



Sustainable Short-Term Production Planning Optimization

Fernando Zanella¹ · Clara Bento Vaz^{1,2,3}

Received: 2 April 2023 / Accepted: 18 August 2023
© The Author(s) 2023

Abstract

This study proposes a framework for short-term production planning of a Portuguese company operating as a tier 2 supplier in the automotive sector. The framework is intended to support the decision-making process regarding a single progressive hydraulic press, which is used to manufacture cold-stamped parts for exhaust systems. The framework consists of two sequential levels: (1) a Mixed-Integer Linear Programming (MILP) model to determine the optimal production quantities per week while minimizing the total cost; (2) a dynamic production sequencing rule for scheduling operations on the hydraulic press. The two levels are combined and implemented in Excel, where the MILP model is solved using the Solver add-in, and the second level uses the optimal production quantities as inputs to determine the production sequence using a dynamic priority rule. To validate the framework, a proposed optimal plan was compared to a real plan executed by the company, and it was found that the framework could save up to 22.1% of the total cost observed in reality while still satisfying demand. To address uncertainties, the framework requires a rolling weekly planning horizon.

Keywords Mixed-integer linear programming · Short-term production planning · Responsible production

Introduction

The United Nations (UN) has set 17 Sustainable Development Goals (SDG), which are a group of guidelines that intend to build a better world by 2030. The topic of these goals involves social (hunger, poverty, equality), economic

(industry, innovation, partnership), and environmental aspects (natural resources, fauna, flora) [1]. The 12th SDG is entitled “Responsible Consumption and Production”, which states that the global economy’s dependence on the natural environment and its resources increased by 65% from 2000 to 2019 [2]. All targets set by the 12th SDG are related to sustainable management and efficient use of resources while reducing all kinds of waste through prevention, reduction, recycling, and reuse. In this context, the Linear Programming (LP) optimization technique is frequently used as an efficient approach to promote sustainable production, since it helps companies to save resources and operate in an optimal way [3]. The most common application of LP involves determining the optimal scenario for a set of activities toward an objective. Most of the time, these activities compete for resources between each other, such as energy and raw materials.

For any company to achieve responsible production, developing and applying a good production plan is essential. However, this task can be complex, since it requires a lot of effort and cooperation between everyone responsible for the decision-making process. A good plan will always lead to a good workflow, promoting efficient resource usage and responsible production. Usually, production planning can be

Fernando Zanella and Clara Bento Vaz contributed equally to this work.

This article is part of the topical collection “Computer Sciences and Artificial Intelligence for Smart Cities” guest edited by Sergio Nesmachnow, Luis Hernández Callejo and Pedro Moreno.

✉ Clara Bento Vaz
clvaz@ipb.pt

Fernando Zanella
a52494@alunos.ipb.pt

- ¹ Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança (IPB), Campus de Santa Apolónia, Bragança 5300-253, Portugal
- ² Laboratório para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança (IPB), Campus de Santa Apolónia, Bragança 5300-253, Portugal
- ³ Centre for Management and Industrial Engineering (CEGI/INESC TEC), Porto, Portugal

separated into three different time horizons: (1) long-range plans, (2) medium-range plans, and (3) short-range plans [4].

Long-range plans consider a period longer than 18 months and are related to new infrastructure, research and development, new products, facility location, and capacity [5]. Medium-range plans are elaborated to always match supply and demand in terms of rough volume and product mix, which means maintaining enough raw material, work in process, and inventory level to meet the demand while also absorbing possible fluctuations. Finding the balance between all these requirements while aligning them with the company's strategic plan is known as Sales and Operations Planning (S &OP). The ultimate result of an S &OP is an Aggregate Production Plan (APP), which involves the determination of the rough quantities and timing of production for a period between 3 and 18 months from the present [5]. The short-range plans can be subdivided into two steps whose order matters: the Master Production Schedule (MPS) and the Detailed Production Schedule (DPS) [6]. The process of breaking down the APP is referred to as the MPS. In practice, this means converting all rough aggregate quantities into a specific manufacturing schedule, usually in terms of weeks, defining what products (or alternative products) should be manufactured, their production timing, and their quantities [5]. Based on the quantities to be produced determined by the MPS, it is possible to develop the Detailed Production Schedule (DPS). For each product manufactured within the planning period, this detailed schedule should indicate start and completion times, machine and employee assignments, required resources, and any other relevant information [6]. This entire process must be done following the Bill of Materials (BOM) structure, which is a document that outlines the hierarchy of all components and their required quantities to manufacture one unit of each product [5].

In order to contribute to the 12th SDG, this work presents a framework to support the short-term production planning and scheduling of a medium-sized Portuguese company, which operates as a tier 2 supplier within the automotive production supply chain. The framework is designed to be used under a weekly rolling planning horizon composed of 4 weeks, meaning that only the first out of the 4 weeks of the production plan should be executed, while the other weeks should be re-planned until they occur, allowing the framework to absorb and anticipate any variation in terms of future demand. The first level of the framework is a Mixed-Integer Linear Programming (MILP) model to optimize the MPS and determine the optimal quantities to be produced per week, considering a single machine. The second level is a DPS related to the same machine, expressed using a Gantt Chart, in which the production schedule is determined by a dynamic sequencing rule based on the weekly quantities to be produced determined on the first level. Both levels

were fully implemented in the same Excel spreadsheet containing a dashboard created to make the tool easier to use by the company's employees. The MILP model is based on the Solver add-in and the Simplex method, whereas the dynamic sequencing rule is based on cell-to-cell formulas and calculations.

This paper is organized as follows: "[Literature Review](#)" encompasses a literature review regarding some common approaches related to short-term production planning optimization. "[Methodology](#)" proposes a framework designed to support the company under study during its short-term production planning process. "[Results and Discussion](#)" compares a real short-term plan developed and executed by the company with an equivalent optimized plan proposed by the framework. Finally, "[Conclusion](#)" includes the study's conclusions and suggestions for future work.

Literature Review

This section presents some of the studies in the field of production planning and scheduling optimization, addressing some methods proposed by many authors to plan and schedule short-term production activities in various manufacturing environments.

Master Production Schedule

Omar et al. [7] disaggregated the APP of a real-world resin manufacturer into the MPS using a model based on Fuzzy Mixed Integer Linear Programming (FMILP), aiming to incorporate the inherent imprecision of the chemical manufacturing process. The production plan obtained using the FMILP model was compared to a similar deterministic model based on Integer Weighted Goal Programming (IWGP). This IWGP model corresponds to the second of the three-level hierarchical production plan proposed by Omar and Teo [8]. The objective of both models was to minimize the total costs, using as inputs some target values determined by the company's APP. LINGO was used to solve both models, while Excel was used to export and import data. The results showed that the FMILP approach leads to cost savings compared to the IWGP model, although the latter is still viable.

Martín et al. [9] developed an optimization model based on MILP for the MPS of a real-world tier 2 supplier of the automotive supply chain. The company produces decorative parts through injection. According to the authors, this manufacturing process is imprecise, and the production time can vary greatly. These uncertainties were incorporated using two techniques proposed by Soyster [10] and Bertsimas and Sim [11]. The objective of both was to minimize the total cost while considering several constraints of the productive

scenario. The decision variables were the optimal quantities to be produced, the number of batches, the net inventory, the idle time, and the required overtime. The models were solved using the software Maximal, and the results showed that the approach proposed by Bertsimas and Sim was better than Soyster's method since the latter was considered by the authors excessively conservative.

Lalami et al. [12] developed an MPS model based on MILP for the automotive powertrain industry. The objective was to minimize the total costs associated with four different productive aspects: (1) meeting forecast demand; (2) satisfying safety stock levels; (3) balancing stock levels between products; (4) leveling the production percentage of each product. The priority and trade-offs among these four aspects were set by the manager using weighting factors. The inputs were related to demand levels, initial inventory, productivity rates, and fixed costs. The constraints involved production capacity, packing size, minimum and maximum quantities to be produced, and specific production on certain days. The main output was the set of all weekly quantities to be produced of each product in each production line, from which some secondary outputs can be calculated, such as estimated inventory levels. The model was solved using CPLEX, and once the correct parameters and weights were determined, the theoretical conclusions matched reality.

Detailed Production Schedule

Sawik et al. [13] developed the DPS based on a MILP model for a real-world hybrid flow shop with multi-capacity machines and batch processing mode. Considering various due dates and production-related data, the objective of the model was to split and distribute customer orders among planning periods so that all orders were fulfilled with the minimum possible deviation over a certain due date. Scheduling both divisible and indivisible orders were addressed while considering capacity constraints. The output was a detailed schedule showing the optimal allocation of orders among planning periods in multiple machines. CPLEX was used to solve the model, and the results indicated that the formulation regarding indivisible orders was stronger than the corresponding formulation for divisible orders.

Cerdá et al. [14] used MILP to develop the DPS of a chemical company that produces in batches, involving a set of single-stage processes that can be performed on parallel machines. The goal was to determine each machine's ideal sequence of operations considering a continuous time domain and constraints related to the BOM structure. Three different minimization objectives were addressed: (1) total lateness, (2) schedule makespan, and (3) the total number of late orders. The models were solved using the General Algebraic Modeling System (GAMS) based on advanced

branch-and-bound algorithms, enhancing the computational performance while providing reliable production schedules.

Another way of determining the production sequence is using a production priority rule. This approach determines the sequence in which products will be processed based on a calculated parameter that represents some key aspect within the production environment. In static environments, the production orders have fixed priority parameters and are scheduled assuming that each workstation is idle and ready to work, leading to a static production sequence [15]. The First Come First Served (FCFS) is a common static priority rule, which states that all production orders are assigned in the sequence they arrive. The Shortest Processing Time (SPT) rule prioritizes the production orders based on the total processing time, from the shortest to the longest, working in the opposite way of the Longest Processing Time (LPT) rule. The Earliest Due Date (EDD) rule prioritizes all production orders considering the due date, starting from the earliest to the latest [6].

In dynamic production environments, all orders have priority parameters that can vary with time and/or production progression, which originates a dynamic production sequence [15]. The Critical Ratio (CR) is a very usual dynamic priority rule where production orders are assigned based on the ratio "due date/remaining lead time" from the lowest to the highest value; if the ratio is below 1.0, then the production is behind schedule; if it is above 1.0, it means that the production is ahead of schedule. Another dynamic priority rule is the Least Work Remaining (LWR), which is a variant of SPT that considers all remaining processing time. The Fewest Operations Remaining (FOR) rule is also a SPT extension; however, the dynamic aspect is related to the number of remaining successive operations. The Slack Time (ST) rule is an extension of EDD that discounts setup and process times.

Integer Programming

As seen through the literature review, most of the proposed models use linear optimization techniques. In operations research, LP is an optimization method where the model's objective is represented using a linear function, the so-called objective function, which can be either maximized or minimized depending on the model's goal. The model's boundary conditions are incorporated based on a set of linear constraints, which limit the feasible solution domain. The decision variables correspond to the values related to the quantifiable decisions to be made. When an LP model is solved, the result is the optimal set of values of the decision variables that generally lies at a corner of the feasible region that has the most favorable value of the objective function. As a result, it can be stated that any LP model has three core

elements: (1) the objective function, (2) the constraints, and (3) the decision variables [3].

When an LP problem allows only integer decision variables, the model is known as Integer Programming (IP). If the problem enables the decision variables to assume both integer and non-integer values, then the model is designated as Mixed-Integer Linear Programming (MILP) [3]. As shown by the literature review, variations of the IP and MILP models also exist. For instance, the Integer Weighted Goal Programming (IWGP) method used by Omar et al. [8] is based on IP; however, some terms of the objective function are multiplied by weights in order to prioritize one aspect over another. Similarly, the Fuzzy Mixed-Integer Linear Programming (FMILP) model also used by Omar et al. [7, 8] is essentially a MILP model that involves uncertainties based on fuzzy mathematics.

Concerning the entire literature review, two main goals were identified as objective functions: cost minimization [7–9, 12] and lateness minimization [13, 14].

Each optimization model's constraints can be unique since different production environments face different restrictions. Despite this, some constraints are notably common among different models proposed in the literature review, such as (1) machine capacity, (2) available workforce, (3) available resources, (4) minimum quantity to be produced, (5) maximum storage capacity, (6) hierarchy of operations, (7) setup time, and (8) due dates.

As stated in the literature review, the decision variables in terms of APP and MPS optimization usually represent the (1) quantities to be produced for each product, in each workstation, during a certain period; the (2) quantities of each product that must be outsourced per period; the (3) number of workers needed per period; the (4) backorder level per product per period; the (5) occurrence of tool change per period (and the consequent setup time), usually expressed using binary values. Some of these decision variables can be found in the works developed by Omar et al. [7, 8], Martín et al. [9], and Lalami et al. [12].

Concerning the DPS, the works from Sawik et al. [13] and Cerdá et al. [14] addressed a combination of the following decision variables: the (1) number of periods that certain order is early or late; the (2) information that certain order is early or late; the (3) information that certain order will start being produced at a determined period; the (4) representation of the production hierarchy between two parts accordingly to its BOM. The decision variables (2), (3), and (4) are expressed using binary variables. It was not found works related to dynamic sequencing rules that were relevant to the scope of the framework.

Methodology

This section aims to describe the methodology used to develop the proposed framework. Firstly, the company under study and its context are briefly described. After, the MPS optimization is presented and discussed, followed by the proposed dynamic production sequencing rule that is the basis of the DPS.

Context Setting

The company under study is a medium-sized firm that produces cold stamping parts of exhaust systems for the automobile industry. These parts are sold to customers who assemble the full exhaust system. Specifically, the company is in the middle of the automobile industry supply chain, and therefore, there is huge pressure on accomplishing due dates.

The company has four progressive hydraulic presses in its production sector: (1) Zani 600 tons, (2) Rovetta I 600 tons, (3) Rovetta II 600 tons, and (4) Cattaneo 1000 tons. Each press has a limited capacity and can produce a specific range of products, but some of them can be manufactured on multiple machines. Despite this, the proposed framework does not have to assign products to a specific press since the company already knows which machine (first option) is more efficient for each product based on unitary costs and productivity rates. The alternative machine (second option) is only used when the primary machine is unavailable because of corrective maintenance or breakdown. Due to the need for high-volume production, limited number of machines, and high setup times, the company's strategy relies on inventory to manage unexpected demand.

The manufacturing process begins by receiving raw materials, which are inspected for quality and stored in a warehouse until needed. When demand occurs, these raw materials are processed using one of the progressive cold stamping machines, transforming them into disks, blanks, or shaped parts. If the outputs are disks or blanks, they must be stored in the corresponding storage area until the next stamping cycle is scheduled. Regarding the shaped parts, after the stamping process, they must be washed in order to remove the lubricant and any other remnants of the process. Depending on the project specifications, some of these shaped parts must be welded and/or trimmed before they can be considered finished products and stored in the warehouse. If no additional manufacturing operations are required after washing, these shaped parts are considered finished products and stored in the warehouse. Throughout this entire

Table 1 Framework notation

<i>Designation</i>	Description
<i>i</i>	Product code to be manufactured ($i=1, 2, 3, \dots, 10$)
<i>j</i>	Week designation ($j= 1, 2, 3, 4$)
<i>Variables</i>	
X_{ij}	Quantity of product <i>i</i> to be produced during week <i>j</i>
ϕ_{ij}	Binary value denoting the tool change related to product <i>i</i> during week <i>j</i> , where 1 indicates that a tool change occurs and 0 the contrary
WPI_{ij}	Weekly priority index for the product <i>i</i> during the week <i>j</i>
Π_{ij}	Initial inventory of product <i>i</i> at the beginning of week <i>j</i> [unit]
<i>Costs</i>	
TC	Total cost [€]
MC_i	Manufacturing cost per unit of product <i>i</i> [€/unit]
IHC_i	Inventory holding cost per week for one unit of product <i>i</i> [€/week·unit]
UMC	Unproductive machine cost [€/h]
<i>Parameters</i>	
SS_i	Safety stock level for product <i>i</i> [unit]
ST_i	Setup time related to product <i>i</i> [h]
D_{ij}	Demand for product <i>i</i> during week <i>j</i> [unit]
S_j	Number of shifts available to work during week <i>j</i> [unit]
P_i	Time required to produce one unit of product <i>i</i> [h/unit]
M_i	Maximum quantity allowed to be produced per product <i>i</i> [unit]
Mt_{ij}	Maximum theoretical quantity of product <i>i</i> produced per week <i>j</i> [unit]

production process, only the cold stamping sector presents bottlenecks, and therefore this is the main scope of the proposed framework.

Notation

This section presents the notation used in the formulation of the framework that is summarized in Table 1.

For both first and second levels of the framework, the sub-index *i* can vary from 1 to 10 since this is the maximum amount of different products a single machine produces per month; and the sub-index *j* ranges from 1 to 4 weeks, corresponding to a planning horizon of 1 month.

First Level

The company works with production forecasts that support the medium-term plan. These forecasts provide aggregated demand volumes that must be met over the next 4 months, although they are subject to change. In other words, this means that the company always has a 4-month APP in place. The production volumes are only disaggregated for the current month, that is, for the next 28 days from the present. This process generates deterministic weekly demand for the next 4 weeks for each product.

Based on the disaggregated demand levels, each product’s weekly quantity to be produced is determined, originating the MPS. This process allows scheduling the machine that

must operate, the necessary tool changes, resource usage, and estimated inventory levels of both finished and unfinished products at the end of each week.

In terms of responsible production, the company always seeks to meet customer demand while minimizing the costs associated with the manufacturing process, ensuring that only the necessary resources are used. In this sense, the first level of the framework is the MPS optimization using a MILP model (1), involving a single machine at a time, with the objective of minimizing the total cost (TC), which is composed of three key components: (1) the manufacturing cost (MC_i), given in €/unit; (2) the inventory holding cost (IHC_i), measured in €/week·unit; (3) the unproductive machine cost (UMC), given in €/h:

$$\begin{aligned}
 \text{Minimize} \quad TC = & \sum_{i=1}^{10} \sum_{j=1}^4 MC_i \cdot X_{ij} \\
 & + \sum_{i=1}^{10} \sum_{j=1}^4 IHC_i \cdot (X_{ij} + \Pi_{ij} - D_{ij}) \\
 & + \sum_{i=1}^{10} \sum_{j=1}^4 UMC \cdot (7.25 \cdot S_j - P_i \cdot X_{ij})
 \end{aligned} \tag{1}$$

$$\begin{aligned}
 s.t. \quad X_{ij} \geq & SS_i + D_{ij} - \Pi_{ij} \quad i = \{1, 2, \dots, 10\} \\
 & j = \{1, 2, 3, 4\}
 \end{aligned} \tag{C1}$$

$$\sum_{i=1}^{10} (X_{ij} \cdot P_i + \phi_{ij} \cdot ST_i) \leq 7.25 \cdot S_j \quad j = \{1, 2, 3, 4\} \quad (C2)$$

$$\sum_{j=1}^4 X_{ij} \leq M_i \quad i = \{1, 2, \dots, 10\} \quad (C3)$$

$$X_{ij} - \phi_{ij} \cdot Mt_{ij} \leq 0 \quad i = \{1, 2, \dots, 10\} \quad j = \{1, 2, 3, 4\} \quad (C4)$$

$$X_{ij} \geq 0 \quad i = \{1, 2, \dots, 10\} \quad j = \{1, 2, 3, 4\} \quad (C5)$$

$$\phi_{ij} \in \{0, 1\} \quad i = \{1, 2, \dots, 10\} \quad j = \{1, 2, 3, 4\} \quad (C6)$$

Among the components of the objective function, the MC_i represents the manufacturing cost of producing one unit of the product i , encompassing the costs of the hydraulic press operating, raw material, and labor. The IHC_i is the estimated weekly inventory holding cost for each surplus unit of the product i . Since the products manufactured by the company are cold-stamped metal parts, they do not require special care other than maintaining them in the warehouse and paying the staff involved. However, stored parts are financial resources that could generate tangible returns for the company if used differently, representing an opportunity cost. The term UMC represents the hourly cost of the machine when it is not producing due to maintenance, tool change (setup time), breakdown, or idle state.

The company assumes that its production environment is based on limited loading, meaning that the machine has a finite available capacity. The main goal of this type of planning is to use the available time as effectively as possible. To achieve this, the two main variables that need to be considered are (1) the quantities to be produced of each product and (2) the setup time attached to the correspondent tool change. The tool change holds significant importance as the machine can only manufacture one product type at a time. Therefore, whenever a different product needs to be manufactured, a corresponding tool change becomes necessary. In terms of the proposed optimization model (1), these two main outputs represent the decision variables and are defined as follows: X_{ij} is the quantity of the product i to be produced during the week j , whereas ϕ_{ij} is a binary value denoting whether or not a tool change related to each product i takes place during the week j , represented by 1 and 0, respectively.

In the real world, the variables X_{ij} are always integer numbers because they denote quantities that need to be produced. However, keeping them as integers in the model (1) often led to hours of resolving time, especially in the case of overload scenarios. To make calculations faster by reducing the computational burden, it was decided to keep X_{ij} as real numbers. This approach made the model capable of being solved in a

few minutes. The effect of this simplification on the result is minimal since the company only produces quantities in the order of thousands of units. Thus, having a non-integer number as a result and then rounding it does not cause any major problem. In the case of ϕ_{ij} , since it is a binary variable, it is always going to be an integer number. Because of these assumptions involving the decision variables, the model (1) is classified as a MILP problem.

To reflect the reality of the company's production process, six constraints are considered: (1) the minimum quantity to be produced to satisfy the demand (C1), (2) the machine availability (C2), (3) the maximum quantity permitted to be produced (C3), (4) the dependence between production and tool setup (C4); (5) the nonnegativity constraint (C5); and (6) the requirement that the variable denoting the tool change is a binary value (C6).

The minimum quantity to be produced per week (C1) is calculated using the safety stock level (SS_i), the demand (D_{ij}), and the week's initial inventory (Π_{ij}). It is important to state that the initial inventory of each product i during the first week (Π_{i1}) is an input. In contrast, for every other week $j > 1$, the corresponding initial inventory is calculated by $\Pi_{ij} = \Pi_{i(j-1)} + X_{i(j-1)} - D_{i(j-1)}$. To absorb potential market fluctuations or unexpected setbacks, the company employs a safety stock level (SS_i).

As previously stated, the company's scheduling approach is based on limited loading, so the total available time to produce (C2) must be sufficient to cover all product units that will be manufactured and the required tool changes and setup time. In practice, knowing the number of available working shifts (S_j) in each week j and the productive hours per shift (7.25 h), one can calculate how many working hours the machine will be able to operate over the planning horizon. Based on the average time to produce one unit of each product (P_i) and the occurrence or not of tool change (ϕ_{ij}), these available working hours affect the upper limit on the total number of parts that can be manufactured.

The maximum quantity allowed to be produced (C3) of each product i during all 4 weeks should be used to limit the domain of the decision variables to a more restrictive level than the time constraint (C2). Hence, it requires the manager's expertise, as this production limit may depend on the combined effect of different factors.

Every time each product i is manufactured, the constraint (C4) ensures that a corresponding tool change is necessary. The main required parameter is denoted by Mt_{ij} , a theoretical value defining how many units would be manufactured if only a certain product i was produced during the week j . As can be seen by (C4), given that Mt_{ij} is always greater than 0, whenever X_{ij} is also greater than 0, then ϕ_{ij} is equal to 1; if X_{ij} is equal to 0, then mathematically ϕ_{ij} could be equal to either 0 or 1 and (C4) still would be satisfied. However,

in the latter case, ϕ_{ij} is always indirectly led to be equal to 0 because of the unproductive machine cost term included in the minimization of the objective function (UMC).

Second Level

The production sequencing rule is based on a Weekly Priority Index (WPI_{ij}). For each week j , the WPI_{ij} of every product i that needs to be produced in a certain machine must be calculated. Then, the production sequence of that machine for the week j is determined based on the higher order of the WPI_{ij}, implying that higher priority is given to products with a higher WPI_{ij} score.

The WPI_{ij} represents the production priority of each product i based on its weekly quantities to be produced in week j , with 1 meaning the highest priority and 0 meaning the lowest. For each week j , the product i with the largest quantity to be produced (X_{ij}) will have a WPI_{ij} equal to 1, whereas the remaining products whose production volumes are lower, will have a WPI_{ij} proportional to their respective X_{ij} . In order to calculate the WPI_{ij}, it is necessary to define the production volume matrix (XM_j), accordingly to expression (2):

$$XM_j = \begin{bmatrix} X_{1j} \\ X_{2j} \\ \dots \\ X_{ij} \end{bmatrix} \quad j = \{1, 2, 3, 4\} \tag{2}$$

In XM_j , each element X_{ij} represents the optimal quantity of product i to be produced during week j , determined by the MILP model (1) on the first level. This optimal quantity takes into account the demand (D_{ij}) for the product and aims to balance the initial inventory (II_{ij}) with the safety stock level (SS_i), as described before. At this stage, the Weekly Priority Index (WPI_{ij}) for each product i , during week j , can be calculated according to expression (3):

$$WPI_{ij} = \begin{cases} \frac{X_{ij}}{MAX(X_{ij})}, & \text{if product } i \text{ is not the first scheduled for week } j+1 \\ 0, & \text{if product } i \text{ is the first scheduled for the following week } j+1 \end{cases} \tag{3}$$

Where $\frac{X_{ij}}{MAX(X_{ij})}$ is the normalized value of X_{ij} based on the largest element of XM_j .

In addition to establishing the production sequence, the WPI_{ij} also links two successive weeks j and $j + 1$ using a common product i that must be manufactured on both weeks. In such case, if the product i is the first scheduled for the week $j + 1$, then, accordingly to expression (3), its corresponding WPI_{ij} must be equal to zero, forcing this product to be the last produced on week j . This way, an unnecessary setup time is avoided since the right tool will already be installed at the beginning of the week $j + 1$.

Since the WPI_{ij} calculation for the week j depends on the production sequence of the following week $j + 1$, the WPI_{ij} must be calculated in reverse order from the last to the first week. However, since there is no information about the production sequence of the fifth week, for every product i produced during the fourth week, the WPI_{i4} must be calculated simply by $\frac{X_{i4}}{MAX(X_{i4})}$. This simplification is not a problem since the framework is intended to be used on a rolling planning horizon at the end of each week, meaning that the production sequence of the fourth week will be revised at least three times until it occurs.

Once the production sequence of the last week is defined, the production sequence for the third, second, and first week can be calculated, in this order, until the entire planning horizon is covered.

Data

The data used to test, validate, and implement the model are related to the company’s products, machines, and inventory. Since this information is known and defined from the start, they are considered parameters, meaning they do not introduce uncertainty into the model. In the end, each parameter is related to one or more of the three components of the total cost: (1) the manufacturing costs (MC_{*i*}); (2) the inventory holding cost (IHC_{*i*}); (3) the unproductive machine cost (UMC). Table 2 shows the statistics about the cost data.

All product structure information was obtained directly through a formal document, containing all levels of each product, from raw material to the finished good. In terms of cold stamping, among the entire company’s portfolio, the vast majority of parts is only submitted to a single stamping step, but on the other hand, there is a range of products that must go through two subsequent stamping steps, as described before. Despite being dependent on each other in terms of manufacturing, these two subsequent processes

can be considered independent in the proposed framework, simply because they occur at two very different moments. In practice, the first stamping process generates disks or blanks

Table 2 Statistics of data related to costs

Parameter	Maximum	Minimum	Average	Std. dev.	Unit
MC _{<i>i</i>}	3.35	0.04	0.60	0.48	€/unit
IHC _{<i>i</i>}	0.20	0.01	0.03	0.03	€/week·unit
UMC	88.16	43.65	61.83	18.77	€/h

that are stocked to be used later, meaning a complete planning cycle. When there is demand, the second stamping process is planned and executed, consuming some of those disks or blanks and leading to the product in its final shape, meaning another complete planning cycle. The same disk type can be used for more than one product, which is why the company adopts the strategy of stocking disks and blanks, seeking to absorb any fluctuations or unforeseen events.

All machine data were easily accessible, as it was only necessary to regroup and rearrange them. The main set of information was obtained from an electronic spreadsheet frequently used by the company. For all hydraulic presses, it was possible to extract what products each of them can produce, their respective average production rates, the costs of each stamping process (MC_i), and their unproductive costs (UMC).

The information about inventory costs (IHC_i) for both finished and unfinished products required processing before they could be used. With the company's support, for each product, finished or not, storage costs were determined based on the volume that each product container occupies inside the warehouse and the logistic costs involved to manipulate it.

Results and Discussion

The framework was validated by comparing a real production plan created and executed by the company with a plan suggested by the framework, demonstrating that the tool provides an optimized and feasible production plan.

The company used the machine Rovetta II to produce four required products during the 4 weeks from December 12th, 2022, to January 6th, 2023. The machine operated for 5 days during the first, second, and fourth weeks; and 4 days during the third week, which was affected by Christmas and New Year's day. During the first week, three working shifts per day were required, in contrast to the regular two shifts per day of the following weeks. Each shift has approximately 7.25 h long, with breaks and preparation time already excluded. Table 3 presents the complete planning horizon description.

Among the four products that needed to be produced, products 1 and 2 frequently have high demands,

Table 3 Time to produce along the entire planning horizon

Week	Start date	Days	Shifts/day	h/shift	h
1	12/12/22	5	3	7.25	108.75
2	19/12/22	5	2	7.25	72.50
3	26/12/22	4	2	7.25	58.00
4	02/01/23	5	2	7.25	72.50

Table 4 Initial inventory and weekly demands over the planning horizon

Product	II_{i1}	D_{i1}	D_{i2}	D_{i3}	D_{i4}
1	3606	12,600	7800	5400	20,400
2	14,415	12,768	8736	6048	19,488
3	69	198	33	0	33
4	0	0	960	480	1920

making them high-runners. In contrast, products 3 and 4 typically have low demand and therefore are categorized as low-runners.

Due to the high demand, high-runners often have a large initial inventory, as they are frequently produced. On the other hand, when a low-runner must be produced, its initial inventory is often low since the company's strategy consists of producing them in quantities much greater than their demand, then gradually consuming the generated inventory until it is necessary to produce these products again. These aspects can be seen in Table 4, which presents the initial inventory (II_{i1}) and demands (D_{ij}) for each product i per week j within the planning horizon.

Based on the initial scenario presented in Table 4, the company developed and executed the production plan shown in Table 5.

Regarding the optimal plan, no safety level (SS_i) was initially set, mainly because this was not clearly defined on the real plan either. However, not considering a safety level resulted in a plan with several unproductive hours and, at some weeks, null inventory levels. Such plans are not practical as they put the company at risk of not meeting due dates if any problem occurs. As a solution, it was decided to assign a safety level equal to 10,000 units for both high-runners (SS_1 and SS_2), and equal to 5000 units for the low-runners (SS_3 and SS_4). This configuration led to the optimal plan presented in Table 6.

Regarding the high-runners, the optimal plan features a more consistent inventory level throughout the planning horizon, especially for product 2. In the optimal plan, the estimated initial inventory for product 2 at the beginning of the fourth week (II_{24}) is 10,000 units, which is equal to the safety stock level for the high-runners. In the real plan, this initial inventory was equal to 47,898 units. This difference proves that the optimal plan only recommends producing enough to meet demand and satisfy the safety stock level.

In terms of the low-runners, by establishing a safety stock level of 5000 units for products 3 and 4 (SS_3 and SS_4), the optimal plan suggests producing them in greater quantities than their total demand, which is aligned with the company's strategy of generating inventory for such products. However, while the optimal plan recommends producing a total of

Table 5 Real production plan

Product	Week 1			Week 2			Week 3			Week 4			$\sum_{j=1}^4 X_{ij}$
	Π_{i1}	D_{i1}	X_{i1}	Π_{i2}	D_{i2}	X_{i2}	Π_{i3}	D_{i3}	X_{i3}	Π_{i4}	D_{i4}	X_{i4}	
1	3606	12,600	22,000	13,006	7800	0	5206	5400	10,000	9806	20,400	31,280	63,280
2	14,415	12,768	9000	10,647	8736	30,000	31,911	6048	22,035	47,898	19,488	0	61,035
3	69	198	7439	7310	33	0	7277	0	0	7277	33	0	7439
4	0	0	0	0	960	5000	4040	480	0	3560	1920	0	5000

Table 6 Optimal production plan

Product	Week 1			Week 2			Week 3			Week 4			$\sum_{j=1}^4 X_{ij}$
	Π_{i1}	D_{i1}	X_{i1}	Π_{i2}	D_{i2}	X_{i2}	Π_{i3}	D_{i3}	X_{i3}	Π_{i4}	D_{i4}	X_{i4}	
1	3606	12,600	18,994	10,000	7800	7800	10,000	5400	13,648	18,248	20,400	121,523	52,594
2	14,415	12,768	8353	10,000	8736	8736	10,000	6048	6048	10,000	19,488	19,488	42,625
3	69	198	5195	5066	33	0	5033	0	0	5033	33	0	5195
4	0	0	5000	5000	960	1781	5821	480	1579	6920	1920	0	8360

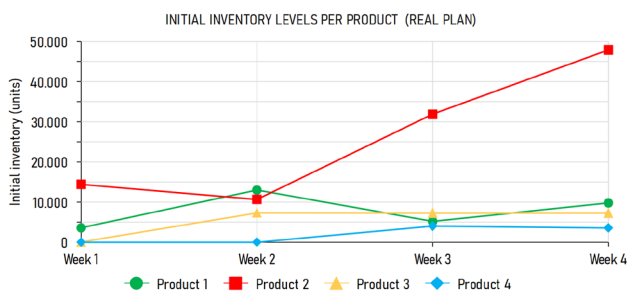


Fig. 1 Initial inventory levels per product *i* during the real plan

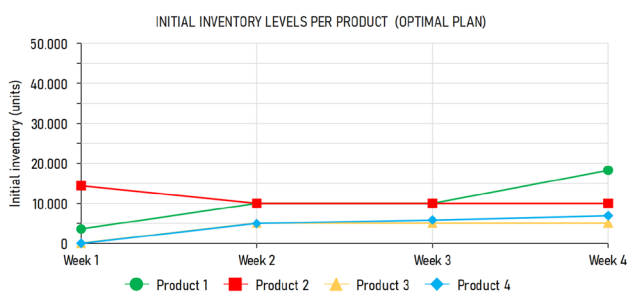


Fig. 2 Initial inventory levels per product *i* during the optimal plan

5195 units of product 3 and 8360 units of product 4, the real plan shows that, in reality, 7439 units of product 3 and 5000 units of product 4 were produced. These differences evidence that the real plan focused on product 3, whereas the optimal plan prioritizes product 4, which is more coherent due to its higher demand. Figures 1 and 2 show the

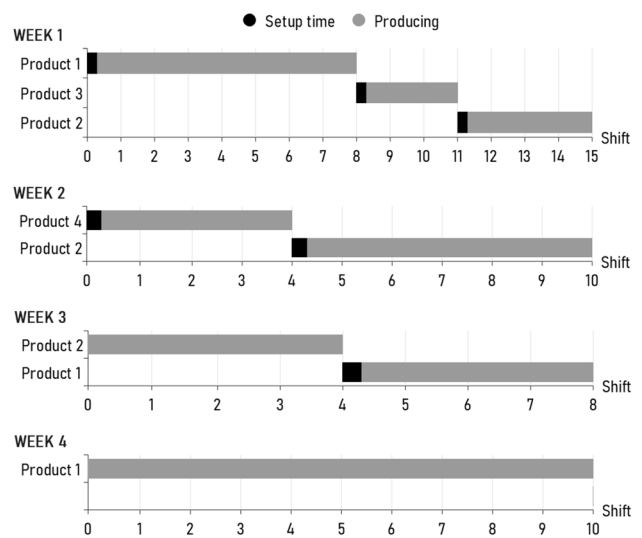


Fig. 3 Production schedule of the real plan

differences on the inventory level for the real plan and the optimal plan, respectively.

The manager manually established the production sequence of the real plan, resulting in the production schedule shown in Fig. 3.

The production schedule shows that the machine was operating most of the time, only stopping during six tool changes, represented by the setup times in Fig. 3.

The production sequence of the optimal plan is determined by the second level of the proposed framework, which

Table 7 Obtained production sequence for the optimal plan

Week 1		Week 2		Week 3		Week 4	
Sequence	WPI _{i1}	Sequence	WPI _{i2}	Sequence	WPI _{i3}	Sequence	WPI _{i4}
1st: Product 1	1.00	1st: Product 2	1.00	1st: Product 1	1.00	1st: Product 2	1.00
2nd: Product 3	0.27	2nd: Product 4	0.20	2nd: Product 4	0.12	2nd: Product 1	0.62
3rd: Product 4	0.26	3rd: Product 1	0.00	3rd: Product 2	0.00	–	–
4th: Product 2	0.00	–	–	–	–	–	–

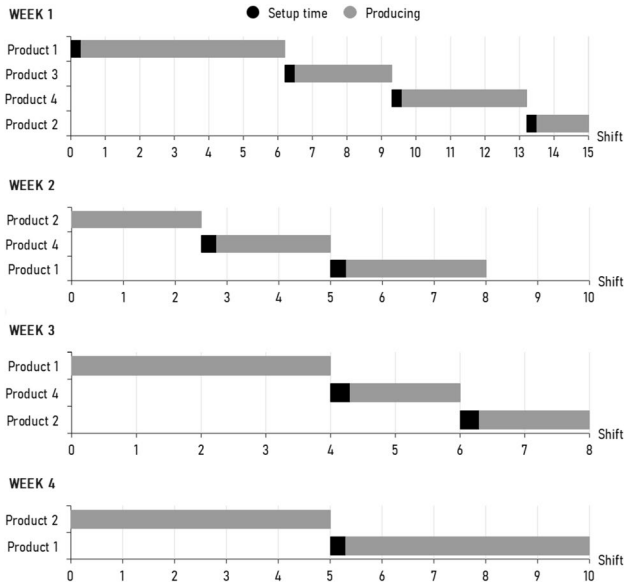


Fig. 4 Production schedule of the optimal plan

Table 8 Total cost for the real and optimal plans

Plan	Costs (€)			
	MC _i	IHC _i	UMC	TC
Real	77,130.09	7164.80	30.23	84,325.12
Optimal	59,916.68	4219.39	1546.69	65,682.76
Variation (%)	-22.3	-41.1	+ 5016.4	-22.1

is based on the Weekly Priority Index (WPI_{ij}). The WPI_{ij} of each product *i* per week *j* is presented in Table 7.

As a result, the framework is capable of presenting the optimal production schedule shown in Fig. 4.

The first aspect seen in the production schedule is that all weeks have their plans connected by a common product, which is the expected outcome of the proposed production sequencing rule. In addition, since the optimal plan aims to produce as close to demand as possible, it proposes nine tool changes and setup times, which is higher than the six tool changes that occurred in the real plan. Lastly, it can be seen that the optimal plan predicts some unproductive time during the second week.

The total cost (TC) to execute the real plan is estimated using an expression equal to the objective function of the MILP model (1) and the data presented in Table 5 as input. The total cost includes the manufacturing costs (MC_i), the inventory holding cost (IHC_i), and the unproductive machine cost (UMC). Table 8 shows the direct cost comparison between both plans.

The results indicate that implementing the optimal plan could result in a 22.1% decrease in the total cost (TC). Specifically, this would involve a reduction of 22.3% in the manufacturing cost (MC_i), a 41.1% decrease in the inventory holding cost (IHC_i), and a 5016.4% increase in the unproductive machine cost (UMC). It is important to note that this significant change in UMC is only due to the residual cost of € 30.23 observed in the real plan.

The decrease in the manufacturing cost (MC_i) is because the optimal plan only recommends producing the necessary quantities. The framework aims to keep every inventory of each product *i* close to its safety level (SS_i), only producing in larger quantities if there is a need to anticipate production. This is evident by comparing the total production of the real plan, which produced 136,754 units, with the optimal plan, which suggested producing only 108,774 units, representing 20.5% less. Since both plans meet all demands, it is clear that, in reality, excessive inventories were generated by the real plan.

This lean inventory approach of the optimal plan directly affects the inventory holding cost (IHC_i) since fewer units stored mean lower cost. In addition, the optimal plan also prioritizes producing as close to demand as possible, which requires more frequent tool changes, increasing the unproductive machine cost (UMC), and reducing the inventory levels even more, and consequently, the inventory holding cost (IHC_i). This tradeoff between the unproductive machine cost (UMC) and inventory holding cost (IHC_i) is a crucial aspect of the optimal plan that helps reducing the total cost (TC) in order to obtain a sustainable production.

Conclusion

The study aimed to create a framework for short-term production planning and scheduling for a medium-sized Portuguese company that operates as a tier 2 supplier within the

automotive production chain, producing cold-stamped parts for exhaust systems using four progressive hydraulic presses. The framework is designed to support the decision-making process during the short-term planning of these machines, determining the optimal quantities to be produced per week and suggesting a production sequence, ultimately resulting in a complete production schedule covering a 4-week planning horizon.

The two levels of the framework include: (1) a MILP model to determine the optimal quantities to be produced during each week in order to minimize total costs; (2) a dynamic production sequencing rule that defines the production schedule based on the first level results. Both levels of the framework were integrated and implemented into a single Excel spreadsheet.

The validation was based on a comparison between an optimal plan proposed by the framework with a plan developed and executed by the company between the end of 2022 and the beginning of 2023. The results indicate that the optimal production plan could save up to 22.1% of the total cost observed in reality, producing as close to the demand. The production sequencing rule was validated and approved by the company's manager.

The framework is capable of absorbing uncertainties due to the rolling planning approach, which means that only the first week of the proposed plan will be executed while the other 3 weeks will be re-planned until they occur, so any variation in demand or available time to produce is incorporated each time the plan is reviewed.

To this point, the framework has yet to be fully implemented in the company's production planning procedure. Before this, the manager must use the framework in parallel with the current procedure, comparing both plans to identify any necessary changes.

The limitations of the framework include: (1) to plan a single machine at a time; (2) to consider a maximum of 10 different products over the planning horizon; (3) the MILP model does not have memory regarding the setup time between two consecutive weeks, despite this being considered by the dynamic production sequencing rule.

Therefore, it is suggested that future work involves monitoring the implementation of the framework and, if necessary, improving it until it can fully replace the current planning procedure. In addition, it would be interesting to incorporate the MILP model into a more powerful solver to evaluate the impacts on the results and conduct sensitivity and post-optimization assessments. Finally, it is suggested to evaluate the effects of a planning horizon longer than 4 weeks, aiming to determine whether a longer time frame would allow the framework to react even more effectively.

Acknowledgements This work has been supported by Foundation for Science and Technology (FCT, Portugal) for financial support through

national funds FCT/MCTES (PIDDAC) to CeDRI (UIDB/05757/2020 and UIDP/05757/2020) and SusTEC (LA/P/0007/2021).

Funding Open access funding provided by FCTIFCCN (b-on).

Data availability Due to the nature of the research, due to commercial supporting, data is not available.

Declarations

Conflict of Interest On behalf of all the authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. United Nations: 17 sustainable development goals. <https://sdgs.un.org/goals>. Accessed 1 Oct 2022.
2. United Nations: sustainable development goal no 12. <https://sdgs.un.org/goals/goal12>. Accessed 1 Oct 2022.
3. Hillier FS, Lieberman GJ. Introduction to operations research: Vol. 9th edn. McGraw-Hill Higher Education; 2010.
4. Stadler H, Kilger C. Supply chain management and advanced planning: concepts, models, software and case studies. Berlin: Springer; 2005.
5. Heizer J, Render B, Munson C. Operations management: sustainability and supply chain management. 12th edn. 2015.
6. Vollmann TE, et al. Manufacturing planning and control systems for supply chain management. New York: McGraw-Hill; 2005.
7. Omar MK, Mohd-Jusoh M, Omar M. Developing a master production schedule using fuzzy mixed integer linear programming. *Int J Comput.* 2021;15:8–13. <https://doi.org/10.46300/9108.2021.15.2>.
8. Omar MK, Teo SC. Hierarchical production planning and scheduling in a multi-product, batch process environment. *Int J Prod Res.* 2007;45(5):1029–47. <https://doi.org/10.1080/00207540600724989>.
9. Martín AG, Díaz-Madroñero M, Mula J. Master production schedule using robust optimization approaches in an automobile second-tier supplier. *Cent Eur J Oper Res.* 2020;28(1):143–66. <https://doi.org/10.1007/s10100-019-00607-2>.
10. Soyster AL. Convex programming with set-inclusive constraints and applications to inexact linear programming. *Oper Res.* 1973;21(5):1154–7.
11. Bertsimas D, Sim M. The price of robustness. *Oper Res.* 2004;52(1):35–53.
12. Lalami I, Frein Y, Gayon J-P. A model for master production scheduling in automotive powertrain plants: a case study. *Int Conf Ind Eng Syst Manag.* 2015;2015:1081–90. <https://doi.org/10.1109/IESM.2015.7380289>.
13. Sawik T. Integer programming approach to production scheduling for make-to-order manufacturing. *Math Comput Model.* 2005;41(1):99–118. <https://doi.org/10.1016/j.mcm.2003.10.053>.

14. Cerdá J, Henning GP, Grossmann IE. A mixed-integer linear programming model for short-term scheduling of single-stage multiproduct batch plants with parallel lines. *Ind Eng Chem Res.* 1997;36(5):1695–707. <https://doi.org/10.1021/ie9605490>.
15. Conway RW, Maxwell WL, Miller LW. *Theory of scheduling.* Dover: Dover Books on Computer Science Series; 2003.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.