An optimization approach to aggregate production planning with combined MTS-MTO production strategy

José Luís Garrido Vila-Chã

Master's Dissertation

Supervisor: Prof. Manuel Augusto de Pina Marques

U.PORTO

FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Mecânica

2021-06-28

Abstract

The correct definition of the production strategy for each product and appropriate medium-term production planning are two crucial aspects that industrial companies need to consider in order to be successful. The decisions made at this level impact operational efficiency and inventory management in opposite directions. While producing to stock with larger lot sizes leads to greater productive efficiency, it also results in higher inventory costs. At the same time, the complete opposite strategy would decrease stock keeping expenses but seriously increase the entropy in the production system by increasing setup operations and lead times. Therefore, the trade-off between both effects has to be considered during the production strategy definition and the creation of production plans.

The present study was conducted in a company operating in the wine industry and was divided in two separate modules. The first one aimed to find the best production strategy for each product considering a hybrid make-to-stock/make-to-order production system. Then, in the second module, an optimization approach to aggregate production planning was developed to generate production plans in the medium term.

The proposed approach for production strategy definition considers the impact that a change in the production strategy of one product has in the whole system by modeling congestion effects using M/G/1 queues. As the problem formulation was proved in the literature to be non-linear, a greedy heuristic was developed to find near-optimal solutions iteratively. The process consists of changing the production strategy of the product that, at each iteration, provides the highest savings. Once a strategy change occurs, costs and congestion effects of the whole system are recalculated and the process is repeated until no more savings can be achieved.

Regarding the aggregate production planning problem, an optimization model based on mixed-integer programming formulation was developed. The planning context includes multiple production lines, a multi-level product structure and significant demand seasonality. The goal was to determine the most efficient and leveled allocation of production resources that meets the fluctuating demand. Three of the main outputs of this model are determining optimized production volumes, finding the most suitable workforce levels and making the best allocation of products to their compatible production lines. The model's objective function consisted in minimizing a global cost function composed by operational and inventory costs.

By applying the developed methodologies the company is expected to achieve significant savings. Preliminary results indicate that the proposed production strategy definition can lead to a reduction of approximately 14% in the total involved costs while at the same time improving the service level by $2.4pp$. Simultaneously, the implementation of the optimized production plans is estimated to result in 3% savings in total costs over a 12-month period compared to current practices.

Resumo

A correta definição da melhor estratégia de produção para cada produto e um planeamento de produção a médio prazo apropriado são dois aspetos cruciais que as empresas industriais têm de ter em consideração para serem bem-sucedidas. As decisões tomadas a este nível impactam a eficiência operacional e a gestão de inventário em sentidos opostos. Enquanto uma estratégia de produção para stock com lotes de produção maiores resulta numa maior eficiência operacional, também resulta em custos de inventário mais elevados. Por sua vez, uma estratégia completamente oposta resultaria numa diminuição nos custos de inventário, mas aumentaria seriamente a entropia do sistema produtivo, através de um maior número de *setups* e maiores tempos de espera. Assim, o compromisso entre ambos os efeitos tem de ser considerado na definição da estratégia de produção e na criação de planos de produção.

O presente estudo foi realizado numa empresa que opera no setor do vinho e foi dividido em dois módulos distintos. O primeiro módulo visou encontrar a melhor estratégia de produção para cada produto, considerando um sistema híbrido de produção *make-to-stock*/*make-to-order*. Por sua vez, no segundo módulo foi desenvolvida uma abordagem de otimização do planeamento agregado da produção capaz de gerar planos de produção a médio prazo.

A abordagem proposta para a definição da estratégia de produção considera o impacto que a mudança na estratégia de produção de um produto tem na totalidade do sistema, modelando os efeitos de congestionamento através de filas de espera M/G/1. Verificando-se na literatura que a formulação do problema é não-linear, foi desenvolvida uma heurística gulosa para encontrar de forma iterativa soluções quase ótimas. O processo consiste em alterar a estratégia de produção do produto que em cada iteração conduz a maiores poupanças. Uma vez definida uma nova estratégia de produção, os custos e os efeitos de congestionamento do sistema são recalculados e o processo é repetido até que não seja possível obter mais poupanças.

Relativamente ao problema de planeamento agregado de produção, foi desenvolvido um modelo de otimização baseado em programação inteira mista. O contexto de planeamento estudado inclui múltiplas linhas de produção, uma estrutura de produtos com múltiplos níveis e significativos efeitos de sazonalidade da procura. O objetivo foi determinar a alocação mais eficiente e nivelada dos recursos produtivos, que respondesse às flutuações da procura. Três dos principais *outputs* do presente modelo são a determinação de volumes de produção otimizados, dos níveis de mão-de-obra adequados, e fazer a melhor alocação possível de produtos às linhas de produção com as quais são compatíveis. A função objetivo do modelo a minimizar incluiu as despesas operacionais e de inventário.

É expectável que a aplicação das metodologias desenvolvidas resulte em poupanças significativas para a empresa. Resultados preliminares indicam que a abordagem de definição da estratégia de produção proposta pode levar a uma redução de aproximadamente 14% dos custos totais envolvidos, e ao mesmo tempo melhorar o nível de serviço em $2,4pp$. De forma semelhante, estima-se que a implementação dos planos de produção otimizados resulte, num período de 12 meses, numa poupança de 3% no total dos custos, em comparação com as práticas atuais.

Acknowledgments

First, I would like to thank my supervisor, Professor Manuel Pina Marques, for all the valuable insights and thorough analysis of my work. I realise that this dissertation would not have the same value without your careful guidance and support.

To Professors Bernardo Almada Lobo, Luís Guimarães and Pedro Amorim, I am truly grateful for the opportunity of working in such a challenging environment as LTPlabs. To all LTPeers, thank you for teaching me so much and welcoming me to such an enthusiastic and encouraging working environment, that truly made work much easier.

I am thankful to Diogo Pereira da Silva, my supervisor at LTPlabs, for the rigorous review of this work and for providing me with such careful guidance. I am forever thankful for your friendship and for all the invaluable advice you gave me along the way.

To Daniel Pereira and Eduardo Ribeiro, who also worked with me in this project, thank you for always being available to answer my doubts and helping me at every step of the way. Your contribution was vital to the success of this work.

None of this would have been possible without an incomparable support back at home. To my mother and father, I cannot express my gratitude enough for the unconditional support and encouragement you have given me throughout my journey. Thank you for always being there for me and for giving me everything I needed to pursue my dreams. To my extended family, thank you for all the incentive you have always given me.

To my friends, you were the ones that uplifted this journey and made it worthwhile. In these five years, we grew together and shared amazing moments that I will cherish forever.

Maria, your endless care, support, and patience made everything much easier. I could not be more grateful for having you by my side and can only hope to have an entire life to repay you.

José Luís Vila-Chã

"Tell me and I forget. Teach me and I remember. Involve me and I learn."

Benjamin Franklin

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 The Project . . . 2
- 1.3 Dissertation structure . . . 3

2 Literature Review **5**
- 2.1 Production strategies . . . 5
  - 2.1.1 Combined MTS-MTO models . . . 6
- 2.2 Aggregate production planning . . . 7
  - 2.2.1 Classification of APP Models . . . 8
  - 2.2.2 Addressed issues and main considerations of APP models . . . 9
  - 2.2.3 Solving approaches . . . 12

3 The Problem **17**
- 3.1 Problem description . . . 17
- 3.2 Planning context . . . 18
  - 3.2.1 Production environment and process . . . 19
  - 3.2.2 Current practices . . . 21
  - 3.2.3 Preliminary analysis . . . 23
- 3.3 The proposed approach . . . 26

4 Methodology **29**
- 4.1 Production strategy definition . . . 29
  - 4.1.1 The model . . . 30
  - 4.1.2 Problem formulation . . . 31
  - 4.1.3 Heuristic . . . 35
- 4.2 Aggregate production planning . . . 37
  - 4.2.1 Model formulation . . . 38

5 Results **51**
- 5.1 Production strategy definition Results . . . 51
- 5.2 Aggregate production planning Results . . . 52
  - 5.2.1 Computational Results . . . 53
  - 5.2.2 Production plans Results . . . 54

6 Conclusions and future work **57**

Acronyms and Symbols

APP	Aggregate production planning
CLSP	Capacited lot sizing problem
EOQ	Economic-order-quantity
FCFS	First-come-first-served
LS	Lot sizing
MIP	Mixed-integer programming
MPQ	Minimum production quantity
MTO	Make-to-order
MTS	Make-to-stock
SKU	Stock keeping unit

List of Figures

1.1 Project's timeline 3

2.1 APP models' classification structure. Source: Cheraghalikhani et al. (2019) . . . 8

3.1 Production process 20
3.2 Production strategy definition process 21
3.3 Current production planning process 22
3.4 ABC analysis 23
3.5 ABC-XYZ analysis 24
3.6 Orders lead time: Core vs. Non-core 25
3.7 Orders lead time: MTS vs. MTO 25
3.8 Stock distribution 25

5.1 Cost comparison between the current practices (As Is) and the optimized strategy definition (To Be) 52
5.2 Comparison between the proposed solution and the As Is situation 52
5.3 Production levels evolution 54
5.4 Stock Evolution 55
5.5 Average number of monthly setups 55
5.6 Cost comparison between the historical practices and optimized plan 56

List of Tables

2.1 Basic issues in APP models. Source: adapted from Cheraghalikhani et al. (2019) 10

3.1 Proposed approach for the production strategy module 26
3.2 Proposed approach for APP module 27

4.1 Table of notation for the production strategy problem formulation 32
4.2 Table of notation for the optimisation model. 40

5.1 Problem data and MIP dimensions 53
5.2 Solution time and MIP gap 54

Chapter 1

Introduction

1.1 Motivation

In a global marketplace with ever increasing competitiveness levels, production planning is crucial for the success of industrial companies. Proper production planning leads to efficient and economical acquisition, consumption and allocation of the production resources to transform raw materials into finished products. Decisions regarding production planning are taken along different hierarchical levels inside an organization and involve multiple functional units, such as production, accounting, and marketing. These decisions are commonly divided into three levels: strategic, tactical and operational (Torabi et al., 2010). Strategic planning typically considers a 2–5 year planning horizon (long-term) and addresses high level decisions such as product design and plant layout. Tactical planning aims for the medium term and intends to provide optimal decisions on production, inventory, workforce levels and overtime usage, leveling these factors and absorbing demand peaks. Finally, at the operational level, short-term decisions like detailed scheduling and sequencing of production jobs are taken to optimize day-to-day operations (Misni and Lee, 2017).

The main focus of this thesis will lie on the medium-term scope, more specifically on Aggregate production planning (APP). As defined by (Jamalnia et al., 2017), APP is a medium-term production and employment planning that usually covers a 3 to 18 months time horizon. In general, its goals include the determination of optimal production volumes, hiring and lay off rates, workforce and inventory levels, backordering and subcontracting quantities, always taking into account the limitations of production resources for each period within the planning horizon. Many companies still make this medium-term planning based on empirical knowledge and not enough analytically backed by a decision support system. Combining this factor with the highly complex nature of the problem and the immense number of possible solutions, there is most certainly an opportunity for improvement. Therefore, the study of optimization models in APP is a very pertinent subject as the potential to improve efficiency levels in these companies is very high.

Today, another fundamental part of a company's ability to be competitive in the marketplace is the service level it can provide to its customers. Delivering the finished products on time, meeting

the established deadlines, is more important than ever as the customer base becomes increasingly demanding. The production strategy is a pivotal factor in achieving this goal. For instance, a pure make-to-stock (MTS) strategy, with all orders being fulfilled from available stock, would lead to perfect service levels. However, this strategy may lead to high inventory levels, which then results in high holding costs. On the other hand, a pure make-to-order (MTO) strategy decreases the holding costs, as production is only initiated after the order arrival, but decreases the service level as the items are not ready to fly off the shelf. Different products have different market demands and production characteristics, like setup times and production rates, so finding the best production strategy for each SKU can be decisive. While some companies prefer to make all their products to stock or all to order, some opt to maintain a hybrid strategy, combining the best of both strategies (Rajagopalan, 2002). According to (Kaminsky and Kaya, 2006), combined MTS-MTO systems offer significantly better performance than pure MTS or MTO systems. This combined approach will also be included in the scope of this work to find the best production strategy for each product and serve as a guideline for the APP model.

1.2 The Project

This dissertation was developed as part of a consulting project for a Portuguese company operating in the beverage industry, specifically in the wine sector. The company is one of the leading producers of premium port wine and one of the top Portuguese wine producers. With two filling plants, both with multiple production lines and a range of around 1200 products across more than 20 different brands, efficient production planning and strategy definition is, for them, of the utmost importance. The issues addressed in this dissertation are part of a larger project that aims to provide them with analytical decision support tools to optimize production planning and raw material procurement.

An earlier diagnosis conducted in the company identified several challenges and opportunities for improvement. The existent production strategy definition process was strongly based on a commercial definition and not enough analytically backed by a cost-service function. Moreover, master production plans were not systematically generated and, instead, some tactical analyses were performed in a rather *ad-hoc* and empirical way.

The main objectives of this project were the development of an algorithm for the definition of the MTS-MTO production strategy for each product and a Mixed-integer programming (MIP) optimization model for the Aggregate Production Planning. These two modules, and a third one regarding material requirements planning that will not be addressed in the present work, will be delivered in the form of a Decision Support System (DSS). This tool will allow the company to leverage data to improve operational efficiency, with improved decision-making.

The key stakeholders of this project were the planning and production departments. However, other divisions, like sales, IT and purchasing, were included in the project to gather all the required information and ensure an effective implementation of the final platform.

Figure 1.1 outlines the main stages of the project, with the respective dates and resulting milestones. The first main module of the project is the definition of a list with the production strategy (MTS-MTO) for each product. This module subdivided into problem formulation, algorithm development and pilot testing (model tuning) will provide inputs, like production lots and safety stocks of MTS products, for the next stage. The second module consists of the development of the MIP optimization model for aggregate production planning and starts by mapping all the requirements and formulating the model constraints. After that, the development of the model itself takes place, followed by a testing phase where preliminary results are generated and validated with the client, in order to perform the necessary adjustments to the model.

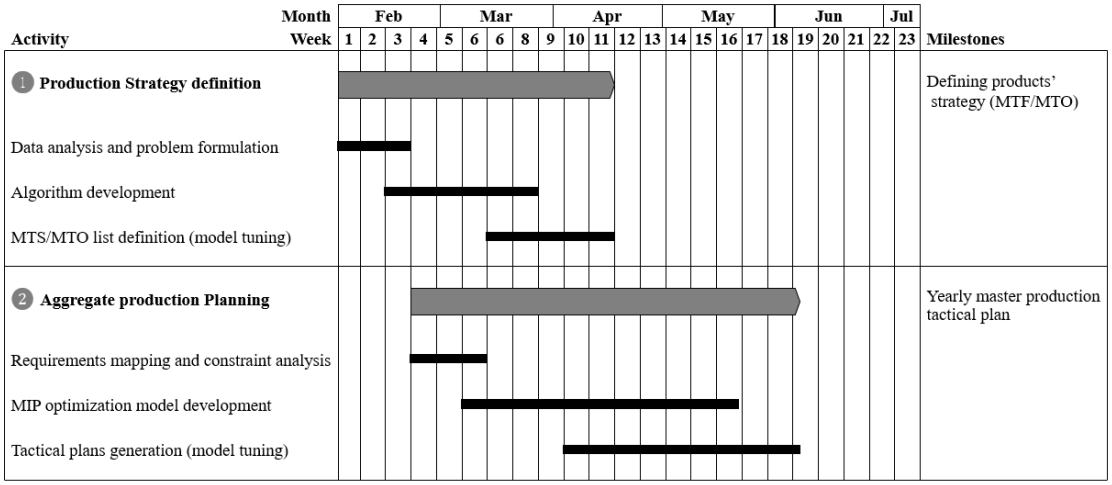


Figure 1.1: Project's timeline

1.3 Dissertation structure

The present dissertation is organized in six chapters as follows. Chapter 2 provides a literature review of key studies on mixed MTS-MTO models and a global overview on APP models, from a model classification structure to the most used solving approaches and formulation. Chapter 3 gives a detailed description of the problem addressed in this project as well as an overview of the company's planning context. Chapter 4 thoroughly elaborates on the developed methodology. Chapter 5 presents the computational and preliminary results of the applied methodology. Lastly, Chapter 6 summarizes the main conclusions of this study, identifying some of the shortcomings of the developed methodology and discussing future improvement opportunities.

Chapter 2

Literature Review

This chapter aims to provide the necessary theoretical background to support the study and elaborate on the current state of the art in hybrid production strategies and aggregate production planning. Section 2.1 aims to provide a brief review of production strategies with a focus on combined MTS-MTO approaches. Section 2.1.1 introduces the main characteristics of the most used production policies, while section 2.1.2 covers some key studies on combined MTS-MTO approaches, emphasizing the strategy selection process for each product. Aggregate production planning models are reviewed in Section 2.2. Firstly, a classification structure for APP models is introduced in Section 2.2.1, followed by an analysis of the main assumptions and addressed issues found in the relevant literature in section 2.2.2. Section 2.2.3 elaborates on solving approaches, with particular attention being given to the study of the capacitated lot sizing problem, due to its similarities in formulation with the problem in hand.

2.1 Production strategies

Nowadays, different production strategies are used by manufacturing companies to meet their customer demands. While some prefer to fulfil their demand through an inventory of finished products, operating under a make-to-stock (MTS) approach, others wait for an order to enter the system to initiate its production, opting for a make-to-order (MTO) strategy (Tashakori, 2015).

In pure MTS systems, also known as push systems, products are made and stored based on the forecasted demand, and the arriving orders are fulfilled from the available stock. This strategy leads to shorter delivery lead times and results in more efficient usage of the production capacity. Moreover, fewer setups are performed and longer production runs are used, enabling companies to take advantage of economies of scale (Nagib et al., 2016). However, even though the time wasted in setups is lower and, consequently, so is the cost associated with this machine setting decision, MTS strategies result in high levels of inventory and, therefore, high inventory costs. Inventory costs are related to storage expenses like warehousing and the cost of opportunity of the invested capital, making this approach more suitable for less expensive products. Furthermore, as highlighted by Soman et al. (2004), MTS systems are less flexible and, as a result, generally offer

a low variety of product specifications. Some of the main performance measures of this policy are the production throughput, inventory levels and line occupation rates.

On the other hand, companies operating under pure MTO strategies, also called pull strategies, have greater flexibility to quickly react to demand changes and offer higher levels of product customization. Even though the MTO policy leads to lower inventory holding costs, as orders are dispatched after production, this happens at the expense of a decrease in the service level. Since no inventory is held, lead times substantially increase, making customers wait longer for their items (Nagib et al., 2016). This is a typical approach for more expensive products and the main indicators to evaluate its performance are the average response time and average order delay (Soman et al., 2004).

Even though pure MTS and MTO strategies are still very common, as global markets get increasingly demanding with customers expecting higher service levels and more diversified product lines, companies are diving into hybrid approaches that combine the advantages of both strategies. Kaminsky and Kaya (2006) concluded that in combined models, costs can be cut dramatically compared to the pure approaches, and far greater performance can be achieved.

2.1.1 Combined MTS-MTO models

Combined MTS-MTO systems are usually implemented in multi-product systems where products experience very different demand patterns. Some products have high and stable demands and others are low and unpredictable. In these combined approaches, some products are made-to-stock, based on the forecasted demand, and others are made-to-order, with production only initiating after an order is received. The difficult part lies in making the decision on which production strategies best suits each product. Some studies have been conducted to address this topic, creating methodologies that evaluate the relative importance of the factors involved. Two of those studies, (Williams, 1984) and (Rajagopalan, 2002), are reviewed next.

Williams (1984) was one of the first to study hybrid MTS-MTO systems, with a focus on the definition of the production strategy for each product. The paper proposes an approach for one-stage systems that considers the stochastic nature of demand and lead times. The interaction between products and capacity is included using queuing theory, with the queue being modelled as a $M/G/m$ system. The first step of the proposed methodology is to perform a Pareto (ABC) analysis to the product range. Then, the higher demand products (A items) are immediately defined as MTS and lower demand products (C items) are classified as MTO. Only the remaining products (B items) are subjected to the decision-making model. Firstly, the items are ordered by ascending demand, and the goal is to determine the line that separates the two policies, i.e., products with demand below that threshold are MTO, and the ones with higher demand are MTS. To make this decision, a costs function, including holding costs, setup costs and stockout costs, is applied to each possible strategy. The impact of the strategy changes on the whole system is evaluated at each step. Apart from the initial division of the products according to their level of demand, the details of the technique used in this study are very similar to the ones used in a more recent study by Rajagopalan (2002) that will be described next.

The model presented by Rajagopalan (2002) addresses the MTS-MTO decision problem in an environment characterized by a large number of items with stochastic demand and varying processing times. The problem context also included significant changeover times, limited production capacity and congestion effects. In this work, an heuristic was developed to solve a non-linear MIP formulation of the strategy definition problem and also to define batch sizes for the MTS items. The inventory control of MTS items used a (q,r) policy, which means that when inventory of product i goes down to a level r_i, a production order for q_i units is initiated. The main goal of the model is to minimize inventory costs of MTS items while ensuring that orders for MTO items are fulfilled within a specified lead time with a specified service level. A first-come-first-served (FCFS) basis is used to queue up the orders for both MTS and MTO products. In order to capture the congestion effects, the queue is modelled as an $M/G/1$ system, enabling the author to obtain the mean queue time. This is then used to calculate the production lead time distribution, which is used to determine the safety stock of MTS items and the service level of MTO items. The model is based on some important trad-offs that deserve to be highlighted. Making an item to order reduces inventory costs but has the downside of needing more setups, which results in higher capacity utilization with non-productive activity. This increases congestion, resulting in higher mean waiting times in the queue, which leads to increased global lead times. Consequently, higher safety stocks for MTS items are needed and service levels for MTO items decrease. A similar trade-off between the inventory cost and the number of setups occurs when determining the lot size for MTS items, as smaller batch sizes lead to lower inventory costs, but a higher number of setups is needed. Concerning the heuristic itself, the starting point considers all products as MTS and then, one by one, follows a greedy approach to find the item in which the change of the production strategy to MTO leads to the highest savings, and the change is made. Once the strategy change on this product is registered, all costs are recalculated, and the procedure is repeated until no more gains in the global system are possible. This highlights an important feature of the proposed approach as it considers the impact that a change in the production strategy of one product has on the global performance of the system.

2.2 Aggregate production planning

Aggregate production planning (APP) is a medium term capacity planning to determine the optimum production volumes and workforce levels required to meet fluctuating customer demands in the near future (Cheraghalikhani et al., 2019).This type of model's common outputs include hiring and layoff rates, inventory levels, backordering and subcontracting quantities. These results are steered by the limitations of production resources for each time period in the planning horizon, which typically ranges from 3 to 18 months (Jamalnia et al., 2017).

Therefore, APP techniques are used to ensure, at a gross level, that adequate resources and capacity exist, to meet fluctuating demand throughout the planning horizon. As the name implies, this type of models works at a more aggregated level, typically dealing with product families instead of the full product range. This greatly reduces the amount of data used during the planning

procedure, therefore making the process faster and enabling plans to be frequently updated (Gomes da Silva et al., 2006).

There are many different APP methodologies and techniques studied in the literature and currently used in industrial companies. A survey conducted by Nam and Logendran (1992) identified some of the most used techniques, which included: Trial and error methods, Graphical techniques, Parametric production planning, Production switching heuristic, Linear programming, Goal programming, Mixed-integer programming, Transportation method and Simulation models. With so many possibilities, it is important to be able to classify them. For this reason, a classification structure for APP models is presented in the following section.

2.2.1 Classification of APP Models

APP models have been classified in different ways by different authors. Nam and Logendran (1992), for example, proposed a classification that divided methodologies into two groups depending on their ability to produce exact optimal or near-optimal solutions. However, in this work a more recent classification, proposed by Cheraghalikhani et al. (2019), will be followed. This classification scheme separates APP models into different structural groups based on two important criteria: level of uncertainty and number of objective functions considered in the model. APP models can be classified as deterministic or uncertain regarding the level of uncertainty of the input parameters. The latter, on its turn, can be subdivided into stochastic and fuzzy models. This creates six different structural groups that are schematically displayed in Figure 2.1. These groups are: deterministic models with single objective, deterministic models with multiple objectives, fuzzy models with single objective, fuzzy models with multiple objectives, stochastic models with single objective and stochastic models with multiple objectives.

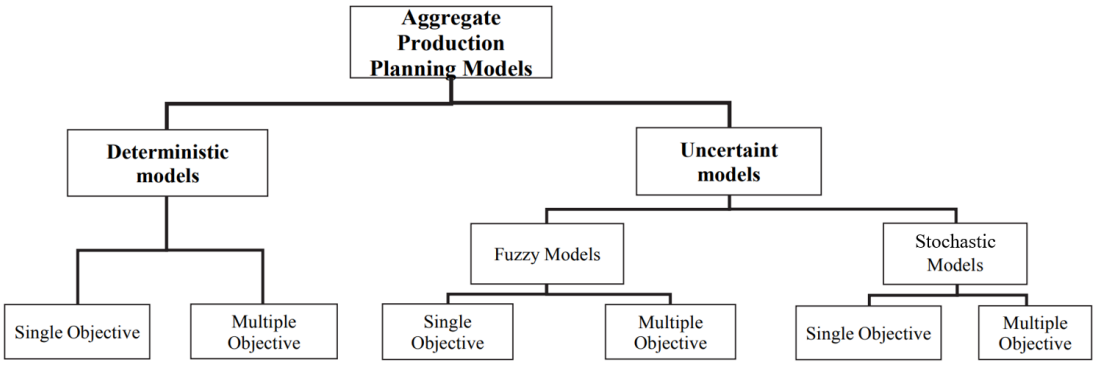


Figure 2.1: APP models' classification structure. Source: Cheraghalikhani et al. (2019)

In APP models formulation, constraints and objective functions are created based on a wide range of input parameters. These parameters are, for example, market demand, production costs, inventory costs, labour costs, subcontracting costs, production rates, maximum labour level and revenue per product. Deterministic models assume that all these parameters are known in advance, making the problem formulation less complex. It is important to point out that many real life

situations will not be accurately represented by this type of model, as the uncertainty of certain parameters is very significant. Customer demand and some types of costs are good examples of parameters that can be imprecise by nature and difficult to forecast across the planning period. Furthermore, some parameters will have incomplete or unobtainable information at the time of the elaboration of the plan. As a result, formulations that deal with these types of uncertainties have been widely studied in the literature. As mentioned above, Cheraghalikhani et al. (2019) divide them into two groups: Fuzzy models and Stochastic Models.

Fuzzy models are generally applied in situations where human judgement, evaluation and decisions are involved. The variables involved are characterized by an attribute called fuzziness, which can be defined as a type of imprecision that lacks well-defined boundaries, commonly associated with decision making, reasoning and learning. These uncertainties can be found in input parameters like fuzzy demands, production capacities with tolerance or imprecise process times. Several authors, like Baykasoglu and Gocken (2010), have developed studies on this subject, highlighting the potential of fuzzy aggregate production planning.

Aside from fuzzy models, uncertainty can also be tackled by stochastic models. These models differ from the previous ones as they are commonly based on randomness and probability concepts, resorting to probability distributions to deal with uncertainties. Simply put, parameters with associated uncertainty are assumed to follow known probability distribution functions. However, according to Lai and Hwang (1992), these models can be computationally less efficient. In some cases, probabilistic distributions may not be able to accurately represent the real uncertainty of a given parameter. Some authors like Jamalnia et al. (2017) further divide stochastic mathematical programming into sub-categories (stochastic linear programming, stochastic nonlinear programming, stochastic multi-objective optimisation and robust optimisation) for more extensive analysis.

Concerning the number of objectives a model takes into consideration, they can be single objective or multi-objective. Single objective models only have one objective function, for example, minimizing total costs. Real world situations, however, usually involve multiple and conflicting objective functions. As stated by Gicquel et al. (2008) these conflicting objectives can be, for example, aiming for an excellent customer service level and, at the same time, minimizing production and inventory costs. Multi-objective production planning aims to find the best trade-off between these opposing goals.

2.2.2 Addressed issues and main considerations of APP models

As mentioned before, the complexity of APP models comes from the need to coordinate interacting resources in order for a company to respond to market demand effectively. Table 2.1 presents the fundamental issues that basic APP models usually take into consideration, alongside a brief description of each one of them.

Table 2.1: Basic issues in APP models. Source: adapted from Cheraghalikhani et al. (2019)

Basic Issue	**Definition**
Market demand	Demand for each period that must be satisfied by production, inventory or backorder
Inventory	Products held in stock in each period
Backorder	Part of the demand that has not been satisfied in the respective period
Production capacity	Maximum amount of products that a system can produce in each period (machines and manpower)
Warehouse space	The capacity of the warehouse for holding inventory
Cost of production	Regular time and overtime production costs as well as the costs of carrying inventory and backorders
Subcontracting	Temporarily hire the capacity of other firms
Labor level	Number of workers in each period (regular and overtime workers)
Hiring and Layoff cost	Additional workers need to be recruited to handle production peaks and redundant workers may be laid-off to reduce overheads
Product Price	Selling price of product

Keeping in mind the basic issues presented, it is also important to understand the basic assumptions made in the simplest APP models. Silver (1976) characterized the assumptions of the most common APP models as follows:

1. Market demand is deterministic.
2. Production costs in any given planning period are strictly linear or are piecewise linear.
3. Costs incurred as a result of changes to production rates in any given period are also linear or piecewise linear.
4. Inventory should be limited over the entire planning horizon.
5. Carrying costs for inventory are known for each period in the planning horizon.
6. Backorders may or may not be allowed.

Models will then take the mentioned considerations and find the combination of production volumes and workforce levels per period that optimizes an objective function like minimizing the sum of all the associated costs. More specific assumptions arise as they are needed in real world planning contexts. Therefore, several other very important, and more specific, issues addressed in these models have already been profoundly studied in the literature, with formulations becoming increasingly complex. A list of those studied issues was compiled in the literature review of Cheraghalikhani et al. (2019), with the most relevant ones being listed below:

- **Labour characteristics**: Labour is a key resource of every APP model and has numerous specific characteristics that, in some cases, are very important to include as they greatly

impact the workforce capacity. The learning curve effect, for example, is very significant in activities that require a lot of manual work, as the time needed to fulfil a job greatly reduces overtime. As workers get more familiar with their tasks, they tend to find more efficient and faster ways of performing them. This has a significant positive increase in the workforce capacity and should, in these cases, be considered in the formulation. Other important aspect to have in mind is related to labour skill and specialization. Basic models assume that all the workers are equally capable of performing the same tasks, which is not always the case. Some APP models have considered different types of labour, allowing a more precise vision of the value brought by each type of worker in terms of the tasks one is able to perform and, possibly, their level of productivity. Furthermore, legal restrictions, like legislation on firing a worker or the maximum overtime allowed for each person, might be important to integrate. Finally, the cost and time spent in labour training per period, as well as the internal policies on labour companies might have, such as efforts to minimize negative impacts on workers' lives by promoting stabilized levels of the workforce instead of frequent hiring and firing, can also be included in the model.

- **Product characteristics**: Some of the predominant product characteristics addressed in APP modelling are the product life cycle, perishability, defectiveness and customer satisfaction level. Implementing a product life cycle indicator in the model allows for a greater understanding of the expected evolution of the sales volume over time and, therefore, better demand forecast. Products in the first stages of their economic life experience a growth in demand while this value will decline as later maturity stages are reached. Similarly, product perishability and defect occurrence are also important factors to include in certain industries where these factors play a major role in planning activities. The customer satisfaction level, which can be defined as the ability of a company to consistently comply with the customer needs and expectations as agreed between parts, is another very important aspect that has been included in APP models. Whether the non-compliance has associated costs that need to be taken into account, or it is the company's objective to increase this metric, there is no doubt that it can be a very relevant element to include.

- **Multiple products, markets and manufacturing plants**: Many examples of APP models in the literature address multiple product families, while more simplified version can include only one. The same happens with markets, as models usually only consider a single market, with a given customer demand and unique sale price. However, many companies have multiple markets to which they can sell the same products, with varying demands and prices. In these cases, it is imperative to include this factor in the model. Lastly, while traditional models tend to assume that all production is made in a single plant, this misrepresents big companies with multiple plants, each one of them with different operational costs.

- **Setup decision**: The action of setting a machine to the required configuration for producing a given product is called a setup. Setups take precious production time, as machines are not

producing, and, therefore, they cost money. For these reasons, several APP models in the literature have included the setup decision as a variable in the model, making it possible to take in consideration the wasted production time and all the associated costs.

- **Financial concepts**: Some studies have presented APP models under financial constraints to model today's worldwide complex financial conditions. Moreover, the time value of money considering the potential interest gains over a given amount of time using compounding interest strategies, should also be considered in some cases.

- **Supply chain concepts**: In some models there is a need to include the network of relationships that transforms intermediate products into finished products. Therefore, special attention has been given to the inclusion of these connections in APP models. This topic will be further developed in Section 2.2.3.1 when multi-level models are introduced.

2.2.3 Solving approaches

When it comes to solving APP models, several approaches have already been extensively studied in the literature. It is important not only to be aware of all those possibilities but also of the individual application of each one of them. In his literature review of APP models, Cheraghalikhani et al. (2019) list a wide range of options and their most common use cases. Metaheuristics, for example, are widely used to solve either deterministic, fuzzy or stochastic models. On the other hand, solving approaches like Heuristics, Linear Decision Rules, Progressive Modelling, Evolutionary Algorithms, Goal Programming and Solver Software are used to tackle deterministic models. Goal Programming and Solver Software are the most frequently used approaches from this last set, with Goal Programming being generally used for multi-objective deterministic models and Solver Software being the go-to approach for single objective models. Lastly, Possibilistic Programming and Fuzzy Programming are used to solve fuzzy models, while Stochastic Programming and Parametric Programming are commonly used in stochastic models.

2.2.3.1 Capacitated Lot Sizing models

The aggregate production planning, in terms of formulation, is very similar to the lot sizing (LS) problems, which have been widely studied in the production planning literature. For that reason, this section will explore its basic formulation and possible extensions.

The main assumptions of the basic capacitated lot sizing models are similar to the mentioned assumptions of the APP models, with slight differences. Production resources are assumed to be limited and can only produce one type of product at a time. Their lack of flexibility results in the need for substantial amounts of setups, to change machine settings between different types of products. Furthermore, demand for all products is considered deterministic and time varying, and needs to be satisfied without backlogging. This means that these plans are built in a way that all the demand is attended in time, achieving perfect customer service level (Gicquel et al., 2008).

This type of model is based on the idea of finding an optimal trade-off between two main types of costs: setup costs and inventory holding costs. As mentioned before, the first relates to the loss of potential production incurred during the time of changing the machine configuration between production sets. Inventory holding costs, in its turn, regard stock keeping costs such as capital opportunity costs and others storing expenses like warehousing and handling labour. On the one hand, if large batches of production are used, setup costs are lower, but inventory costs increase. On the other hand, smaller and more frequent batches will lead to higher setup costs, but lower holding costs as small levels of stock are held from a period to another (Gicquel et al., 2008).

An important dimension to the formulation of lot sizing models is the time structure used for the planning horizon discretization. A classification in big bucket and small bucket models is commonly used. As explained by Belvaux and Wolsey (2000), in big bucket models several setups per period can occur, while in small bucket periods only one setup is allowed. In other studies, such as Guimarães et al. (2014) these classifications are called macro-period and micro-period models, respectively.

CLSP - Capacitated lot sizing problem

The capacitated lot sizing problem (CLSP), in terms of the time discretization structure, is a big bucket model and can be considered the traditional formulation of the lot sizing problem (Guimarães et al., 2014). It assumes that only one machine is available and that raw material directly transforms into finished products, with no intermediary items. The goal of the classical CLSP is to determine, for each period in the planning horizon, the optimal production quantity (lot size) for each item, in a way that the total cost of the production plan is minimized.As described by Gicquel et al. (2008), a MIP formulation for this problem is presented bellow. The formulation plans a set of N items, indexed by $i = \{1, ..., N\}$, over a planning horizon of T periods, indexed by $t = \{1, ..., T\}$.

The input parameters of the model are the following:

- D_{it}: Deterministic demand for item i in period t, [in units];
- P_t: Available capacity on the resource in period t, [in time units];
- v_{it}: Capacity needed to produce one unit of i in period t, [in time units];
- h_i: Holding costs per unit and period for item i, [in monetary units];
- c_{it}: Setup costs for item i in period t, [in monetary units].

The next step is to define the decision variables, which will be the output of the model:

- I_{it}: Inventory level of item i at the end of period t, [in units];
- x_{it}: Production quantity for item i in period t, [in units],

- y_{it}: Binary setup variable. Takes the value of 1 if the resource setup is done for item i in period t, and 0 otherwise.

Finally, the formulation of the objective function and the model constraints is as follows:

$$min \sum_{i=1}^{N} \sum_{t=1}^{T} (h_{it} I_{it} + c_{it} y_{it}) \tag{2.1}$$

$$I_{it} = I_{i,t-1} + x_{it} - D_{it}, \ \forall i,t \tag{2.2}$$

$$v_{it} x_{it} \leq P_t y_{it}, \ \forall i,t \tag{2.3}$$

$$\sum_{i=1}^{N} v_{it} x_{it} \leq P_t, \ \forall t \tag{2.4}$$

$$I_{it} \geq 0, \ \forall i,t \tag{2.5}$$

$$x_{it} \geq 0, \ \forall i,t \tag{2.6}$$

$$y_{it} \in \{0,1\}, \ \forall i,t \tag{2.7}$$

Equation 2.1 is the model objective function, which is to minimize the sum of the holding and the setup costs. The inventory balance is ensured by Equation 2.2, while constraints 2.3 enforces the occurrence of a setup if the product i is to be produced in period t. The role of constraints 2.4 is to prevent the production time for each period to exceed the available machine capacity. Together, constraints 2.2 and 2.5 prevent backlogging, while constraints 2.6 impose the non-negativity condition on the production quantities. Lastly, constraint 2.7 defines the setup variable as binary.

It is important to note that, in the CLSP, different products can have different production rates, as the parameter v_{it} gives the time consumed by each product on the resource (machine).

CLSP extension: Multi-level and Multi-resource

Lot-sizing models vary in complexity with the number of resources and the number of production levels considered. If the products to be planned can only be made in one resource (i.e. machine), the model is single-resource, whereas if multiple machines are possible, a multi-resource model is used. Multi-resource models add a new layer of complexity to the process as a new decision, related to the assignment of production lots to machines, has to be made.

In addition to the number of resources, the number of levels also influences the complexity of LS models. A production system can be single-level or multi-level. Single-level systems are systems where the finished products are directly obtained from the raw materials. In multi-level models, on the other hand, there are intermediate products between the raw material and the finished products. This last case creates the need to map the parent-component relationships between products. Moreover, while in single-level models demand can be directly assessed from confirmed orders and forecasts, in multi-level models the demand for one level depends on the lot sizing decisions made for the following levels and incorporation factors need to be considered.

Regarding the modelling techniques of parent-component relationships between items, given by the bill of materials in multi-level models, different types of product structures can be identified. Serial product structures, for example, relate to cases where each item has a single predecessor and a single successor in the network. In other cases, when a product can be made from several precedent components and each item only has one successor, a product assembly structure is used. Not rarely, though, in the context of various industrial firms, each item can have multiple predecessors and multiple successors. This is called a general product structure and results in several end-products having the same components, further complexifying the problem (Gicquel et al., 2008).

The CLSP is a single-level multi-resource model, thus being often insufficient to represent real world situations where intermediary products exist and parallel machines are available. Several studies have been conducted in the field of multi-level and multi resource lot sizing models, to bridge that gap. It is possible to verify that most authors tackling this issue use big bucket models, which makes it possible to consider these approaches extensions of the basic CLSP formulation.

Chapter 3

The Problem

This chapter introduces the problem addressed in this project. Section 3.1 contains a brief explanation of the problem at hand and highlights the opportunities for improvement identified by a previous diagnosis project in the company. Section 3.2 provides an overview of the company's planning context. Firstly, the production process is described in section 3.2.1, followed by a description of the company's current practices for production strategy definition and medium-term planning in section 3.2.2. Next, in section 3.2.3, a set of preliminary analysis and performance indicators of the AS-IS situation are presented. Finally, section 3.3 describes the main steps of the proposed approach to tackle this challenge.

3.1 Problem description

Aggregate production planning is a pivotal process for several industrial companies. This type of capacity planning allows companies to plan the resources they will need in the medium term. Knowing in advance the optimized production volumes, workforce and inventory levels for each period to meet the fluctuating demand leads to more efficient management of the production resources, such as material requirements planning, for example. Productive performance and inventory management have naturally opposite objectives. Therefore, a holistic approach such as this one is of the utmost importance. While the production of small batches benefits inventory management by keeping lower stock levels, it makes the production process less efficient as more setups are needed.

The trade-off between holding and setup costs is also the main driver of the production strategy definition. The correct separation of the products into the MTS and MTO classification is vital to achieving good levels of performance in a hybrid production system. A wrongly classified MTS item takes unnecessary inventory space with items that may never be sold, and an improperly classified MTO item needlessly generates unwanted stress on the production line, incurring, respectively, in higher holding costs and setup costs. For that reason, this project also aims to provide a decision support system that suggests the most suitable production strategy for each product and also, in the case of the MTS items, to define the minimum batch size and safety stock.

These decisions have to take in consideration the desired service level and the goal of minimizing a global cost function composed by inventory costs and operational costs.

As stated before, in the company where this study was undertaken, several improvement opportunities were identified by a previous diagnosis project. Regarding the production strategy definition, despite already operating under a hybrid system, items were classified as MTS or MTO based on an existent commercial definition, which will be further explained in section 3.2.3. Consequently, the process was too empirical, lacking an exact understanding of the impact of decisions on costs for the business. The methodology was also inconsistent between products as there were lots of exceptions to the established rule. The definition of safety stocks and minimum production quantities (MPQs) for MTS products was not enough analytically backed. Furthermore, some relevant dimensions like the impact that changing the strategy of a given product has on the whole system were not taken into consideration. Instead, a single product analysis that did not consider line congestion effects was used. All these factors combined with the inexistence of a decision support system for this purpose, highlighted the need for the development of such tools.

Concerning the medium-term planning in the company, the diagnosis concluded that master production and tactical plans were not systematically generated on the basis of a global cost-service function. Instead, single constraint analysis were performed in a rather *ad-hoc* and incomplete way, thus not reaching optimum cost. In addition, the MTO forecast was not included as reserved capacity in the plans and a very reduced planning horizon was used. These factors made it difficult to plan production capacity with the anticipation needed to manage all the involved resources. Hence the company will, undoubtedly, benefit from an optimization model capable of generating master production plans, while respecting all production constraints and minimizing a global cost function. This can be made in a 12-month rolling horizon and enables the company to generate plans more frequently and test different scenarios.

3.2 Planning context

The company under study has two filling plants with a total of 22 production lines and a range of more than 1200 products across more than 20 different brands. Given that the company's product assortment is divided between both facilities, with each product being produced in only one factory, it is possible to separate production planning between each factory. Nonetheless, the company's planning context is still very complex, with multiple dimensions to be considered. Some of those features of the planning context are listed below.

1. Demand patterns are variable during the year (high seasonality);
2. Multiple lines with distinct bottlenecks;
3. Different customers can have different order's lead times;
4. Wine and dry goods (e.g., bottles, corks and labels) have significant preparation lead times;
5. High workforce availability and flexibility (every worker can perform any job).

All factors considered, pressure on the supply chain may cause capacity problems, both in internal operation but also on suppliers capacity to replenish the business. Hence, proper medium term capacity planning and the correct production strategy definition for each product are very important and have to take in consideration all these aspects.

There are also business restrictions that derive from the nature of the business itself. Being a wine company, there are many enology restrictions and considerations that need to be addressed to ensure quality of the finished product. Specific types of wine, like Port wine, for example, have different production processes and aging stages between the harvest of the grapes and bottling procedure. However, in the context of this study we were only concerned with the process after the wines arrive at the filling plant. At this point, the constrains include, for instance, bottling schedules that vary with the wine type. White wines, for example, cannot be bottled in the second semester due to the higher temperatures and the same happens for all consumer wines in the harvesting period, as the enologists are occupied with vineyard duties, and cannot be present in the filling plant as required.

Regarding the demand forecast, before the present study, a project was undertaken in the company to develop a forecasting model that predicts the demand of a 12-month time horizon. The output of this tool is the quantity of each SKU, in wine boxes, that is expected to be ordered in each month. This is an important aspect of the planning context as demand can be considered deterministic, in the planner's perspective.

3.2.1 Production environment and process

Each of the company's roughly 1200 SKUs corresponds to a box of bottles of wine from a certain brand and wine type. The wine was harvested in a given year, and the item is destined for a specific market (e.g. country). They also differ in the number of bottles and the capacity of each bottle. In addition, each SKU has a specific type of pallet in which it will be delivered to the customer that varies with the market it is destined to. It only takes one of the previous aspects to be different for two SKUs to be considered different. This defines the highest level of granularity the company works with, and each of these products has its own assigned code.

However, for certain uses within the company's structure, it is useful to aggregate the SKUs in groups of similar characteristics. For that purpose, the company uses an aggregator code system that groups items of the same wine type and brand, that have the same bottle capacity and the same number of bottles per box. This means that as long as two SKUs have these characteristics in common, they have the same aggregator code, even if they came from harvests from different years or if they are palletized on a different pallet. This aggregator code will hereafter be referred simply as Aggregator.

The production process in the filling plants has two main steps that all finished items must undergo: Bottling and Finishing. The bottling process is the first one and consists of filling a glass bottle with the wine and placing the cork. The finishing process is the final step of the production and consists of applying a label, that depends on the target market, and packing the bottles in the correct boxes. These two stages can immediately follow one another in the same production line,

a process hereafter referred as direct production, or they can be performed on different occasions. In the latter, once the bottling phase is complete, the bottles with no label are stored in the wine cellar of the production plant, to be finished later. In some cases, big bottling batches are used (1 or 2 fillings per year) and then the items are labeled as needed. The unlabeled bottles stored at the wine cellar are called semi-finished items and have their own specific code, different from the ones the final products have. This is fundamental as the same semi-finished product can originate different finished products, since it can be labeled for different markets or integrate wine boxes with a different number of bottles, for example.

There can be a third stage that is only required by some products, and it is called manipulation. The manipulation stage consists in taking a finished product and manually adding something to complement it. For example, some manipulated products consist of finished wine bottles to which is added a wooden box and a wine glass to improve the final package. This is a very time-consuming stage, as workers have to handle these items one by one. The complete production process is represented in Figure 3.1.

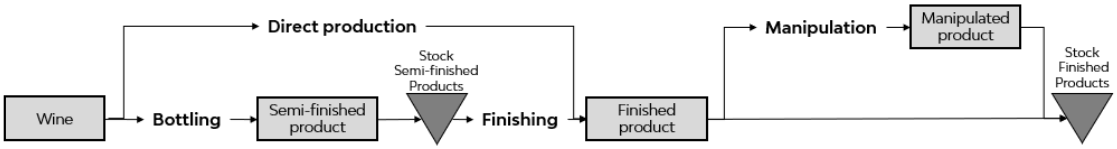


Figure 3.1: Production process

While semi-finished products are stored in the wine-cellar of the production plant waiting to go through the finishing process, the finished products are stored at a customs warehouse under the responsibility of a logistics partner. From there they are delivered to the customers.

As already stated, within the same aggregator, the exact same items can be found in different pallets. In many cases, when a given pallet has low demand, which will consequently lead to short production batches, it is preferable to produce that same quantity in a pallet that already has high production volumes. When this happens, at the finished products warehouse, the products are moved to the correct pallet as needed. The pallet with the highest production volume is referred as production pallet, while the other one is the transfer pallet. If the entire content of one pallet is changed to another, the process is called a transfer. On the other hand, if only a few items are needed in the other pallet, they are moved one by one, and the process is simply called picking. The safety stock of the transfer pallet is always kept in the production pallet, and only the quantities needed to fulfil incoming orders are moved to the correct pallet. The decision on which pallets to classify as production pallets and which to classify as transfer pallets involves, therefore, a trade-off between the picking and transfer costs and the cost of the extra setups in the case of producing

directly in that pallet. Today this decision is made empirically by the planning department of the company.

In addition to the different pallets, within the same aggregator code, there can also be harvests of the same wine from different years, also called variants. Regarding this aspect, it is important to make sure that once a newer variant is available, the older ones are no longer produced. If, eventually, the demand of an older variant is greater than its available stock, then it will have to be fulfilled from the production of the most recent one.

Another very important aspect of the production process is that the same product can be made in different production lines. At the same time, the same production line, can make several different products. This fact, combined with the existence of intermediate products that can originate more than one final product, make this production environment a multi-resource and multi-level problem.

3.2.2 Current practices

During the diagnosis project, the processes currently followed in the company were mapped to clearly understand all the procedures and the reasons behind the choices they made.

The current process used by the company to define the production strategy for each product is heavily dependent on a commercial definition. This definition consisted of the separation of the products into two categories: core and non-core. The current process is mapped in Figure 3.2.

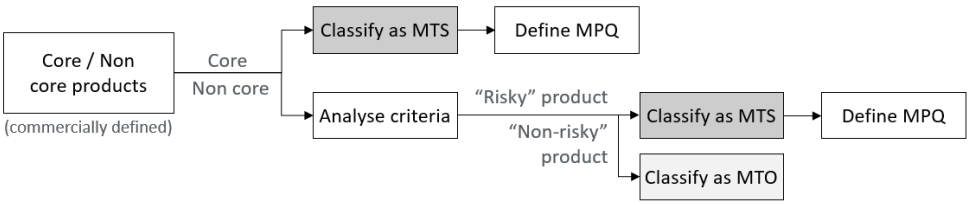


Figure 3.2: Production strategy definition process

Firstly, the commercial department decides whether a product should be considered core and non-core based on their experience of what products are crucial to the company, based on their historical demand and on the criticality of not complying with the lead times established by some customers. For core products, the company commits to a lead time of up to 5 working days, while for non-core products, this value increases to 15 days. Core products are immediately classified as MTS and an MPQ (minimum production quantity) is defined. In contrast, if the product is non-core, a series of criteria are evaluated to assess the risk associated with not producing this item to stock. The factors taken into account to perform this evaluation include the productive flexibility of the item, the risk of replenishment of the components and the wine, the sales volume and stability, and the requirements of its final customers. If the non-core item was deemed "risky", then it

would also be made to stock and MPQ was defined. On the other hand, if the commercial risk was considered to be low, the item would be classified as MTO. Certain types of wines, corresponding to 17% of total production, were an exception to this rule, with production strategy and quantities being defined *ad-hoc* by the commercial department. This process was repeated every six months and the result was a list with the production strategy (MTS or MTO) for each SKU, and the MPQs of the make-to-stock items.

Regarding the production planning, as mentioned in section 3.1, a very small planning horizon was used, considering the involved decisions. The most comprehensive layer of planning in the AS-IS situation was a one-month planning horizon, with the process being illustrated in Figure 3.3.

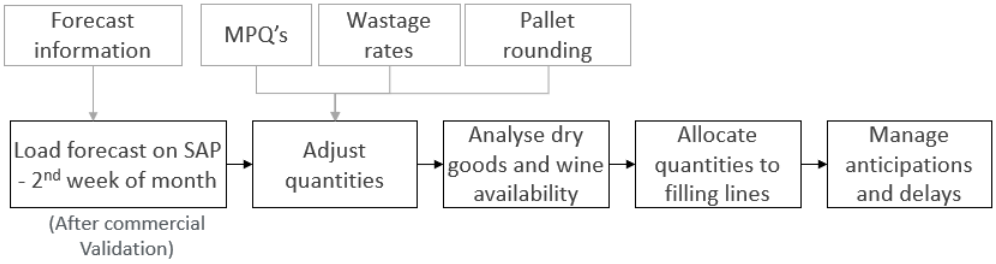


Figure 3.3: Current production planning process

In the beginning of the second week of each month, the planning department received the forecasted demand for the next 4 weeks, delivered by the forecast tool. That information was then loaded on SAP, the ERP used by the company, and an adjustment to the quantities was performed. This adjustment aims to account for the defined MPQs and the wastage rates, and also to make sure the quantities are rounded to the pallet. The next step is to analyze the dry goods availability and confirm the proposed plan is feasible in terms of material requirements. Then, production lots are allocated to production lines, having in mind the need to optimize the orders' due dates fulfilment and to minimize setups. Finally, there is a last layer of planning where the planning department can use the remaining capacity to allocate production resources to manage anticipations and delays. It is at this stage, for example, that the seasonal needs for manipulated products are leveled. As previously mentioned, manipulated products are very time consuming and require high levels of dedicated labor. Since these products have a seasonal demand peak in the last months of the year, it is crucial to anticipate their production. This is made to flatten resource utilization throughout the year and also because it is most likely not possible to produce everything during the demand peak.

As the main trade-off studied in this project is between the holding costs and setup costs, it is important, now that the current process is explained, to clarify the components involved in these costs. The inventory cost includes a unitary cost per pallet charged by the logistics partner in the finished product warehouse, a unitary cost for each 9 liters of wine in the wine cellar, which includes the warehouse related labor, and the capital opportunity cost of the total inventory value. The setup costs, on the other hand, have three main components. The most significant one is the

wine wastage cost, which concerns the wine that is still in the pipe system of a given production line in the transition between two production runs of different wine types. Even though this wine is not entirely lost, it is redirected to a reservoir of lower quality wine, being valued at a lower price. In addition to this, the nonproductive time lost in this machine setting operation must also be considered and includes the cost of operating the machine itself and the manpower required to perform the operation.

3.2.3 Preliminary analysis

At the beginning of the project, a set of preliminary descriptive analysis was performed to understand the planning context and the current performance of the company. The focus fell on three main aspects: Demand Pattern, Service Level and Stock.

Demand patterns

When analyzing the demand characteristics, a strong Pareto effect is noticeable. An ABC analysis, with sales data from the previous year, concluded that 10% of products concentrate 80% of the total sales. This analysis is depicted in Figure 3.4 and leads to the conclusion that the company's product range has very different demand patterns, allowing the classification of items according to the demand level.

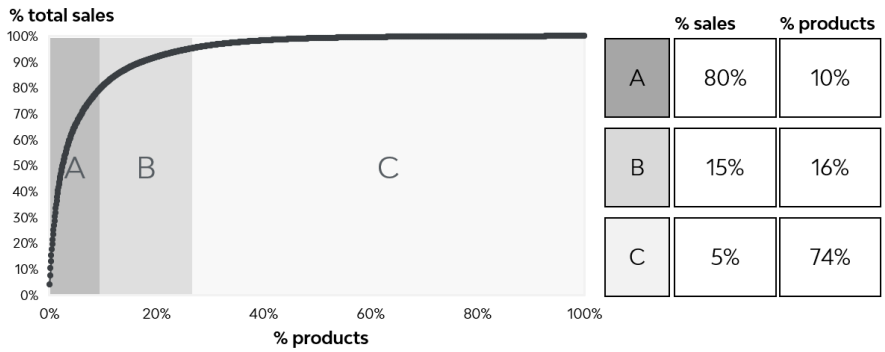
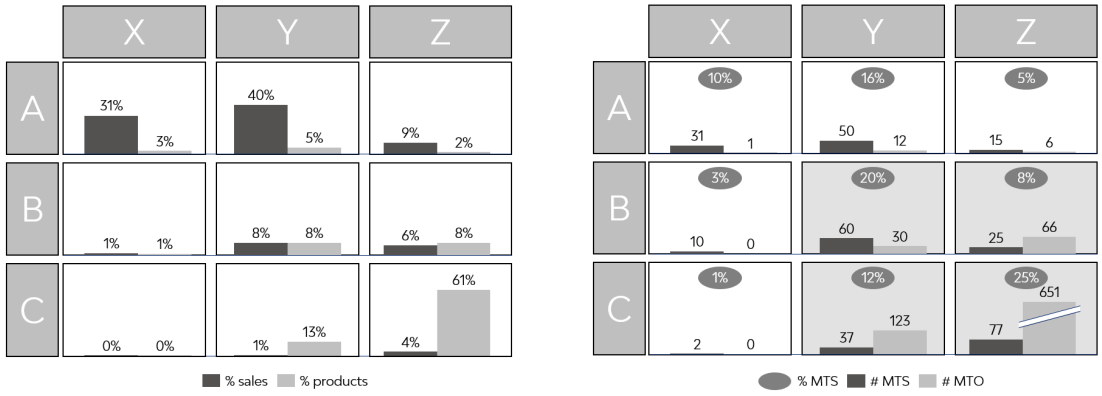


Figure 3.4: ABC analysis

To further complete the assessment on demand, a XYZ analysis was performed. The XYZ analysis is a way of classifying inventory items according to the variability of their demand, with X items being characterized by a steady demand with little variation and Z items having strong demand fluctuations. The indicator used to access demand variability is the Coefficient of variation, CV. Products with $CV < 0.7$, $0.7 < CV < 1.5$ and $CV \geq 1.5$ were, respectively, classified as X, Y and Z items. Figure 3.5a combines this analysis and ABC analysis into a single graph, displaying the percentage of products, and its corresponding percentage of sales, belonging to each of the quadrants formed by the combinations of these two classifications. From this analysis, it can be

concluded, for example, that 3% steady products are responsible for 31% of the sales, whereas 61% highly volatile products lead to 4% of the sales.

It was interesting to see how the current MTS-MTO production strategy classification fit into the quadrants of the previous analysis. Figure 3.5b shows the same double-entry table as in Figure 3.5a, but this time the information in each quadrant exhibits the number of MTS and MTO items in each division. It also has the percentage of the total MTS products in each group. As highlighted in the figure, approximately 65% of MTS products are in categories with reduced sales volume and high sales volatility. This fact further stresses the potential for improvement of the current practices since a very significant part of the stock is concentrated on the products with fewer sales.



(a) Percentage of sales vs. Percentage of products

(b) MTS vs. MTO

Figure 3.5: ABC-XYZ analysis

Service level

The global service level of the company in the AS-IS situation is 93.6%, but the company's goal is to reach 96%. Order lead time, which is the difference between the requested date and the order creation date, is one of the factors that impacts service level the most. The longer a customer is willing to wait for a product, the higher the probability of fulfilling it on time. The shorter the order lead time of a given product, the more likely the company will have to keep it in stock to be able to respond on time to the demand. As the current definition of the production strategy was based on the core, non-core commercial definition, it was pertinent to verify the orders lead time of these two groups. Figure 3.6 shows the percentage of the total number of orders from both core and non-core items, that were placed with a given order lead time. As we can see, non-core orders were placed, on average, with a lower lead time, which further comes to prove there might be potential for improvement as these are the items being classified as MTO. The average customer lead time was 7.2 and 10.0 working days for, respectively, non-core and core orders. The same analysis is made in Figure 3.7 for the current MTS-MTO classification. Here we can see that MTO orders are placed, on average, with a lower lead time, with the average customer lead time being 5.3 and 10.5 working days for, respectively, MTO and MTS orders.

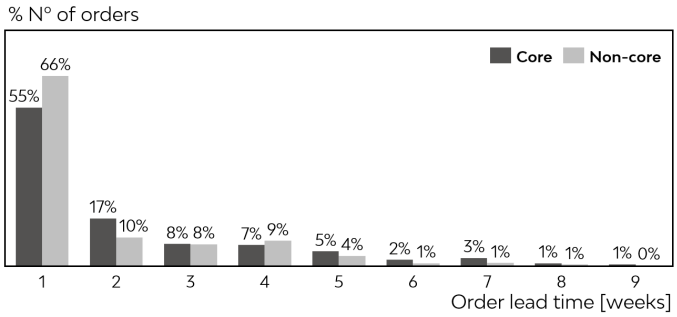


Figure 3.6: Orders lead time: Core vs. Non-core

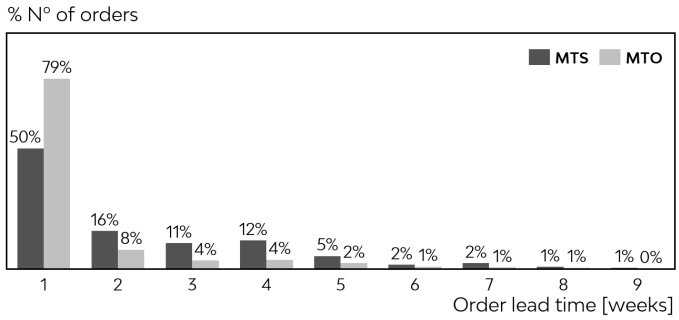


Figure 3.7: Orders lead time: MTS vs. MTO

Stock

The company's stock level at the time the analysis was made, in the beginning of the project, is represented in Figure 3.8, with the already familiar division of the products in the ABC and XYZ classifications. Besides the quantity kept in stock in each group, the average product rotation was also calculated and is represented within the circles in the table. It is possible to conclude that approximately 36% of the current stock is composed by products in categories with reduced sales volume and high sales volatility (highlighted cells).

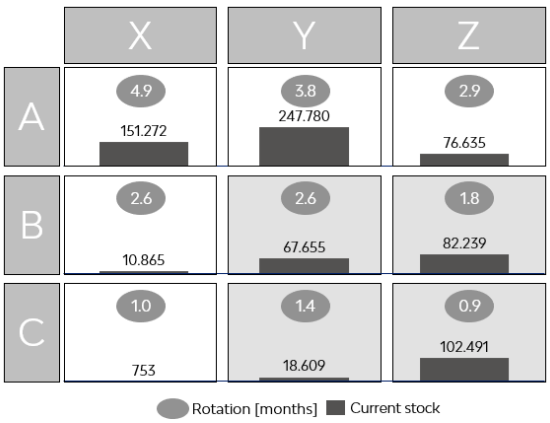


Figure 3.8: Stock distribution

3.3 The proposed approach

Given the improvement opportunities identified in the diagnostic project, the approach to the problem was divided into two modules. The first module addressed the definition of the production strategy for each product. The final goal for this module was to create a list that indicated the best policy (MTS or MTO) for every item and, in the case of MTS products, provided the corresponding safety stock and MPQ. The second module aimed to create a 12-month rolling horizon production plan, to plan capacity and production resources in the medium term.

The proposed approach for the production strategy module followed a four-fold methodology, as seen in table 3.1, that started by mapping the AS-IS production strategy definition process and identifying all the variables involved in the decision. Then, in the second step, all the relevant information was gathered, analyzed and treated to be supplied to the model. Next, the third step was to model the congestion effects, the costs involved in the change of strategy and the impact a given change has in the whole system. Then, in the fourth step an algorithm (heuristic) was developed with the goal of defining, in a near-optimal way, the production strategy for each product that minimizes the total cost of the system. Finally, on the last step, an Excel visualization dashboard was developed to evaluate the model results, identify potential issues, and polish the approach.

Table 3.1: Proposed approach for the production strategy module

Process mapping	Map the current production strategy definition process
	Identify all the involved variables
Data Treatment	Data request
	Data analysis and preparation
System modeling	Model congestion effects through a queueing system
	Identify and quantify the costs of each strategy
Heuristic development	Minimize total costs
	For each strategy change identify the expected savings
	For each product define the best production strategy
Results visualization	Create an interface to analyze the final MTS-MTO list

Regarding the aggregate production planning module, the proposed approach was divided into five steps, as seen in table 3.2. It aimed to obtain optimal medium term production plans that minimize the company's cost structure. This approach was very similar to the one used in the previous module.

In the first step, a thorough analysis of the existing processes was performed, enabling the mapping of the current production planning methodology and all the steps in the production pro-

cess. At this stage, all the operational and business constraints and particularities were identified so that they could be included in the model. The second step, similarly to what was done for the first module, was to gather all the necessary input data to the optimization model and also performing some preliminary analysis to assess the AS-IS condition of the company.

The third and main step of this module was the development of the optimization model. First, the gathered information had to be structured in a format that could be used as input to the model. Then, the model constraints and objective function were formulated based on the information collected in step 1, with the end goal being the minimization of total production and inventory costs.

Finally, in order to analyse the generated plans, an excel interface that received and transformed the model output files was created. The goal was to have a display tool that presented the model results in a detailed and clear way, allowing the user to understand the recommendations of the model. By doing this, quick problem identification and effective model refinement were possible.

Table 3.2: Proposed approach for APP module

Process mapping	Map the current planning procedures
	Analyze in detail the production process
	Identify all the operational constraints
Data treatment	Data request
	Data analysis
Optimization model development	Create an input structure with all necessary data
	Implement all the model constraints
	Minimize production and inventory related costs
	Provide a 12 month production plan and present the production resources needed to fulfil it
Results visualization	Create an interface to analyze the proposed production plans

Chapter 4

Methodology

This chapter elaborates on the methodology developed to tackle the production strategy definition problem and the production planning problem presented in the previous chapter. The first is presented in section 4.1, with the main assumptions and characteristics of the approach being presented in section 4.1.1. Next, sections 4.1.2 and 4.1.3 describe, respectively, all the involved formulation and the developed heuristic. The aggregate production planning methodology is presented in section 4.2 and starts by outlining the main features of the model, as well as the assumptions it relies on. Section 4.2.1 presents the model formulation, elaborating on the model's objective function and the implemented constraints.

4.1 Production strategy definition

The final output of the production strategy definition module is a list with the production strategy (MTS or MTO) for each of the company's products. Additionally, it has to define the minimum production lots and safety stocks of the MTS items. In fact, some MTO items may also have a minimum production lot defined due to a limitation imposed by the company, regarding the number of setups. This topic will be clarified in section 4.1.2

The end goal is to determine the combination of production strategies that leads to minimum global costs, while ensuring that the service level the company desires is reached. They had already a considerably high service level of 93.6%, indicating that they have been able to deliver most of their orders on time. Nevertheless, the company's objective is to reach a 96% service level, which may be achieved with the help of this decision support tool while, at the same time, optimizing inventory and operational costs.

The first step of this module is detailed in the previous chapter and consisted in mapping the current production strategy definition process and identifying all the involved variables. Once this was done, one important decision had to be made, which was to define the level of granularity at which the model was going to operate. Until this moment, only two levels of granularity have been discussed, the SKU and the Aggregator. For this purpose, a solution in the middle was adopted, with the decision being made at the level Aggregator+Pallet. This is the result of the separation

of the SKUs contained in each Aggregator in groups according to their pallet. To put it simply, within each Aggregator+Pallet the only varying factor is the year of the harvest, which does not influence the production strategy decision.

Since the company suffers from high demand seasonality, experiencing much higher levels of demand in the second half of the year, the production strategy should be reviewed at the beginning of each semester. This means that the proposed methodology will consider a 6-month time horizon in which the production strategy of each product will remain the same, until the following review.

4.1.1 The model

The formulation adopted in this project was mostly based on Rajagopalan (2002)'s work, with slight differences arising from the particularities of the company's production context.

One of the most important aspects to incorporate in the decision support tool was a measure of the impact that a change in one product's strategy has in the whole system. For this reason, congestion effects were incorporated by modelling the queue of each production line through a M/G/1 approximation. As explained before, multiple production lines are available, hence the occupation rate of each one of them must be taken into account. For the purpose of this module, only direct production was considered and each line was assumed to be a one-stage system with the throughput being given by its bottleneck. Customer orders, whether MTS or MTO, are fulfilled in a first-come-first-served (FCFS) basis. In this environment, the more products are classified as MTO, the higher the corresponding lines' occupation rates, due to the higher number of required setups. This results in higher mean queue waiting times with greater variability, which consequently increases the global production lead times and deteriorates the service level.

The inventory replenishment policy for the MTS items considered in the presented approach uses a (q, r) system which means that when the stock of product i goes down to a level r_i, a production order for q_i units is initiated. This was already the inventory control policy used by the company with the r_i replenishment level being the safety stock of each product. However, even though replenishment requirements are triggered immediately by the ERP, it can take up to a week for the planning department to respond. Therefore, a conservative review time of one week was also considered to ensure that safety stocks are met.

The Economic-order-quantity (EOQ) was used to determine the minimum lot size of each product. The EOQ is the batch size that aims to minimize the costs associated with maintaining inventory and with machine setups, while meeting customer demand (Sanni et al., 2020).

In addition to all the considerations stated above, the approach was developed under the following assumptions:

1. Demand is deterministic and directly extracted from the forecasting tool used by the company.
2. Setup times and processing times are deterministic.
3. Setups are sequence-independent.

4. Raw materials needed for production are immediately available.
5. Lot sizes can be rounded to a full pallet without loss of optimality.
6. Given the large number of items, we approximate the arrival process for the queuing system by a Poisson process, even if this may not be appropriate when system utilization and the coefficient of variation in arrivals are high. (Albin, 1982).

As demonstrated by Rajagopalan (2002), an optimization model based on MIP formulation of this problem results in a non-linear mixed-integer program, which is very difficult to solve since it is not generally convex. For this reason, it was decided to use an heuristic that would lead to good results in a shorter period of time. The next section will elaborate on the formulation used to model each step of the problem.

4.1.2 Problem formulation

Before diving into the heuristic, it is necessary to understand the formulation used to calculate each one of the parameters involved. First, we will present the formulation used to account for the congestion effects, and the calculations used to determine the safety stocks and batch sizes of MTS items. Then, the definition of the inventory and setups costs is explained, bearing in mind the company's cost structure. The variables and parameters used in the formulation are presented in Table 4.1. The unit of time considered in the following calculations is a month.

Production lots and average number of setups per period

As mentioned above, to determine the batch sizes of MTS items, the EOQ was used in order to obtain the quantity that minimizes inventory and setup costs while fulfilling the forecasted demand. Thus, the lot size of each production, LS_i, is given by equation 4.1. C_{Setup_i} and $C_{Inventory_i}$ correspond, respectively, to the cost of setup and the inventory cost of a given product per unit of time. The determination of both these costs will be explored in the costs segment of the present section.

$$LS_i = \sqrt{\frac{2 \cdot D_i \cdot C_{Setup_i}}{C_{Inventory_i}}} \tag{4.1}$$

Once the batch size is determined, its value is rounded to a full pallet by the requirement of the company. Next, knowing the final batch sizes, it is possible to estimate the number of batches per unit of time, λ_i. In the case of MTS items, this number is obtained by dividing the average forecasted demand per unit of time, D_i, by the batch size, LS_i (equation 4.2). The MTO items, on the other hand, by definition do not have a defined production lot since they are made to fulfil the specific orders that arrive at the company. For this reason, the number of production orders per unit of time for MTO items is estimated by dividing their expected demand per unit of time, D_i, by the average order size, $Avg_\ orderMTO_i$ (equation 4.3).

$$\lambda_{i_MTS} = \frac{D_i}{LS_i} \tag{4.2}$$

$$\lambda_{i_MTO} = \frac{D_i}{Avg_\, orderMTO_i} \tag{4.3}$$

Table 4.1: Table of notation for the production strategy problem formulation

Parameters	
D_i	Average demand per unit of time of product i
σ_i	Standard deviation of demand per unit time for product i
α_i	Setup time of product i
β_i	Processing rate of product i
$Avg_\, orderMTO_i$	Average order size of MTO product i
R	Stock revision period
k	Inverse of the cumulative normal distribution function computed for the desired service level
h_i	Holding costs per unit of time of one unit of product i
UC_i	Unitary cost of product i
$wastage$	Quantity of wasted wine, in liters, during the setup of a given production line
WV_i	Value of each liter of the wine of product i
DV_i	Downgraded value of each liter of wasted wine of product i
MAQ	Hourly machine operating
$SetupWorkers_i$	Number of workers required to perform the setup operation for product i
Avg_labour_cost	Average hourly labour cost per worker
$WACC$	Weighted average cost of capital
Variables	
LS_i	Lot size of product i
λ_i	Number of batches of product i per unit of time
λ	Total number of batches per unit time for all items
T_i	Average order processing time of product i
w	Waiting time in the queue
ρ	Line utilization
ss_i	Safety stock of product i
$Stock_i$	Total stock of product i

There is, however, an exception to the non-existence of lot sizes in MTO products. The company did not want to execute more than two setups of the same product each month. For this reason, if the expected number of setups per unit of time of an MTO item, calculated using equation 4.3, is higher than two, a lot size is created for that item, in order to comply with this constraint. Accordingly, this lot size is given by $LS_{i_MTO} = \frac{D_i}{2}$ and these items are regarded as non-pure MTO's, even though we will continue to include them in the MTO group.

Congestion effects

The formulation of the congestion effects is the same for each one of the production lines. Thus, in the following explanation, no differentiation will be made and the presented calculations refer to the process applied to one production line.

To model the congestion effects through a queuing system, we first need to calculate the average order processing time, T_i . For MTS items, this is given by equation 4.4 and consists of adding the setup time, α_i, to the production time, LS_i/β_i . In the case of MTO items, the size of the production lots is approximated by dividing the forecasted demand, D_i by the expected number of setups, λ_i, which, as shown by equation 4.3, corresponds to the average order size, $Avg_orderMTO_i$. Therefore, their average order processing time is calculated as shown in equation 4.5. Despite not being presented here, this last equation has also been adapted to the case of the non-pure MTOs discussed earlier, which was done by simply replacing the lot size expression D_i/λ_i with $D_i/2$. As there are multiple lines in which a single product can be produced, an assumption had to be made in order to define the line in which the item would be produced. This was important in order to know the correct production rate to use for each product. Thus, for the purpose of this calculation, the line which historically was the most used for each product, hereafter referred as the preferential line, was used. Furthermore, as in this module we are only assuming direct production, the production throughput, β_i, considers the full production process, from the filling to the finishing of the product.

$$T_{i_MTS} = \alpha_i + LS_i/\beta_i \tag{4.4}$$

$$T_{i_MTO} = \alpha_i + (D_i/\lambda_i)/\beta_i \tag{4.5}$$

Now, Let $\lambda = \sum_i \lambda_i$ be the total number of orders per unit time for all items, to be processed in a given line. The mean service time is given by:

$$\bar{T} = \mathbb{E}(T) = \frac{\sum_i \lambda_i T_i}{\lambda} \tag{4.6}$$

and:

$$\mathbb{E}(T^2) = \frac{\sum_i \lambda_i T_i^2}{\lambda} \tag{4.7}$$

Then the mean waiting time in the queue in a $M/G/1$ system is (Allen, 1990):

$$\mathbb{E}(w) = \frac{\lambda \mathbb{E}(T^2)}{2(1-\rho)} \tag{4.8}$$

where ρ corresponds to the line utilization and is defined as $\rho = \sum_i \lambda_i T_i$. Observing that $\mathbb{E}(T^3) = \sum_i \lambda_i T_i^3 / \lambda$, the variance in the waiting time in the queue is given by:

$$v(w) = \mathbb{E}(T^2) + \frac{\lambda \mathbb{E}(T^3)}{3(1-\rho))} \tag{4.9}$$

From equations 4.8 and 4.9 we can obtain the parameters for the distribution of the waiting time in the queue, which will be needed to calculate the safety stock of MTS items.

Holding Costs

To calculate the holding costs of MTS products, we first need to calculate both the cycle stock and the safety stock, which together form the total stock level of finished products. As full direct production is being assumed, the stock of intermediate products stored in the wine cellar will not be considered in this module. The safety stock is the component that accounts for the variability in the system, allowing the company to respond to lead time and demand fluctuations while still ensuring the desired service level. The average production lead time for a batch of the item i is given by ($LT_i = \mathbb{E}(w) + T_i$) and the variance is v_w. As it has already been mentioned, accounting for the fact that the company the stock revision is not exactly a continuous process, a revision period, R, was also added to the equation. Therefore, the mean demand during the lead time and revision period for item i, is given by ($D_i(LT_i + R)$) and its safety stock is defined as:

$$ss_i = k\sqrt{(LT_i + R)\sigma_i^2 + v(w)D_i^2} \tag{4.10}$$

where k is the inverse of the cumulative standard normal distribution function computed for the desired service level.

In contrast, the average cycle stock of a certain product is simply the result of $\frac{LS_i}{2}$. Therefore, the average total stock level of a MTS item is defined as $Stock_{MTS_i} = \frac{LS_i}{2} + ss_i$. The previous calculation is only referred to the stock of MTS products since, in theory, only these items are kept in stock. However, as mentioned before, in the context of this company and under certain conditions, there is the possibility of using lot sizes in the non-pure MTO items, that are larger than the specific orders. In this situation, a certain amount of stock is kept and, on average, is equal to $Stock_{MTO_i} = \frac{LS_{i_MTO}}{2}$.

Once the stock level of each product is known, it is possible to calculate the associated costs per unit of time. These costs are divided into two components: capital cost and storage cost. The storage costs regard warehousing expenses and are given by equation 4.11, which simply considers the total stock level and the unitary holding cost, h_i. The capital cost addresses the opportunity cost of the value invested in stock and considers the unitary cost of each product, UC_i, (equation 4.12). Again, as the time unit considered in these calculation is a month, the weighted average capital cost appears divided by 12 months.

$$C_{storage_i} = h_i \cdot Stock_i \tag{4.11}$$

$$C_{capital_i} = UC_i \cdot Stock_i \cdot WACC/12 \tag{4.12}$$

Setup Costs

The setup costs of a given item per unit of time are calculated by multiplying its expected number of setups in the same period by the unitary setup cost. The unitary setup cost has three components: wine wastage, machine operating cost and labour cost. The wine wastage is given by equation 4.13 and relates to the wine that is still in the machine when a production order for a different wine will be initiated. This wine is not entirely lost but is redirected to a reservoir of lower quality wine, being valued at a lower price, DV_i.

$$C_wastage\ _i = wastage \cdot (WV_i - DV_i) \tag{4.13}$$

Furthermore, the setup operation takes time and requires labour to configure the machine for the next production run. During this time, even though it is not producing, the machine has an operating cost per unit of time defined by the *MAQ* parameter. Thus, the setup cost related to the machine operating expenses is given by equation 4.14. Finally, considering the number of workers required to perform the setup operation for product i, $SetupWorkers_i$, and the average labour cost per worker and unit of time, Avg_labour_cost, it is possible to calculate the associated labour costs (equation 4.15) .

$$C_SetupMaq_i = \alpha_i \cdot MAQ \tag{4.14}$$

$$C_SetupLabour_i = \alpha_i \cdot SetupWorkers_i \cdot Avg_labour_cost \tag{4.15}$$

4.1.3 Heuristic

Once all the involved variables are formulated we can initiate the description of the heuristic procedure. This heuristic has the goal of minimizing the sum of inventory and setup costs by determining the best production strategies for each product.

The heuristic procedure is summarized in the algorithm 1. At the beginning, all products are classified as MTS and that list is given to the system. Then, the algorithm enters a loop that only ends when no more global cost savings can be added to the system by changing the production strategy of the products.

Inside the loop, the first goal is to determine, for each product, the batch size and the expected number of production orders per period, according to its current production strategy. With this information, and assuming that each product is produced in its preferential line, the next step is to compute the utilization of the production lines. Then, we are able to determine the mean

waiting time in the queues and the respective variance, through the queuing theory formulation previously presented. Next, the safety stocks of MTS items can be updated and added to the average cycle stock to determine the average total stock level of each item. In the case of the non-pure MTO items with defined lot sizes, their stock is also calculated at this stage. Then, for each product, the costs (holding and setup) of producing it as MTS and as MTO are calculated. Once these calculations are done, the costs are compared, and the products are ordered by the highest savings (if there are any) that would result from changing the strategy from MTS to MTO. The product on the top of the list, the one that provides the highest savings, has its production strategy changed from MTS to MTO. Once the change of that product's strategy is recorded, the updated list re-enters the loop for the next iteration. The procedure is repeated until no item has savings associated with a strategy change.

Let $|S_i|$ be the potential savings of changing the production strategy of product i, which is given by the difference between the total costs of producing it as MTS and the total costs of producing it as MTO. Similarly, let $|S_{max}|$ be the potential savings obtained by the product on the top of the list, at the end of each iteration, i.e., the one that leads to greater savings by changing its production policy. In its turn, the product on top of the list at the end of each iteration will be referred as $Product_top$.

Result: List with the production strategy for each product;
Initialization:
- For each product i, set initial production policy to MTS;
- Calculate queues and costs;
- Determine S_{max} and $Product_top$;
while $|S_{max}| > 0$ **do**
  - Change the production strategy of $Product_top$ and register that change;
  - Recalculate LS_i and λ_i for every item according to the updated production strategies ;
  - Recalculate the average processing times T_i ;
  - Recalculate production lines utilization;
  - Recalculate mean waiting times and variance in the queues;
  - Recalculate safety stocks of MTS items;
  - Calculate the costs of producing each product as MTS and as MTO in the recalculated scenario;
  - Calculate the savings, S_i;
  - Sort the products in descending order of $|S_i|$;
  - $|S_{max}| \leftarrow$ Value of $|S_i|$ on the top of the list after sorting;
  - $Product_top \leftarrow$ Product on the top of the list after sorting;
end

Algorithm 1: Heuristic for production strategy definition

The main concept of this algorithm, which follows a greedy approach, is that it takes in consideration the impact that the strategy change from the previous iteration had on the whole system. As the algorithm progresses and items start changing from MTS to MTO, the congestion effects increase due to the higher time spent in setup operations. With lead times becoming higher, the safety stocks of MTS items have to increase in order to ensure that the established service level

is maintained. The process stops when the savings in holding costs that result from changing an item's strategy to MTO, are less than the increase in costs this causes in the remaining MTS items due to increased congestion effects and consequent increases in safety stocks.

To the final list provided by the heuristic, a few more layers of decisions were applied, which this work will not examine in detail. Nonetheless, these extra decision layers include strategic and business decisions such as the ones presented next.

For products classified as MTO that are commercially classified as Core products, it is verified if their expected lead time is less than the one the company commits to in these items (5 days). If the expected lead time is greater than 5 days, the production strategy for these items is changed to MTS. This subsequent decision aims to ensure the service level compliance in MTO items, as no penalty for lost sales is being considered in the heuristic. The reason for not including such penalty was that the company could not quantify the cost of lost sales. Therefore, this was the solution adopted to ensure that the agreed lead times were met.

Next, for MTS items, an assessment is made to assess if it is more advantageous to produce the item on its own pallet or to produce it on a pallet that has a higher production volume and then transfer it to the right one when needed. This topic and the difference between production pallet and transfer pallet is explained in section 3.2.1. To do this, a trade-off analysis is performed that considers the increased setup costs of producing it on its own pallet versus the costs of moving the items later to the correct pallet, through picking and/or transfer operations. The result of this analysis divides the MTS items into two types: MTS with production on its own pallet and MTS with transfer.

4.2 Aggregate production planning

An optimization model based on a linear mixed-integer programming formulation was developed to address the medium-term production planning challenge. Since the goal was to determine and level the production resources required in each period at a macro level, a monthly time discretization was adopted.

Before initiating the model formulation, all the aspects the company wanted the model's output to include had to be established. The final version of this tool starts by presenting the monthly optimized production volumes of each product and the occupation rates of the production lines. It also provides the stock level of each product, both for finished and intermediate products. In terms of the workforce resources, it determines the number of workers required in each month, who may be permanent employees of the company or temporary employees hired for one-month periods. Furthermore, the model also includes the possibility of using overtime and opening a second shift, if the user decides to activate these options. In both cases, the output includes the number of hours used of such resources.

As the initial goal was this medium-term capacity and resource allocation planning, at a more aggregated level, we started by defining the model's granularity as the Aggregator code. The added complexity of further detailed granularity didn't seem to bring significant benefits to the

decisions at stake, as they are taken at a more aggregated level. However, as the project unfolded, it was concluded that further detail had to be included for the company to be able to upload the production plans on SAP. For this reason the final formulation of the model, that will be presented next, is more detailed than the typical APP model. The final production plans were generated at the SKU level, even though some considerations were made at more aggregated levels, which will be explained in the model formulation section.

The model will be used once a month to generate the production plan for the following 12 months, using a rolling horizon approach.

4.2.1 Model formulation

The goal of the mixed-integer programming formulation of this model is to generate production plans that minimize total costs while complying with all the operational and business constraints. Thus, to ensure a comprehensive mapping of all constraints and faithful modeling of the production process, a detailed process mapping was conducted at the beginning of the project.

Once the process was well understood, it was time to gather all the necessary data and prepare an input structure to the model. The input structure includes the list of all products and their relevant characteristics such as the production strategy, lot size and safety stock defined in the previous module. These parameters coming from the production strategy module, as already mentioned, will be updated at the beginning of each semester. For this reason, in the model implementation, the corresponding parameters appear duplicated, corresponding respectively to the values of the 1^{st} and 2^{nd} semesters. For simplicity, in the formulation presented below, only one parameter will be used (representing the value corresponding to the semester of the period in consideration at any point).

The complete list of the available production lines is also supplied to the model as input, with their operational specifications and costs. Moreover, as one product can be produced in more than one production line, a complete list of all the compatible combinations is also provided. Additionally, in this list, for each product-line combination, the production throughput and number of necessary workers in the production process are included. The main driver of the production process is the demand forecast for each period, which is extracted from the company's forecasting tool and supplied to the model.

The input information described above is not exhaustive but has covered the most relevant topics. All parameters and variables involved and their mathematical notation are listed in Table 4.2.

As the problem product's structure is multi-level, due to the existence of intermediate products that can originate more than one final product, all the possible production flows had to be discriminated. From the diagram in Figure 3.1, it was observed that products can derive from four different production flows:

- **Production flow 0**: Intermediate products that only go through the bottling process. Once produced, they are stored in the wine cellar of the production plant while waiting to be finished.
- **Production flow 1**: Finished products that result from direct production (bottling and finishing in a single run). Once produced, if no manipulation is required, they are stored at the finished goods warehouse by the logistics partner.
- **Production flow 2**: Finished products that result from the finishing process of previously bottled intermediate products (production flow 0). Once produced, if no manipulation is required, they are stored at the finished goods warehouse by the logistics partner.
- **Production flow 3**: Manipulated finished products which result from the manipulation process of finished goods (production flow 1 or production flow 2). When a finished product is going to be manipulated, no stock of that finished product is kept in between the two stages as the manipulation is performed immediately after the finishing process is concluded. Once the manipulation process is performed, they are stored at the finished goods warehouse by the logistics partner.

In the present formulation, when an intermediate product is transformed into another, it is assigned a new code, being considered, in each stage, a different item.

One last classification is needed to further clarify the formulation ahead. When a product leaves its final production stage, it is called a *final product*. This differentiation between final and finished products has not been made so far but, for the purpose of the model formulation, it is an important distinction to make. A final product can result from one of the following scenarios: a finished product resulting from production flow 1 or 2 that does not need manipulation, or a manipulated product from production flow 3. These are the market-ready products and are sent to be stored in the finished goods warehouse. A finished product that still needs to be manipulated is not a final product.

Another important aspect to be noted concerns the planning of MTO items. In these cases, the production volumes in each month provided by the model consist in a mere capacity reservation to address the orders that are forecasted to be placed. However, only after the actual orders arrive the production orders are uploaded to the ERP.

In addition to what has already been mentioned, the main assumptions of the model are the following:

1. Demand is deterministic and directly extracted from the forecasting tool.
2. There is only one setup of the same Aggregator in a given production line within certain period, which is used to process all the SKUs of that Aggregator that need to be produced.
3. Setup times and processing times are deterministic.
4. Setups are sequence-independent, even though the same setup is considered for all the SKUs of the same Aggregator in a given month.

5. Raw materials needed for production are immediately available.
6. Lot sizes can be rounded to a full pallet without loss of optimality.
7. All workers are equally specialized and equally capable of performing the same jobs with the same efficiency.
8. Backordering is not allowed.

Other assumptions and approximations are discussed throughout the following chapters when required.

Table 4.2: Table of notation for the optimisation model.

Sets	
N	Set of all products (indexed by i)
T	Set of all periods (indexed by t)
M	Set of all production lines (indexed by m)
A	Set of all Aggregators (indexed by a)
AP	Set of all Aggregators+Pallet (indexed by p)
S_a	Set of all products that belong to Aggregator a
S_p	Set of all products that belong to Aggregator+Pallet p
S_m	Set of all products that can be produced in line m
S_f0	Set of all products from Production flow 0
S_f1	Set of all products from Production flow 1
S_f2	Set of all products from Production flow 2
S_f3	Set of all products from Production flow 3
Sa_f0	Set of all Aggregators from Production flow 0
Sa_f1	Set of all Aggregators from Production flow 1
Sa_f2	Set of all Aggregators from Production flow 2
Sa_f3	Set of all Aggregators from Production flow 3
S_final	Set of all final products
S_mto	Set of all MTO items
S_cw	Set of all products corresponding to consumer wines
S_ww	Set of all products corresponding to white wines
Der_i	Set of all products preceded by product i
S_oldV	Set of all products that correspond to old variants
$Pl_i\ (=\ Pl_a)$	Set of all production lines where product i can be produced. It is the same set for all products within Aggregator a, so it can also be indexed at this level
Parameters	
$Forecast_{i,t}$	Expected demand for product i in period t
$Safety_Stock_p$	Safety stock of Aggregator+Pallet p
$Stock_init_i$	Stock of product i at the beginning of the first period

MPQ_a	Minimum production quantity of Aggregator a
$Cd_{i,m}$	Production rate of the product i in production line m
h_fin	Holding cost per stock unit at the finished goods warehouse
h_int	Holding cost per stock unit at the intermediate goods wine cellar
UC_i	Unitary cost of product i
MAQ_m	Cost of operating production line m per unit of time
$WACC$	Weighted average cost of capital
PW_wage_t	Wage of a permanent worker in period t
TW_wage_t	Wage of a temporary worker in period t
$Overtime_Cost$	Cost of each man-hour of overtime
$SecondShift_Cost$	Cost of each man-hour of second shift
$wastage_m$	Quantity of wine wasted in a bottling setup on prodction line m
$Wval_a$	The unitary value of the wine (liquid) of Aggregator a
$DGval_a$	The downgraded unitary value of the wasted wine (liquid) in a bottling setup of the Aggregator a
$SetupTime_{a,m}$	Setup time for Aggregator a in production line m
$SetupCrew_{a,m}$	Number of workers needed to execute the setup for Aggregator a in production line m
$LineCrew_{i,m}$	Number of workers needed to produce product i in production line m
WD_t	Number of working days of period t
Dwh	Number of daily working hours in regular time
$Dwh2s$	Number of working hours in second shift
$MaxOT$	Maximum number of daily overtime hours per worker
$MaxOT_annual$	Maximum number of annual overtime hours per worker
Cap_TW	Maximum number of temporary workers allowed
$Maintenance_{m,t}$	Maintenance days of production line m in period t
Bot_i	Number of bottles that compose product i
$MaxInv$	Maximum inventory allowed in the finished products warehouse
Pal_i	Number of products in a full pallet of product i
M	Big number. A number large enough to enforce certain constraints
S	Small number. A number very close to zero, but slightly above
Decision Variables	
PW_t	Number of active permanent workers in period t
TW_t	Number of active temporary workers in period t
$Overtime_t$	Overtime used in period t (man-hours)
$SecondShift_t$	Time of second shift used in period t (man-hours)
$I_{i,t}$	Inventory of product i in period t
$P_{a,m,t}$	Binary variable indicating if production line m was set to the Aggregator a in period t

$Qt_r_{i,m,t}$	Quantity produced of product i on line m in period t in regular time
$Qt_ot_{i,m,t}$	Quantity produced of product i on line m in period t in overtime
$Qt_2s_{i,m,t}$	Quantity produced of product i on line m in period t in second shift
Auxiliary Variables	
$Full_{m,t}$	Binary variable that equals 1 if the regular time capacity of production line m in period t is fully used, and 0 otherwise
$Qta_r_{a,m,t}$	Total quantity of products from Aggregator a produced on line m in period t in regular time
$Qta_ot_{a,m,t}$	Total quantity of products from Aggregator a produced on line m in period t in overtime
$Qta_2s_{a,m,t}$	Total quantity of products from Aggregator a produced on line m in period t in second shift

4.2.1.1 Objective function

In this mathematical formulation, the objective function is to minimize the sum of all inventory, workforce and production-related costs, for the total planning horizon.

Inventory Costs

Inventory related costs correspond to the costs associated with warehousing and the opportunity cost of all the value invested in stock. The warehousing costs are composed by two components. The first one addresses the expenses at the finished products warehouse and is obtained by multiplying the total stock of final products by the holding cost charged by the logistics partner per unit of stock (Equation 4.16).

$$Finished\ Inventory\ Cost = \sum_{t \in T} \sum_{i \in S_final} h_fin \cdot I_{i,t} \tag{4.16}$$

The other component of the storage cost concerns the expenses of storing the intermediate products at the wine cellar. This corresponds to the total stock of intermediate products that result from production flow 0, multiplied by the unitary storage costs indicated by the company (Equation 4.17).

$$WineCellar\ Inventory\ Cost = \sum_{t \in T} \sum_{i \in S_f0} h_int \cdot I_{i,t} \tag{4.17}$$

To complete the inventory costs, the cost of capital is obtained by multiplying the unitary cost of each product by its stock level, taking into account the weighted average cost of capital (Equation 4.18). As the latter is annually defined and the considered unit of time is a month, the division by 12 is necessary.

$$Capital\ Cost = \sum_{t \in T} \sum_{i \in N} I_{i,t} \cdot UC_i \frac{WACC}{12} \tag{4.18}$$

Workforce Costs

The total workforce costs of the planning horizon result from paying the workers' wages, the costs of overtime labour and the cost of opening a second shift. The costs related to the wages of permanent and temporary workers is given by Equation 4.19 and results from multiplying the number of worker of each type that is active in each period by their respective wage. Overtime and second shift are both formulated in man-hour units. Therefore, their cost is obtained by multiplying their unitary cost by the total number of man-hours of these resources activated by the model (Equation 4.20 and 4.21, respectively).

$$Workforce\ Cost = \sum_{t \in T} PW_t \cdot PW_wage_t \ + TW_t \cdot TW_wage_t \tag{4.19}$$

$$Total\ Overtime\ Cost = \sum_{t \in T} Overtime_t \cdot Overtime_Cost \tag{4.20}$$

$$Total\ SecondShift\ Cost = \sum_{t \in T} SecondShift_t \cdot SecondShift_Cost \tag{4.21}$$

Operational costs

The operational costs considered in this model consist in the cost of operating the production lines and costs incurred during setups. The hourly cost of operating each production line varies from one production line to another and is given by the *MAQ* parameter. Thus, the total cost of operating the production lines is obtained by multiplying the *MAQ* by the total production time in each line. In it's turn, the total production time can be determined by dividing the total production quantity by the respective throughput. Having this in mind, the total cost of operating these production lines throughout the year, according to the plan proposed by the model, will be given by:

$$ProdLines\ Operating\ Cost = \sum_{t \in T} \sum_{m \in M} \sum_{i \in S_m} \frac{Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t}}{Cd_{i,m}} \cdot MAQ_m \tag{4.22}$$

Regarding the cost of setups, a distinction must be made. Bottling setups are invariably different from finishing setups as the first type has to include the cost of wine wastage, as explained in previous sections. Therefore, finishing setups take in consideration the cost of maintaining the production line active during the non-productive period and also the labour costs of the crew needed to perform the operation (Equation 4.23). Bottling setups, on the other hand, take into consideration the two previous factors plus the wastage cost (Equation 4.24). Production flows 0 and 1 undergo the bottling procedure, hence their setup times have to be calculated by the bottling setup expression, while production flows 2 and 3 use the finishing setup expression. The

AvgLabourCost, which is the average labour cost per man-hour, was calculated by dividing the average salary of the company workers by the average number of working hours in a month.

As stated in the model assumptions, the setups in the production lines were defined at the Aggregator level. Since SKUs within the same Aggregator are almost identical, having the same wine type and bottle format, it was assumed that no setup cost existed between them. Therefore, in a given period, a product can only be produced in a given production line if there is a setup for that product's Aggregator. This also means that we are considering that if, in a given month, more than one SKU of the same Aggregator is produced in the same production line, their productions will follow one another.

$$Finishing\ Setup\ Cost = \sum_{t \in T} \sum_{a \in [Sa_f2 \cup Sa_f3]} \sum_{m \in Pl_a} (AvgLabourCost \cdot SetupCrew_{a,m} \cdot SetupTime_{a,m} + MAQ_m \cdot SetupTime_{a,m}) \cdot P_{a,m,t} \tag{4.23}$$

$$Bottling\ Setup\ Cost = \sum_{t \in T} \sum_{a \in [Sa_f0 \cup Sa_f1]} \sum_{m \in Pl_a} [wastage_m \cdot (Wval_a - DGval_a) + MAQ_m \cdot SetupTime_{a,m} + AvgLabourCost \cdot SetupCrew_{a,m} \cdot SetupTime_{a,m}] \cdot P_{a,m,t} \tag{4.24}$$

Considering all the terms above, the objective function to minimize is given by:

$$\begin{aligned} Minimize \sum & FinishedInventoryCost + WineCellarInventoryCost + CapitalCost \\ & + WorkforceCost + TotalOvertimeCost + TotalSecondShift\ Cost \\ & + ProdLinesOperatingCost + BottlingSetupCost + FinishingSetupCost \end{aligned} \tag{4.25}$$

4.2.1.2 Model constraints

Once the objective function has been determined, we can move on to the formulation of the problem's constraints. In this section, all the relevant constraints included in the model are presented. To make it simpler, the constraints have been grouped into sets that target common domains, which are presented below.

Workforce allocation

Regarding capacity and resource allocation, there are two main resources to allocate: workforce and time in the production lines. Regarding the workforce, the first step is to ensure that the existing capacity is greater than or equal to the capacity required to meet the production needs. This is accomplished by imposing that the man-hours required to fulfil the plan have to be less or equal to the available capacity expressed in man-hours, for each of the three possible production schedules

(regular time, overtime and second shift). Constraints 4.26 address the regular production period, the combination of constraints 4.27 and 4.28 address the overtime period and, finally, constraints 4.29 and 4.30 address the second shift.

$$\sum_{i \in N} \sum_{m \in Pl_i} \frac{Qt_r_{i,m,t}}{Cd_{i,m}} \cdot LineCrew_{i,m} \leq (PW_t + TW_t) \cdot Dwh \cdot WD_t \ , \quad \forall t \tag{4.26}$$

$$\sum_{i \in N} \sum_{m \in Pl_i} \frac{Qt_ot_{i,m,t}}{Cd_{i,m}} \cdot LineCrew_{i,m} \leq Overtime_t \ , \quad \forall t \tag{4.27}$$

$$Overtime_t \leq (PW_t + TW_t) \cdot MaxOT \cdot WD_t \ , \quad \forall t \tag{4.28}$$

$$\sum_{i \in N} \sum_{m \in Pl_i} \frac{Qt_2s_{i,m,t}}{Cd_{i,m}} \cdot LineCrew_{i,m} \leq SecondShift_t \ , \quad \forall t \tag{4.29}$$

$$SecondShift_t \leq (PW_t + TW_t) \cdot Dwh2s \cdot WD_t \ , \quad \forall t \tag{4.30}$$

The maximum number of temporary workers used in the model can be configured through the variable Cap_TW and is enforced by the following constraints:

$$TW_t \leq Cap_TW \ , \quad \forall t \tag{4.31}$$

In order to respect the company's human resources policy of not letting any worker undergo more than $MaxOT_annual$ overtime hours in a year, constraints 4.32 are applied. This imposes that the total anual man-hours of overtime are less or equal to the average number of workers per period multiplied by the $MaxOT_annual$.

$$\sum_{t \in T} Overtime_t \leq MaxOT_annual \cdot \frac{\sum_{t \in T} (PW_t + TW_t)}{12} \ , \quad \forall t \tag{4.32}$$

Production lines occupation

The allocation of the time in the production lines is the next step in terms of resource allocation. The goal is to make sure that the time needed to produce all the items in a given line does not exceed the available time in the production lines. Even though the lines could operate 24 hours a day, they need the workers to operate them. Therefore, the maximum available time in the machine on a given period corresponds to the total working hours, in either regular time, overtime or second shift. Constraints 4.33 to 4.35 ensure just that, for the three different schedules.

$$\sum_{i \in S_m} \frac{Qt_r_{i,m,t}}{Cd_{i,m}} \leq (WD_t - Maintenance_{m,t}) \cdot Dwh \ , \quad \forall m,t \tag{4.33}$$

$$\sum_{i \in S_m} \frac{Qt_ot_{i,m,t}}{Cd_{i,m}} \leq (WD_t - Maintenance_{m,t}) \cdot MaxOT \ , \quad \forall m,t \tag{4.34}$$

$$\sum_{i \in S_m} \frac{Qt_2s_{i,m,t}}{Cd_{i,m}} \leq (WD_t - Maintenance_{m,t}) \cdot Dwh2s \ , \quad \forall m,t \tag{4.35}$$

Then, to make sure that overtime and second shift are only used once the regular time is completely full, the following set of constraints is used. The auxiliary binary variable $Full_{m,t}$ tells us if the regular time on line m in period t is already fully occupied ($Full_{m,t} = 1$) or if there is still available capacity ($Full_{m,t} = 0$). Let $Cap_line_regular_{m,t} = (WD_t - Maintenance_{m,t}) \cdot Dwh$ be the total capacity in hours of production line m, in period t and in regular time. Additionally, $Prod_time\-regular_{m,t} = Qt_r_{i,m,t}/Cd_{i,m}$ represents the used capacity, in hours, on line m, in period t, in regular time. Similarly, for the capacity used in overtime and second shift, the same logic is applied: $Prod_time_overtime_{m,t} = Qt_ot_{i,m,t}/Cd_{i,m}$ and $Prod_time_SecondShift_{m,t} = Qt_2s_{i,m,t}/Cd_{i,m}$.

$$Cap_line_regular_{m,t} - Prod_time_regular_{m,t} <= M \cdot (1 - Full_{m,t}) \ , \ \forall m,t \tag{4.36}$$

$$Cap_line_regular_{m,t} - Prod_time_regular_{m,t} >= S - M \cdot Full_{m,t} \ , \ \forall m,t \tag{4.37}$$

$$Prod_time_overtime_{m,t} \leq M \cdot Full_{m,t} \ , \ \forall m,t \tag{4.38}$$

$$Prod_time_SecondShift_{m,t} \leq M \cdot Full_{m,t} \ , \ \forall m,t \tag{4.39}$$

Ensuring demand fulfilment

The main driver of the model's production orders is the fact that demand of each product must be fulfilled in time, as no backordering is allowed. For this reason, in a given period, if the available stock of a given product, carried over from the previous period, is lower than its forecasted demand, a quantity equal or greater to the difference between these two values must be produced. This is enforced by the following constraints:

$$\sum_{m \in Pl_i} (Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t}) \geq Forecast_{i,t} - I_{i,t-1} \ , \quad \forall i,t \tag{4.40}$$

Enforce setups and minimum production quantities

One of the most important constraints for the operation of the model is the one that enforces that the production of a given product in a given line requires a setup for its Aggregator in that period. This is achieved through constraints 4.41 and 4.42, which link the total quantity of a given product produced in each period in a given line, with the setup variable $P_{a,m,t}$. Constraint 4.42 also has the role of enforcing the minimum quantity to be produced, the MPQ_a. This minimum quantity results from the production lots defined in the previous module, which in the case of MTS items

corresponds to the EOQ and in the case of MTO is generally zero. The exception of the previous rule are the production lots that can occur in some non-pure MTO items, as previously explained. Like the setup decision, the obligation to meet the minimum quantity is enforced at the Aggregator level. Thus, if there is production of SKUs of a given Aggregator in a given line and period, their total must be at least equal to the MPQ_a.

$$\sum_{i \in S_a} (Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t}) \leq M \cdot P_{a,m,t} \ , \quad \forall a,t,m \in Pl_a \tag{4.41}$$

$$\sum_{i \in S_a} (Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t}) \geq MPQ \cdot P_{a,m,t} \ , \quad \forall a,t,m \in Pl_a \tag{4.42}$$

Inventory related constraints

Regarding the inventory related constraints, there are many aspects to address, the most important of all being the inventory balance equation. Assuming all the demand forecast of a given period directly translates into sales, the model must know that this quantity is no longer available in stock. Furthermore, as this is a multi-level problem with parent-relationships between products, the consumption of a preceding product whenever a product that incorporates it is produced must be enforced. This needs to consider the incorporation factor, which in the context of this company consists in the number of bottles of the preceding material that are required to produce the wine box of the final SKU. Products resulting from production flows 0, 1 and 2 are the ones which may originate a different product and their inventory balance is given by equation 4.43. This constraint enforces that the number of bottles of product i produced in period t added to the stock carried over from period $t-1$ must be equal to the demand for product i and for its by-products plus the remaining stock at the end of period t.

$$(Qt_total_{i,t} + I_{i,t-1}) \cdot Bot_i = (Forecast_{i,t} + I_{i,t}) \cdot Bot_i + ForecastDer_{i,t} \ , \quad \forall i \in [S_f0 \cup S_f1 \cup S_f2], t \tag{4.43}$$

where $Qt_total_{i,t}$ is the total quantity of product i produced in period t and given by

$$Qt_total_{i,t} = \sum_{m \in Pl_i} Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t} \tag{4.44}$$

and $ForecastDer_{i,t}$ is the total number of bottles of product i that is needed to fulfil the demand of its derived products. It is obtained by the following expression:

$$ForecastDer_{i,t} = \sum_{k \in Der_i} Forecast_{k,t} \cdot Bot_k \tag{4.45}$$

Products that come from production flow 3 are always final products, thus their inventory balance does not need to consider by-products (Constraint 4.46).

$$Qt_total_{i,t} + I_{i,t-1} = Forecast_{i,t} + I_{i,t} \ , \quad \forall i \in S_f3, t \tag{4.46}$$

Next, compliance with the safety stocks in all periods must be ensured for all products. Safety stocks are ensured at the Aggregator+Pallet level since it was at this granularity level that they were defined.

$$\sum_{i \in S_p} I_{i,t} \geq Safety_Stock_p \ , \quad \forall p, t \tag{4.47}$$

The previous safety stock equation imposes the target stock in each period for the generality of products. However, some specific cases arose during the project, where it was necessary to impose different restrictions that will be described next.

Regarding the last period of the planning horizon, as there is no forecast information after that moment, the model would assume that no demand exists after that and the recommended stocks would be much lower. To avoid this and ensure stable stock levels, it was imposed that in the last period the quantity in stock, for MTS final products, should be equal to its average level, which was approximated by the sum of the Safety Stock with half of the MPQ.

Another case relates to the intermediate items that result from production flows 1 and 2 and undergo manipulation. As mentioned before, these items do not constitute stock. The reason is that these products immediately go from one stage to the other, without being stored in between. Therefore a constraint which enforced the inventory of such products to be zero in every period was also added to the model.

Finally, the last inventory related constraint addresses the maximum warehousing capacity. While the wine cellar where intermediate items are stored has more than enough space and no maximum capacity is considered, the same does not happen with the finished goods warehouse. In the finished goods warehouse, a limit is established with the logistics partner and cannot be exceeded. This means that the total inventory of final products has to be lower or equal to the defined limit, as imposed by constraint 4.48.

$$\sum_{i \in S_final} I_{i,t} >= MaxInv \ , \quad \forall t \tag{4.48}$$

Prevent MTO antecipation

To prevent the model from over-anticipating the production of MTO items, which goes against the purpose of this production strategy, a new constraint was added. This constraint prevents the quantity of MTO items produced in a given month from being greater than the forecasted demand for the same period. The two exceptions to this rule are the non-pure MTO items where there is an established production lot and when the demand forecast of a certain period is less than a full pallet. The latter case, for example, happens with some items that have forecasted orders throughout several months, but in each month only a few cases are expected. In these situations, instead of producing a very small quantity every month, the company prefers to produce a full

pallet and use the limited stock that remains to fulfil the orders of the following periods. Therefore, the quantity to produce each month is equal to the maximum value between the period forecast, the MPQ and a full pallet (constraints 4.49).

$$\sum_{m \in Pl_i} (Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t}) \leq Max(Forecast_{i,t}, MPQ_i, Pal_i) \ , \quad \forall i \in S_mto, t \quad (4.49)$$

Enology constraints

The intrinsic nature of the wine industry means that specific constraints are necessary to meet more particular enology requirements. When a new harvest of wine is available for a given product, the company no longer produces the older variants. Eventually, if the demand forecast for older variants is greater than the available stock, the surplus is added to the forecast of the newer one. Having made this change to the inputted forecast, to ensure that no demand is left unfulfilled, all that remains is to impose that the quantity produced each month for these old variants is zero (constraints 4.50).

$$\sum_{i \in S_oldV} \sum_{m \in Pl_i} Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t} = 0 \ , \quad \forall t \quad (4.50)$$

Another restriction that has a significant impact on the production plans is related to the white wines' bottling process. The higher temperatures typically felt on the second half of the year are not adequate for the bottling process of this type of wine. Therefore, production flows 0 and 1 (the ones that include the bottling stage) of white wines cannot occur in the second semester (constraints 4.51). Similarly, consumer wines, in general, cannot be bottled during the harvesting period, which comprises the months of September and October. The reason behind this limitation is that enologists are busy with vineyard duties and, consequently, they are not available to monitor the bottling process as it is required. As in the case of the white wines, the restriction to be imposed consists in forcing the total produced quantities of these wines to be zero in these periods (constraints 4.52).

$$\sum_{m \in Pl_i} \sum_{i \in S_ww} Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t} = 0 \ , \quad \forall t \in 2nd\ semester \quad (4.51)$$

$$\sum_{m \in Pl_i} \sum_{i \in S_cw} Qt_r_{i,m,t} + Qt_ot_{i,m,t} + Qt_2s_{i,m,t} = 0 \ , \quad \forall t \in [Sept, Oct] \quad (4.52)$$

Chapter 5

Results

The experimental results obtained by the application of the proposed methodology are presented in this chapter. Section 5.1 presents the results of the production strategy definition heuristic and the expected costs of the final strategy partition. Then, section 5.2 provides the results of the optimization model for medium-term production planning with regard to the computational performance and the preliminary results of the generated plans.

5.1 Production strategy definition Results

Using the proposed approach, a list with the production strategy for each item in the company's product assortment was generated. The partition of products into the MTS and MTO strategies was based on the minimization of a global cost function that comprises inventory and setup costs.

Preliminary results show that by using the optimized solution the company would experience a 14% reduction in the total considered costs. Figure 5.1 establishes a comparison between the current strategy definition (As Is) and the optimized solution (To Be), for all the analysed components of the cost function. The inventory related costs in the proposed solution, which include the capital and storage costs, are higher than in the present situation, which indicates higher stock levels. On the other hand, the setup related costs, which include the wine wastage and the expenses with machine operation and labor during the setup operation, are significantly lower, resulting in the total costs decrease.

This decrease in total costs was achieved while at the same time increasing the service level to the company's target of 96%, which represents a 2.4pp increase over the current service level (Figure 5.2a). This leads to the increase of safety stocks and, consequently, contributes to the higher global stock level.

When comparing the actual production strategies suggested by the heuristic with the ones the company uses today, as expected by the lower setups costs and higher inventory expenses, there was an increase in the number of MTS items. Figure 5.2b shows that 13% of the previously classified MTS products had their production strategy changed to MTO in the optimized solution. On the other hand, a number of products equivalent to 18% of the initial quantity of MTS items

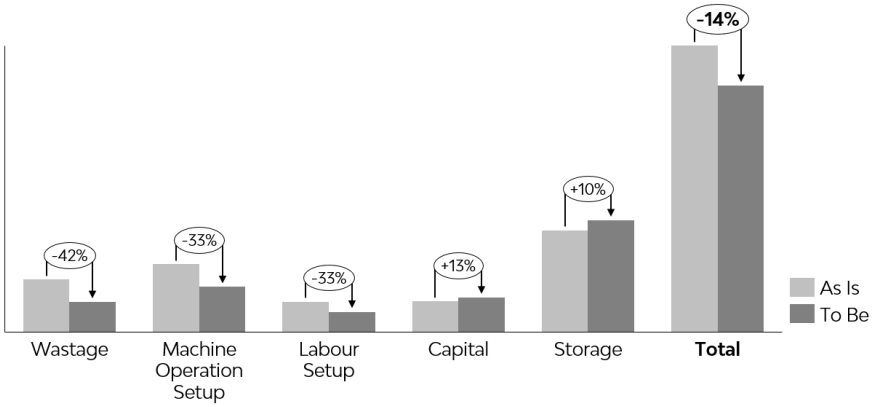
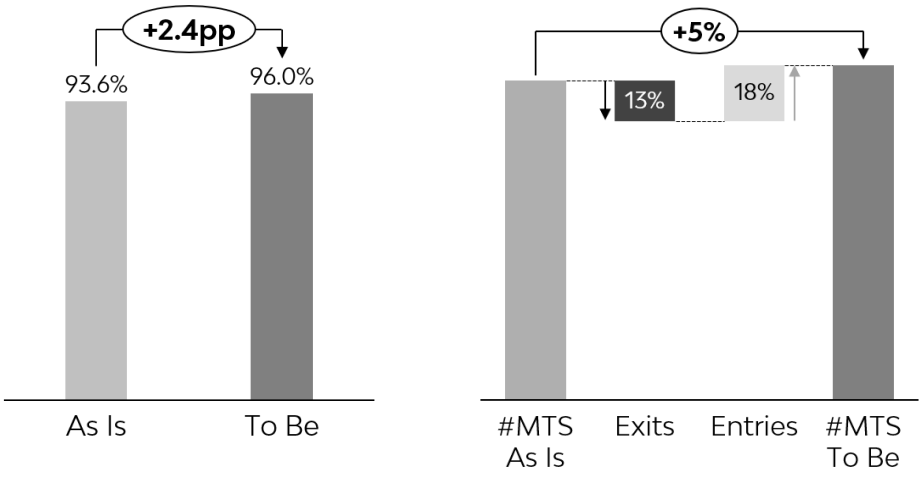


Figure 5.1: Cost comparison between the current practices (As Is) and the optimized strategy definition (To Be)

took the opposite direction, changing their production strategy from MTO to MTS. Therefore, the final balance was a 5% increase in total MTS products.



(a) Service level

(b) Number of MTO items

Figure 5.2: Comparison between the proposed solution and the As Is situation

5.2 Aggregate production planning Results

During the implementation of the model presented in section 4.2, it was developed a feature to allow the user to run different scenarios. This includes the possibility of defining whether or not the model can allocate resources to overtime, open a second shift and define the data set to run. It is possible to run the entire data set of the company, comprehending all the products and the production lines of both plants, or generate plans for each plant separately. The latter configuration was implemented to achieve faster results, since the size of the data set proved to

have a significant impact on the computational performance of the model. Lastly, it is possible to define a minimum number of permanent workers, in order to generate plans according to the company's current permanent workforce reality.

As intended, the output of the generated plans includes the quantity of each product to produce in each month, and the production line allocated to each production order. It also indicates the number of permanent and temporary workers needed in each period, as well as the man-hours of overtime and second shift. Moreover, it presents to the user the stock level of each product in each period of the planning horizon and the total stock levels for both the finished and intermediate products warehouses. A lot more information can be extracted from the model output, such as the occupation rate and the number of setup operations for each production line in each period of the planning horizon.

To make the results somewhat comparable with the company's current practices and demonstrate the improvements that an optimization model can bring, a similar scenario was used to generate the results presented in this section. Thus, the present results consider a scenario where overtime and second shift were not allowed to be used, and the only way to increase the workforce capacity was to hire temporary workers. This is in line with the company's current policy, which only works regular hours and hires temporary workers to cover the demand peaks. Regarding the data-set, the results include the company's global information, with all products and production lines included.

5.2.1 Computational Results

The optimization model formulated in Chapter 4 was implemented in Python using the PuLP library. The presented instance was solved using the commercial solver CPLEX, version 12.7.0, on an Intel®Xeon®E5-2640 v3 2.60GHz processor with 78.1 GB of available RAM, limited to the utilization of 10 threads. Default CPLEX parameters were used.

The instance shown in Table 5.1 is one of the instances that was tested and refers to the scenario described above. The table presents the instance dimensions and the resulting dimensions of the MIP problem. The used data set comprehended 976 products, 1596 compatible combinations between items and its production lines, and a time horizon of 12 periods.

Table 5.1: Problem data and MIP dimensions

Problem Data			MIP Dimensions		
Products	Compatible combinations	Periods	Constraints	Total variables	Binary variables
976	1596	12	33462	47879	13240

By testing different instances, it was possible to conclude that the number of products, compatible combinations, and periods significantly impact the dimension of the MIP problem, which is strictly related to the computational performance of each instance. In this case, a time limit of one hour was established for the solving process, which proved to be insufficient to reach optimality. The solution time and the gap at the end of the solving process are presented in Table 5.2. Even

though optimality was not reached within the time limit, the proposed production plans proved to be good solutions, as it will be demonstrated in the following section.

Table 5.2: Solution time and MIP gap

Solution time	MIP gap
3601.16 sec	0.16%

5.2.2 Production plans Results

An Excel dashboard was developed to analyse the generated plans, which enabled to draw the conclusions presented in this section.

Besides minimizing a global cost function, this model, and APP models in general, had the goal of leveling the production volumes and, consequently, the allocation of resources throughout the planning horizon. This approach intends to absorb the effect of demand peaks, which in this case occur in the last four months of the year, by producing in advance in the previous months. Figure 5.3, shows that this was achieved by plotting the curve of the production levels proposed by the model against the demand forecast. As it can be seen, the demand peak in the forecast curve is flattened when translated to the production evolution curve. At the same time, the levels of production in the months leading up to the demand peak are higher than the expected demand, meaning that the production of the last months is being anticipated. Here, the pattern of production from the historical data is also included, to show that the optimized plans are much more leveled throughout the planning horizon.

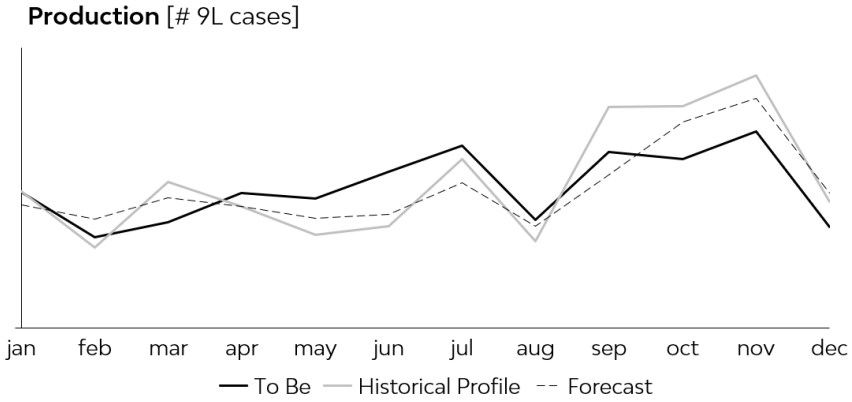


Figure 5.3: Production levels evolution

The evolution of the stock of finished goods is presented in Figure 5.4 as well as the historical data for comparison. As we can see, the proposed plans generate higher levels of stock which is both justified by the increased number of MTS items and also by the anticipation of production. As items are produced earlier to balance the demand fluctuations, they have to spend more time in stock. The safety stock is also plotted in Figure 5.4. As the production strategy is recalculated at

the beginning of each new semester to account for the seasonality effects, the safety stock levels are higher in the second semester due to the peaks in demand that occur in this period. At the end of December, the safety stock returns to levels congruent with those of the first semester.

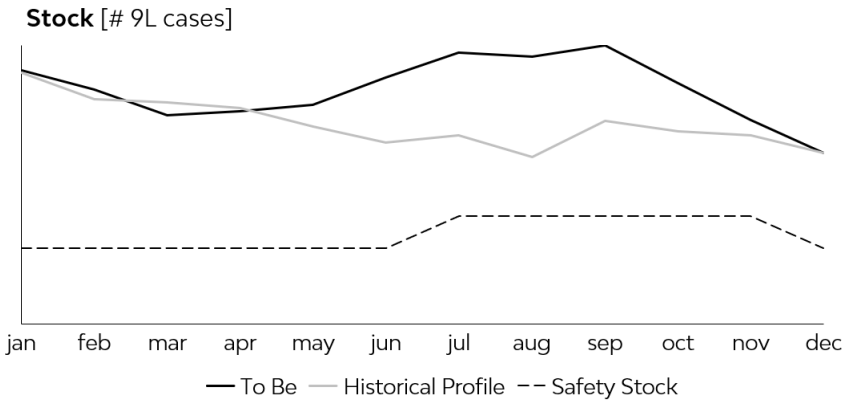


Figure 5.4: Stock Evolution

In terms of the average number of monthly setups, the proposed solution results in an 18% decrease, when compared to the historical practices (Figure 5.5). This results from the higher number of MTS items and also because the proposed lot sizes are greater than the ones the company historically has been using.

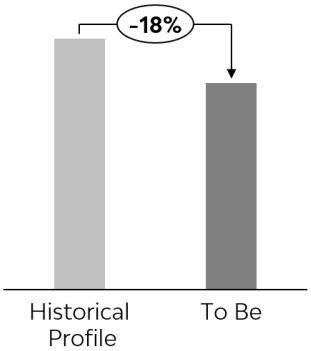


Figure 5.5: Average number of monthly setups

Regarding cost analysis, simply comparing the optimized plan with a historical plan for the same time period would not make sense because they address different demands and, therefore, cannot be directly compared. Furthermore, the plans generated by the model already consider the optimization performed by the strategy definition module. Hence, as both effects would be overlapped, it would not be possible to know which module generated the eventual gains. To address this issue, a new scenario imposed the production levels to follow the historical production pattern as seen in the grey line of Figure 5.3. By doing this, it was possible to simulate what the result would be if current practices were applied to the same sales forecast.

Figure 5.6 presents the cost comparison between the optimized (To be) plans and the plan that followed the historical practices (Historical Profile). As expected by the higher stock levels, the inventory related costs are higher in the proposed solution. Nevertheless, the savings in operational costs more than make up for it, resulting in a estimated 3% decrease in total costs. Wastage costs, machine operating costs, and labor costs have all been reduced. The most impactful factor was the 12% reduction in labour costs, because the historical practices result in the hiring of many temporary works to accommodate the demand fluctuations.

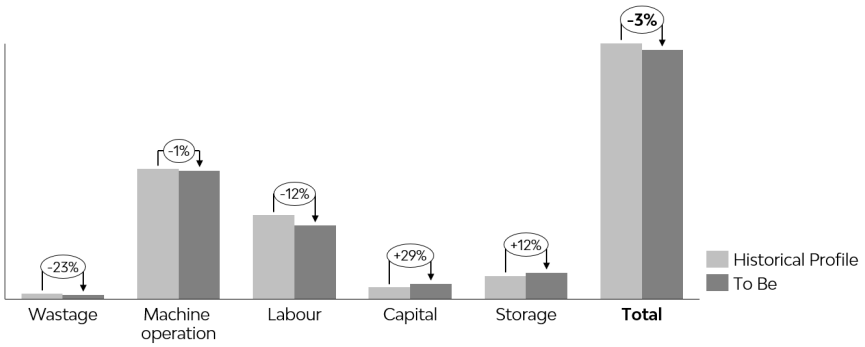


Figure 5.6: Cost comparison between the historical practices and optimized plan

Chapter 6

Conclusions and future work

The work carried out in this project aimed to improve two correlated aspects of the planning process of a Portuguese company operating in the wine sector: production strategy definition and aggregate production planning. The objective was to develop a methodology to be implemented in a decision support system capable of assisting the company's planning department to choose the best production strategy for each product and to plan capacity and allocate resources in the medium term. The project was motivated by the lack of analytical foundations of the current decision-making process.

The work was divided into two separate modules. The first one had the objective of developing a methodology capable of creating a list with the optimal production strategy for each product. The second module, in its turn, aimed to develop an optimization approach to generate aggregate production plans. Both problems were based on the search for the optimal trade-off between operational and inventory costs, as these two factors benefit from opposite strategies. A predominantly MTS strategy with larger production lots benefits operational efficiency at the expense of higher inventory costs. On the other hand, in a dominantly MTO environment, the inventory costs are lower, but the entropy in the production system greatly increases, with more setup operations and intensified congestion effects.

Regarding the production strategy definition module, a hybrid MTS/MTO system was implemented. As the formulation of an optimization model for this cost minimization problem proved in the literature to be non-linear, a greedy heuristic was developed, and near-optimal solutions were determined. The approach considered the impact that a change in the production strategy of one product has on the whole system by modeling congestion effects using M/G/1 queues.

Experimental results have shown that the proposed methodology is expected to bring approximately 14% savings when implemented by the company, while at the same time increasing the service level by 2.4pp. The final solution suggested a 5% increase in total number of MTS products, which resulted in significant operational savings, that surpassed the increase in inventory costs.

In the aggregate production planning module, the proposed methodology used a mixed-integer programming optimization model to generate plans that minimize inventory and operational costs.

The goal was to develop a tool that helped the company to plan capacity and allocate resources in the medium term. Production volumes, workforce levels and production lines utilization had to be balanced in an environment characterized by strong seasonality effects. The company's planning context included multiple production lines, many of which could produce the same products, and a multi-level product structure where the same intermediate item could originate several different finished products. Demand was considered deterministic since the company already had in place a forecasting tool that predicted the monthly demand of each SKU. All these factors combined resulted in a multi-level, multi-resource, single objective deterministic problem. The developed formulation was an extension to the lot sizing models found in the literature, specifically tailored to meet the specifications of the problem in hand.

Even though the solution was not yet fully implemented by the company, preliminary results show that the application of the proposed methodology can reduce total operational and inventory costs by 3%. The most significant factor in this cost reduction was a more efficient allocation of the workforce, which allowed for a 12% reduction in labour related costs. The experimental results also revealed a more leveled production throughout the year, smoothing the demand peak of the last four months and relieving pressure on the entire system. Consequently, as some production gets anticipated, higher stock levels were obtained. Nevertheless, the costs associated with the higher inventory level, are less than the operational savings achieved. This increased operational efficiency was also noted in the reduction of the average number of monthly setups by 18%.

It can be concluded that optimizing the allocation of production resources is of the utmost importance for several reasons. Firstly, a significant cost reduction can be achieved. Secondly, a leveled allocation of resources promotes stability in the production plant as the workload is balanced throughout the year. Moreover, in terms of the workforce management, a balanced allocation prevents constant hirings and firings of temporary workers. Instead, a constant number of workers is maintained throughout the planning horizon, which also brings stability to the workers lives.

By using the developed tools, the planning department will be able to perform these planning tasks much faster and dedicate more time to testing different scenarios and take better decisions with more information and confidence.

Despite the expected gains, some limitations and opportunities for improvement can be found in the proposed methodologies. For example, the solutions provided for both modules are highly dependent on the correct modelling of the company's cost structure. Hence, accurate cost calculation is crucial, which in many cases is not straightforward as there are many hidden costs that are hard to quantify. Furthermore, in both cases, several assumptions that relaxed real-world constrains were considered, which may result in the complexity of the system not being completely represented.

In the case of the definition of the production strategy, the highly combinatorial nature of the problem is not fully captured by the proposed heuristic. Even though it accounts for the effect a strategy change in one product has on the remaining products, it does not address the problem as a whole. In each step of the process, the strategy change is performed to the product that by itself

leads to the highest savings, hence not considering the effect of using different combinations. To overcome this problem the production environment could be modelled in a simulation program, and a simulation-optimization approach could be implemented, to account for the combinatorial effects.

Concerning the optimization model, the computational performance was proven to be highly dependent of the size of the instance, and given the complexity of the present problem, the solver struggled to deliver an optimal solution in a reasonable amount of time. This problem could be addressed by employing a fix-and-optimize heuristic that reduces the size of the problem size by fixing some decision variables between iterations. The application of such technique could deliver significantly more agile and still high quality results.

To conclude, the main contribution of the present work to the literature of this field was the development of an integrated approach that combined optimization in aggregate production planning with MTS/MTO production strategy definition. Furthermore, it was developed a tailor-made methodology capable of addressing the specific needs of companies operating in the wine sector.

Bibliography

Albin, S. (1982). On poisson approximations for superposition arrival processes in queues. *Management Science*, 28(2):126–137.

Allen, A. (1990). *Probability, Statistics, and Queueing Theory*. Academic Press.

Baykasoglu, A. and Gocken, T. (2010). Multi-objective aggregate production planning with fuzzy parameters. *Advances in Engineering Software*, 41(9):1124–1131.

Belvaux, G. and Wolsey, L. A. (2000). A Specialized Branch-and-Cut System for Lot-Sizing Problems. *Management Science*, 46(5):724–738.

Cheraghalikhani, A., Khoshalhan, F., and Mokhtari, H. (2019). Aggregate production planning: A literature review and future research directions. *International Journal of Industrial Engineering Computations*, 10(2):309–330.

Gicquel, C., Minoux, M., Dallery, Y., and Gicquel (2008). Capacitated lot sizing models : a literature review. *hal-00255830*.

Gomes da Silva, C., Figueira, J., Lisboa, J., and Barman, S. (2006). An interactive decision support system for an aggregate production planning model based on multiple criteria mixed integer linear programming. *Omega*, 34(2):167–177.

Guimarães, L., Klabjan, D., and Almada-Lobo, B. (2014). Modeling lotsizing and scheduling problems with sequence dependent setups. *European Journal of Operational Research*, 239(3):644–662.

Jamalnia, A., Yang, J. B., Xu, D. L., and Feili, A. (2017). Novel decision model based on mixed chase and level strategy for aggregate production planning under uncertainty: Case study in beverage industry. *Computers and Industrial Engineering*, 114(August):54–68.

Kaminsky, P. M. and Kaya, O. (2006). MTO-MTS Production Systems in Supply Chains.

Lai, Y.-J. and Hwang, C.-L. (1992). A new approach to some possibilistic linear programming problems. *Fuzzy Sets and Systems*, 49(2):121–133.

Misni, F. and Lee, L. S. (2017). A Review on Strategic, Tactical and Operational Decision Planning in Reverse Logistics of Green Supply Chain Network Design. *Journal of Computer and Communications*, 05(08):83–104.

Nagib, A. N. M., Adnan, A. N., Ismail, A., Halim, N. H. A., and Khusaini, N. S. (2016). The Role of Hybrid Make-to-Stock (MTS) - Make-to-Order (MTO) and Economic Order Quantity (EOQ) Inventory Control Models in Food and Beverage Processing Industry. *IOP Conference Series: Materials Science and Engineering*, 160(1):0–11.

Nam, S. and Logendran, R. (1992). Aggregate production planning — a survey of models and methodologies. *European Journal of Operational Research*, 61(3):255–272.

Rajagopalan, S. (2002). Make to order or make to stock: Model and application. *Management Science*, 48(2):241–256.

Sanni, S., Jovanoski, Z., and Sidhu, H. (2020). An economic order quantity model with reverse logistics program. *Operations Research Perspectives*, 7:100133.

Silver, E. A. (1976). 6 - medium range aggregate production planning: State of the art. In IBRAHIM, I., SEO, K., and VLACHOS, P., editors, *Readings in Managerial Economics*, Pergamon International Library of Science, Technology, Engineering and Social Studies, pages 227–254. Pergamon.

Soman, C. A., Van Donk, D. P., and Gaalman, G. (2004). Combined make-to-order and make-to-stock in a food production system SOM-theme A: Primary processes within firms. *Int. J. Production Economics*, 90:223–235.

Tashakori, L. (2015). Modeling ordered decision making in MTO/MTS production industries. *Decision Science Letters*, 4(4):453–464.

Torabi, S. A., Ebadian, M., and Tanha, R. (2010). Fuzzy hierarchical production planning (with a case study). *Fuzzy Sets and Systems*, 161(11):1511–1529.

Williams, T. (1984). Special products and uncertainty in production/inventory systems. *European Journal of Operational Research*, 15(1):46–54.