UNIVERSIDADE DO VALE DO RIO DOS SINOS (UNISINOS)
UNIDADE ACADÊMICA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA
DE PRODUÇÃO E SISTEMAS
NÍVEL MESTRADO

DANIEL LUIS KLEIN

SEQUENCIAMENTO INTELIGENTE:
Uma Proposta de Integração entre o Aprendizado por Reforço e o Tambor-Pulmão-Corda

São Leopoldo
2024

DANIEL LUIS KLEIN

SEQUENCIAMENTO INTELIGENTE:
Uma Proposta de Integração entre o Aprendizado por Reforço e o Tambor-Pulmão-Corda

Dissertação apresentada como requisito para obtenção do título de Mestre em Engenharia de Produção e Sistemas, pelo Programa de Pós-Graduação em Engenharia de Produção e Sistemas da Universidade do Vale do Rio dos Sinos (UNISINOS).

Orientador: Prof. Dr. Leandro Gauss
Coorientador: Prof. Dr. Daniel Pacheco Lacerda

São Leopoldo
2024

K64s Klein, Daniel Luis.
Sequenciamento inteligente : uma proposta de integração entre o aprendizado por reforço e o tambor-pulmão-corda / Daniel Luis Klein. – 2024.
107 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Engenharia de Produção e Sistemas, 2024.
"Orientador: Prof. Dr. Leandro Gauss
Coorientador: Prof. Dr. Daniel Pacheco Lacerda".

1. Teoria das restrições (Administração). 2. Aprendizado por reforço. 3. Administração da produção. 4. Planejamento da produção. 5. Controle de produção. 6. Manufatura. I. Título.

CDU 658.5

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Amanda Schuster Ditbenner – CRB 10/2517)

AGRADECIMENTOS À CAPES

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

SUMÁRIO

1 INTRODUÇÃO[1]

Este capítulo destaca o papel do sequenciamento avançado de produção na sincronização entre demanda e capacidade, bem como os desafios enfrentados para a sua realização em processos estocásticos. A seção ainda enfatiza o aprendizado por reforço como uma abordagem promissora para superar esses desafios e levanta a possibilidade de explorar a integração dessa técnica com a Teoria das Restrições para aprimorar o sequenciamento de produção sob o ponto de vista pragmático e teórico.

1.1 CONTEXTUALIZAÇÃO E QUESTÃO DE PESQUISA

O sequenciamento de produção desempenha um papel fundamental na gestão dos sistemas de produção, determinando o volume e a ordem em que as tarefas são executadas, de maneira a maximizar a utilização de recursos e minimizar o tempo de atravessamento (Hopp; Spearman, 2008; Pinedo, 2016-). Essa atividade não apenas influencia diretamente a produtividade, mas também afeta a eficiência operacional e a capacidade de resposta ao mercado (Zhou *et al.*, 2022). Contudo, os sistemas de produção enfrentam desafios significativos devido à variabilidade, tanto interna quanto externa (Harjunkoski *et al.*, 2014; Hopp; Spearman, 2008). A variabilidade interna se refere a flutuações nos tempos de processamento das máquinas, à disponibilidade de materiais e à capacidade de trabalho (Benkel; Jørnsten; Leisten, 2016; Davis *et al.*, 2012), enquanto a variabilidade externa inclui mudanças nas demandas do mercado, interrupções na cadeia de suprimentos e variações ambientais (Oztemel; Gursev, 2020).

Para enfrentar essas questões, pode-se utilizar o Advanced Planning and Scheduling (APS), uma abordagem avançada que permite sincronizar eficientemente a capacidade produtiva com a demanda, melhorando a operação e reduzindo custos (Chen; Ji, 2007; Márquez; Ribeiro, 2022). Na manufatura, o uso de APS reduziu o tempo de sequenciamento e impediu o planejamento em excesso (Chen *et al.*, 2017), além de aprimorar a utilização de robôs autônomos ao reduzir a quantidade de pedidos em atraso (Liang; Zhou; Jiang, 2024). Apesar dos avanços proporcionados pelo APS, alcançar resultados ideais ainda é complexo devido ao

[1] A estrutura da dissertação foi desenvolvida de acordo com a tese de Gauss (2023) e o Instituto de Tecnologia de Eindhoven.

processamento computacional intensivo e à sensibilidade aos fatores estocásticos inerentes aos sistemas de produção (Missbauer; Uzsoy, 2022).

Uma solução potencial para os desafios enfrentados pelo APS surge com o aprendizado por reforço (RL), um paradigma da inteligência artificial que permite aos sistemas aprenderem com base na interação com o ambiente (Sutton; Barto, 2018). O RL demonstra capacidade para lidar com processos estocásticos, adaptando-se dinamicamente às mudanças e incertezas inerentes aos sistemas produtivos (Esteso *et al.*, 2023; Panzer; Bender, 2022). Recentemente, o RL tem sido explorado em estudos de sequenciamento em ambientes de manufatura. Por exemplo, um agente de RL foi utilizado para sequenciar um job-shop flexível, mantendo-se efetivo mesmo com a reconfiguração do ambiente (Liu; Piplani; Toro, 2022). Outro estudo utilizou múltiplos agentes para o sequenciamento adaptativo na produção de semicondutores, reduzindo o retrabalho e elevando a produtividade (Sakr *et al.*, 2023). O RL também foi aplicado a um ambiente de manufatura flexível, ultrapassando o desempenho de heurísticas fixas de sequenciamento (Shiue; Lee; Su, 2020). No entanto, muitos desses estudos combinam RL com abordagens de sequenciamento que, em certa medida, negligenciam a variabilidade dos sistemas.

A Teoria das Restrições (TOC), especialmente o Drum-Buffer-Rope (DBR), oferece uma abordagem que reconhece e gerencia eficazmente a variabilidade nos sistemas de produção (Goldratt, 2006). O DBR protege a restrição que determina o desempenho do sistema por meio do uso de pulmões, garantindo uma produção mais estável e eficiente (Schragenheim, 2010). Em uma linha automotiva de alta variabilidade, por exemplo, o DBR elevou a produtividade ao reduzir as filas nos recursos e o nível de ordens em produção (Golmohammadi, 2015). Um setor de remanufatura elevou a utilização dos recursos com a introdução de pulmões de tempo antes da restrição (Ma; Chen, 2009), enquanto o DBR elevou a produtividade e reduziu o lead-time em outra fábrica automotiva (Dohale; Ambilkar; Bilolikar, 2021).

Apesar das sinergias potenciais existentes entre TOC e RL para melhorar o desempenho do sequenciamento em processos estocásticos, a integração desses dois campos de estudo ainda não foi explorada. A hipótese central desta pesquisa é que a combinação dessas abordagens pode oferecer benefícios substanciais que vão além dos aspectos puramente operacionais. Portanto, esta dissertação tem como objetivo investigar as seguintes questões de pesquisa: (i) É possível integrar

TOC e RL no sequenciamento de produção de processos estocásticos? (ii) Em caso afirmativo, quais são os benefícios pragmáticos e teóricos dessa integração?

1.2 ORGANIZAÇÃO DA PESQUISA

Para responder às questões de pesquisa, esta dissertação está organizada da seguinte maneira: O Capítulo 2 descreve os procedimentos metodológicos adotados e fornece uma visão geral dos demais capítulos. O Capítulo 3 apresenta uma revisão sistemática da literatura, usada para desenvolver um modelo conceitual que explora possíveis sinergias entre a TOC e o RL. Para tanto, são revisados e analisados 50 artigos, com destaque às variáveis utilizadas na modelagem dos agentes de RL e nas simulações baseadas no modelo Drum-Buffer-Rope (DBR). O Capítulo 4 modela e avalia dois modelos de sequenciamento: o primeiro segue as premissas do DBR, enquanto o segundo integra conceitos da TOC em um agente de RL. A análise compara o desempenho produtivo dos modelos e examina o comportamento do agente de RL para entender e explicar suas decisões. Por fim, o Capítulo 5 discute os resultados encontrados nos Capítulos 3 e 4, e o Capítulo 6 destaca as implicações teóricas e práticas, além de indicar limitações e sugestões para pesquisas futuras.

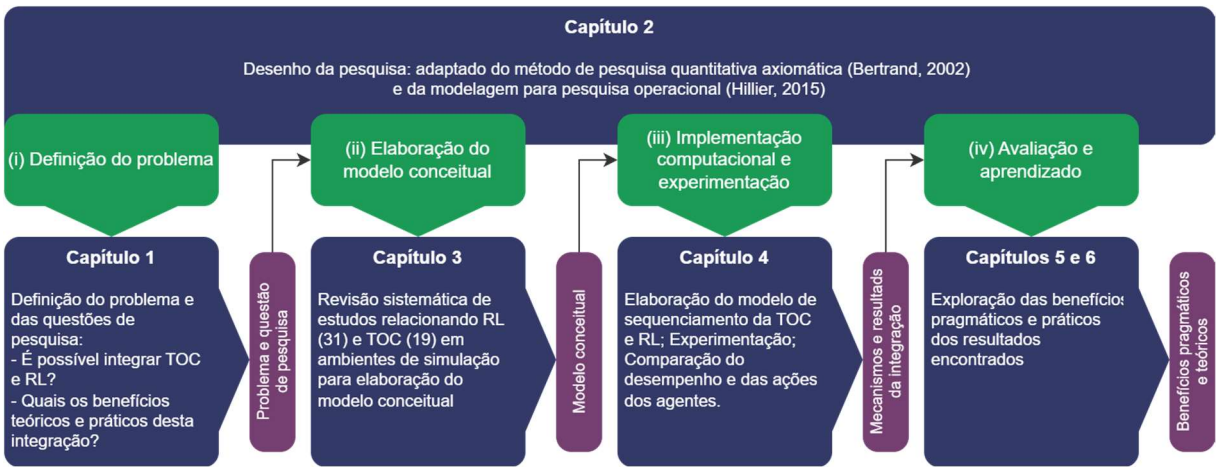
2 DESENHO DA PESQUISA

Este capítulo apresenta o desenho da pesquisa, detalhando a lógica e os procedimentos adotados na condução do estudo. Além disso, explica a interação entre os capítulos da dissertação, ilustrando o escopo de cada um e suas respectivas entradas e saídas.

2.1 MODELAGEM QUANTITATIVA

Este trabalho tem como objetivo explorar a integração da TOC com o RL para o sequenciamento da produção, utilizando modelos de simulação. Para alcançar tal objetivo, adotou-se a modelagem quantitativa como método de pesquisa. A modelagem quantitativa utiliza técnicas matemáticas e estatísticas para representar e analisar computacionalmente sistemas e processos, permitindo simular diferentes cenários e avaliar seus impactos (Morabito; Pureza, 2012). Ela possibilita a interação entre variáveis e permite abordar problemas complexos para a extração de informações e tomadas de decisões (Dresch; Lacerda; Antunes Junior, 2015; Morabito; Pureza, 2012). Assim, o trabalho utiliza de maneira conjunta as abordagens propostas por Bertrand e Fransoo (2002) e Hillier e Liberman (2015), contemplando quatro etapas: (i) Definição do problema; (ii) Elaboração do modelo conceitual; (iii) Implementação computacional e experimentação; e (iv) Avaliação e aprendizado. Cada etapa corresponde a um capítulo desta dissertação, seguindo a sequência lógica da pesquisa. A próxima seção apresenta um resumo de cada etapa, e a Figura 1 ilustra o relacionamento entre elas.

Figura 1 – Desenho da pesquisa



Fonte: Adaptado de Bertrand e Fransoo (2002) e Hillier e Lieberman (2015)

2.1.1 Definição do Problema (Capítulo 1)

O problema de pesquisa é definido como sequenciamento da produção em ambientes complexos e de elevada variabilidade. Como o RL tem capacidade de lidar com cenários complexos (Sutton; Barto, 2018) e a TOC tem proposições para absorver a variabilidade, as questões de pesquisa são definidas por: (i) É possível integrar TOC e RL no sequenciamento de produção de processos estocásticos? (ii) Em caso afirmativo, quais são os benefícios pragmáticos e teóricos dessa integração? Para responder às questões de pesquisa, a modelagem quantitativa inicia por um modelo conceitual (Bertrand; Fransoo, 2002; Hillier; Lieberman, 2015), sendo este introduzido no próximo capítulo desta dissertação.

2.1.2 Elaboração do modelo conceitual (Capítulo 3)

Para a modelagem conceitual, conduziu-se uma revisão sistemática da literatura seguindo o método LGT (*Literature Grounded Theory* – Literatura Fundamentada na Literatura) (Ermel *et al.*, 2021-). Para tanto, foram revisados 19 artigos de TOC e 31 de RL, estudos abordando sequenciamento da produção utilizando simulação por eventos discretos. Os objetivos e as variáveis das simulações da TOC foram mapeados, assim como as variáveis utilizadas na modelagem dos elementos do RL. A análise temática revelou a força e a fragilidade dos grupos, permitindo associar os elementos do RL com os conceitos e métodos da TOC. Com base no resultado, elaborou-se um modelo conceitual com as possíveis sinergias entre as áreas. A fim de validá-las, o próximo capítulo implementa os conhecimentos adquiridos e testa as proposições em um ambiente de simulação.

2.1.3 Implementação computacional e experimentação (Capítulo 4)

O modelo conceitual é utilizado como referência para simulações. Para isso, elaborou-se um modelo de simulação fabril sobre o paradigma da simulação por eventos discretos, o qual foi selecionado da literatura do DBR e modificado para elevar sua variabilidade (Machado *et al.*, 2023). O modelo de simulação foi utilizado como base para elaboração de dois modelos de sequenciamento, sendo que o primeiro foi implementado conforme o DRB (Goldratt, 2006; Schragenheim, 2010) e

o segundo de acordo com os conceitos da TOC em um agente de RL. Ambos foram integrados ao modelo de simulação fabril e submetidos à simulação. Durante a simulação, dados referentes à produtividade e ao comportamento dos sequenciadores foram registrados para análises distintas. Uma análise pragmática comparou o desempenho produtivo por meio de técnicas estatísticas (Law, 2015), evidenciando redução no tempo de atravessamento, no inventário e no nível de ordens em processamento durante o sequenciamento do RL. O comportamento dos sequenciadores foi explorado com técnicas que explicam modelos de aprendizado de máquina (Kuhnle *et al.*, 2022; Rudin *et al.*, 2022), salientando a importância do WIP para a tomada de decisão no sequenciamento. As descobertas revelam aprendizados para ambas as áreas, os quais são discutidos no próximo capítulo.

2.1.4 Avaliação e aprendizado (Capítulos 5 e 6)

O capítulo 5, que aborda a avaliação e as percepções do modelo, resume os resultados das etapas anteriores, discutindo as sinergias do modelo híbrido e seu desempenho à luz dos conceitos da TOC. No capítulo 6, enquanto as implicações teóricas expõem as características do DBR que podem ser aprimoradas e os impactos da introdução de uma teoria de gestão na modelagem do RL, as implicações práticas apresentam os benefícios dessa integração para futuras implementações tecnológicas. As limitações da pesquisa e sugestões para trabalhos futuros finalizam o capítulo.

3 SINERGIAS ENTRE A TEORIA DAS RESTRIÇÕES E O APRENDIZADO POR REFORÇO: UMA REVISÃO SISTEMÁTICA DA LITERATURA

Na manufatura, a competitividade e a complexidade do cenário desafiam as empresas a buscar soluções que elevam a capacidade de análise e de tomada de decisão. As técnicas de aprendizagem de máquina surgem como uma alternativa viável para enfrentar esses desafios. O aprendizado por reforço (RL – Reinforcement Learning) tem sido empregado em atividades de planejamento, controle e sequenciamento da produção. No entanto, essas aplicações desconsideram a natureza estocástica dos processos de manufatura. Nesse contexto, a Teoria das Restrições (TOC - Theory of Constraints) propõe o método Tambor, Pulmão e Corda (DBR – Drum-Buffer-Rope), que utiliza pulmões de tempo para mitigar os efeitos indesejados decorrentes de processos estocásticos. Assim, este capítulo busca identificar e analisar as possíveis sinergias existentes entre o RL e a TOC em ambientes de manufatura por meio de uma revisão sistemática da literatura. Os resultados evidenciam a escassez de estudos que aplicam teorias de gestão a modelos de RL, assim como, a ausência de estudos no campo da TOC que utilizem RL para aprimorar resultados. Em termos de contribuições, são elencados os impactos de se desconsiderar teorias de gestão de operações na modelagem dos agentes de RL. Além disso, apresenta-se um quadro síntese de possíveis sinergias entre a TOC o RL.

Palavras-chave: Teoria das Restrições; Aprendizado por Reforço; Sequenciamento da Produção; Planejamento e Controle da Produção; Manufatura

3.1 INTRODUÇÃO

O sequenciamento da produção tem como objetivo sincronizar a demanda e a capacidade, definindo o volume e a ordem da produção (Hopp; Spearman, 2008; Pinedo, 2016-). A otimização dessa atividade afeta tanto a produtividade quanto a eficiência do sistema, permitindo o direcionamento dos esforços para atender às necessidades do mercado (Zhou *et al.*, 2022). No entanto, é preciso lidar com as dificuldades associadas às variabilidades internas e externas ao sistema produtivo (Harjunkoski *et al.*, 2014). Enquanto as variabilidades internas envolvem a disponibilidade dos recursos, a complexidade dos fluxos de produção e a característica estocástica dos processos (Benkel; Jørnsten; Leisten, 2016), as externas ocorrem em função da oscilação de demanda, do fornecimento de matéria prima e da exigência dos clientes (Oztemel; Gursev, 2020). Em razão de necessidades como redução de custos e alto padrão de qualidade, os desafios do sequenciamento exigem tomadas de decisão mais ágeis e assertivas (Zhou *et al.*, 2022).

Abordagens de otimização têm se destacado como solução para tais desafios, uma vez que conseguem lidar com as múltiplas variáveis presentes no sequenciamento (Neufeld; Schulz; Buscher, 2023; Zhang *et al.*, 2023). Essas abordagens são conhecidas por Advanced Planning and Scheduling (APS), pois conseguem elaborar combinações ótimas, ou quase ótimas, para as ordens de produção, permitindo reduzir custos e elevar a produtividade (Chen; Ji, 2007; Márquez; Ribeiro, 2022). Na manufatura, a implementação de um APS reduziu 75% do esforço para sequenciamento (Chen *et al.*, 2017), minimizou a quantidade de pedidos em atraso ao elevar a utilização de robôs autônomos (Liang; Zhou; Jiang, 2024) e reduziu em 34% o tempo para fabricação de peças pré-moldadas (Liu *et al.*, 2023). Apesar dos avanços, a combinação de múltiplas variáveis com eventos estocásticos eleva exponencialmente o tempo de processamento para se obter uma solução razoável, dificultando a implementação do APS nos cenários descritos (Missbauer; Uzsoy, 2022).

Uma solução potencial para lidar com esses desafios é a utilização de aprendizado por reforço (RL – Reinforcement Learning), um dos campos de estudo do aprendizado de máquina, em que um agente aprende ao interagir com o ambiente (Sutton; Barto, 2018). A capacidade do RL de tomar decisões sequenciais e de se adaptar à dinâmica do sistema é a principal motivação para que seja aplicado no sequenciamento da produção (Del Real Torres *et al.*, 2022; Li *et al.*, 2022; Wang; Pan; Wang, 2022). Estudos demonstraram a eficácia do RL em contextos de manufatura flexível, job shop, flow shop e fluxos reentrantes, resultando em melhorias em prazos de entrega, tempo de ciclo e utilização de recursos (Chen *et al.*, 2022; Gerpott *et al.*, 2022; Liu; Piplani; Toro, 2022; Tang; Salonitis, 2021; Woo *et al.*, 2021). Apesar dessa eficácia, poucos estudos utilizam teorias de gestão de operações para modelar e avaliar agentes de RL (Zhou *et al.*, 2022). Por exemplo, um sistema multiagente foi desenvolvido com conceitos de sustentabilidade do Lean (Paraschos; Koulinas; Koulouriotis, 2023). O RL foi empregado em ambientes produtivos baseados no CONWIP (Constant Work in Processes) para ajustar o volume de trabalho em processamento (Silva; Azevedo, 2019) e determinar quando antecipar ou atrasar a liberação das ordens (Xanthopoulos; Chnitidis; Koulouriotis, 2019). No entanto, os estudos desconsideram a possibilidade de absorver as variabilidades do sistema, de modo que o RL é utilizado, apenas, como um otimizador, não integrando plenamente os conceitos das teorias de gestão.

A Teoria das Restrições (TOC - *Theory of Constraints*), em especial o método Tambor, Pulmão e Corda (DBR – Drum, Buffer and Rope), oferece uma solução eficaz para lidar com a variabilidade, enfatizando a importância das restrições (gargalos) no gerenciamento dos sistemas produtivos (Goldratt, 2006). O DBR utiliza pulmões para proteger a restrição das variabilidades do sistema, balanceando o fluxo de trabalho e, consequentemente, estabilizando a produção (Schragenheim, 2010). Na manufatura, o DBR elevou a produtividade de uma linha automotiva de alta variabilidade (Golmohammadi, 2015), melhorou a utilização dos recursos em um setor de remanufatura (Ma; Chen, 2009) e reduziu o *lead-time* em uma fábrica automotiva (Dohale; Ambilkar; Bilolikar, 2021). Adicionalmente, a TOC compreende que a meta das organizações é gerar lucro tanto no presente quanto no futuro, propondo métricas que norteiam a tomada de decisão nos níveis estratégico, tático e operacional (Gupta; Ko; Min, 2002).

Apesar da efetividade da TOC e do RL para lidar com o sequenciamento da produção, as potenciais sinergias dessas áreas de estudo ainda não foram exploradas. Assim, esta pesquisa busca integrar os conceitos desses dois campos no sequenciamento da produção, investigando as seguintes questões de pesquisa: (i) Quais são as potenciais sinergias entre TOC e RL? (ii) Como essas sinergias podem ser integradas em um modelo conceitual?

Este capítulo, portanto, responde às questões de pesquisa ao conduzir uma revisão sistemática da literatura (RSL) a fim de analisar a modelagem dos agentes de RL em ambientes discretos e as simulações de processos baseadas no DBR. Para tanto, foram analisados 31 estudos relacionados ao RL e 19 envolvendo simulações do DBR. A análise temática evidencia e discute a utilização de teorias de gestão, em especial a TOC, como base para a modelagem de agentes de RL, assim como os possíveis benefícios do uso de RL para aprimorar o DBR. Por fim, desenvolve-se um modelo conceitual, relacionando os conceitos da TOC com os principais elementos do RL.

O presente capítulo está estruturado da seguinte forma: a seção 2 descreve os procedimentos metodológicos; a seção 3 apresenta os principais resultados; a seção 4 discute, analisa e sintetiza os resultados; e a seção 5 evidencia as contribuições, as limitações e as oportunidades para novas pesquisas.

3.2 PROCEDIMENTOS METODOLÓGICOS

Para alcançar o objetivo da pesquisa, conduziu-se uma revisão sistemática da literatura, a partir do *Literature Grounded Theory* (LGT) (Ermel *et al.*, 2021-). A Figura 2 expõe os procedimentos metodológicos da pesquisa. A relação entre a TOC e o RL foi definida como tema centra deste trabalho. Assim, a primeira etapa compreende uma busca preliminar nas bases Scopus, Science Direct, Web of Science e IEEE para identificar estudos que relacionem os dois temas (passo 1 da Figura 2). Na etapa 2, os documentos resultantes foram submetidos à leitura inspecional (Adler; Van Doren, 2010), constatando-se a ausência da relação entre os temas. Assim, um protocolo de pesquisa (Tabela 10 do Apêndice I) foi desenvolvido com base em leituras preliminares e validado por seis especialistas na área (passo 3 da Figura 2). Os especialistas foram selecionados pelos seguintes critérios: (i) ter publicado revisões sistemáticas ou ter conhecimento sobre o tema; e (ii) ser pesquisador com doutorado na área de pesquisa.

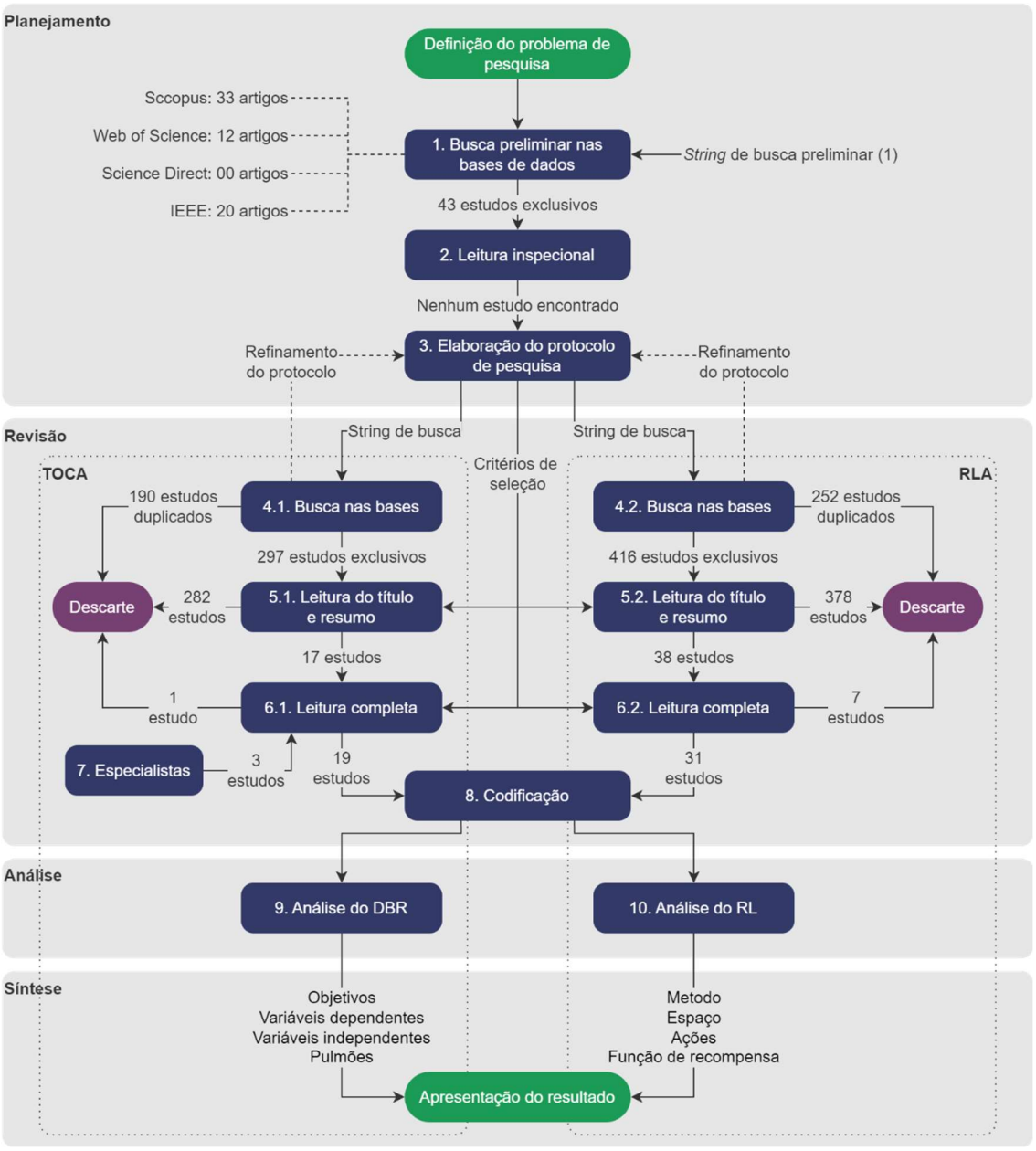
A partir desse ponto, a revisão foi segmentada para facilitar a identificação dos estudos: uma abordagem para TOC (TOCA) e uma abordagem para RL (RLA). Nas etapas 4.1 e 4.2 (Figura 2), as buscas foram executadas nas mesmas bases da busca preliminar, e as duplicatas foram removidas, resultando em 297 itens para TOCA e 416 para RLA. Os critérios de seleção do protocolo foram aplicados nas etapas 5.1 e 5.2 (Figura 2), durante a leitura do título, do resumo e das palavras-chave. Após a última etapa, 38 artigos para RLA e 17 artigos para TOCA foram selecionados para leitura completa (etapas 6.1 e 6.2 da Figura 2). Na etapa 7 (Figura 2), foram consultados especialistas e incluídos 3 artigos para leitura completa. Ao final da leitura, o corpus de análise estava composto por 19 artigos para TOCA e 31 para RLA (Figura 2). As estatísticas de exclusões das etapas 4.1, 4.2, 5.1, 5.2, 6.1 e 6.2, para TOCA e RLA, são apresentadas, respectivamente, nas Tabelas 12 e Tabela 13 do Apêndice I.

O processo de codificação do corpus de análise (etapa 8 da Figura 2) utilizou o método de codificação mista, com códigos categóricos e abertos, assim como a criação de categorias *a priori* e *a posteriori* (Ermel *et al.*, 2021-). Os códigos categóricos e *a priori* foram definidos previamente à leitura completa dos artigos, sendo baseados nas referências utilizadas para construção do protocolo da RSL. Os códigos abertos e as categorias a *posteriori* foram criados durante a leitura analítica

(Strauss; Corbin, 1990). A visão geral dos códigos está disponível na Tabela 14 do Apêndice I.

As análises foram realizadas em três etapas. O passo 9 refere-se à TOCA, e contém a classificação do objetivo das simulações, as variáveis dependentes, as variáveis independentes e os buffers. Os objetivos e as variáveis dependentes e independentes foram categorizados *a posteriori*. Os buffers foram definidos com base na literatura do DBR (Goldratt, 2006). No passo 10, a análise contempla os estudos do RLA e foi executada com adaptação do framework proposto por Esteso *et al.* (2023). A modelagem do agente foi mapeada com o algoritmo utilizado, a linguagem de programação, o espaço observável, as ações do agente e a função de recompensa (Esteso *et al.*, 2023). Por fim, apresenta-se um modelo funcional das possíveis sinergias entre a TOC e o RL.

Figura 2 - Etapas da revisão sistemática da literatura.



Fonte: Adaptado de (Ermel *et al.*, 2021-).

3.3 ANÁLISE DOS RESULTADOS

Esta seção apresenta a análise dos dados coletados durante a revisão sistemática da literatura. A análise foi conduzida em duas etapas. A primeira contempla os artigos com o tema Teoria das Restrições (TOCA), e a segunda aborda os artigos com o tema Aprendizado por Reforço (RLA). Por fim, um modelo funcional das simulações é apresentado.

3.3.1 Teoria das Restrições (TOCA)

Os resultados indicam que as simulações se concentram principalmente na análise da aplicabilidade do DBR (47%), tanto em casos empíricos quanto hipotéticos. Estudos com o objetivo de otimizar parâmetros do DBR (21%) e identificar a restrição (16%) são os mais frequentes nesse contexto.

Quanto às variáveis independentes, a mais utilizada é o sequenciamento da produção (47%), que envolve a avaliação de regras de sequenciamento para o gargalo, a modificação da programação com base na mudança do recurso gargalo ou a análise do impacto do sequenciamento proposto pelo DBR. Além do sequenciamento, alguns estudos investigam em que medida o tamanho dos pulmões influencia o desempenho do sistema (26%). A taxa de chegada das ordens determina o volume de trabalho, sendo manipulada para controlar a capacidade protetiva e o impacto dos produtos que não passam pelo gargalo (21%). Alterações na posição da restrição (16%) e no tamanho do lote (16%) são simuladas para avaliar o impacto que causam no tempo de atravessamento, no atraso da entrega e no tempo de espera nas filas.

As variáveis dependentes, por sua vez, são empregadas para monitorar a simulação e medir o impacto das mudanças impostas pelas variáveis independentes. Devido à natureza metodológica dos estudos e à dificuldade em modelar custos, as variáveis mais utilizadas representam unidades de tempo ou quantidade de ordens. No contexto temporal, há variações na interpretação do tempo de atravessamento, representado como o tempo total decorrido entre o recebimento e a entrega do pedido, assim como o tempo transcorrido entre a liberação para o chão de fábrica e a entrega (Thürer; Stevenson, 2018). Enquanto o tempo de atravessamento (35%) avalia o tempo de resposta do sistema, o atraso das ordens (26%) e a quantidade de ordens atrasadas e antecipadas (19%) avaliam a capacidade de entregar os pedidos no prazo (Atwater; Chakravorty, 2002). O volume total produzido (13%) representa, em cada ciclo da simulação, a produtividade total do sistema. A porcentagem de utilização dos recursos pode indicar o recurso restritivo e se ele está sendo explorado, no entanto não deve ser avaliada como um parâmetro de desempenho para a produtividade (Gupta; Ko; Min, 2002).

Entre os pulmões elencados na literatura (Goldratt, 2006), o pulmão da restrição é o mais utilizado (73%) devido à capacidade que tem de absorver as variabilidades e

de proteger o gargalo. No entanto, alguns estudos não fazem referência à utilização desse recurso (Al-Aomar, 2006; Castro; Godinho-Filho; Tavares-Neto, 2022; Dohale; Ambilkar; Bilolikar, 2021; Gupta; Ko; Min, 2002). Os pulmões de expedição e de montagem protegem, respectivamente, os prazos de entrega e o fluxo das ordens, mas não são essenciais para o sucesso do DBR, podendo ser aplicados conforme a necessidade (Betterton; Cox, 2009). Enquanto o pulmão de capacidade indica como o sistema poderá lidar com variabilidades no futuro, impactando o tempo de atravessamento e as entregas no prazo (Atwater; Chakravorty, 2002), o pulmão de espaço monitora a saída do gargalo para evitar bloqueios (Betterton; Cox, 2009). O pulmão de estoque não foi identificado nos estudos.

Os cinco passos de focalização são citados em cinco estudos, porém, apenas dois deles desenvolvem esses passos (Gupta; Ko; Min, 2002; Schragenheim; Ronen, 1990). O primeiro passo de focalização é considerado elemento fundamental das simulações, contudo poucos trabalhos utilizam a análise da capacidade produtiva para identificar o gargalo (Golmohammadi, 2015; Gupta; Ko; Min, 2002; Schragenheim; Ronen, 1990; Wu; Morris; Gordon, 1994). O bloqueio ou a inanição do gargalo é objeto de estudo de apenas uma pesquisa (Betterton; Cox, 2009). As métricas de desempenho propostas pela TOC, como o ganho, a despesa operacional e o inventário não são exploradas nas simulações (Cox III; Schleier Jr., 2010; Goldratt, 2006).

3.3.2 Aprendizado por reforço (RLA)

O aprendizado por reforço é uma técnica de aprendizado de máquina que treina um agente para tomar decisões interagindo com um ambiente (Dogan; Birant, 2021). O objetivo do RLA é aprender a política ideal que maximiza a recompensa acumulada, identificando o estado atual do ambiente e agindo conforme as ações disponíveis (Sutton; Barto, 2018). Nesse contexto, esta seção apresenta os resultados da análise RLA. A análise expõe a modelagem do espaço observável e as ações do agente, assim como a função de recompensa e a linguagem de programação (Esteso *et al.*, 2023). A Tabela 18, do Apêndice I, evidencia a consolidação das categorias de análise.

O espaço observável representa o estado atual do ambiente virtual, interpretado pelo modelo de RL para orientar a escolha da ação. Esse estado deve estar alinhado com a política do ambiente e refletir uma relação de causalidade com

as ações (Esteso *et al.*, 2023). A seleção das variáveis não é trivial. Enquanto alguns estudos optam por espaços reduzidos, comprometendo a precisão do modelo (Jeon *et al.*, 2022; Marchesano *et al.*, 2022), outros propõem espaços combinados e extensos, aumentando a complexidade da solução e exigindo treinamentos prolongados (Hu *et al.*, 2020; Liu *et al.*, 2022). Nesse contexto, variáveis relativas aos recursos, como a quantidade de ordens na fila, o estado e a taxa de utilização, indicam ao modelo o estado físico do sistema, enquanto variáveis temporais, como o tempo restante para conclusão da ordem e a data de entrega, revelam a urgência.

As ações do modelo constituem a interface de interação com o ambiente, podendo ser diretas ou indiretas. À medida que a complexidade do ambiente aumenta, as ações indiretas tendem a melhorar o desempenho do modelo (Samsonov; Ben Hicham; Meisen, 2022). Consequentemente, as ações predominantemente modeladas são regras de despacho. Alguns estudos adotam ações diretas, como a seleção de uma ordem na fila, a indicação do recurso para processamento ou mesmo a decisão de não realizar nenhuma ação e aguardar outra tomada de decisão. Por envolverem recursos ou ordens específicas, as ações diretas devem representar cada elemento do sistema, aumentando o número de ações e podendo comprometer a escalabilidade da solução (Kuhnle *et al.*, 2022).

A definição da função de recompensa é crucial para o aprendizado do modelo, fornecendo *feedback* sobre as ações escolhidas e delineando seu objetivo (Del Real Torres *et al.*, 2022). No contexto do planejamento e controle da produção, os principais objetivos incluem a redução do tempo de atravessamento, o cumprimento de prazos de entrega e a minimização do estoque (Lödding, 2013). Nesse sentido, as funções de recompensa foram predominantemente modeladas com base em dados de atraso nas entregas e em taxa de produção, penalizando o agente pelos atrasos e desconsiderando as ordens realizadas no prazo. Quanto aos níveis de inventário, são representados pela quantidade de ordens em aberto e em processamento, sem considerar os estoques. Apesar de poder distorcer o desempenho do modelo, a taxa de utilização dos recursos é empregada para a função de recompensa.

3.4 DISCUSSÃO

A restrição do sistema, ou o gargalo, é o principal conceito da TOC, porém, apenas um dos modelos de aprendizado de máquina considerou a existência da

restrição, desenvolvendo uma rede neural que a identifica e eleva a capacidade do recurso restritivo até eliminá-lo (Thomas *et al.*, 2018). Além de superestimada, essa abordagem é, por vezes, irreal, pois um recurso é considerado gargalo justamente por não ter opções viáveis para elevar sua capacidade (Goldratt, 2006). A falta de consciência sobre a restrição expõe que as modelagens do espaço observável, as ações e a função de recompensa dos modelos de RL podem ser falhas.

O espaço observável representa o estado atual do sistema para o agente treinado e deve incluir as variáveis relacionadas ao objetivo do modelo (Sutton; Barto, 2018). Ainda que o monitoramento de filas, recursos e ordens em produção seja comum, nem sempre essas variáveis são as mais representativas do estado atual. A fila (inventário) que precede a restrição é crucial no sistema produtivo, pois protege o ganho, enquanto as filas dos recursos não gargalos representam desperdício de tempo (Goldratt, 2006). Embora seja teoricamente possível monitorar todas as filas em um ambiente hipotético e virtual, na prática isso é custoso e, muitas vezes, inviável. Ao desenvolver um agente inteligente, o objetivo é implantá-lo em uma linha de produção real, portanto, seu espaço observável deve ser factível. Sob o ponto de vista da TOC, o monitoramento da linha ocorre por meio do estado dos pulmões, embora nenhum modelo de RL incorpore esse conceito em seu espaço observável. Além dos pulmões convencionais, pode ser importante monitorar o pulmão de capacidade para identificar quando as oscilações do mercado deslocam a restrição de um recurso gargalo para outro com capacidade restritiva.

As ações dos modelos frequentemente se concentram no sequenciamento de toda a linha de produção, negligenciando a importância do gargalo. O sequenciamento é crucial para otimizar o aproveitamento da restrição, minimizando os tempos de preparação e assegurando que ela opere no processo de modo a resultar em maiores ganhos (Goldratt, 2006). No entanto, ao sequenciar todos os recursos, é provável que o modelo convirja para ótimos locais em vez de alcançar o ótimo global. Da mesma forma, ao focar apenas no *backlog* de ordens, podem surgir filas desnecessárias e uma ativação inadequada de recursos. Um recurso não gargalo não deve acumular longas filas e nem ser ativado constantemente, a menos que seja importante para expandir a capacidade a jusante (Cox III; Schleier Jr., 2010). Em relação aos pulmões, nenhum estudo até o momento se preocupou em ajustar dinamicamente as filas dos recursos, uma vez que todas são consideradas infinitas. Consequentemente, recursos que não

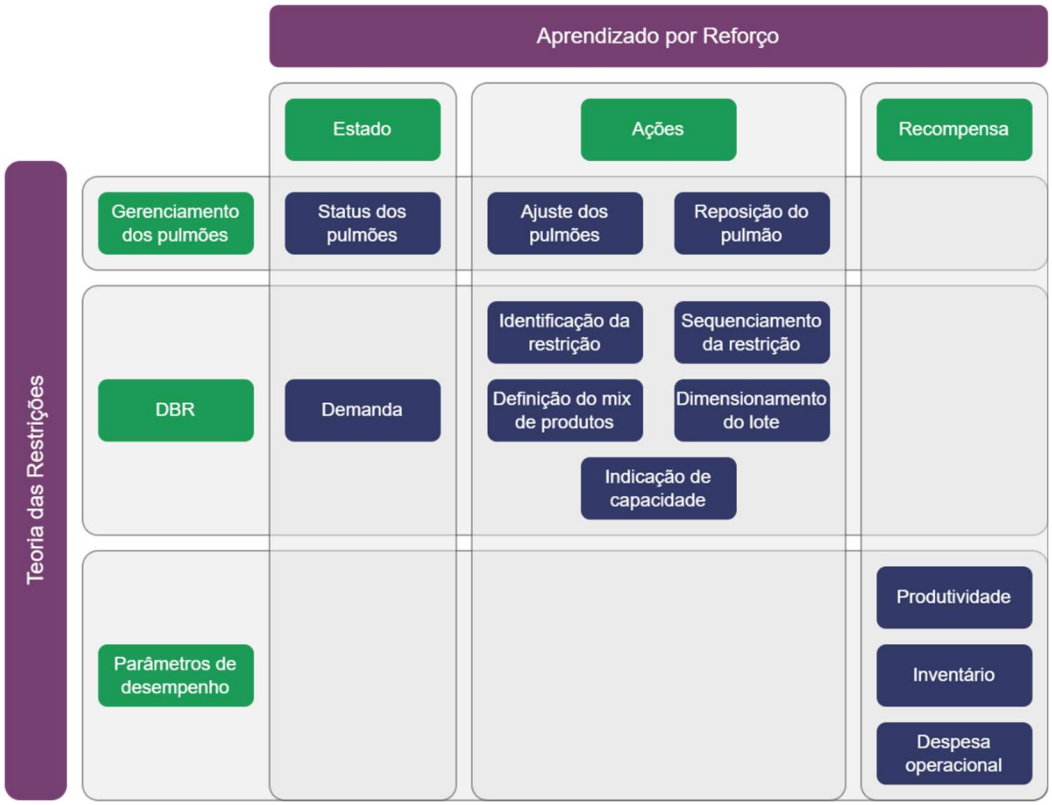
são gargalos podem operar com grandes filas, enquanto o recurso gargalo pode enfrentar bloqueios ou ociosidade (Betterton; Cox, 2009).

A função de recompensa nos modelos de RL muitas vezes não reflete métricas de desempenho alinhadas a filosofias de gestão, o que pode distorcer o aprendizado do modelo e dificultar sua aplicação em cenários reais. Além disso, as métricas usadas tendem a ser generalizadas para todos os produtos e processos. Não há consideração, em nenhum algoritmo, sobre a redução da produção de certos itens para aproveitar ao máximo a restrição com outros que proporcionam maiores ganhos em relação ao tempo utilizado pela restrição. Aspectos como o ganho e a despesa operacional não são explorados nos estudos analisados, possivelmente devido à escassez de dados ou ao foco em questões metodológicas. O preço de venda e o custo de produção são particularmente difíceis de quantificar quando se lida com dados sintéticos, pois estão sujeitos às flutuações do mercado. Sem um contexto real, é provável que esses valores sejam imprecisos e, consequentemente, prejudiquem o aprendizado do modelo. Embora métricas como o atraso de pedidos e o tamanho das filas sejam indicadores de desempenho que visam reduzir o inventário e garantir entregas pontuais, não asseguram, necessariamente, o aumento dos ganhos, uma vez que não estão diretamente ligadas à restrição do sistema (Kim *et al.*, 2021; Sakr *et al.*, 2023).

As simulações baseadas no DBR focam em validar a aplicabilidade do método, mas nem todas seguem os cinco passos de focalização propostos pela TOC (Cox III; Schleier Jr., 2010). A detecção da restrição não é considerada na maioria dos estudos, sendo este um ponto em que aplicações de algoritmos de RL podem auxiliar. Além disso, os demais passos também poderiam se beneficiar de modelos de RL, que podem contribuir para explorar a restrição, decidir o mix de produtos e o tamanho do lote e definir o sequenciamento. Na etapa de subordinação, a ativação dos recursos no momento correto, a transferência de ordens da restrição para recursos não restritivos e o ajuste do tamanho do lote nos recursos não restritivos também são exemplos de ações que podem ser treinadas com RL. Não tão diretamente como nas demais ações, na etapa de elevação da restrição um modelo de RL pode indicar os pontos de oscilação de demanda que possam requerer elevação da capacidade da restrição, utilização de horas extras ou terceirização. No último passo, a melhoria contínua, há relação direta com o conceito de CI/CD (*Continuous integration/Continuous delivery/Continuous deploymen*) proposto para modelos de aprendizado de máquina (Kreuzberger; Kühl; Hirschl, 2023).

A otimização de parâmetros do DBR é um dos objetivos do grupo de artigos da TOCA, em que o parâmetro mais estudado é o tamanho dos pulmões (Chen; Chen; Ma, 2007; Ma; Chen, 2009; Schragenheim; Ronen, 1991; Zhang; Du, 2015). Contudo, mesmo com a aplicação de algoritmos para otimização, os valores obtidos são definidos como estáticos para o cenário simulado, geralmente representados pela média do período. Em processos estocásticos, podem existir períodos em que os valores configurados não sejam os necessários e, nesse caso, é preciso avaliar novamente o cenário. Nesse contexto, um modelo de RL poderia interpretar o estado atual do ambiente, executar o dimensionamento dinâmico dos pulmões e definir os tempos de reposição mais adequados, assim como sequenciar a fila de ordens para o gargalo (Cox III; Schleier Jr., 2010; Schragenheim; Ronen, 1991). De modo geral, as ações e métricas propostas pela TOC demonstram sinergia com os algoritmos de RL, e a implementação dos conceitos da TOC nos modelos de RL pode elevar o ganho e reduzir a necessidade computacional. A Figura 3 sintetiza os conceitos da TOC e sua relação com o RL.

Figura 3 – Modelo conceitual da integração entre TOC e RL.



Fonte: Elaborado pelo autor.

3.5 CONCLUSÃO

Este estudo identificou e analisou os pontos de sinergia entre a TOC e o RL, destacando as limitações e as áreas passíveis de aprimoramento dessas abordagens. Foram analisados 50 artigos que abordam simulações de linhas de produção, dentre os quais 19 são relacionados à TOC e 31 ao RL. Como resultado, os pontos de interseção entre a TOC e o RL foram identificados e organizados em um modelo conceitual. Foi possível, ainda, compreender como o Gerenciamento dos Pulmões, o DBR e os parâmetros de desempenho da TOC se relacionam com o espaço observável (i.e., estado), as ações e a função de recompensa do RL. Além disso, observou-se uma escassez de estudos de RL que incorporam modelos teóricos de gestão de operações em suas soluções. Ao mesmo tempo, não se identificou nenhum estudo no campo da TOC que utilize técnicas de aprendizado de máquina para melhorar resultados.

No que diz respeito às limitações, a revisão da literatura se restringe a simulações baseadas em eventos discretos (DES), excluindo outras técnicas, como a dinâmica de sistemas. Além disso, não foram considerados estudos que envolvam linhas de produção com processos contínuos, nem outros que estão além do escopo do sequenciamento da produção. No contexto da revisão da literatura, também não foi empregada a técnica de bola de neve para ampliar a base de dados. Em termos de desenvolvimento, não houve a elaboração de um modelo de RL para validar a sinergia proposta e comparar os resultados com estudos já realizados, o que mantém este estudo no âmbito teórico.

As limitações indicadas introduzem oportunidades para pesquisas futuras. Dentre as sugestões, estão a ampliação do escopo com a generalização dos processos de fabricação e a inclusão de demais técnicas de simulação. Outra opção é o desenvolvimento de um modelo de aprendizado por reforço baseado nos conceitos da Teoria das Restrições. Paralelamente, pode-se, ainda, comparar esse modelo com uma linha de produção gerenciada pelo método do Tambor-pulmão-corda.

4 INTEGRAÇÃO DO TAMBOR, PULMÃO E CORDA COM APRENDIZADO POR REFORÇO

Os sistemas de manufatura enfrentam desafios devido à crescente demanda por produtos customizados e à natureza estocástica dos processos produtivos, o que exige gerenciamento otimizado da produção. Tecnologias avançadas como sistemas ciberfísicos, IoT e IA são aplicadas para elevar a eficiência do sequenciamento da produção, com o objetivo de sincronizar a capacidade com a demanda. O aprendizado por reforço (RL – *Reinforcement Learning*) surge como uma alternativa promissora para automatizar a tomada de decisão e melhorar a eficiência produtiva dos sistemas de manufatura. No entanto, as abordagens tradicionais de RL negligenciam a utilização de teorias de gestão e sua capacidade de absorver as variabilidades inerentes ao sistema produtivo. A Teoria das restrições (TOC – *Theroy of Constraints*) propõe o método Tambor, Pulmão e Corda (DBR – *Drum, Buffer and Rope*) para lidar com as variabilidades do sistema produtivo e sequenciar a produção. Nesse contexto, este capítulo integra os conceitos da TOC em um modelo de RL para sequenciamento da produção, bem como, compara os resultados pragmáticos desse modelo com o desempenho do DBR. Uma análise exploratória é conduzida para extrair implicações teóricas do modelo de RL. Como principal contribuição, os resultados evidenciam a capacidade do modelo proposto de melhorar o desempenho produtivo, em especial decompondo o pulmão de expedição proposto pelo DBR e salientando a importância do nível de ordens em produção (WIP – *Work in Process*).

Palavras-chave: Teoria das Restrições; Aprendizado por Reforço; Sequenciamento da Produção; Planejamento e Controle da Produção; Manufatura.

4.1 INTRODUÇÃO

Os sistemas de manufatura enfrentam desafios associados à complexidade, uma vez que o aumento da demanda por produtos customizados aliado à natureza estocástica dos processos exige constantes tomadas de decisão e gerenciamento otimizado da linha de produção (Guo *et al.*, 2024). Tecnologias como sistemas cyber físicos, internet das coisas e inteligência artificial são utilizadas para elevar a produtividade e a eficiência, atuando, também, para consolidar a Indústria 4.0 (da Silva *et al.*, 2022; Zheng *et al.*, 2021). Devido à capacidade de impactar positivamente a eficiência do sistema produtivo (Zhou *et al.*, 2022), o sequenciamento da produção tem sido foco de implementações tecnológicas (Zheng *et al.*, 2021), aprimorando a habilidade de tomada de decisão para sincronizar a demanda do mercado à capacidade produtiva do sistema (Hopp; Spearman, 2008).

O sequenciamento da produção atribui tarefas a recursos durante intervalos de tempo definidos, visando atender a demanda com o menor esforço possível (Goldratt,

2006; Pinedo, 2016-). Esse objetivo é desafiado por variabilidades inerentes ao sistema produtivo, como a volatilidade do mercado, a disponibilidade dos recursos e o suprimento de matérias-primas (Harjunkoski *et al.*, 2014). Nesse contexto, sistemas avançados de planejamento e sequenciamento (APS – *Advanced Planning and Scheduling*) são utilizados como alternativa para elevar a eficiência (Sousa *et al.*, 2014). Entretanto, ainda que um APS possa atingir uma solução ótima ou quase ótima, seu custo computacional pode ser alto e, por vezes, inviável (Márquez; Ribeiro, 2022). Assim, alternativas como o aprendizado por reforço (RL – *Reinforcement Learning*) são utilizadas para aperfeiçoar o sequenciamento da produção (Panzer; Bender, 2022), permitindo automatizar a tomada de decisão ao fornecer respostas precisas e em tempo adequado à tomada de decisão (Panzer; Bender, 2022).

O RL é um campo de estudo do aprendizado de máquina, em que um agente interage com um ambiente para maximizar sua recompensa de longo prazo (Sutton; Barto, 2018). A capacidade do RL de tomar decisões sequenciais e de se adaptar à dinâmica do sistema é a principal motivação para que seja aplicado no sequenciamento da produção (Del Real Torres *et al.*, 2022; Li *et al.*, 2022; Wang; Pan; Wang, 2022). Estudos demonstram a eficácia do RL em contextos de manufatura flexível, *job shop, flow shop* e fluxos reentrantes, com obtenção de melhorias no atendimento a prazos de entrega, no tempo de atravessamento e na utilização de recursos (Chen *et al.*, 2022; Gerpott *et al.*, 2022; Liu; Piplani; Toro, 2022; Tang; Salonitis, 2021; Woo *et al.*, 2021). Apesar da eficácia do RL, poucos estudos incorporam bases teóricas de gestão de operações para modelá-lo e avaliá-lo (Zhou *et al.*, 2022). Por exemplo, um sistema multiagente foi desenvolvido com conceitos de sustentabilidade do Lean (Paraschos; Koulinas; Koulouriotis, 2023). Em outros estudos, o RL foi utilizado em um ambiente baseado no CONWIP (*Constant Work in Processes*) para ajustar o volume de trabalho em processamento de uma produção sob encomenda (Silva; Azevedo, 2019), para determinar quando antecipar ou atrasar a liberação das ordens (Xanthopoulos; Chnitidis; Koulouriotis, 2019) e para parametrizar dinamicamente o método de controle da produção DDMRP (*Demand Driven Material Requirements Planning*) (Lahrichi *et al.*, 2023). No entanto, ainda que efetivos, esses estudos desconsideram a possibilidade de absorver as variabilidades do sistema, além de utilizar o RL apenas como um otimizador e não como técnica plenamente integrada aos conceitos das teorias de gestão.

A Teoria das Restrições (TOC – *Theory of Constraints*) assume que o desempenho de qualquer sistema é limitado por uma restrição e propõe pulmões para absorver as variabilidades (Goldratt; Cox, 1984). Em tarefas de sequenciamento, a TOC utiliza os métodos Tambor Pulmão e Corda (DBR – *Drum, Buffer and Rope*) e Gerenciamento de Pulmões (BM – *Buffer Management*) (Goldratt, 2006; Schragenheim; Ronen, 1990). Juntos, esses métodos permitem gerenciar sistemas com alta interdependência, monitorando a restrição do sistema e incluindo pulmões para absorver a variabilidade (Golmohammadi, 2015). O DBR se mostrou eficiente em diversos cenários, melhorando a eficiência em uma produção sob encomenda do setor aeronáutico (Telles *et al.*, 2020) e elevando a produtividade de pequenas e médias empresas de manufatura (Buestán Benavides; Van Landeghem, 2015). Do mesmo modo, o DBR foi utilizado para aprimorar a fabricação de peças automotivas (Golmohammadi, 2015) e de painéis LCD (Wu *et al.*, 2010) e para restaurar cilindros de motor (Ma; Chen, 2009). Além do DBR, a TOC propõe que três métricas de desempenho são suficientes para tomadas de decisão, sendo elas: (i) ganho; (ii) inventário; e (iii) despesa operacional (Gupta; Ko; Min, 2002).

Diante desse contexto, este capítulo visa explorar os resultados da integração do RL com a TOC a fim de aprimorar o desempenho do sequenciamento da produção, respondendo às seguintes questões de pesquisa: (i) Como integrar a TOC e o RL? (ii) Quais são os resultados pragmáticos e teóricos dessa integração? Para responder a essas perguntas, foram elaborados dois métodos de sequenciamento, um baseado no DBR e o outro integrando os conceitos da TOC e do RL. Os métodos foram testados em um ambiente de simulação e avaliados comparativamente. Uma análise do comportamento foi elaborada para compreender as ações do RL, objetivando identificar e propor avanços para a teoria do DBR. Em termos de contribuições, este estudo demonstra que a integração da TOC com o RL aprimora a eficiência do sequenciamento por meio da decomposição do pulmão de expedição, em especial do nível de ordens em processamento. Concomitantemente, aproxima os modelos de RL às teorias de gestão de operações, avançando a modelagem de agentes inteligentes a cenários mais compreensíveis a profissionais e exigindo menor esforço de monitoramento (Neufeld; Schulz; Buscher, 2023).

As demais seções deste trabalho estão estruturadas da seguinte maneira: A Seção 2 apresenta a fundamentação teórica, tanto para o DBR quanto para o RL. A Seção 3 descreve o método de pesquisa, seguida pela Seção 4, que evidencia os

modelos computacionais desenvolvidos. Os cenários simulados são abordados na Seção 5, e, os respectivos resultados e análises, na Seção 6. Por fim, a Seção 7 discute os resultados, e a Seção 8 contextualiza as contribuições e limitações da pesquisa, bem como as oportunidades para novos estudos.

4.2 FUNDAMENTAÇÃO TEÓRICA

4.2.1 Tambor, pulmão e corda

O Tambor, pulmão e corda (DBR – *Drum, Buffer and Rope*) é o método proposto pela TOC para lidar com ambientes produtivos que operam com variabilidade e interdependência, buscando balancear o fluxo da produção e absorver a variabilidade (Goldratt, 2006; Telles *et al.*, 2022). No DBR, o Tambor é o elemento que dita o ritmo do sistema, considerado a restrição do sistema quando a capacidade de produção é inferior à demanda (Cox III; Schleier Jr., 2013). O tambor deve ser monitorado, sequenciado e protegido das variabilidades para proporcionar que o sistema produtivo tenha desempenho adequado (Goldratt, 2006).

Os Pulmões suportam a proteção em relação às variabilidades. São classificados em (Cox III, Schleier Jr., 2010; Goldratt, 2006): (i) pulmão da restrição – protege a operação da restrição, é utilizado como orientação para a liberação das ordens de produção; (ii) pulmão de expedição – protege a data da entrega das ordens, é utilizado como monitoramento das ordens processadas pela restrição; (iii) pulmão de montagem – protege a montagem dos produtos processados na restrição, possibilitando que não fiquem aguardando produtos que não passaram na restrição; (iv) pulmão de capacidade – trata-se da diferença entre a capacidade da restrição e os demais recursos, suportando o reabastecimento do pulmão da restrição em caso de paradas não planejadas à montante ou à jusante da restrição; e (v) pulmões de espaço – espaço físico situado logo após a restrição para garantir que a restrição não pare em caso de paradas não planejadas à jusante da restrição.

Por fim, a Corda é o elemento de liberação do material para as operações iniciais, conforme o ritmo da restrição. Em síntese, a Corda é o elemento de sincronização das operações por meio da liberação de materiais (Schragenheim; Ronen, 1990).

Juntamente ao DBR, a TOC propõe o gerenciamento dos pulmões (BM – *Buffer Management*). O BM é o método para monitorar e ajustar os pulmões do sistema, devendo ser implementado em conjunto ao DBR (Cox III; Schleier Jr., 2010). O monitoramento utiliza o valor de referência do pulmão para classificar seu estado atual em três níveis: (i) verde – superior a 66% do valor de referência; (ii) amarelo – entre 66% e 33% do valor de referência; e (3) vermelho – inferior a 33% do valor de referência (Goldratt, 2006). O valor de referência inicial para os pulmões pode ser considerado 50% do *lead-time* do sistema, entretanto o BM atua para ajustar o pulmão. De modo geral, pode-se reduzir o valor de referência do pulmão quando ele está operando a maior parte do tempo no verde, assim como se deve elevar o valor de referência quando está por muito tempo no vermelho (Schragenheim, 2010).

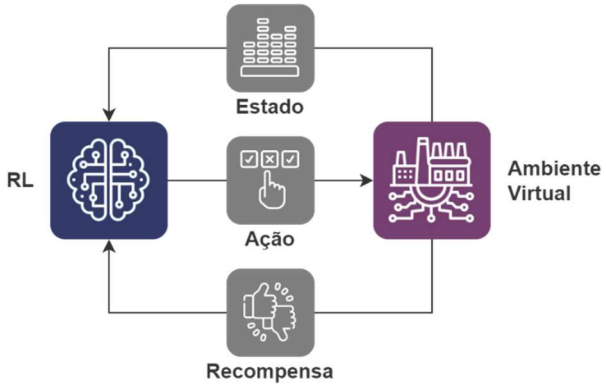
4.2.2 Aprendizado por reforço

O aprendizado por reforço (RL – *Reinforcement Learning*) é uma das áreas do aprendizado de máquina, tendo como premissa a existência de dois componentes: (1) agente e (2) ambiente. O agente interpreta o estado atual do ambiente, interage e recebe uma recompensa pela sua ação (Sutton; Barto, 2018), aprendendo, por meio de tentativa e erro, a política que maximiza a recompensa de longo prazo (Esteso *et al.*, 2023). Diferente das abordagens tradicionais de aprendizado de máquina, com bases de dados fixas, no RL o agente interage com um ambiente virtual, gerando novos dados de aprendizado (Panzer; Bender, 2022). Por se tratar de um ambiente virtual, este é comumente modelado como uma simulação de eventos discretos (DES – *Discrete Event Simulation*) (Kuhnle *et al.*, 2022; Nguyen *et al.*, 2022).

Algoritmos de RL são desenvolvidos com base no processo de decisão de Markov (MDP – *Makov Decision Process*), em que o estado atual do ambiente contém todas as informações necessárias à tomada da decisão que leva ao próximo estado. Logo, os elementos principais do MDP são (S, A, P, R, γ): o estado S; as ações A; a probabilidade de transição de estados ou a dinâmica do ambiente P; a função de recompensa R; e o fator de desconto γ, o qual representa o *trade-off* entre as recompensas futuras e as imediatas (Liu; Piplani; Toro, 2022). Assim, em cada passo t, o agente observa o estado $s_t \in S$, executa a ação $a_t \in A$, recebe a recompensa $r_t \in R$ e o próximo estado $s_{t+1} \in S$. A ação aplicada pelo agente é baseada na sua política $\pi(a|s)$, representada por uma distribuição de probabilidade

que mapeia uma ação a_t para o estado s_t (Sutton; Barto, 2018). A Figura 4 ilustra essa relação.

Figura 4 - Interação entre RL e ambiente virtual



Fonte: Adaptado de Sutton e Barto (2018).

O aprendizado do agente ocorre com a atualização da política $\pi(a|s)$ ou da função de valor. A função de valor estima o quão benéfico é estar naquele estado ou tomar certa ação, e pode ser: função de estado-valor $v_\pi(s)$ – informa ao agente a recompensa total que pode receber pela escolha de um determinado caminho; ou função de ação-valor $q_\pi(s,a)$ – informa ao agente o retorno esperado para tomar a ação a_t no estado s_t (Hayes *et al.*, 2022; Sutton; Barto, 2018). Porém, estimar o estado atual não é suficiente para o aprendizado. Conforme o agente interage com o ambiente, suas recompensas atualizam a função de valor, possibilitando, com o passar do tempo, estimar os estados subsequentes do ambiente (Panzer; Bender, 2022).

A fim de ajustar a função de valor para estimar valores do cenário, é preciso percorrer os pares de estado-ação. Para cenários de complexidade reduzida, é possível ajustar a função com abordagens computacionais diretas. Porém, em cenários com alto grau de complexidade, a solução mais aplicada é o uso de redes neurais como estimador das funções do RL, o que é caracterizado pelo termo aprendizado por reforço profundo (DRL – *Deep Reinforcement Learning*) (Esteso *et al.*, 2023).

4.3 DESIGN DA PESQUISA

Esta pesquisa foi conduzida com base no método de pesquisa de modelagem quantitativa. A modelagem quantitativa é o método de pesquisa que visa representar a realidade por meio da simplificação, possibilitando a interação entre variáveis e a abordagem de problemas complexos (MORABITO; PUREZA, 2012; DRESCH;

LACERDA; ANTUNES JUNIOR, 2015). Assim, a pesquisa é estruturada em cinco blocos, conforme ilustra a Figura 5.

Figura 5 - Desenho da pesquisa

Fonte: Elaborado pelo autor

O primeiro bloco (1) contém o modelo de simulação de fábrica. Na etapa 1.1, definiu-se o modelo base de simulação em conformidade com a literatura do DBR. Como referência, utilizou-se o modelo DES de Atwater e Chakravorty (2002) por contemplar compartilhamento de recursos em fluxos complexos e por ser explorado em outros estudos relacionados ao DBR (Atwater; Stephens; Chakravorty, 2004; Atwater; Chakravorty, 2002; Chakravorty; Atwater, 2005). O modelo foi implementado computacionalmente na etapa 1.2, utilizando-se o framework Simpy (versão 4.1.1) como linguagem de simulação (Law, 2015) e os padrões da OpenAI Gym (Brockman *et al.*, 2016). Em seguida, na etapa 1.3, o modelo foi validado por meio de teste-t (Machado *et al.*, 2023; Telles *et al.*, 2020), empregando-se a ocupação dos recursos como métrica. Na última etapa (1.4), a variabilidade do modelo foi elevada, visando avaliar as contribuições do RL em um ambiente com maior variabilidade e proximidade com a realidade (Machado *et al.*, 2023). Nesse sentido, foram atribuídas variáveis estocásticas relacionadas à manutenção, como o tempo entre falhas (TBF — *Time Between Failure*) e o tempo para reparo (TTR — *Time to Repair*), além da necessidade de preparação de máquinas entre as trocas de produto. Por fim, o modelo de simulação foi utilizado para desenvolver os modelos de sequenciamento da produção, tanto do DBR quanto do RL. Os detalhes da construção do modelo são apresentados na seção 4.1.

O sequenciamento do DBR foi desenvolvido no segundo bloco. Na etapa 2.1, foi conduzida a análise da produção, a fim de avaliar a capacidade da fábrica e de identificar a restrição (Goldratt, 2006; Schragenheim, 2010). Os resultados da análise embasaram a implementação do DBR (etapa 2.2) e do BM (etapa 2.3), sendo que esse processo foi elaborado conforme os conceitos propostos por Schragenheim (2010). Por fim, para que o sequenciamento represente seu melhor desempenho, os pulmões foram otimizados na etapa 2.4. A seção 4.2 contém os detalhes do modelo de sequenciamento do DBR.

Simultaneamente, o sequenciamento do RL foi elaborado no terceiro bloco, contemplando a modelagem do estado, das ações e da função de recompensa (Sutton; Barto, 2018). O estado (etapa 3.1) compreende o nível de produtos em processamento (WIP — *Work in process*), de produtos acabados (FG — *Finished goods*) e do pulmão da restrição (CB — *Constraint buffer*) (Goldratt, 2006). As

ações (etapa 3.2) representam a quantidade que deve ser produzida de cada produto, enquanto a função de recompensa (etapa 3.3) foi modelada com base nas métricas de desempenho da TOC (Gupta; Ko; Min, 2002; Machado *et al.*, 2023). Por fim, a etapa 3.4 contempla o treinamento do agente, no qual foi utilizado o algoritmo *Proximal Policy Optimization* (PPO) (Schulman *et al.*, 2017). O algoritmo PPO foi escolhido por duas razões: (i) apresenta resultados adequados em ambientes com decisões sequenciais (OpenAI *et al.*, 2019; Schulman *et al.*, 2017), como em uma produção MTS, em que a recompensa (venda) não ocorre imediatamente após a ação do agente, mas em momentos no futuro; e (ii) foi aplicado com sucesso em atividades de sequenciamento da produção (Chen *et al.*, 2022; Rummukainen; Nurminen, 2019), em planejamento da manutenção (Kuhnle; Jakubik; Lanza, 2019) e em reposição de inventário (Vanvuchelen; Gijsbrechts; Boute, 2020). Os detalhes da modelagem do RL são apresentados na seção 4.3.

Ambos os métodos de sequenciamento foram acoplados ao modelo de simulação da fábrica para simulação de cenários no bloco 4 (Law, 2015). Durante a simulação do DBR (etapa 4.1) e do RL (etapa 4.2), variáveis foram registradas (Tabela 1) visando tanto comparar a produtividade quanto compreender o comportamento dos métodos de sequenciamento. Além das métricas de desempenho, as variáveis utilizadas para a tomada de decisão e a respectiva decisão foram registradas para avaliar o comportamento dos métodos de sequenciamento, de modo que a soma da variável WIP com FG representa o pulmão de expedição no DBR (Goldratt, 2006).

Tabela 1 – Variáveis coletadas para análise da produtividade e do comportamento

Tipo da variável	Variável	Definição	Unidade de medida	Referência
Métrica de desempenho	Tempo de atravessamento (FT)	Tempo médio entre a liberação e a conclusão das ordens de produção.	Hora	(Castro; Godinho-Filho; Tavares-Neto, 2022; Thürer; Stevenson, 2018)
	Tamanho do lote (LS)	Média da quantidade total de peças por ordem de produção.	Peça	(Golmohammadi, 2015)
	Produtos em processamento (WIP)	Média da quantidade total de peças em produção.	Peça	(Castro; Godinho-Filho; Tavares-Neto, 2022)
	Produtos acabados em estoque (FG)	Média da quantidade de peças no estoque de produtos acabados.	Peça	(Castro; Godinho-Filho; Tavares-Neto, 2022)
	Inventário total (IV)	Média do total de peças no sistema produtivo (WIP + FG).	Peça	(Gupta; Ko; Min, 2002)
	Produtos vendidos (SP)	Produtos entregues quando requisitado pelo mercado (clientes).	Peça	(Schragenheim, 2010)
	Demanda (DM)	Quantidade de produtos solicitados pelo mercado.	Peça	(Schragenheim, 2010)
Registro do comportamento	CB	Quantidade de trabalho liberado para produção, a ser executado pela restrição ou o pulmão da restrição.	Hora	(Goldratt, 2006)
	WIP_n	Quantidade de peças em produção para o produto “n”.	Peça	(Goldratt, 2006)
	FG_n	Quantidade de peças no estoque de produtos acabado para o produto “n”.	Peça	(Goldratt, 2006)
	AC_n	Quantidade de peças liberadas para produção para o produto “n”.	Peça	(Goldratt, 2006)

Fonte: Elaborado pelo autor

No último bloco, foram desenvolvidas as análises com os resultados dos cenários. A primeira etapa (5.1) é a análise da produtividade, em que as métricas de desempenho foram comparadas a partir do teste de hipótese de Wilcoxon, com intervalo de confiança de 95% e significância de 5% (Law, 2015; Yamane, 1973). Na etapa 5.2, técnicas para explicar modelos de aprendizado de máquina foram aplicadas para compreender as ações dos modelos de sequenciamento, como a correlação de Spearman, a regressão linear e a árvore de decisão (Rudin *et al.*, 2022). A correlação de Spearman mede a relação monotônica entre dois conjuntos de dados, variando de −1 (correlação negativa) a 1 (correlação positiva), com 0 (zero) indicando não haver correlação (Yu *et al.*, 2021). A regressão linear determina

a contribuição das variáveis independentes sobre a variável dependente, além de evidenciar a proporção de variância da variável dependente, explicada pelas variáveis independentes (Nathans; Oswald; Nimon, 2012). Similarmente, a Árvore de decisão expõe um conjunto de regras interligadas, permitindo a visualização dos dados como uma sequência de condições legíveis por humanos (Kuhnle *et al.*, 2022).

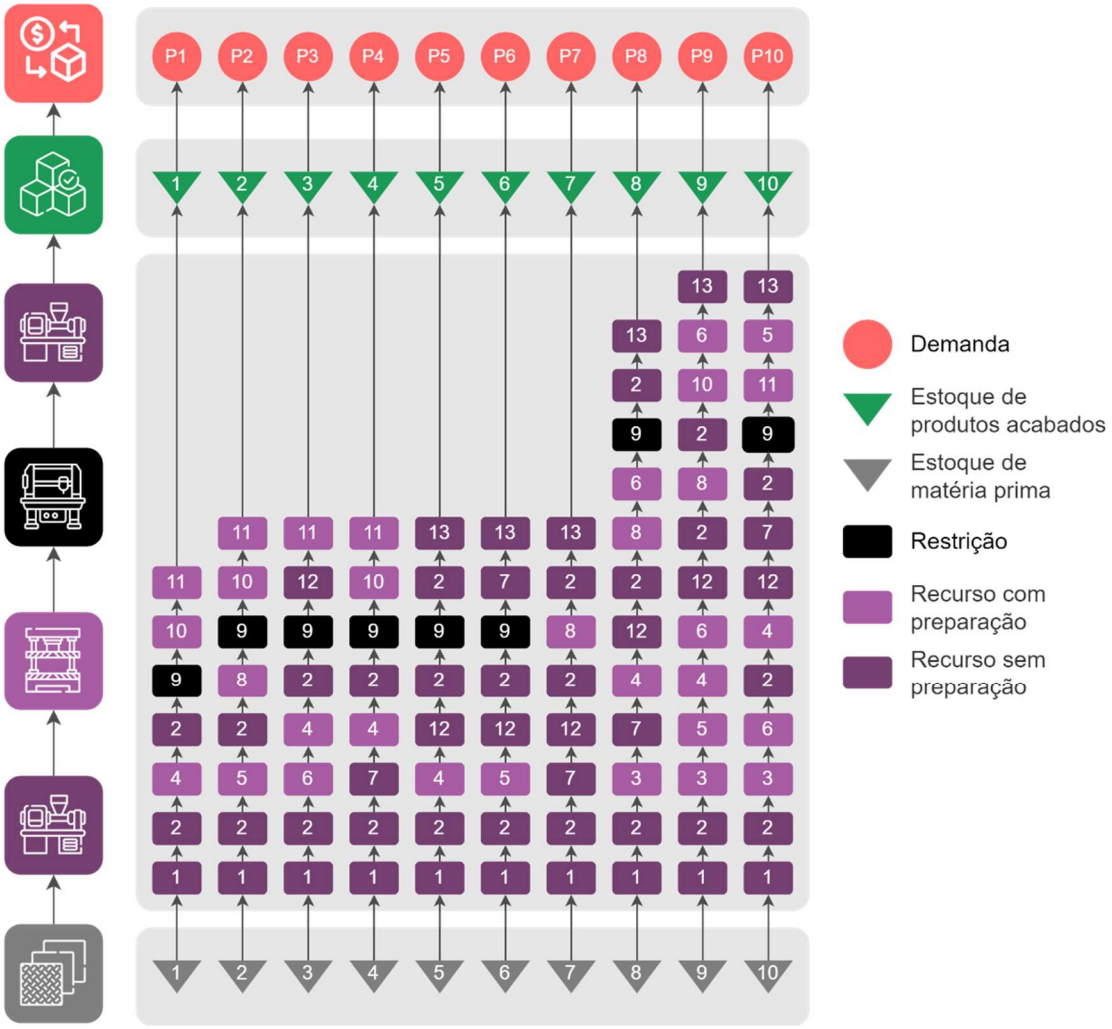
A Árvore de Decisão é um dos algoritmos de aprendizado de máquina mais compreensíveis, capaz de representar em um fluxograma os passos tomados pelo modelo (Myles *et al.*, 2004). Sua contribuição reside no fato de que, em cada decisão, os dados são desagregados, resultando na menor variância possível em cada conjunto. Assim, a variável que melhor explica a desagregação é determinada como a condição de escolha (Russel; Norvig, 2010). Além de evidenciar as decisões, a árvore de decisão permite calcular a importância de cada variável na decisão final. Logo, para cada ação, pode-se determinar quais variáveis têm maior influência na decisão (Rudin *et al.*, 2022). Por fim, foi analisada a distribuição das variáveis para evidenciar as diferenças de comportamento entre o DBR e o RL.

4.4 MODELOS COMPUTACIONAIS

4.4.1 Modelo de simulação / Ambiente virtual

Esta seção descreve o modelo de simulação da fábrica, apresentando as etapas do seu desenvolvimento. A primeira etapa é a seleção do modelo base de simulação, passando pela implementação do modelo e pela validação. Por fim, são apresentadas as modificações propostas para elevar a variabilidade do modelo. A Figura 6 ilustra os fluxos do modelo final de simulação.

Figura 6 - Modelo de simulação da fábrica



Fonte: Adaptado de Atwater e Chakravorty (2002)

4.4.1.1 Definição do modelo base de simulação

O modelo de simulação utilizado como base deste estudo foi proposto, inicialmente, por Bitran e Tirupati (1988) e utilizado por Atwater e Chakravorty (2002), com o objetivo de avaliar o impacto da variação da capacidade protetiva em sistemas de manufatura coordenados pelo DBR. O modelo consiste em um *job shop* que opera sob encomenda, com treze recursos e dez produtos, sendo que cada produto tem uma rotina única de processamento com tempos estocásticos e demanda conforme apresentado na Tabela 2. O recurso 9 é a restrição do sistema. A escolha desse modelo ocorreu por contemplar fluxos complexos e por ser explorado em outros estudos relacionados ao DBR (Atwater; Stephens; Chakravorty, 2004; Atwater; Chakravorty, 2002; Chakravorty; Atwater, 2005).

Tabela 2 - Demanda dos produtos

Produto	Distribuição	Média	Desvio Padrão
1	Erlang 3	10	5,77
2	Erlang 2	10	7,07
3	Uniforme	10	5,77
4	Erlang 3	10	5,77
5	Erlang 4	10	5,00
6	Erlang 2	10	7,07
7	Erlang 4	10	5,00
8	Uniforme	10	5,77
9	Erlang 4	10	5,00
10	Erlang 2	10	7,07

Fonte: Adaptado de Atwater e Chakravorty (2002)

Atwater e Chakravorty (2002) modelaram o DBR com um pulmão de restrição e um pulmão de expedição, considerando a metade do *lead-time* do sistema em cada pulmão. O *lead-time* foi especificado como a data de vencimento dos pedidos recebidos, tido, para todos, como 61 unidades de tempo, de modo que cada pulmão representa 30,5 unidades de tempo. O sequenciamento dos produtos processados pela restrição foi elaborado subtraindo-se, da data de vencimento, o pulmão de expedição, o tempo de processamento da restrição e o pulmão de restrição. Os pedidos que não são processados pela restrição são liberados assim que recebidos. Entretanto, considerando que a data de vencimento é fixa e que ambos os pulmões são metade dela, o sequenciamento dos pedidos processados pela restrição passa a ser o mesmo dos não processados. Por ser um modelo de simulação e considerar simplificações reais (Law, 2015), a preparação de máquinas e as falhas de equipamentos não foram implementadas no modelo base.

4.4.1.2 Implementação do modelo

O modelo de simulação foi desenvolvido como uma simulação por eventos discretos (DES – *Discrete Event Simulation*). A DES é uma das principais técnicas para analisar e testar sistemas de fabricação (Jahangirian *et al.*, 2010; Jeon; Kim, 2016; Nguyen *et al.*, 2022). No contexto do DBR, a DES foi utilizada para avaliar o impacto do deslocamento da restrição (Thürer; Stevenson, 2018), da variação na capacidade protetiva (Betterton; Cox, 2009) e do sequenciamento dos produtos

(Chakravorty; Atwater, 2005). Paralelamente, a DES é usada para desenvolver ambientes de treinamento de agentes de RL para sequenciamento da produção (Eriksson *et al.*, 2022; Liu; Piplani; Toro, 2022; Marchesano *et al.*, 2022). Assim, com base em estudos relacionados e com a finalidade de facilitar a integração com o agente de RL, o modelo de simulação foi desenvolvido em Python 3.11, com o módulo Simpy e conforme os padrões da OpenAI Gym para ambientes de treinamento de RL (Brockman *et al.*, 2016; Devanga; Badilla; Dehghanimohammadabadi, 2022a; Thürer; Fernandes; Stevenson, 2022; Thürer; Stevenson, 2018; Woo *et al.*, 2021).

4.4.1.3 Validação

O modelo base de simulação foi validado comparando-se as médias de utilização dos recursos, tanto para a simulação quanto para o estudo original (Tabela 3). A simulação foi executada por 200.000 unidades de tempo, sendo que os dados coletados são referentes às últimas 100.000, conforme o estudo inicial (Atwater; Chakravorty, 2002). Foram executadas 50 replicações e aplicado teste t para duas amostras, com intervalo de confiança de 95% e 5% de significância, para testar a hipótese nula ($h0$) de que não há diferença significativa entre o experimento e os valores reportados pelo estudo original. As variáveis de produtividade não foram consideradas na validação por três motivos: (i) o estudo inicial reportou as métricas em gráficos, dificultando a obtenção do valor real; (ii) o objeto de interesse é a dinâmica da simulação e não os resultados obtidos pelo estudo original; (iii) o modelo de simulação será modificado para ampliar a variabilidade, logo, suas métricas não serão reproduzidas.

Tabela 3 – Validação do modelo base de simulação

	Utilização no estudo original	Utilização no modelo base (média)
Recurso 01	0,78	0,780
Recurso 02	0,90	0,907
Recurso 03	0,80	0,801
Recurso 04	0,74	0,741
Recurso 05	0,80	0,801
Recurso 06	0,84	0,836
Recurso 07	0,71	0,706
Recurso 08	0,75	0,749
Recurso 09	0,94	0,939
Recurso 10	0,72	0,720
Recurso 11	0,72	0,721
Recurso 12	0,81	0,810
Recurso 13	0,87	0,870
Média	0,798	0,798
Desvio padrão	0,070	0,071
Erro (5%)	0,043	0,044
p-valor	0,999	
Teste-t (95%)	h0	

Fonte: Elaborado pelo autor

4.4.1.4 Elevação da variabilidade

No modelo base, os pedidos recebidos contêm uma unidade do respectivo produto, sendo liberados assim que recebidos. Em um cenário livre de *setup*, liberar ordens de produção unitárias não compromete o sistema, pois os recursos não necessitam alterar sua configuração para processar o novo produto. Contudo, ao se incluir a necessidade de preparação, reduz-se o tempo disponível para processamento de ordens de produção unitárias, visto que parte do tempo passa a ser consumido pelas preparações (Goldratt, 2006). Nesse contexto, as preparações de máquina foram adicionadas com o objetivo de que o agrupamento das ordens de produção seja uma ação necessária no modelo. Entretanto, para não alterar o recurso gargalo, somente alguns recursos foram modificados (Tabela 4).

Tabela 4 – Elementos de variabilidade adicionados ao modelo de simulação

	Preparação			Tempo entre falhas (TBF)			Tempo para reparo (TTR)		
Recurso	Distribuição	Média	σ*	Distribuição	Média	σ*	Distribuição	Média	σ*
1	Constante	0	0	Gamma	7.200	1.440	Gamma	4	2
2	Constante	0	0	Gamma	7.200	1.440	Gamma	4	2
3	Gamma	0,2	0,1	Gamma	7.200	1.440	Gamma	4	2
4	Gamma	0,2	0,1	Gamma	7.200	1.440	Gamma	4	2
5	Gamma	0,2	0,1	Gamma	7.200	1.440	Gamma	4	2
6	Gamma	0,2	0,1	Gamma	7.200	1.440	Gamma	4	2
7	Constante	0	0	Gamma	7.200	1.440	Gamma	4	2
8	Gamma	0,2	0,1	Gamma	7.200	1.440	Gamma	4	2
9	Gamma	0,5	0,1	Gamma	7.200	1.440	Gamma	8	2
10	Gamma	0,2	0,1	Gamma	7.200	1.440	Gamma	4	2
11	Gamma	0,2	0,1	Gamma	7.200	1.440	Gamma	4	2
12	Constante	0	0	Gamma	7.200	1.440	Gamma	4	2
13	Constante	0	0	Gamma	7.200	1.440	Gamma	4	2

Fonte: Adaptado de Atwater e Chakravorty (2002).

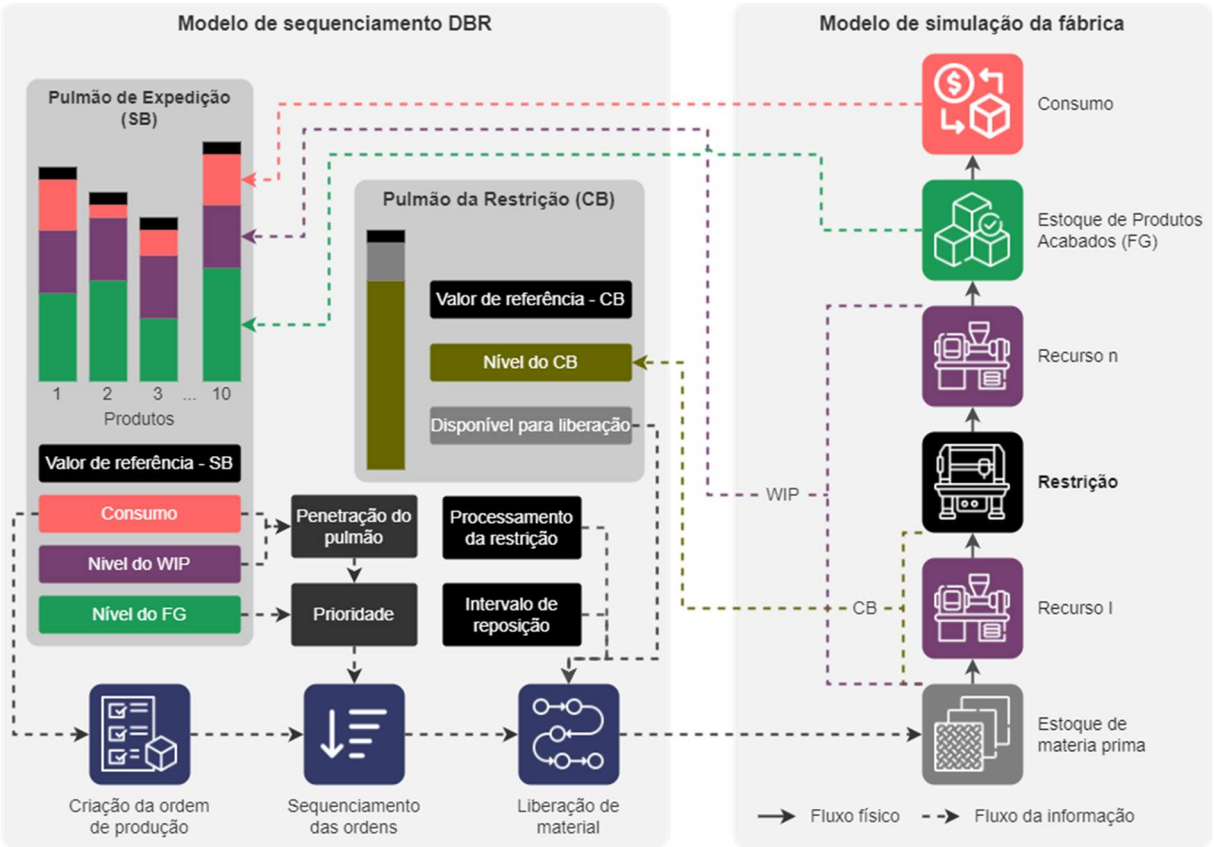
A falha de equipamento foi considerada para todos os recursos, pois, em um cenário real, nenhum recurso está livre de falhas. As falhas foram modeladas com as variáveis de tempo entre falhas (TBF – *Time Between Fail*) e tempo para reparo (TTR – *Time To Repair*). A ausência de dados históricos impossibilitou a definição de distribuições de probabilidade conforme a manutenção baseada em confiabilidade (Sellitto; Pinho, 2022). No entanto, foi utilizada a distribuição gamma (Tabela 4) por ser uma das mais empregadas em estudos relacionados à manutenção (Fogliato; Ribeiro, 2009; Sellitto; Brusius, 2017).

O modelo base não determina a relação da unidade de tempo com um evento real, contudo, analisando as médias e os desvios padrão dos parâmetros utilizados, pode-se considerar que uma unidade de tempo seja equivalente a uma hora. Assim, a média do tempo entre falhas foi dimensionada para considerar uma quebra a cada 300 dias trabalhados, com desvio padrão de 60 dias. A mesma lógica foi utilizada para o tempo de reparo, dimensionado para 4 horas, com desvio padrão de 2 horas. A fim de salientar a existência da restrição, tanto o tempo de preparação quanto o tempo de reparo foram elevados na restrição, sendo, respectivamente, 0,5 e 8 horas.

4.4.2 Modelo de sequenciamento pelo DBR

O sequenciamento proposto pela TOC compreende a utilização do DBR e do BM. O DBR é responsável por sequenciar e liberar as ordens de produção conforme o ritmo da restrição, enquanto o BM monitora os pulmões do sistema para operarem em condições ideais (Goldratt, 2006). A integração entre o DBR e BM é essencial para a produtividade do sistema em ambientes com elevada oscilação de demanda e variabilidade (Schragenheim, 2010). Entretanto, ainda que exista um algoritmo para o BM, Schragenheim (2010) indica que o ajuste dos pulmões deve ser realizado sob avaliação humana, ponderando-se os eventos que causaram a necessidade de ajuste com a demanda prevista. Assim, aliado ao regime estocástico estacionário do modelo de simulação (Law, 2015), o BM foi implementado com o objetivo de obter os valores ideais dos pulmões, fixando-os durante a simulação dos cenários. Nesse contexto, a Figura 7 ilustra a integração do modelo de simulação com o sequenciamento do DBR, o qual é detalhado nas seções subsequentes.

Figura 7 – Modelo de sequenciamento do DBR



Fonte: Elaborado pelo autor

4.4.2.1 Análise da operação

O primeiro passo do processo de focalização é a identificação da restrição (Goldratt; Cox, 1984). Para tanto, mesmo sendo conhecida a restrição do sistema, a utilização dos recursos foi calculada com base na demanda de cada produto. Considerando que uma unidade de tempo da simulação seja equivalente a 1 hora, o intervalo de tempo usado para o cálculo da utilização dos recursos foi de 72 horas, representando 3 dias de operação ininterrupta. Simultaneamente, esse intervalo foi empregado como intervalo de reposição para confecção das ordens de produção.

O sequenciamento proposto por Atwater e Chakravorty (2002) atende o objetivo do modelo base de simulação, no entanto deixa de atender a demanda com a adição da preparação de máquina. No modelo base, o recurso 9 é considerado o recurso com capacidade restritiva, operando com 94% de utilização. Implementando-se as preparações e mantendo-se a lógica de sequenciamento, o recurso 9 precisa ampliar sua capacidade em 34% e operar 100% do tempo para atender a demanda (Tabela 5). Assim, mudanças na operação devem ser executadas para que a fábrica atenda a demanda sem ampliar sua capacidade.

Tabela 5 – Tempo de processamento e utilização e dos recursos

	Tempo de processamento			Sequenciamento unitário		
Recurso	Distribuição	μ	σ	Produtos por intervalo (média)	Utilização sem preparação	Utilização com preparação
1	Gamma	0,780	0,450	72,0	78,0%	78,0%
2	Gamma	0,363	0,209	180,0	90,8%	90,8%
3	Gamma	2,667	1,885	21,6	80,0%	86,0%
4	Gamma	1,057	1,057	50,4	74,0%	88,0%
5	Gamma	2,000	1,155	28,8	80,0%	88,0%
6	Gamma	1,400	0,700	36,0	70,0%	80,0%
7	Gamma	1,775	1,025	36,0	88,8%	88,8%
8	Gamma	1,875	1,083	28,8	75,0%	83,0%
9	Gamma	1,175	0,831	57,6	94,0%	134,0%
10	Gamma	1,800	1,039	28,8	72,0%	80,0%
11	Gamma	1,440	1,440	36,0	72,0%	82,0%
12	Gamma	1,157	0,579	50,4	81,0%	81,0%
13	Gamma	1,450	0,837	43,2	87,0%	87,0%

Fonte: Adaptado de Atwater e Chakravorty (2002)

A principal mudança proposta é a troca do modo de operação da fábrica, a qual opera em formato sob encomenda (MTO — *make-to-order*), para um formato de fabricação para estoque (MTS – *make-to-stock*). Como no modelo base de simulação a demanda é totalmente atendida (Chakravorty; Atwater, 2005), essa alteração permite que a fábrica agrupe ordens de produção e mantenha estoques suficientes para atender a demanda sem alterar a capacidade de produção.

4.4.2.2 Implementação do DBR

O pulmão de expedição (SB) é responsável por proteger a disponibilidade dos produtos. A unidade de medida do SB é em peças, de modo que seu nível representa a soma do estoque de acabados (FG) com a quantidade de produtos em produção (WIP) (Schragenheim, 2010). O pulmão da restrição (CB – *Constraint Buffer*) atua como limitador de liberação de materiais, além de proteger a restrição contra variabilidades do sistema (Schragenheim, 2010). A atividade de sequenciamento deve não apenas sequenciar as ordens de produção, mas também definir a quantidade de cada produto (Castro; Godinho-Filho; Tavares-Neto, 2022; Schragenheim, 2010). Assim, as ordens são geradas a cada intervalo de reposição (72 horas), conforme a diferença entre o nível do SB e seu valor de referência, para cada produto. O sequenciamento das ordens ocorre conforme a relação entre a penetração do SB e seu valor de referência, representando a prioridade da ordem. A penetração é definida pela diferença entre o valor de referência e o nível do estoque de produtos acabados (FG). Assim, quanto maior é a penetração do SB, maior é a prioridade da ordem, e antes ela é liberada. Entretanto, nem todas as ordens são liberadas para produção, pois elevariam a quantidade de setup a ponto de comprometer a disponibilidade da restrição (Schragenheim, 2010). Para lidar com esse problema, o tempo disponível da restrição até o próximo intervalo de reposição e o pulmão da restrição são utilizados como linha de corte, de modo que as ordens são liberadas enquanto houver tempo disponível na restrição e enquanto não excederem o pulmão da restrição.

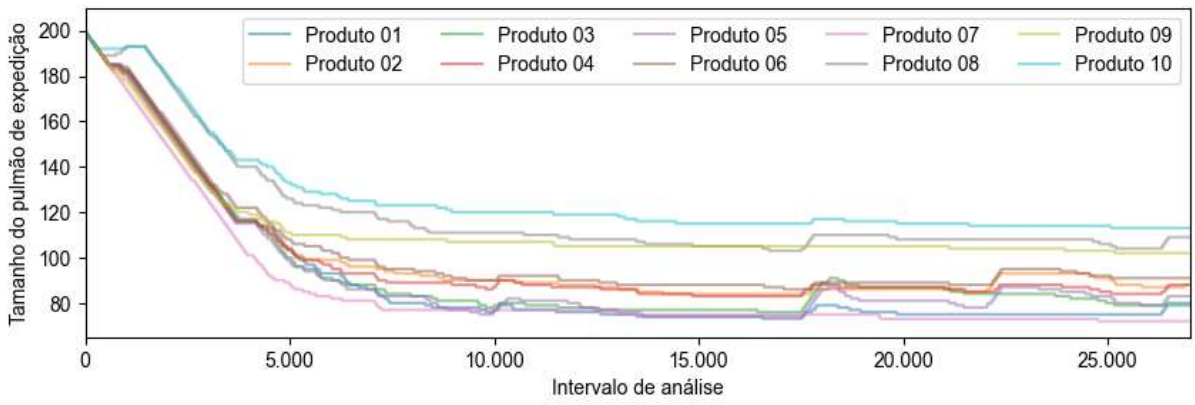
4.4.2.3 Implementação do BM

A definição do tamanho do pulmão de expedição (SB) é crucial por impactar diretamente a prioridade da ordem. Assim, o valor inicial de metade do lead-time deve ser ajustado ao longo das interações (Schragenheim, 2010; Schragenheim; Ronen, 1990). O valor de referência do SB é elevado quando, no intervalo de reposição, o total de penetração for maior do que a zona vermelha do pulmão, assim como o valor de referência do SB é reduzido quando estiver na zona verde por um tempo igual ou maior do que o dobro do intervalo de reposição (Schragenheim, 2010). Schragenheim (2010) sugere que, quando necessário, o pulmão seja elevado 20% ou reduzido 15%, entretanto, como o objetivo é obter o valor ideal, utilizou-se o valor de uma unidade do pulmão.

4.4.2.4 Otimização dos pulmões

A otimização dos pulmões visou identificar os valores ideias dos pulmões de expedição (SB) e do pulmão da restrição (CB). O SB foi otimizado acoplando-se o sequenciamento do DBR e do BM (seção 4.2.3) no modelo de simulação da fábrica, que foi configurado para iniciar com 200 unidades durante a simulação de 27.500 intervalos de reposição (referente a 82.500 dias). A Figura 8 apresenta o decaimento dos pulmões ao longo dessa etapa. Após a simulação, os valores finais de cada pulmão foram configurados como iniciais para um novo experimento. Nele, os valores máximos e mínimos dos pulmões foram registrados, e as médias arredondadas foram utilizadas como valores definitivos (Tabela 6).

Figura 8 - Ajuste dos pulmões de expedição ao longo das interações



Fonte: Elaborado pelo autor

Tabela 6 - Valores ajustados para os pulmões de expedição

Pulmão	Amostras	Média	Desvio Pad.	Min	Max	Valor definido
Produto 01	27.500	73,92	2,52	72,00	82,00	74,00
Produto 02	27.500	83,76	3,57	79,00	90,00	84,00
Produto 03	27.500	77,20	3,30	73,00	86,00	78,00
Produto 04	27.500	78,78	1,01	77,00	82,00	79,00
Produto 05	27.500	75,59	3,95	72,00	84,00	76,00
Produto 06	27.500	87,95	2,72	85,00	94,00	88,00
Produto 07	27.500	69,00	0,00	69,00	69,00	69,00
Produto 08	27.500	104,45	3,65	99,00	111,00	105,00
Produto 09	27.500	89,00	0,00	89,00	89,00	89,00
Produto 10	27.500	112,78	1,65	111,00	118,00	113,00

Fonte: Elaborado pelo autor

Com os pulmões de expedição ajustados, procedeu-se ao acerto do pulmão da restrição. Para essa etapa, a quantidade de produtos vendidos (SP) e o inventário (IV) foram utilizados como métricas de desempenho. Conduziu-se experimentos para os níveis de pulmão de 250 (E1), 300 (E2), 350 (E3) e 400 (E4) unidades de tempo, com 33 replicações cada. Cada replicação rodou por 800.000 unidades de tempo, com dados coletados para as últimas 400.000. A distribuição de probabilidade *t Students* foi utilizada para quantificar o tamanho da amostra e o erro associado às métricas, com intervalo de confiança de 95% e 5% de significância (Law, 2015). O maior tamanho de amostra calculado é 11, relativo aos produtos vendidos, com o pulmão de 250 unidades. Logo, as 33 replicações são suficientes para o experimento. A fim de confirmar a existência de diferença estatística entre os experimentos, foram aplicados testes de hipótese em pares sucessivos (Tabela 7). Os resultados demonstraram que, para todos os experimentos, a demanda foi a mesma. Entretanto, o experimento E1 vendeu menos produtos do que os demais, indicando que o valor de 250 unidades para o pulmão da restrição não é suficiente. A respeito do inventário, os testes indicam que os experimentos são estatisticamente diferentes. Assim, por atender a demanda com o menor nível de inventário, o pulmão da restrição foi configurado para 300 unidades (E2).

Tabela 7 - Teste de hipótese nas métricas

	Teste de hipótese					
	h0: E1=E2 h1: E1≠E2		h0: E2=E3 h1: E2≠E3		h0: E3=E4 h1: E3≠E4	
Métrica	Resultado	p-valor	Resultado	p-valor	Resultado	p-valor
Demanda	h0*	0,28042	h0	0,37805	h0	0,52638
Produtos vendidos	h1*	≈0,00000	h0	0,27492	h0	0,36908
Inventário	h1*	≈0,00000	h1*	0,00003	h1*	0,00254

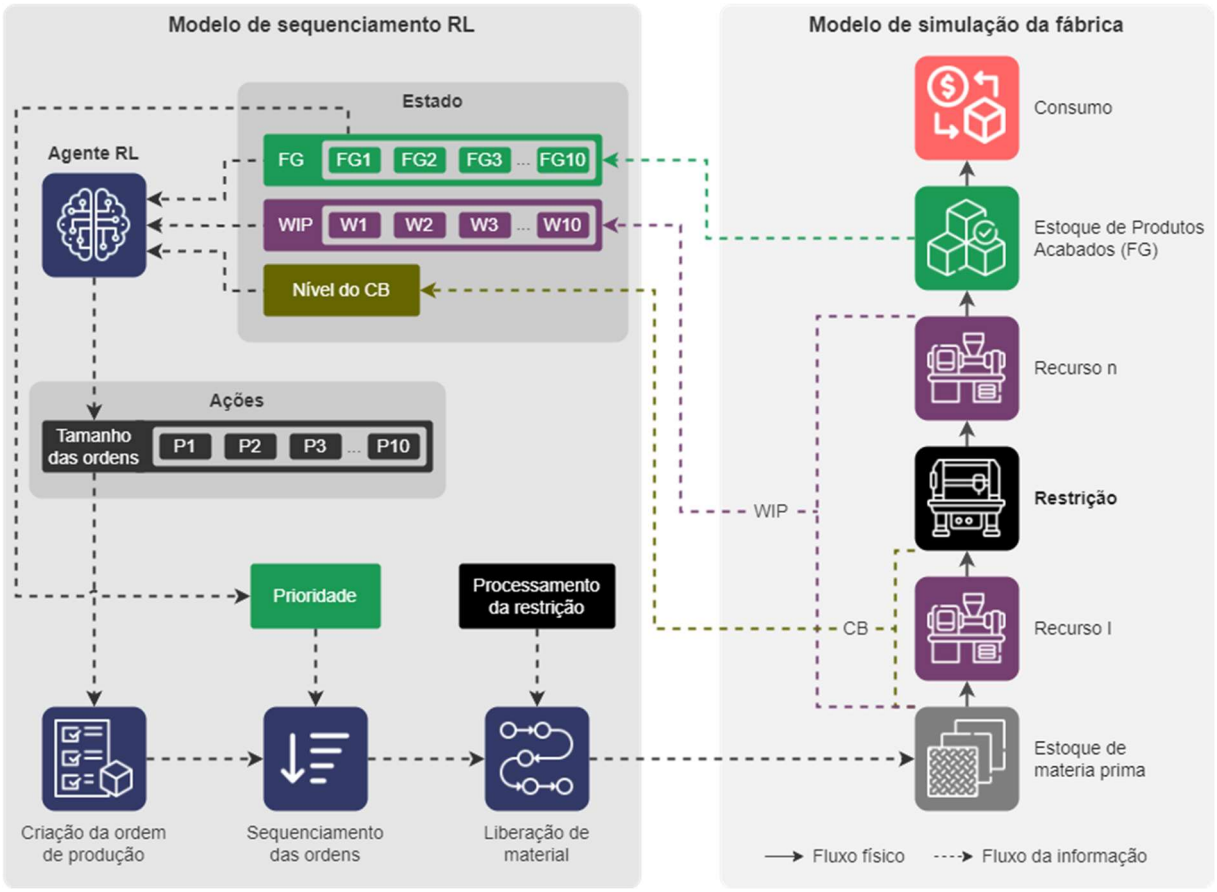
*Wilcox para amostras não pareadas

Fonte: Elaborado pelo autor.

4.4.3 Modelo de sequenciamento por aprendizagem por reforço

Os dois principais elementos para o desenvolvimento de uma abordagem por RL são: o agente e o ambiente (Devanga; Badilla; Dehghanimohammadabadi, 2022a). A Figura 9 ilustra a interação entre o agente de RL e o modelo de simulação da fábrica (ambiente). O agente foi desenvolvido com o algoritmo *Proximal Policy Optimization* (PPO) (Schulman *et al.*, 2017), implementado pela biblioteca Stable Baselines 3 (Raffin *et al.*, 2021). No que diz respeito ao ambiente, foi utilizado o modelo de simulação descrito na Seção 4.1. Como o ambiente dita a dinâmica do sistema, sua modelagem deve contemplar os elementos do MDP para permitir a aprendizagem da política que leva ao maior retorno (Sutton; Barto, 2018). Assim, as próximas seções detalham os elementos do MDF, que são o estado, as ações e a função de recompensa.

Figura 9 – Modelo de sequenciamento por RL



Fonte: Elaborado pelo autor

4.4.3.1 Modelagem do estado

A modelagem do estado foi elaborada conforme as variáveis utilizadas para sequenciamento do DBR (Schragenheim, 2010). Assim, é possível estabelecer uma relação entre as decisões do modelo de RL e do DBR. Ao todo, o estado é composto por três variáveis, representadas em um vetor de 21 posições. A primeira posição representa o pulmão da restrição, e é obtida pela soma do tempo de processamento na restrição das ordens liberadas que não passaram na restrição, acrescido do tempo de preparação para cada ordem. A quantidade de produtos em processamento para cada item é representada pelas posições de 2 a 11. Por fim, a quantidade de produtos acabados disponíveis para venda é representada pelas posições de 12 a 21. A elucidação do estado está presente na Figura 9.

4.4.3.2 Modelagem da ação

O agente de RL deve definir a quantidade de cada produto do sistema que será produzida. A quantidade de cada produto é determinística entre 0 (zero) e 20 unidades, sendo que o 0 (zero) representa não produzir nada. Assim, o espaço de ações representa uma matriz de 10x21, sendo 21 posições de escolha (0 a 20) para cada produto (10 produtos). O valor de 20 unidades foi definido com base nos resultados iniciais do DBR e por representar um valor suficientemente grande para o volume de ordens em comparação com a demanda. O agente de RL não utiliza pulmões, pois define a quantidade de cada produto para produção. Entretanto, o conceito do SB foi utilizado para sequenciar as ordens elaboradas pelo agente. Considerando que o SB de cada produto representa a soma das ordens em produção com o nível do estoque de produtos finalizados, as ordens definidas pelo agente são sequenciadas pelo menor nível do respectivo SB. Assim, as ordens liberadas em primeiro lugar são referente aos produtos em menor quantidade no sistema, e as demais ordens são liberadas conforme o ritmo de produção da restrição (Goldratt, 2006).

4.4.3.3 Modelagem da recompensa

A função de recompensa é a função que o agente de RL irá maximizar, devendo representar o objetivo de longo prazo (Sutton; Barto, 2018). Assim, a função modelada representa duas das três métricas de desempenho propostas pela TOC: (i) ganho e (ii) inventário. A função com base na terceira métrica da TOC (despesa operacional) não foi modelada, pois, além de o ambiente ser um cenário hipotético sem despesas reais, maximizar a produção reduzindo inventário é suficiente para a tomada de decisão. Dessa forma, a função de recompensa (R) é representada pela equação 1.

$$R = \frac{\sum_{i=1}^{q} 1 - W(wp_i) - F(fg_i) - P(l_i)}{q} \qquad (1),$$

sendo que $q = 10$ representa a quantidade de produtos produzidos na fábrica e $W(wp_p)$ representa a penalidade associada aos produtos em produção, sendo calculada pela equação 2. wp_p é a quantidade atual do produto p em produção e pt_p

é o tempo de processamento sem esperas para o produto p. wp_p é dividido por pt_p com objetivo de normalizar a penalidade entre os produtos, assim, produtos que necessitam de mais tempo de produção podem operar com um wp_p maior. Eleva-se toda a equação ao quadrado para reduzir a variação em valores baixos de wp_p e proporcionar penalidades maiores em valores elevados de wp_p. A mesma lógica se aplica para a penalidade associada ao estoque de produtos acabados $F(fg_p)$. Esta é calculada pela equação 3, em que fg_p é a quantidade de produtos acabados no estoque para o produto p. A penalização para os produtos acabados é menor, a fim de permitir que o agente mantenha níveis maiores em favor de não perder vendas.

$$W(wp_p) = \left(\frac{0.02 \times wp_p}{pt_p}\right)^2 \quad (2)$$

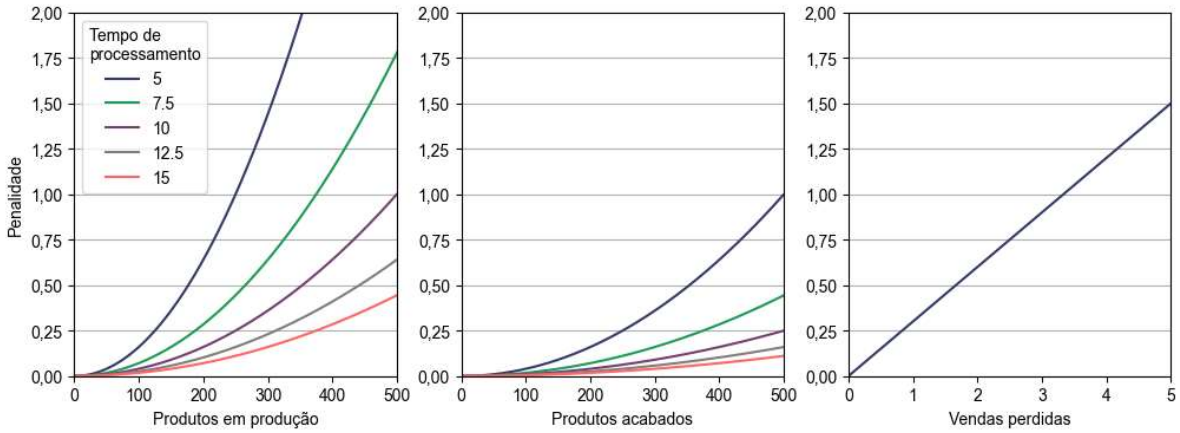
$$F(fg_p) = \left(\frac{0.01 \times fg_p}{pt_p}\right)^2 \quad (3)$$

Por fim, $P(l_p)$ representa a penalidade associada à perda de vendas, e é calculada pela Equação 4, em que l_p é a quantidade de vendas perdidas. Como o objetivo é garantir a disponibilidade de produtos (Schragenheim, 2010), a máxima produtividade é associada a não perder nenhuma venda. Nesse contexto, uma venda perdida se configura quando um pedido para um determinado produto é recebido e não há estoque do produto disponível.

$$P(l_p) = 0.3 \times l_p \quad (4)$$

A Figura 10 elucida o comportamento de cada componente de penalidade, ilustrando o impacto que o tempo de processamento total dos produtos exerce sobre os produtos em produção e os produtos acabados. Adotou-se essa abordagem pois recompensas entre os valores 0 e 1 apresentam melhor desempenho para algoritmos baseados em redes neurais (Russel; Norvig, 2010; Sutton; Barto, 2018). Para que a recompensa não exceda o limite inferior, sempre que o agente atingir uma recompensa igual a 0 (zero), o ambiente é reiniciado.

Figura 10 - Comportamento dos componentes da função de recompensa



Fonte: Elaborado pelo autor

4.4.3.4 Treinamento do agente

Ao interagir com o ambiente de simulação, o agente de RL otimiza sua política a fim de maximizar a recompensa (Kuhnle *et al.*, 2022; Schulman *et al.*, 2017). Cada interação do agente representa um intervalo de reposição (72 horas), de modo que, primeiro, ele recebe o estado do ambiente para, então, tomar uma ação. A simulação é reproduzida até o próximo intervalo de reposição com a ação tomada pelo agente e, a partir disso, a recompensa associada ao estado-ação é calculada. A recompensa e um novo estado são enviados ao agente, sendo que a recompensa é armazenada para o aprendizado do agente, e o novo estado é utilizado para uma nova tomada de decisão (Sutton; Barto, 2018).

O agente foi treinado com o algoritmo *Proximal Policy Optimization* (PPO) (Schulman *et al.*, 2017). O PPO é um algoritmo que otimiza sua política por meio de pequenos incrementos, impedindo que o aprendizado degrade ou convirja para um ótimo local (Vanvuchelen; Gijsbrechts; Boute, 2020). O processo de aprendizagem consiste em interagir com o ambiente por determinado volume de interações, armazenando os estados, as ações e as recompensas. Os dados armazenados são utilizados para o aprendizado do agente e para a atualização da política. Essa característica faz do PPO um algoritmo capaz de lidar com recompensas de longo prazo, uma vez que ele não utiliza estados isolados para o aprendizado, mas todo o percurso trilhado durante as interações (Wu *et al.*, 2024).

A versão do algoritmo utilizada foi implementada pelo repositório Stable Baselines3 (Raffin *et al.*, 2021), que emprega redes neurais para aproximar a política

e a função de valor (Russel; Norvig, 2010; Schulman *et al.*, 2017). No PPO, a atualização da política ocorre em intervalos de interações definidos, sendo necessário que o ambiente represente episódios (Raffin *et al.*, 2021). Um episódio é representado pelo intervalo entre o início e o fim da simulação da fábrica, sendo que um novo se inicia quando o anterior é finalizado. Assim, durante o treinamento, a simulação foi configurada para finalizar quando o agente receber uma recompensa igual a zero ou ao atingir 13.888 interações (1.000.000 unidades de tempo). Logo, a atualização da política foi configurada para 122.880 interações. A fim de se reduzir o tempo de treinamento, foram utilizadas 15 simulações paralelas (multiprocessamento). Os parâmetros usados para o treinamento são apresentados na Tabela 8. Foram empregados valores padronizados para implementação de parâmetros não citados, os quais podem ser consultados no repositório (Raffin *et al.*, 2021).

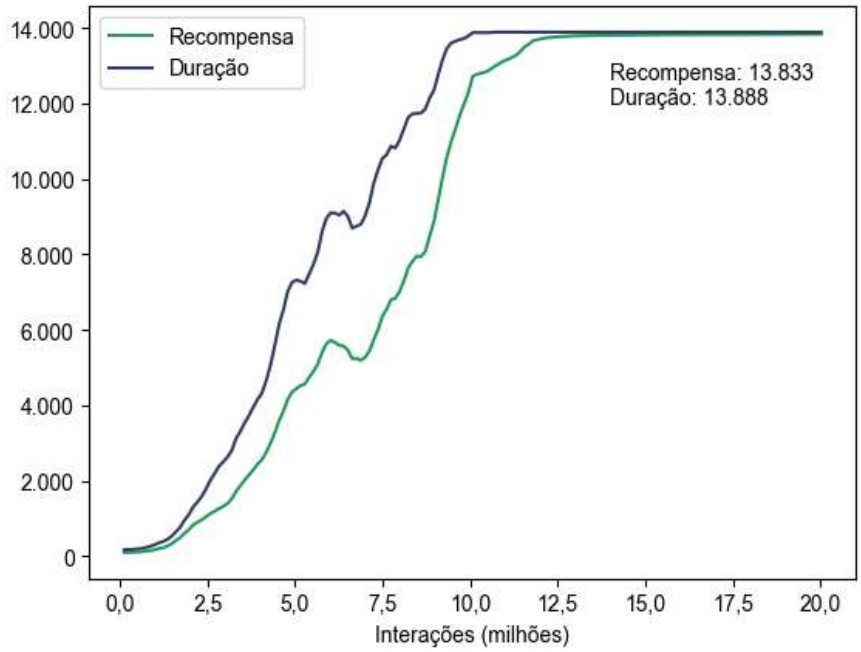
Tabela 8 - Parâmetros de treinamento do agente

Parâmetro	Valor
Taxa de aprendizado (*Learning rate*)	0.0002
Batch	1.024
Épocas	10
Gamma	0.99
Ambientes em paralelo	15
Interações entre atualização	122.880

Fonte: Elaborado pelo autor

O treinamento foi executado com um processador i5-13450HX, com 16 GB de memória e placa gráfica NVIDIA Geforce RTX 4050, por 20.000.000 de interações. A Figura 11 apresenta a evolução da duração e da recompensa média, por episódio, ao longo das duas horas de treinamento. Como a recompensa máxima de cada interação é igual a 1, o máximo de recompensa que o agente pode adquirir por episódio é a própria duração (13.888 interações). Assim, ao final do treinamento, a média da recompensa obtida pelo agente foi de 0,996.

Figura 11 – Avanço na duração e recompensa por episódio no treinamento.



Fonte: Elaborado pelo autor

4.5 SIMULAÇÃO DOS CENÁRIOS

Tanto o método de sequenciamento do DBR quanto o agente de RL treinado foram simulados. Cada cenário foi simulado por 50 replicações, com 800.000 unidades de tempo cada, sendo a primeira metade considerada aquecimento. Assim, os dados foram coletados durante a segunda metade da simulação. Para cada métrica de desempenho, foi calculado o erro e o tamanho da amostra com base na distribuição de *t Student,* com intervalo de confiança de 95% e 5% de significância. Entre as métricas, o tamanho máximo calculado para amostragem para o DBR e para o RL foram, respectivamente, 4 e 2, logo, as 50 replicações foram suficientes para representar a distribuição da população. Os resultados dos cenários são apresentados na Tabela 9.

Tabela 9 – Resultado dos cenários simulados

Métrica		Média	Desvio Padrão	Erro (95%)	Erro/Média	Tamanho amostra
Tempo de atravessamento (FT)	DRB	203,157	10,838	2,570	0,01265	4
	RL	180,570	3,048	0,723	0,00400	1
Tamanho do lote (LS)	DBR	7,370	0,022	0,005	0,00072	1
	RL	7,486	0,015	0,004	0,00048	1
Produtos em processamento (WIP)	DBR	209,071	11,258	2,669	0,01277	4
	RL	171,507	2,998	0,711	0,00414	1
Produtos acabados em estoque (FG)	DBR	577,531	11,433	2,711	0,00469	1
	RL	338,656	13,696	3,247	0,00959	2
Inventário total (IV)	DBR	786,602	0,225	0,053	0,00007	1
	RL	510,163	10,831	2,568	0,00503	1
Produtos vendidos (SP)	DBR	399.913,860	321,188	76,154	0,00019	1
	RL	399.967,700	314,344	74,531	0,00019	1
Demanda (DM)	DBR	399.938,360	326,313	77,369	0,00019	1
	RL	400.053,340	357,060	84,659	0,00021	1

Fonte: Elaborado pelo autor (2024)

4.6 ANÁLISE DOS RESULTADOS

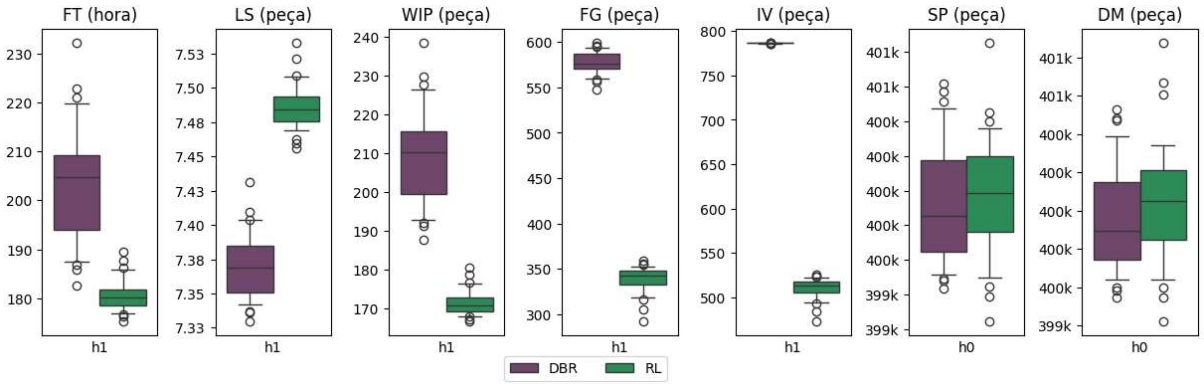
Os resultados foram analisados em duas etapas: (i) análise de produtividade e (ii) análise de comportamento. A análise de produtividade visa comparar o desempenho produtivo entre as duas abordagens de sequenciamento, enquanto a análise de comportamento visa explicar como os resultados foram alcançados.

4.6.1 Análise da produtividade

A análise da produção utilizou testes de hipótese para verificar se há diferenças estatísticas significativas entre o desempenho de cada método de sequenciamento (Figura 12). O teste de hipótese verificou a hipótese nula de as amostras de cada método de sequenciamento pertencerem à mesma distribuição de probabilidade. Foi utilizado o teste de Wilcoxon para amostras não pareadas, com 95% de confiança e 5% de significância. Ainda que o teste mais utilizado para comparação de médias seja o teste-t para duas amostras (Law, 2015), seus pressupostos de normalmente e homoscedasticidade devem ser atendidos (Yamane, 1973). Assim, o teste de Wilcoxon foi usado, pois uma ou mais amostras, em cada avaliação, não passaram nos testes de Shapiro-Wilk (normalidade) ou

Levene (homoscedasticidade) (Telles *et al.*, 2020). Os resultados detalhados são expostos na Tabela 19 do Apêndice II.

Figura 12 – Comparativo de produtividade entre métodos de sequenciamento



Fonte: Elaborado pelo autor.

O resultado evidencia que, estatisticamente, não há diferença significativa na demanda (DM) e na quantidade de produtos vendidos (SP), indicando que ambas as abordagens atenderam igualmente à demanda. Entretanto, o tempo de atravessamento (FT), os produtos em processamento (WIP), os produtos acabados em estoque (FG) e o inventário total (IV) são inferiores no RL.

4.6.2 Análise de comportamento

A análise de comportamento utilizou três técnicas para avaliar a relação entre as variáveis do estado e as decisões dos métodos de sequenciamento. A primeira técnica foi a correlação de Spearman, a qual avalia a relação monotônica entre duas variáveis quantitativas. Essa técnica foi utilizada, pois, diferentemente da correlação de Pearson, não requer que as relações sejam lineares. Em vez disso, avalia se as variáveis tendem a aumentar ou a diminuir juntas, independentemente da sua relação (Barbosa; Azevedo, 2018; Xiao *et al.*, 2015). A Figura 16 e a Figura 17 do Apêndice II apresentam, respectivamente, a matriz de correlação para o DBR e para o RL.

A análise de correlação evidencia que o DBR tem relação negativa entre os níveis de FG e WIP para um mesmo produto. Esse comportamento é um efeito do DBR, que mantém o SB no valor de referência indiferentemente do estágio de

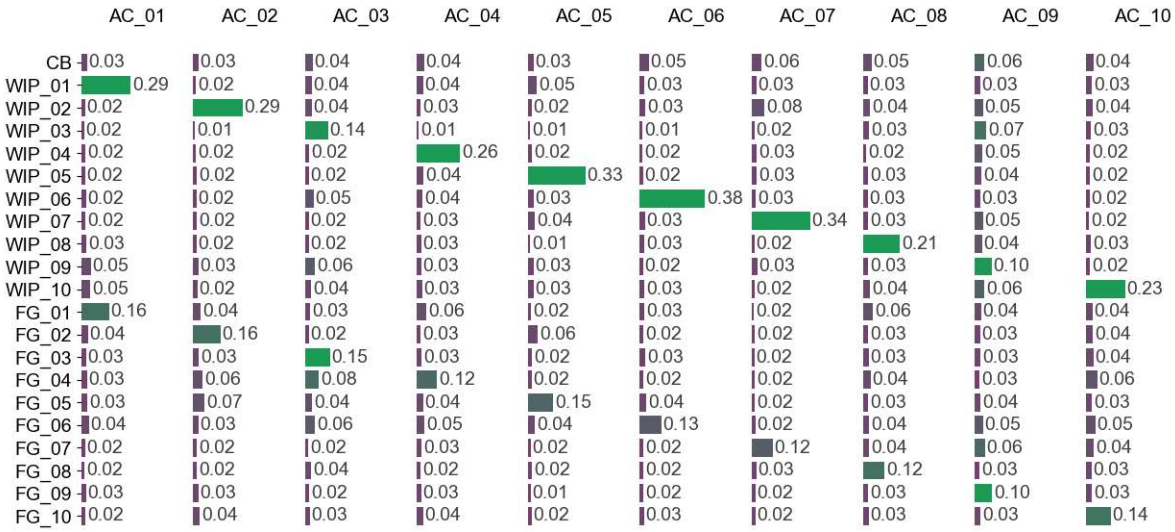
produção dos produtos, proporcionando equilíbrio entre FG e WIP. Os níveis de FG e WIP entre diferentes produtos têm comportamento similar, com exceção dos produtos que não são processados pela restrição (7 e 9). Isso ocorre devido ao limite de liberação de materiais nos produtos processados pela restrição, enquanto os produtos 7 e 9 são liberados conforme o consumo. As ações não evidenciaram correlações significativas com as variáveis, com exceção da 2, 6 e 7, que têm correlação fraca com os respectivos FG, assim como a ação 9 em relação ao respectivo WIP. Referente ao agente de RL, foi evidenciado comportamento similar ao DBR, tanto com respeito à relação do FG com o WIP quanto ao comportamento dos produtos 7 e 9. Entretanto, a intensidade da correlação é menor, indicando que apesar de o agente de RL buscar manter os níveis de FG e WIP equilibrados, outros fatores influenciam na decisão das ações. Diferente das ações do DBR, o agente de RL tem correlações negativas fracas e moderadas com os respectivos WIP, com exceção da ação 9. A correlação negativa das ações com o WIP indica que o agente de RL impõe prioridade maior para o WIP do que para o FG, liberando material conforme o WIP é reduzido. Esse fenômeno não é identificado no DBR.

A segunda abordagem utilizou a regressão linear para identificar a contribuição do WIP e do FG sobre a quantidade de produtos liberados pelos métodos de sequenciamento. Foi conduzida uma regressão linear para cada produto, utilizando o estado como variável independente e a ação do respectivo produto como variável dependente. O resultado da análise é apresentado no Apêndice II, na Tabela 20 para o DBR, e na Tabela 21 para o RL. O resultado evidencia que as ações do DBR têm relação linear com as variáveis do estado, apresentando coeficiente de determinação superior a 80%. Entretanto, o RL não compartilha o mesmo resultado, evidenciando coeficientes de determinação abaixo de 55%. Para ambos os métodos de sequenciamento, as variáveis que influenciam a tomada de decisão são WIP e FG dos respectivos produtos analisados, sendo que o CB não impacta na tomada de decisão. Porém, ao se comparar a influência do WIP e do FG em relação ao DBR e ao RL, a influência é maior no DBR.

Com o objetivo de complementar a análise e tornar as decisões do RL legíveis, um modelo de Árvore de Decisão foi treinado para cada ação do histórico de estado-ação do agente de RL. Similar à regressão linear, a Árvore de Decisão do DBR evidenciou que o WIP e o FG são as variáveis de maior influência para decisão sobre a quantidade de produtos. Ambas as variáveis representam, juntas, mais de

85% do impacto nas decisões. Sobre o RL, o WIP apresentou maior influência do que o FG, porém o impacto conjunto é inferior a 50% em todas as ações (Figura 13). Diferente do DBR, em que as variáveis de outros produtos têm baixa ou nenhuma interferência na decisão, no RL são evidenciados valores próximos a 5%.

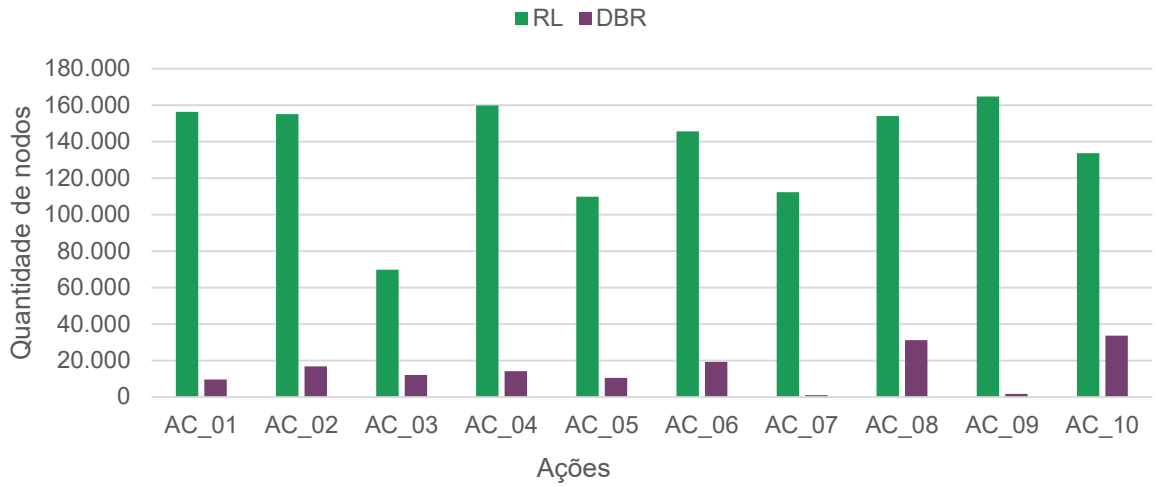
Figura 13 - Influência das variáveis nas ações do RL, pela árvore de decisão



Fonte: Elaborado pelo autor

A quantidade de nodos da árvore de decisão indica o total de decisões necessárias para explicar o conjunto de dados (Myles *et al.*, 2004; Rudin *et al.*, 2022). Nesse contexto, a árvore do DBR foi ajustada com menos nodos do que o RL (Figura 14 e Tabela 22, Apêndice II), indicando que as decisões tomadas pelo RL são mais complexas do que as do DBR. No DBR, as ações dos produtos 7 e 9 evidenciam a menor complexidade, pois, por serem esses os produtos não processados pela restrição, a liberação de material considera apenas o respectivo SB. Em oposição, as ações dos produtos 8 e 10 apresentam a maior quantidade de nodos, representando as decisões mais complexas do DBR, o que corrobora o fato de esses serem os produtos processados pela restrição com a maior quantidade de processos (Figura 6). Entretanto, no RL não é evidenciado um padrão ou uma relação entre a quantidade de nodos e a dinâmica do modelo de simulação (Tabela 23, Apêndice II). Inclusive, o produto 9 contém a maior quantidade de nodos mesmo não sendo processado pela restrição.

Figura 14 - Quantidade de nodos da árvore de decisão



Fonte: Elaborado pelo autor

4.7 DISCUSSÕES

A integração entre RL e DBR reduziu o inventário e o tempo de atravessamento, permitindo elevar a eficiência do sistema produtivo pela decomposição do SB em WIP e FG. Do ponto de vista da TOC, a redução do inventário contribui para elevar o retorno sobre o investimento (ROI – *Return on investment*) e melhorar o fluxo de caixa (Gupta; Ko; Min, 2002; Machado *et al.*, 2023). Ademais, colabora para diluir a variabilidade do sistema ao reduzir as filas em frente aos recursos (Goldratt, 2006; Machado *et al.*, 2023). Consequentemente, filas menores impactam positivamente o lead-time, permitindo respostas mais rápidas às oscilações de demanda e a eventuais paradas de máquinas (Thürer; Fernandes; Stevenson, 2022). A decomposição do SB utiliza o nível atual do WIP e do FG para tomada de decisão, descartando a necessidade de um valor de referência para o SB e, por consequência, de ajuste constante do pulmão. Sob o ponto de vista empresarial, ao desconsiderar os pulmões, pode-se reduzir a carga de trabalho para análise e tomada de decisão e elevar a eficiência do sistema produtivo, ao passo que se mantém decisões assertivas.

Os conceitos da TOC auxiliaram a modelagem do RL, tanto para o estado quanto para a função de recompensa. Uma vez que o estado deve ser suficiente

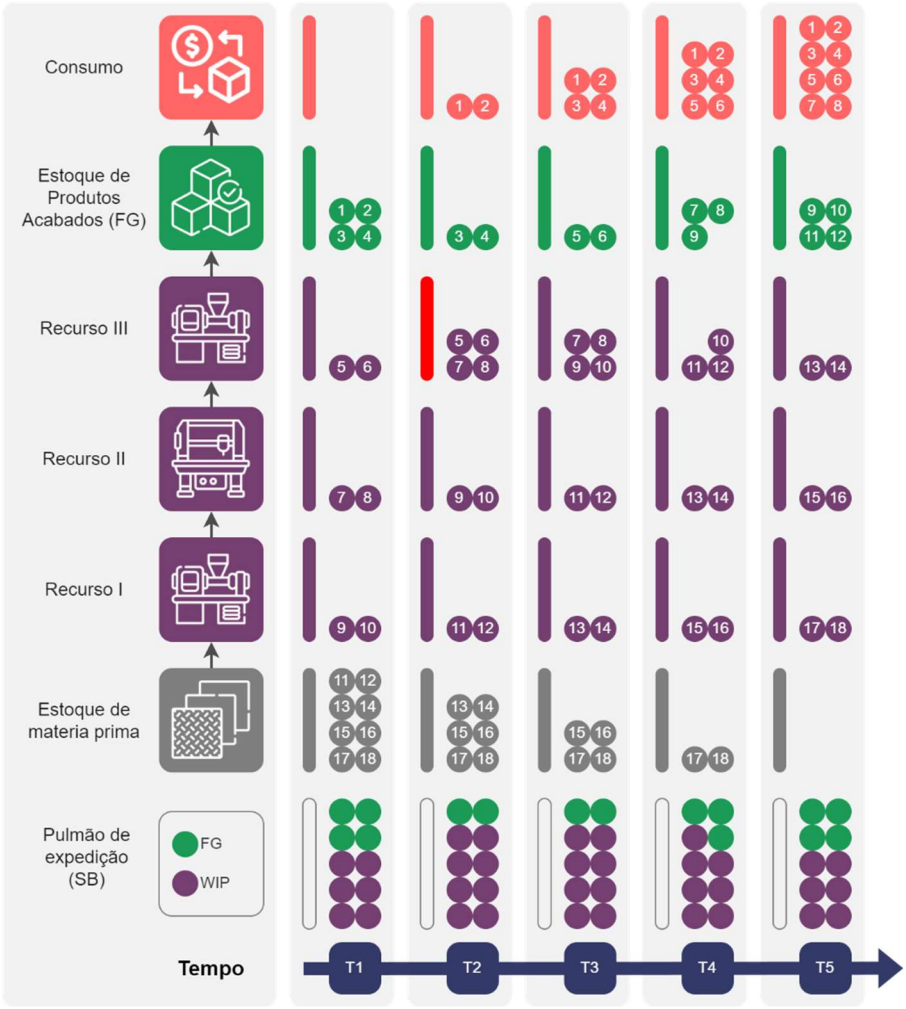
para a tomada de decisão do agente (Sutton; Barto, 2018), ao utilizar o conceito do SB, o agente de RL consegue compreender a necessidade produtiva do sistema e atender a demanda. Essa abordagem descarta a necessidade de monitorar o sistema produtivo por completo, como, por exemplo, os produtos na fila (Shiue; Lee; Su, 2020; Thomas *et al.*, 2018), as características da ordem de produção (Liu; Piplani; Toro, 2022; Marchesano *et al.*, 2022), a utilização e a condição dos recursos (Liu *et al.*, 2022; Sakr *et al.*, 2023). Ainda nesse contexto, a função de recompensa elaborada com as métricas de desempenho da TOC (ganho e inventário) é suficiente para o aprendizado do algoritmo de RL, demonstrando sua eficácia na tomada de decisões.

As aplicações do RL como método de sequenciamento limitam-se a sequenciar ordens de produção definidas previamente (Esteso *et al.*, 2023). Entretanto, a integração com o DBR possibilita que o agente defina o tamanho das ordens, enquanto o sequenciamento ocorre com base no volume de produtos acabados no estoque. Essa abordagem permite utilizar o RL em um ambiente MTS, diferente das tradicionais abordagens concentradas em MTO (Panzer; Bender, 2022). Da mesma forma, eleva a autonomia do sequenciamento para tomada de decisão autônoma no contexto da indústria 4.0 (Jan *et al.*, 2023; Oztemel; Gursev, 2020). A aplicação do DBR como método de sequenciamento do modelo de simulação foi eficaz, porém a integração com RL atingiu a mesma produtividade com menor inventário e tempo de atravessamento (Figura 12).

Tanto a análise de correlação quanto a regressão linear e a árvore de decisão foram empregadas visando explicar as decisões do agente de RL, sendo aplicadas ao DBR para comparação com um cenário conhecido. Assim, ambas as técnicas evidenciaram o comportamento do DBR conforme a sua modelagem, ressaltando a influência do SB na definição da quantidade de produtos na ordem (Schragenheim, 2010). No contexto do RL, o WIP apresentou maior influência do que o FG na tomada de decisão. Esse resultado pode ser explicado pela característica do RL de aprender a dinâmica do sistema, permitindo maior sensibilidade em relação à tomada de decisão e a perturbações no fluxo, como, por exemplo, quebras de máquinas. Sempre que um recurso está inoperante, as ordens não são processadas e, por consequência, o WIP não é reduzido. O agente de RL pode interpretar esse cenário e definir quantidades menores para a próxima ordem. No DBR isso não ocorre, pois as ordens são definidas pela diferença entre o valor de referência e o

nível atual do SB, tendendo a estabilidade sempre próximo ao valor de referência. Assim, quando um recurso está inoperante, o mercado continua a solicitar produtos e o nível do SB é reduzido, logo, as ordens são elaboradas para elevar o nível do SB. Esse fenômeno é ilustrado na Figura 15, de modo que, quando o recurso III fica inoperante (barra vermelha, no tempo T2), peças são acumuladas (5, 6, 7 e 8) e o mercado consome o FG. Com a redução do FG, uma nova ordem é liberada para equalizar o nível do SB e, consequentemente, o WIP é elevado. Nos momentos T3, T4 e T5, o nível de WIP e FG retornam ao inicial devido à capacidade protetiva dos recursos, e os produtos continuam a ser liberados conforme o nível do SB.

Figura 15 - Quebra de máquinas no DBR.



Fonte: Elaborado pelo autor

Ao evidenciar o desempenho do RL, pode-se afirmar que há espaço para aprimoramentos no DBR, em especial, considerando o WIP como uma variável de tomada de decisão e não apenas como parte do SB. Ainda que não seja possível indicar a relação direta entre o WIP e a quantidade de produtos liberados, a quantidade de nodos da árvore de decisões evidencia uma diferença expressiva entre os dois métodos de sequenciamento. Essa diferença pode indicar que outras variáveis ou combinações sejam responsáveis pelas decisões em determinados estados do sistema, de modo que definir tais relações pode ser uma tarefa exaustiva e, em alguns casos, impossível (Kuhnle *et al.*, 2022; Rudin *et al.*, 2022).

4.8 CONCLUSÕES

Este estudo apresentou uma primeira tentativa de integrar o RL com o DBR para sequenciamento da produção. O modelo de RL foi elaborado conforme os conceitos da TOC e treinado em um ambiente DES. O ambiente foi modelado com base em estudos do DBR, representando um *job-shop* com processos estocásticos. Modificações foram implementadas para avaliar o desempenho da solução em cenários de alta variabilidade, com preparação de máquina e quebra de equipamentos. Os resultados foram comparados com o sequenciamento do DBR, sendo que o RL atingiu a mesma produtividade com menores níveis de inventário.

A principal contribuição desta pesquisa reside na possibilidade de se elevar a eficiência do sequenciamento da produção ao integrar RL e TOC, decompondo o SB em WIP e FG. Simultaneamente, a análise do comportamento do RL indicou o WIP como a principal variável na tomada de decisão para liberação dos produtos, abrindo portas para a exploração de novas abordagens e aprimoramentos no DBR. Ainda que não tenha sido possível identificar a total influência do WIP nas ações do RL, sua integração com o DBR melhorou o sistema produtivo, reduzindo o inventário e o tempo de atravessamento. Ademais, a introdução dos conceitos da TOC na modelagem do RL permitiu a operação do sistema produtivo sem o monitoramento exaustivo da produção, considerando, apenas, o nível de WIP e de FG para cada produto.

Com relação a limitações, o modelo de simulação é teórico, representando simplificações da realidade que desconsideram o custo de operação e, consequentemente, o ganho do sistema (Goldratt, 2006). Assim, não foram

conduzidas análises econômicas, de modo que a priorização do mix de produtos propostos pela TOC não foi avaliada. Adicionalmente, o modo de operação definido como MTS desconsidera a possibilidade de operar em MTO. As oportunidades para trabalhos futuros incluem a implementação da solução em ambientes de maior complexidade, a utilização de dados empíricos, a comparação da abordagem proposta para o modelo de RL com outros modelos e a exploração do DBR utilizando variáveis isoladas para o WIP e FG.

5 DISCUSSÕES

Os resultados dos Capítulos 3 e 4 são apresentados e discutidos. De modo geral, os resultados evidenciam ser possível a integração entre a TOC e o RL. Os conceitos da TOC aplicados a um modelo de RL contribuíram para elevar o desempenho do sequenciamento, reduzindo o tempo de atravessamento, o inventário e a quantidade de ordens em produção. A análise exploratória do comportamento do agente demonstrou que o nível das ordens em produção (WIP – Work in process) é a variável de maior relevância para a tomada de decisão, seguida pelo nível do estoque de produtos acabados. Com o resultado do estudo, conclui-se que: é possível integrar a TOC com o RL para sequenciamento da produção; os conceitos da TOC contribuem para tornar o modelo de RL mais legível a gestores, além de facilitar a introdução do modelo em um ambiente operacional, pois não é necessário monitorar a linha de produção exaustivamente; e a utilização do RL junto ao DBR descarta a necessidade de dimensionamento e monitoramento contínuo dos pulmões.

5.1 DESCOBERTAS DO CAPÍTULO 3

O capítulo 3 revisou a literatura associada ao DBR e ao RL, visando compreender como os elementos dessas áreas podem se relacionar para proporcionar melhorias no desempenho produtivo. A revisão abordou os pontos em que a introdução de conceitos de gestão de operações, em especial da TOC, pode auxiliar na modelagem de agentes de RL. Os principais pontos de sinergia são relacionados aos elementos do RL, sendo eles: (i) as ações; (ii) o estado observável; e (iii) a função de recompensa (Sutton; Barto, 2018).

As ações do agente estão diretamente relacionadas à atividade de sequenciamento, em que é possível definir tanto a serialidade do processo quanto a quantidade a ser produzida (Harjunkoski *et al.*, 2014; Pinedo, 2016-). Entretanto, as ações dos modelos de RL estão focadas em sequenciar a linha de produção, ignorando a possibilidade de redimensionar os lotes. Essa abordagem obriga o agente a associar um volume de trabalho, por vezes desnecessário, a um recurso. Do ponto de vista da TOC, apenas a restrição deve operar constantemente, enquanto outros recursos devem ser ativados conforme o ritmo da restrição para impedir que ela sofra com bloqueios ou inanição (Betterton; Cox, 2009; Goldratt, 2006).

O espaço observável das implementações de RL utiliza a maior quantidade possível de informações da linha. Ainda que tal abordagem possa ser significativa para o agente de RL, ela é viável apenas em cenários virtuais e teóricos, pois

monitorar detalhadamente o sistema produtivo é custoso e, por vezes, impossível. Na prática, para integrar o agente de RL a cenários reais, as variáveis do espaço observável devem ser factíveis de monitoramento.

Quanto à função de recompensa dos modelos de RL, suas métricas não referenciam teorias de gestão, podendo distorcer o aprendizado do modelo e dificultar sua inserção em cenários reais. O principal objetivo dos estudos mapeados é a redução do atraso nas ordens e do tamanho das filas, contudo, mesmo que os modelos de RL proporcionem a redução do inventário, não garantem a elevação do ganho, pois não estão associados à restrição do sistema (Kim *et al.*, 2021; Sakr *et al.*, 2023).

5.2 DESCOBERTAS DO CAPÍTULO 4

No capítulo 4, os conceitos da TOC foram utilizados para modelar um agente de RL com capacidade de sequenciar a produção de um modelo de simulação. Paralelamente, o sequenciamento foi executado pelo método proposto pela TOC, o DBR. Assim, foi possível comparar o desempenho dos métodos de sequenciamento e analisar seus comportamentos.

A principal descoberta desse capítulo foi a capacidade do RL de melhorar o desempenho do sistema produtivo desacoplando o WIP e o FG para tomada de decisão. O resultado foi obtido em um ambiente simulado, modelado como um sistema produtivo operando no formato de produção para estoque (MTS). A abordagem com RL operou com níveis menores de estoque de produtos acabados e em processamento, assim como reduziu o tempo de atravessamento. Essas melhorias contribuem para elevar o retorno sobre o investimento e o fluxo de caixa (Gupta; Ko; Min, 2002). Abordagens tradicionais de RL utilizam a quantidade de peças produzidas ou o tempo de entrega para avaliar o desempenho do modelo (Esteso *et al.*, 2023). No entanto, por se tratar de um sistema produtivo MTS e em função de o objetivo da produção estar associado a garantir a disponibilidade de produtos, a quantidade de peças produzidas será similar para ambos os métodos de sequenciamento.

O resultado alcançado pelo modelo de RL foi explorado com auxílio de técnicas explicativas para modelos de aprendizado de máquina (Kuhnle *et al.*, 2022; Rudin *et al.*, 2022). Durante esse processo, identificou-se o WIP como a variável

mais relevante para a tomada de decisão do agente. Tal descoberta diverge do método proposto pelo DBR, no qual a quantidade de produtos da ordem é definida pela diferença entre o valor de referência do pulmão de expedição e seu valor atual (Schragenheim, 2010).

Como uma primeira abordagem de combinação do RL com a TOC, os conceitos da TOC colaboraram para a modelagem do agente de RL, permitindo o aprendizado do modelo e atingindo resultados melhores que o DBR. As variáveis utilizadas no estado são suficientes para o agente compreender a necessidade produtiva e atender a demanda (Hopp; Spearman, 2008), assim como as métricas de desempenho da TOC permitem que a função de recompensa indique o caminho para a atualização da política (Sutton; Barto, 2018). Essa integração reduz o esforço para monitorar o sistema produtivo durante uma possível implementação empírica do agente de RL, uma vez que quantificar as peças em produção (WIP) e no estoque de produtos acabados (FG) é significativamente mais simples do que monitorar as filas dos recursos (Shiue; Lee; Su, 2020; Thomas *et al.*, 2018), as características das ordens de produção (Liu; Piplani; Toro, 2022; Marchesano *et al.*, 2022) ou, ainda, a utilização e o status dos recursos (Liu *et al.*, 2022; Sakr *et al.*, 2023).

Empregar uma teoria de gestão como base na modelagem do RL permitiu o desenvolvimento de uma abordagem completa para o ambiente MTS. A maioria das aplicações do RL no sequenciamento se limita a sequenciar ordens de produção definidas previamente (Esteso *et al.*, 2023; Panzer; Bender, 2022), ignorando a possibilidade de redimensionar o tamanho do lote. Assim, uma solução que permita dimensionar o lote e sequenciar a produção eleva a autonomia do sequenciamento, possibilitando avançar no contexto da indústria 4.0 (Jan *et al.*, 2023; Oztemel; Gursev, 2020).

6 CONCLUSÕES

Este capítulo destaca as implicações teóricas das descobertas para a literatura da TOC e do RL. Também destaca as implicações práticas para profissionais implementarem modelos de RL em sistemas produtivos. Por fim, apresenta as limitações e orientações para pesquisas futuras.

6.1 IMPLICAÇÕES TEÓRICAS PARA A TOC

Esta dissertação apresenta uma primeira tentativa de integrar a Teoria das Restrições a um algoritmo de Aprendizagem por Reforço. No capítulo 3, a revisão da literatura comprovou a inexistência de estudos relacionando ambas as áreas, além de discutir como o RL pode auxiliar no sequenciamento proposto pelo DBR por meio de um modelo conceitual. O modelo relacionou os conceitos da TOC aos do RL e demonstrou suas sinergias. Entretanto, foi no capítulo 4 que a implementação computacional foi elaborada, comparando-se o sequenciamento do DBR com um agente de RL informado pela TOC. Os resultados do modelo indicaram que o DBR tem espaço para melhorias em sua abordagem, uma vez que o RL atingiu resultados superiores ao DBR. A evidência de que o nível de ordens em processamento é mais relevante para a tomada de decisão pode orientar os próximos avanços em relação ao DBR. Por fim, o uso de RL integrado à TOC descarta a necessidade de dimensionamento e monitoramento dos pulmões de expedição, o que possibilita uma operação mais dinâmica e eficiente do ponto de vista da operação.

6.2 IMPLICAÇÕES TEÓRICAS PARA O RL

A revisão da literatura conduzida no capitulo 3, em conjunto com revisões já existentes no campo do Aprendizado por Reforço (Esteso *et al.*, 2023; Panzer; Bender, 2022; Rolf *et al.*, 2023; Samsonov; Ben Hicham; Meisen, 2022), permitiu compreender as principais abordagens de modelagem do RL, em especial, a do espaço observável, das ações e da função de recompensa. O modelo conceitual elaborado indicou os conceitos da TOC aplicáveis aos elementos do RL. No capítulo 4, o modelo conceitual foi utilizado como base para introduzir os conceitos da TOC à modelagem do RL, possibilitando reduzir a quantidade de variáveis monitoradas pelo espaço observável, de modo que foram empregados, apenas, o nível de produtos em produção e o estoque de produtos acabados. As métricas de desempenho

propostas pela TOC foram eficazes em nortear o aprendizado do modelo, pois permitiram o avanço da política na direção de maior eficiência produtiva. O conceito de sequenciamento com base na penetração do pulmão e do consumo da restrição habilitou a solução proposta para sequenciar as ordens, além de dimensionar os lotes de cada produto. É evidente que a utilização de teorias de gestão como base para a elaboração de modelos de RL contribui para atingir os objetivos e permitiu desenvolver soluções mais eficientes e legíveis para profissionais e gestores.

6.3 IMPLICAÇÕES PRÁTICAS

As descobertas desta dissertação norteiam aplicações práticas em duas direções. A primeira perspectiva corrobora estudos prévios que utilizaram teorias de gestão como base para modelagem e implementação de tecnologias (Paraschos; Koulinas; Koulouriotis, 2023; Xanthopoulos; Chnitidis; Koulouriotis, 2019), demonstrando que tais teorias buscam atingir resultados eficientes e enxutos. A segunda perspectiva aponta para a possibilidade de introdução de agentes de RL em ambientes industriais sem a necessidade de monitoramento extenso do fluxo de produção. Ao introduzir um agente de RL em um sistema autônomo, a principal dificuldade é o monitoramento constante das variáveis necessárias para a tomada de decisão (Kuhnle *et al.*, 2022). Entretanto, o modelo proposto neste estudo reduz drasticamente essa dificuldade ao utilizar, genericamente, o nível de produtos em processamento e o estoque de produtos acabados no espaço observável do agente de RL.

6.4 LIMITAÇÕES E TRABALHOS FUTUROS

As limitações desta pesquisa sugerem sua ampliação e o desenvolvimento de trabalhos futuros. Duas classes de limitações são identificadas neste trabalho: (i) temáticas e (ii) técnicas. Quanto ao tema de estudo, este foi limitado à TOC, desconsiderando a utilização de outras teorias de gestão para auxiliar na construção do modelo de RL. Além disso, esta pesquisa limitou-se a simulações de eventos discretos e descartou aplicações em linhas de processos contínuos. As limitações técnicas dizem respeito à abordagem dos cenários simulados. Ao considerar outras teorias de gestão, novos cenários podem ser simulados, permitindo a comparação

do desempenho em um amplo escopo. Ainda que o agente de RL tenha apresentado bons resultados, outros algoritmos de aprendizado podem ser implementados e testados em relação à mesma teoria de gestão. Ainda, outras abordagens para modelagem do estado observável podem ser avaliadas, comparando-se o impacto de suas variáveis nas ações e no resultado do sequenciamento. Por fim, a análise de comportamento do agente de RL permitiu compreender as principais variáveis que impactam a tomada de decisão, entretanto a análise pode ser ampliada ao se utilizar algoritmos mais complexos e ao se avaliar o comportamento do agente por uma lente temporal, em busca de comportamentos com maior granularidade.

6.5 OBSERVAÇÕES FINAIS

As abordagens de RL são focadas em propor algoritmos e estados complexos como solução para desafios industriais. No entanto, este estudo apresenta uma abordagem integradora, incluindo uma teoria de gestão a um modelo de RL para a resolução do problema. A TOC permitiu, ao agente de RL, atingir um desempenho superior ao DBR. Concomitantemente, o RL demonstrou lacunas no desempenho do DBR, em especial, na utilização da quantidade de ordens em processamento (WIP) para a tomada de decisão. Por fim, esta pesquisa avaliou a integração da TOC com um algoritmo de aprendizado de máquina, um campo de estudo, até então, pouco explorado pelo meio acadêmico.

REFERÊNCIAS

ADLER, M. J.; VAN DOREN, C. **Como Ler Livros**. 2. ed. São Paulo: É Realizações, 2010.

AL-AOMAR, R. Capacity-constrained production scheduling of multiple vehicle programs in an automotive pilot plant. **International Journal of Production Research**, Industrial Engineering, Jordan University of Science and Technology, Irbid-22110, Jordan, v. 44, n. 13, p. 2573–2604, 2006. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-33744980921&doi=10.1080%2F00207540500521212&partnerID=40&md5=27cf440ec86d31a5179d37dd312d61f8.

ALTENMÜLLER, T. *et al.* Reinforcement learning for an intelligent and autonomous production control of complex job-shops under time constraints. **Production Engineering**, [*s. l.*], v. 14, n. 3, p. 319–328, 2020. Disponível em: https://link.springer.com/10.1007/s11740-020-00967-8.

ATWATER, J. B.; STEPHENS, A. A.; CHAKRAVORTY, S. S. Impact of scheduling free goods on the throughput performance of a manufacturing operation. **International Journal of Production Research**, [*s. l.*], v. 42, n. 23, p. 4849–4869, 2004. Disponível em: https://www.tandfonline.com/doi/full/10.1080/00207540412331285805.

ATWATER, J. B.; CHAKRAVORTY, S. S. A study of the utilization of capacity constrained resources in drum-buffer-rope systems. **Production and Operations Management**, [*s. l.*], v. 11, n. 2, p. 259–273, 2002. Disponível em: https://onlinelibrary.wiley.com/doi/10.1111/j.1937-5956.2002.tb00495.x.

AYDIN, M. E.; ÖZTEMEL, E. Dynamic job-shop scheduling using reinforcement learning agents. **Robotics and Autonomous Systems**, [*s. l.*], v. 33, n. 2–3, p. 169–178, 2000. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0921889000000877.

BARBOSA, C.; AZEVEDO, A. Assessing the impact of performance determinants in complex MTO / ETO supply chains through an extended hybrid modelling approach. [*s. l.*], v. 7543, 2018.

BENKEL, K.; JØRNSTEN, K.; LEISTEN, R. Variability aspects in flowshop scheduling systems. **Proceedings of 2015 International Conference on Industrial Engineering and Systems Management, IEEE IESM 2015**, [*s. l.*], n. October, p. 118–127, 2016.

BERTOLINI, M. *et al.* Machine Learning for industrial applications: A comprehensive literature review. **Expert Systems with Applications**, [*s. l.*], v. 175, n. December 2020, p. 114820, 2021. Disponível em: https://doi.org/10.1016/j.eswa.2021.114820.

BERTRAND, J. W. M.; FRANSOO, J. C. Operations management research methodologies using quantitative modeling. **International Journal of Operations and Production Management**, [*s. l.*], v. 22, n. 2, p. 241–264, 2002.

BETTERTON, C. E.; COX, J. F. Espoused drum-buffer-rope flow control in serial lines: A comparative study of simulation models. **International Journal of Production Economics**, ["School of Business Administration, The Citadel, Charleston, SC 29409, United States", “Terry College of Business, University of Georgia, Athens, GA 30602, United States”], v. 117, n. 1, p. 66–79, 2009. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0925527308002909.

BITRAN, G. R.; TIRUPATI, D. Multiproduct queueing networks with deterministic routing: Decomposition approach and the notion of interference. **Management Science**, [*s. l.*], v. 32, p. 75–100, 1988.

BROCKMAN, G. *et al.* OpenAI Gym. [*s. l.*], p. 1–4, 2016. Disponível em: http://arxiv.org/abs/1606.01540.

BUESTÁN BENAVIDES, M.; VAN LANDEGHEM, H. Implementation of S-DBR in four manufacturing SMEs: a research case study. **Production Planning & Control**, [*s. l.*], v. 26, n. 13, p. 1110–1127, 2015. Disponível em: https://www.tandfonline.com/doi/full/10.1080/09537287.2015.1015060.

CASTRO, R. F.; GODINHO-FILHO, M.; TAVARES-NETO, R. F. Dispatching method based on particle swarm optimization for make-to-availability. **Journal of Intelligent Manufacturing**, Department of Industrial Engineering, Federal University of São Carlos – UFSCAR, São Carlos, Brazil, v. 33, n. 4, p. 1021–1030, 2022. Disponível em: https://link.springer.com/10.1007/s10845-020-01707-6.

CHAKRAVORTY, S. S.; ATWATER, J. B. The impact of free goods on the performance of drum-buffer-rope scheduling systems. **International Journal of Production Economics**, [*s. l.*], v. 95, n. 3, p. 347–357, 2005. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S092552730400009X.

CHEN, J. C. *et al.* Comparison of simulated annealing and tabu-search algorithms in advanced planning and scheduling systems for TFT-LCD colour filter fabs. **International Journal of Computer Integrated Manufacturing**, [*s. l.*], v. 30, n. 6, p. 516–534, 2017. Disponível em: http://dx.doi.org/10.1080/0951192X.2016.1145805.

CHEN, Z. *et al.* Optimal Design of Flexible Job Shop Scheduling Under Resource Preemption Based on Deep Reinforcement Learning. **Complex System Modeling and Simulation**, [*s. l.*], v. 2, n. 2, p. 174–185, 2022. Disponível em: https://ieeexplore.ieee.org/document/9841531/.

CHEN, J.; CHEN, D.; MA, Y. Study on hybrid production scheduling based on TOC. **Zhongguo Jixie Gongcheng/China Mechanical Engineering**, Nanjing University of Science and Technology, Nanjing 210094, China, v. 18, n. 20, p. 2433–2439, 2007. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-36248972323&partnerID=40&md5=73ff551acbba76deeb906f892e592e9d.

CHEN, K.; JI, P. A mixed integer programming model for advanced planning and scheduling (APS). **European Journal of Operational Research**, [*s. l.*], v. 181, n. 1, p. 515–522, 2007.

COX III, J. F. *et al.* **The Theory of Constraints International Certification Organization Dictionary. TOCICO,**. Washington, DC: [*s. n.*], 2012.

COX III, J. F.; SCHLEIER JR., J. G. **Handbook da teoria das restrições**. [*S. l.: s. n.*], 2013. Disponível em: www.grupoa.com.br.

COX III, J. F.; SCHLEIER JR., J. G. **Theory of Constraints Handbook**. Illustrateed. [*S. l.*]: McGraw-Hill, 2010.

CREIGHTON, D. C.; NAHAVANDI, S. Optimising discrete event simulation models using a reinforcement learning agent. *In*: , 2002. **Proceedings of the Winter Simulation Conference**. [*S. l.*]: IEEE, 2002. p. 1945–1950. Disponível em: http://ieeexplore.ieee.org/document/1166494/.

DA SILVA, N. A. *et al.* Industry 4.0 and micro and small enterprises: systematic literature review and analysis. **Production and Manufacturing Research**, [*s. l.*], v. 10, n. 1, p. 696–726, 2022. Disponível em: https://doi.org/10.1080/21693277.2022.2124466.

DAVIS, J. *et al.* Smart manufacturing, manufacturing intelligence and demand-dynamic performance. **Computers and Chemical Engineering**, [*s. l.*], v. 47, p. 145–156, 2012. Disponível em: http://dx.doi.org/10.1016/j.compchemeng.2012.06.037.

DEL REAL TORRES, A. *et al.* A Review of Deep Reinforcement Learning Approaches for Smart Manufacturing in Industry 4.0 and 5.0 Framework. **Applied Sciences (Switzerland)**, [*s. l.*], v. 12, n. 23, 2022.

DEVANGA, A.; BADILLA, E. D.; DEHGHANIMOHAMMADABADI, M. Applied Reinforcement Learning for Decision Making in Industrial Simulation Environments. *In*: , 2022a. **2022 Winter Simulation Conference (WSC)**. [*S. l.*]: IEEE, 2022. p. 2819–2829. Disponível em: https://ieeexplore.ieee.org/document/10015282/.

DEVANGA, A.; BADILLA, E. D.; DEHGHANIMOHAMMADABADI, M. Applied Reinforcement Learning for Decision Making in Industrial Simulation Environments. *In*: , 2022b. **2022 Winter Simulation Conference (WSC)**. [*S. l.*]: IEEE, 2022. p. 2819–2829. Disponível em: https://link.springer.com/10.1007/s40747-022-00844-0.

DOGAN, A.; BIRANT, D. Machine learning and data mining in manufacturing. **Expert Systems with Applications**, [*s. l.*], v. 166, n. February 2019, p. 114060, 2021. Disponível em: https://doi.org/10.1016/j.eswa.2020.114060.

DOHALE, V.; AMBILKAR, P.; BILOLIKAR, V. Application of TOC Strategy Using Simulation: Case of the Indian Automobile Component Manufacturing Firm. *In*: , 2021, Mumbai. **Proceedings of the International Conference on Industrial Engineering and Operations Management**. Mumbai: [*s. n.*], 2021. p. 500–508. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85123363088&partnerID=40&md5=d029d07316dc87a490527e0e08897e79.

DRESCH, A.; LACERDA, D. P.; ANTUNES JUNIOR, J. A. V. **Design Science Research: Método de Pesquisa para Avanço da Ciência e Tecnologia**. 1. ed. [*S. l.*]: Bookman, 2015.

ERIKSSON, K. *et al.* Conceptual framework of scheduling applying discrete event simulation as an environment for deep reinforcement learning. **Procedia CIRP**, [*s. l.*], v. 107, p. 955–960, 2022. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85132264077&doi=10.1016%2Fj.procir.2022.05.091&partnerID=40&md5=32ccbf9b3a144f7259cab6aa154477ea.

ERMEL, A. P. C. *et al.* **Literature Reviews: Modern Methods for Investigating Scientific and Technological Knowledge**. Cham: Springer, 2021-. ISSN 15497879.

ESTESO, A. *et al.* Reinforcement learning applied to production planning and control. **International Journal of Production Research**, [*s. l.*], v. 61, n. 16, p. 5772–5789, 2023. Disponível em: https://doi.org/10.1080/00207543.2022.2104180.

FOGLIATO, F.; RIBEIRO, J. L. D. **Confiabilidade e Manutenção Industrial**. Rio de Janeiro: Elsevier Brasil, 2009.

GAREY, M. R.; JOHNSON, D. S.; SETHI, R. The Complexity of Flowshop and Jobshop Scheduling. **Mathematics of Operations Research**, [*s. l.*], v. 1, n. 2, p. 117–129, 1976. Disponível em: https://pubsonline.informs.org/doi/10.1287/moor.1.2.117.

GAUSS, L. **Exploring Mechanisms as Boundary Objects for Connecting Design with Science in Operations Management Research**. 2023. - Universidade do Vale do Rio dos Sinos, [*s. l.*], 2023.

GERPOTT, F. T. *et al.* Integration of the A2C Algorithm for Production Scheduling in a Two-Stage Hybrid Flow Shop Environment. **Procedia Computer Science**, [*s. l.*], v. 200, p. 585–594, 2022. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85127829514&doi=10.1016%2Fj.procs.2022.01.256&partnerID=40&md5=842244e7e05fdbaa4f125e10fbcb9089.

GOLDRATT, E. M. **The Haystack Syndrome: Sifting Information Out of the Data Ocean**. [*S. l.*]: North River Press, 2006.

GOLDRATT, E. M.; COX, J. **The Goal: Excellence In Manufacturing**. Croton-on-Hudson, N.Y: North River Press, 1984.

GOLMOHAMMADI, D. A study of scheduling under the theory of constraints. **International Journal of Production Economics**, Management Science and Information Systems Department, University of Massachusetts Boston, United States, v. 165, p. 38–50, 2015. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0925527315000833.

GRAVES, S. C. *et al.* Scheduling of re-entrant flow shops. **Journal of Operations Management**, [*s. l.*], v. 3, n. 4, p. 197–207, 1983. Disponível em: https://onlinelibrary.wiley.com/doi/10.1016/0272-6963%2883%2990004-9.

GUIDE, V. D. R. A Simulation Model of Drum-Buffer-Rope for Production Planning and Control at a Naval Aviation Depot. **SIMULATION**, Department of Graduate Logistics Management, Graduate School of Logistics & Acquisition Management, Air Force Institute of Technology, WPAFB, OH 45433-7765, United States, v. 65, n. 3, p. 157–168, 1995. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0029371528&doi=10.1177%2F003754979506500302&partnerID=40&md5=8b71d999d0968880f31f18158eadfc66.

GUO, P. *et al.* Multi-objective scheduling of cloud-edge cooperation in distributed manufacturing via multi-agent deep reinforcement learning. **International Journal of Production Research**, [*s. l.*], 2024. Disponível em: https://doi.org/10.1080/00207543.2024.2329316.

GUPTA, M.; KO, H.-J.; MIN, H. TOC-based performance measures and five focusing steps in a job-shop manufacturing environment. **International Journal of Production Research**, Department of Management, School of Business, University of Louisville, Louisville, KY 40292, United States, v. 40, n. 4, p. 907–930, 2002. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0037051313&doi=10.1080%2F00207540110097185&partnerID=40&md5=88913c6339308afa3238cc04964096b5.

HARJUNKOSKI, I. *et al.* Scope for industrial applications of production scheduling models and solution methods. **Computers & Chemical Engineering**, [*s. l.*], v. 62, p. 161–193, 2014. Disponível em: http://dx.doi.org/10.1016/j.compchemeng.2013.12.001.

HAYES, C. F. *et al.* A practical guide to multi-objective reinforcement learning and planning. **Autonomous Agents and Multi-Agent Systems**, [*s. l.*], v. 36, n. 1, p. 26, 2022. Disponível em: https://doi.org/10.1007/s10458-022-09552-y.

HEGER, J.; VOSS, T. Dynamically Changing Sequencing Rules with Reinforcement Learning in a Job Shop System With Stochastic Influences. *In*: , 2020. **2020 Winter Simulation Conference (WSC)**. [*S. l.*]: IEEE, 2020. p. 1608–1618. Disponível em: https://ieeexplore.ieee.org/document/9383903/.

HILLIER, F. S.; LIEBERMAN, G. J. **Introduction to Operational Research**. 45. ed. New York: McGraw-Hill, 2015.

HOPP, W.; SPEARMAN, M. **Factory Physics: Foundations of Manufacturing Management**. [*S. l.: s. n.*], 2008.

HU, L. *et al.* Petri-net-based dynamic scheduling of flexible manufacturing system via deep reinforcement learning with graph convolutional network. **Journal of Manufacturing Systems**, [*s. l.*], v. 55, p. 1–14, 2020. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0278612520300145.

IDREES, H. D.; SINNOKROT, M. O.; AL-SHIHABI, S. A Reinforcement Learning Algorithm to Minimize the Mean Tardiness of a Single Machine with Controlled Capacity. *In*: , 2006. **Proceedings of the 2006 Winter Simulation Conference**. [*S. l.*]: IEEE, 2006. p. 1765–1769. Disponível em: https://ieeexplore.ieee.org/document/4117811/.

JAHANGIRIAN, M. *et al.* Simulation in manufacturing and business: A review. **European Journal of Operational Research**, [*s. l.*], v. 203, n. 1, p. 1–13, 2010. Disponível em: http://dx.doi.org/10.1016/j.ejor.2009.06.004.

JAN, Z. *et al.* Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. **Expert Systems with Applications**, [*s. l.*], v. 216, n. November 2021, p. 119456, 2023. Disponível em: https://doi.org/10.1016/j.eswa.2022.119456.

JEON, S.-W. *et al.* Design and Implementation of Simulation-Based Scheduling System with Reinforcement Learning for Re-Entrant Production Lines. **Machines**, [*s. l.*], v. 10, n. 12, p. 1169, 2022. Disponível em: https://www.mdpi.com/2075-1702/10/12/1169.

JEON, S. M.; KIM, G. A survey of simulation modeling techniques in production planning and control (PPC). **Production Planning & Control**, [*s. l.*], v. 27, n. 5, p. 360–377, 2016. Disponível em: http://dx.doi.org/10.1080/09537287.2015.1128010.

KARDOS, C. *et al.* Dynamic scheduling in a job-shop production system with reinforcement learning. **Procedia CIRP**, [*s. l.*], v. 97, p. 104–109, 2021. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S221282712031430X.

KIM, T. *et al.* On Scheduling a Photolithograhy Toolset Based on a Deep Reinforcement Learning Approach with Action Filter. *In*: , 2021. **2021 Winter Simulation Conference (WSC)**. [*S. l.*]: IEEE, 2021. p. 1–10. Disponível em: https://ieeexplore.ieee.org/document/9715450/.

KREUZBERGER, D.; KÜHL, N.; HIRSCHL, S. Machine Learning Operations (MLOps): Overview, Definition, and Architecture. **IEEE Access**, [*s. l.*], v. 11, n. March, p. 31866–31879, 2023. Disponível em: https://ieeexplore.ieee.org/document/10081336/.

KUHNLE, A. *et al.* Explainable reinforcement learning in production control of job shop manufacturing system. **International Journal of Production Research**, [*s. l.*], v. 60, n. 19, p. 5812–5834, 2022. Disponível em: https://doi.org/10.1080/00207543.2021.1972179.

KUHNLE, A.; JAKUBIK, J.; LANZA, G. Reinforcement learning for opportunistic maintenance optimization. **Production Engineering**, [*s. l.*], v. 13, n. 1, p. 33–41, 2019. Disponível em: http://dx.doi.org/10.1007/s11740-018-0855-7.

LAHRICHI, Y. *et al.* A first attempt to enhance Demand-Driven Material Requirements Planning through reinforcement learning. **IFAC-PapersOnLine**, [*s. l.*], v. 56, n. 2, p. 1797–1802, 2023. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S2405896323023017.

LANG, S. *et al.* Integration of Deep Reinforcement Learning and Discrete-Event Simulation for Real-Time Scheduling of a Flexible Job Shop Production. *In*: , 2020. **2020 Winter Simulation Conference (WSC)**. [*S. l.*]: IEEE, 2020. p. 3057–3068. Disponível em: https://ieeexplore.ieee.org/document/9383997/.

LAW, A. M. **Simulation modeling and analysis**. 5. ed. Tucson: McGraw-Hill, 2015.

LI, X. *et al.* Review on Learning-based Methods for shop Scheduling problems. *In*: , 2022. **2022 IEEE International Conference on e-Business Engineering (ICEBE)**. [*S. l.*]: IEEE, 2022. p. 294–298. Disponível em: https://ieeexplore.ieee.org/document/10035066/.

LIANG, T.; ZHOU, L.; JIANG, Z. Integrated scheduling of production and material delivery for the intelligent manufacturing system. **International Journal of Production Research**, [*s. l.*], 2024. Disponível em: https://doi.org/10.1080/00207543.2024.2363435.

LIU, J. *et al.* Dynamic scheduling for semiconductor manufacturing systems with uncertainties using convolutional neural networks and reinforcement learning. **Complex & Intelligent Systems**, [*s. l.*], v. 8, n. 6, p. 4641–4662, 2022. Disponível em: https://link.springer.com/10.1007/s40747-022-00844-0.

LIU, W. *et al.* Scheduling optimization for production of prefabricated components with parallel work of serial machines. **Automation in Construction**, [*s. l.*], v. 148, n. February, p. 104770, 2023. Disponível em: https://doi.org/10.1016/j.autcon.2023.104770.

LIU, R.; PIPLANI, R.; TORO, C. Deep reinforcement learning for dynamic scheduling of a flexible job shop. **International Journal of Production Research**, [*s. l.*], v. 60, n. 13, p. 4049–4069, 2022. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129220142&doi=10.1080%2F00207543.2022.2058432&partnerID=40&md5=6c8b8c2cc7a1fcf9163b551778d0c574.

LÖDDING, H. **Handbook of Manufacturing Control**. [*S. l.: s. n.*], 2013.

MA, L.; CHEN, J. Modeling and Simulating of Time Buffer Control for Engine Remanufacturing System. *In*:, 2009. **2009 International Conference on Information Engineering and Computer Science**. [*S. l.*]: IEEE, 2009. p. 1–4. Disponível em: http://ieeexplore.ieee.org/document/5365624/.

MACHADO, M. P. *et al.* Exploratory decision robustness analysis of the theory of constraints focusing process using system dynamics modeling. **International Journal of Production Economics**, [*s. l.*], v. 260, n. March, p. 108856, 2023. Disponível em: https://doi.org/10.1016/j.ijpe.2023.108856.

MARCHESANO, M. G. *et al.* Dynamic scheduling of a due date constrained flow shop with Deep Reinforcement Learning. **IFAC-PapersOnLine**, [*s. l.*], v. 55, n. 10, p. 2932–2937, 2022. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S2405896322021917.

MÁRQUEZ, C. R. H.; RIBEIRO, C. C. Shop scheduling in manufacturing environments: a review. **International Transactions in Operational Research**, [*s. l.*], v. 29, n. 6, p. 3237–3293, 2022. Disponível em: https://onlinelibrary.wiley.com/doi/10.1111/itor.13108.

MISSBAUER, H.; UZSOY, R. Order release in production planning and control systems: challenges and opportunities. **International Journal of Production Research**, [*s. l.*], v. 60, n. 1, p. 256–276, 2022. Disponível em: https://doi.org/10.1080/00207543.2021.1994165.

MORABITO, R.; PUREZA, V. Modelagem e Simulação. *In*: CAUCHICK MIGUEL, P. A. (org.). **Metodologia de Pesquisa em Engenharia de Produção e Gestão de Operações**. 2. ed. Rio de Janeiro: Elsevier, 2012.

MYLES, A. J. *et al.* An introduction to decision tree modeling. [*s. l.*], p. 275–285, 2004.

NATHANS, L. L.; OSWALD, F. L.; NIMON, K. Interpreting Multiple Linear Regression: A Guidebook of Variable Importance - Practical Assessment, Research & Evaluation. **Practical Assessment Research & Evaluation**, [*s. l.*], v. 17, n. 9, p. 1–19, 2012. Disponível em: https://pareonline.net/getvn.asp?v=17&n=9.

NEUFELD, J. S.; SCHULZ, S.; BUSCHER, U. A systematic review of multi-objective hybrid flow shop scheduling. **European Journal of Operational Research**, [*s. l.*], v. 309, n. 1, p. 1–23, 2023. Disponível em: https://doi.org/10.1016/j.ejor.2022.08.009.

NGUYEN, Tiep *et al.* Knowledge mapping of digital twin and physical internet in Supply Chain Management: A systematic literature review. **International Journal of Production Economics**, [*s. l.*], v. 244, n. July 2021, p. 108381, 2022. Disponível em: https://doi.org/10.1016/j.ijpe.2021.108381.

OPENAI *et al.* Dota 2 with Large Scale Deep Reinforcement Learning. [*s. l.*], 2019. Disponível em: http://arxiv.org/abs/1912.06680.

OZTEMEL, E.; GURSEV, S. Literature review of Industry 4 . 0 and related technologies. **Journal of Intelligent Manufacturing**, [*s. l.*], v. 31, n. 1, p. 127–182, 2020. Disponível em: https://doi.org/10.1007/s10845-018-1433-8.

PANZER, M.; BENDER, B. Deep reinforcement learning in production systems: a systematic literature review. **International Journal of Production Research**, [*s. l.*], v. 60, n. 13, p. 4316–4341, 2022. Disponível em: https://doi.org/10.1080/00207543.2021.1973138.

PARASCHOS, P. D. *et al.* Machine learning integrated design and operation management for resilient circular manufacturing systems. **Computers & Industrial Engineering**, [*s. l.*], v. 167, n. January, p. 107971, 2022. Disponível em: https://doi.org/10.1016/j.cie.2022.107971.

PARASCHOS, P. D.; KOULINAS, G. K.; KOULOURIOTIS, D. E. A reinforcement learning/ad-hoc planning and scheduling mechanism for flexible and sustainable manufacturing systems. **Flexible Services and Manufacturing Journal**, [*s. l.*], n. 0123456789, 2023. Disponível em: https://doi.org/10.1007/s10696-023-09496-9.

PATERNINA-ARBOLEDA, C. D.; DAS, T. K. A multi-agent reinforcement learning approach to obtaining dynamic control policies for stochastic lot scheduling problem. **Simulation Modelling Practice and Theory**, [*s. l.*], v. 13, n. 5, p. 389–406, 2005. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S1569190X04001406.

PINEDO, M. L. **Scheduling**. Cham: Springer International Publishing, 2016-. ISSN 1079-9737. Disponível em: http://link.springer.com/10.1007/978-3-319-26580-3.

RABELO, L. C.; JONES, A.; YIH, Y. Development of a real-time learning scheduler using reinforcement learning concepts. *In*: , 1994. **Proceedings of 1994 9th IEEE International Symposium on Intelligent Control**. [*S. l.*]: IEEE, 1994. p. 291–296. Disponível em: http://ieeexplore.ieee.org/document/367802/.

RAFFIN, A. *et al.* Stable-baselines3: Reliable reinforcement learning implementations. **Journal of Machine Learning Research**, [*s. l.*], v. 22, p. 1–8, 2021.

ROLF, B. *et al.* A review on reinforcement learning algorithms and applications in supply chain management. **International Journal of Production Research**, [*s. l.*], v. 61, n. 20, p. 7151–7179, 2023. Disponível em: https://doi.org/10.1080/00207543.2022.2140221.

RUDIN, C. *et al.* Interpretable machine learning: Fundamental principles and 10 grand challenges. **Statistics Surveys**, [*s. l.*], v. 16, p. 1–85, 2022.

RUMMUKAINEN, H.; NURMINEN, J. K. Practical Reinforcement Learning - Experiences in Lot Scheduling Application. **IFAC-PapersOnLine**, [*s. l.*], v. 52, n. 13, p. 1415–1420, 2019. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S2405896319313783.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence: A modern Approach**. [*S. l.*]: Pearson, 2010.

RUSSELL†, R. S.; DAR-EL‡, E. M.; TAYLOR, B. W. A comparative analysis of the COVERT job sequencing rule using various shop performance measures. **International Journal of Production Research**, [*s. l.*], v. 25, n. 10, p. 1523–1540, 1987. Disponível em: https://www.tandfonline.com/doi/full/10.1080/00207548708919930.

SAKR, A. H. *et al.* Simulation and deep reinforcement learning for adaptive dispatching in semiconductor manufacturing systems. **Journal of Intelligent Manufacturing**, [*s. l.*], v. 34, n. 3, p. 1311–1324, 2023.

SAMSONOV, V.; BEN HICHAM, K.; MEISEN, T. Reinforcement Learning in Manufacturing Control: Baselines, challenges and ways forward. **Engineering Applications of Artificial Intelligence**, [*s. l.*], v. 112, 2022. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85129460925&doi=10.1016%2Fj.engappai.2022.104868&partnerID=40&md5=10fbafc35a1620d540d11cf9f2b50898.

SCHRAGENHEIM, E. Managing Make-to-Stock and the Concept of Make-to-Availability. *In*: III COX, J. F.; JR. SCHLEIER, J. G. (org.). **Theory of Constraints Handbook**. [*S. l.*]: McGraw-Hill, 2010.

SCHRAGENHEIM, E.; RONEN, B. Buffer management. A diagnostic tool for production control. **Production and Inventory Management Journal**, [*s. l.*], v. Second Qua, n. 2, p. 74–79, 1991.

SCHRAGENHEIM, E.; RONEN, B. Drum-Buffer-Rope Shop Floor Control. **Production and inventory management**, [*s. l.*], v. 31, n. 2, p. 18–22, 1990.

SCHUH, G. *et al.* Towards an Automated Application for Order Release. **Procedia CIRP**, [*s. l.*], v. 107, p. 1323–1328, 2022. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S221282712200436X.

SCHULMAN, J. *et al.* Proximal Policy Optimization Algorithms. [*s. l.*], p. 1–12, 2017. Disponível em: http://arxiv.org/abs/1707.06347.

SELLITTO, M. A.; BRUSIUS, W. Maintenance strategy based on reliability analytical models for three parallel mechanical transformation machines. **IEEE Latin America Transactions**, [*s. l.*], v. 15, n. 5, 2017.

SELLITTO, M. A.; PINHO, B. Maintenance Strategy Choice Supported by the Failure Rate Function: Application in a Serial Manufacturing Line. **Periodica Polytechnica Social and Management Sciences**, [*s. l.*], v. 31, n. 1, p. 38–51, 2022. Disponível em: https://pp.bme.hu/so/article/view/18627.

SHEREMETOV, L. *et al.* Optimization Algorithm for Dynamic Multi-Agent Job Routing. *In*: EMERGING SOLUTIONS FOR FUTURE MANUFACTURING SYSTEMS. Boston: Kluwer Academic Publishers, 2005. v. 159, p. 183–192. Disponível em: http://link.springer.com/10.1007/0-387-22829-2_19.

SHI, D. *et al.* Intelligent scheduling of discrete automated production line via deep reinforcement learning. **International Journal of Production Research**, [*s. l.*], v. 58, n. 11, p. 3362–3380, 2020. Disponível em: https://www.tandfonline.com/doi/full/10.1080/00207543.2020.1717008.

SHIUE, Y.-R.; LEE, K.-C.; SU, C.-T. A Reinforcement Learning Approach to Dynamic Scheduling in a Product-Mix Flexibility Environment. **IEEE Access**, [*s. l.*], v. 8, p. 106542–106553, 2020. Disponível em: https://ieeexplore.ieee.org/document/9110908/.

SHIUE, Y.-R.; LEE, K.-C.; SU, C.-T. Development of dynamic scheduling in semiconductor manufacturing using a Q-learning approach. **International Journal of Computer Integrated Manufacturing**, [*s. l.*], v. 35, n. 10–11, p. 1188–1204, 2022. Disponível em: https://www.tandfonline.com/doi/full/10.1080/0951192X.2021.1946849.

SHIUE, Y.-R.; LEE, K.-C.; SU, C.-T. Real-time scheduling for a smart factory using a reinforcement learning approach. **Computers & Industrial Engineering**, [*s. l.*], v. 125, p. 604–614, 2018. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S036083521830130X.

SILVA, T.; AZEVEDO, A. Production flow control through the use of reinforcement learning. **Procedia Manufacturing**, [*s. l.*], v. 38, n. 2019, p. 194–202, 2019. Disponível em: https://doi.org/10.1016/j.promfg.2020.01.026.

SOUSA, T. B. de *et al.* AN OVERVIEW OF THE ADVANCED PLANNING AND SCHEDULING SYSTEMS. **Independent Journal of Management & Production**, [*s. l.*], v. 5, n. 4, 2014. Disponível em: http://www.ijmp.jor.br/index.php/ijmp/article/view/239.

STRAUSS, A.; CORBIN, J. M. **Basics of qualitative research: Grounded theory procedures and techniques.** Thousand Oaks, CA, US: Sage Publications, Inc, 1990.

SUN, G. E. *et al.* Research on DBR Production Scheduling System Application Based on Simulation. **Advanced Materials Research**, Huazhong University of Science and Technology, Wuhan, China, v. 889–890, p. 1207–1212, 2014. Disponível em: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84896879031&doi=10.4028%2Fwww.scientific.net%2FAMR.889-890.1207&partnerID=40&md5=d86ec8ec96f482b3f96fee644f6fe5cb.

SUN, P., LI, K. (2018). Methodology – A Review of Intelligent Manufacturing: Scope, Strategy and Simulation. In: Wang, S., Price, M., Lim, M., Jin, Y., Luo, Y., Chen, R. (eds) Recent Advances in Intelligent Manufacturing . ICSEE IMIOT 2018 2018. Communications in Computer and Information Science, vol 923. Springer, Singapore. https://doi.org/10.1007/978-981-13-2396-6_33.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning: An Introduction**. [*S. l.*]: Bradford Books, 2018. v. 258

TANG, J.; SALONITIS, K. A Deep Reinforcement Learning Based Scheduling Policy for Reconfigurable Manufacturing Systems. **Procedia CIRP**, [*s. l.*], v. 103, p. 1–7, 2021. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S2212827121008398.

TELLES, E. S. *et al.* Drum-Buffer-Rope in an engineering-to-order productive system: a case study in a Brazilian aerospace company. **Journal of Manufacturing Technology Management**, [*s. l.*], v. 33, n. 6, p. 1190–1209, 2022. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/JMTM-10-2021-0420/full/html.

TELLES, E. S. *et al.* Drum-buffer-rope in an engineering-to-order system: An analysis of an aerospace manufacturer using data envelopment analysis (DEA). **International Journal of Production Economics**, [*s. l.*], v. 222, n. February 2019, p. 107500, 2020. Disponível em: https://doi.org/10.1016/j.ijpe.2019.09.021.

THOMAS, T. E. *et al.* Minerva: A reinforcement learning-based technique for optimal scheduling and bottleneck detection in distributed factory operations. *In*: , 2018. **2018 10th International Conference on Communication Systems & Networks (COMSNETS)**. [*S. l.*]: IEEE, 2018. p. 129–136. Disponível em: http://ieeexplore.ieee.org/document/8328189/.

THÜRER, M.; FERNANDES, N. O.; STEVENSON, M. Production planning and control in multi-stage assembly systems: an assessment of Kanban, MRP, OPT (DBR) and DDMRP by simulation. **International Journal of Production Research**, [*s. l.*], v. 60, n. 3, p. 1036–1050, 2022. Disponível em: https://doi.org/10.1080/00207543.2020.1849847.

THÜRER, M.; STEVENSON, M. Bottleneck-oriented order release with shifting bottlenecks: An assessment by simulation. **International Journal of Production Economics**, [*s. l.*], v. 197, p. 275–282, 2018. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0925527318300409.

THÜRER, M.; STEVENSON, M. On the beat of the drum: improving the flow shop performance of the Drum–Buffer–Rope scheduling mechanism. **International Journal of Production Research**, [*s. l.*], v. 56, n. 9, p. 3294–3305, 2018. Disponível em: https://doi.org/10.1080/00207543.2017.1401245.

TURGUT, Y.; BOZDAG, C. E. Deep Q-Network Model for Dynamic Job Shop Scheduling Problem Based on Discrete Event Simulation. *In*: , 2020. **2020 Winter Simulation Conference (WSC)**. [*S. l.*]: IEEE, 2020. p. 1551–1559. Disponível em: https://ieeexplore.ieee.org/document/9383986/.

VANVUCHELEN, N.; GIJSBRECHTS, J.; BOUTE, R. Use of Proximal Policy Optimization for the Joint Replenishment Problem. **Computers in Industry**, [*s. l.*], v. 119, n. August, p. 103239, 2020. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0166361519308218.

WANG, L.; PAN, Z.; WANG, J. A Review of Reinforcement Learning Based Intelligent Optimization for Manufacturing Scheduling. **Complex System Modeling and Simulation**, [*s. l.*], v. 1, n. 4, p. 257–270, 2022.

WASCHNECK, B. *et al.* Optimization of global production scheduling with deep reinforcement learning. **Procedia CIRP**, [*s. l.*], v. 72, p. 1264–1269, 2018. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S221282711830372X.

WOO, J. H. *et al.* Development of a Reinforcement Learning-Based Adaptive Scheduling Algorithm for Block Assembly Production Line. *In*: , 2021. **2021 Winter Simulation Conference (WSC)**. [*S. l.*]: IEEE, 2021. p. 1–12. Disponível em: https://ieeexplore.ieee.org/document/9715509/.

WU, X. *et al.* A deep reinforcement learning model for dynamic job-shop scheduling problem with uncertain processing time. **Engineering Applications of Artificial Intelligence**, [*s. l.*], v. 131, n. December 2022, p. 107790, 2024. Disponível em: https://doi.org/10.1016/j.engappai.2023.107790.

WU, H. H. *et al.* Simulation and scheduling implementation study of TFT-LCD Cell plants using Drum-Buffer-Rope system. **Expert Systems with Applications**, [*s. l.*], v. 37, n. 12, p. 8127–8133, 2010. Disponível em: http://dx.doi.org/10.1016/j.eswa.2010.05.075.

WU, S.-Y.; MORRIS, J. S.; GORDON, T. M. A simulation analysis of the effectiveness of drum-buffer-rope scheduling in furniture manufacturing. **Computers & Industrial Engineering**, [*s. l.*], v. 26, n. 4, p. 757–764, 1994. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/0360835294900108.

XANTHOPOULOS, A. S.; CHNITIDIS, G.; KOULOURIOTIS, D. E. Reinforcement learning-based adaptive production control of pull manufacturing systems. **Journal of Industrial and Production Engineering**, [*s. l.*], v. 36, n. 5, p. 313–323, 2019. Disponível em: https://doi.org/10.1080/21681015.2019.1647301.

XIAO, C. *et al.* Using Spearman ' s correlation coef fi cients for exploratory data analysis on big dataset. [*s. l.*], 2015.

YADAV, A.; JAYSWAL, S. C. Modelling of flexible manufacturing system: a review. **International Journal of Production Research**, [*s. l.*], v. 56, n. 7, p. 2464–2487, 2018. Disponível em: http://doi.org/10.1080/00207543.2017.1387302.

YAMANE, T. **Statistics: An Introductory Analysis**. New York: Harper & Row, 1973.

YU, Q. *et al.* Evaluation System and Correlation Analysis for Determining the Performance of a Semiconductor Manufacturing System. **Complex System Modeling and Simulation**, [*s. l.*], v. 1, n. 3, p. 218–231, 2021. Disponível em: https://ieeexplore.ieee.org/document/9600619/.

YU, X.; RAM, B. Bio-inspired scheduling for dynamic job shops with flexible routing and sequence-dependent setups. **International Journal of Production Research**, [*s. l.*], v. 44, n. 22, p. 4793–4813, 2006. Disponível em: http://www.tandfonline.com/doi/abs/10.1080/00207540600621094.

ZHANG, N. *et al.* A Review of Robust Machine Scheduling. **IEEE Transactions on Automation Science and Engineering**, [*s. l.*], p. 1–12, 2023. Disponível em: https://ieeexplore.ieee.org/document/10053636/.

ZHANG, X. M.; DU, Y. L. Research of Production Scheduling Based on Theory of Constraints. *In*: , 2015. **Proceedings of the 2015 International Conference on Electrical, Automation and Mechanical Engineering**. [*S. l.: s. n.*], 2015. p. 142–145. Disponível em: https://www.atlantis-press.com/article/22251.

ZHENG, T. *et al.* The applications of Industry 4.0 technologies in manufacturing context: a systematic literature review. **International Journal of Production Research**, [*s. l.*], v. 59, n. 6, p. 1922–1954, 2021. Disponível em: https://doi.org/10.1080/00207543.2020.1824085.

ZHOU, L. *et al.* Production and operations management for intelligent manufacturing: a systematic literature review. **International Journal of Production Research**, [*s. l.*], v. 60, n. 2, p. 808–846, 2022. Disponível em: https://doi.org/10.1080/00207543.2021.2017055.

APÊNDICE I

Tabela 10 - Protocolo de pesquisa da revisão da literatura.

	Protocolo de pesquisa		
1.	*Framework* conceitual	1.1.	A Teoria das Restrições (TOC) é uma filosofia de gestão associada ao gerenciamento do recurso gargalo. Desenvolvida por Goldratt e Cox (1984), a TOC reconhece a variabilidade inerente aos processos produtivos e visa equilibrar o fluxo de produção (Costa et al., 2022). Por meio de um processo de melhoria contínua, ela envolve cinco etapas: (1) identificar a restrição; (2) explorar a restrição; (3) subordinar todo o resto à restrição; (4) elevar a restrição e (5) se, em qualquer etapa anterior, a restrição for removida, retornar à primeira etapa e não deixar que a inércia tome conta do sistema (Goldratt; Cox, 1984)
		1.2.	O método Tambor, Pulmão e corda (DBR - *drum-buffer-rope*) foi projetado para operar em ambientes caracterizados por alta dependência e variabilidade (Schragenheim; Ronen, 1991; Goldratt, 2006; Cox III; Schleier, 2010). O tambor, que representa o gargalo ou a restrição do sistema, define o ritmo de produção. À medida que novos pedidos são introduzidos, o sequenciamento da restrição é ajustado para que não haja perda de tempo. A corda é o canal de comunicação entre o início do processo (liberação de material) e a restrição, proporcionando a sincronização da taxa de saída do gargalo com a liberação de materiais (Thürer; Stevenson, 2018b). O Pulmão é definido como o volume de trabalho (em unidade de tempo) em frente à restrição. Seu propósito é evitar que a restrição tenha interrupções não planejadas para não prejudicar o ganho (Goldratt, 2006). O pulmão pode, ainda, ser classificado como pulmão de montagem, de expedição, de espaço ou de capacidade (Schragenheim; Ronen, 1991).
		1.3.	O sequenciamento da produção é uma etapa crucial em sistemas de manufatura, pois é responsável por sincronizar a capacidade com a demanda (Hopp; Spearman, 2008). No entanto, devido às variabilidades do processo, o sequenciamento é considerado uma tarefa de alta complexidade (Panzer; Bender, 2022). Nos últimos anos, os métodos de otimização, determinísticos ou por aproximação, apresentaram um crescimento acentuado em suas aplicações às tarefas de planejamento (Bertolini *et al.*, 2021). Em especial, a utilização de aprendizado de máquina se destacou como uma solução capaz de proporcionar resultados próximos do ótimo e com tempo de resposta reduzido (Jan *et al.*, 2023; Paraschos *et al.*, 2022).
		1.4.	Dentre as técnicas de aprendizado de máquina, o Aprendizado por Reforço (RL – Reinforcement Learning) despertou o interesse da comunidade acadêmica nos últimos anos (Esteso et al., 2023), em especial para automatizar a atividade de sequenciamento em ambientes produtivos (Wang; Pan; Wang, 2022). O RL é treinado em um ambiente virtual, tomando decisões e recebendo recompensas conforme suas ações. Seu objetivo é maximizar a recompensa de longo prazo, ao passo que atualiza sua política de tomada de decisões (Russel; Norvig, 2010; Sutton; Barto, 2018). Diferente das abordagens tradicionais de aprendizado de máquina, no RL não são necessários dados rotulados, pois os dados são gerados pela interação com o ambiente virtual (Panzer; Bender, 2022). Esse ambiente é comumente modelado como um espaço de ações discretas, em que as decisões são tomadas em momentos discretos (Del Real Torres et al., 2022;

			Panzer; Bender, 2022).
		1.5.	A técnica mais utilizada para modelar ambientes virtuais de manufatura é a simulação por eventos discretos (Jahangirian et al., 2010; Jeon; Kim, 2016). Ela é comumente usada para verificar o planejamento da produção e estimar o impacto das alterações no sistema (Sun; Li, 2018). Entretanto, tem se destacado como a principal abordagem para a elaboração de ambientes de treinamento para algoritmos de RL (Del Real Torres *et al.*, 2022).
2.	Questões de pesquisa	2.1.	Quais são as principais variáveis e quais seus impactos em sistemas geridos pelo DBR?
		2.2.	Quais são as principais variáveis utilizadas para modelar o espaço observável, as ações e a recompensa em aplicações de RL?
		2.3.	Como os elementos do RL podem se relacionar com os métodos e conceitos propostos pela TOC?
3.	Tipo da revisão	3.1.	Configurativa
4.	Horizonte de tempo	4.1.	Até 2023
5.	Método de busca	5.1.	Busca em bases de dados
		5.2.	Contato com especialistas
6.	*String* de busca	6.1.	("reinforcement learning") AND ("theory of constraints" OR "theory of constraint" OR (drum AND buffer AND rope) OR bottleneck) AND (manufacture OR industrial OR production OR fabrication)
		6.2.	("theory of constraints" OR "theory of constraint" OR (drum AND buffer AND rope)) AND (manufacture OR industrial OR production OR fabrication) AND (simulation OR simulated)
		6.3.	("reinforcement learning") AND (scheduling OR "order release" OR dispatching) AND (manufacture OR industrial OR production OR fabrication) AND (simulation OR simulated)
7.	Bases de dados	7.1.	Scopus
		7.2.	Web of Science
		7.3.	Science Direct
		7.4.	IEEE
8.	Critérios de inclusão	8.1.	Utilização do DBR com simulações por eventos discretos.
		8.2.	Aplicações de aprendizado por reforço em atividades de sequenciamento com simulações por evento discreto.
9.	Critérios de exclusão		<u>General:</u>
		9.1.	Estudos duplicados.
		9.2.	Estudos indisponíveis.
		9.3.	Estudos não escritos em inglês ou português.
		9.4.	Estudos não relacionados ao sequenciamento da produção.
		9.5.	Estudos em processos contínuos.
		9.6.	Estudos sem utilização de simulações por eventos discretos
			<u>Tambor, Pulmão e Corda (DBR):</u>
		9.7.	Estudos não relacionados ao DBR.
		9.8.	Estudos comparativos ou fusão de heurísticas.
			<u>Aprendizado por reforço (RL):</u>
		9.9.	Estudos não relacionados ao RL.
		9.10.	Otimizações para redução de energia.
		9.11.	Estudos relacionados à cadeia de suprimentos, logística ou robótica.
10.	Data Analysis	10.1.	Desenvolvimento científico
		10.2.	Análise temática
11.	Data Synthesis	11.1.	Meta síntese

Tabela 11 - Artigos do corpus de análise

Tag.	Grupo	Título	Autor, ano
A01	TOC	Drum-buffer-rope shop floor control	Schragenheim and Ronen, 1990
A02	TOC	Buffer management. A diagnostic tool for production control	Schragenheim and Ronen, 1991
A03	TOC	A simulation analysis of the effectiveness of drum-buffer-rope scheduling in furniture manufacturing	Wu et al., 1994
A04	TOC	A Simulation Model of Drum-Buffer-Rope for Production Planning and Control at a Naval Aviation Depot	Guide, 1995
A05	TOC	TOC-based performance measures and five focusing steps in a job-shop manufacturing environment	Gupta et al., 2002
A06	TOC	Impact of scheduling free goods on the throughput performance of a manufacturing operation	Atwater et al., 2004
A07	TOC	The impact of free goods on the performance of drum-buffer-rope scheduling systems	Chakravorty and Atwater, 2005
A08	TOC	Capacity-constrained production scheduling of multiple vehicle programs in an automotive pilot plant	Al-Aomar, 2006
A09	TOC	Study on hybrid production scheduling based on TOC	Chen et al., 2007
A10	TOC	A study of the utilization of capacity constrained resources in drum-buffer-rope systems*	ATWATER and CHAKRAVORTY, 2002
A11	TOC	Espoused drum-buffer-rope flow control in serial lines: A comparative study of simulation models	Betterton and Cox, 2009
A12	TOC	Modeling and Simulating of Time Buffer Control for Engine Remanufacturing System	Ma and Chen, 2009
A13	TOC	Research on DBR Production Scheduling System Application Based on Simulation	Sun et al., 2014
A14	TOC	A study of scheduling under the theory of constraints	Golmohammadi, 2015
A15	TOC	Research of Production Scheduling Based on Theory of Constraints	Zhang and Du, 2015
A16	TOC	Bottleneck-oriented order release with shifting bottlenecks: An assessment by simulation	Thürer and Stevenson, 2018b
A17	TOC	On the beat of the drum: improving the flow shop performance of the Drum–Buffer–Rope scheduling mechanism	Thürer and Stevenson, 2018a
A18	TOC	Application of TOC Strategy Using Simulation: Case of the Indian Automobile Component Manufacturing Firm	Dohale et al., 2021
A19	TOC	Dispatching method based on particle swarm optimization for make-to-availability	Castro et al., 2022
A20	RL	Development of a real-time learning scheduler using reinforcement learning concepts	Rabelo et al., 1994
A21	RL	Dynamic job-shop scheduling using reinforcement learning agents	Aydin and Öztemel, 2000
A22	RL	Optimising discrete event simulation models using a reinforcement learning agent	Creighton and Nahavandi, 2002
A23	RL	A multi-agent reinforcement learning approach to obtaining dynamic control policies for stochastic lot scheduling problem	Paternina-Arboleda and Das, 2005
A24	RL	Optimization Algoritmo for Dynamic Multi-Agent Job Routing	Sheremetov et al., 2005
A25	RL	A Reinforcement Learning Algoritmo to Minimize the Mean Tardiness of a Single Machine with Controlled Capacity	Idrees et al., 2006
A26	RL	Bio-inspired scheduling for dynamic job shops with flexible routing and sequence-dependent setups	Yu; Ram, 2006
A27	RL	Minerva: A reinforcement learning-based technique for optimal scheduling and bottleneck detection in distributed factory operations	Thomas et al., 2018
A28	RL	Optimization of global production scheduling with deep reinforcement learning	Waschneck et al., 2018
A29	RL	Real-time scheduling for a smart factory using a reinforcement learning approach	Shiue et al., 2018
A30	RL	Practical Reinforcement Learning -Experiences in Lot	Rummukainen and

Tag.	Grupo	Título	Autor, ano
		Scheduling Application	Nurminen, 2019
A31	RL	A Reinforcement Learning Approach to Dynamic Scheduling in a Product-Mix Flexibility Environment	Shiue et al., 2020
A32	RL	Deep Q-Network Model for Dynamic Job Shop Scheduling Problem Based on Discrete Event Simulation	Turgut and Bozdag, 2020
A33	RL	Dynamically Changing Sequencing Rules with Reinforcement Learning in a Job Shop System With Stochastic Influences	Heger and Voss, 2020
A34	RL	Integration of Deep Reinforcement Learning and Discrete-Event Simulation for Real-Time Scheduling of a Flexible Job Shop Production	Lang et al., 2020
A35	RL	Intelligent scheduling of discrete automated production line via deep reinforcement learning	Shi et al., 2020
A36	RL	Petri-net-based dynamic scheduling of flexible manufacturing system via deep reinforcement learning with graph convolutional network	Hu et al., 2020
A37	RL	Reinforcement learning for an intelligent and autonomous production control of complex job-shops under time constraints	Altenmüller et al., 2020
A38	RL	A Deep Reinforcement Learning Based Scheduling Policy for Reconfigurable Manufacturing Systems	Tang and Salonitis, 2021
A39	RL	Development of a Reinforcement Learning-Based Adaptive Scheduling Algoritmo for Block Assembly Production Line	Woo et al., 2021
A40	RL	Dynamic scheduling in a job-shop production system with reinforcement learning	Kardos et al., 2021
A41	RL	On Scheduling a Photolithograhy Toolset Based on a Deep Reinforcement Learning Approach with Action Filter	Kim et al., 2021
A42	RL	Applied Reinforcement Learning for Decision Making in Industrial Simulation Environments	Devanga et al., 2022
A43	RL	Deep reinforcement learning for dynamic scheduling of a flexible job shop	R. Liu et al., 2022
A44	RL	Design and Implementation of Simulation-Based Scheduling System with Reinforcement Learning for Re-Entrant Production Lines	Jeon et al., 2022
A45	RL	Development of dynamic scheduling in semiconductor manufacturing using a Q-learning approach	Shiue et al., 2022
A46	RL	Dynamic scheduling for semiconductor manufacturing systems with uncertainties using convolutional neural networks and reinforcement learning	J. Liu et al., 2022
A47	RL	Dynamic scheduling of a due date constrained flow shop with Deep Reinforcement Learning	Marchesano et al., 2022
A48	RL	Integration of the A2C Algoritmo for Production Scheduling in a Two-Stage Hybrid Flow Shop Environment	Gerpott et al., 2022
A49	RL	Towards an Automated Application for Order Release	Schuh et al., 2022
A50	RL	Simulation and deep reinforcement learning for adaptive dispatching in semiconductor manufacturing systems	Sakr et al., 2023

Fonte: Elaborado pelo autor

Tabela 12 - Critérios de exclusão do grupo TOCA

Excluídos	Porcentagem	Critério de exclusão
85	30,0%	Estudos não relacionados ao DBR
62	21,9%	Estudos não relacionados ao sequenciamento da produção
54	19,1%	Estudos comparativos ou fusão de heurísticas
52	18,4%	Estudos sem utilização de simulações por eventos discretos
16	5,7%	Estudos não escritos em inglês ou português
11	3,9%	Estudos indisponíveis
2	0,7%	Atas de conferência
1	0,4%	Estudos em processos contínuos

Fonte: Elaborado pelo autor

Tabela 13 - Critérios de exclusão do grupo RLA

Excluídos	Porcentagem	Critério de exclusão
197	51,2%	Estudos não relacionados ao sequenciamento da produção
59	15,4%	Estudos sem utilização de simulações por eventos discretos
52	13,5%	Estudos relacionados à cadeia de suprimentos, logística ou robótica
21	5,5%	Atas de conferências
19	4,9%	Patentes
11	2,9%	Estudos indisponíveis
11	2,9%	Estudos não relacionados ao RL
6	1,6%	Estudos em processos contínuos
5	1,3%	Otimizações para redução de energia
4	1,0%	Estudos não escritos em inglês ou português

Fonte: Elaborado pelo autor

Tabela 14 - Códigos da análise de conteúdo

Grupo	Código	Definição
Algoritmo	A2C	Algoritmo Advantage actor critic (Gerpott *et al.*, 2022)
Algoritmo	A3C	Algoritmo Asynchronous Advantage Actor Critic (Liu *et al.*, 2022)
Algoritmo	DQN	Algoritmo Deep-Q-Network (Sakr *et al.*, 2023)
Algoritmo	DDQN	Algoritmo Double Deep-Q-Network (Devanga; Badilla; Dehghanimohammadabadi, 2022b; Jeon *et al.*, 2022)
Algoritmo	DDDQN	Algoritmo Dueling Double Deep-Q-Learning (Tang; Salonitis, 2021)
Algoritmo	PPO	Algoritmo Proximal Policy Optimization (Chen *et al.*, 2022; Rummukainen; Nurminen, 2019)
Algoritmo	Q-learning	Algoritmo Q-learning (Shiue; Lee; Su, 2020, 2022)
Pulmão DBR	Assembly buffer	Pulmão de tempo antes da montagem. Aplicado a peças que não passam pela restrição e que serão montadas nas peças que passaram pela restrição (Cox III *et al.*, 2012).
Pulmão DBR	Constraint buffer	Pulmão de tempo antes da restrição. Aplicado para proteger o gargalo da variabilidade. (Cox III *et al.*, 2012)
Pulmão DBR	Shipping buffer	Pulmão de tempo antes da expedição. Aplicado para proteger a data de entrega das ordens. (Cox III *et al.*, 2012)
Pulmão DBR	Space buffer	Pulmão de espaço físico após a restrição. Aplicado para garantir alocação das peças processadas pela restrição. (Cox III *et al.*, 2012)
Pulmão DBR	Capacity buffer	Pulmão de capacidade extra. Representa a diferença entre a capacidade dos recursos em relação à restrição (Cox III *et al.*, 2012)
Pulmão DBR	Stock buffer	Material armazenado em locais específicos do processo para

Grupo	Código	Definição
		proteger a demanda da variabilidade do suprimento de matéria prima. (Cox III *et al.*, 2012).
Regra de sequenciamento	C over T (COVERT)	Seleciona a ordem de acordo com a maior proporção entre o atraso esperado da ordem e do tempo de processamento da operação. (RUSSELL†; DAR-EL‡; TAYLOR, 1987).
Regra de sequenciamento	Critical Ratio (CR)	Seleciona o trabalho com a menor proporção de tempo até a data de vencimento e o tempo total restante de processamento (Liu; Piplani; Toro, 2022).
Regra de sequenciamento	Creation time (CT)	Seleciona o trabalho com a data de criação mais antiga (Sakr *et al.*, 2023).
Regra de sequenciamento	Earliest due date (EDD)	Seleciona o trabalho com a data de vencimento mais próxima (Shiue; Lee; Su, 2022)
Regra de sequenciamento	First in first out (FIFO)	Seleciona o trabalho que chegou primeiro na fila (Shiue; Lee; Su, 2022)
Regra de sequenciamento	Fewest number of operations remaining (FOPR)	Seleciona o trabalho com o menor número de operações restantes (Jeon *et al.*, 2022).
Regra de sequenciamento	Last due date (LDD)	Seleciona o trabalho com a data de vencimento mais recente (Rabelo; Jones; Yih, 1994).
Regra de sequenciamento	Last in first out (LIFO)	Seleciona o último trabalho que chegou à fila (Rabelo; Jones; Yih, 1994).
Regra de sequenciamento	Longest processing time (LPT)	Seleciona o trabalho com o tempo de processamento mais longo (Gerpott *et al.*, 2022).
Regra de sequenciamento	Largest setup time (LST)	Seleciona o trabalho com o maior tempo de preparação (Rabelo; Jones; Yih, 1994).
Regra de sequenciamento	Least slack time (LST)	Seleciona o trabalho com a menor diferença entre o tempo até a data de vencimento e o tempo de processamento restante (tempo livre) (Liu; Piplani; Toro, 2022).
Regra de sequenciamento	Longest transfer time (LTT)	Seleciona o trabalho com o tempo de transferência mais longo (Chen *et al.*, 2022).
Regra de sequenciamento	Most number of operations remaining (MOPR)	Seleciona o trabalho com o maior número de operações restantes (Jeon *et al.*, 2022).
Regra de sequenciamento	Product importance (PI)	Seleciona o trabalho de maior importância (Sakr *et al.*, 2023).
Regra de sequenciamento	Similar setup preferred (SIMSET)	Seleciona o trabalho com a configuração semelhante à da operação real (Heger; Voss, 2020).
Regra de sequenciamento	Shortest imminent operation time (SIO)	Seleciona o trabalho com o menor tempo de processamento no próximo recurso (Shiue; Lee; Su, 2022).
Regra de sequenciamento	Shortest processing time (SPT)	Seleciona o trabalho com o menor tempo de processamento neste recurso (Liu; Piplani; Toro, 2022).
Regra de sequenciamento	Shortest remaining processing time (SRPT)	Seleciona o trabalho com o menor tempo de processamento restante (Shiue; Lee; Su, 2020).
Regra de sequenciamento	Shortest setup time (SST)	Seleciona o trabalho com o menor tempo de configuração (Rabelo; Jones; Yih, 1994).
Regra de sequenciamento	Shortest transfer time (STT)	Seleciona o trabalho com o menor tempo de transferência (Chen *et al.*, 2022)
Regra de sequenciamento	Work in next queue (WINQ)	Seleciona o trabalho com base na menor fila do recurso subsequente (Liu; Piplani; Toro, 2022).
Variáveis independentes	Buffer size	O nível alvo de um buffer específico, medido em tempo, peças ou capacidade (Cox III; Schleier Jr., 2010)
Variáveis independentes	Demand	O número de produtos encomendados pelos clientes em um período de tempo específico (Rummukainen; Nurminen, 2019).
Variáveis independentes	Due date	A data em que a ordem deve ser concluída.
Variáveis independentes	Prioridade da ordem	A importância relativa ou a urgência atribuída a uma ordem específica dentro da programação de produção
Variáveis	Jobs arrival rate	A taxa na qual novas ordens entram no sistema de produção

Grupo	Código	Definição
independentes		
Variáveis independentes	Lot size	A quantidade de unidades processadas em uma única ordem de produção.
Variáveis independentes	Lot size Tansferation	A quantidade de unidades transferidas entre centros de trabalho.
Variáveis independentes	Process/Job Specifics	Especificidades de uma ordem ou processo.
Variáveis independentes	Processing Time	A quantidade de tempo necessária para processar uma ordem com um determinado recurso.
Variáveis independentes	Product mix	A variedade e a combinação de diferentes produtos produzidos em um determinado período de tempo.
Variáveis independentes	Resource Capacity	A capacidade produtiva de um recurso medida em tempo.
Variáveis independentes	Setup Time	O tempo necessário para preparar um recurso para processar um trabalho específico.
Variáveis independentes	TBF	Tempo entre falhas.
Variáveis independentes	TTR	Tempo para reparo.
Variáveis independentes	Transfer time	O tempo necessário para transferir uma ordem entre centros de trabalho.
Variáveis dependentes	Constraint utilization	A relação entre o tempo utilizado e disponível na restrição.
Variáveis dependentes	Early deliveries jobs	Ordens concluídas e entregues antes das datas de vencimento programadas.
Variáveis dependentes	Finished jobs	Ordens concluídas no sistema.
Variáveis dependentes	Flow Time	O tempo gasto para que uma ordem atravesse o sistema de produção, excluindo a espera do backlog.
Variáveis dependentes	Inventory	A quantidade de matérias-primas, ordens em andamento ou produtos acabados presentes na fábrica em um determinado momento.
Variáveis dependentes	Late deliveries jobs	Ordens concluídas e entregues após as datas de vencimento programadas.
Variáveis dependentes	Lead Time	O tempo total gasto para que uma ordem passe por todo o sistema de produção, incluindo o backlog.
Variáveis dependentes	Mean cycle time	O tempo médio para processar uma ordem.
Variáveis dependentes	Net profit	O lucro total obtido pela fábrica após a dedução de todas as despesas e custos operacionais
Variáveis dependentes	Number of jobs in backlog	A contagem de ordens pendentes que ainda não foram liberadas para o chão de fábrica.
Variáveis dependentes	Number of jobs queue	O número total de ordens aguardando em filas.
Variáveis dependentes	Operation Expenses	Os custos incorridos pela fábrica em suas operações, incluindo custos indiretos, mão de obra e outros custos operacionais.
Variáveis dependentes	Protective capacity	A capacidade adicional incorporada ao sistema de produção para lidar com interrupções ou flutuações na demanda.
Variáveis dependentes	Remaining Time	O tempo restante para concluir um processo de uma ordem.
Variáveis dependentes	Resource Availability	O grau de disponibilidade de um recurso.
Variáveis dependentes	Resource State	O estado atual de um recurso.
Variáveis dependentes	Resource utilization	O grau de utilização de um recurso.

Grupo	Código	Definição
Variáveis dependentes	ROI	Retorno sobre o investimento. A relação entre o lucro líquido e o investimento total.
Variáveis dependentes	Scrap Rates	A porcentagem de produtos defeituosos ou inutilizáveis gerados durante um processo.
Variáveis dependentes	Slack time	A quantidade de tempo em que uma ordem pode ser atrasada sem afetar a data de vencimento.
Variáveis dependentes	Tardiness	O grau em que as ordens concluídas excedem suas datas de vencimento
Variáveis dependentes	Throughput	A taxa na qual os produtos são entregues.
Variáveis dependentes	Time untill due date	O tempo restante entre a data de vencimento e a hora atual.
Variáveis dependentes	Wait time in queue	O tempo gasto pelas ordens em uma fila
Variáveis dependentes	Work in process	O número total de ordens que estão sendo processados no momento
Área	Automobile assembly line	Uma linha de montagem síncrona, na qual os automóveis são transferidos simultaneamente por uma série de estações de trabalho de montagem (Al-Aomar, 2006; Jeon *et al.*, 2022)
Área	Automotive	Fabricação de componentes específicos para uso na fabricação de automóveis (Dohale; Ambilkar; Bilolikar, 2021; Golmohammadi, 2015)
Área	Circuit board	Produção ou montagem de componentes ou dispositivos eletrônicos e semicondutores (Altenmüller *et al.*, 2020; Sakr *et al.*, 2023).
Área	Forniture	Fabricação e montagem de itens de mobiliário (Wu; Morris; Gordon, 1994)
Área	Machine fabrication	Fabricação de peças para máquinas-ferramenta (Zhang; Du, 2015).
Área	Remanufacture	O processo de desmontagem, restauração e reconstrução de produtos ou componentes usados (Guide, 1995; Ma; Chen, 2009).
Área	Shipyard	Produção de peças para navios e outras embarcações marítimas (Woo *et al.*, 2021).
Tipo de modelo	Real problems	Simulação de um problema real com dados reais (Jahangirian *et al.*, 2010).
Tipo de modelo	Hypothetical Problem	Simulação de um problema real com dados sintéticos (Jahangirian *et al.*, 2010).
Tipo de modelo	Methodological papers	Simulação de um problema teórico com dados sintéticos (Jahangirian *et al.*, 2010).
Tipo de manufatura	Flexible Manufactory	Uma produção projetada para se adaptar facilmente a mudanças no tipo, volume ou design do produto, usando equipamentos e processos versáteis (Yadav; Jayswal, 2018).
Tipo de manufatura	Flow Shop	Produção em que os processos são organizados em uma sequência linear, com cada etapa levando à próxima em um fluxo contínuo (Garey; Johnson; Sethi, 1976).
Tipo de manufatura	Job shop	Produção em que os produtos seguem diferentes fluxos de processamento (Garey; Johnson; Sethi, 1976).
Tipo de manufatura	Reentrant flow	Produção em que os produtos podem retornar a etapas anteriores do proceso (Graves *et al.*, 1983).
Variável genérica	Tempo atual do sistema	O tempo atual do sistema (Liu *et al.*, 2022)
Variável genérica	Completion to Schedule	A proporção de completude do planejamento (Al-Aomar, 2006; Guide, 1995)
Variável genérica	Mean Slack time of queue	A quantidade média de tempo disponível (antes da data de vencimento) para as ordens em fila (Shiue; Lee; Su, 2022).
Variável genérica	Number of unprocesed jobs	A quantidade de ordens que não foram processadas (Guide, 1995)
Variável	Part features	Característica específica de uma peça (Sakr *et al.*, 2023)

Grupo	Código	Definição
genérica		
Variável genérica	Raw material Units	A quantidade de matéria-prima disponível (Paternina-Arboleda; Das, 2005; Tang; Salonitis, 2021)
Variável genérica	Released Jobs	A quantidade de ordens liberadas para produção (Chen *et al.*, 2022; Liu *et al.*, 2022)
Variável genérica	Remaining parts per order	A quantidade de peças aguardando processamento por ordem de produção (Jeon *et al.*, 2022)
Variável genérica	Setup cost	Custos associados à preparação de máquinas (Rummukainen; Nurminen, 2019)
Variável genérica	Quantidade de preparação da ordem	Número de preparações de máquina em uma ordem (Yu; Ram, 2006)
Variável genérica	Shortage	Falta de materiais para processamento (Al-Aomar, 2006)
Variável genérica	Target Level	O tamanho definido do pulmão na Teoria das Restrições (Cox III; Schleier Jr., 2010)
Variável genérica	Time constraint	O limite de tempo imposto à duração de um trabalho (Altenmüller *et al.*, 2020)
Variável genérica	Buffer status	O estado atual do pulmão na Teoria das Restrições (Schragenheim; Ronen, 1991)
Variável genérica	Constraint Resource	O recurso que limita a produção do sistema (Goldratt, 2006)
Variável genérica	Job ID	A identificação da ordem de produção (Thomas *et al.*, 2018)
Variável genérica	Production Scheduling	O planejamento agendado da produção (Lang *et al.*, 2020; Liu *et al.*, 2022)
Variável genérica	Resource ID	A identificação do recurso (Thomas *et al.*, 2018)

Fonte: Elaborado pelo autor

Tabela 15 - Variáveis independentes nas simulações

Tag	Variável	Tipo da fonte de variabilidade
A01	Demanda	Determinística
	Capacidade dos recursos	Determinística
	Tempo de processamento	Estocástica
	Tempo de preparação	Estocástica
A02	Tempo de processamento	Estocástica
	Tempo de preparação	Estocástica
	TBF	Estocástica
	TTR	Estocástica
A03	Demanda	Determinística
	Tamanho do lote	Determinística
	Lote de transferência	Determinística
	Tempo de processamento	Determinística
A04	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
	Tempo de preparação	Estocástica
A05	Data de entrega	Determinística
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A06	Demanda	Determinística

Tag	Variável	Tipo da fonte de variabilidade
	Tamanho do lote	Determinística
	Tempo de processamento	Determinística
	Tempo de preparação	Determinística
	TBF	Determinística
A07	Data de entrega	Determinística
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A08	Demanda	Determinística
	Tempo de processamento	Estocástica
	TBF	Estocástica
	TTR	Estocástica
A09	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A10	Demanda	Determinística
	Tempo de processamento	Determinística
	Tempo de preparação	Estocástica
	TBF	Estocástica
	TTR	Estocástica
A11	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
	TBF	Estocástica
	TTR	Estocástica
A12	Tempo de processamento	Estocástica
	TBF	Estocástica
	TTR	Estocástica
A13	Tempo de processamento	Determinística
	Taxa de chegada de ordens	Estocástica
A14	Tempo de processamento	Estocástica
	Tempo de preparação	Estocástica
A15	Demanda	Determinística
	Tempo de processamento	Determinística
	Tempo de preparação	Determinística
A16	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A17	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A18	Tempo de processamento	Determinística
A19	Demanda	Estocástica
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A20	Demanda	Estocástica
	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica

Tag	Variável	Tipo da fonte de variabilidade
	Especificidade do processo	Estocástica
	Tempo de processamento	Estocástica
	Tempo de preparação	Estocástica
A21	Data de entrega	Determinística
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A22	TBF	Estocástica
	TTR	Estocástica
A23	Tempo de processamento	Determinística
	Tempo de preparação	Determinística
	Demanda	Estocástica
A25	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A26	Data de entrega	Determinística
	Tempo de processamento	Determinística
	Taxa de chegada de ordens	Estocástica
A27	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
	Mix de produtos	Estocástica
A28	Tempo de processamento	Determinística
	Taxa de chegada de ordens	Estocástica
	Tempo de transferência	Estocástica
A29	Tempo de processamento	Estocástica
A30	Demanda	Determinística
	Tempo de processamento	Determinística
	Tempo de preparação	Determinística
	Demanda	Estocástica
	Taxa de chegada de ordens	Estocástica
A31	Taxa de chegada de ordens	Determinística
	Tempo de processamento	Estocástica
	Tempo de preparação	Estocástica
	TBF	Estocástica
	TTR	Estocástica
A32	Data de entrega	Determinística
	Tempo de preparação	Determinística
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
	Tempo de transferência	Estocástica
A33	Taxa de chegada de ordens	Estocástica
A34	Data de entrega	Determinística
	Tempo de processamento	Determinística
A35	Tempo de processamento	Determinística
	Tempo de transferência	Não especificada

Tag	Variável	Tipo da fonte de variabilidade
	Tempo de processamento	Estocástica
A36	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica
A37	Data de entrega	Determinística
	Demanda	Estocástica
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A38	Tempo de processamento	Determinística
	Tempo de preparação	Determinística
	Demanda	Estocástica
A39	Data de entrega	Determinística
	Tempo de preparação	Determinística
	Demanda	Estocástica
	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A40	Tempo de processamento	Determinística
	Tempo de preparação	Determinística
A41	Tempo de processamento	Determinística
	Demanda	Estocástica
A42	Tempo de processamento	Determinística
	Tempo de transferência	Não especificada
A43	Data de entrega	Determinística
	Taxa de chegada de ordens	Não especificada
	Tempo de processamento	Estocástica
	Tempo de preparação	Estocástica
A44	Tempo de processamento	Não especificada
	Tempo de preparação	Não especificada
A45	Tempo de processamento	Estocástica
	TTR	Estocástica
A46	Tempo de processamento	Estocástica
A47	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica
	Tempo de processamento	Estocástica
A48	Tempo de processamento	Estocástica
A49	Data de entrega	Estocástica
	Taxa de chegada de ordens	Estocástica
A50	Demanda	Não especificada
	Taxa de chegada de ordens	Não especificada
	Tempo de processamento	Não especificada
	Capacidade dos recursos	Não especificada
	Taxa de chegada de ordens	Estocástica
	TBF	Estocástica
	TTR	Estocástica

Fonte: Elaborado pelo autor

Tabela 16 - Resumo da análise do ambiente virtual

Tag	Grupo	Manufatura flexível	Flow Shop	Job shop	Fluxo reentrante	Validation	Software
A01	TOC	-	-	X	X	Genérico	Não especificado
A02	TOC	-	-	X	X	Genérico	Não especificado
A03	TOC	-	X	-	-	Moveleira	Siman
A04	TOC	-	X	-	-	Remanufatura	Slam II
A05	TOC	-	-	X	-	Genérico	Não especificado
A06	TOC	-	-	X	-	Genérico	ARENA
A07	TOC	-	-	X	-	Genérico	AWESIM
A08	TOC	-	X	-	-	Linha de montagem automotiva	Witness
A09	TOC	-	X	-	-	Genérico	Não especificado
A10	TOC	-	X	-	-	Fabricação peças automotivas	Witness
A11	TOC	-	X	-	-	Genérico	ARENA
A12	TOC	-	-	X	-	Remanufatura	ARENA
A13	TOC	-	X	-	-	Genérico	Simio
A14	TOC	-	-	X	-	Fabricação peças automotivas	ARENA
A15	TOC	-	-	-	-	Fabricação de máquinas ferramenta	Flexim
A16	TOC	-	X	-	-	Genérico	SimPy
A17	TOC	-	X	-	-	Genérico	SimPy
A18	TOC	-	X	-	-	Fabricação peças automotivas	ARENA
A19	TOC	-	X	-	-	Genérico	SimPy
A20	RL	-	-	X	-	Genérico	C-linguage
A21	RL	-	-	X	-	Genérico	Não especificado
A22	RL	-	X	-	-	Genérico	Batch Control Language
A23	RL	X	-	-	-	Genérico	ARENA
A24	RL	-	X	-	-	Genérico	JADE AP
A25	RL	-	X	-	-	Genérico	Java - Linguage Prog
A26	RL	-	-	X	-	Genérico	Promodel 6.0
A27	RL	X	-	-	-	Genérico	Tecnomatix Plant Simulation
A28	RL	-	-	X	-	Genérico	AnyLogic
A29	RL	X	-	X	-	Genérico	Matlab
A30	RL	-	-	-	-	Genérico	Python
A31	RL	-	-	X	X	Fabricação de eletrônicos	Não especificado
A32	RL	-	-	X	X	Genérico	AnyLogic
A33	RL	-	-	X	-	Genérico	Não especificado
A34	RL	-	-	X	-	Genérico	Salabim
A35	RL	-	X	-	X	Genérico	Python
A36	RL	X	-	X	-	Genérico	Tecnomatix Plant Simulation
A37	RL	-	-	X	-	Genérico	Python
A38	RL	-	-	X	-	Genérico	SimPy
A39	RL	X	-	-	X	Fabricação de eletrônicos	Não especificado
A40	RL	X	-	-	-	Genérico	Não especificado
A41	RL	-	X	-	-	Embarcações	SimPy
A42	RL	-	X	-	-	Genérico	Simio

Tag	Grupo	Manufatura flexível	Flow Shop	Job shop	Fluxo reentrante	Validation	Software
A43	RL	-	X	-	-	Fabricação de eletrônicos	Salabim
A44	RL	-	-	X	X	Linha de montagem automotiva	Siemens Plant Simulator
A45	RL	-	X	-	-	Genérico	Não especificado
A46	RL	X	-	X	-	Genérico	SimPy
A47	RL	-	X	-	-	Genérico	AnyLogic
A48	RL	-	-	X	-	Genérico	SimPy
A49	RL	-	-	X	-	Genérico	Tecnomatix Plant Simulation
A50	RL	-	-	X	-	Fabricação de eletrônicos	ARENA

Fonte: Elaborado pelo autor

Tabela 17 - Resumo da análise da Teoria das Restrições

Tag	Objetivos	Pulmões utilizados	Variáveis independentes	Variáveis dependentes
A01	Aplicabilidade	Montagem, Restrição, Expedição	Sequenciamento	Produtividade
A02	Otimização de parâmetros do DBR	Montagem, Restrição, Expedição	Tamanho do pulmão	Status do pulmão
A03	Aplicabilidade	Restrição	Sequenciamento	Tempo de atravessamento, Utilização dos recursos
A04	Aplicabilidade	Montagem	Sequenciamento	Proporção de completude do planejamento, Ordens não processadas
A05	Avaliar o impacto da capacidade protetiva	Restrição, Expedição, Capacidade	Sequenciamento, Capacidade protetiva, Taxa de chegada das ordens, Utilização dos recursos	Tempo de atravessamento, Ordens entregue atrasadas, Atraso das ordens, Ordens em processamento
A06	Aplicabilidade	Não especificado	Tamanho do lote, Tempo de processamento, Tempo de preparação, TBF	Lucro líquido, Despesa operacional, ROI, Ordens finalizadas, Tempo médio de ciclo, Utilização dos recursos, Produtividade, Ordens em processamento
A07	Análise do impacto de produtos livres	Restrição, Expedição	Utilização da restrição, Produtos livres, Taxa de chegada das ordens	Tempo de atravessamento, Atraso das ordens, Ordens em processamento
A08	Aplicabilidade	Não especificado	Sequenciamento	Falta de material, Proporção de completude do planejamento, Ordens entregue antecipadas, Ordens entregue atrasadas, Atraso das ordens, Produtividade
A09	Análise do impacto de produtos livres, Avaliar o impacto da capacidade protetiva	Restrição, Expedição, Capacidade	Capacidade protetiva, Taxa de chegada das ordens	Ordens entregue antecipadas, Ordens finalizadas
A10	Otimização de parâmetros do DBR, Identificação da restrição	Restrição	Tamanho do pulmão, Lote de transferência	Tempo de atravessamento, Ordens entregue atrasadas, Atraso das ordens
A11	Aplicabilidade	Montagem, Restrição, Expedição, Space Buffer, Capacidade	Restrição, Sequenciamento	Inventário, Tempo de atravessamento, Quantidade de ordens na fila, Produtividade, Tempo de espera na fila
A12	Otimização de parâmetros do DBR	Restrição	Tamanho do pulmão	Tempo de atravessamento, Utilização dos recursos, Ordens em processamento
A13	Aplicabilidade	Restrição	Sequenciamento	Tempo de atravessamento, Ordens entregue atrasadas, Atraso das ordens
A14	Aplicabilidade, Análise do impacto de produtos livres, Variables impact	Montagem, Restrição	Release time, Taxa de chegada das ordens, Tamanho do lote	Lucro líquido, Ordens finalizadas, Utilização dos recursos, Ordens em processamento
A15	Otimização de parâmetros do DBR, Identificação da restrição	Montagem, Restrição, Expedição	Tamanho do pulmão, Tamanho do lote	Tempo de atravessamento, Atraso das ordens
A16	Melhorias no DBR	Restrição	Tamanho do pulmão, Regra de liberação, Sequenciamento	Tempo de atravessamento, Ordens entregue atrasadas, Lead Time, Atraso das ordens
A17	Identificação da restrição	Restrição, Capacidade	Restrição	Tempo de atravessamento, Ordens entregue atrasadas, Lead Time, Atraso das ordens
A18	Aplicabilidade	Não especificado	Identificação da restrição, Sequenciamento	Ordens finalizadas, Utilização dos recursos
A19	Melhorias no DBR	Expedição	Restrição, Sequenciamento	Valor de referência do pulmão, Tempo de preparação, Tempo de atravessamento, Utilização dos recursos, Ordens em processamento

Fonte: Elaborado pelo autor

Tabela 18 - Resumo da análise de Aprendizado por Reforço

Tag	Algoritmo	Variáveis utilizadas no estado observável	Ações do agente	Variáveis utilizadas na recompensa
A20	Q-learning	Data de entrega, Taxa de chegada das ordens, Especificidades do processo, Tempo de processamento, Tempo de preparação	CR - Taxa de criticidade, EDD - Menor tempo de entrega, FIFO - Primeiro a entrar, primeiro a sair, LDD - Maior tempo de entrega, LIFO - Último a entrar, primeiro a sair, LPT - Maior tempo de processamento, LST - Maior tempo de preparação, LST - Menor tempo disponível até entrega, SPT - Menor tempo de processamento, SST - Menor tempo de preparação	Ordens em processamento
A21	Q-learning	Tempo médio disponível na fila, Quantidade de ordens na fila	CONVERT - C over T, CR - Taxa de criticidade, SPT - Menor tempo de processamento	Quantidade de ordens na fila, Tempo disponível até entrega
A22	Não especificado	Ordens em processamento	Não especificado	Custo de preparação, Inventário, Despesa operacional
A23	Last Mean Square	Quantidade de ordens para liberação, Quantidade de ordens na fila	Troca do produto no recurso	Quantidade de ordens para liberação, Quantidade de ordens na fila
A24	Q-learning	Não especificado	Não especificado	Não especificado
A25	Não especificado	Fila dos recursos, Quantidade de ordens na fila, Queue	EDD - Menor tempo de entrega, FIFO - Primeiro a entrar, primeiro a sair, SPT - Menor tempo de processamento, Quantidade de operadores	Utilização dos recursos, Atraso das ordens
A26	Não especificado	Não especificado	Não especificado	Não especificado
A27	Q-learning	Fila dos recursos, Especificidades do processo, Tempo de processamento, Quantidade de ordens na fila, Tempo restante de processamento, Utilização dos recursos, Tempo disponível até entrega, Atraso das ordens, Ordens em processamento	EDD - Menor tempo de entrega, LST - Menor tempo disponível até entrega, SIO - Menor tempo da próxima operação, SPT - Menor tempo de processamento, SRPT - Menor tempo de processamento restante	Produtividade
A28	Deep-Q-Learning	Quantidade de ordens na fila, Estado dos recursos	Identificador das ordens	Produtividade
A29	Deep-Q-Learning	Identificador dos recursos, Identificador das ordens, Tempo restante de processamento, Disponibilidade dos recursos, Estado dos recursos	Identificador dos recursos	Capacidade dos recursos, Ordens em processamento
A30	PPO	Ordens finalizadas, Estado dos recursos	Identificador das ordens, Não fazer nada	Custo de preparação, Inventário, Ordens entregue atrasadas
A31	Deep-Q-Learning	Identificador dos recursos, Fila dos recursos, Quantidade de preparação da ordem, Prioridade da ordem, Quantidade de ordens na fila, Estado dos recursos	Identificador das ordens, Não fazer nada	Restrição de tempo, Ordens em processamento
A32	Deep-Q-Learning	Fila dos recursos, Quantidade de ordens na fila, Utilização dos recursos, Atraso das ordens	EDD - Menor tempo de entrega, FIFO - Primeiro a entrar, primeiro a sair, SIMSET - Preparação mais similar, SPT - Menor tempo de processamento	Atraso das ordens
A33	Deep-Q-Learning	Tempo atual do sistema, Estado dos recursos	Não especificado	Tempo de processamento, Tempo de atravessamento

Tag	Algoritmo	Variáveis utilizadas no estado observável	Ações do agente	Variáveis utilizadas na recompensa
A34	Deep-Q-Learning	Fila dos recursos, Data de entrega, Especificidades do processo, Tempo de processamento, Tempo médio de ciclo, Quantidade de ordens na fila, Tempo restante de processamento	Alocação ao recurso, Selecionar a sequencia de operação	Tempo de atravessamento, Atraso das ordens
A35	Deep-Q-Learning	Especificidades do processo, Estado dos recursos	Não fazer nada, Transferir uma ordem	Não especificado
A36	Q-learning	Fila dos recursos, Tempo de processamento, Quantidade de ordens na fila, Tempo restante de processamento, Utilização dos recursos, Atraso das ordens	EDD - Menor tempo de entrega, LST - Menor tempo disponível até entrega, SIO - Menor tempo da próxima operação, SPT - Menor tempo de processamento, SRPT - Menor tempo de processamento restante	Ordens entregue atrasadas, Tempo médio de ciclo, Produtividade, Parâmetros de performance
A37	Deep-Q-Learning	Fila dos recursos, Data de entrega, Identificador das ordens, Especificidades do processo, Tempo de processamento, Quantidade de ordens na fila, Disponibilidade dos recursos, Estado dos recursos, Operações da ordem	Identificador das ordens	Não especificado
A38	Deep-Q-Learning	Identificador dos recursos, Identificador das ordens, Ordens em processamento	Identificador dos recursos	Lead Time
A39	Deep-Q-Learning	Fila dos recursos, Data de entrega, Especificidades do processo, Quantidade de ordens na fila, Estado dos recursos, Tempo de espera na fila	Identificador das ordens	Atraso das ordens
A40	Dueling double deep-Q-Learning	Quantidade de matéria-prima, Especificidades do processo, Ordens finalizadas, Estado dos recursos	Não fazer nada, Solicitar material bruto	Ordens finalizadas, Quantidade de ordens para liberação, Estado dos recursos
A41	Deep-Q-Learning, Double Deep-Q-Learning	Tempo de processamento, Tempo restante de processamento	Identificador das ordens	Tempo de processamento, Lead Time
A42	Double Deep-Q-Learning	Não especificado	Prioridade da fila	Tempo de atravessamento
A43	A2C	Tempo de processamento, Quantidade de ordens na fila, Utilização dos recursos, Atraso das ordens	CR - Taxa de criticidade, EDD - Menor tempo de entrega, LPT - Maior tempo de processamento, LST - Menor tempo disponível até entrega, SPT - Menor tempo de processamento	Tempo de atravessamento, Atraso das ordens
A44	Double Deep-Q-Learning	Peças restantes por ordem, Ordens em processamento	FIFO - Primeiro a entrar, primeiro a sair, FOPR - Menor número de operações faltantes, MOPR - Maior número de operações faltantes	Tempo de atravessamento
A45	A3C, CNN-A3C	Tempo atual do sistema, Ordens liberadas, Identificador dos recursos, Fila dos recursos, Data de entrega, Identificador das ordens, Especificidades do processo, Ordens finalizadas, Tempo restante de processamento, Estado dos recursos, Tempo de espera na fila, Ordens em processamento	Prioridade da ordem, Liberar uma parcela da ordem	Data de entrega

Tag	Algoritmo	Variáveis utilizadas no estado observável	Ações do agente	Variáveis utilizadas na recompensa
A46	Deep-Q-Learning	Tempo de processamento, Quantidade de ordens na fila, Tempo restante de processamento, Disponibilidade dos recursos, Tempo disponível até entrega, Tempo até a data de entrega, Ordens em processamento	CR - Taxa de criticidade, LST - Menor tempo disponível até entrega, SPT - Menor tempo de processamento, WINQ – Quantidade de ordens na fila	Tempo disponível até entrega
A47	Deep-Q-Learning	Data de entrega, Prioridade da ordem	Identificador das ordens	Data de entrega, Prioridade da ordem, Tempo de processamento
A48	Deep-Q-Learning	Não especificado	Não especificado	Data de entrega, Produtividade
A49	Q-learning	Tempo médio disponível na fila, Tempo de processamento, Quantidade de ordens na fila, Utilização dos recursos, Tempo disponível até entrega, Atraso das ordens, Tempo até a data de entrega, Ordens em processamento	EDD - Menor tempo de entrega, FIFO - Primeiro a entrar, primeiro a sair, LST - Menor tempo disponível até entrega, SIO - Menor tempo da próxima operação, SPT - Menor tempo de processamento, SRPT - Menor tempo de processamento restante	Ordens entregue atrasadas, Tempo médio de ciclo, Produtividade
A50	Deep-Q-Learning	Características da peça, Fila dos recursos, Especificidades do processo, Flow Time, Quantidade de ordens na fila, Estado dos recursos	CR - Taxa de criticidade, CT – Data de criação, EDD - Menor tempo de entrega, PI – Importância do produto	Fila dos recursos, Especificidades do processo, Quantidade de ordens na fila, Utilização dos recursos, Tempo de espera na fila

Fonte: Elaborado pelo autor

APENDICE II

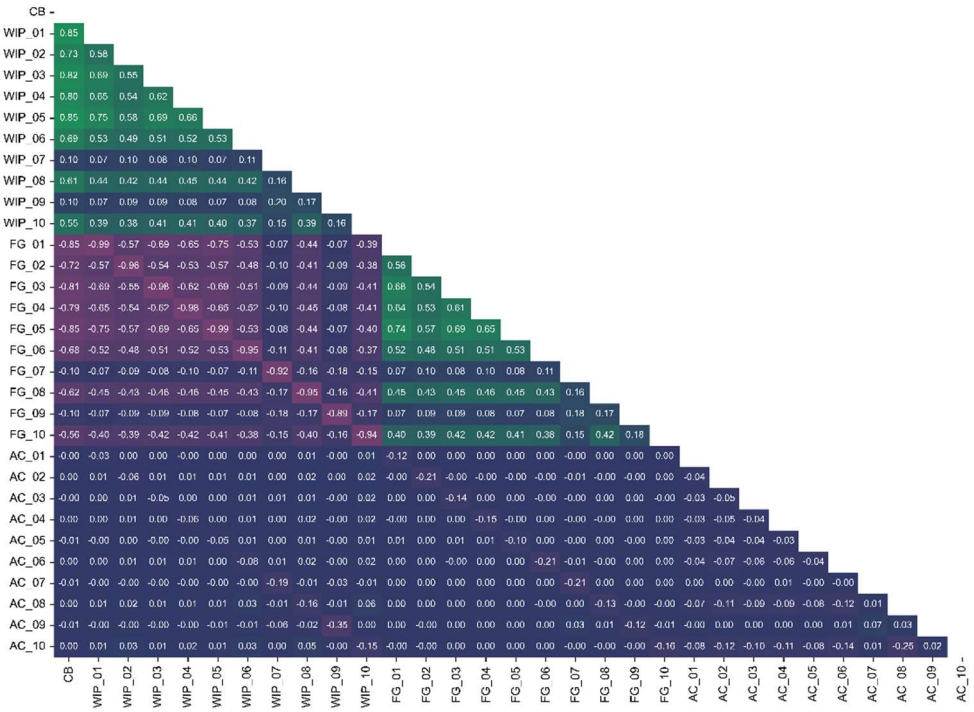
Tabela 19 - Teste de hipótese para o desempenho dos métodos de sequenciamento

	DBR		RL		h0: DBR = RL h1: DBR ≠ RL	
Métrica	Média	Desvio Padrão	Média	Desvio Padrão	Resultado	p-valor
FT	203,157	10,838	180,570	3,048	h1*	≈0,000
LS	7,370	0,022	7,486	0,015	h1*	0,001
WIP	209,071	11,258	171,507	2,998	h1*	≈0,000
FG	577,531	11,433	338,656	13,696	h1*	≈0,000
IV	786,602	0,225	510,163	10,831	h1*	≈0,000
SP	399.913,860	321,188	399.967,700	314,344	h0*	0,399
DM	399.938,360	326,313	400.053,340	357,060	h0*	0,099

*Wilcox para amostras não pareadas

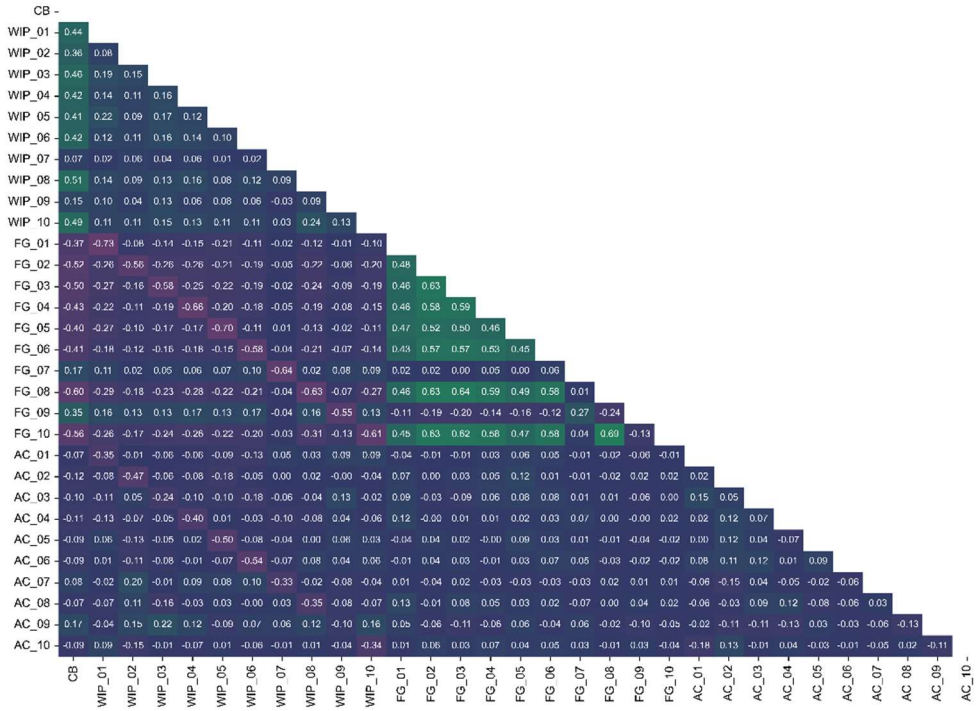
Fonte: Elaborado pelo autor

Figura 16 - Correlação de Person para variáveis no DBR



Fonte: Elaborado pelo autor

Figura 17 - Correlação de Person para variáveis no RL



Fonte: Elaborado pelo autor

Tabela 20 - Resultado das regressões lineares para o DBR

Ações	AC_01		AC_02		AC_03		AC_04		AC_05		AC_06		AC_07		AC_08		AC_09		AC_10	
r²	0,925		0,907		0,904		0,891		0,896		0,894		0,999		0,815		0,998		0,833	
	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*
Constante	7,195	0,00	7,199	0,00	7,196	0,00	7,198	0,00	7,198	0,00	7,199	0,00	7,201	0,00	7,198	0,00	7,200	0,00	7,200	0,00
CB	0,065	0,00	0,056	0,00	0,051	0,00	0,061	0,00	0,070	0,00	0,050	0,00	-0,002	0,00	0,072	0,00	-0,003	0,03	0,034	0,01
WIP_01	**-10,649**	**0,00**	0,407	0,00	0,258	0,00	0,325	0,00	0,216	0,00	0,483	0,00	-0,003	0,00	0,955	0,00	-0,012	0,00	1,069	0,00
WIP_02	0,143	0,00	**-7,665**	**0,00**	0,204	0,00	0,256	0,00	0,172	0,00	0,369	0,00	-0,002	0,00	0,733	0,00	-0,006	0,00	0,833	0,00
WIP_03	0,186	0,00	0,352	0,00	**-9,132**	**0,00**	0,297	0,00	0,236	0,00	0,458	0,00	-0,003	0,00	0,895	0,00	-0,009	0,00	0,989	0,00
WIP_04	0,178	0,00	0,314	0,00	0,251	0,00	**-8,553**	**0,00**	0,197	0,00	0,415	0,00	-0,002	0,00	0,861	0,00	-0,010	0,00	0,903	0,00
WIP_05	0,250	0,00	0,427	0,00	0,290	0,00	0,317	0,00	**-10,380**	**0,00**	0,505	0,00	-0,003	0,00	1,007	0,00	-0,014	0,00	1,101	0,00
WIP_06	0,151	0,00	0,310	0,00	0,218	0,00	0,240	0,00	0,163	0,00	**-7,362**	**0,00**	-0,002	0,00	0,738	0,00	-0,007	0,00	0,865	0,00
WIP_07	0,001	0,59	0,001	0,73	0,005	0,08	-0,004	0,17	0,003	0,32	0,003	0,43	**-4,311**	**0,00**	-0,006	0,30	0,005	0,00	-0,002	0,69
WIP_08	0,203	0,00	0,395	0,00	0,289	0,00	0,314	0,00	0,235	0,00	0,493	0,00	-0,003	0,00	**-7,356**	**0,00**	-0,016	0,00	1,132	0,00
WIP_09	-0,001	0,57	-0,004	0,18	-0,003	0,21	-0,007	0,01	-0,008	0,00	-0,008	0,02	0,002	0,00	0,018	0,00	**-5,919**	**0,00**	0,016	0,00
WIP_10	0,197	0,00	0,395	0,00	0,287	0,00	0,344	0,00	0,223	0,00	0,489	0,00	-0,002	0,00	1,066	0,00	-0,011	0,00	**-7,778**	**0,00**
FG_01	**-10,743**	**0,00**	0,427	0,00	0,281	0,00	0,348	0,00	0,239	0,00	0,509	0,00	-0,004	0,00	0,993	0,00	-0,013	0,00	1,113	0,00
FG_02	0,155	0,00	**-7,879**	**0,00**	0,222	0,00	0,273	0,00	0,188	0,00	0,390	0,00	-0,002	0,00	0,769	0,00	-0,008	0,00	0,863	0,00
FG_03	0,208	0,00	0,377	0,00	**-9,251**	**0,00**	0,321	0,00	0,256	0,00	0,483	0,00	-0,004	0,00	0,937	0,00	-0,011	0,00	1,021	0,00
FG_04	0,200	0,00	0,343	0,00	0,274	0,00	**-8,676**	**0,00**	0,221	0,00	0,437	0,00	-0,003	0,00	0,899	0,00	-0,011	0,00	0,931	0,00
FG_05	0,274	0,00	0,461	0,00	0,309	0,00	0,347	0,00	**-10,438**	**0,00**	0,538	0,00	-0,003	0,00	1,051	0,00	-0,014	0,00	1,139	0,00
FG_06	0,160	0,00	0,323	0,00	0,229	0,00	0,258	0,00	0,176	0,00	**-7,567**	**0,00**	-0,002	0,00	0,771	0,00	-0,009	0,00	0,889	0,00
FG_07	0,010	0,00	0,013	0,00	0,015	0,00	0,006	0,08	0,011	0,00	0,014	0,00	**-4,329**	**0,00**	0,005	0,35	0,002	0,00	0,011	0,07
FG_08	0,212	0,00	0,405	0,00	0,295	0,00	0,322	0,00	0,246	0,00	0,498	0,00	-0,003	0,00	**-7,383**	**0,00**	-0,016	0,00	1,125	0,00
FG_09	0,004	0,05	-0,001	0,74	0,003	0,32	0,000	0,97	-0,003	0,20	0,000	0,98	0,002	0,00	0,021	0,00	**-5,585**	**0,00**	0,025	0,00
FG_10	0,205	0,00	0,406	0,00	0,295	0,00	0,353	0,00	0,231	0,00	0,500	0,00	-0,002	0,00	1,082	0,00	-0,012	0,00	**-7,875**	**0,00**

Fonte: Elaborado pelo autor

Tabela 21 - Resultado das regressões lineares para o RL

Ação	AC_01		AC_02		AC_03		AC_04		AC_05		AC_06		AC_07		AC_08		AC_09		AC_10	
r²	0,463		0,504		0,321		0,406		0,487		0,505		0,351		0,391		0,314		0,436	
	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*	Coef.	*p*
Constante	7,193	0,00	7,196	0,00	7,201	0,00	7,203	0,00	7,199	0,00	7,197	0,00	7,199	0,00	7,205	0,00	7,200	0,00	7,199	0,00
CB	0,419	0,00	0,510	0,00	0,555	0,00	0,513	0,00	0,536	0,00	0,799	0,00	0,558	0,00	0,589	0,00	0,464	0,00	0,284	0,00
WIP_01	**-3,207**	**0,00**	-0,059	0,00	-0,235	0,00	-0,099	0,00	0,565	0,00	0,236	0,00	-0,086	0,00	0,135	0,00	0,089	0,00	0,456	0,00
WIP_02	-0,001	0,84	**-2,741**	**0,00**	0,047	0,00	-0,324	0,00	-0,275	0,00	-0,400	0,00	0,629	0,00	0,226	0,00	0,399	0,00	-0,467	0,00
WIP_03	-0,257	0,00	-0,061	0,00	**-1,540**	**0,00**	-0,077	0,00	-0,018	0,00	0,003	0,59	-0,131	0,00	-0,471	0,00	0,437	0,00	0,142	0,00
WIP_04	0,047	0,00	-0,089	0,00	-0,073	0,00	**-2,106**	**0,00**	0,144	0,00	0,014	0,04	0,298	0,00	0,000	0,95	0,136	0,00	-0,010	0,07
WIP_05	0,058	0,00	-0,174	0,00	-0,050	0,00	0,190	0,00	**-3,320**	**0,00**	-0,012	0,12	0,240	0,00	0,239	0,00	-0,392	0,00	0,211	0,00
WIP_06	-0,259	0,00	-0,138	0,00	-0,319	0,00	0,073	0,00	-0,106	0,00	**-3,323**	**0,00**	0,193	0,00	-0,059	0,00	-0,049	0,00	-0,117	0,00
WIP_07	0,404	0,00	0,108	0,00	-0,177	0,00	-0,044	0,00	-0,006	0,33	-0,088	0,00	**-1,696**	**0,00**	0,051	0,00	0,415	0,00	0,215	0,00
WIP_08	-0,007	0,30	-0,132	0,00	-0,167	0,00	-0,359	0,00	-0,142	0,00	0,193	0,00	0,033	0,00	**-1,828**	**0,00**	0,363	0,00	0,106	0,00
WIP_09	0,784	0,00	0,265	0,00	0,547	0,00	0,250	0,00	0,449	0,00	0,276	0,00	-0,205	0,00	-0,060	0,00	**-1,283**	**0,00**	-0,002	0,73
WIP_10	0,365	0,00	-0,123	0,00	-0,128	0,00	-0,252	0,00	0,056	0,00	0,258	0,00	-0,153	0,00	-0,267	0,00	0,476	0,00	**-2,345**	**0,00**
FG_01	**-2,947**	**0,00**	0,464	0,00	0,154	0,00	0,592	0,00	0,397	0,00	0,293	0,00	-0,068	0,00	0,647	0,00	0,333	0,00	0,499	0,00
FG_02	0,139	0,00	**-2,739**	**0,00**	-0,141	0,00	-0,299	0,00	0,380	0,00	0,065	0,00	0,360	0,00	0,262	0,00	0,225	0,00	0,004	0,64
FG_03	-0,347	0,00	0,386	0,00	**-1,819**	**0,00**	0,227	0,00	0,240	0,00	0,551	0,00	-0,018	0,01	0,118	0,00	-0,334	0,00	0,494	0,00
FG_04	0,549	0,00	0,542	0,00	0,503	0,00	**-1,780**	**0,00**	0,193	0,00	-0,077	0,00	0,070	0,00	0,289	0,00	-0,254	0,00	0,579	0,00
FG_05	0,494	0,00	0,662	0,00	0,366	0,00	0,388	0,00	**-2,429**	**0,00**	0,504	0,00	0,126	0,00	0,410	0,00	-0,132	0,00	0,453	0,00
FG_06	0,416	0,00	0,197	0,00	0,367	0,00	0,517	0,00	0,353	0,00	**-2,404**	**0,00**	-0,021	0,00	0,099	0,00	-0,251	0,00	0,383	0,00
FG_07	0,340	0,00	-0,041	0,00	-0,179	0,00	0,166	0,00	-0,007	0,28	0,254	0,00	**-1,289**	**0,00**	-0,176	0,00	0,747	0,00	0,313	0,00
FG_08	-0,184	0,00	-0,062	0,00	0,230	0,00	-0,214	0,00	0,054	0,00	0,279	0,00	0,212	0,00	**-1,748**	**0,00**	0,351	0,00	0,252	0,00
FG_09	0,384	0,00	0,340	0,00	0,204	0,00	0,261	0,00	0,155	0,00	0,325	0,00	-0,018	0,00	0,033	0,00	**-1,442**	**0,00**	0,217	0,00
FG_10	0,450	0,00	0,309	0,00	0,083	0,00	0,007	0,34	-0,096	0,00	0,519	0,00	0,025	0,00	-0,090	0,00	0,314	0,00	**-2,692**	**0,00**

Fonte: Elaborado pelo autor

Tabela 22 - Resultado da Árvore de decisão para o DBR

Ações	AC_01	AC_02	AC_03	AC_04	AC_05	AC_06	AC_07	AC_08	AC_09	AC_10
r²	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
Nodos	9.537	16.691	12.097	14.203	10.441	19.211	957	31.221	1.607	33.651
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
CB	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,02	0,00	0,02
WIP_01	**0,49**	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,00	0,01
WIP_02	0,00	**0,47**	0,00	0,01	0,01	0,01	0,00	0,01	0,00	0,01
WIP_03	0,00	0,00	**0,50**	0,01	0,00	0,01	0,00	0,01	0,00	0,01
WIP_04	0,00	0,00	0,00	**0,49**	0,00	0,01	0,00	0,01	0,00	0,01
WIP_05	0,00	0,00	0,00	0,00	**0,46**	0,01	0,00	0,01	0,00	0,01
WIP_06	0,00	0,01	0,00	0,01	0,00	**0,46**	0,00	0,01	0,00	0,01
WIP_07	0,00	0,00	0,00	0,01	0,00	0,00	**0,55**	0,01	0,00	0,01
WIP_08	0,00	0,01	0,01	0,01	0,01	0,01	0,00	**0,43**	0,00	0,02
WIP_09	0,00	0,01	0,01	0,01	0,00	0,01	0,00	0,01	**0,46**	0,01
WIP_10	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,02	0,00	**0,48**
FG_01	**0,43**	0,00	0,00	0,00	0,01	0,00	0,00	0,01	0,00	0,01
FG_02	0,00	**0,42**	0,01	0,01	0,01	0,01	0,00	0,01	0,00	0,01
FG_03	0,00	0,00	**0,40**	0,00	0,01	0,01	0,00	0,01	0,00	0,01
FG_04	0,00	0,01	0,01	**0,39**	0,00	0,01	0,00	0,01	0,00	0,01
FG_05	0,00	0,00	0,00	0,00	**0,43**	0,01	0,00	0,01	0,00	0,01
FG_06	0,00	0,00	0,01	0,01	0,01	**0,42**	0,00	0,01	0,00	0,01
FG_07	0,00	0,00	0,00	0,00	0,00	0,00	**0,45**	0,01	0,00	0,01
FG_08	0,01	0,01	0,01	0,01	0,01	0,01	0,00	**0,34**	0,00	0,02
FG_09	0,00	0,00	0,00	0,01	0,01	0,00	0,00	0,01	**0,54**	0,01
FG_10	0,01	0,01	0,01	0,01	0,01	0,01	0,00	0,02	0,00	**0,32**

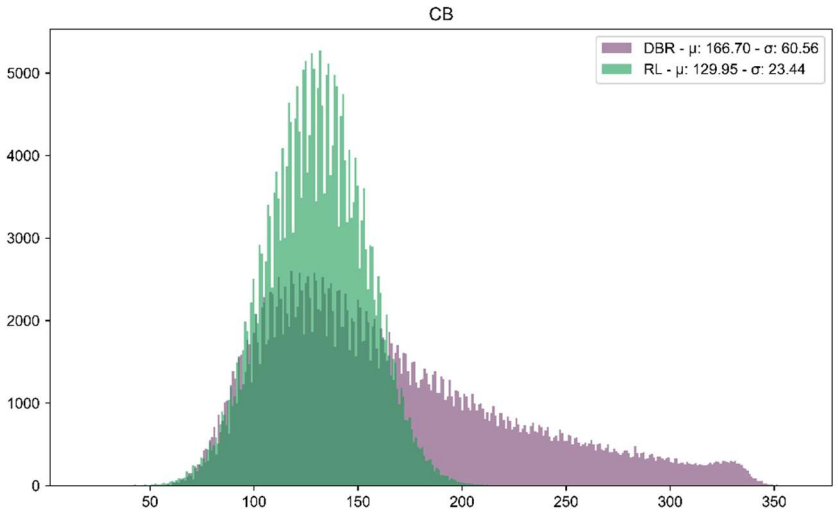
Fonte: Elaborado pelo autor

Tabela 23 - Resultado da árvore de decisão para o RL

Ação	AC_01	AC_02	AC_03	AC_04	AC_05	AC_06	AC_07	AC_08	AC_09	AC_10
r²	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0	1,0
Nodos	156.233	155.093	69.693	159.749	109.731	145.639	112.269	154.091	164.679	133.625
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
CB	0,033	0,029	0,045	0,044	0,032	0,052	0,061	0,049	0,057	0,041
WIP_01	**0,287**	0,016	0,040	0,037	0,051	0,028	0,028	0,028	0,033	0,033
WIP_02	0,020	**0,287**	0,040	0,025	0,021	0,032	0,078	0,043	0,049	0,038
WIP_03	0,016	0,011	**0,136**	0,012	0,013	0,015	0,017	0,035	0,067	0,026
WIP_04	0,023	0,017	0,025	**0,255**	0,023	0,016	0,032	0,019	0,046	0,020
WIP_05	0,019	0,023	0,023	0,042	**0,334**	0,019	0,032	0,030	0,040	0,023
WIP_06	0,025	0,020	0,050	0,036	0,032	**0,379**	0,027	0,028	0,030	0,018
WIP_07	0,023	0,023	0,024	0,025	0,040	0,029	**0,341**	0,029	0,048	0,017
WIP_08	0,029	0,019	0,023	0,028	0,012	0,028	0,018	**0,214**	0,042	0,026
WIP_09	0,054	0,031	0,061	0,027	0,030	0,020	0,034	0,029	**0,098**	0,022
WIP_10	0,052	0,024	0,036	0,034	0,032	0,027	0,016	0,041	0,056	**0,225**
FG_01	**0,161**	0,043	0,031	0,057	0,017	0,026	0,016	0,058	0,037	0,036
FG_02	0,038	**0,158**	0,024	0,025	0,057	0,023	0,021	0,030	0,030	0,037
FG_03	0,026	0,033	**0,146**	0,029	0,024	0,032	0,017	0,032	0,033	0,042
FG_04	0,032	0,057	0,082	**0,121**	0,021	0,020	0,021	0,044	0,027	0,064
FG_05	0,033	0,068	0,044	0,039	**0,147**	0,040	0,021	0,030	0,036	0,032
FG_06	0,042	0,030	0,058	0,046	0,040	**0,126**	0,023	0,037	0,047	0,054
FG_07	0,020	0,023	0,022	0,030	0,016	0,022	**0,123**	0,037	0,058	0,044
FG_08	0,019	0,022	0,038	0,023	0,023	0,021	0,028	**0,124**	0,030	0,030
FG_09	0,026	0,030	0,025	0,026	0,014	0,021	0,024	0,030	**0,101**	0,032
FG_10	0,024	0,037	0,028	0,039	0,019	0,026	0,023	0,032	0,034	**0,141**

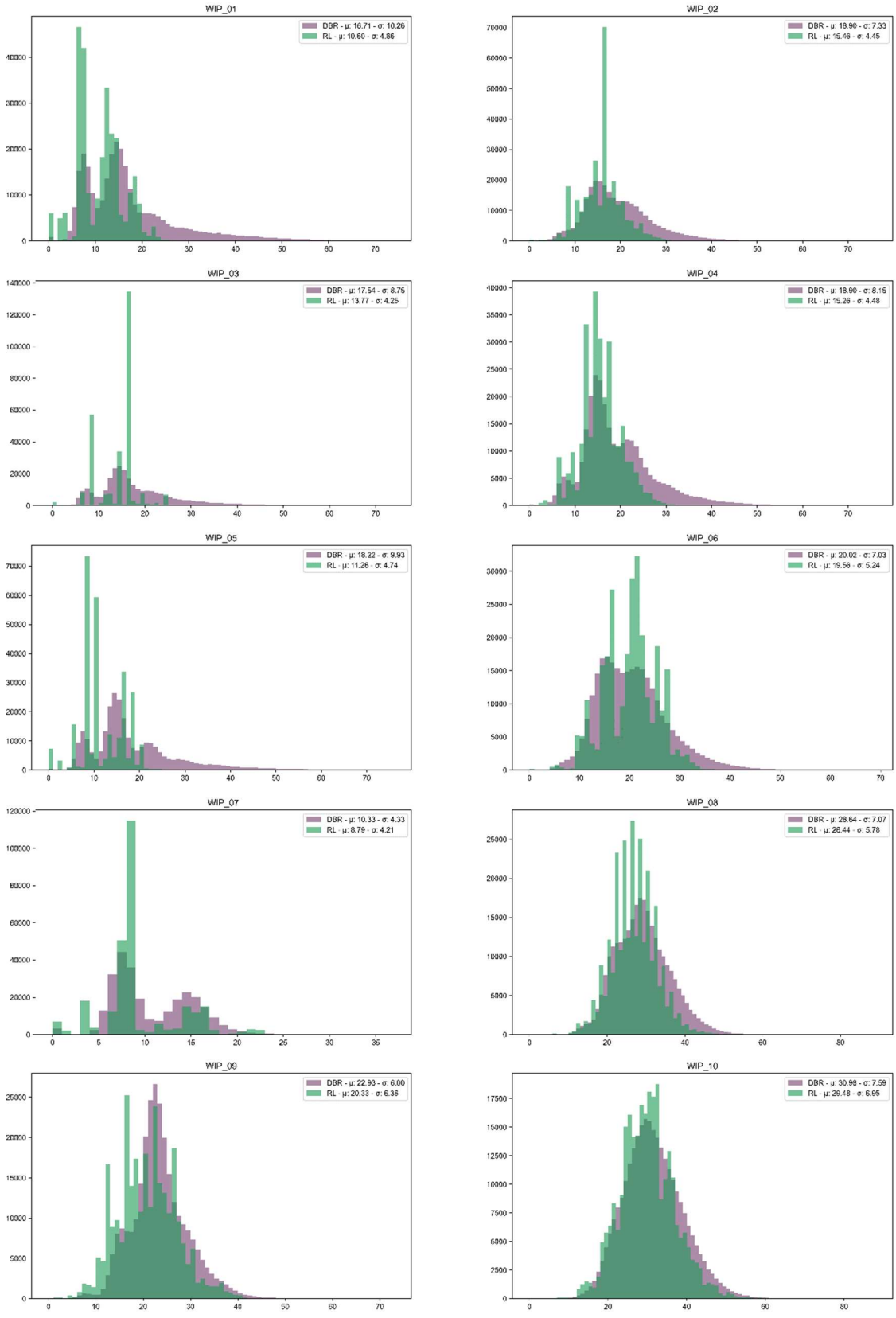
Fonte: Elaborado pelo auto

Figura 18 - Comparativo distribuição da variável CB



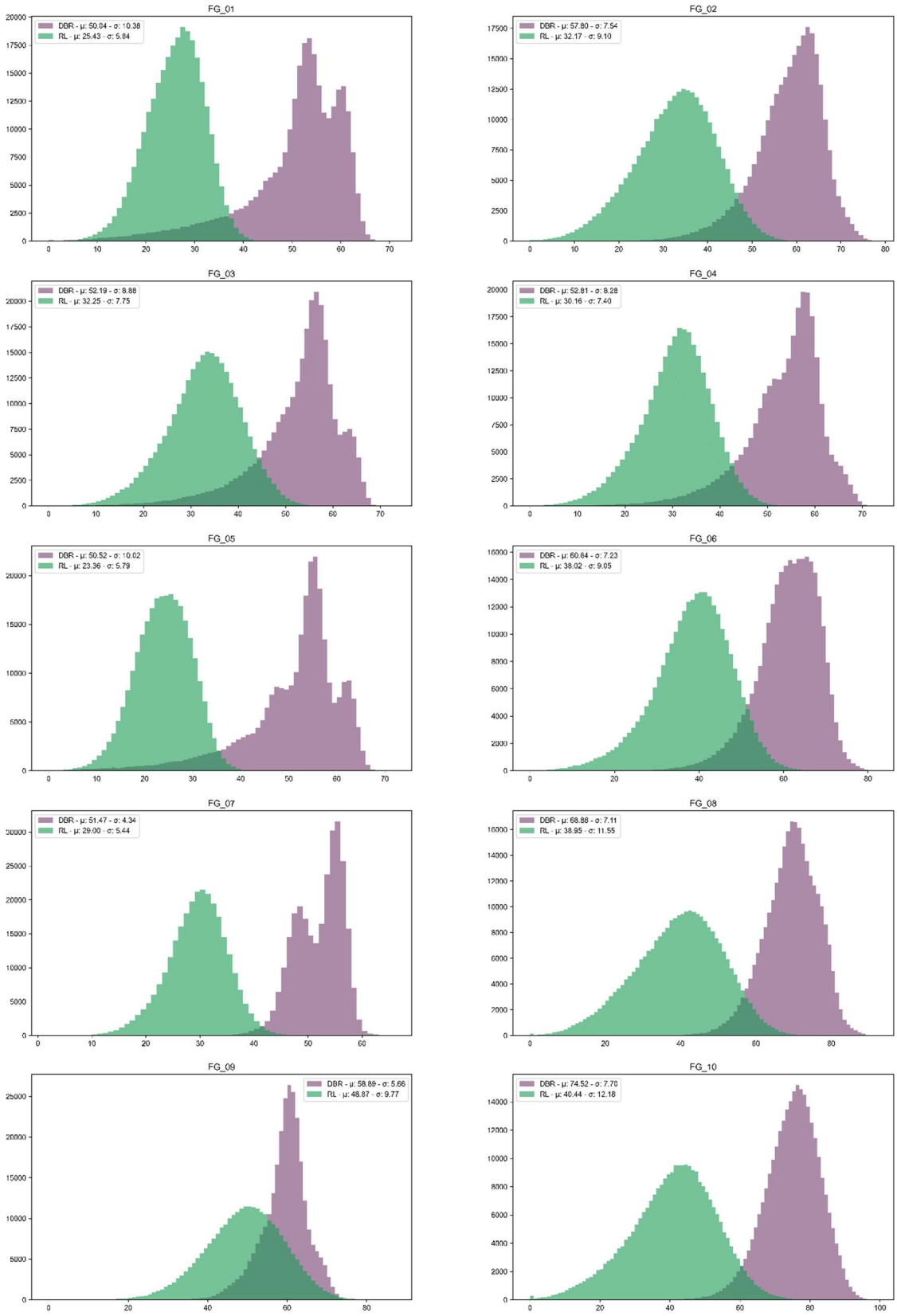
Fonte: Elaborado pelo autor

Figura 19 – Comparativo da distribuição da variável WIP



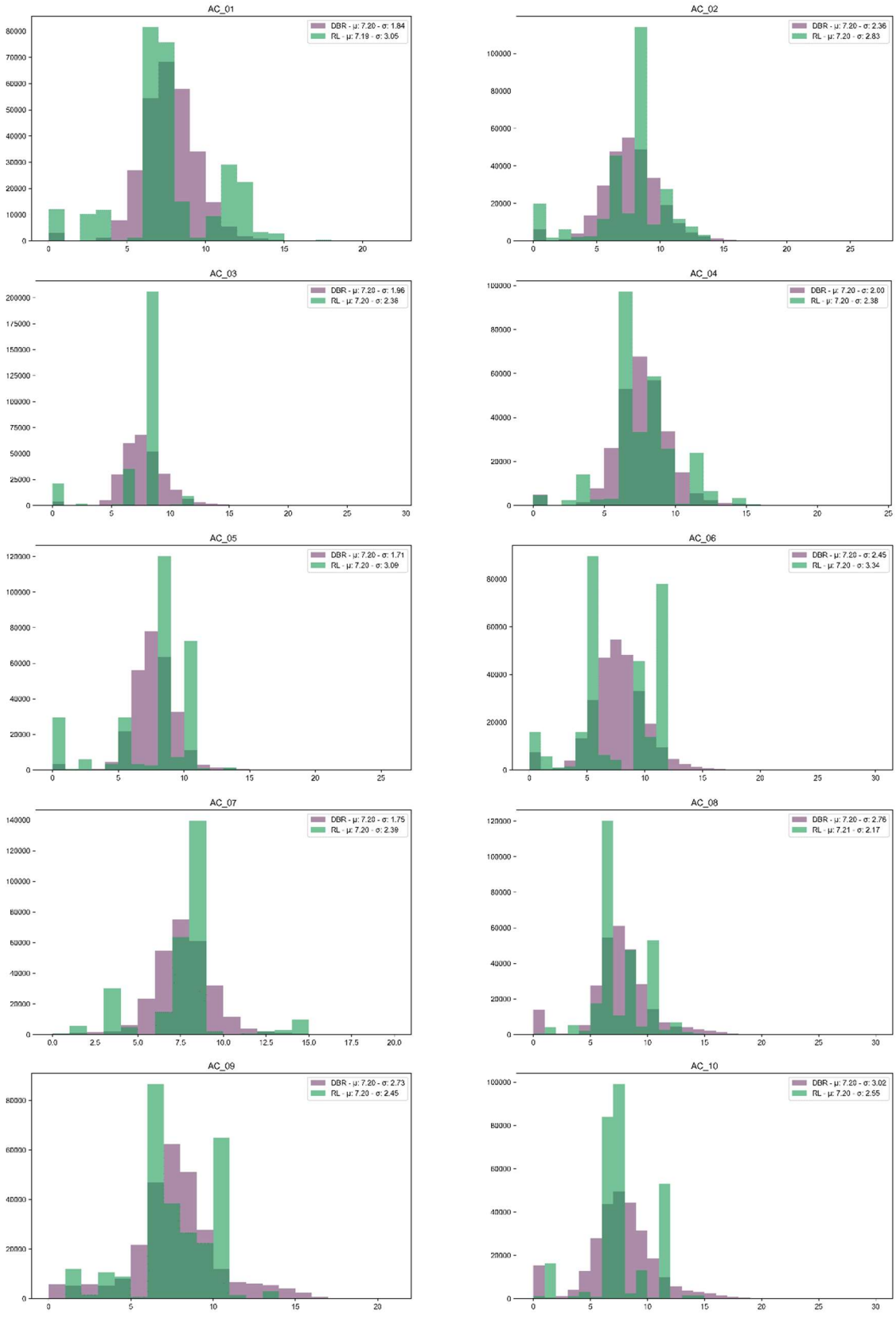
Fonte: Elaborado pelo autor

Figura 20 - Comparativo da distribuição da variável FG



Fonte: Elaborado pelo autor

Figura 21 - Comparativo da distribuição da variável AC



Fonte: Elaborado pelo autor