) in the time horizon, and the y-axis gives the orange production scaled by the production in period $t = 0$, to protect the data provided by our partner company. Hence, in period $t = 0$ the scaled production is 1, and, for example, in period $t = 10$ the orange production in the solution of model variant No-No is almost 60% greater than the production in period $t = 0$, as the scaled production is close to 1.6. Details on computation times and solution gaps for each model variant are presented in Appendix E.

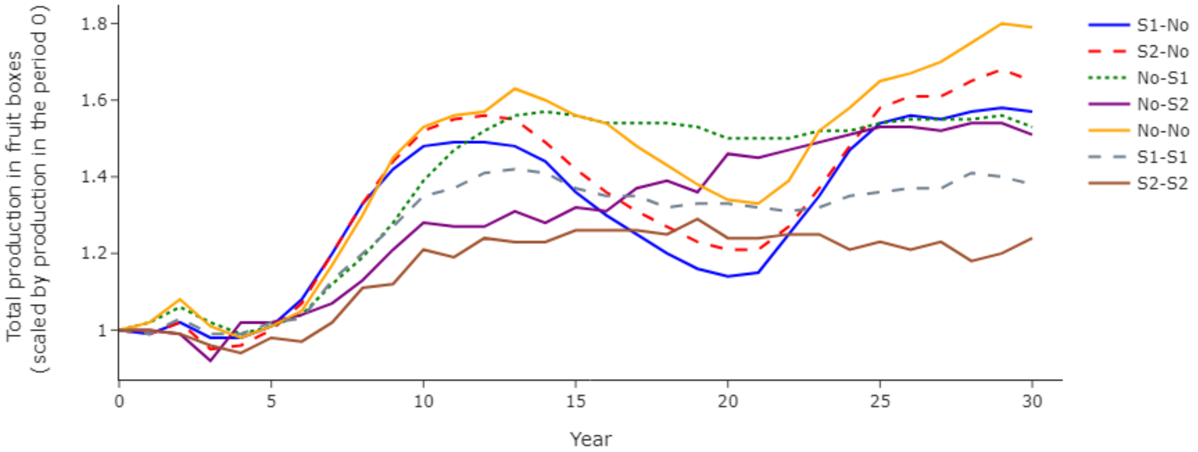

Figure 3: *Total production in orange boxes in the solutions of the SPMC variants.*

Notably, in all SPMC variants, the production increased significantly until the 10th period. This increase was not monotonic, as we observed a slight reduction in production around the 3rd period, corresponding to eradication due to advanced age and low productivity. From periods 5 to 10, the increase in orange production accentuates in the solutions of all variants, as the new strata planted in the initial years of the planning horizon start to be productive (new plantations require at least three years until they reach a significant productivity level). In some variants, such as No-S1, S1-S1, and S2-S2, the production stabilizes after period 15. However, in variants S1-No, S2-No and No-No, the production drops significantly after period 13, indicating that age balance control plays an important role in keeping the production stable, in addition to bringing several other benefits to recourse management, as mentioned before.

The best result regarding total production is observed for the model variant No-No, as expected. However, it does not lead to an attractive solution in practice as it disregards the desired varietal and age control. As mentioned before, we only included this variant in the experiments to see what would be the largest total production (considering the summation over all periods) when the varietal and age balance requirements are not imposed. A negative aspect of the solution provided by this variant is the oscillation in production throughout the years, with two peaks of production around periods 13 and 29 and a valley around period 20. According to the partner company, this oscillatory behavior should be avoided in practice since it unbalances the use of resources from one year to another. We see similar behavior in the results of variants S2-No and S1-No, which are the second and third largest total productions. These observations indicate that ignoring the age balance requirements may increase the total production over the time horizon but result in large oscillations in production, which negatively affect the management of resources in practice.

The model variants considering both age and varietal control requirements simultaneously, namely S1-S1 and S2-S2, result in solutions that promote stability in the production amount from period 13 onward. Even though their corresponding total productions were not as high as in the other variants, they still resulted in a significant increase with respect to the production at the beginning of the time horizon, without deviating too much from varietal and age balance levels, as presented in the following subsections. These were the main features sought by the company, as they wanted to maximize their overall production while having a stable production amount over the years and satisfying varietal and age balance requirements as much as possible.

5.4. Varietal Balance Analysis

For the varietal balance analysis, we focus on observing the production percentages according to a variety of types, considering the desired levels specified by the company. Fig. 4 presents the production percentages of each variety at each period in the solutions obtained by the variants No-No, S1-S1, and S2-S2, considering one of the poles of the partner company. The charts for the other variants are presented in Appendix A. As the varietal balance is enforced by pole, and we have 10 poles in total, we arbitrarily selected one representative pole and showed the analysis for this pole only. The other nine poles show similar behavior as in the one analyzed here. Additionally, the dashed horizontal lines in the charts show the minimum and maximum values (5 and 35%, respectively) that are used in the definition of the desired varietal balance. For each variety, the company defines specific values inside this interval, but we are not allowed to reveal them.

The results presented in the charts show that the varietal balance requirements tend to be significantly violated if not properly accounted for in the model. We observe that in the solution obtained with the No-No variant, one of the varieties reaches almost 100%, given its high productivity, while others are not even produced in the pole. Conversely, variant S1-S1 was very effective in providing a solution that satisfies the desired varietal balance levels. After period 15, the production quantities of all variety types satisfy the desired levels and remain stable until the end of the planning horizon. Note that the variant S2-S2 was not as successful in ensuring the varietal balance requirements. Hence, the penalization strategy S1 was the most effective for the varietal balance.

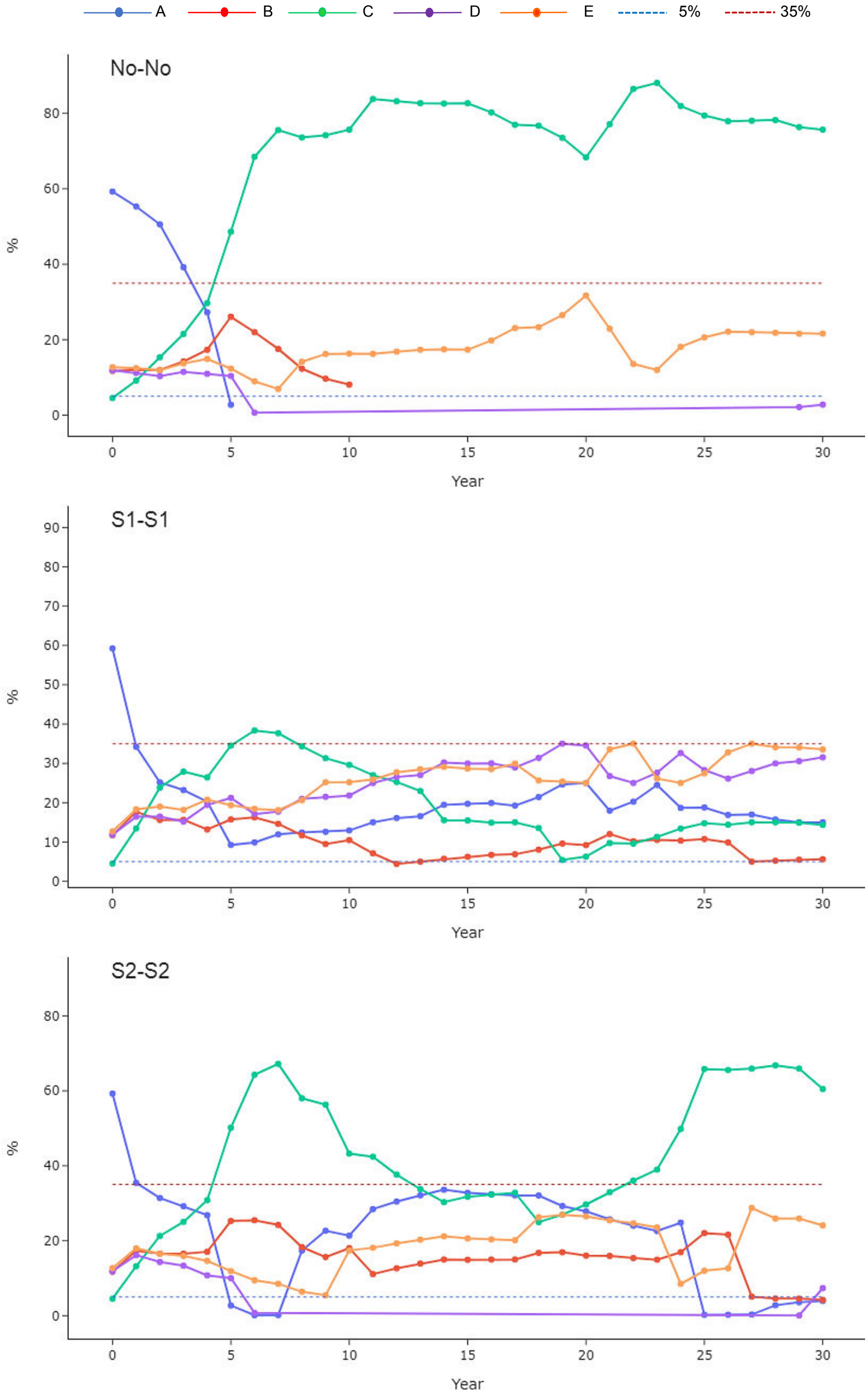

Figure 4: *Production percentages of each variety at each period for a given pole, in the solutions obtained with the model variants No-No, S1-S1, and S2-S2.*

5.5. Age Balance Analysis

We now analyze the age balance levels of the solutions provided by the model variants. As mentioned before, satisfying the desired minimum and maximum levels is a tough requirement in practice. The company's plantation at the beginning of the planning horizon, which was used as the boundary condition in the model, cannot satisfy these levels throughout the horizon. Additionally, new plantations take years to reach maximum productivity. Hence, finding an appropriate combination of plantation and eradication that promotes the age balance without significantly reducing production can be extremely challenging.

Figure 5 presents charts with the percentage of plants in age groups G1 to G5 at each period of the time horizon in the solutions provided by model variants No-No, S1-S1, and S2-S2. The charts for the other variants are presented in Appendix B. The dashed horizontal lines in the charts correspond to the minimum and maximum values for desired percentages (15 and 25%, respectively), but different specific values may be imposed by the company for each group within these two levels. In these charts, we observe large violations of these percentages at the first time periods, as during them, the age balance requirements are relaxed or lightly penalized in the objective function. For variant No-No, these violations are observed throughout the whole time horizon, while we see that S1-S1 and S2-S2 could better control these violations after the middle of the time horizon. In particular, S1-S1 was very effective in satisfying the age balance levels, reducing the violation of these requirements to zero from period 18 onward. Variant S2-S2 was not so effective, as we observe that the percentage of certain age groups above or below the desired percentages in all periods of the time horizon. Therefore, strategy S1 is the most appropriate for imposing the minimum and maximum age balance requirements, as it resulted in a solution that reaches the levels desired by the company.

5.6. Planting and Eradication Analysis

In this analysis, we focus on managerial insights regarding the influence of varietal and age balance on planting and eradication. Planting requires high investments in machinery, agricultural inputs, workforce, and seeds. For this reason, the annual planting is limited to 4 million plant seedlings. Likewise, eradication is limited to 5.000 ha/year. Note that the measuring unit used in the planting effort is the number of plant seedlings, while the unit in the eradication effort is an area in hectares.

Figures 6 and 7 show the total amount of new plantations and eradication, respectively, according to the solutions obtained with the SPMC variants No-No, S1-S1, S2-S2 (see Appendix C and D for the other variants). In these results, we observe that the solutions of all model variants present similar behavior up to period 5 because of the system's initial state before the optimization process. Moreover, planting is directly related to the value of the objective function through the constraints (2). Thus, SPMC prioritizes planting the most productive strata, maximizing the productivity required in the objective function (1). Furthermore, the results indicate that variant S1-S1 achieves the best result, as this is the only one that promotes a stabilized planting (after period 5), keeping the amount of planting seedlings around 2 million every year. We observe that this variant, is the only one resulting in a stabilized eradication level too (from period 13 onward). Again, these results were evaluated as extremely positive by the company, because this solution allows them to effectively manage the resources related to planting and eradication, including a better cash flow management to invest in the annual harvest.

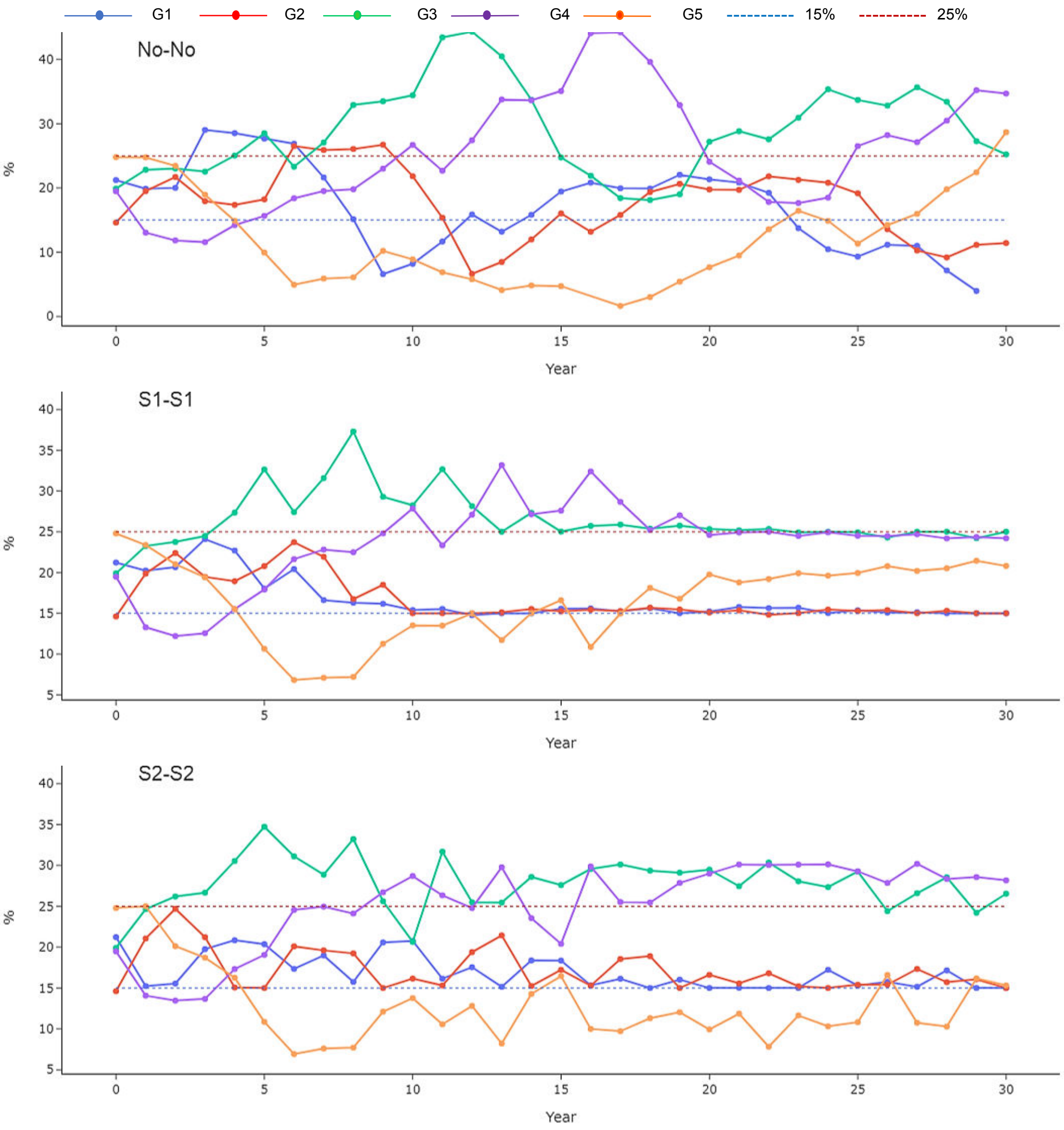

Figure 5: *Percentages of plants in the five age groups at each time period in the solutions obtained with the model variants No-No, S1-S1, and S2-S2.*

6. Conclusion

We have proposed an effective optimization approach to support strategic planning decisions in citrus planting. It was motivated by the partnership with one of the largest orange juice producers in the world. This company was interested in optimizing its fruit production while satisfying important requirements such as varietal and age balance, maximum plantation, and eradication efforts, among others. To support this complex decision-making process using formal and scientific techniques, we introduced a mixed-integer linear optimization model, named the strategic planning model for citriculture (SPMC), considering the main characteristics, requirements, and goals of the company's strategic plan. It is worth mentioning that this model is not restricted to the partner company only but can be used by other companies in the citrus and related sectors.

We analyzed the proposed approach through computational experiments using a realistic, large-scale instance created from real-world data provided by the partner company. We

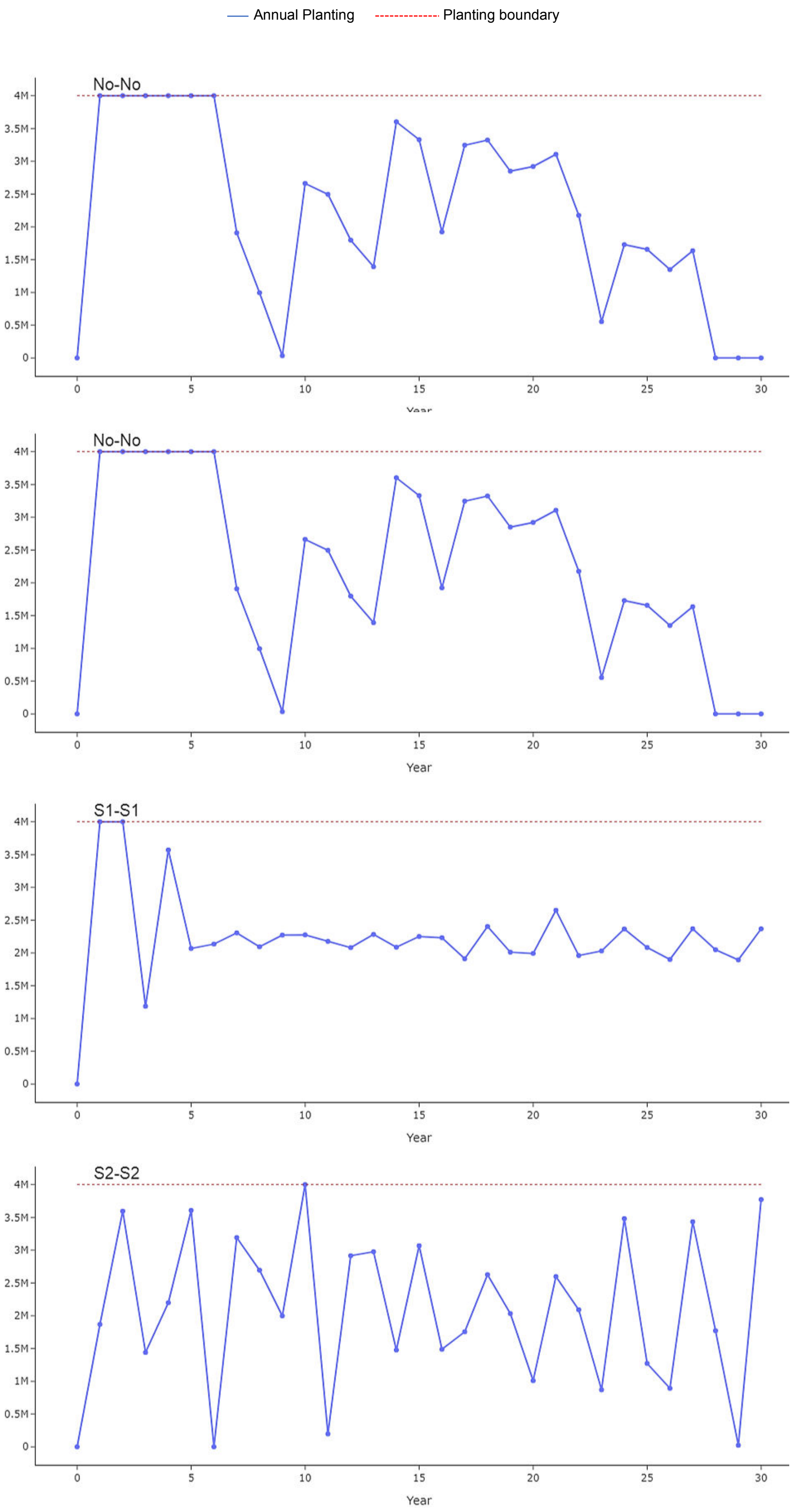

Figure 6: *Planting at each period in the solutions of model variants No-No, S1-S1, and S2-S2.*

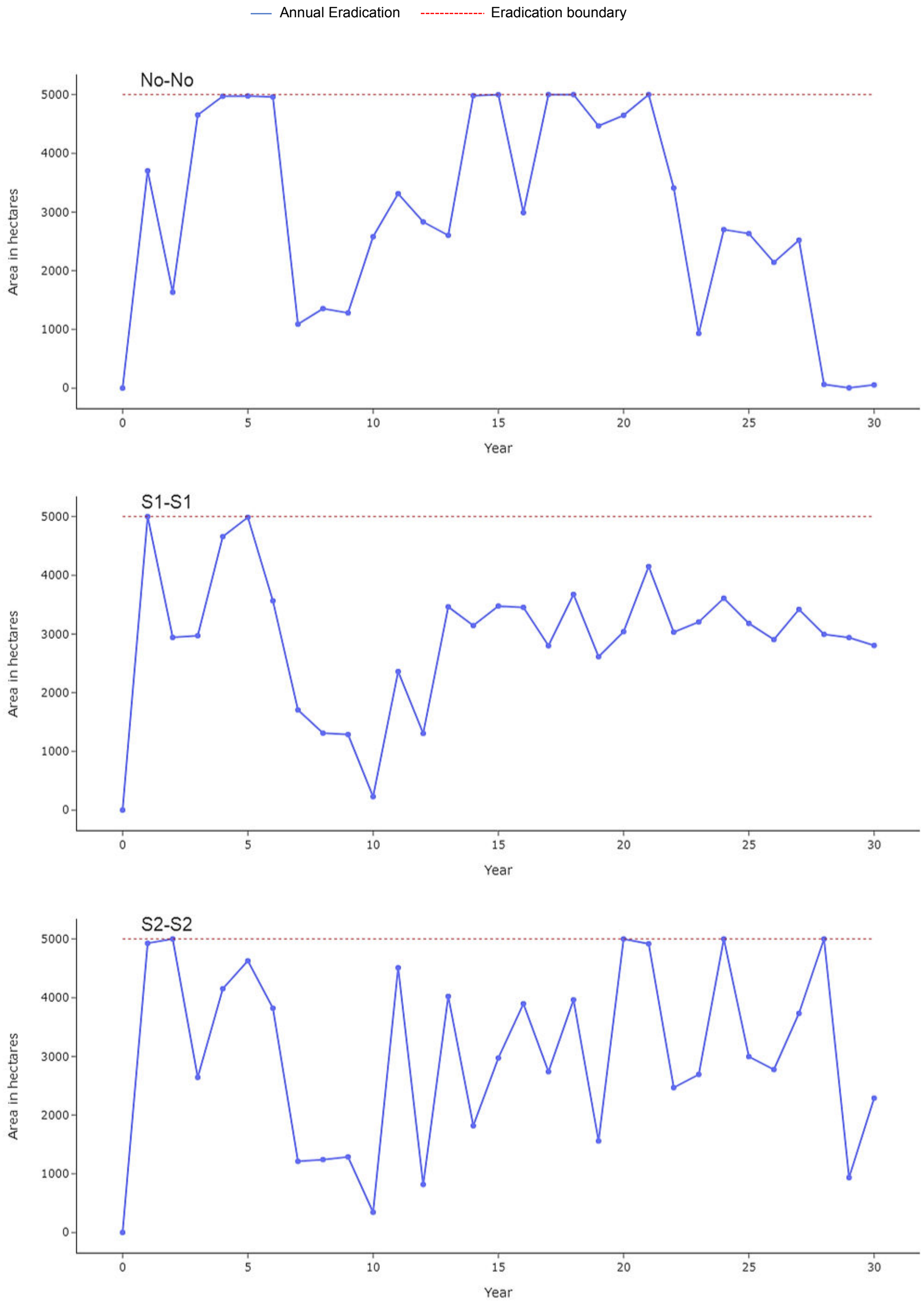

Figure 7: *Eradication at each period in the solutions of model variants No-No, S1-S1, and S2-S2.*

created seven variants of the SPMC model using different ways of enforcing the varietal and age balance requirements in order to evaluate the influence of these requirements on the solutions. As expected, the solution with the largest total orange production was obtained when we completely ignored the varietal and age balance requirements in the model. However, this solution does not meet the company goals, as the production significantly violates the desired minimum and maximum levels related to the varietal and age balance. Additionally, the production showed a highly oscillatory behavior through the years of the planning horizon, which has a negative impact on resource management.

Conversely, gradually penalizing the violation of the balance and age requirements through the years in the objective function of the SPMC resulted in a solution that was considered very effective by the partner company. Although it does not result in the largest total production, it still significantly increases the annual orange production with respect to the production observed in the plantation of the company at the beginning of the planning horizon. More importantly, this is achieved while completely satisfying the desired varietal and age balance requirements after a few periods from the beginning of the planning horizon. This solution also results in stable production values after period 13 and in stable plantation and eradication efforts. The resulting varietal and age balance, together with stability in production, planting, and eradication, were the main goals of the partner company. Therefore, the proposed optimization approach offers a successful tool to aid decision-making in their strategic planning process.

As is typical with optimization models, the solutions derived from the SPMC are significantly influenced by the input data and initial conditions. For example, the specified percentages for varietal and age balance can heavily impact the decisions and may lead to different production outcomes. Consequently, it is important for decision-makers utilizing this approach to accurately define input parameters and explore different scenarios that encompass relevant variations in the input data and parameters. The use of a computer-aided optimization tool like the one suggested in this paper offers an advantage as it enables managers to swiftly conduct experiments using diverse input data, thus allowing them to make decisions based on the best outcomes for the most likely scenarios. As indicated, each run of the SPMC can be performed within practical computation times. Given the periodicity and importance of the strategic planning process, the results of many different scenarios can be evaluated beforehand to support informed decision-making.

We can foresee interesting research opportunities regarding extensions of our study. First, one can investigate more efficient ways of ensuring the varietal and age balance requirements using soft constraints. One possibility would be using multi-objective techniques to deal with the conflicting objectives/requirements when maximizing production and minimizing varietal and age balance violation. Moreover, to improve the computational performance of solving the SPMC model, one should resort to decomposition approaches (e.g., decomposing by strata or periods). Alternatively, one can rely on model-based heuristics, such as rolling horizon approaches, *Relax-and-Fix*, and *Fix-and-Optimize*. All these approaches may help to resolve larger and more challenging instances.

Acknowledgements

This work was partially supported by São Paulo Research Foundation (FAPESP) [grant numbers 2013/07375-0, 2022/05803-3], Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) [Finance Code 001], the National Council for Scientific and Technological Development (CNPq) [grant number 313220/2020-4, 405702/2021-3].

References

[1] USDA, Citrus: World markets and trade, Available on: https://downloads.usda.library.cornell.edu/usdaesmis/files/w66343603/bv73d549r/1v53m4335/citrus.pdf, 2022.

[2] CEPEA, Afinal, quanto o agronegócio representa no PIB brasileiro?, Available in: https://www.cepea.esalq.usp.br/br/opiniao-cepea/afinal-quanto-o-agronegociorepresenta-no-pibbrasileiro.aspx, 2022.

[3] E. A. G. et al., Guia de reconhecimento dos citros em campo, FUNDECITROS 1 (2021) 83p.

[4] SECEX, Estatísticas de comércio exterior em dados abertos, Avaliable on: https://www.gov.br/produtividade-e-comercio-exterior/pt-br/assuntos/comercio-exterior/estatisticas/base-de-dados-bruta, 2023.

[5] D. A. Kimball, Description of citrus fruit, in: Citrus Processing: A Complete Guide, Springer, 1999, pp. 7–42.

[6] L. P. Catalá, G. A. Durand, A. M. Blanco, J. Alberto Bandoni, Mathematical model for strategic planning optimization in the pome fruit industry, Agricultural Systems 115 (2013) 63–71.

[7] J. Rajakal, R. Tan, V. Andiappan, Y. Wan, M. Pang, Does age matter? a strategic planning model to optimise perennial crops based on cost and discounted carbon value, Journal of Cleaner Production 318 (2021) 128526.

[8] EMBRAPA, Controle alternativo das doenças dos citros, Available in: http://www.infoteca.cnptia.embrapa.br/infoteca/handle/doc/1038594, 2016.

[9] FUNDECITROS, Levantamento da incidência das doenças dos citros: greening, cvc e cancro cítrico, FUNDECITROS 1 (2020b) 67p. URL: `https://www.fundecitrus.com.br/pdf/levantamentos/levantamento-doencas-2020.pdf`.

[10] G. Guerreiro Neto, S. R. F. Figueira, Maior dificuldade fitossanitária à produção da laranja no principal cinturão citrícola brasileiro - safras de 2017 a 2019., Citrus Res. Technol. 42 (2021) 10p.

[11] G. AGRO, Laranja e sucos de frutas mais caros: entenda o que tem elevado os preços, Available on: `https://g1.globo.com/google/amp/economia/agronegocios/noticia/2023/09/19/laranja-e-sucos-de-frutas-mais-caros-entenda-o-que-tem-elevado-os-precos.ghtml`. Acess at: October, 8th of 2023, 2023.

[12] J. Caixeta-Filho, Orange harvesting scheduling management: A case study, Journal of the Operational Research Society 57 (2006) 637–642.

[13] J. R. Munhoz, R. Morabito, Optimization approaches to support decision making in the production planning of a citrus company: A brazilian case study, Computers and Electronics in Agriculture 107 (2014) 45–57.

[14] E. C. Conforto, D. C. Amaral, S. L. d. Silva, Roteiro para revisão bibliográfica sistemática: aplicação no desenvolvimento de produtos e gerenciamento de projetos, in: Congresso Brasileiro de Gestão de Desenvolvimento de Produto - CBGDP, 2011, p. 12p.

[15] N. K. Tsolakis, C. A. Keramydas, A. K. Toka, D. A. Aidonis, E. T. Iakovou, Agrifood supply chain management: A comprehensive hierarchical decision-making framework and a critical taxonomy, Biosystems Engineering 120 (2014) 47–64.

[16] W. E. Soto-Silva, E. Nadal-Roig, M. C. Gonzàlez-Araya, L. M. Pla-Aragones, Operational research models applied to the fresh fruit supply chain, European Journal of Operational Research 251 (2016) 345–355.

[17] R. Jain, L. Malangmeih, D. S S Raju, S. Srivastava, I. Kingsly, A. Kaur, Optimization techniques for crop planning: A review, Indian Journal of Agricultural Sciences 88 (2018) 1826–1861.

[18] T. Taskiner, B. Bilgen, Optimization models for harvest and production planning in agri-food supply chain: A systematic review, Logistics 5 (2021).

[19] T.-D. Nguyen, T. Nguyen-Quang, U. Venkatadri, C. Diallo, M. Adams, Mathematical programming models for fresh fruit supply chain optimization: A review of the literature and emerging trends, AgriEngineering 3 (2021) 519–541p.

[20] A. R. Ravindran, V. V. P. Paul M. Griffin and, Service Systems Engineering and Management (Operations Research Series), 1st ed., CRC Press, 2018.

[21] A. Higgins, R. Muchow, A. Rudd, A. Ford, Optimising harvest date in sugar production: A case study for the mossman mill region in Australia: I. development of operations research model and solution, Field Crops Research 57 (1998) 153–162.

[22] A. Higgins, Optimizing cane supply decisions within a sugar mill region, Journal of Scheduling 2 (1999) 229–244.

[23] P. He, J. Li, D. Zhang, S. Wan, Optimisation of the harvesting time of rice in moist and non-moist dispersed fields, Biosystems Engineering 170 (2018) 12–23.

[24] M. Escallón-Barrios, D. Castillo-Gomez, J. Leal, C. Montenegro, A. Medaglia, Improving harvesting operations in an oil palm plantation, Annals of Operations Research 314 (2022) 411–449.

[25] H. Florentino, A. De Lima, L. De Carvalho, A. Balbo, T. Homem, Multiobjective 0-1 integer programming for the use of sugarcane residual biomass in energy cogeneration, International Transactions in Operational Research 18 (2011) 605–615.

[26] H. O. Florentino, M. Pato, A bi-objective genetic approach for the selection of sugarcane varieties to comply with environmental and economic requirements, Journal of the Operational Research Society 65 (2014) 842–854.

[27] S. C. Poltroniere, A. Aliano Filho, A. S. Caversan, A. R. Balbo, H. de Oliveira Florentino, Integrated planning for planting and harvesting sugarcane and energy-cane for the production of sucrose and energy, Computers and Electronics in Agriculture 184 (2021) 105956.

[28] M. Osaki, M. Batalha, Optimization model of agricultural production system in grain farms under risk, in sorriso, brazil, Agricultural Systems 127 (2014) 178–188.

[29] A. Aliano Filho, H. O. Florentino, M. V. Pato, S. C. Poltroniere, J. F. S. Costa, Exact and heuristic methods to solve a bi-objective problem of sustainable cultivation, Annals of Operations Research 314 (2022) 347–376.

[30] K. Darby-Dowman, S. Barker, E. Audsley, D. Parsons, A two-stage stochastic programming with recourse model for determining robust planting plans in horticulture, Journal of the Operational Research Society 51 (2000) 83–89.

[31] N. Brulard, V.-D. Cung, N. Catusse, C. Dutrieux, An integrated sizing and planning problem in designing diverse vegetable farming systems, International Journal of Production Research 57 (2019) 1018–1036.

[32] J. R. Munhoz, R. Morabito, Um modelo baseado em programação linear e programação de metas para análise de um sistema de produção e distribuição de suco concentrado de laranja, Gestão e Produção (UFSCar) 8 (2001c) 139–159p.

[33] J. R. Munhoz, R. Morabito, Otimização no planejamento agregado de produção em indústrias de processamento de suco concentrado congelado de laranja, Gestão Produção (UFSCar) 17 (2010) 465–481p.

[34] J. R. Munhoz, R. Morabito, Uma abordagem de otimização robusta no planejamento agregado de produção na indústria cítrica, Produção 23 (2013) 422–435p.

[35] S. Seminara, S. Bennici, M. Di Guardo, M. Caruso, A. Gentile, S. La Malfa, G. Distefano, Sweet orange: Evolution, characterization, varieties, and breeding perspectives, Agriculture 13 (2023) 264.

[36] J. R. Munhoz, R. Morabito, Um modelo baseado em programação linear e programação de metas para análise de um sistema de produção e distribuição de suco concentrado congelado de laranja, Anais do XXXIII Simpósio Brasileiro de Pesquisa Operacional 1 (2001A) 1–10p.

[37] FUNDECITROS, Inventário de Árvore do cinturão citrícola de são paulo e triângulo/sudoeste mineiro - retrato dos pomares em março/2021, FUNDECITROS 1 (2021) 107p. URL: `https://www.fundecitrus.com.br/pdf/pes_relatorios/2021_07_30_Inventario_e_Estimativa_do_Cinturao_Citricola_2021-2022.pdf`.

[38] M. M. Caputo, Avaliação de doze cultivares de laranja doce de maturação precoce na região sudoeste do Estado de São Paulo, Masters Thesis, Universidade de São Paulo, 2012.

[39] G. Schafer, M. Bastianel, A. L. C. Dornelles, Porta-enxertos utilizados na citricultura, Ciência Rural 31 (2001).

[40] E. F. Coelho, A. S. de Oliveira, A. F. de Jesus Magalhães, IRRIGACAO E FERTIRRIGACÃO EM CITROS., 38 ed., Embrapa Mandioca e Fruticu/tur, 2000.

[41] A. C. R. Cruz, Consumo de Água por Cultura de Citros Cultivada em Latossolo Vermelho Amarelo, Master's thesis, Universidade de São Paulo, https://www.teses.usp.br/teses/disponiveis/11/11140/tde-20102003-153219/publico/antonio.pdf, 2003.

[42] J. W. Kendall, Hard and soft constraints in linear programming, Omega 3 (1975) 709–715.

[43] M. Tamiz, D. F. Jones, E. El-Darzi, A review of goal programming and its applications, Annals of operations Research 58 (1995) 39–53.

[44] C.-L. Hwang, A. S. M. Masud, Multiple objective decision making—methods and applications: a state-of-the-art survey, volume 164, Springer Science & Business Media, 2012.

[45] D. Jones, M. Tamiz, Pratical Goal Programming, 1 ed. ed., Springer, 2010.

Appendix A. Varietal balance results for other model variants

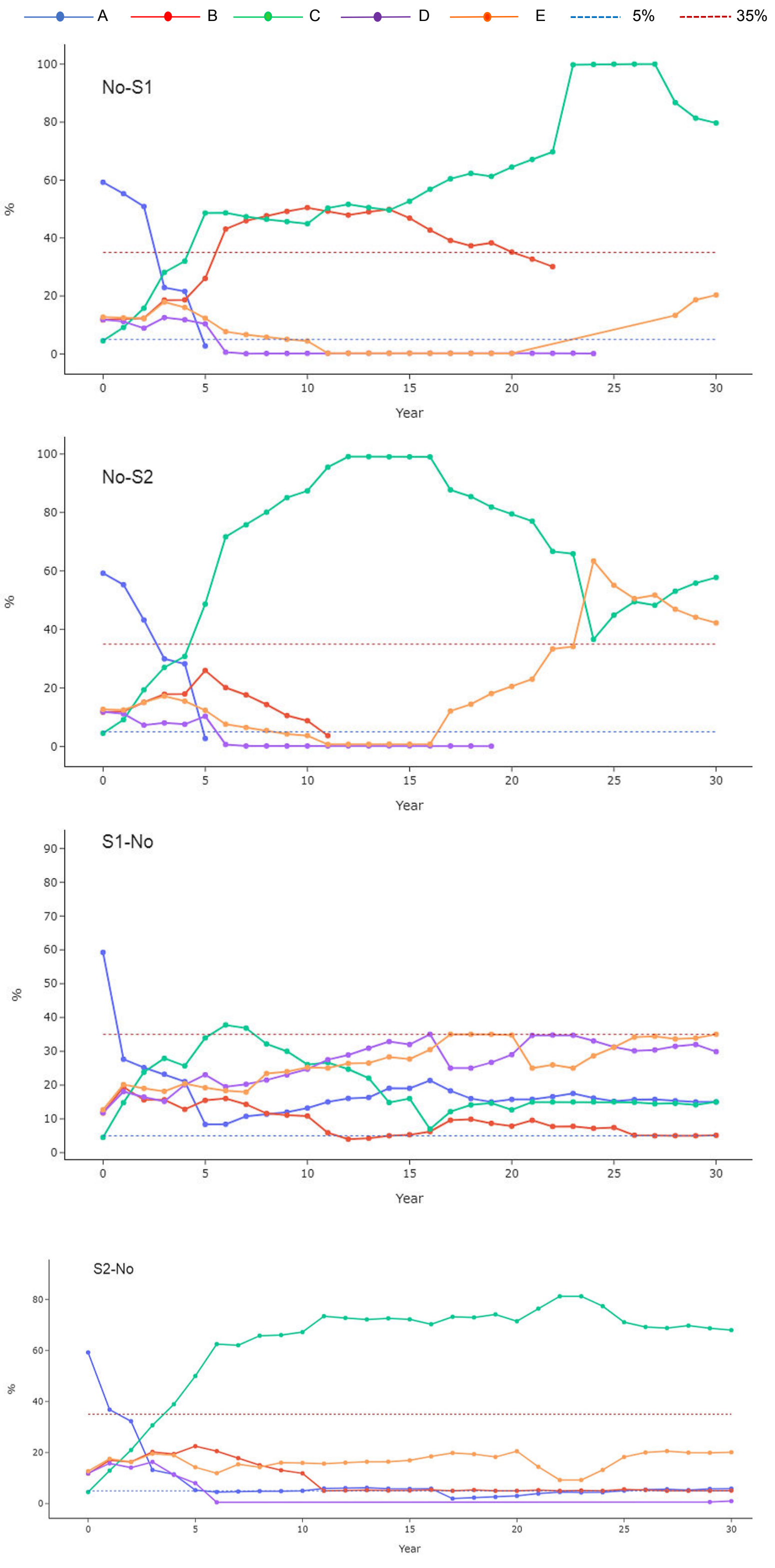

Figure A.8: *Production percentage of each variety at each time period in a given pole of the company, in the solutions obtained with the model variants No-S1, No-S2, S1-No, and S2-No.*

Appendix B. Age balance results for other model variants

Figure B.9: *Percentages of plants in the five age groups at each time period in the solutions obtained with the model variants No-S1, No-S2, S1-No, and S2-No.*

Appendix C. Planting results for other model variants

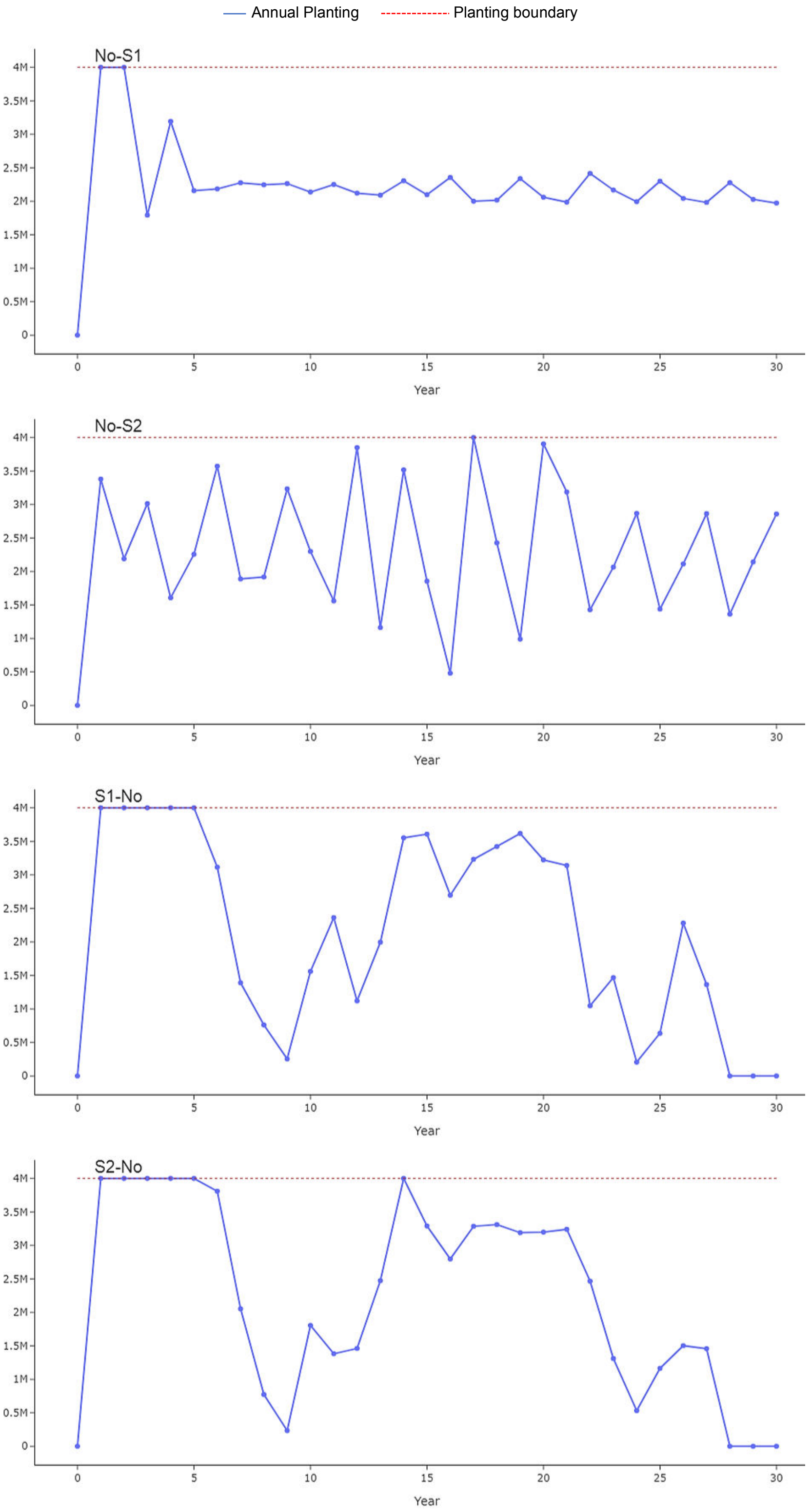

Figure C.10: *Planting of new seedlings at each period in the solutions of model variants No-S1, No-S2, S1-No, and S2-No.*

Appendix D. Eradication results for other model variants

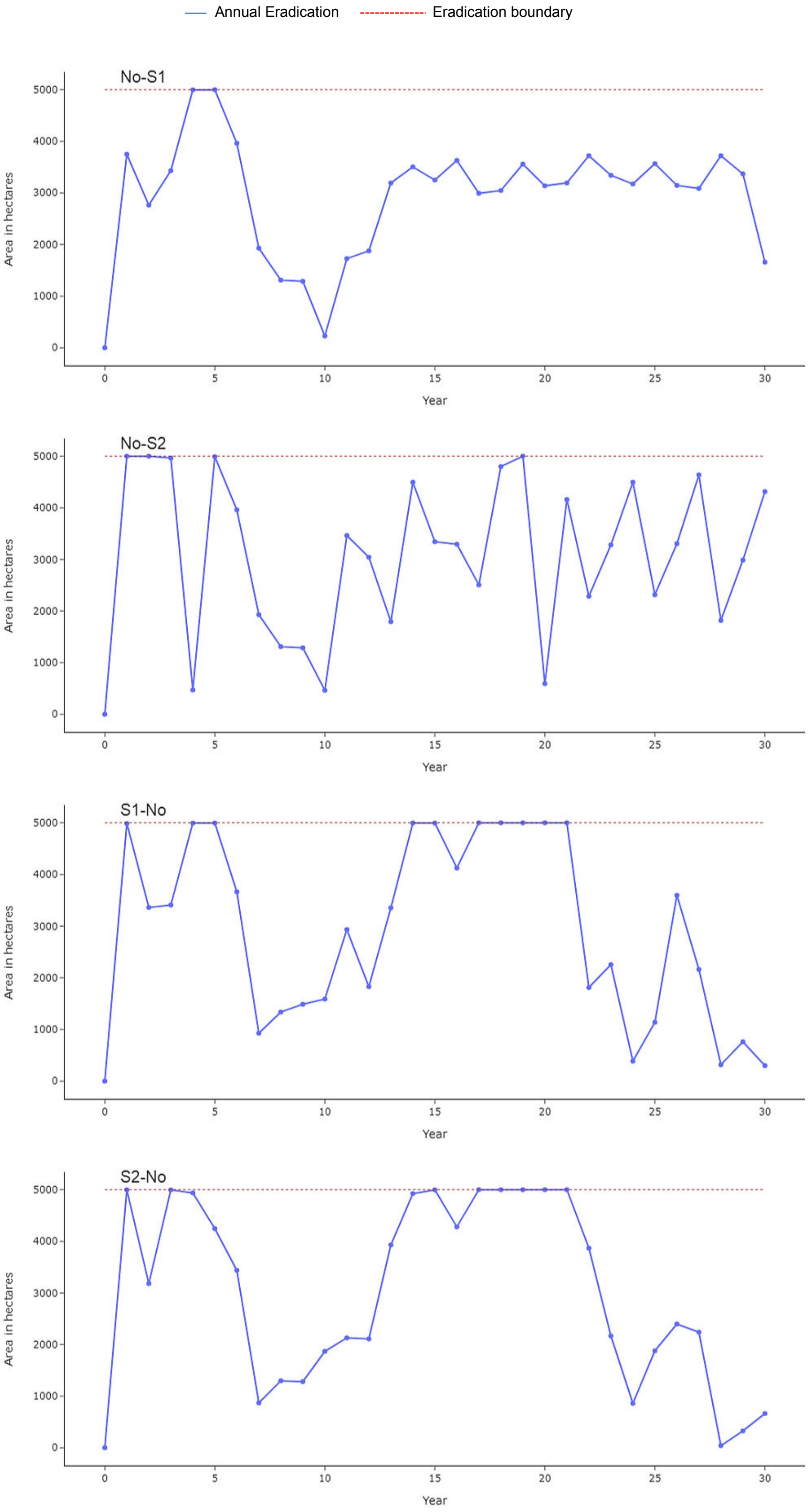

Figure D.11: *Eradication at each period in the solutions of model variants No-S1, No-S2, S1-No and S2-No.*

Appendix E. Details of computational experiments with the SPMC

We present the details regarding the experiments with different variants of the SPMC model. Table E.3 presents, for each variant, the final value of the objective function and the relative gap of the best solution obtained when the solver achieved the time limit of 1 hour. Note that the solver could not prove optimality in any variant and, hence, all experiments finished due to the time limit. For this reason, we do not present computation times in Table E.3.

Table E.3: Computational results of different variants of the SPMC model.

Variant	Obj. Func.	Relative gap
No-No	1.98E+21	0.1883%
No-S1	1.87E+21	0.7659%
S1-No	1.78E+21	0.5090%
S1-S1	1.67E+21	1.4736%
S2-No	-2.16E+21	0.6385%
No-S2	-6.47E+21	21.3896%
S2-S2	-1.14E+22	15.1255%

We observe in Table E.3 that the model variants S2-No, No-S2, and S2-S2 have negative values in their respective objective functions. This occurs because, in these experiments, these variants models highly violates the varietal and age balance requirements, which are penalized in the objective function. Moreover, the last two variants still show large relative gaps when the solver terminates due to the achieving the time limit. These results reinforce that the penalization strategy S2 did not provide good solutions. For the other four model variants, we observe that S1-S1 has the smallest objective value, mainly because it effectively satisfies the varietal and age balance requirements in the last periods. Additionally, it has a slightly larger gap among the four, indicating the difficulty of solving the SPMC with penalizations on both varietal and age balance violation.