

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
PROGRAMA DE PÓS-GRADUAÇÃO
MESTRADO PROFISSIONAL EM ENGENHARIA DE PRODUÇÃO**

HENRIQUE DEMARCO

**MODELO MATEMÁTICO PARA DIMENSIONAMENTO DE LOTE E
SEQUENCIAMENTO DE PRODUÇÃO COM MÁQUINAS PARALELAS**

PORTO ALEGRE – RS

2023

HENRIQUE DEMARCO

**MODELO MATEMÁTICO PARA DIMENSIONAMENTO DE LOTE E
SEQUENCIAMENTO DE PRODUÇÃO COM MÁQUINAS PARALELAS**

Dissertação submetida ao Programa de Pós-Graduação Mestrado Profissional em Engenharia de Produção da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do título de Mestre em Engenharia de Produção, modalidade Profissional, na área de concentração em Sistemas de Produção.

Orientador: Professor, Dr. Ricardo Augusto Cassel

PORTO ALEGRE – RS

2023

HENRIQUE DEMARCO

**MODELO MATEMÁTICO PARA DIMENSIONAMENTO DE LOTE E
SEQUENCIAMENTO DE PRODUÇÃO COM MÁQUINAS PARALELAS**

Esta dissertação foi julgada adequada para a obtenção do título de Mestre em Engenharia de Produção, na modalidade Profissional, e aprovada em sua forma final pelo Orientador e pela Banca Examinadora designada pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Rio Grande do Sul.

Prof. Dr. Ricardo Augusto Cassel

Orientador PMPEP/UFRGS

Prof^a. Dr^a. Christine Tessele Nodari

Coordenadora PMPEP/UFRGS

Banca Examinadora:

Professora Giovana Savitri Pasa, Dr^a. (UFRGS)

Professor João Carlos Furtado, Dr. (UNISC)

Professor Michel José Anzanello, Ph.D. (UFRGS)

AGRADECIMENTOS

Meus mais sinceros agradecimentos ao contribuinte brasileiro, que com o seu esforço e trabalho torna-se o principal fomentador da ciência em nosso país, ainda que, para isto, deixe de prover de sua própria subsistência.

RESUMO

A programação da produção é atividade primordial dentro das indústrias e passa a ser mais importante quanto mais complexa é a operação de cada empresa. À medida que o ambiente produtivo aumenta em recursos, produtos e restrições, torna-se mais difícil alcançar a programação ótima dada a quantidade de possibilidades e combinações distintas. A execução desta atividade com o devido cuidado pode gerar economia de recursos, maximização das horas de maquinário convertidas em produção, entre outros benefícios. O uso de modelos matemáticos e computadores para esta atividade possibilita testar uma grande quantidade de combinações das mais diferentes variáveis e, no limite, encontrar a proposta de solução ótima e que faz a melhor alocação dos recursos escassos. Dentro deste escopo, o presente trabalho apresenta um modelo matemático, acompanhado do código computacional em *Python*. Este é aplicável em indústrias que utilizam máquinas paralelas heterogêneas de capacidade finita, com *setups* dependentes (tanto em custo quanto em tempo) da sequência de programação, com produtos em estágio único de processamento, observando restrições de estoque de componentes, horas-homem disponíveis e capacidade de armazenagem de produtos acabados. Complementando as restrições mencionadas, o modelo tem por objetivo reduzir o custo de execução da programação. Para isto, considera os custos de *setup*, os custos de manutenção de estoque, a penalidade financeira por não atender o pedido no prazo especificado pelo cliente e o custo de produção do item vendido considerando que as máquinas são heterogêneas e, por consequência, apresentam custos diferentes para produzir o mesmo item. O modelo executado resolveu o problema para uma indústria com 18 máquinas, 350 produtos acabados, 4576 componentes, 2 macro períodos de programação, cada um contendo 5 micro períodos variáveis. A solução ótima foi encontrada em 10h de processamento utilizando um computador com processador Intel i5-2410M 2.30GHz *quad-core* com 6GB de memória RAM, *solver* Gurobi, tendo sido programado em linguagem *Python* com a biblioteca PuLP. Muito embora este tempo total de processamento possa ser considerado alto para uma aplicação prática, o modelo atingiu o *gap* de 1% do melhor resultado em aproximadamente 2h.

Palavras-chave: programação da produção; dimensionamento; sequenciamento; modelo matemático; Python; Gurobi.

ABSTRACT

Production scheduling is a primary activity within industries and becomes more important the more complex each company's operation is. As the productive environment increases in resources, products and constraints, it becomes harder to achieve the optimal lot-sizing and scheduling given the number of possibilities and different combinations. Carrying out this activity can save resources, maximize machine hours converted into production, among other benefits. The use of mathematical models and computers for this activity makes it possible to test a large number of combinations of different variables and, in the limit, to find the optimal solution that makes the best allocation of scarce resources. Within this scope, this dissertation presents a mathematical model, accompanied by computational code in Python, applicable in industries that use capacitated heterogeneous parallel machines, with dependent setups (both in cost and time) of the scheduling sequence, single-stage processing, observing component availability constraints, available man-power, and finished goods storage capacity. Complementing the mentioned restrictions, the model aims to reduce the cost of production plan execution. For this, it considers setup costs, inventory maintenance costs, back-ordering penalty and finished good production cost considering that the machines are heterogeneous and, consequently, have different costs to produce the same item. The model solved the problem for an industry with 19 machines, 480 products, 4576 components, 2 macro periods, each one containing 5 variable micro periods. The optimal solution was found in 10h of processing using a computer with an Intel i5-2410M 2.30GHz quad-core processor with 6GB of RAM, solved by Gurobi package, programmed using Python language and PuLP library. Although total processing time can be considered high for a practical application, the model reached the gap of 1% of the best result in approximately 2h.

Keywords: scheduling; lot-sizing; mathematical model; Python; Gurobi.

LISTA DE FIGURAS

Figura 1 – Pêndulo da estratégia para condução da pesquisa.....	21
Figura 2 – Esquema representativo das heurísticas abordadas.....	41
Figura 3 - Gráfico do lote econômico	50

LISTA DE TABELAS

Tabela 1 - Comparativo das etapas da Design Science Research	27
Tabela 2 - Contribuições consideradas no modelo final.....	66
Tabela 3 - Comparativo de testes.	83
Tabela 4 - Comparativo de custos do plano de produção na baixa sazonalidade.....	84
Tabela 5 - Comparativo de custos do plano de produção na alta sazonalidade.	85

LISTA DE ABREVIATURAS E SIGLAS

B&B – *Branch and Bound*

B&C – *Branch and Cut*

C&B – *Cut and Branch*

CLSD – *Capacitated Lot-sizing Problem with Sequence-Dependent Setups*

CLSP – *Capacitated Lot-Sizing Problem*

CSLP – *Continuous Setup Lot-sizing Problem*

DLSP – *Discrete Lot-sizing and Scheduling Problem*

DSR – *Design Science Research*

ENEGEP – *Encontro Nacional de Engenharia de Produção*

ERP – *Enterprise Resource Management*

GA – *Genetic Algorithm*

GH – *Greedy Heuristics*

GLSP – *General Lot-sizing and Scheduling Problem*

GLSPMS – *General Lot-sizing and Scheduling Problem Multi Stage*

GLSPPL – *General Lot-sizing and Scheduling Problem for Parallel Production Lines*

JSON – *Java Script Object Notation*

JSSP – *Job Shop Scheduling Problem*

LP – *Linear Programming*

LS – *Local Search*

MIP – *Mixed Integer Programming*

MLCLSP – *Dynamic Multi-Level Capacitated Lot-sizing Problem*

MLCLSP – *Multilevel Capacitated Lot-sizing Problem with Linked Lot Sizes*

MRP – *Material Requirements Planning*

MRPII – *Manufacturing Resources Planning*

PLSP – *Proportional Lot-sizing and Scheduling Problem*

PMP – *Plano Mestre de Produção*

PO – *Pesquisa Operacional*

R&F – *Relax and Fix*

RH – *Rounding Heuristics*

S&OP – *Sales & Operation Plan*

SA – *Simulated Annealing*

SQL – *Structured Query Language*

TS – *Tabu Search*

TSP – *Travelling Salesman Problem*

SUMÁRIO

1. INTRODUÇÃO.....	13
1.1. PROBLEMA DA PESQUISA.....	14
1.2. OBJETIVOS.....	16
1.3. JUSTIFICATIVA.....	17
1.4. DELIMITAÇÕES.....	18
1.5. ESTRUTURA DO TRABALHO.....	19
2. METODOLOGIA.....	20
2.1. MÉTODOS DE PESQUISA.....	24
2.2. MÉTODO DE TRABALHO.....	26
2.2.1. Identificação do problema.....	27
2.2.2. Revisão da literatura.....	28
2.2.3. Desenvolvimento do artefato.....	29
2.2.4. Aplicação do artefato.....	30
2.2.5. Conclusões.....	31
2.2.6. Publicação dos resultados.....	31
2.3. TÉCNICAS DE COLETA E ANÁLISE DE DADOS.....	31
3. FUNDAMENTAÇÃO TEÓRICA.....	36
3.1. PLANEJAMENTO E CONTROLE DA PRODUÇÃO.....	36
3.2. PESQUISA OPERACIONAL.....	37
3.2.1. Programação linear.....	39
3.2.2. Programação inteira.....	39
3.3. HEURÍSTICAS.....	40
3.3.1. Heurísticas de programação matemática.....	41
3.3.2. Heurísticas lagrangeanas.....	43
3.3.3. Decomposição e agregação.....	44
3.3.4. Metaheurísticas.....	44
3.3.5. Heurística gananciosa.....	46
3.4. SETUP.....	47
3.5. DIMENSIONAMENTO DE LOTES.....	50
3.6. SEQUENCIAMENTO DA PRODUÇÃO.....	52
3.7. CLASSES DE MODELOS.....	53
3.7.1. Dados de entrada.....	53
3.7.2. Períodos.....	54
3.7.3. Capacidade dos recursos.....	54

3.7.4.	Ambiente produtivo	54
3.7.5.	Estágios de produção	55
3.7.6.	Tipos de recursos.....	56
3.7.7.	Evolução dos modelos na literatura.....	57
3.8.	PUBLICAÇÕES RECENTES	62
4.	CONSTRUÇÃO DO ARTEFATO	64
4.1.	MODELAGEM MATEMÁTICA.....	66
4.2.	PROGRAMAÇÃO EM LINGUAGEM COMPUTACIONAL	70
5.	APLICAÇÃO DO ARTEFATO	77
5.1.	TESTES REALIZADOS	82
5.2.	RESULTADOS	85
6.	CONCLUSÕES	88
6.1.	LIMITAÇÕES.....	89
6.2.	TRABALHOS FUTUROS.....	90
	REFERÊNCIAS BIBLIOGRÁFICAS	92

1. INTRODUÇÃO

A mente humana é capaz de realizar tarefas complexas e incríveis ao mesmo tempo, mas apesar disto, em algumas tarefas como cálculos, não conseguimos competir com um computador. A título de comparação, o computador mais rápido do mundo conhecido no momento em que esta pesquisa é realizada, o *Frontier*, localizado no Laboratório Nacional de Oak Ridge, Tennessee, Estados Unidos, consegue realizar $1,1 \times 10^{18}$ operações matemáticas com números decimais por segundo, superando muito a capacidade do cérebro humano (CHOI, 2022; OAK RIDGE NATIONAL LABORATORY, 2022). Desta forma, o uso do computador permite otimizar o recurso tempo, sendo base para ganho de eficiência à medida em que a capacidade criativa da mente humana pode ser utilizada em outras tarefas. Ainda no contexto do uso do tempo, o computador consegue entregar melhores resultados dentro do mesmo espaço de tempo ou, alterando-se a restrição, entregar o mesmo resultado no menor tempo.

Dentre as tarefas que exigem um alto grau de processamento ou uma grande quantidade de cálculos, inclui-se a análise de alternativas para tomar uma decisão quando dispõe-se de métodos para que a decisão seja quantitativa. De acordo com Wang et al. (2008), a tomada de decisão sem um suporte quantitativo pode proporcionar riscos significativos ao negócio conforme o número de possibilidades disponíveis aumenta.

Desta lacuna emerge a Pesquisa Operacional, que significa abordagem científica para tomada de decisões e que procura determinar como melhor projetar e operar um sistema, usualmente sob condições que requerem a alocação de recursos escassos (ARENALES et al., 2015). Ainda de acordo com Arenales et al. (2015), com o aprimoramento dos métodos de solução dos modelos e o desenvolvimento tecnológico dos computadores, tem sido possível resolver modelos cada vez mais complexos, outrora intratáveis. Esta última afirmação corrobora com o parágrafo inicial deste capítulo, no sentido de que a performance dos computadores trouxe a superação da velocidade de cálculo da mente humana, gerando resultados mais condizentes com a necessidade da indústria.

Um dos problemas mais estudados pela Pesquisa Operacional é o de dimensionamento de lotes e sequenciamento da produção. O sequenciamento é definido como o processo de atribuição de recursos restritos a um conjunto de tarefas

que precisam ser cumpridas (ABEDINNIA et al., 2017). As técnicas de sequenciamento desempenham um papel importante na redução dos tempos ociosos, agilizando o processo de produção e reduzindo os custos, melhorando os processos operacionais (FU et al., 2011). Apesar do sequenciamento ser uma necessidade de outras áreas, dentro da produção torna-se altamente importante e exponencialmente complexo à medida em que restrições vão sendo adicionadas. Da complexidade surge a dificuldade de simular todas as possibilidades para, então, alcançar o menor custo de produção e melhorias operacionais.

A primeira abordagem sistemática para os problemas de sequenciamento foi realizada em meados da década de 1950. Desde então, milhares de artigos sobre diferentes problemas de programação apareceram na literatura. Inicialmente os modelos simplificavam a realidade, como os que negligenciavam o custo e tempo de *setup*, ou que não consideravam algumas restrições definitivamente impactantes para viabilidade de uso da resposta. Ao passo que a capacidade computacional foi evoluindo, novas propostas surgiram e tornaram os modelos mais adequados e capazes de resolver os problemas reais (ALLAHVERDI et al., 2008).

1.1. PROBLEMA DA PESQUISA

Na realidade de uma fábrica, dos mais diversos segmentos, as decisões de programação são as mais variadas e precisam responder muitos questionamentos como: qual máquina deve ser utilizada para fazer determinado produto? Qual troca de molde ou matriz deve ser realizada? Quanto de cada produto deve ser produzido? É viável pagar horas extraordinárias para a manutenção ou deixar a máquina parada em caso de falha? Quantas pessoas alocar em cada recurso produtivo? É viável pagar horas extraordinárias para a produção ou incorrer em ruptura de estoque? Quando deve ser realizado o setup? Além disto, é necessário lidar com os mais variados interesses dentro da organização, que nem sempre são plenamente alinhados com a redução de custos e aumento de produtividade.

A atividade de programação da produção é, em geral, realizada por uma pessoa ou por um grupo de pessoas. A primeira dificuldade emerge da falta de visibilidade e capacidade de calcular todas as possibilidades combinatórias, considerando as restrições existentes em cada parque fabril. Ainda que as combinações impossíveis sejam excluídas da população estatística, é fácil concluir que, ainda assim, a quantidade de possibilidades para avaliar pode ser muito além do

que o tempo disponível pode permitir ao ser humano estimar. A falta de visibilidade é causa de boa parte das decisões antieconômicas que são tomadas na tarefa de programação da produção (TUBINO, 2009).

Nas empresas em que uma equipe realiza a programação, geralmente cada colaborador do setor é responsável por um grupo de máquinas. Nesta situação, quando existe algum recurso crítico compartilhado, é necessário que haja um critério de utilização pré-estabelecido ou, no limite inferior, ao menos uma conversa de alinhamento entre os responsáveis pela programação. Então tem-se uma restrição que requer mais atenção e que, por vezes, torna iterativa a programação da produção, onde uma área depende da outra simultaneamente.

A influência dos diversos setores da empresa na programação da produção pode ser impactante no seu resultado. A definição de qual dos atores o programador da produção deve priorizar passa, invariavelmente, pela qualidade dos relacionamentos interpessoais existentes dentro das empresas. A própria natureza dos diversos atores envolvidos no processo acaba por adicionar emoções à tomada de decisão. Sendo assim, é previsível que uma medida antieconômica pode ser tomada para satisfazer o ator que exerce maior influência ou possui maior prestígio, algo que passa completamente inerte no caso da programação ser realizada por um computador, considerando apenas aspectos quantitativos e econômicos previamente parametrizados no modelo matemático.

A falta de insumos também é fonte de incertezas no processo decisório da programação. Quando a ausência é declarada e previsível, o prejuízo é mínimo na tomada de decisão, dada a possível e provável antecipação ao problema. Entretanto os insumos repentinamente indisponíveis podem causar distorções no planejamento, necessitando adaptações de rumo sem chance de um estudo mais detalhado para pautar a tomada de decisão. São exemplos de causas a falta de material por atraso do fornecedor ou por erro de estoque. Esta última é potencializada quando descoberta tardiamente, pois muitas vezes a produção já está em andamento, conta-se com uma quantidade de insumos baseado no sistema informatizado e, na prática, descobre-se que não existe tal quantidade.

A falta de colaboradores incorre em problema semelhante à falta de insumos, quer seja previamente conhecida ou repentina. Isto porque colaboradores podem passar por problemas que não são previsíveis, como um acidente de trânsito, e que impedem a presença no trabalho. Também é importante mencionar as

exigências da legislação brasileira, que devem ser respeitadas, como as férias remuneradas, licenças, ausências justificadas, entre outras. Quanto mais intensiva em mão de obra é a atividade industrial, mais impactante torna-se a restrição de colaboradores. Portanto, absolutamente útil e necessário é considerar a quantidade de horas-homem disponível para realizar a programação da produção que seja passível de execução.

Por fim, a capacidade de armazenagem pode representar uma restrição para a indústria. Exceto àquelas que possuem capacidade muito acima da necessidade, para as demais o tamanho do armazém tende a ser uma restrição importante em qualquer modelo. Ainda que, para alguns tipos de negócios, a expansão possa ser feita rapidamente e de forma paliativa nos períodos de alta produção com baixa demanda, é evidente que ultrapassar o limite de armazenagem é uma situação proibitiva. Sendo assim, não pode o programador exceder este limite e, a principal decisão nestes casos é parar a produção. Esta decisão, entretanto, gera outros efeitos colaterais, como a ociosidade de recursos ou a promoção de vendas, em geral com redução de preços e de margem.

Diante de todas essas dúvidas e conflitos, o problema de sequenciamento da produção torna-se não somente desafiador para os profissionais da área, mas também um popular campo de estudo na academia, sendo um dos mais frequentes tópicos dentro dos problemas de otimização de gestão e engenharia (ABEDINNIA et al., 2017). Portanto, o que se pretende responder com o presente trabalho é: como deve ser um modelo matemático que otimize o dimensionamento dos lotes e sequencie a produção com menor custo, considerando as disponibilidades de componentes, capacidade de armazenagem e restrições de mão de obra?

1.2. OBJETIVOS

O objetivo geral é propor um modelo matemático para o planejamento da produção, que minimize os custos e considere, de forma concomitante, as restrições de mão de obra, capacidade de armazenagem e disponibilidade de componentes. Este modelo deve entregar a definição do tamanho dos lotes e o sequenciamento da produção com o menor custo, respeitando as restrições existentes, de forma totalmente autônoma, ou seja, sem a intervenção humana.

Para atingir o objetivo geral, os seguintes objetivos específicos são necessários:

- i. Avaliar os modelos disponíveis na literatura e sua aderência ao problema da pesquisa;
- ii. Consolidar as restrições oriundas de pesquisas distintas em um único modelo;
- iii. Programar o modelo em linguagem computacional;
- iv. Validar o modelo proposto através de estudo de caso.

1.3. JUSTIFICATIVA

A competição acirrada exige que as empresas tenham custos reduzidos, estoques disponíveis para rápido atendimento dos pedidos e um *mix* grande de produtos para satisfazer os clientes (SLACK; BRANDON-JONES; JOHNSTON, 2018). De certa forma são necessidades antagônicas, pois o aumento de *mix* e dos estoques, sem flexibilidade na produção, gera aumento de custos.

As decisões de dimensionamento de lotes afetam diretamente os custos operacionais e a eficiência da empresa em relação ao mercado. Para responder às variações de mercado de forma mais abrangente, rápida e com mais segurança, pode-se recorrer ao uso de modelos matemáticos e computacionais para o planejamento, programação e controle da produção (TOSO, 2008).

O presente trabalho justifica-se, pelo prisma empresarial, pela dificuldade atual de se chegar na programação da produção que tenha o menor custo, respeitando as restrições, conforme mencionado anteriormente. A dificuldade é, principalmente, advinda da quantidade de decisões que o programador da produção precisa tomar. Estimativas para uma empresa com 800 produtos em linha, 45 recursos produtivos (máquinas), 60 moldes para injeção e cerca de 180 funcionários somente no âmbito da produção, dão conta de que 2×10^9 decisões diferentes são necessárias para se chegar ao menor custo, em apenas um dia de programação. A ordem de grandeza demonstra a impossibilidade da mente humana verificar todas as combinações possíveis, fortalecendo ainda mais a justificativa profissional do presente trabalho. Entende-se que um modelo de otimização do dimensionamento de lotes e programação da produção pode ser útil para fazer testes em diferentes cenários, compreendendo previsões de demanda, limitações de capacidade, número de máquinas, permitindo assim que o programador teste diferentes alternativas antes de escolher um programa de produção para os próximos dias ou as próximas semanas (PAGLIARUSSI, 2013).

Partindo para o âmbito acadêmico, o tema é amplamente abordado pela literatura recente, conforme mencionado anteriormente. Craighead e Meredith (2008) apontam que existe uma estabilização das pesquisas envolvendo modelagem e simulação nos principais periódicos internacionais. Já Nakano (2010) pontua que, ao se analisar os anais do ENEGEP (Encontro Nacional de Engenharia de Produção), existe pouca aplicação da modelagem e da simulação computacional. Desta lacuna emerge a possibilidade de atender a comunidade local, oferecendo ferramentas que possam suportar o aumento da produtividade e competitividade das empresas brasileiras.

Mesmo sendo um tema difundido, não foram encontrados, na literatura, trabalhos que façam a união das diversas restrições encontradas na vida real para a maior parte das empresas. A necessidade de abordar o problema desta forma emerge da potencial falta de utilidade que outros modelos apresentam ao não contemplar dificuldades reais da programação da produção, ou de contemplá-las de forma isolada. A junção das restrições em modelo único dá caráter inovador ao presente trabalho, avançando além das fronteiras que outros trabalhos já chegaram.

1.4. DELIMITAÇÕES

A aplicação prática será realizada em uma fábrica que produz descartáveis plásticos através do processo de injeção, sendo que este trabalho não se propõe a discutir aspectos técnicos do processo produtivo em si, limitando-se ao que foi julgado suficiente para explicar as restrições encontradas.

Como a modelagem matemática do dimensionamento de lotes e sequenciamento da programação da produção é um tópico avançado dentro da pesquisa operacional e da engenharia, este trabalho não visa discutir aspectos básicos dessas áreas do conhecimento.

Apesar de haver apresentação e conceituação de diversas heurísticas presentes na literatura, não é objetivo deste trabalho desenvolvê-las para a busca da melhor solução. Para isto já existem algoritmos difundidos na literatura e que serão aplicados em momento oportuno. As referências literárias não buscam exaurir o assunto, mas apenas situar o leitor de sua existência.

1.5. ESTRUTURA DO TRABALHO

O presente trabalho conta com seis capítulos inter-relacionados entre si. O primeiro contempla a parte introdutória, contextualizando o leitor sobre o que será apresentado, tratando do problema da pesquisa, objetivos gerais e específicos, delimitações e justificativa.

Na seção dois a metodologia é apresentada, trazendo ao leitor não só conceitos úteis que embasam a escolha da *Design Science Research*, mas também conceitos de outros métodos científicos que, de alguma forma, estão contidos na apresentação. Ainda neste capítulo são explicadas as etapas metodológicas utilizadas, delineamento da pesquisa, planejamento e ao final, conceitos que abrangem a construção do artefato.

No terceiro capítulo o referencial teórico é construído, tratando de conceitos importantes para o entendimento do trabalho como Planejamento e Controle da Produção e a sua importância nas organizações industriais, Pesquisa Operacional como área científica dentro das Engenharias, algumas das heurísticas mais utilizadas para resolução dos problemas de engenharia, *setup* e seus impactos no ambiente fabril, o problema de dimensionamento de lotes e sequenciamento da produção, bem como um breve histórico e abordagem das classes de modelos existentes para resolver este tipo de problema.

No capítulo quatro é apresentado o desenvolvimento do artefato, com o modelo matemático e a programação computacional em linguagem *Python*, contendo explicação de cada etapa de construção.

No quinto capítulo é apresentada a aplicação do artefato, iniciando pela caracterização da empresa utilizada como exemplo, histórico, produtos e capacidades, bem como os tempos para resolução do modelo tanto na busca pela solução ótima quanto na busca pela solução viável em tempo computacional razoável.

No sexto capítulo são discutidos os resultados encontrados, as conclusões, limitações encontradas e as sugestões para trabalhos futuros sobre o tema pesquisado.

Por fim, são listadas as referências bibliográficas.

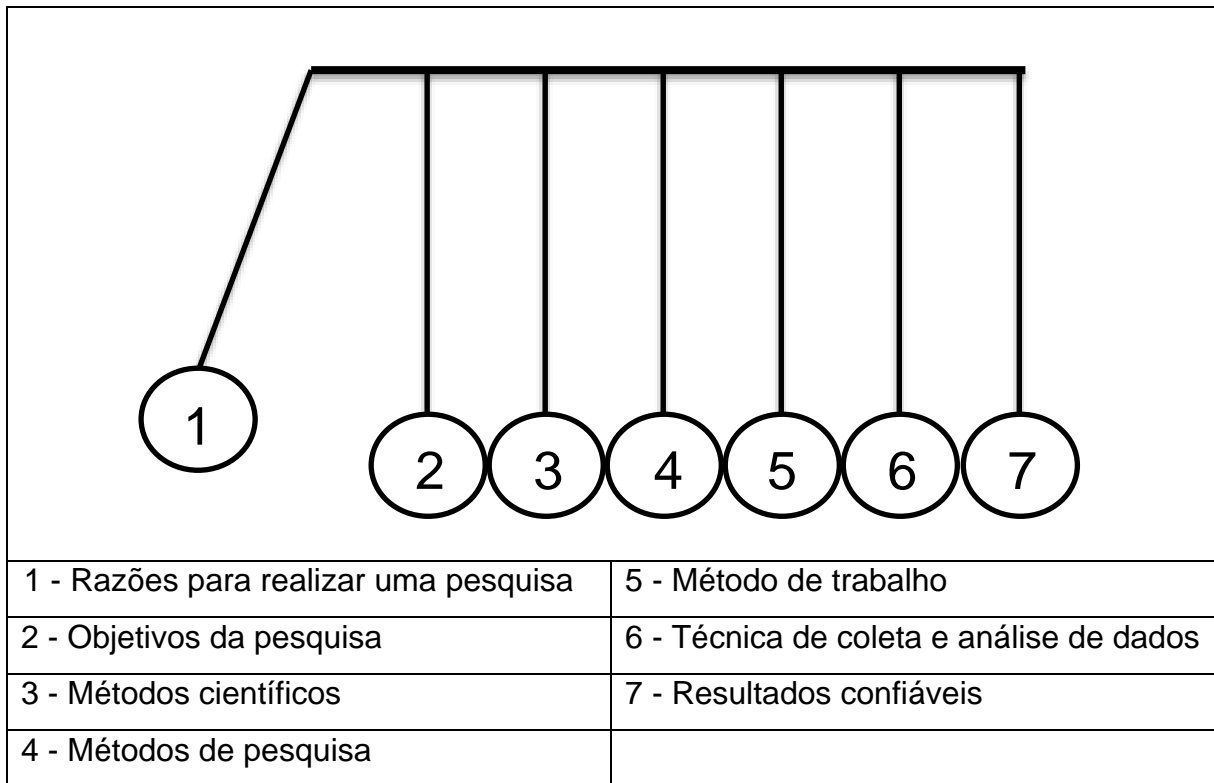
2. METODOLOGIA

De acordo com Appolinário (2004, p. 150), a pesquisa pode ser entendida como o processo pelo qual a ciência busca responder problemas que surgem e que são apresentados para a humanidade. No mesmo sentido, Gil (2002) conceitua a pesquisa como sendo o processo formal e sistemático de desenvolvimento do método científico. Sendo assim, a metodologia é considerada um mapa pelo qual a pesquisa científica será conduzida.

A importância da metodologia adequada é que esta traduz-se em fator exitoso de qualquer pesquisa, pois as estratégias definidas irão balizar e levar ao rumo desejado, além de obter a validação da comunidade científica na construção do conhecimento (SILVA; MENEZES, 2005). A validação, por sua vez, decorre do uso de método estruturado, que previne falhas ou vieses e aceito pela comunidade científica. A adequada condução do método de pesquisa é um dos pré-requisitos para a construção de um conhecimento científico confiável (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

A pesquisa científica precisa seguir alguns procedimentos para garantir a sua confiabilidade, conforme pontuado no início deste capítulo. A estrutura tradicionalmente utilizada, de acordo com Dresch et al (2015), possui sete etapas que se relacionam e dependem umas das outras como um pêndulo, conforme exemplificado na Figura 1. As etapas 4, 5 e 6 contam com seções específicas neste capítulo, dada a sua complexidade.

Figura 1 – Pêndulo da estratégia para condução da pesquisa.



Fonte: adaptado de Dresch et al (2015).

Considerando a etapa 1 da proposta de Dresch et al (2015), as razões para realizar uma pesquisa são as mais variadas. Para Booth, Colomb e Williams (1997), muitos pesquisadores de todos os níveis escrevem como se a única tarefa é responder uma questão de seu próprio interesse. Estes pesquisadores falham em entender que a pesquisa deve resolver um problema que a comunidade enfrenta e que esta, por sua vez, também deseja resolver.

Conforme já explicitado anteriormente no capítulo introdutório, uma das motivações é que existe um problema e ele é comum dentro das indústrias: a programação da produção. Complementa-se com a falta de abordagem anterior que considere as diferentes restrições de forma conjunta. Ou seja, dada a existência de um problema comum no meio produtivo e a ausência de uma solução apropriada, sendo este problema o fato causador de perdas, por si só já seriam razões suficientes para esta pesquisa. Neste sentido, Dresch et al (2015) definem que o processo de pesquisa também pode ser motivado pela observação da realidade ou pelo desejo de, com base na literatura e nos conhecimentos prévios, encontrar uma lacuna que sirva como ponto de partida para a pesquisa.

Ainda há que se considerar o potencial inovador de uma pesquisa neste âmbito, já que o artefato almejado pode ser convertido em patente, produto e em um novo negócio. Isto porque resolver um problema tão complexo, dentro de empresas também complexas, agregando valor ao processo decisório da programação da produção, pode ensejar ganhos produtivos e financeiros que podem ser compensados através do livre mercado.

Na etapa 2 da definição da pesquisa, é necessário que o pesquisador defina o objetivo que deseja atingir com a investigação, ou seja, se deseja explorar, descrever, explicar ou, ainda, predizer algum comportamento do fenômeno que está estudando (DRESCH; LACERDA; ANTUNES JUNIOR, 2015). Quanto aos objetivos, esta pesquisa é caracterizada como uma pesquisa exploratória. De acordo com Gil (2002, p. 41), pesquisas deste tipo buscam maior aprofundamento no problema, tornando-o mais explícito e construindo hipóteses. Em verdade, o problema da programação da produção não é novidade. Porém este trabalho propõe-se a explorá-lo através da aglutinação das diferentes restrições que existem na indústria e que já foram abordadas separadamente em outros modelos.

A definição dos objetivos baseou-se na compreensão do problema, suas características e na finalidade que precisaria ser alcançada para suprimir o problema. Os objetivos específicos foram definidos como sendo pontos de controle, dos quais a pesquisa precisa atingir para que fosse possível alcançar o objetivo geral. Neste aspecto, o objetivo geral da pesquisa responde, de certa forma, qual problema deseja-se resolver. Este assunto foi abordado no capítulo introdutório, ao qual o leitor pode remeter-se em caso de dúvidas.

No que tange à etapa 3, para Dresch et al (2015) o método científico é uma premissa de como o conhecimento será construído. Sua escolha é de competência exclusiva do pesquisador e sofrerá influência direta do ponto de partida da pesquisa em si, isto é, se a pesquisa teve início com a observação da realidade ou com a identificação de uma lacuna teórica, bem como dos objetivos que se deseja alcançar. Na busca pelo rigor metodológico na condução de estudos científicos, é necessário que o pesquisador defina, logo no início das suas atividades, qual será o método por ele utilizado. Também é fundamental que sejam explicados os motivos que levaram a essa escolha (DRESCH; LACERDA; CAUCHICK MIGUEL, 2015).

Para Saunders et al (2007), há duas abordagens principais de pesquisa: indutiva e dedutiva. A primeira é definida pela coleta de dados e posterior construção

de uma teoria em cima do conhecimento gerado pela coleta. Com a dedução, a teoria e as hipóteses são formuladas e a pesquisa tem por objetivo testar as hipóteses. Adicionalmente temos o método hipotético-dedutivo, sugerido por Popper (2005), que se caracteriza por identificar um problema, propor e testar hipóteses que poderão resultar em previsões e explicações, partindo de conhecimento prévio. Por fim, Peirce (1978) apresenta o método abduativo, o qual define como o processo de formar e explicar hipóteses, sendo o único operador lógico a introduzir uma nova ideia. A abdução meramente sugere que algo pode ser, utilizando o raciocínio criativo.

Os métodos científicos podem ser complementares. Sua complementariedade tem como objetivo ampliar a compreensão sobre os fenômenos e, principalmente, fundamentar racionalmente nossas conclusões. Além disso, diversos autores têm apresentado métodos científicos alternativos (LACERDA et al., 2013).

Para o presente trabalho, percebe-se uma forte intersecção entre os métodos hipotético-dedutivo e abduativo. Ao primeiro, a sinergia deve-se principalmente à identificação de um problema, que está relacionado com a qualidade do serviço executado internamente (programação da produção). Já ao segundo, a sinergia está na possibilidade de propor uma solução criativa e de caráter inovador, conforme explicado no capítulo introdutório.

A pesquisa deste trabalho é de natureza aplicada, uma vez que busca solucionar problemas específicos e de abordagem prática. Para Da Silva e Menezes (2005, p. 20), a pesquisa aplicada busca gerar conhecimentos práticos e dirigidos para resolver problemas específicos, envolvendo verdades e interesses locais. No mesmo sentido, Dresch et al. (2015) afirmam que a pesquisa de ordem prática é também chamada de pesquisa aplicada, e seu principal interesse é que os resultados auxiliem os profissionais na solução de problemas do cotidiano.

Já em relação à abordagem, Da Silva e Menezes (2005, p. 20) apontam que as pesquisas quantitativas têm a capacidade de traduzir em números os dados e informações que levarão à validação ou negação da hipótese; enquanto pesquisas qualitativas contam com fatores de subjetividade que não podem ser traduzidas em números. O fato deste trabalho propor um modelo matemático, por si só não significa que é do tipo quantitativo. Entretanto, no caso específico, as informações devem ser convertidas para unidades monetárias, pois são estas unidades que conduzirão para

a programação de menor custo. Sendo assim, a abordagem é classificada como sendo quantitativa.

Para reforçar a correta escolha do método científico, Dresch et al (2015) afirmam que o método hipotético-dedutivo pode ser encontrado nas pesquisas em gestão quando o problema a ser investigado tem relação com a medição da qualidade de produtos ou serviços, por exemplo. O pesquisador gera hipóteses e as coloca à prova para verificar se são falseáveis ou podem ser corroboradas. O método abduutivo é necessário quando o investigador está propondo possíveis soluções para resolver o problema em estudo. A pesquisa aqui apresentada enquadra-se muito bem nesta citação.

2.1. MÉTODOS DE PESQUISA

De acordo com Gil (2002, p. 43), é necessário traçar um modelo conceitual e operativo da pesquisa. Desta forma torna-se possível analisar os fatos empiricamente, confrontando a visão teórica com os dados realidade. A literatura científica da língua inglesa trata este modelo como *design*, do qual Gil sugere que seja traduzido para delineamento.

Os métodos podem ser classificados como (GIL, 2002):

- Pesquisa bibliográfica: quando elaborada a partir de material já publicado, principalmente artigos científicos;
- Pesquisa documental: semelhante à pesquisa bibliográfica, porém é baseada em material que ainda não receberam tratamento analítico.
- Pesquisa experimental: quando há um objeto de estudo e variáveis capazes de influenciá-lo, definindo formas de controle e observando os efeitos da manipulação das variáveis;
- Pesquisa *expost-facto*: feita “a partir de fato passado”, ou seja, quando o experimento se realiza após os fatos, para verificar a relação entre as variáveis;
- Estudo de corte: geralmente utilizada nas pesquisas da área da saúde, tem por objetivo isolar uma amostra e acompanhá-la por certo período de tempo;
- Levantamento: caracterizam-se pela interrogação direta das pessoas cujo comportamento se deseja conhecer;

- Estudo de caso: quando envolve o estudo profundo e exaustivo de um ou poucos objetos de maneira que se permita o seu amplo e detalhado conhecimento;
- Pesquisa-ação: quando a pesquisa está estreitamente ligada à resolução de um problema coletivo;
- Pesquisa participante: quando diversos pesquisadores e membros participam de forma integrada.

Complementarmente, Dresch et al (2015) reforçam que existem diferentes métodos e a lista não é exaurida neste trabalho, mas que cinco merecem destaque, dos quais três já foram previamente conceituados: estudo de caso, pesquisa-ação e *survey* (equivalente ao Levantamento proposto por Gil em 2002). Adicionalmente temos:

- Modelagem: modelos são representações simplificadas da realidade e permitem uma compreensão do ambiente que está sendo estudado (BURGESS; PIDD, 1999). Ainda este mesmo autor defende a classificação em abordagem *hard*, que é fundamentada em bases matemáticas e o problema está compreendido; e a abordagem *soft*, que deve ser utilizada quando há elementos comportamentais envolvidos;
- *Design Science Research*: método que fundamenta e operacionaliza a condução da pesquisa, que objetiva a criação de um artefato ou uma prescrição. Como método de pesquisa orientado à solução de problemas, a *design science research* busca, a partir do entendimento do problema, construir e avaliar artefatos que permitam transformar situações, alterando suas condições para estados melhores ou desejáveis. Ela é utilizada nas pesquisas como forma de diminuir o distanciamento entre teoria e prática (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

Este trabalho caracteriza-se por ser *design science research*, já que permite ao pesquisador explorar um determinado fenômeno, como também projetar soluções para um dado problema (DRESCH; LACERDA; CAUCHICK MIGUEL, 2015). Diferentemente de outros métodos de pesquisa, a *design science research* busca produzir conhecimento na forma de uma prescrição para apoiar a solução de um

determinado problema real, ou um projeto, para construir um novo artefato. Cabe ressaltar que, ao longo da pesquisa, haverá uma forte interação com o método de modelagem, muito embora este não figure como principal método dentro da classificação da pesquisa. A *design science research* contribui diretamente para a redução da lacuna existente entre a teoria e a prática, uma vez que esse método trata de problemas de interesse dos profissionais presentes nas organizações e também dos acadêmicos (HUGHES et al., 2011).

Para Simon (2019), existe diferença entre uma solução ótima ou ideal e uma solução satisfatória: a decisão ótima em um modelo simplificado só raramente será ótima no mundo real. A própria natureza da simplificação do mundo real, característica da modelagem, distancia da solução para o mundo real. O tomador de decisão pode escolher entre decisões ótimas em um mundo simplificado ou decisões suficientemente boas, que o satisfazem, num mundo mais próximo da realidade.

Uma característica fundamental da pesquisa que utiliza a *design science research* como método é ser orientada à solução de problemas específicos, não necessariamente buscando a solução ótima, mas a solução satisfatória para a situação. Assim, buscam-se soluções suficientemente boas para problemas em que a solução ótima seja inacessível ou de implantação inviável. Isso exige uma definição clara do que seriam resultados satisfatórios e pode ser obtido de duas formas: consenso entre as partes envolvidas no problema; ou avanço da solução atual em comparação com as soluções geradas pelos artefatos anteriores (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

A partir dessa análise comparativa, podemos constatar que a *design science research* é o método de pesquisa mais indicado quando o objetivo do estudo é projetar e desenvolver artefatos, seja em um ambiente real ou não. Conforme explicitado anteriormente, projetar um artefato definido como um sistema computacional que faça a programação da produção através de modelos matemáticos é a entrega desejada.

2.2. MÉTODO DE TRABALHO

O método de pesquisa escolhido remete ao método de trabalho proposto por Dresch et al (2015), conforme a Tabela 1, contendo as adaptações realizadas nesta pesquisa. Cada etapa será detalhada na sequência.

Tabela 1 - Comparativo das etapas da Design Science Research

Etapa	Método original	Método adaptado
1	Identificação do problema	Identificação do problema
2	Conscientização do problema	
3	Consulta às bases de conhecimento	Revisão sistêmica da literatura
4	Busca por classes de problemas e artefatos existentes	
5	Proposição do artefato	Desenvolvimento do artefato
6	Projeto do artefato	
7	Desenvolvimento do artefato	
8	Análise do artefato	Aplicação do artefato
9	Explicitação das aprendizagens	Conclusões
10	Conclusões	
11	Generalização	Publicação dos resultados
12	Comunicação dos resultados	

Fonte: adaptado de Dresch et al (2015).

2.2.1. Identificação do problema

Para Dresch et al (2015), no momento da identificação do problema o pesquisador precisa justificar a importância de estudá-lo. Identificado e justificado em termos de relevância, é necessário que o problema seja compreendido e definido clara e objetivamente, sendo a saída dessa etapa a questão de pesquisa formalizada. Já a etapa de conscientização do problema, é proposta por Dresch et al (2015) como uma forma de expandir a compreensão do contexto, das facetas e das causas do problema, bem como entender os requisitos para o correto funcionamento do artefato. É uma fase que se complementa à anterior e à posterior, uma vez que identificar o problema também passa por estar consciente dos impactos e, à medida que a revisão da literatura acontece, mais profundidade é trazida para o problema. Pode-se afirmar que é até difícil de dizer a ordem correta com que essas etapas acontecem na prática. Indo mais além, difícil também afirmar as fronteiras de cada uma delas. Por esta razão, no presente trabalho ambas foram unificadas.

Boa parte desta etapa já foi tratada no capítulo introdutório. Adicionalmente, cabe ressaltar que a identificação do problema nasceu da experiência do autor como Gerente de Produção de empresa do ramo de descartáveis plásticos, quando percebeu que a programação dispndia grande quantidade de tempo de uma pessoa altamente qualificada, que mesmo assim não conseguia entregar um resultado

consistente em termos de custo. Não foram raras as vezes em que outras pessoas do quadro de funcionários conseguiam fazer sugestões que, ao primeiro momento e sem grande aprofundamento, minimizavam o custo. Percebeu-se, então, que o programador nem sempre tomava a decisão que seria a mais econômica para a empresa.

Além disso, a substituição da pessoa que ocupava a função de programador da produção era reconhecida por todos como a troca mais problemática de toda a empresa, dada a quantidade de variáveis e conhecimento do negócio que o novo substituto precisava ter antes de ser considerado apto para realizar a tarefa.

Diante dos desafios, foi-se verificando que o artefato precisaria atender a necessidade de entregar a programação da produção, a qual entende-se por informar quanto produzir de cada produto, levando em conta a disponibilidade dos mais variados recursos, bem como as restrições apresentadas explícitas e implicitamente, objetivando o melhor resultado econômico para a empresa. Mais além, percebeu-se que é de suma importância ter disponível o aumento de custo ou redução de lucro que as restrições estão causando, o que se chama, na pesquisa operacional, de preço sombra.

2.2.2. Revisão da literatura

Revisões sistemáticas da literatura são estudos secundários utilizados para mapear, encontrar, avaliar criticamente, consolidar e agregar os resultados de estudos primários relevantes acerca de uma questão ou tópico de pesquisa específico, bem como identificar lacunas a serem preenchidas, resultando em um relatório coerente ou em uma síntese (DRESCH; LACERDA; ANTUNES JUNIOR, 2015). Todo projeto de pesquisa deve considerar uma etapa de revisão sistemática da literatura (SAUNDERS; LEWIS, 2019), como forma de suportar o estudo primário.

A revisão sistemática da literatura, no presente caso, é demonstrada no capítulo 3. Foram utilizadas buscas em bases científicas reconhecidamente confiáveis pela comunidade acadêmica, validadas através de indicações, como *Scopus*, *Google Scholar*, *Web of Science*, *Science Direct*, portal de periódicos da Universidade Federal do Rio Grande do Sul e também da Universidade Federal de Santa Catarina. O aprofundamento do conhecimento demonstrou que existe uma lacuna a ser preenchida, especificamente para o setor produtivo, considerando as diversas restrições existentes no planejamento, programação e execução da produção.

Ainda dentro da revisão da literatura, foram estudados os artefatos criados até o momento e como podem ser melhorados os resultados já publicados. É possível que o pesquisador encontre um artefato que atenda plenamente às suas necessidades para solucionar o problema. Nesse caso, sua pesquisa poderá continuar na medida em que o novo artefato traga melhores soluções em comparação aos existentes (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

A natureza da presente pesquisa exige que esta etapa seja amplamente explorada, pois a dificuldade em construir o artefato proposto pode ser substancialmente reduzida à medida em que se encontra outro artefato similar, ou que tenha sido projetado para uma finalidade parecida.

2.2.3. Desenvolvimento do artefato

O desenvolvimento do artefato passa pela proposição, planejamento e desenvolvimento propriamente dito. A união em torno de uma etapa única deu-se pela praticidade, uma vez que as fronteiras não são claramente delineadas e, de certa forma, existem intersecções na execução das atividades.

A proposição de artefatos é, por natureza, essencialmente criativa. Porquanto o método abduutivo é o mais apropriado. A criatividade e precisão na proposição é suportada pelas etapas anteriores na medida em que a identificação de classes de problemas e de artefatos desenvolvidos tratava da visualização de possíveis artefatos genéricos para resolver um problema também genérico. No entanto, mesmo tais soluções, quando consolidadas, precisam ser adaptadas à realidade em estudo (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

No planejamento do artefato são definidas as métricas pelas quais o artefato será avaliado. Os métodos e ferramentas que serão usados para construir o artefato são definidos nesta etapa também. É importante estabelecer os objetivos avaliados antes de estabelecer o processo de avaliação em si, de forma a não criar viés de avaliação que pode direcionar o que será avaliado com base em como será avaliado (ALTURKI; GABLE; BANDARA, 2011). A partir desta etapa, apenas um artefato é selecionado. Até a etapa anterior, permite-se ter mais de uma opção. Aliás, de acordo com Dresch et al (2015) é até recomendado que haja mais do que um. Componentes, relacionamentos internos e externos, limites e os procedimentos de construção devem compor o projeto do artefato (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

Desenvolvimento é o processo de construir um artefato para um propósito específico (MARCH; SMITH, 1995). Ao longo do desenvolvimento o pesquisador pode utilizar as mais variadas abordagens como algoritmos computacionais, representações gráficas, protótipos, maquetes, etc (SIMON, 2019).

A solução apresentada, no entanto, difere na essência com base no que o pesquisador entende por artefato. Em especial na área da tecnologia da informação, alguns artigos definem o artefato como um código de programação executável, enquanto outros artigos definem como sendo o conhecimento embutido no código executável (ALTURKI; GABLE; BANDARA, 2011). Neste mesmo sentido Dresch et al (2015) afirmam que, ao final do desenvolvimento, dois entregáveis são esperados: o artefato em seu estado funcional e a heurística de construção, sendo a heurística uma das contribuições da *design science* para o avanço do conhecimento.

Todo o trabalho de desenvolvimento do artefato está descrito no capítulo 4.

2.2.4. Aplicação do artefato

Avaliação é o processo de determinar quão bem um artefato funciona (MARCH; SMITH, 1995). O artefato, uma vez desenvolvido, passa a ser o objeto de avaliação do pesquisador e o rigor da avaliação é o que distingue o trabalho acadêmico do trabalho prático exercido dentro das empresas. É na avaliação que a comunidade acadêmica mais concentra a atenção (ALTURKI; GABLE; BANDARA, 2011).

A avaliação pode ser conduzida tanto em ambiente experimental quanto no contexto real. No caso desta pesquisa, em que o artefato é do tipo instanciado, necessariamente precisa ser realizado em ambiente real, em que é esperada a interação entre o pesquisador, os usuários e as pessoas da organização no qual o artefato será instanciado (DRESCH; LACERDA; ANTUNES JUNIOR, 2015). Considerando que o desempenho do artefato depende do uso desejado, o processo de avaliação é complexo. O novo sistema deve ser verificado se funciona corretamente, além de entregar o resultado desejado (ALTURKI; GABLE; BANDARA, 2011).

As saídas esperadas são: o artefato avaliado e a formalização das heurísticas contingenciais, explicitando os limites e condição de utilização. Entretanto, caso o artefato não atinja os requisitos desejados, remetendo-se ao tópico de critérios de avaliação, o pesquisador pode verificar em qual etapa pode ter ocorrido a falha e

retomar o desenvolvimento a fim de solucionar o problema (DRESCH; LACERDA; ANTUNES JUNIOR, 2015). Esta etapa da metodologia está descrita no capítulo 5.

2.2.5. Conclusões

Ao atingir esta etapa da *design science research*, significa que o artefato passou pelos testes na etapa de avaliação e pode ser publicado. Para isto, a conclusão deve abranger os resultados, contribuições, limitações e os novos conhecimentos adquiridos ao longo da pesquisa. Esta explicitação vai servir de base do conhecimento para outros pesquisadores e praticantes, que poderão dar sequência no desenvolvimento de outros artefatos a partir deste ponto (ALTURKI; GABLE; BANDARA, 2011).

Além de fomentar novos pesquisadores e praticantes, é possível que o próprio pesquisador tenha novos *insights*. Por isso, essas etapas podem guiá-lo a novos problemas que mereçam ser estudados, e, assim, a *design science research* tem um novo início (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

2.2.6. Publicação dos resultados

A generalização permite que o conhecimento adquirido na geração de um artefato específico possa, posteriormente, ser aplicado a outras situações similares e que são enfrentadas por diversas organizações (DRESCH; LACERDA; ANTUNES JUNIOR, 2015). Este aspecto reforça a sinergia e aderência que existe entre a *design science research* e a prática dentro das empresas, sendo argumento importante para a escolha como método de pesquisa.

Por fim, é essencial que haja a comunicação dos resultados, por meio da publicação em *journals*, revistas setoriais, seminários, congressos, etc., com o intuito de atingir o maior número possível de interessados na temática, tanto na academia como nas organizações. A disseminação do conhecimento gerado contribui significativamente para o avanço do conhecimento geral (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

2.3. TÉCNICAS DE COLETA E ANÁLISE DE DADOS

As técnicas de coleta e análise de dados são essenciais para garantir que os métodos de pesquisa e de trabalho, definidos pelo pesquisador, serão

operacionalizados. A correta observação e reflexão das fontes de dados possíveis de serem utilizadas, bem como de sua disponibilidade, são fatores importantes na escolha de qual técnica poderá ser empregada. Por fim, o pesquisador deve considerar em qual comunidade acadêmica está inserido o seu trabalho, utilizando técnicas que são aceitas e coerentes com o objetivo do trabalho (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

As técnicas aqui apresentadas são as que foram utilizadas na presente pesquisa e não se pretende conceituar todas as existentes. A ordem de apresentação teórica segue a cronologia dos fatos. A coleta de dados pode ser particularmente dificultada a depender do sigilo da informação, especialmente para informações que são exclusivas de determinada empresa ou governo (SAUNDERS; LEWIS; THORNHILL, 2007). Para obter êxito neste processo, duas qualidades são fundamentais: a paciência e a persistência (SILVA; MENEZES, 2005).

A **observação** é uma técnica de coleta de dados que acontece quando se utilizam os sentidos na obtenção de dados de determinados aspectos da realidade. De acordo com Da Silva e Menezes (2005), este primeiro momento do desenvolvimento da pesquisa é entendido como observação assistemática, já que não contava com planejamento e controle previamente elaborados; não-participante, uma vez que o pesquisador presencia o fato, mas não participa; e observação na vida real, uma vez que o registro de dados é realizado à medida que ocorre.

Ao ingressar no segmento de descartáveis plásticos, o autor percebeu e interessou-se pelo método de programação da produção que era utilizado naquele tempo, em que a área era subordinada ao setor de Logística. Através da observação do cotidiano foi possível perceber os diversos atores envolvidos no processo. Muito embora a ideia de realizar esta pesquisa ainda não existia, já havia um problema a ser resolvido, o conhecimento estava sendo formado e os dados estavam sendo coletados.

A base **documental** costuma ser a primeira etapa de uma pesquisa. São classificadas em primárias, se produzidas pelo pesquisador no âmbito da pesquisa, ou secundárias, se tem origem em outras fontes como bancos de dados informatizados. Também são classificadas em verbais e não verbais (SAUNDERS; LEWIS; THORNHILL, 2007).

Este trabalho também contou com a técnica documental, que envolveu a busca dos dados históricos de demanda, fichas técnicas de produção, disponibilidade

de insumos, histórico de absenteísmo, etc, como forma de validar as diferentes origens que causavam os problemas na programação da produção. Através de consulta ao *Enterprise Resource Management - ERP* da empresa, muitos dados foram disponibilizados e utilizados. Além de colaborar para o entendimento do problema, estes dados são essenciais para o funcionamento do artefato, como será demonstrado nos capítulos seguintes.

A **pesquisa bibliográfica** é desenvolvida com base em material já elaborado, constituído principalmente de livros e artigos científicos. A principal vantagem da pesquisa bibliográfica reside no fato de permitir ao investigador a cobertura de uma gama de fenômenos muito mais ampla do que aquela que poderia pesquisar diretamente. Essa vantagem torna-se particularmente importante quando o problema de pesquisa requer dados muito dispersos pelo espaço (GIL, 2002). A pesquisa bibliográfica procura levar o pesquisador a ter contato com o que foi dito ou escrito a respeito de determinado assunto, permitindo o estudo sob novo enfoque e mesmo novas descobertas sobre o assunto. Nessa técnica de coleta de dados, o pesquisador pode utilizar livros, artigos em periódicos científicos e anais de congressos, entre outros (DRESCH; LACERDA; ANTUNES JUNIOR, 2015).

Em termos práticos, quando o problema é de difícil solução a exemplo deste caso, é altamente recomendável que seja feita uma busca por artigos científicos que tratem do tema. Encontrar o ponto de partida para a solução pode ser muito facilitado, pois é altamente provável que outras empresas tenham passado pelo mesmo problema, incentivando pesquisadores ao redor do planeta. As adaptações das eventuais soluções encontradas na bibliografia naturalmente devem acontecer, e a base de conhecimento adquirida através da leitura de artigos colabora para o sucesso nesta etapa. Adicionalmente, por se tratar de uma pesquisa científica, é absolutamente natural que faça parte da metodologia.

Entrevista é a estratégia mais familiar entre as disponíveis para coletar dados qualitativos. A técnica é amplamente utilizada em pesquisas da área médica e relativamente comum naquelas da área de gestão. Podem ser classificadas em não-estruturadas, quando o entrevistador não possui um roteiro e os assuntos podem ser explorados de maneira mais ampla; semi-estruturada, quando existe um roteiro pré-definido mas não há perguntas pré-estabelecidas; estruturada, quando as perguntas são pré-definidas e não há margem para que a entrevista extrapole o roteiro pré-estabelecido (DICICCO-BLOOM; CRABTREE, 2006).

Nesta pesquisa a entrevista não-estruturada foi amplamente utilizada como forma de conhecer os fatores que cada área e cada gestor defendiam aos questionar a programação da produção. A técnica foi de suma importância para o aprofundamento do problema, entender os diferentes *stakeholders* envolvidos, seus anseios, necessidades e aflições. Por se tratar de entrevista não-estruturada, a documentação desta etapa está dispersa neste trabalho, vindo à tona pontualmente de acordo com a apresentação da situação.

Dresch et al (2015) sugerem o uso de três técnicas em pesquisas de *design science research*:

- **análise de discurso:** procura entender os mecanismos obscuros no uso da linguagem, não sendo uma técnica para descrever ou explicar algum fenômeno, mas criticar o que já existe;
- **análise de conteúdo:** busca inferir conclusões em relação ao conteúdo das mensagens proferidas por alguém, podendo explicar as causas da mensagem (motivações do interlocutor) e as consequências (efeitos) que essa mensagem terá;
- **estatística multivariada:** utilizada na geração de informações úteis a partir de dados coletados, com finalidade de orientar a tomada de decisão e gerar conhecimentos.

No presente trabalho foi utilizada a técnica da análise de conteúdo, que é descrita como um agrupamento de instrumentos metodológicos que se adapta constantemente para melhorar e que se aplica aos mais variados discursos (BARDIN, 2016). A análise de dados quantitativos pode ser codificada em diferentes níveis, de acordo com a necessidade da pesquisa, após ser coletada, possibilitando a remoção de dados indesejados (SAUNDERS; LEWIS; THORNHILL, 2007). Conforme explicado em capítulo futuro, poucos dados quantitativos precisaram ser inseridos no sistema. A empresa já contava com as informações dentro do próprio sistema de informática. No entanto, a utilização requereu tratamento prévio e análise para remover ruídos estatísticos.

Por outro lado, para os dados qualitativos a grande contribuição veio da revisão bibliográfica e das entrevistas. A análise de ambos possibilitou a criação de um *framework*, onde foi possível fazer uso de teorias existentes para formular as próprias questões de pesquisa e objetivos. As proposições teóricas contidas no

framework retroalimentam as análises qualitativas, possibilitando a indução e inferência dentro do processo dedutivo (SAUNDERS; LEWIS, 2019).

Para atingir os resultados esperados, após a identificação do problema no ambiente fabril e em termos práticos, o trabalho foi dividido nas seguintes macro etapas:

- Revisão da literatura, objetivando entender os limites de conhecimento já estabelecidos por outros autores;
- Desenvolvimento do modelo matemático;
- Coletar informações sobre as restrições como tempos e custos de *setup*, suas dependências, tempos padrões de produção, eficiência média dos recursos, necessidades de mão de obra, etc;
- Programação computacional do modelo;
- Execução do modelo e verificação de sua aderência à realidade.

3. FUNDAMENTAÇÃO TEÓRICA

Serão abordados neste capítulo os conceitos utilizados no desenvolvimento do modelo proposto, para a definição do lote, sequenciamento e programação da produção. Além de abordar conceitos da programação da produção, tais como: programação e controle da produção, definição de lotes econômicos e lotes mínimos e pesquisa operacional; pretende-se situar o leitor quanto às abordagens mais recentes sobre o tema e o posicionamento do que está sendo pesquisado em nível global.

3.1. PLANEJAMENTO E CONTROLE DA PRODUÇÃO

Os problemas de planejamento e programação da produção são muito complexos para serem resolvidos de forma monolítica. Portanto, os sistemas de planejamento da produção geralmente distinguem o planejamento de longo, médio e curto prazo. Obviamente, essas fases devem ser conectadas de alguma forma a fim de entregar planos viáveis (QUADT, 2004).

Com base nos registros de estoque, previsão de vendas e pedidos em carteira, é realizado o Plano Mestre de Produção (PMP). A partir do PMP, a programação da produção terá como resultado a emissão de ordens de produção ou ordens de compra, realizando o seu sequenciamento dentro do ambiente fabril, sempre buscando atingir um objetivo que pode ser: reduzir o *lead time*, reduzir estoques, reduzir custos de *setup*, maximizar o número de pedidos atendidos, etc. (TUBINO, 2000).

Dentro da indústria tem-se alguns itens com modelo de produção em massa, o qual consiste em itens com alta e estável demanda, produção em grande escala e alto grau de padronização. A demanda alta justifica ter recursos dedicados, de forma que não ocorre troca de molde nessas máquinas, mas sim a troca de cor do produto. Existem também os produtos com sua produção classificada como repetitivo em lotes, pois não possuem demanda que justifique a dedicação exclusiva de recursos, além de contarem com certo grau de customização, majoritariamente atrelados à embalagem.

Uma outra classificação importante relaciona os itens que são produzidos para estoque, de forma que estão disponíveis para atender o cliente imediatamente. Neste caso a etapa de produção já foi realizada no momento em que o pedido entra

no fluxo do processo. Também há alguns itens que podem ser classificados como produção sob encomenda, pois para estes itens o início da produção só acontece com a entrada do pedido no sistema.

Tubino (2009) também classifica os sistemas de produção em puxados e empurrados. Os sistemas puxados funcionam como a prateleira de um supermercado: é definido um estoque na prateleira que deve ser suficiente para atender a demanda entre os intervalos de abastecimento. Se os clientes comprarem x unidades do produto, no momento da reposição será reabastecido com x unidades a prateleira. Já na programação empurrada, são geradas ordens de produção e sequenciadas sem observar a demanda efetivamente realizada, mas sim baseado numa previsão ou expectativa, que podemos também chamar de necessidades. Para o presente caso tem-se um sistema de produção empurrada, pois existe uma alta variação de demanda e períodos de forte sazonalidade onde a venda supera largamente a capacidade produtiva. Em tentativas anteriores de implementar a produção puxada, foram nestes momentos que o sistema entrou em colapso.

3.2. PESQUISA OPERACIONAL

A pesquisa operacional (PO) engloba o desenvolvimento e a aplicação de métodos e técnicas de resolução de problemas aplicados na busca de uma tomada de decisão com eficiência, como otimização matemática, simulação, teoria de filas e outros modelos estocásticos. Os métodos e técnicas de PO envolvem a construção de modelos matemáticos que visam descrever um problema. Devido à natureza computacional e estatística da maioria das técnicas, PO também tem fortes laços com a ciência da computação. Por causa de sua ênfase na interação homem-tecnologia e de seu foco em aplicações práticas, a PO tem sobreposição com outras disciplinas, em particular engenharia de produção e gerenciamento de operações (INTERNATIONAL FEDERATION OF OPERATIONAL RESEARCH SOCIETIES, 2022).

Os problemas abordados pela PO, dentro do escopo deste trabalho, são classificados conforme o grau de dificuldade para resolução. Vamos supor um problema booleano contendo 100 variáveis, das quais desejamos encontrar a solução em que todas sejam verdadeiras. Se cada variável pode assumir dois valores (verdadeiro ou falso), então temos 2^{100} possibilidades ou 10^{30} . Já para o problema do caixeiro viajante (*TSP – Travelling Salesman Problem*), que consiste em determinar a

sequência de visita de cidades permutadas duas a duas em que a distância de ida é a mesma da volta (simétrico), considerando n cidades temos um espaço de $(n - 1)!/2$ possibilidades, ou seja, para um problema de 10 cidades tem-se 181.000 soluções. Já para um problema de 20 cidades, tem-se 10.000.000.000.000.000 de opções (MICHALEWICZ; FOGEL, 2004). O crescimento exponencial do espaço amostral é o que determina um problema ser do tipo *NP-hard*.

No problema de sequenciamento da produção em *job shop* (*job shop scheduling problem* - JSSP), para n -tarefas e m -máquinas com tempo de *setup*, o número máximo de possíveis soluções é $(n!)^m$. Sendo assim, mesmo para um problema pequeno 5 x 5, o número de soluções possíveis é de $2,4 \times 10^{10}$. Isto torna o problema *NP-hard* no sentido mais forte (SHARMA; JAIN, 2016).

Nos problemas de otimização matemática, busca-se atribuir valores para as variáveis desconhecidas inicialmente. Estas variáveis são chamadas de variáveis de decisão. Elas possuem um domínio, o qual é uma representação compacta dos valores que podem ser assumidos em cada variável. Os tipos de variáveis de decisão são referências para os objetos dos quais a exata natureza depende de cada modelo (INTERNATIONAL BUSINESS MACHINES, 2022). No exemplo deste trabalho, uma variável de decisão importante é quanto produzir de cada item. Inicialmente desconhecido, é o valor que se deseja descobrir. O domínio da variável deve ser qualquer valor não negativo, uma vez que a natureza da produção exige um número maior ou igual a zero. Nesta condição, um número negativo não é passível de aceitação, pois seria o mesmo que destruir produtos. Além disso, não é um valor válido produzir fração de uma caixa, ou seja, esta variável de decisão só pode assumir número inteiro.

Outro conceito importante na pesquisa operacional é a restrição, que é o limite de um modelo. No ambiente fabril temos as restrições difíceis (*hard constraints*) que são aquelas que, se não satisfeitas, tornam a solução inválida. No caso deste trabalho, cada operador de injeção consegue cuidar de até cinco máquinas e, se for atribuído seis máquinas, a restrição não é atendida e a solução é inválida. Já para as restrições fáceis (*soft constraints*), tem-se como exemplo as horas extras, que devem ser evitadas sob pena de incorrer em custos adicionais. Entretanto é possível que, em alguns momentos, seja vantajoso ter perdas com horas extras à perdas com vendas não atendidas por falta de produção. Neste caso a restrição deve ser evitada, muito embora a sua consideração verdadeira não invalide a solução proposta.

3.2.1. Programação linear

A programação linear ou *linear programming* (LP) é um caso especial de programação matemática, consistindo numa técnica para a otimização de uma função objetivo na forma linear, sujeita a restrições de igualdade linear e desigualdade linear. Sua região viável é um poliedro convexo, que é um conjunto definido como a interseção de um número finito de espaços, cada um dos quais contido em alguma desigualdade linear. Sua função objetivo é uma função linear de valor real definida neste poliedro. Um algoritmo de programação linear encontra um ponto no poliedro onde esta função tem o menor ou maior valor se tal ponto existir.

A programação linear pode ser aplicada a vários campos de estudo. É amplamente utilizado em matemática, negócios, economia e engenharia. As indústrias que usam modelos de programação linear incluem transporte, energia, telecomunicações e manufatura. Ele provou ser útil na modelagem de diversos tipos de problemas no planejamento, roteamento, agendamento, atribuição e design (MATOUŠEK; GÄRTNER, 2007).

3.2.2. Programação inteira

Se todas ou algumas das variáveis em um modelo LP são restritas a valores inteiros, então temos um modelo inteiro puro ou misto, do inglês *mixed integer programming* (MIP). Como em outros ramos da matemática, os modelos envolvendo números inteiros são muito mais difíceis de resolver do que os modelos envolvendo apenas números reais. Na maioria dos modelos práticos as variáveis inteiras são restritas aos dois valores 0 e 1, o que é comumente chamado de variável binária.

Existem duas razões principais para usar variáveis inteiras ao modelar problemas como um programa linear:

- As variáveis inteiras representam quantidades que só podem ser inteiras. Por exemplo, não é possível construir carros 3,7;
- As variáveis inteiras representam decisões e devem assumir apenas o valor 0 ou 1. Por exemplo, visitar determinada cidade num problema de roteirização de veículo.

Essas considerações ocorrem com frequência na prática e, portanto, a programação linear inteira pode ser usada em muitas áreas de aplicação como

transportes, planejamento da produção, redes de telecomunicações, divisão territorial, portfólio de projetos, alocação de caixa, etc. (WILLIAMS, 2009).

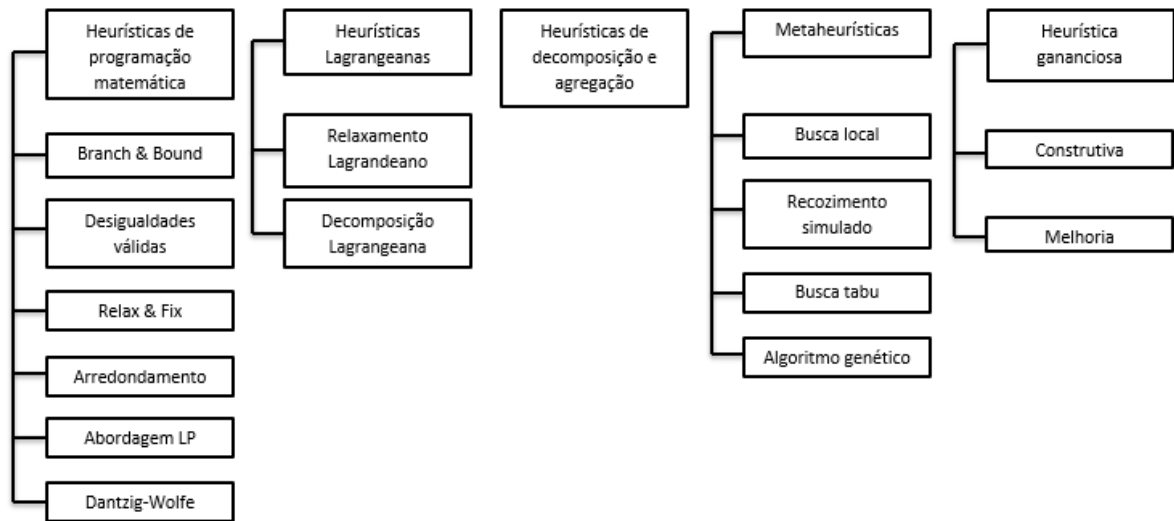
3.3. HEURÍSTICAS

Nos problemas que envolvem cálculos computacionais é necessário distinguir os procedimentos de solução ótima das heurísticas. Os procedimentos de solução ótima encontram a melhor solução para um determinado critério. A desvantagem é seu tempo computacional. Para uma classe de problemas chamados *NP-hard*, não se pode desenvolver um procedimento de solução ótima que resolva médios ou grandes problemas em uma quantidade razoável de tempo de computação. Isto porque o tempo necessário para a resolução cresce exponencialmente à medida que variáveis são adicionadas (GAREY; JOHNSON, 1979). Em contraste, as heurísticas dificilmente entregam o ideal, porém entregam soluções adequadas em um período de tempo moderado. Alguns deles são aplicáveis para problemas de tamanho médio e até mesmo grandes como aparecem na vida real (QUADT, 2004).

Existem muitos algoritmos clássicos que são projetados para procurar uma solução ideal e completa. Sempre que um algoritmo trata soluções completas, você pode interrompê-lo a qualquer momento e sempre terá pelo menos uma resposta potencial que pode tentar. Em contraste, se você interromper um algoritmo que funciona com soluções parciais, talvez não consiga usar nenhum de seus resultados.

Para algoritmos de soluções completas, quando uma nova solução é encontrada e tem uma avaliação melhor do que a melhor solução anterior, ela substitui aquela solução anterior. Os exemplos incluem pesquisa local e métodos de otimização numérica baseados em gradiente. Alguns métodos heurísticos modernos, como recozimento simulado, busca tabu e algoritmos evolutivos, também se enquadram nesta classe (MICHALEWICZ; FOGEL, 2004). A Figura 2 apresenta um esquema de como os conceitos estão divididos na seção seguinte.

Figura 2 – Esquema representativo das heurísticas abordadas



Fonte: autor (2023).

3.3.1. Heurísticas de programação matemática

O método *branch and bound* (B&B) é um procedimento de solução exata que enumera implicitamente soluções viáveis. Este método consiste em duas partes: ramificação (*branch*) e limitação (*bound*). Enquanto durante a etapa de ramificação novos subconjuntos disjuntos do espaço solução são gerados, os menos promissores são removidos durante a etapa de limitação. Para modelos MIP com variáveis binárias, a ramificação é baseada na fixação subsequente das variáveis binárias em 0 e 1.

Uma versão relaxada de cada subproblema é resolvida para determinar um limite. Diferentes opções são propostas para relaxar a formulação do modelo como, por exemplo, na etapa de LP remover a restrição de integralidade das variáveis binárias originais do MIP. Uma grande tendência de pesquisa no campo do dimensionamento de lotes é reformular o modelo matemático, redefinir as variáveis de decisão correspondentes e para introduzir desigualdades válidas (BUSCHKÜHL et al., 2008).

As desigualdades válidas (*valid inequalities*) reduzem o tamanho do espaço de solução cortando partes irrelevantes. Três métodos podem ser utilizados (BUSCHKÜHL et al., 2008):

- se as desigualdades válidas são geradas dinamicamente para cortar soluções não inteiras atuais, isso é chamado de método do plano de corte (*cutting plane*);
- desigualdades válidas podem ser introduzidas no decorrer de um algoritmo B&B, método chamado de *branch & cut* (B&C) (QIU et al., 2018);
- o procedimento de *cut & branch* (C&B) incorpora todas as desigualdades geradas na formulação do modelo antes de iniciar o algoritmo B&B.

As heurísticas *relax & fix* (R&F) reduzem o número de variáveis binárias a serem tratadas simultaneamente, dividindo o problema em vários subproblemas. Três conjuntos de variáveis binárias podem ser verificados: o primeiro é resolvido até a otimalidade, o segundo é relaxado e terceiro conjunto (aplicados à conservação dos *setups*) são fixados aos valores de uma iteração anterior (FERREIRA; MORABITO; RANGEL, 2010).

Na heurística de arredondamento ou *rounding heuristics* (RH), o relaxamento LP é resolvido e as variáveis binárias fracionárias são posteriormente arredondadas. No caso de problemas de dimensionamento de lote capacitado, essas soluções são muitas vezes inviáveis, pois a capacidade pode não ser suficiente para atender a solução. Assim, durante o RH, as variáveis binárias fracionárias são geralmente arredondadas para cima e para baixo apenas em relação a um determinado limite (BUSCHKÜHL et al., 2008).

Os métodos da abordagem *Linear Programming* são de solução iterativa e exploram o fato de que a omissão das variáveis binárias e inteiras leva a um problema significativamente mais fácil (BUSCHKÜHL et al., 2008). Uma vez que o problema foi resolvido dentro do modelo linear, assume-se que a solução está muito próxima daquela que seria obtida com as restrições binárias e inteiras. Desta forma reduz-se o espaço amostral e a resolução fica mais rápida (AGHEZZAF; ARTIBA; ELMAGHRABY, 1995).

A decomposição de Dantzig–Wolfe é um algoritmo para resolver problemas de programação linear com estrutura especial, contendo um conjunto de restrições de conexão, e se baseia na geração de colunas (variáveis) atrasadas para melhorar a tratabilidade de programas lineares de grande escala. Para a maioria dos programas

lineares resolvidos através do algoritmo simplex revisado, em cada etapa, a maioria das colunas (variáveis) não está na base.

Nesse esquema, um problema mestre contendo pelo menos as colunas atualmente ativas (a base) usa um ou vários subproblemas para gerar colunas para entrada na base, de modo que sua inclusão melhore a função objetivo. As restrições restantes precisam ser agrupadas em submatrizes independentes, de modo que, se uma variável tiver um coeficiente diferente de zero em uma submatriz, ela não terá um coeficiente diferente de zero em outra submatriz.

Após identificar a forma requerida, o problema original é reformulado em um programa mestre e n subprogramas. Cada coluna no novo programa mestre representa uma solução para um dos subproblemas. O programa mestre impõe que as restrições de conexão sejam satisfeitas, dado o conjunto de soluções de subproblemas que estão atualmente disponíveis. O programa mestre então solicita soluções adicionais do subproblema de modo que o objetivo geral do programa linear original seja melhorado (DANTZIG; WOLFE, 1960).

3.3.2. Heurísticas lagrangeanas

No relaxamento lagrangeano as restrições *hard* de um problema são relaxadas e sua violação é punida com custo de penalidade na função objetivo. A cada iteração, um limite inferior é calculado com base no relaxamento e valores dados dos multiplicadores de Lagrange. Uma solução viável é construída e serve como o novo limite superior. Por fim, os multiplicadores lagrangeanos são atualizados.

As heurísticas lagrangeanas são baseadas na convergência do limite inferior e do limite superior através da adaptação dos multiplicadores lagrangeanos. A ideia básica é que, com os multiplicadores Lagrangianos adequados à solução para o problema relaxado, será muito próxima da solução ótima do problema original. Então apenas as pequenas modificações terão que ser feitas para obter uma solução viável próxima da solução ótima (BUSCHKÜHL et al., 2008).

A decomposição lagrangeana é um caso especial de relaxamento lagrangeano, em que o problema relaxado se decompõe aditivamente em dois subproblemas, cada um tendo uma das duas matrizes do problema original como suas restrições, com um multiplicador lagrangeano por variável. A técnica utilizada consiste em criar uma (ou mais) cópias idênticas dos vetores das variáveis de decisão, utilizar uma dessas cópias em cada conjunto de restrições e em dualizar a condição para que

sejam idênticas. Esta proposta é interessante porque os subproblemas lagrangeanos mantêm todas as restrições originais (GUIGNARD; KIM, 1987).

3.3.3. Decomposição e agregação

A ideia das abordagens de decomposição e agregação é resolver subproblemas de tamanho reduzido e então coordenar as soluções individuais. As abordagens de agregação reduzem o tamanho do problema omitindo os detalhes primeiro e dividindo a solução posteriormente. As abordagens de decomposição dividem o problema original em subproblemas e coordenam os cronogramas posteriormente.

Na decomposição baseada em itens, as restrições de capacidade podem ser negligenciadas para os subproblemas. A decomposição baseada no tempo é principalmente combinada com cronogramas contínuos. A ideia é dividir o horizonte de planejamento em janelas de tempo mais curtas (geralmente sobrepostas). Uma vez encontrada uma solução para a janela atual, a solução é corrigida por alguns períodos no início da janela e o problema resolvido novamente para a próxima janela de tempo (BUSCHKÜHL et al., 2008).

3.3.4. Metaheurísticas

Metaheurísticas são estratégias gerais que orientam o processo para resolver problemas de otimização. A esperança por trás dessa abordagem é ganhar flexibilidade e capacidade de lidar com problemas grandes e complexos. Metaheurísticas são geralmente não determinísticas e podem incorporar mecanismos para evitar ficar preso em áreas confinadas do espaço de busca. Além disso, o espaço de busca também pode incluir soluções inviáveis, onde a violação de restrições é cobrada com custo de penalidade. As metaheurísticas pertencem ao grupo de procedimentos de melhoria a partir de uma dada solução inicial. Os dois princípios básicos que determinam o comportamento de uma metaheurística são intensificação e diversificação. O último potencializa a exploração do espaço de busca, enquanto o primeiro permite a exploração da experiência de busca acumulada. A maioria das metaheurísticas para problemas de dimensionamento de lotes usa uma representação de solução direta, como por exemplo variáveis binárias para *setups* e variáveis contínuas para decisões de produção. Além disso, algumas heurísticas são restritas

a encontrar uma grade binária próxima da ótima. O respectivo plano de produção pode então ser derivado por meio de programação linear, abordagens heurísticas ou reotimização dupla (GLOVER; KOCHENBERGER, 2003).

A técnica de busca local ou *local search* (LS) é baseada na ideia geral de que uma solução melhor pode estar contida na vizinhança de uma solução possível para o problema. Desta forma, partindo de uma solução viável, varia-se minimamente o vetor de solução de forma a continuar dentro do espaço amostral de soluções viáveis. Se esta nova solução apresentar resultado melhor que a anterior, repete-se o processo de iteração para tentar localizar novas soluções melhores. Caso não seja encontrada solução melhor após testar as possibilidades da vizinhança, assume-se que esta é a solução localmente ótima, o que não garante que seja a solução ótima geral (FOCACCI; LABURTHE; LODI, 2003).

O recozimento simulado ou *simulated annealing* (SA) utiliza uma técnica parecida com a LS: consiste em gerar, aleatoriamente, soluções que estejam próximas da vizinhança da solução inicial e avaliar o resultado, fazendo iterações até chegar na solução aceitável. A vantagem de gerar soluções aleatórias, dentro do espaço amostral das soluções viáveis, é que para os problemas que possuem vários pontos ótimos locais, o SA é capaz de testá-los, enquanto a LS pode ficar “preso” na vizinhança que não necessariamente apresenta a melhor solução, também conhecida como região sub-ótima. O nome do algoritmo vem da metalurgia, em que o tratamento térmico de mesmo nome consegue alterar as propriedades do material de acordo com parâmetros previamente testados (HENDERSON; JACOBSON; JOHNSON, 2003).

A metaheurística de busca tabu ou *tabu search* (TS) melhora o desempenho da LS suprimindo a principal restrição. Primariamente, em cada passo, movimentos que pioram o resultado podem ser aceitos se nenhum movimento de melhoria estiver disponível, a exemplo do que acontece quando a LS chega no ótimo local. Além disso, são introduzidas proibições para desencorajar o algoritmo a voltar para soluções previamente visitadas. A implementação da TS usa estruturas de memória que descrevem as soluções visitadas ou conjuntos de regras fornecidos pelo usuário (GENDREAU, 2006). Se uma solução potencial foi visitada anteriormente em um determinado período de curto prazo ou se violou uma regra, ela é marcada como para que o algoritmo não considere essa possibilidade repetidamente (GLOVER, 1989).

O algoritmo genético ou *genetic algorithm* (GA) faz parte do grande grupo dos algoritmos evolucionários. O nome decorre da relação com as teorias da evolução, como a seleção natural e a mutação. Em um algoritmo genético, uma população de soluções candidatas (chamadas de fenótipos) para um problema de otimização evolui para soluções melhores. Cada solução candidata tem um conjunto de propriedades (genótipos) que podem ser mutados e alterados. A evolução geralmente começa a partir de uma população de indivíduos gerados aleatoriamente e é um processo iterativo, com a população em cada iteração chamada de geração. Em cada geração, a aptidão (valor da função objetivo) de cada indivíduo da população é avaliada. Os indivíduos mais aptos são estocasticamente selecionados da população atual, e o genoma de cada indivíduo é modificado para formar uma nova geração. A nova geração de soluções candidatas é então usada na próxima iteração do algoritmo. Comumente, o algoritmo termina quando um número máximo de gerações foi produzido ou um nível de aptidão satisfatório foi alcançado para a população (WHITLEY, 1994).

3.3.5. Heurística gananciosa

Um conjunto de algoritmos heurísticos bastante intuitivos são as heurísticas gananciosas ou *greedy heuristics* (GH). Começando do zero e trabalhando período a período, ou a partir de uma determinada solução inicial, aumenta-se os tamanhos dos lotes sucessivamente para obter economia de custos. GH geralmente consiste em uma rotina de viabilidade e um índice de prioridade para selecionar o melhor candidato para tal movimento. Enquanto o primeiro garante a viabilidade do cronograma geral, ou seja, que toda a demanda seja atendida sem atrasos e violação de capacidade, o segundo serve como critério de custo para manipular as decisões de agendamento.

Dois tipos de rotinas de viabilidade podem ser encontrados: mecanismos de *feedback* e mecanismos de antecipação: os primeiros são geralmente usados em rotinas para trás, empurrando quantidades de produção inviáveis para períodos anteriores; já os mecanismos de antecipação, normalmente usados em rotinas antecipadas, calculam o estoque mínimo necessário para evitar futuras violações de capacidade e programam os lotes de produção de acordo (BUSCHKÜHL et al., 2008).

As heurísticas construtivas ou *constructive heuristics* geram uma solução do zero adicionando componentes a uma solução parcial inicialmente vazia, até que a solução esteja completa. A maioria trabalha período a período para frente ou para

trás. Eles são míopes, de modo que a cada passo o tamanho do lote de produção atual é aumentado, ignorando o custo futuro. Essa abordagem míope pode ter o efeito benéfico da estabilidade se apenas as decisões para os primeiros períodos forem implementadas em um sistema de planejamento com horizontes rolantes (KOULAMAS, 1998).

Heurísticas de melhoria ou *improvement heuristics* geram uma solução melhor viável a partir de uma solução inicial inferior geralmente inviável por simples rotinas de deslocamento. No caso do problema de definição de lote e sequenciamento, inicialmente pretende-se que um sequenciamento viável seja alcançado em um passe para frente. Em todas as etapas, é calculada a capacidade residual após o atendimento da demanda. Se for positivo, o lote aumenta ainda mais, se negativo, a produção é deslocada para períodos anteriores. Finalmente, a solução é melhorada pela fusão de lotes (WU et al., 2019).

3.4. SETUP

Setup é o trabalho realizado para preparar o recurso produtivo para uma nova produção (OSTWALD, 1992). O tempo de *setup* é um tempo necessário para preparar os recursos, como máquinas, para executar uma tarefa (ALLAHVERDI; SOROUSH, 2008). Isso inclui a obtenção de ferramentas, o posicionamento do trabalho no material do processo, a devolução de ferramentas, a limpeza, a definição dos gabaritos e acessórios necessários, o ajuste das ferramentas e a inspeção do material (ALLAHVERDI; GUPTA; ALDOWAISAN, 1999).

Houve um aumento significativo no interesse em problemas de sequenciamento envolvendo tempos de *setup* no passado recente. Uma média de 40 trabalhos por ano foram adicionados à literatura desde 2003. Esse aumento decorre do fato de que há uma enorme economia quando os custos de *setup* são incorporados explicitamente nas decisões de programação em vários ambientes do mundo real (ALLAHVERDI et al., 2008).

Os *setups* podem ser classificados em dois grupos quanto a sua dependência, de acordo com Gupta et al (2000):

- Sequência **independente**: quando depende exclusivamente do próximo produto.

- Sequência **dependente**: quando depende tanto do produto que está sendo finalizado no recurso quanto do produto que entrará no recurso.

Na indústria plástica os *setups* são dependentes com a produção de itens de cores diferentes. Um conjunto de ordens constituídas por itens de cores diferentes deve ser atribuído a um conjunto de máquinas. Quando a cor de um item previamente processado na máquina não é a mesma que foi processada anteriormente, uma quantidade de plástico é desperdiçada até que a cor padrão seja alcançada. A sequência de itens atribuídos a cada máquina afetará a quantidade total de desperdício de material (popularmente chamada de *apara*), tempo de máquina e (por consequência) custo (FRANÇA et al., 1996).

Uma outra classificação categoriza os problemas de *setup* como **lote** ou **não lote**: um problema de *setup* em lote ocorre quando os tipos de peças são agrupados em famílias de produtos e um tempo de *setup* principal é incorrido ao alternar entre tipos de peças pertencentes a diferentes famílias. Em paralelo, em algumas situações, um *setup* secundário é incorrido ao alternar entre tipos de peças dentro da mesma família de produtos (ALLAHVERDI; GUPTA; ALDOWAISAN, 1999). No mesmo sentido, Tang (1990) afirma que o tempo de *setup* principal depende apenas da família que está sendo substituída e de que o tempo de *setup* secundário depende apenas do tipo de peça que está sendo trocada.

Na prática esta situação traduz-se como uma troca de molde em máquinas injetoras sendo o *setup* principal, enquanto que uma troca de cor é o *setup* secundário. Já em máquinas extrusoras, o *setup* principal refere-se à troca de matriz enquanto que o *setup* secundário pode ser uma troca de cor ou até mesmo variação do diâmetro ou comprimento do produto, sem uma troca de matriz.

Allahverdi et al (2008) também distinguem os *setups* em antecipatórias ou não-antecipatórias. Um *setup* é antecipatório se puder ser iniciado antes que a tarefa ou lote correspondente fique disponível na máquina. Caso contrário, o *setup* é não-antecipatório. Quando não mencionado explicitamente, assume-se que é não-antecipatório.

Além da classificação anterior, os mesmos autores mencionam uma característica bem importante para o caso prático que está sendo analisado neste trabalho: o caso em que os *setups* são exercidos por uma única equipe. A preparação de um trabalho em uma máquina, cujo tempo é chamado de tempo de *setup*, é

executado por um único servidor. Desta forma, solicitações simultâneas da equipe de *setup* resultarão em tempo ocioso da máquina (ALLAHVERDI et al., 2008). Esta situação é típica do presente trabalho, pois a equipe de manutenção só consegue trocar um molde por vez. Se houver solicitações ao mesmo tempo, a máquina precisa aguardar a conclusão do *setup* anterior.

Outra consideração importante sobre os *setups* é que, em alguns casos, passar do produto A para B e posteriormente de B para C pode ser mais vantajoso, em termos de custo, do que passar do A para C diretamente. Isto acontece com as trocas de cores (*setups intra batch* ou *minor setups*) no processo de injeção, porém para o caso de uma indústria química, tal situação pode ser verdadeira em termos de custos, mas impossível no sentido prático, por exemplo, pelo uso de químicos incompatíveis entre si. Em alguns modelos matemáticos são estabelecidos lotes mínimos para superar esta restrição, de forma a forçar que a produção ocorra efetivamente, evitando que o modelo apenas sugira a realização do *setup* (STAGGEMEIER; CLARK, 2001).

Por fim, os *setups* podem ser **conservativos**, quando o recurso está preparado para iniciar novo período com a produção de um item que terminou de ser produzido no período anterior, ou **não-conservativos**, quando o início de um período produtivo obrigatoriamente exige a preparação do recurso. A situação de *setup* conservativo é bem adequada para a produção de descartáveis injetados, pois apesar de haver um pequeno esforço ao iniciar o período produtivo (segunda-feira, por exemplo), este é menor e menos custoso do que fazer uma troca de molde. Já para a produção de descartáveis extrusados, o sistema não-conservativo é mais adequado, pois religar a máquina é tão custoso quanto trocar um produto.

Dois casos podem ser mencionados: em primeiro lugar, uma conservação de *setup* só pode ser permitido se uma operação de *setup* correspondente tiver ocorrido no período imediatamente anterior; em segundo lugar, se forem permitidas conservações de *setups* consecutivas, o estado pode ser transferido por vários períodos seguidos sem *setups* adicionais. No último caso, nenhuma outra atividade de preparação é permitida, o que suspenderia o estado de atual (BUSCHKÜHL et al., 2008).

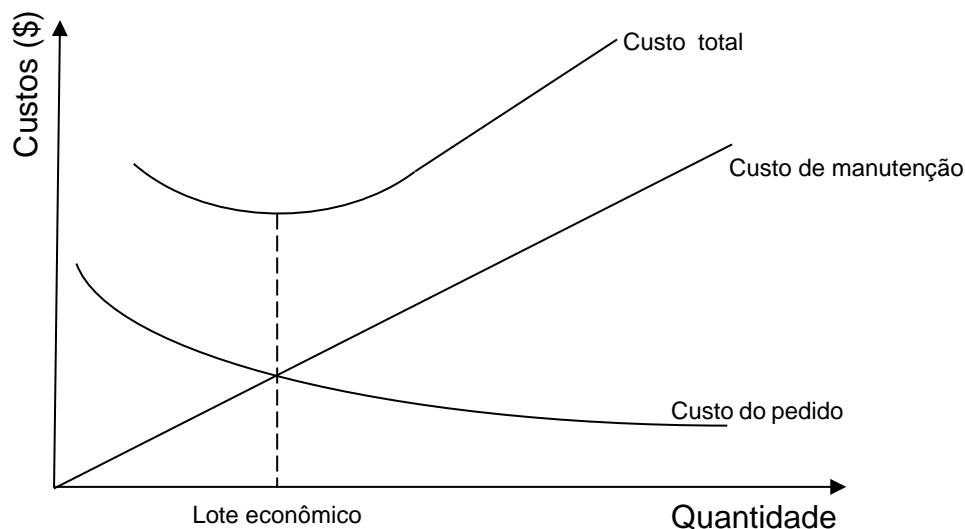
3.5. DIMENSIONAMENTO DE LOTES

As pesquisas de dimensionamento de lotes datam do início do século XIX. Um grande número de diferentes problemas foram identificados, dos quais uma quantidade ainda maior de soluções, propostas de modelos e algoritmos foram desenvolvidos (BUSCHKÜHL et al., 2008).

O trabalho pioneiro de Harris (1913) define o problema do dimensionamento de lotes em seu título: quantas peças fazer de uma vez. O autor elenca que as variáveis envolvidas na definição são: custo unitário, custos de setup, juros e depreciação do estoque, movimentação e intervalo de produção. Foi este o primeiro trabalho de dimensionamento de lotes, concluído com a proposição do modelo de lote econômico de produção ou de compra.

De forma consolidada, a determinação do tamanho dos lotes de reposição é obtida através da análise dos custos que estão envolvidos nessa tomada de decisão. Existem três componentes que determinam o custo do lote: custos diretos, custos de preparação e custos de armazenagem (TUBINO, 2009). O ponto de inflexão da curva de custos, obtido pela soma destes três componentes, é o que vai determinar o lote econômico de produção.

Figura 3 - Gráfico do lote econômico



Fonte: adaptado de Assaf Neto (2005).

O objetivo do dimensionamento de lotes é encontrar as quantidades de produção ideais que equilibrem a compensação entre preparação, manutenção de estoque e custos de pedidos pendentes. Os volumes de demanda de vários períodos

são combinados em um período. Isso é útil porque configurar uma máquina para produzir um determinado produto pode consumir tempo e dinheiro. Por outro lado, produzir um produto em um período diferente de seu período de demanda incorre em custos de retenção de estoque ou de pedidos pendentes, dependendo se ele é produzido antes ou depois de seu período de demanda (QUADT, 2004).

Em muitas situações práticas a quantidade de estoque disponível, resultado das decisões de dimensionamento de lote, são indicadores importantes, constantemente monitorados pelas áreas de logística e finanças. Conseqüentemente, busca-se focar nos modelos de definição de lotes padrão, que encontram o ponto ótimo da curva do lote econômico. Então os custos de manutenção de estoque podem servir como parâmetro para selecionar o produto que deve ter a produção antecipada, no caso de haver capacidade produtiva escassa e não seja objetivo desperdiçar horas nos recursos produtivos (COPIL et al., 2017).

A evolução da informática e aplicação nas empresas propiciou uma grande evolução na forma como os lotes são calculados. O modelo baseado no cálculo das necessidades de materiais foi expandido nos anos 60 do século passado, com o surgimento do *Material Requirements Planning* (MRP). Este modelo parte do princípio da dependência da demanda, ou seja, partindo da demanda de produtos acabados em cada período, é possível obter a necessidade de todos os demais itens da produção, de acordo com a estrutura de composição de cada produto final (TUBINO, 2009).

Infelizmente, o modelo padrão de MRP e os seus sistemas derivados dão muita ênfase para a estrutura do produto e ignoram a capacidade limitada dos recursos, quando trata-se exclusivamente de dimensionamento de lote. Isto resulta na incapacidade de gerar planos de produção exequíveis, uma vez que as necessidades podem superar as capacidades dos recursos disponíveis, tendo baixa ou nenhuma aplicabilidade no mundo real (BUSCHKÜHL et al., 2008).

A inclusão dos cálculos da necessidade de capacidade fez com o *Manufacturing Resources Planning* (MRPII) fosse criado, um sistema que calcula a necessidade de outros recursos do processo de manufatura, além da necessidade de materiais. O MRPII engloba também as decisões de como produzir, ou seja, com que recurso. Entretanto se o recurso não tiver capacidade para atender, será necessário utilizar alguma forma de superar a restrição, por exemplo, utilizando horas extras ou antecipando a produção para período ocioso imediatamente anterior. Se o problema

de capacidade for muito grande, será necessário reiniciar o processo de planejamento (CORRÊA; GIANESI; CAON, 2014).

No limite, pode ser impossível resolver a restrição de capacidade se a necessidade de recursos for muito acima da disponibilidade. Além disso, o MRPII não leva em conta as perdas de capacidade causadas pelas trocas de produtos. Sendo assim, modelos que tratam somente do dimensionamento de lotes não atendem à necessidade desta pesquisa, muito embora tenham sido o ponto de partida de diversas pesquisas.

3.6. SEQUENCIAMENTO DA PRODUÇÃO

A programação está relacionada ao problema detalhado de quando e em qual máquina produzir uma unidade de produto, que na literatura leva o nome de tarefa (*job*). Ele determina os horários de início e término da produção precisos para todas as tarefas e, portanto, a sequência de execução em cada máquina (QUADT, 2004).

As decisões de dimensionamento de lote e programação são interdependentes. Um procedimento de dimensionamento de lote calcula quantas unidades produzir em um período. Para decidir isso, ele precisa de informações sobre os tempos de preparação (*setups*). Os tempos de preparação são determinados pela atribuição da máquina e sequência de produção, que são fixas pelo procedimento de agendamento. Sendo assim, para obter soluções ótimas é necessário considerar simultaneamente a quantidade que será produzida, a atribuição de recursos para execução e a sequência de execução (STADTLER, 2005).

No mesmo sentido, Meyr e Mann (2013) afirmam que para calcular tamanhos de lotes viáveis, a capacidade deve ser respeitada. Devido à dependência dos tempos de preparação, a sequência dos lotes influencia a capacidade disponível. No entanto, decidir sobre os tamanhos dos lotes também significa decidir sobre o número de preparações, que influencia novamente as capacidades. Por causa dessa inter-relação crucial, tanto o dimensionamento quanto a programação dos lotes de produção devem ser feitos simultaneamente.

Se os tempos de preparação forem dependentes da sequência, a relação é óbvia. No entanto, a relação é verdadeira mesmo se os tempos de *setup* forem independentes da sequência, pois o último produto de um período pode muitas vezes ser continuado no próximo período sem incorrer em uma configuração adicional, mas sim o desligamento e religamento da máquina. Portanto, o procedimento de

dimensionamento de lote precisa de uma entrada do procedimento de sequenciamento. Por outro lado, o sequenciamento precisa dos volumes de produção da fase de dimensionamento do lote para encontrar uma atribuição e uma sequência.

Um objetivo típico do dimensionamento de lotes é minimizar todos os custos relacionados. Em nosso caso, são custos de manutenção de estoque, pedidos pendentes e *setups*. Os procedimentos de sequenciamento geralmente têm um objetivo orientado para o tempo. Como ambas as decisões andam juntas, os problemas de dimensionamento de lotes e sequenciamento da programação são classificados conjuntamente, conforme o tipo de solução a que se propõe.

3.7. CLASSES DE MODELOS

Existem vários modelos envolvendo diferentes recursos: a presença de uma ou várias máquinas (paralelas ou não); linhas de produção com um único estágio ou linhas em sequência com vários estágios; formulação que pode envolver custos de preparação, que por sua vez podem ser fixos, variar por produto ou depender do sequenciamento. Outro recurso é a demanda, que pode ser constante, variar em períodos regulares ou variar em períodos irregulares. Ainda pode ser determinística ou probabilística (STAGGEMEIER; CLARK, 2001). A seguir serão discutidas as variáveis mais relevantes no desenvolvimento do problema de dimensionamento de lotes e sequenciamento da produção, bem como suas classificações.

3.7.1. Dados de entrada

Com o objetivo de retratar a realidade da forma mais fiel possível, os modelos propostos requerem diversos dados como entrada para que se alcance as variáveis da função. Entre eles podemos citar demandas dos produtos, tempos de *setup*, taxas de produção, etc. Ocorre que estas entradas, na vida real, são majoritariamente **estocásticas**. Entretanto os modelos, para fins de simplificação, tratam como **determinísticas**.

A simplificação tem por objetivo reduzir a necessidade computacional, porém pode gerar algumas distorções nos resultados. A variabilidade nos tempos de *setup* existe devido a vários motivos. É possível que o processo de *setup* não seja padronizado ou que ocorram erros durante este processo, o que aumenta o tempo total de execução. Atribuir uma taxa de produção constante para determinado produto

em recurso específico não traduz a realidade. Diante disso, a vida real nos mostra que há sempre a variabilidade inerente ao tempo de execução de uma atividade específica (TAŞ et al., 2019).

3.7.2. Períodos

Os modelos propostos diferenciam-se naqueles que utilizam **macroperíodos**, em que múltiplos *setups* são permitidos dentro do intervalo; e os que adotam **microperíodos**, em que apenas um *setup* é permitido dentro de cada intervalo (COPIL et al., 2017).

O horizonte de planejamento é decomposto em intervalos relativamente pequenos ou microperíodos dentro dos quais um item pode ser produzido no máximo, enquanto vários itens podem ser produzidos durante cada período ou macroperíodo (FERREIRA et al., 2012). A conexão entre os conceitos, que mencionam número de *setups* e número de itens produzidos por período fica evidenciada uma vez que a produção de itens diferentes requer a preparação do recurso produtivo.

3.7.3. Capacidade dos recursos

Um modelo realista deve considerar uma capacidade de produção **finita**, vários produtos compartilhando essa capacidade e uma situação de planejamento dinâmica, que parte de um estoque dado, status inicial de produção e envolve um intervalo de planejamento finito com demandas flutuantes (FLEISCHMANN; MEYR, 1997).

Em algumas situações que ocorrem em indústrias específicas, a capacidade do recurso é tão grande que pode ser assumida como **infinita**, aproximação que reduz uma restrição importante nos modelos e facilita a resolução, em termos de tempo computacional.

3.7.4. Ambiente produtivo

Gupta *et al* (1999) classificam os problemas quanto ao seu ambiente produtivo:

- **Máquina única:** as tarefas utilizam um único recurso disponível e são executadas uma de cada vez. Cada tarefa possui propriedades

diferentes como tempo de processamento, data de entrega, tempo de *setup* e importância;

- **Máquinas paralelas:** as tarefas podem ser processadas em qualquer uma das várias máquinas disponíveis. Cada tarefa, com características diferentes, tem uma única operação que pode ser executada em qualquer máquina. Allahverdi et al (2008) também define que as máquinas paralelas podem ser idênticas, ou ter apenas velocidades diferentes, ou serem completamente não relacionadas;
- **Flow shop:** em uma fábrica de *m-máquinas*, as tarefas precisam ser processadas seguindo uma ordem pré-definida. Cada tarefa tem *m* operações e a operação *i* de cada tarefa deve ser executada na máquina *i*. Os tempos para cada operação em diferentes máquinas podem divergir entre si;
- **Job shop:** consiste num ambiente de várias máquinas diferentes entre si e uma tarefa pode necessitar de uma ou de todas as máquinas para ser concluída, de forma que só poderá passar pela máquina uma única vez, seguindo uma ordem pré-estabelecida;
- **Open shop:** diverge do *job shop* pelo fato de que as tarefas podem ser executadas em qualquer ordem.

Aqui vale ressaltar a diferença de conceito de um *flow shop* em relação ao que mencionam os autores Quadt e Kuhn (2005), afirmando que cada máquina paralela tem a mesma taxa de produção para determinado produto. Esta restrição é difícil de encontrar na prática, pois como o maquinário evolui e torna-se mais produtivo com o passar do tempo, aliado ao fato de que as empresas, geralmente, possuem uma expansão quase que linear, é absolutamente normal que uma nova máquina seja adquirida para fazer um mesmo produto. Então esta nova máquina geralmente possui uma taxa de produção maior do que uma máquina antiga, que já esteja em uso no parque fabril.

3.7.5. Estágios de produção

Os problemas de dimensionamento de lote e sequenciamento da produção são divididos em dois grandes grupos: *multi-stage* e *single-stage*. *Single-stage* são

problemas em que os produtos só tem uma etapa produtiva ou que, mesmo que tenham mais de uma etapa, podem ser simplificados em uma única, como por exemplo uma linha com múltiplos recursos que tem taxas de produção similares. *Multi-stage* são os produtos que necessitam passar por mais de uma etapa, dentro da mesma fábrica, até ficarem prontos. Tal diferença é substancial na formulação dos problemas.

Considere o exemplo de uma empresa com duas linhas de produção que faz 20 unidades do produto P1 na máquina M1 e 15 unidades do produto P2 na máquina M2. Neste caso estamos falando de duas máquinas que funcionam paralelamente, com um único estágio produtivo. Agora adicione ao exemplo uma terceira máquina M3, que atua após as máquinas M1 e M2, com capacidade de 30 unidades por período. A simples adição de M3 tornou o problema como sendo de múltiplos estágios, adicionando um recurso que passou a ser o gargalo, dada sua produção inferior à soma do primeiro estágio. Se esta restrição não for modelada corretamente, levará a decisões incorretas (WÖRBELAUER; MEYR; ALMADA-LOBO, 2019).

3.7.6. Tipos de recursos

Os modelos simultâneos de dimensionamento de lotes e sequenciamento da produção consideram a capacidade limitada do sistema, respeitando um tempo máximo que as respectivas máquinas ou linhas de produção podem estar disponíveis. No entanto, a maioria desses modelos considera a capacidade de produção dessas máquinas ou linhas de produção como o único fator limitante. Em outras palavras, atender a uma determinada demanda de produtos com um custo mínimo é limitado apenas pelas capacidades de produção disponíveis das máquinas ou linhas de produção. Entretanto outras limitações podem surgir pela escassa disponibilidade de, por exemplo, ferramentas de *setup*, operadores ou matérias-primas que, portanto, não podem ser negligenciadas em modelos de otimização (WÖRBELAUER; MEYR; ALMADA-LOBO, 2019).

Na literatura, as máquinas ou linhas de produção são chamados de recursos ou fatores primários, enquanto os recursos adicionais escassos, citados anteriormente, são chamados de recursos secundários. Apenas alguns dos modelos também levam em consideração recursos secundários que limitam a quantidade de produção dos produtos (COPIL et al., 2017).

Os recursos secundários, por sua vez, são divididos em:

- Cumulativos, quando a soma de algum recurso não atende a totalidade da produção, por exemplo, no plano para produzir 30 itens A e 20 itens B, mas a matéria prima só é suficiente para fazer 40 unidades;
- Dedicados, quando o item a ser produzido necessita de algum recurso adicional que limita a produção, por exemplo, quando há diversas injetoras em paralelo capaz de fazer o item, mas há somente um molde daquele item. Não importa quantas injetoras haja e suas capacidades somadas, o limitante vai ser o molde.

Uma outra diferença importante entre ambos é que os cumulativos não se renovam automaticamente a cada período, tornando obrigatória a sua reposição no horizonte de planejamento, enquanto os dedicados renovam-se automaticamente, ou seja, o molde do exemplo anterior poderá ser usado em cada período sem a necessidade de reposição. Em suma, é importante reconhecer que, se os recursos secundários fossem completamente negligenciados no processo de planejamento, os planos resultantes poderiam se tornar inviáveis para aplicações industriais do mundo real. No entanto, não é necessário modelar recursos secundários em todos os detalhes (WÖRBELAUER; MEYR; ALMADA-LOBO, 2019).

3.7.7. Evolução dos modelos na literatura

Os aspectos históricos abordados nesta seção são a consolidação de diversas revisões de literatura apresentadas em artigos que descrevem a evolução dos modelos no tempo. Cabe ressaltar que foram suprimidos os modelos que não envolvem a decisão conjunta do dimensionamento de lotes e sequenciamento da produção, por não estarem no escopo desta pesquisa. Além disso, eventuais modelos que estariam na definição anterior, mas que não foram considerados relevantes para conceituar o leitor, também foram suprimidos.

Um dos trabalhos iniciais que descreve o *trade-off* entre os custos de *setup* e de manutenção de estoque para um cenário de demanda dinâmica é fornecido por Wagner e Within (1958). A introdução de limitações de capacidade e a consideração de vários produtos simultaneamente levaram ao *capacitated lot-sizing problem* (CLSP), que Bitran e Yanasse (1982) mostraram ser *NP-hard*.

O problema *dynamic multi-level capacitated lot-sizing problem* (MLCLSP) foi introduzido por Billington et al (1983). Ele descreve o seguinte cenário: o horizonte de planejamento é finito e dividido em T períodos de tempo discretos. Existem K itens com demandas externas específicas do período que devem ser atendidas obrigatoriamente dentro do período. Os itens são produzidos em M recursos não idênticos com capacidades limitadas específicas do período. Os recursos podem ser agrupados grupo de recursos, como máquinas ou trabalhadores semelhantes, que são tratados como uma única entidade. A capacidade de um recurso por período é definida pelo produto do número de unidades e o tempo disponível. Para cada item, existe uma atribuição exclusiva a um único recurso. A produção de cada unidade de um item leva uma quantidade constante de tempo de processamento. Sempre que a quantidade de produção em um período for maior que zero, o modelo assume que um determinado tempo de preparação é incorrido. Além disso, os tempos de setup e o custo são considerados independentes da sequência de produtos durante um período. Uma característica específica do padrão MLCLSP é que cada quantidade de produção positiva durante um período induz a um setup, mesmo nos casos em que a produção de um item ocorre em dois períodos consecutivos (BUSCHKÜHL et al., 2008).

Buschkühl et al (2008) mencionam outros modelos de definição de lotes e sequenciamento com objetivo de reduzir os custos de *setup* e de estoques, considerando datas de entregas dos pedidos e capacidades finitas:

- *Continuous setup lotsizing problem* (CSLP) proposto por Karmarkar e Schrage (1985), que não considerava o sequenciamento, mas somente a definição do lote (CLARK; CLARK, 2000).
- *Discrete lot sizing and scheduling problem* (DLSP) proposto por Fleischmann (1990). A grande restrição do DLSP era que a capacidade de um determinado período deve ser completamente exaurida com *setups* e/ou produção de um único produto ou não é produzida dentro deste período, algo que o autor chama de “tudo ou nada” (MEYR; MANN, 2013);
- *Proportional lotsizing and scheduling problem* (PLSP) proposto por Drexel e Haase (1995), que permitia somente um *setup* por período, em qualquer momento dentro do período (CLARK; CLARK, 2000).
- *Capacitated lotsizing problem with sequence-dependent setups* (CLSD) proposto por Haase (1996), em que a sequência de lotes é

descrita pela enumeração dos produtos produzidos no período, como se fosse uma volta dentro do problema do caixeiro-viajante (FERREIRA et al., 2012);

- *General lotsizing and scheduling problem* (GLSP) proposto por Fleischmann e Meyr (1997), o qual tratava do problema em situações com um único recurso produtivo. O termo *General* é baseado no fato de que vários modelos bem conhecidos para o problema diferem do GLSP apenas por restrições adicionais que restringem a estrutura de tempo das soluções.

De acordo com Meyr e Mann (2013), os problemas de máquinas paralelas heterogêneas e *setup* dependente da sequência foram introduzidos em De Matta e Guignard (1994). O modelo PSP também apresentava a restrição “tudo ou nada” (FLEISCHMANN; MEYR, 1997), mas como até aquele momento os desenvolvimentos não consideraram a perda com *setups*, esta foi a principal inovação trazida pelo trabalho que expandiu o DLSP.

Posteriormente o problema PLSP foi expandido por Clark e Clark (2000), quando os autores permitiram *backlogging* entre os períodos, ou seja, uma produção não realizada no período anterior poderia ser realizada no período atual. Na prática, isso se traduz como a possibilidade de entregar ao cliente com atraso. Entretanto este modelo não trazia os custos de *setups* por serem considerados negligenciáveis diante dos custos de penalidade por *backlogging* e de manutenção de estoque, com a ressalva de os autores explicam como incluir os custos de *setup* e de produção na função objetivo. A restrição de uso deste modelo era a necessidade de se fazer o primeiro *setup* no início do período, muito embora fosse possível realizar outros dentro do mesmo intervalo. Ainda de acordo com os autores, os resultados computacionais sugerem que para problemas de médio a grande porte, as quantidades de tempo que serão gastas para identificar uma solução viável são impraticáveis. Note que apenas mencionam identificar solução viável e não encontrar uma solução ótima ou próxima da ideal.

As generalizações vieram com Meyr (2002), através do modelo *General Lotsizing and Scheduling Problem for Parallel Production Lines* (GLSPPL). Até então, as pesquisas majoritariamente concentravam-se em sistemas produtivos com uma única linha, sendo que alguns poucos trabalhos abordavam os sistemas com máquinas paralelas. O GLSPPL considera custos de inventário, preparação de

máquinas e o custo de produção de cada unidade, considerando que pode haver diferenciação se mesmo produto for produzido em diferentes máquinas. Caso as linhas de produção sejam homogêneas, este custo pode ser ignorado uma vez que não há impacto prático.

Suerie e Stadtler (2003) apresentaram o modelo *multilevel capacitated lot-sizing problem with linked lot sizes* (MLCLSP), o qual apresentou uma nova formulação para resolver o problema dos estados conservativos dos *setups* com a característica de conter produtos com mais de uma etapa produtiva. Os pesquisadores tentam considerar classes mais realistas de MLCLSP, incluindo tempos de *setup*, estado conservativo dos *setups* e *backlogging*. No entanto, ainda hoje a maioria dos modelos e algoritmos propostos para MLCLSP baseiam-se em uma das duas seguintes premissas (ALMEDER et al., 2015):

- os *lead times* são negligenciados, permitindo que predecessores e sucessores sejam produzidos no mesmo período;
- os *lead times* representam pelo menos um período para cada componente, forçando o tempo de processamento (em número de períodos) dos produtos acabados a ser pelo menos igual ao número de níveis do produto.

Diante do problema apresentado, Almeder et al (2015) propuseram duas extensões de MLCLSP representando os dois casos extremos:

- *batching*, quando um item sucessor só pode ser produzido se os lotes completos dos itens predecessores estiverem concluídos, dentro do mesmo período;
- *lote-streaming*, quando o item sucessor pode ser produzido simultaneamente com seu predecessor, desde que use um recurso diferente.

Seeanner (2013) estende sua formulação GLSP para o problema multiestágio (GLSPMS) e para considerar também as matérias-primas. As constantes definem o consumo de cada matéria-prima durante a produção de uma unidade de um produto específico, como uma ficha técnica de materiais. A produção global não pode exceder uma determinada disponibilidade de cada matéria-prima. Este modelo também considera vários operadores de *setup* diferentes em que, para cada *setup*, pode-se definir quais operadores são capazes de realizar esta tarefa, ou seja, são possíveis substitutos. As variáveis binárias registram qual operador realmente executa

um *setup*, tornando simples garantir que um operador seja atribuído, no máximo, a uma única linha por microperíodo. Essa abordagem é válida porque o GLSPMS possui uma estrutura de microperíodo idêntica em todas as linhas. Ainda é possível iniciar uma configuração no microperíodo anterior e terminá-la no microperíodo atual, contando com restrições adicionais que garantem que um microperíodo seja longo o suficiente para terminar o *setup* que foi iniciado no microperíodo anterior (SEEANNER, 2013; SEEANNER; MEYR, 2013).

Wörbelauer, Meyr e Almada-Lobo (2019) basearam-se no GLSPPL e no GLSPMS para criar um modelo capacitado, *single-stage*, com máquinas heterogêneas e paralelas, que é capaz de considerar as restrições de recursos secundários como matéria-prima ou operadores de *setup*, mas sem levar em conta a qualificação de cada operador, de forma a diferenciar se algum operador está apto a realizar determinado *setup* ou não.

Já no ambiente da moldagem por injeção plástica, a primeira tentativa de otimizar o CSLP com máquinas paralelas para vários itens foi realizada por Dastidar e Nagi (2005), que propõem uma formulação para minimizar os custos de estoque, *backorder* e *setup*, e uma abordagem de solução baseada em um esquema de decomposição de duas fases. Masoud e Mason (2016) estenderam esse modelo otimizando simultaneamente as decisões de transporte e também desenvolveram um algoritmo híbrido de SA para resolver o problema proposto voltado para a indústria automotiva. Ibarra-Rojas et al (2011) formulam um modelo de programação inteiro quadraticamente restrito para resolver o problema de uma perspectiva de fabricação peça-molde-máquina, na qual moldes limitados são configurados em diferentes máquinas. A formulação também é decomposta em dois subproblemas: um para dimensionamento de lotes e outro para verificação de sequenciamento viável. Rios-Solis et al (2020) relatam uma abordagem heurística para resolver um problema de dimensionamento de lotes e sequenciamento em um ambiente de produção de injeção plástica onde as peças são produzidas usando moldes para formar produtos acabados, ou seja, múltiplos estágios. Por fim, Mula et al (2021) propuseram um modelo que trata de um produto injetado com cavidades que produzem, ao mesmo tempo, o lado esquerdo e o lado direito da porta de um veículo, trazendo uma complexidade adicional aos modelos: quando um item é produzido, obrigatoriamente é gerado estoque do outro. Neste mesmo segmento de produção, Huang et al (2011)

apresentaram um modelo para a indústria de calçados, que injeta o par ao produzir a sola.

O JSSP flexível é uma variação do modelo clássico e consiste em dois sub-problemas: atribuir uma tarefa à determinada máquina e sequenciar as operações em cada máquina. Os objetivos comuns dos problemas de JSSP são minimizar alguma unidade de desempenho como fila de produção, número de pedidos atrasados, tempo médio de atraso, tempo médio de *setup* ou números de *setups*, etc. (SHARMA; JAIN, 2016).

3.8. PUBLICAÇÕES RECENTES

Esta seção busca posicionar o leitor sobre os trabalhos publicados nos últimos três anos, como forma de mostrar quais caminhos as pesquisas relacionadas estão seguindo. Não se pretende pontuar as inovações de cada publicação, mas sim o escopo de cada pesquisa.

Irwan et al. (2022) apresentam uma revisão recente da literatura relacionada ao problema de definição de lote e sequenciamento da produção. Através de pesquisa nas bases *Science Direct*, *Springer* e *Google Scholar*, os autores demonstram que foram publicados somente três trabalhos relacionados às palavras chaves *integration lot sizing and scheduling*, *scheduling with uncertainty demand* e *parallel line capacitated* a partir do ano de 2020. Neste mesmo sentido, Soler et al. (2021) mencionam que as revisões mais atuais são os trabalhos de Copil et al. (2017) e Guimarães et al. (2014), o que corrobora no sentido de que as pesquisas recentes têm se dedicado à aplicação propriamente dita do que a apresentação de novos modelos ou técnicas de resolução.

Mohammadi et al. (2020) propõem um modelo que possibilita a definição dos lotes de compra, a partir de múltiplos fornecedores, quando descontos são aplicados a partir de determinada quantidade, integrado ao problema de definição de lotes e sequenciamento da produção. O modelo considera múltiplos recursos produtivos, *multi-level* e *setups* complexos. Para resolver, são aplicadas técnicas de agregação de dados e posterior desagregação, de forma a minimizar o tempo computacional necessário, mas sem perder a informação em nível primário.

Alves et al. (2021) apresentam um trabalho comparativo que busca mensurar as diferenças entre resolver o problema de dimensionamento de lotes e sequenciamento da produção de forma iterativa, afirmando que a resolução dos

problemas de forma independente pode levar a planos impossíveis de serem executados ou que não alcançam o menor custo. Por outro lado, os modelos que propõe resoluções simultâneas são impraticáveis em termos de tempo computacional quando utilizados em situações reais que criam problemas relativamente grandes.

Soler et al. (2021) apresentam um modelo que foi aplicado na indústria alimentícia e que traz, na sua formulação, o uso de recursos escassos. Esta situação faz com que apenas um subconjunto de máquinas possa ser utilizado ao mesmo tempo, dado que algum recurso é compartilhado entre elas e não permite a utilização de todas as máquinas simultaneamente. Para resolver o problema os autores propõe a utilização das heurísticas *branch-and-bound* e *relax-and-fix*, comprovando ser mais eficiente do que o *solver* CPLEX.

Alimian et al. (2022) propuseram um modelo que incluiu a manutenção preventiva e a data de entrega de cada lote, considerando a idade do equipamento desde a última manutenção preventiva e atribuindo determinado risco de falha através de distribuição exponencial. Inicialmente o modelo ganhou equações não lineares e que posteriormente foram linearizadas. O modelo foi resolvido usando duas heurísticas de horizonte rolante, as quais foram apresentadas comparativamente ao *solver* CPLEX, obtendo resultados expressivamente melhores.

A lista de modelos disponíveis na literatura é demasiadamente extensa. Em alguns casos, apesar de reivindicar inovação, é possível encontrar modelos que tratam da mesma situação e que foram publicados no espaço curto de tempo. Tal situação sugere que há uma corrida para publicar modelos cada vez mais adequados às situações reais das empresas e que, dado o alto número de trabalhos existentes, torna-se absolutamente difícil conhecer todos aqueles que existem.

4. CONSTRUÇÃO DO ARTEFATO

Diante das restrições e dificuldades para programar a produção, já explicitadas nos capítulos introdutórios, aliado ao aprofundamento do tema através da revisão bibliográfica, a proposição de um artefato capaz de reduzir (ou até mesmo suprimir) os problemas desta rotina específica da indústria restou-se como solução adequada.

O modelo proposto precisa contemplar as restrições mais comuns na indústria:

- Capacidade de armazenagem de produtos acabados limitada pela estrutura física disponível, que geralmente é medida em posições palete;
- Capacidade limitada pela mão de obra disponível, geralmente medida em horas-homem, considerando também a possibilidade de fazer horas extras, dentro dos limites da legislação brasileira;
- Capacidade limitada pelas horas de máquina disponíveis, geralmente sendo este o principal limitante dentro da indústria;
- Impossibilidade de produzir quando não há insumos (componentes) suficientes para realizar o planejado;
- Respeito às regras de lote mínimo, muito embora essa regra possa ser flexibilizada no caso de ser causa comprovada de aumento de custos;

A medição de desempenho precisa atender dois critérios quantitativos: tempo e custo.

Por tempo, entende-se o lapso temporal que o computador levará até chegar à solução viável apropriada. Note que o critério não necessariamente passa por encontrar a solução ótima, uma vez que a própria literatura sobre o tema, por diversos autores distintos, menciona que a busca da solução ótima pode exigir demasiado tempo computacional.

Já no aspecto do custo, entende-se por alcançar o resultado mais próximo da solução ótima, qual seja, uma programação da produção que resulte no menor custo possível levando-se em conta custos de manutenção de inventário, lucros cessantes por não atender os clientes no prazo desejado e acabar perdendo o pedido para a concorrência, custos com *setups*, custos adicionais por não utilizar a máquina

mais econômica para produzir determinado item e custos adicionais por utilizar horas extras.

Por solução viável apropriada entende-se aquela que apresenta um tempo razoável de solução e um custo relativamente próximo do ótimo. São critérios que variam de indústria para indústria. Por exemplo: uma indústria siderúrgica, que tem por natureza o horizonte de planejamento longo, carteira de pedidos estável, antecipada e previsível, pode dispender um maior tempo computacional para chegar no resultado mais próximo do ótimo. Em termos práticos e no exemplo hipotético, é aceitável que o computador leve 24 horas para planejar um mês de produção e, nesse tempo computacional, muito provavelmente chegará à solução ótima. Por outro lado, uma indústria de descartáveis plásticos, que tem uma carteira de pedidos altamente volátil e precisa programar a produção a cada dia, não pode esperar 24 horas de tempo computacional para chegar à solução ótima. Desta forma, pode-se limitar o tempo de execução.

Dentre os artefatos existem na literatura, alguns se destacam por apresentarem a formulação matemática contida no escopo desta pesquisa:

Meyr (2002) aborda o dimensionamento de lote e o sequenciamento da produção de forma simultânea, de diversos produtos em linhas de produção paralelas não idênticas, ou seja, com máquinas heterogêneas. A capacidade limitada das linhas de produção pode ser ainda mais reduzida por tempos de *setup* dependentes da sequência. A demanda dinâmica e determinística de produtos padrão deve ser atendida sem atrasos, com o objetivo de minimizar os custos de *setup*, manutenção de estoques e produção dependentes da sequência.

Józefowska e Zimniak (2008) apresentaram a modelagem matemática para considerar a margem de contribuição de cada produto em cada máquina. Esta contribuição é significativa para este trabalho uma vez que a penalidade de custo por não produzir o item na máquina mais eficaz conta com a mesma lógica, apenas invertendo o sinal matemático da constante. Além disso, apresentaram também a restrição de lote mínimo para cada produto, situação que já foi comentada anteriormente neste capítulo.

Ulbricht e Volpi (2015) basearam-se no modelo de Meyr (2012) para introduzir a possibilidade de uso de horas extras quando o modelo extrapola a capacidade de tempo disponível em cada máquina. Neste trabalho os autores

concebem a utilização prática deste incremento de disponibilidade ao considerar que as horas extras podem ser aplicadas em apenas uma máquina, período a período.

Wörbelauer, Meyr e Almada-Lobo (2019) consideram outras limitações de capacidade produtiva que surgem da escassa disponibilidade de, por exemplo, ferramentas ou operadores de *setup* ou de matérias-primas que, portanto, não podem ser negligenciadas nos modelos de otimização. Esses fatores de produção são chamados de recursos secundários. A proposta traz, pela primeira vez, uma visão unificada dos fatores de produção escassos. Os *insights* sobre diferentes tipos de recursos secundários ajudam a desenvolver uma nova formulação de modelo, generalizando e estendendo as abordagens atualmente usadas que são específicas para alguns contextos. Desta forma, o modelo de Wörbelauer, Meyr e Almada-Lobo colaborou para a construção das restrições de componentes ou matéria-prima.

Tabela 2 - Contribuições consideradas no modelo final.

Autores:	Contribuições:
Meyr	Máquinas paralelas heterogêneas e <i>setup</i> dependente
Józefowska e Zimniak	Margem de contribuição e lote mínimo
Ulbricht e Volpi (baseado em Meyr)	Utilização de horas-extras
Wörbelauer, Meyr e Almada-Lobo	Restrição de componentes

Fonte: autor (2023).

Sendo assim, o presente modelo é uma expansão do proposto por Meyr (2002), com as adaptações propostas por Ulbricht e Volpi (2015), Józefowska e Zimniak (2008), Wörbelauer, Meyr e Almada-Lobo (2019), conforme demonstrado na Tabela 2.

4.1. MODELAGEM MATEMÁTICA

O modelo deve apresentar as seguintes características:

- Minimizar os custos envolvidos na produção através da penalidade por falta de produto quando há pedido em carteira, redução de estoque acabado, otimização de *setups* e produzir na máquina mais

econômica até o limite da sua capacidade e só então produzir em outra máquina mais cara;

- Multi recursos: máquinas paralelas e heterogêneas;
- Estágio único: os produtos passam por apenas uma etapa de produção;
- *Setup* dependente: os custos de preparação não podem ser ignorados e estes dependem do sequenciamento;
- Capacitado: os recursos são limitados e as capacidades devem ser respeitadas.

Considerando os produtos $i, j = 1, \dots, P$ a serem sequenciados nas máquinas $m = 1, \dots, M$, utilizando os componentes $c = 1, \dots, C$, num horizonte finito de planejamento com macro períodos $t = 1, \dots, T$ de duração definida e micro períodos s pertencentes ao macro período t na máquina m .

Constantes e dados fornecidos:

S_{mt}	Conjunto de micro períodos s pertencentes ao macro período t e à máquina m
C_{mt}	Capacidade (em tempo) da máquina m no macro período t
a_{im}	Processamento (em tempo) do produto i na máquina m .
mp_{im}	Mão de obra necessária para produzir o produto i na máquina m .
cp_{im}	Penalidade de custo extra por produzir o produto i na máquina m , quando a produção não for feita na máquina m mais eficiente.
l_i	Lote mínimo do produto i
d_{it}	Demanda do produto i no período t
h_i	Custos de manutenção de estoque do produto i para cada período t
p_i	Penalidade por não ter estoque do produto i para cada período t
ins_{ct}	Quantidade recebida do insumo c no final do período t
q_{ci}	Quantidade aplicada do insumo c para produzir o produto i
sc_{ijm}	Custos de preparação e mudança produto i para o produto j na máquina m
st_{ijm}	Tempo de preparação e mudança produto i para o produto j na máquina m
o_m	Custo de cada hora extra na máquina m
Mo_{mt}	Quantidade máxima de horas extras que podem ser feitas na máquina m no período t

I_{i0}	Estoque do produto i no período $t = 0$, início do período de planejamento
e_{c0}	Estoque do componente c no período $t = 0$
Pl_i	Quantidade de unidades do produto i em cada paleta para armazenagem
y_{mi0}	Assume valor 1 se a máquina m está preparada para iniciar a produção do produto i no período $t = 0$, início do período de planejamento
Q	Capacidade máxima de estocagem em paletes

Variáveis de decisão:

x_{ims}	Quantidade produzida do produto i , na máquina m no micro período s
I_{it}^+	Estoque do produto i ao final do período t
I_{it}^-	Demanda não atendida do produto i ao final do período t
e_{ct}	Estoque do insumo c ao final do período t
y_{mis}	Variável binária que denota se a máquina m está preparada para iniciar a produção do produto i no micro período s
$z_{mij s}$	Variável que denota se há troca do produto i para o produto j , na máquina m , no micro período s
o_{mt}	Quantidade da horas extras utilizadas pela máquina m no período t

Função objetivo:

$$\begin{aligned} \min \sum_{i=1}^P \sum_{t=1}^T h_i \times I_{it}^+ + \sum_{i=1}^P \sum_{t=1}^T p_i \times I_{it}^- + \sum_{m=1}^M \sum_{i=1}^P \sum_{s=1}^S cp_{im} \times x_{ims} \\ + \sum_{m=1}^M \sum_{i=1}^P \sum_{j=1}^P \sum_{s=1}^S sc_{ijm} \times z_{mij s} + \sum_{m=1}^M \sum_{t=1}^T che_m \times o_{mt} \end{aligned} \quad (1)$$

Sujeito a:

$$I_{it}^+ - I_{it}^- = I_{it-1}^+ - I_{it-1}^- + \sum_{m,s \in S_{mt}} x_{ims} - d_{it} \quad \forall t, i \quad (2)$$

$$\sum_i N_i \leq Q \quad \forall t \quad (3)$$

$$N_i \geq \frac{I_{it}^+}{Pl_i} \quad \forall t \quad (4)$$

$$\sum_{i,s \in S_{mt}} a_{im} \times x_{ims} + \sum_{i,j,s \in S_{mt}} st_{ijm} \times z_{mij s} \leq C_{mt} + o_{mt} \quad \forall m, t \quad (5)$$

$$\sum_{i,m,s \in S_{mt}} a_{im} \times mp_{im} \times x_{ims} \leq Cmp_t + \sum_m o_{mt} \quad \forall t \quad (6)$$

$$o_{mt} \leq Mo_{mt} \quad \forall m, t \quad (7)$$

$$x_{ims} \leq \frac{C_{mt}}{a_{im}} \times y_{mis} \quad \forall m, i, s \quad (8)$$

$$\sum_i y_{mis} = 1 \quad \forall m, s \quad (9)$$

$$x_{ims} \geq l_i \times (y_{mis} - y_{mis-1}) \quad \forall m, i, s \quad (10)$$

$$z_{mij_s} \geq y_{mis-1} + y_{mjs} - 1 \quad \forall m, i, j, s \quad (11)$$

$$e_{ct-1} + ins_{ct} - \sum_{i,m} q_{ci} \times x_{ims} = e_{ct} \quad \forall c, t \quad (12)$$

$$y_{mis} \in \{0,1\}; x_{ims}, I_{it}^+, I_{it}^-, N_i \in \mathbb{N}; z_{mij_s}, o_{mt} \geq 0 \quad (13)$$

A função objetivo (1) trata da minimização dos custos de manutenção dos itens em estoque, penalidade por não entregar o produto ao cliente, produção considerando as diferentes máquinas, *setups* e custos com horas extras. Pode parecer óbvio que $p_i > h_i \forall i$, pois do contrário o modelo poderia optar por arcar com a penalidade da falta do que produzir. O mesmo acontece com $p_i > cp_{im} \forall i, m$, o que obrigatoriamente exige que cp_{im} seja uma penalidade por não produzir no recurso mais eficaz e que $cp_{im} = 0$ quando a produção do lote for alocada no recurso mais econômico.

A restrição (2) faz o balanço do estoque e unidades pendentes de entrega (*backorders*) com aquilo que foi produzido e demandado no período, de forma a manter os dados no horizonte de planejamento.

As restrições (3) e (4) limitam a quantidade máxima de paletes em estoque, que é decorrente da capacidade de armazenagem. Na prática, a limitação é dada em posições-paleta, sendo necessário arredondar para cima sempre que um paleta fica incompleto, pois mesmo contendo poucas caixas o paleta continua ocupando uma posição. Outro ponto importante a se considerar é que, se há dois paletes de produtos diferentes e ambos têm metade da quantidade, não significa que a soma das duas metades contará como somente um paleta ocupado. Dois paletes incompletos, ainda que a soma deles seja menor que um, resultam em duas posições paleta ocupadas.

A restrição (5) limita a capacidade às horas máquina disponíveis, descontando tempos de setup e adicionando horas extras. Já a restrição (6) limita a

produção às horas-homem disponíveis, também adicionando eventuais horas extras. As horas extras são limitadas pela restrição (7), que é necessária considerando que nem todas as máquinas possuem disponibilidade todos os dias, como por exemplo os recursos que já estão disponíveis 24 horas por dia, ou seja, não existem mais horas para capacidade adicional.

As restrições (8) e (9) garantem que só haverá produção se a máquina estiver preparada e que cada máquina só estará preparada para produzir um produto em cada micro período. Já a restrição (11) faz a conexão entre os *setups* e o estado do recurso no que tange à disponibilidade para fazer os produtos.

A restrição (10) estabelece o lote mínimo de produção de cada produto sempre que houver *setup*. Aqui cabe reforçar que, em alguns segmentos da indústria, é possível que $sc_{ijm} + sc_{jkm} \leq sc_{ikm}$. Mesmo assim, optou-se por manter esta restrição considerando que, se a origem dos dados históricos partir dos apontamentos de produção, pode acontecer destes dados não serem altamente confiáveis, principalmente se os apontamentos não forem automáticos. Caso seja este o caso, é recomendável que seja utilizada matriz determinística para os custos de *setups*. Avaliando a base de dados utilizada na validação do modelo proposto, percebe-se que para uma mesma troca de i para j na máquina m , há *setups* em que o custo varia dez vezes. Tal variabilidade é explicado pela inconsistência da informação que o operador da máquina insere no sistema, muito embora uma pequena variação seja esperada e inerente ao processo de *setup*.

A restrição (12) impede que seja programada uma quantidade de produção sem que haja insumos. Na prática, não são raras as situações que existe pedido do produto e que não é possível produzi-lo por falta de alguma matéria-prima. Isto desencadeia a necessidade de reprogramar a produção e movimentar várias áreas para não perder horas de recurso produtivo.

Por fim, a restrição (13) determina os limites das variáveis e seu tipo.

4.2. PROGRAMAÇÃO EM LINGUAGEM COMPUTACIONAL

A fim de demonstrar, de forma sucinta, as etapas para criação do artefato, o código computacional será explicitado e explicado a seguir. Para cumprir a etapa de transformar o modelo matemático em programação computacional, foi utilizada a linguagem *Python*, com a biblioteca PuLP e o *solver* Gurobi. A escolha do *Python* e

PuLP deveu-se à grande flexibilidade, conhecimento prévio do autor, disponibilidade de documentos de suporte e da abrangência da comunidade acadêmica em fóruns de discussão. Já para o *solver* os fatores foram a possibilidade de obter uma licença acadêmica sem custo, o excelente desempenho reconhecido pela própria comunidade acadêmica e a sua integração com a biblioteca PuLP.

A primeira etapa consiste em importar as bibliotecas necessárias para o código. A biblioteca Pandas serve para o tratamento dos dados e apresentação ao usuário, que serão explicados na última etapa. Já a biblioteca PuLP é a que faz a construção do arquivo representativo do modelo, previamente ao início da etapa de solução do modelo.

```
import pandas as pd
import pulp
```

A criação da variável do problema é a próxima etapa necessária, em que `prob` assume ser uma classe de problema definida na biblioteca PuLP, com nome `GLSSPP` e com finalidade de minimizar a função objetivo.

```
prob = pulp.LpProblem("GLSSPP", pulp.LpMinimize)
```

A próxima etapa consiste na definição das variáveis envolvidas no problema, seus tipos e limites, conforme a equação (13). A variável binária y , acompanhada dos seus respectivos índices, pode ser verificada a seguir. Interessante notar que, após criação da variável, o estado inicial de configuração das máquinas é atualizado.

```
#Binary
y_m_t_s_i = {(m,t,s,i):
pulp.LpVariable(cat=pulp.LpBinary,
name="y_{0}_{1}_{2}_{3}".format(m,t,s,i))
for (i,m) in set_i_m for (t,s) in set_t_s if s>0}
y_m_t_s_i.update(Y_m_0_0_i)
```

Subsequentemente são definidas as variáveis inteiras e contínuas, seguindo a mesma lógica de aplicar os limites para cada índice, bem como o estado inicial quando aplicável.

```
#Integer
n_i_t = {(i,t): # pallets
```

```

pulp.LpVariable(cat=pulp.LpInteger, lowBound=0,
name="n_{0}_{1}".format(i,t))
for i in set_i for t in set_t}

z_m_i_j_t_s = {(m,i,j,t,s): # setup
pulp.LpVariable(cat=pulp.LpInteger, lowBound=0,
name="Z_{0}_{1}_{2}_{3}_{4}".format(m,i,j,t,s))
for (i,j,m) in set_i_j_m for (t,s) in set_t_s if s>0}

#Continuous
x_i_m_t_s = {(i,m,t,s): # produced quantity
pulp.LpVariable(cat=pulp.LpContinuous, lowBound=0,
name="x_{0}_{1}_{2}_{3}".format(i,m,t,s))
for (i,m) in set_i_m for (t,s) in set_t_s}

Ip_i_t = {(i,t): # inventory to hold
pulp.LpVariable(cat=pulp.LpContinuous, lowBound=0,
name="Ip_{0}_{1}".format(i,t))
for i in set_i for t in set_t}
Ip_i_t.update(Ip_i_0)

Im_i_t = {(i,t): # backorders
pulp.LpVariable(cat=pulp.LpContinuous, lowBound=0,
name="Im_{0}_{1}".format(i,t))
for i in set_i for t in set_t}
Im_i_t.update({(i, -1): 0 for i in set_i})

e_c_t = {(c,t): # inventory of components
pulp.LpVariable(cat=pulp.LpContinuous, lowBound=0,
name="e_{0}_{1}".format(c,t))
for c in set_c for t in set_t}
e_c_t.update(e_c_0)

o_m_t = {(m,t): # overtime
pulp.LpVariable(cat=pulp.LpContinuous, lowBound=0, upBound=2,
name="o_{0}_{1}".format(m,t))
for m in set_m for t in set_t}

```

A função objetivo é definida através da variável de mesmo nome. Após a definição, é atribuída ao problema.

```

#Objective
objective = pulp.lpSum(P_i[i] * Im_i_t[i,t] for i in set_i for t in
set_t) + pulp.lpSum(H_i[i] * Ip_i_t[i,t] for i in set_i for t in
set_t) + pulp.lpSum(CP_i_m[i,m] * x_i_m_t_s[i,m,t,s] for (i,m) in
set_i_m for t in set_t for s in set_s) + pulp.lpSum(SC_i_j_m[i,j,m]
* z_m_i_j_t_s[m,i,j,t,s] for (t,s) in set_t_s if s>0 for (i,j,m) in
set_i_j_m) + pulp.lpSum(CO * o_m_t[m,t] for m in set_m for t in
set_t)

```



```
prob.setObjective(objective)
```

Uma vez definida a função objetivo, cada restrição foi inserida. Para facilitar o relacionamento entre a restrição e a equação apresentada na formulação matemática, o leitor pode basear-se no número entre parênteses que consta na primeira linha de cada restrição, precedido pelo símbolo #, que na linguagem Python estabelece a linha de comentário, ou seja, esta linha é ignorada na execução do programa, servindo apenas para referência e comunicação durante a escrita do código.

```
# warehouse capacity (3)
constraints = {(t): prob.addConstraint(
pulp.LpConstraint(
    e=(pulp.lpSum(n_i_t[i,t] for i in set_i) - Q),
    sense=pulp.LpConstraintLE,
    rhs=0,
    name="warehouse_cap_{0}".format(t)))
    for t in set_t}

# palet round up (4)
constraints = {(i,t) : prob.addConstraint(
pulp.LpConstraint(
    e=(n_i_t[i,t] * PL_i[i] - Ip_i_t[i,t]),
    sense=pulp.LpConstraintGE,
    rhs=0,
    name="palet_round_up_{0}_{1}".format(i,t)))
    for i in set_i for t in set_t}

# Capacity of machine (5)
constraints = {(m,t) : prob.addConstraint(
pulp.LpConstraint(
    e=(pulp.lpSum(A_i_m[i,m] * x_i_m_t_s[i,m,t,s] for
(i,m_) in set_i_m if m_ == m for s in set_s) +
pulp.lpSum(ST_i_j_m[i,j,m] * z_m_i_j_t_s[m,i,j,t,s] for (i,j,m_) in
set_i_j_m if m_ == m for s in set_s) - C_m_t[m,t] - o_m_t[m,t]),
    sense=pulp.LpConstraintLE,
    rhs=0,
    name="capacity_machine_{0}_{1}".format(m,t)))
    for m in set_m for t in set_t}

# Capacity of manpower (6)
constraints = {(t) : prob.addConstraint(
pulp.LpConstraint(
    e=(pulp.lpSum(A_i_m[i,m] * x_i_m_t_s[i,m,t,s] *
MP_i_m[i,m] for (i,m) in set_i_m for s in set_s) - Cmp_t[t] -
pulp.lpSum(o_m_t[m,t] for m in set_m)),
```

```

        sense=pulp.LpConstraintLE,
        rhs=0,
        name="capacity_menpower_{0}".format(t)))
    for t in set_t}

# Maximum overtime by machine (7)
constraints = {(m,t) : prob.addConstraint(
pulp.LpConstraint(
    e=(o_m_t[m,t] - MOT_m_t[m,t]),
    sense=pulp.LpConstraintLE,
    rhs=0,
    name="max_overtime_{0}_{1}".format(m,t)))
    for t in set_t for m in set_m}

# Only produces if setup ready (8)
constraints = {(m,i,t,s) : prob.addConstraint(
pulp.LpConstraint(
    e=(x_i_m_t_s[i,m,t,s] * A_i_m[i,m] - C_m_t[m, t] *
y_m_t_s_i[m,t,s,i]),
    sense=pulp.LpConstraintLE,
    rhs=0,
    name="setup_ready_{0}_{1}_{2}_{3}".format(m,i,t,s)))
    for (i,m) in set_i_m for (t,s) in set_t_s}

# Only produces one product each s (9)
constraints = {(m,t,s) : prob.addConstraint(
pulp.LpConstraint(
    e=(pulp.lpSum(y_m_t_s_i[m,t,s,i] for (i,m_) in set_i_m
if m_ == m)),
    sense=pulp.LpConstraintEQ,
    rhs=1,
    name="one_per_s_{0}_{1}_{2}".format(m,t,s)))
    for m in set_m for (t,s) in set_t_s if s>0}

# connection between setups and resource state (11)
constraints = {(m,i,j,t,s) : prob.addConstraint(
pulp.LpConstraint(
    e=(z_m_i_j_t_s[m,i,j,t,s] -
y_m_t_s_i[m,set_t_s[set_t_s.index((t,s))-
1][0],set_t_s[set_t_s.index((t,s))-1][1],i] - y_m_t_s_i[m,t,s,j]),
    sense=pulp.LpConstraintGE,
    rhs=-1,
    name="setup_state_{0}_{1}_{2}_{3}_{4}".format(m,i,j,t,s)
)))
    for (i,j,m) in set_i_j_m for (t,s) in set_t_s if s>0}

# minimal lot size (10)
constraints = {(m,i,t,s) : prob.addConstraint(
pulp.LpConstraint(

```

```

        e=(x_i_m_t_s[i,m,t,s] - L_i[i] * (y_m_t_s_i[m,t,s,i] -
y_m_t_s_i[m,set_t_s[set_t_s.index((t,s))-
1][0],set_t_s[set_t_s.index((t,s))-1][1],i))),
        sense=pulp.LpConstraintGE,
        rhs=0,
        name="min_lotsize_{0}_{1}_{2}_{3}".format(m,i,t,s)))
    for (i,m) in set_i_m for (t,s) in set_t_s if s>0}

# component consumption (12)
constraints = {(c,t) : prob.addConstraint(
pulp.LpConstraint(
    e=(pulp.lpSum(QA_c_i[c,i] * x_i_m_t_s[i,m,t,s] for
(i_,m) in set_i_m for (i,c_) in set_i_c if c_ == c if i == i_ for
(t_,s) in set_t_s if t_ == t and s>0) + e_c_t[c,t] - e_c_t[c,t-1] -
CR_c_t[c,t-1]),
    sense=pulp.LpConstraintEQ,
    rhs=0,
    name="components_{0}_{1}".format(c,t)))
    for c in set_c for t in set_t}

```

Neste momento, todos os requisitos do modelo estão definidos. O próximo passo é chamar o *solver*. É possível perceber que existem alguns parâmetros informados ao Gurobi, quer seja manter os arquivos do modelo e o arquivo com as soluções, parar a busca quando o encontrar uma solução que seja até 1% do valor ótimo ou quando atingir o tempo de 1200 segundos.

```

prob.solve(pulp.GUROBI_CMD(keepFiles=True,gapRel=0.01,timeLimit=1200
))

```

Por fim, é necessário que o resultado seja formatado para o usuário. Esta etapa é cumprida gravando os dados finais em planilha do *Microsoft Excel* através da biblioteca Pandas, deixando visíveis os dados que são efetivamente relevantes: quantidade de cada produto a ser produzida em cada máquina, na sequência de execução; *setups* que devem realizados considerando sequência, máquina, produto anterior e produto posterior.

```

with pd.ExcelWriter('./solution.xlsx') as writer:
    for res in [{'x_i_m_t_s': x_i_m_t_s},
{'z_m_i_j_t_s':z_m_i_j_t_s}, {'y_m_t_s_i': y_m_t_s_i}]:
        key = list(res.keys())[0]
        df = pd.DataFrame.from_dict(res.get(key), orient="index",
columns=["var"])
        name = key.split('_')
        df.index = pd.MultiIndex.from_tuples(df.index,
names=name[1:])

```

```
df.reset_index(inplace=True)
df[name[0]] = df["var"].apply(lambda item: item.varValue if
type(item) != int else item)
df.drop(columns=["var"], inplace=True)
df.drop(df[df[name[0]] <= 0.001].index, inplace=True)
print(df)
df.to_excel(writer, sheet_name=key)
```

O código apresentado neste capítulo é uma adaptação do modelo matemático para uma linguagem que o computador consegue processar. A inserção de dados depende de transformação prévia, que varia conforme cada caso, principalmente dependente do formato que os dados são disponibilizados. Popularmente é a etapa chamada de “de/para” em rotinas de trabalhos computacionais, com objetivo de informar a origem de cada variável necessária ao modelo.

5. APLICAÇÃO DO ARTEFATO

A avaliação do modelo deve ser realizada em ambiente real, conforme enunciado na etapa de definição da metodologia. Para realizar a aplicação, foi selecionada uma indústria do ramo de descartáveis plásticos, com ambiente produtivo adequado e contendo as restrições previamente definidas no modelo.

O ambiente de produção objetivo de modelagem conta com 18 injetoras, 60 moldes e 350 produtos acabados. A fábrica funciona 24 horas por dia, com algumas injetoras sendo desligadas aos domingos por que não há demanda suficiente para justificar seu funcionamento pleno, enquanto outras não desligam. O setor de injeção conta com 40 funcionários em cada turno, divididos entre embaladores, operadores de injeção e líderes.

A maioria dos itens são produzidos no modelo *make to stock*, em que a estocagem precede a venda e o cliente espera disponibilidade imediata ao colocar o pedido. Estes itens geram custo de armazenagem e a sua falta pode gerar tanto o custo de venda perdida ou *backorder*, que é o componente tangível, quanto o descontentamento do cliente, componente intangível. Também existem itens *make to order*, que são produzidos para clientes específicos, na embalagem própria do cliente (*private label*), os quais a empresa não mantém estoque e os clientes aceitam um período ajustado entre a inserção da ordem de compra e a efetiva produção.

Anualmente, cerca de 3 meses antes do próximo exercício fiscal, a empresa realiza o planejamento das vendas. Esta etapa é conhecida como *Sales & Operation Plan (S&OP)* e envolve as áreas de Compras, Vendas, Produção e Logística. A equipe de Vendas prepara a previsão por grupo, subgrupo e classe de produtos, dividida mensalmente, e repassa para a Produção. Por sua vez, a equipe de Produção avalia se existe capacidade de recursos para atingir a expectativa de vendas, identificando itens com forte sazonalidade e que obrigatoriamente vão requerer antecipação da produção. Eventuais itens que superem a disponibilidade anual são tratados com a Direção da empresa, no que tange a novos investimentos para expansão, ou então são devolvidos para Vendas como produtos que não devem ser estimulados. Nestes casos, Vendas estabelece um limite no ano, podendo, inclusive, aumentar o preço para desestimular a demanda. Uma vez que a capacidade produtiva está respeitada, Logística avalia a capacidade de armazenagem, com base na produção e previsão de vendas, dando especial atenção para os produtos

sazonais. Compras identifica potenciais gargalos e começa a tratar com os fornecedores. Estes dois últimos setores não tem poder para alterar o plano, mas devem participar do desenvolvimento para que possam identificar potenciais problemas.

Já no médio prazo a previsão de demanda é informada pelo setor de Vendas e *Marketing* em período semanal, para as próximas 12 semanas. A cada semana o número é revisado, descartando a semana que encerrou e adicionando a última semana na série, no modelo conhecido como previsão rolante. Esta previsão é feita agregando produtos da mesma família, que na prática são diferenciados somente pela cor do item. Ademais, assume-se como premissa do modelo que os insumos necessários para a produção no período sempre estão disponíveis no início.

No curto prazo, a Produção tem conhecimento dos recursos que não pode desligar, mesmo que esteja com o estoque em níveis altos. A programação é feita observando a disponibilidade de cada item e a programação de carregamento de veículos: se algum item falta, este ganha prioridade máxima na produção. Para os demais itens, trabalha no modelo de supermercado, abastecendo os depósitos com a quantidade que foi vendida desde a última produção do item e respeitando os lotes mínimos.

Ainda na operacionalização da programação da produção, com base na observação e entrevistas que se estenderam ao longo de meses, foi possível perceber algumas decisões de sequenciamento sendo questionadas pela equipe operacional da fábrica. Percebeu-se que boa parte das decisões antieconômicas eram por falta de visibilidade das combinações possíveis, por parte do programador da produção, que dificilmente alcança o melhor plano de produção utilizando o refinamento por tentativa e erro. Este obstáculo é ainda mais aparente em colaboradores novos na função, pois é natural que, à medida que a experiência vá aumentando, emergem decisões mais aderentes ao aumento da rentabilidade do negócio.

Outro fator que pode ser observado na atividade de planejar a produção é o conflito de interesses entre as áreas que circundam a esfera de decisão do programador. São áreas (ou departamentos) com as quais este colaborador precisa dialogar a fim de resolver conflitos, decorrentes das decisões inerentes à função. Como exemplo prático baseado em caso específico, percebeu-se que o setor de Vendas interferia na programação para atender um cliente especial, em detrimento de todo o planejamento previamente realizado. Nesta situação, Vendas utiliza-se de sua

esfera de influência para atender o cliente, mesmo que, no curtíssimo prazo, tal decisão possa dar causa ao aumento dos custos produtivos, gerando uma medida antieconômica. Ainda que pudesse ser defendida com pensamento de longo prazo, argumentando sobre a importância de determinado cliente, momentaneamente é uma distorção naquilo que os gestores envolvidos na atividade produtiva buscam: redução nos custos produtivos sem ignorar aspectos de qualidade do produto.

Outra área que demonstrou exercer forte influência na decisão da programação foi o setor de Manutenção: foi estabelecido o limite de três trocas de molde por dia e definidos horários nos quais não poderia haver troca de molde. Ficavam fora do plano os sábados, domingos e segundas-feiras até ao meio-dia. Os dois primeiros pela indisponibilidade de equipe, já que o setor de Manutenção não funcionava nos finais de semana. O terceiro dia sob a alegação de que, ao iniciar a semana, muitas máquinas que foram desligadas no domingo requeriam a atenção dos técnicos até que entrassem em funcionamento pleno. Aqui cabe ressaltar que ultrapassar o limite de trocas de molde estabelecido era uma raridade da ordem de dez ocorrências ao longo do ano todo, muito embora não houvesse parâmetro para afirmar ou negar que a possibilidade de fazer mais do que três *setups* na máquina seria fonte de economia. Esta restrição é negociável se houver justificativa para tal. Em algumas épocas do ano, com alto impacto da sazonalidade, são necessárias mais trocas no mesmo dia. Como esta restrição é passível de flexibilização, não será considerada no modelo.

A equipe de Manutenção não é dedicada para *setups*, então existem duas restrições que impactam o modelo. A primeira refere-se à impossibilidade de realizar dois *setups* simultaneamente, conforme mencionado durante a revisão bibliográfica. A segunda refere-se à limitação de até três *setups* por dia.

O setor de Compras também é fonte de influência na programação da produção, entretanto contava com uma característica impositiva: as alterações causadas no planejamento eram oriundas da falta de alguma matéria prima, ou seja, o programador não podia decidir pela situação mais econômica dado que não haveria insumo. De certa forma podemos caracterizar como uma restrição de recursos, não uma decisão por influência.

A produção incorre em custos de *setup* da injetora em vários casos, mesmo que seja somente a troca da pigmentação para produzir item similar, entretanto de outra cor. Alguns *setups* tem seus custos negligenciados, pois é apenas alterado o

fracionamento do item (pacotes para granel e vice-versa). Já outros *setups* são mais complexos e custosos, pois envolvem a troca do molde de injeção.

Em relação à capacidade produtiva, temos uma combinação de produto e máquina, ou seja, o mesmo produto pode ter taxas de produção diferentes entre as máquinas. Isto se deve ao avanço tecnológico durante o intervalo de aquisição dos equipamentos. A ocorrência de falhas no maquinário também influencia o resultado final, sendo que neste caso assume-se uma média de falhas e adiciona-se esta variável à taxa de produção ideal, com base no histórico do par ordenado produto e máquina.

A empresa possui sistema informatizado que centraliza as informações. Os apontamentos de produção acontecem em tempo real, de forma manual, na medida em que um palete de produto é completado ou quando a produção do item é encerrada. Os apontamentos de perda (refugo ou avara) são feitos ao final de cada turno ou na ocorrência de situação específica, como exemplo a troca de molde, sempre informando o motivo que deu causa à geração do refugo. Os apontamentos de parada de máquina são capturados pela injetora e inseridos no sistema automaticamente, sendo estes os que apresentam maior confiabilidade.

Com os dados históricos disponíveis na base de dados da empresa, iniciou-se o desafio de conectar o modelo a esta base. A etapa foi vencida utilizando biblioteca própria do banco de dados *Oracle* para *Python*, que permitiu criar as consultas (*queries*) em formato *Structured Query Language* (SQL) e converter o resultado em formato *Java Script Object Notation* (JSON), o qual é facilmente manipulável através do através do *Python*. Por uma questão de sigilo, as consultas e os dados não serão publicados, uma vez que permitiriam identificar a empresa bem como a exposição de dados sensíveis do negócio.

A sequência de construção das informações iniciou-se pelo conjunto de recursos m , no qual foram importados para o modelo os recursos cadastrados e com situação ativa. Cabe mencionar que a base de dados dispõe de informações dos últimos 20 anos, na qual muitos outros recursos foram cadastrados e hoje não estão mais ativos. Para o propósito deste trabalho, os inativos precisam ser ignorados. Juntamente com a informação do recurso, foi importada da base de dados qual é o produto i que foi produzido por último no recurso, sendo crucial para posteriormente determinar o estado inicial de cada máquina, ou seja, para qual produto o recurso está preparado.

Posteriormente foi construído o conjunto de produtos i , seguindo a mesma lógica de ignorar os que foram descontinuados. Nesta consulta foi possível obter o lote mínimo de produção, quantidade armazenada em cada palete, estoque inicial e demanda diária de cada item.

Com os dois conjuntos de dados iniciais, foi possível criar os pares ordenados (m, i) de forma a definir quais os itens são produzidos em cada máquina, qual é a taxa de produção padrão, a mão de obra necessária para produzir uma unidade de i em m e o custo. O custo foi convertido em penalidade, subtraindo o menor valor obtido para cada i , fazendo com que a penalidade fosse 0 no recurso mais eficiente. Também partindo-se do menor custo de produção foi possível estabelecer o custo de manutenção de estoque, aplicando a taxa de custo de capital definida pela empresa, com capitalização diária, bem como o lucro cessante por venda perdida.

Além disso, também foi possível criar a matriz de *setups*, combinando todas as trocas de i para j em cada recurso m , desde que fizesse parte do conjunto (m, i) e $i \neq j$, afinal de contas é irrelevante a troca de um produto para ele mesmo. Assim originou-se a matriz de *setups* (m, i, j) , que por característica dessa indústria é assimétrica, conforme explicitado anteriormente. Esta técnica possibilitou obter dados históricos de tempo do *setup* através dos apontamentos de máquina parada, bem como do custo do *setup* através dos apontamentos de perda. Entretanto, surgiu o problema de que nem todas as combinações de trocas de produtos estavam no histórico. Percebeu-se que menos de 35% das possibilidades estavam preenchidas na matriz, o que demonstra que existe um padrão consistente na sequência da programação, muito embora não tenha sido possível comprovar que esta sequência é a de menor custo. Para superar esta dificuldade foram analisados *setups* de produtos similares, o que possibilitou preencher cerca de 80% da matriz. O restante foi estimado pelo programador da produção, em conjunto com os líderes de turno e gestor da Produção.

O conjunto dos componentes c foi obtido partindo-se do conjunto i e consultando a lista de materiais aplicados para cada i . Posteriormente foi extraído da base de dados a quantidade de estoque inicial para cada c e a programação de recebimento dos insumos, conforme as ordens de compra cadastradas no sistema e o prazo de entrega acordado com cada fornecedor.

Por fim, foi definido o intervalo de planejamento, em conjunto com os gestores envolvidos no processo de programação da produção, sendo que o horizonte de dois dias é suficiente para uso do modelo. Atualmente a empresa faz a programação de curto prazo somente para um dia, dada a volatilidade da carteira de pedidos. Mas apesar do modelo ter sido definido para dois períodos, a programação deve ser revista a cada dia, ou seja, novamente deve ser executado o modelo. Em relação aos microperíodos, foi estabelecido que cinco são suficientes, com base no histórico. Raros os casos em que este número de microperíodos foi plenamente utilizado, o que se pode traduzir como o número de produtos distintos a serem produzidos no mesmo recurso dentro do macroperíodo. De posse destas definições de períodos, foi extraído da base de dados as capacidades (em horas) de cada recurso, conforme definido no calendário de funcionamento da fábrica.

O máximo de horas extras em cada recurso foi limitado a duas horas por dia, considerando que recursos com 24 horas de produção disponíveis no calendário de funcionamento da fábrica, por óbvio, não dispõe de mais horas no dia para aplicação de horas extraordinárias. A capacidade de armazenagem informada pelo setor de Logística é de 2600 paletes. O custo adicional para cada hora extra foi informado pelo departamento de Gestão de Pessoas como sendo de R\$ 200,00/h. Já o número de colaboradores, multiplicado pela jornada diária, resultou em uma capacidade de força de trabalho de 850 horas por dia, já descontando a mão de obra indireta.

5.1. TESTES REALIZADOS

Com objetivo de demonstrar o nível de eficiência do modelo, foram realizados diversos testes variando o tamanho do problema, aumentando o número de períodos, e também variando o tempo limite ou o intervalo relativo entre a solução encontrada e a solução ótima. A Tabela 3 demonstra que o modelo, apesar de necessitar considerável tempo computacional para chegar na solução ótima, consegue atingir resultados viáveis em prazo relativamente curto, quando se avalia a necessidade da indústria para obter a solução, bem como o horizonte de tempo na solução apresentada.

Tabela 3 - Comparativo de testes.

Instância	Macroperíodos	Microperíodos	Tempo (min)	Gap (%)
1	3	10	20	18,8
2	3	10	250	4,1
3	3	5	20	12,3
4	3	5	120	7,2
5	2	5	120	1
6	2	5	600	0

Fonte: autor (2023).

A fim de demonstrar a eficácia do modelo no aspecto qualitativo, foi realizada também um teste durante 7 dias de programação. O ponto de partida foi exatamente o mesmo, tanto para o caso em que o programador da produção realizou o planejamento quanto para o teste através do modelo. Considerando o período 0, em que a carteira de pedidos, os níveis de estoque, a capacidade de produção, a taxa de ocupação do armazém e os custos estão disponíveis e são conhecidos, o modelo foi executado para 2 macroperíodos contendo 5 microperíodos, sem limite de tempo de execução.

Subsequentemente, para o período 1 foram agregadas as demandas igualmente e considerados os planos de produção executados separadamente. A programação foi realizada seguindo a lógica de agregar somente a demanda diária, tanto no plano de produção executado manualmente quanto no plano executado pelo modelo, de forma rolante, até atingir o período 7. Desta forma, os componentes de incerteza relativos ao mercado não exerceram influência no resultado final.

Tabela 4 - Comparativo de custos do plano de produção na baixa sazonalidade.

Período	Custo relativo acumulado (%)
0	-0,05%
1	-0,18%
2	-0,93%
3	-1,35%
4	-1,89%
5	-2,29%
6	-2,84%

Fonte: autor (2023).

O modelo conseguiu realizar a programação acumulando uma redução de custo de 2,84% em relação ao programador da produção. Dois fatos chamaram atenção quando comparadas as programações: o nível final de estoque de produtos acabados, que no resultado do modelo foi menor que o apresentado pelo programador; em nenhum momento, tanto no teste quanto na realidade, houve penalidade por perda de pedido.

Ao final do primeiro teste foi levantada a hipótese do modelo, por ter um horizonte mais curto de planejamento e ter sido simulado em período com capacidade produtiva suficiente para atender a demanda, ter dado causa para a redução do estoque sem que houvesse penalidade por falta de produtos. Tal hipótese foi testada através de uma nova simulação, utilizando os mesmos parâmetros de 2 macropérios contendo 5 micropérios, sem limite de tempo de execução, porém aplicada a outro momento do ano.

Neste novo teste o período escolhido para o ponto de partida foi tal que a empresa, historicamente, tem alto volume de pedidos, incorre em rupturas de estoques e capacidade insuficiente de produção. Seguindo a mesma lógica usada no teste anterior, o momento zero já continha pedidos atrasados, ou seja, tanto o programador quanto o modelo precisaram tratar esta situação que é considerada crítica na indústria objeto do estudo de caso.

Tabela 5 - Comparativo de custos do plano de produção na alta sazonalidade.

Período	Custo relativo acumulado (%)
0	0,15%
1	-0,08%
2	-0,16%
3	0,02%
4	-0,12%
5	-0,25%
6	-0,08%

Fonte: autor (2023).

Neste segundo teste percebe-se que o modelo entregou um resultado muito próximo daquele que o programador também atingiu. Durante todo o período de teste, os custos com *backlog* foram altos se comparados ao teste durante a baixa sazonalidade. Sem capacidade produtiva para recuperar os estoques, tanto o modelo quanto o programador não conseguem melhorar os custos da função objetivo. Isto demonstra a importância de evitar a chegada nessa situação de constantes rupturas.

O tempo para se chegar na solução ótima foi praticamente o mesmo do teste anterior, com 9h52min de execução. Os níveis de estoque finais apresentaram ligeira variação, sendo que o modelo terminou com 0,26% a mais de estoque, em valores monetários.

5.2. RESULTADOS

O modelo resolveu o problema para uma indústria com 18 máquinas, 350 produtos acabados, 4576 componentes, 2 macro períodos de programação, cada um contendo 5 micro períodos variáveis e com armazém de 2600 posições palete. A solução ótima foi encontrada em 10h de processamento utilizando um computador com processador Intel i5-2410M 2.30GHz *quad-core* com 6GB de memória RAM, *solver* Gurobi, programado em linguagem *Python* utilizando a biblioteca PuLP. Muito embora este tempo total de processamento possa ser considerado alto para uma aplicação prática, o modelo atingiu o *gap* de 1% do melhor resultado em aproximadamente 2h.

O teste comparativo com o atual método de programação, aplicada ao caso prático, demonstrou que o modelo conseguiu atingir critérios qualitativos, reduzindo o custo do plano de produção em 2,84% quando considerados 7 macroperíodos consecutivos de programação, na baixa sazonalidade. Dentre as saídas entregues pelo artefato também foram obtidos os planos de produção, que responderam às perguntas feitas na introdução deste trabalho, trazendo ao programador da produção quanto deveria produzir de cada item, em qual máquina deveria ser alocado, quando deveria ser realizada produção sob regime de hora extra, quando realizar os *setups* e quais produtos substituir, etc. Os números foram suprimidos deste trabalho por questão de sigilo, uma vez que fazendo a engenharia reversa seria possível descobrir dados sensíveis da indústria.

Durante a avaliação final com os envolvidos da empresa, foi destacado que o programador não sofre cobrança para reduzir os níveis de estoque acabado e que, por consequência, é esperado que tome decisões mais conservadoras no sentido de produzir além do necessário, contribuindo para elevar os custos de capital de manutenção de inventário. Isto se deve à cultura organizacional e solidez financeira da empresa.

No quesito financeiro, a empresa opera exclusivamente com capital próprio e, desta forma, não inclui no custo do produto a remuneração do capital envolvido na operação. Entretanto, no método de avaliação utilizado pela Controladoria, o resultado da empresa é comparado com outros investimentos, ponderando os riscos da atividade empresarial, de forma que deve retornar este custo de capital próprio mais o prêmio de risco. Além disso, o aspecto do custo de capital é majoritariamente decorrente do valor investido em maquinário, uma vez que as injetoras são todas importadas e de alta tecnologia, deixando em segundo plano os custos de estoque. Mesmo que os custos de estoque fossem considerados relevantes pela alta gestão da empresa, ainda assim o custo dos insumos supera, com larga vantagem, o custo do estoque acabado.

No aspecto da cultura organizacional é perceptível a valorização e busca do atendimento ao cliente em nível de excelência. Os gestores são cobrados (e com as cobranças potencializadas) se algum cliente não tem o seu pedido atendido, mas não reportaram situações em que o estoque de produto acabado tenha dado causa a algum tipo de ação por parte da direção. Como o modelo considera que o não atendimento de um pedido gera a penalidade pelo lucro cessante daquele item que,

em teoria, deixou de ser recebido pela empresa, numa situação em que o estoque não seja suficiente para atender a demanda é, potencialmente, uma situação geradora de conflito. Neste caso, pode-se atribuir ao modelo um fator multiplicativo para o lucro cessante ou, no limite, assumir a falta como um custo infinito. Isto forçaria o modelo a buscar, de todas as formas, atender à demanda.

6. CONCLUSÕES

Este trabalho surgiu com base na identificação de um problema real e frequente nas indústrias de diversos segmentos, oriundo da dificuldade de se realizar a programação de produção de forma mais eficaz na gestão dos custos, utilizando um modelo matemático capaz de considerar a disponibilidade de componentes, capacidade de armazenagem e restrições de mão de obra. Com o aprofundamento do estudo do problema, percebeu-se que a proposição de um artefato poderia sanar a dificuldade rotineira na programação da produção. A aplicação da metodologia *Design Science Research* propiciou obter a robustez científica desejada, bem como tornou mais claras as etapas da pesquisa para alcançar os objetivos.

Dentro da revisão bibliográfica, esta pesquisa analisou os conceitos relacionados à programação da produção, pesquisa operacional e os modelos existentes na literatura que tratam das restrições habitualmente encontradas em empresas industriais de único estágio, etapa que resultou na conclusão do primeiro objetivo específico. Ao verificar os modelos já existentes na literatura, foi possível encontrar a modelagem das restrições de disponibilidade de componentes. Já a restrição de capacidade de armazenagem e mão de obra foram desenvolvidas pelo próprio autor e agregadas ao modelo, cumprindo o segundo objetivo específico proposto.

Com o conhecimento teórico organizado e tendo em vista que o problema recorrente, qual seja a dificuldade de avaliar todas as combinações que envolvem a programação da produção, de forma que seja possível entregar o plano de menor custo, desenvolveu-se um artefato capaz de realizar esta tarefa de forma autônoma. O artefato foi projetado para atender os requisitos de eficácia na solução bem como eficiência no tempo computacional para se chegar na solução viável, muito embora esta não seja a solução ótima. O artefato foi, então, programado em linguagem *Python*, utilizando a biblioteca PuLP e o *solver* Gurobi, cumprindo o terceiro objetivo específico.

Durante a execução desta pesquisa foi escolhido o caso de uma empresa de descartáveis plásticos para aplicar o modelo. A disponibilidade de dados, por parte da empresa, foi ampla e irrestrita, o que colaborou muito para o avanço durante as etapas de aplicação e validação do modelo. Mas apesar do acesso irrestrito, a organização dos dados requereu muitas horas de trabalho, uma vez que a base de

dados não estava, necessariamente, organizada conforme o modelo foi construído. Algumas conversões precisaram ser realizadas, além da busca de dados que não estavam no histórico recente da empresa. A etapa descrita marcou a conclusão do quarto objetivo específico.

Ao longo da programação em linguagem computacional, a maior dificuldade foi obter suporte da biblioteca PuLP, que foi construída e é aperfeiçoada pela comunidade de interessados em programação linear e não possui profissionais com dedicação exclusiva para suas melhorias. Nas vezes em que foram necessárias respostas para as dificuldades e que não estavam previamente esclarecidas nos fóruns de discussão, o lapso temporal foi impactante para o cronograma de desenvolvimento. Por outro lado, a documentação da biblioteca e os exemplos previamente existentes foram fonte importante para a geração do conhecimento e, sem dúvida, servem de base sólida para qualquer pessoa que tenha desejo de iniciar uma pesquisa nesta área de programação linear.

O artefato superou os critérios de avaliação previamente definidos, validados através de aplicação prática na indústria, e atingiu o objetivo geral da presente pesquisa. O modelo funciona de forma autônoma, buscando na base de dados da empresa as informações necessárias e, conforme demonstram os resultados, encontrou a solução com *gap* de 1% em relação ao resultado ótimo em 2h de tempo computacional. Ao final, o programa computacional projetou na tela o tamanho dos lotes de cada produto e como estes devem ser sequenciados nos recursos produtivos, cumprindo o objetivo geral da presente pesquisa.

6.1. LIMITAÇÕES

Dentre as limitações da pesquisa, cabe destacar que o estudo de caso para validar o artefato foi realizado em uma única empresa. A validação em outras empresas poderia gerar *insights* e oportunidades de melhoria no código computacional, além de testar os resultados em novos negócios. Entretanto, caso seja utilizada uma empresa com base de dados diferente, provavelmente seja necessária a adaptação do bloco de código que importa as informações para o modelo.

O modelo negligencia a limitação de trocas de molde, que existe na indústria objeto de estudo. Tal restrição não é de difícil implantação, sendo necessário incluir uma variável binária que sinaliza quando há troca de molde na combinação de dois produtos. A soma desta variável binária no macroperíodo não deveria ser superior

a três, para o caso específico. Optou-se por suprimi-la dada sua absoluta especificidade para esta indústria de descartáveis plásticos, sem potencial utilidade quando o modelo for usado em outras indústrias.

6.2. TRABALHOS FUTUROS

A título de sugestão para trabalhos futuros, a validação da redução de custos pode ser executada com prazo maior, recomendando-se 360 macroperíodos, para englobar a sazonalidade de demanda presente na indústria em que foi aplicado o modelo. A validação do modelo em outros tipos de indústria também é recomendada pelo autor.

Outra sugestão é implantar, dentro da solução do modelo, o mesmo algoritmo utilizado atualmente e que é válido para diversas indústrias:

1. Programar os itens que estão programados pela logística para carregamento e que não possuem estoque disponível;
2. Programar os itens que estão em carteira (pedidos firmes) e que não possuem disponibilidade de estoque. Na prática, estes itens irão gerar ruptura tão logo sejam enviados para carregamento dos veículos de transporte;
3. Com a capacidade remanescente, programar conforme a política de reposição de estoque que gere o menor custo;

Desta forma, acredita-se que o modelo poderá alcançar a solução de forma mais rápida, pois haveria uma divisão de um grande problema em três problemas menores, muito embora não seja possível garantir que, com essas regras, seria gerada a programação global de menor custo. Além disso, num cenário de demanda muito acima da capacidade, é esperado que o modelo entregue excessivas trocas de produtos em cada macro período.

Ainda dentro do caso específico dos descartáveis plásticos, recomenda-se aplicar a restrição de três trocas de molde por dia, seguindo a lógica exposta na seção anterior e realizando testes para aprovar ou reprovar a hipótese de que esta restrição gera economia para a empresa. Esta restrição é fortíssima na cultura da empresa, muito embora não haja comprovação da sua eficácia.

Por fim, seria muito bem vinda a adaptação para indústrias multi estágio. Um vez que este sistema produtivo engloba a maioria dos negócios, o modelo poderia ser utilizado numa quantidade maior de indústrias. Reconhecidamente, o modelo de

estágio único é limitante na utilização prática, no sentido de restringir o número de empresa abrangidas pela solução, muito embora não prejudique o resultado desta pesquisa.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABEDINNIA, H. et al. Machine scheduling problems in production: A tertiary study. **Computers and Industrial Engineering**, v. 111, p. 403–416, 1 set. 2017.
- AGHEZZAF, E. H.; ARTIBA, A.; ELMAGHRABY, S. E. **Hybrid flowshops: an LP based heuristic for the planning level problems**. IEEE Symposium on Emerging Technologies & Factory Automation. **Anais...**1995.
- ALIMIAN, M. et al. Solving a parallel-line capacitated lot-sizing and scheduling problem with sequence-dependent setup time/cost and preventive maintenance by a rolling horizon method. **Computers & Industrial Engineering**, v. 168, p. 108041, 1 jun. 2022.
- ALLAHVERDI, A. et al. A survey of scheduling problems with setup times or costs. **European Journal of Operational Research**, v. 187, n. 3, p. 985–1032, jun. 2008.
- ALLAHVERDI, A.; GUPTA, J. N. D.; ALDOWAISAN, T. A review of scheduling research involving setup considerations. **Omega**, v. 27, n. 2, p. 219–239, abr. 1999.
- ALLAHVERDI, A.; SOROUSH, H. M. The significance of reducing setup times/setup costs. **European Journal of Operational Research**, v. 187, n. 3, p. 978–984, 2008.
- ALMEDER, C. et al. Lead time considerations for the multi-level capacitated lot-sizing problem. **European Journal of Operational Research**, v. 241, n. 3, p. 727–738, 16 mar. 2015.
- ALTURKI, A.; GABLE, G. G.; BANDARA, W. A Design Science Research Roadmap. Em: Milwaukee: Springer-Verlag, 2011. v. 6629 LNCSp. 107–123.
- ALVES DE OLIVEIRA SOLER, W.; SANTOS, M. O.; RANGEL, S. OPTIMIZATION MODELS FOR A LOT SIZING AND SCHEDULING PROBLEM ON PARALLEL PRODUCTION LINES THAT SHARE SCARCE RESOURCES. **Res**, v. 55, p. 1949–1970, 2021.
- ALVES, F. F. et al. Approaches for the joint resolution of lot-sizing and scheduling with infeasibilities occurrences. **Computers and Industrial Engineering**, v. 155, 1 maio 2021.

APPOLINÁRIO, F. **Dicionário de metodologia científica: um guia para a produção do conhecimento científico**. São Paulo: Atlas, 2004.

ARENALES, M. et al. **Pesquisa Operacional**. 2ª ed. São Paulo: Elsevier, 2015.

BARDIN, L. **Análise de conteúdo**. 1. ed. São Paulo: Atlas, 2016.

BILLINGTON, P. J.; MCCLAIN, J. O.; THOMAS, L. J. Mathematical Programming Approaches to Capacity-Constrained MRP Systems: Review, Formulation and Problem Reduction. <http://dx.doi.org/10.1287/mnsc.29.10.1126>, v. 39, n. 10, p. 1126–1141, 1 out. 1983.

BITRAN, G. R.; YANASSE, H. H. Computational Complexity of the Capacitated Lot Size Problem. <http://dx.doi.org/10.1287/mnsc.28.10.1174>, v. V 28, n. N 10, p. 1174–1186, 1 out. 1982.

BOOTH, W. C.; COLOMB, G. G.; WILLIAMS, J. M. **The Craft of Research**. 3ª ed. ed. Chicago: The University of Chicago Press, 1997. v. 48

BURGESS, T.; PIDD, M. **Computer Simulation in Management Science, 4th Edition**. 5ª ed. ed. Lancaster: John Wiley & Sons, 1999. v. 50

BUSCHKÜHL, L. et al. Dynamic capacitated lot-sizing problems: a classification and review of solution approaches. **OR Spectrum** 2008 32:2, v. 32, n. 2, p. 231–261, 22 out. 2008.

CHENG, T. C. E.; GUPTA, J. N. D.; WANG, G. A review of flowshop scheduling research with setup times. **Production and Operations Management**, v. 9, n. 3, p. 262–282, 2000.

CHOI, C. Q. **The Beating Heart of the World's First Exascale Supercomputer - IEEE Spectrum**. Disponível em: <<https://spectrum.ieee.org/frontier-exascale-supercomputer>>. Acesso em: 21 out. 2022.

CLARK, A. R.; CLARK, S. J. Rolling-horizon lot-sizing when set-up times are sequence-dependent. **International Journal of Production Research**, v. 38, n. 10, p. 2287–2307, 14 jul. 2000.

COPIL, K. et al. Simultaneous lotsizing and scheduling problems: a classification and review of models. **OR Spectrum**, v. 39, n. 1, p. 1–64, 1 jan. 2017.

CORRÊA, H. L.; GIANESI, I. G. N.; CAON, M. **Planejamento, programação e controle da produção: MRP II/ERP : conceitos, uso e implantação: base para SAP, Oracle Applications e outros Softwares integrados de gestão**. São Paulo: Atlas, 2014.

CRAIGHEAD, C. W.; MEREDITH, J. Operations management research: evolution and alternative future paths. **International Journal of Operations & Production Management**, v. 28, n. 8, p. 710–726, 18 jul. 2008.

DANTZIG, G. B.; WOLFE, P. Decomposition Principle for Linear Programs. <https://doi.org/10.1287/opre.8.1.101>, v. 8, n. 1, p. 101–111, 1 fev. 1960.

DE MATTA, R.; GUIGNARD, M. Studying the effects of production loss due to setup in dynamic production scheduling. **Eur J Oper Res**, v. 72, n. 1, p. 62–73, 6 jan. 1994.

DICICCO-BLOOM, B.; CRABTREE, B. F. The qualitative research interview. **Medical Education**, v. 40, n. 4, p. 314–321, 1 abr. 2006.

DRESCH, A.; LACERDA, D. P.; ANTUNES JUNIOR, J. A. V. **Design science research: método de pesquisa para avanço da ciência e tecnologia**. Porto Alegre: Bookman, 2015.

DRESCH, A.; LACERDA, D. P.; CAUCHICK MIGUEL, P. A. A Distinctive Analysis of Case Study, Action Research and Design Science Research. **Review of Business Management**, v. 17, n. 56, p. 1116–1133, 24 nov. 2015.

DREXL, A.; HAASE, K. Proportional lotsizing and scheduling. **International Journal of Production Economics**, v. 40, n. 1, p. 73–87, 1 jun. 1995.

FERREIRA, D. et al. Single-stage formulations for synchronised two-stage lot sizing and scheduling in soft drink production. **International Journal of Production Economics**, v. 136, n. 2, p. 255–265, abr. 2012.

FERREIRA, D.; MORABITO, R.; RANGEL, S. Relax and fix heuristics to solve one-stage one-machine lot-scheduling models for small-scale soft drink plants. **Computers and Operations Research**, v. 37, n. 4, p. 684–691, abr. 2010.

FLEISCHMANN, B. The discrete lot-sizing and scheduling problem. **European Journal of Operational Research**, v. 44, n. 3, p. 337–348, 15 fev. 1990.

FLEISCHMANN, B.; MEYR, H. The general lotsizing and scheduling problem. **Operations-Research-Spektrum 1997 19:1**, v. 19, n. 1, p. 11–21, 1 mar. 1997.

FOCACCI, F.; LABURTHE, F.; LODI, A. Local Search and Constraint Programming. **Handbook of Metaheuristics**, p. 369–403, 2 fev. 2003.

FRANÇA, P. M. et al. A tabu search heuristic for the multiprocessor scheduling problem with sequence dependent setup times. **International Journal of Production Economics**, v. 43, n. 2–3, p. 79–89, jun. 1996.

FU, M. et al. Batch production scheduling for semiconductor back-end operations. **IEEE Transactions on Semiconductor Manufacturing**, v. 24, n. 2, p. 249–260, 2011.

GAREY, M. R.; JOHNSON, D. S. **Computers and Intractability: A Guide to the Theory of NP-Completeness (Series of Books in the Mathematical Sciences)**. [s.l.: s.n.].

GENDREAU, M. An Introduction to Tabu Search. **Handbook of Metaheuristics**, p. 37–54, 2 fev. 2006.

GHOSH DASTIDAR, S.; NAGI, R. Scheduling injection molding operations with multiple resource constraints and sequence dependent setup times and costs. **Computers & Operations Research**, v. 32, n. 11, p. 2987–3005, 1 nov. 2005.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002.

GLOVER, F. Tabu Search—Part I. **ORSA Journal on Computing**, v. 1, n. 3, p. 190–206, 1989.

GLOVER, F.; KOCHENBERGER, G. A. **Handbook of Metaheuristics**. Boston, MA: Springer US, 2003. v. 57

GUIGNARD, M.; KIM, S. Lagrangean decomposition: A model yielding stronger lagrangean bounds. **Mathematical Programming** 1987 **39:2**, v. 39, n. 2, p. 215–228, 1 jun. 1987.

GUIMARÃES, L.; KLABJAN, D.; ALMADA-LOBO, B. Modeling lotsizing and scheduling problems with sequence dependent setups. **European Journal of Operational Research**, v. 239, n. 3, p. 644–662, 16 dez. 2014.

HAASE, K. Capacitated lot-sizing with sequence dependent setup costs. **Operations-Research-Spektrum** 1996 **18:1**, v. 18, n. 1, p. 51–59, 1 mar. 1996.

HARRIS, F. W. How many parts to make at once. **Factory, The Magazine of Management**, p. 135–136, fev. 1913.

HENDERSON, D.; JACOBSON, S. H.; JOHNSON, A. W. The Theory and Practice of Simulated Annealing. **Handbook of Metaheuristics**, p. 287–319, 2 fev. 2003.

HUANG, J. et al. Genetic algorithm for rotary machine scheduling with dependent processing times. **Journal of Intelligent Manufacturing** 2011 **23:5**, v. 23, n. 5, p. 1931–1948, 2 mar. 2011.

HUGHES, T. et al. Scholarship that matters: Academic-practitioner engagement in business and management. **Academy of Management Learning and Education**, v. 10, n. 1, p. 40–57, set. 2011.

IBARRA-ROJAS, O. J. et al. A decomposition approach for the piece–mold–machine manufacturing problem. **International Journal of Production Economics**, v. 134, n. 1, p. 255–261, 1 nov. 2011.

INTERNATIONAL BUSINESS MACHINES. **Decision variables - IBM Documentation**. Disponível em: <<https://www.ibm.com/docs/en/icos/12.9.0?topic=types-decision-variables>>. Acesso em: 18 set. 2022.

INTERNATIONAL FEDERATION OF OPERATIONAL RESEARCH SOCIETIES. **What is OR – IFORS, Operations Research, Operational Research, Management Science**. Disponível em: <<https://www.ifors.org/what-is-or/>>. Acesso em: 22 set. 2022.

IRWAN, H. et al. A REVIEW OF INTEGRATION MODEL OF LOT-SIZING-SCHEDULING PROBLEM. **Malaysian Construction Research Journal**, v. 17, n. Special issue 3, p. 160–174, 2022.

JÓZEFOWSKA, J.; ZIMNIAK, A. Optimization tool for short-term production planning and scheduling. **International Journal of Production Economics**, v. 112, n. 1, p. 109–120, 1 mar. 2008.

KARMAKAR, U. S.; SCHRAGE, L. The Deterministic Dynamic Product Cycling Problem. <https://doi.org/10.1287/opre.33.2.326>, v. 33, n. 2, p. 326–345, 1 abr. 1985.

KOULAMAS, C. A new constructive heuristic for the flowshop scheduling problem. **European Journal of Operational Research**, v. 105, n. 1, p. 66–71, 16 fev. 1998.

LACERDA, D. P. et al. Design Science Research: Método de pesquisa para a engenharia de produção. **Gestão e Produção**, v. 20, n. 4, p. 741–761, 2013.

MARCH, S. T.; SMITH, G. F. Design and natural science research on information technology. **Decision Support Systems**, v. 15, n. 4, p. 251–266, set. 1995.

MASOUD, S. A.; MASON, S. J. Integrated cost optimization in a two-stage, automotive supply chain. **Computers & Operations Research**, v. 67, p. 1–11, 1 mar. 2016.

MATOUŠEK, J.; GÄRTNER, B. **Understanding and Using Linear Programming**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

MEYR, H. Simultaneous lotsizing and scheduling on parallel machines. **European Journal of Operational Research**, v. 139, n. 2, p. 277–292, 1 jun. 2002.

MEYR, H.; MANN, M. A decomposition approach for the General Lotsizing and Scheduling Problem for Parallel production Lines. **European Journal of Operational Research**, v. 229, n. 3, p. 718–731, 16 set. 2013.

MICHALEWICZ, Z.; FOGEL, D. B. **How to Solve It: Modern Heuristics**. 2nd. ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004.

MOHAMMADI, M.; ESMAELIAN, M.; ATIGHEHCHIAN, A. Design of mathematical models for the integration of purchase and production lot-sizing and scheduling problems under demand uncertainty. **Applied Mathematical Modelling**, v. 84, p. 1–18, 1 ago. 2020.

MULA, J. et al. A capacitated lot-sizing model with sequence-dependent setups, parallel machines and bi-part injection moulding. **Applied Mathematical Modelling**, v. 100, p. 805–820, 1 dez. 2021.

NAKANO, D. Métodos de Pesquisa Adotados na Engenharia de Produção e Gestão de Operações. **Metodologia de pesquisa em engenharia de produção e gestão de operações**, p. 145–163, 2010.

OAK RIDGE NATIONAL LABORATORY. **Frontier supercomputer debuts as world's fastest, breaking exascale barrier | ORNL**. Disponível em: <<https://www.ornl.gov/news/frontier-supercomputer-debuts-worlds-fastest-breaking-exascale-barrier>>. Acesso em: 21 out. 2022.

OSTWALD, P. F. **Engineering Cost Estimating**. 3ª ed. Englewood Cliffs: Prentice Hall, 1992.

PAGLIARUSSI, M. **Contribuições para a otimização da programação da produção de bebidas a base de frutas**. São Carlos, SP: UFSCar, 2013.

PEIRCE, C. S. Pragmatism and Abduction. **Collected Papers of Charles Sanders Peirce, Volumes V and VI: Pragmatism and Pragmaticism and Scientific Metaphysics**, p. 180–212, 1978.

POPPER, K. The logic of scientific discovery. **The Logic of Scientific Discovery**, p. 1–516, 1 jan. 2005.

QIU, Y. et al. Formulations and branch-and-cut algorithms for multi-product multi-vehicle production routing problems with startup cost. **Expert Systems with Applications**, v. 98, p. 1–10, 15 maio 2018.

QUADT, D. **Lot-sizing and scheduling for flexible flow lines**. Berlin: Springer-Verlag, 2004.

QUADT, D.; KUHN, H. Conceptual framework for lot-sizing and scheduling of flexible flow lines. **International Journal of Production Research**, v. 43, n. 11, p. 2291–2308, 1 jun. 2005.

RÍOS-SOLÍS, Y. Á. et al. A heuristic based on mathematical programming for a lot-sizing and scheduling problem in mold-injection production. **European Journal of Operational Research**, v. 284, n. 3, p. 861–873, 1 ago. 2020.

SAUNDERS, M. A.; LEWIS, P. **Research Methods for Business Students Eights Edition Research Methods for Business Students**. 4^a ed. ed. Londres: Pearson Education Limited, 2019.

SAUNDERS, M.; LEWIS, P.; THORNHILL, A. **Research methods for business students**. 4. ed. Londres: Pearson Education Limited, 2007.

SEEANNER, F. **Multi-Stage Simultaneous Lot-Sizing and Scheduling: Planning of Flow Lines with Shifting Bottlenecks**. [s.l: s.n.].

SEEANNER, F.; MEYR, H. Multi-stage simultaneous lot-sizing and scheduling for flow line production. **OR Spectrum**, v. 35, n. 1, p. 33–73, 2013.

SHARMA, P.; JAIN, A. A review on job shop scheduling with setup times. **Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture**, v. 230, n. 3, p. 517–533, 19 mar. 2016.

SILVA, E. L. DA; MENEZES, E. M. Metodologia da Pesquisa e Elaboração de Dissertação. 2005.

SIMON, H. A. **The Sciences of the Artificial, reissue of the third edition with a new introduction by John Laird**. 3^aed ed. Cambridge: MIT Press, 2019.

SLACK, N.; BRANDON-JONES, A.; JOHNSTON, R. **Administração da Produção**. 8. ed. São Paulo: Atlas, 2018.

SOLER, W. A. O.; SANTOS, M. O.; AKARTUNALI, K. MIP approaches for a lot sizing and scheduling problem on multiple production lines with scarce resources, temporary workstations, and perishable products. **Journal of the Operational Research Society**, v. 72, n. 8, p. 1691–1706, 2021.

STADTLER, H. Supply chain management and advanced planning - Basics, overview and challenges. **European Journal of Operational Research**, v. 163, n. 3, p. 575–588, set. 2005.

STAGGEMEIER, A. T.; CLARK, A. R. A survey of lot-sizing and scheduling models. **23rd Symposium of the Brazilian Operational Research Society (SOBRAPO)**, p. 8, 1 nov. 2001.

SUERIE, C.; STADTLER, H. The capacitated lot-sizing problem with linked lot sizes. **Management Science**, v. 49, n. 8, p. 1039–1054, 2003.

TANG, C. S. Scheduling batches on parallel machines with major and minor set-ups. **European Journal of Operational Research**, v. 46, n. 1, p. 28–37, set. 1990.

TAŞ, D. et al. A capacitated lot sizing problem with stochastic setup times and overtime. **European Journal of Operational Research**, v. 273, n. 1, p. 146–159, 16 fev. 2019.

TOSO, E. A. V. **Dimensionamento e Sequenciamento de Lotes de Produção na Indústria de Suplementos para Nutrição Animal**. São Carlos: UFSCar, 2008.

TUBINO, D. F. **Manual de planejamento e controle da produção**. São Paulo: Atlas, 2000.

TUBINO, D. F. **Planejamento e controle da produção: teoria e prática**. 2. ed. São Paulo: Atlas, 2009.

ULBRICHT, G.; VOLPI, N. M. P. Um Problema De Dimensionamento E Sequenciamento De Lotes Em Máquinas Paralelas Com Utilização De Capacidade Extra. **XLVII Simposio Brasileiro de Pesquisa Operacional SBPO.**, 2015.

WAGNER, H. M.; WHITIN, T. M. Dynamic Version of the Economic Lot Size Model. <http://dx.doi.org/10.1287/mnsc.5.1.89>, v. 5, n. 1, p. 89–96, 1 out. 1958.

WANG, Y. M. et al. A two-stage genetic algorithm for large size job shop scheduling problems. **International Journal of Advanced Manufacturing Technology**, v. 39, n. 7–8, p. 813–820, 2008.

WHITLEY, D. A genetic algorithm tutorial. **Statistics and Computing** 1994 4:2, v. 4, n. 2, p. 65–85, jun. 1994.

WILLIAMS, H. P. Integer programming. **International Series in Operations Research and Management Science**, v. 130, p. 25–70, 2009.

WÖRBELAUER, M.; MEYR, H.; ALMADA-LOBO, B. Simultaneous lotsizing and scheduling considering secondary resources: a general model, literature review and classification. **OR Spectrum**, v. 41, n. 1, p. 1–43, 11 mar. 2019.

WU, Y. et al. Learning Improvement Heuristics for Solving Routing Problems. **IEEE Transactions on Neural Networks and Learning Systems**, 12 dez. 2019.