

O Problema Integrado de Produção e Roteamento com Flexibilidade nos Prazos de Entrega

A. A. N. MACHADO^{1*} e M. O. SANTOS²

Recebido em 3 de março de 2025 / Aceito em 25 de maio de 2026

RESUMO. Este artigo investiga o problema integrado de produção, estoque e roteamento com prazos de entrega flexíveis, ou seja, as entregas podem acontecer em um conjunto de períodos, porém sujeitas a penalizações. Para lidar com o problema, desenvolve-se um modelo de otimização linear inteira que flexibiliza os prazos de entrega dos itens demandados. Considera-se um ambiente produtivo com uma única fábrica, capaz de produzir e estocar uma quantidade finita de produtos distintos que compartilham recursos produtivos. Há um conjunto de clientes com estoques próprios e uma frota limitada de veículos heterogêneos. Cada veículo realiza no máximo uma rota por período. O objetivo do problema é determinar o plano de produção, estoque e roteamento que minimize os custos ao longo de um horizonte finito, atendendo às demandas dentro dos períodos possíveis (*flexible due dates*). Análises considerando um conjunto de instâncias artificiais de frota heterogênea, geradas aleatoriamente e baseadas nos dados da literatura, conforme especificado no início da Seção 3, mostram que explorar a flexibilidade nos prazos de entrega pode reduzir significativamente os custos, especialmente os de estoque. Além disso, avaliam-se diferentes cenários com variações na capacidade veicular e prazos de entrega.

Palavras-chave: problema integrado de produção, estoque e roteamento; *flexible due dates*; modelos de otimização inteira; análise de sensibilidade.

1 INTRODUÇÃO

No mundo moderno, organizações de diferentes setores, tanto públicas quanto privadas, buscam otimizar o desempenho de suas operações, especialmente na gestão de cadeias de suprimentos (SCM). Segundo Mostafa e Eltawil (2015) [16], SCM consiste em um conjunto de abordagens para integrar fornecedores, produtores, armazéns e lojas, garantindo que a mercadoria seja produzida e distribuída nas quantidades, locais e prazos adequados. O objetivo é minimizar o custo total do sistema, assegurando os níveis de serviço demandados.

*Autor correspondente: Alex Augusto Nunes Machado – E-mail: augusttoalex@gmail.com

¹Departamento de Matemática Aplicada e Estatística, ICMC, Universidade de São Paulo, Av. Trabalhador São-carlense, 400, 13566-590, São Carlos, SP, Brasil – E-mail: augusttoalex@gmail.com <https://orcid.org/0009-0008-5375-2253>

²Departamento de Matemática Aplicada e Estatística, ICMC, Universidade de São Paulo, Av. Trabalhador São-carlense, 400, 13566-590, São Carlos, SP, Brasil – E-mail: mari@icmc.usp.br <https://orcid.org/0000-0003-0822-3439>

Chandra e Fisher (1994) [5] destacam que a coordenação entre os níveis de uma organização gera ganhos de desempenho inacessíveis de outra forma. A integração pode ocorrer de duas formas: a coordenação geral, que unifica decisões de diferentes funções da cadeia produtiva, como localização de instalações, estoques, produção, distribuição e marketing; e a coordenação multi-fábrica, que ocorre entre escalões de uma mesma função. Os autores ilustram esta última com fábricas de produtos semiacabados que, combinados, formam o produto final. Neste artigo, considera-se a coordenação geral, integrando decisões de produção, estoque e roteamento.

Os primeiros registros da integração geral entre produção, estoque e roteamento aparecem nos artigos *Dynamic Distribution Model with Warehouse and Customer Replenishment Requirements*, de Pankaj Chandra (1993) [4], e *Coordination of Production and Distribution Planning*, de Pankaj Chandra e Marshall L. Fisher (1994) [5], conforme apontam [16], [8], [20], [13], [14] e [9].

Chandra e Fisher (1994) [5] buscam entender os impactos da coordenação, comparando-a com uma abordagem desacoplada, na qual os problemas de produção e estoque são otimizados separadamente, e o plano de roteamento é determinado com base na disponibilidade de estoque. Entre os resultados, concluem que: (1) o aumento do horizonte de planejamento e da quantidade de produtos e clientes reduziu os custos; (2) a abordagem integrada gerou maiores economias quando os custos de roteamento predominam sobre os de produção e estoque; (3) o problema integrado levou o dobro do tempo computacional em comparação com a estratégia não coordenada, considerando o ambiente *DecStation 2100* da época.

Recentemente, em [19] os pesquisadores estudaram um problema de produção, estoque e roteamento (PRP), no qual uma fábrica com capacidade limitada gerencia a produção de diversos itens, distribuição e controla os estoques de clientes ao longo de um horizonte de planejamento (*Vendor Managed Inventory* - VMI). O objetivo é minimizar os custos de preparação para a produção (*setup*), os custos de produção de vários itens, os custos de estoque e os custos de roteamento, considerando que, em cada período, é permitida apenas uma rota por veículo e que cada cliente recebe no máximo uma única visita. Os autores destacam a escassez de abordagens evolucionárias para o problema e propõem um novo algoritmo memético. Comparações entre esse algoritmo e uma abordagem híbrida com um *solver* comercial, utilizando a solução inicial do algoritmo memético, foram feitas em instâncias de 5, 10, 20 e 30 clientes com frotas homogêneas. A abordagem híbrida obteve melhorias médias nas soluções de 0,74% e 0,26% nas instâncias de 20 e 30 clientes, respectivamente, sugerindo seu potencial para futuras pesquisas.

Os trabalhos de [16], [20], [9], [1], [21], [15] e [17] destacam o caráter recente da pesquisa sobre integração das funções da cadeia produtiva, apontando a existência de espaço para novas investigações, especialmente na incorporação de flexibilização dos prazos de entrega (*flexible due dates*) em problemas integrados de produção, estoque e distribuição.

Considerando a possibilidade de atrasos, em [11], faz-se a distinção entre janelas de tempo (*time windows*) e prazos de entrega (*due dates*). Conforme destacam, problemas envolvendo janelas de tempo referem-se geralmente a um intervalo de tempo dentro do qual as entregas devem

ser realizadas em cada período do horizonte de planejamento, ou seja, cada período possui um horário inicial e final (*earliest* e *latest*) dentro do qual os clientes estarão disponíveis para o recebimento das mercadorias. Assim, em cada período do horizonte de planejamento, a janela de tempo deve iniciar e terminar dentro do período. Por exemplo, pode-se ter um horizonte de planejamento de sete dias em que as janelas de entregas de cada dia são das 8 h às 17 h.

Em contrapartida, outra possibilidade de flexibilização de tempo envolve um conjunto de períodos que se inicia a partir de um período inicial combinado para o recebimento do produto até um certo período máximo (*due date*) estipulado, que ainda é aceitável para a entrega deste. Assim, nesta abordagem, as *due dates* também são considerados períodos. Por exemplo, considerando um horizonte de planejamento de sete dias, se a demanda de um produto prevista para entrega no terceiro dia puder ser entregue após dois períodos, isso significa que ela pode ser entregue até o quinto dia, com a *due date* ocorrendo no quinto período.

Essa última proposta de flexibilização nos períodos de entrega introduzida por [3] também foi explorada em [12], [7], [6] e [10] e será a mesma sobre a qual o modelo desenvolvido neste artigo, e apresentado na Seção 2, se apoia. Em particular, em [3] um depósito (que pode ser uma fábrica) é responsável por atender um conjunto de clientes por meio de uma frota homogênea de veículos. Caso o horizonte de planejamento não seja respeitado, uma penalização é cobrada. Em cada período, os pedidos de cada cliente devem ser atendidos por no máximo um veículo e entregas antes das datas da entrega combinada do pedido não são permitidas. Neste caso, o objetivo é a minimização dos custos totais relacionados ao transporte, estoque e eventuais penalizações no problema de roteamento de veículos com prazos de entrega.

Na literatura também encontram-se os trabalhos de [13] e [14] que consideram a permissão de estender entregas por mais de um período. Deste modo, as pesquisas apresentadas em [13] e [14] consideram as janelas de tempo (*time windows*) dentro das quais as entregas devem ocorrer, podendo, no entanto, abranger mais de um período, desde que respeitem um prazo máximo pré-estabelecido (*due date*). Assim, embora essa janela de tempo se assemelhe à noção de janela de tempo apresentada em [11], o fato de [13] e [14] abordarem situações em que as rotas podem exigir mais de um período para serem concluídas pode fazer com que essa noção se “misture” com o conceito de prazo de entrega (*due date*) descrito em [11]. Isso ocorre porque pode haver situações em que as entregas sejam realizadas em períodos posteriores à sua data de lançamento, devido à impossibilidade de completar a rota em um único período.

Neste trabalho, a terminologia de prazos de entrega flexíveis (*flexible due dates*) será entendida no sentido introduzido por [3], ou seja, refere-se ao período máximo no qual poderá ocorrer a entrega a partir da data de lançamento do pedido.

Este artigo aborda um problema que integra decisões de produção, estoque e roteamento de produtos para os clientes. Considera-se um ambiente com uma única fábrica, que possui capacidade limitada de estocagem e produz uma quantidade finita de categorias de produtos utilizando recursos produtivos comuns e limitados. Os clientes, distribuídos geograficamente, também possuem estoques próprios. O transporte é realizado por uma frota heterogênea de veículos com capa-

cidade limitada em número de itens. Assim, embora o problema envolva diferentes categorias de produtos, todas consomem a capacidade veicular da mesma forma. Além disso, cada veículo pode realizar no máximo uma rota por período, que deve começar e terminar dentro desse mesmo período. O objetivo é determinar o plano de produção, estoque e roteamento que minimize os custos associados, incluindo penalizações por atrasos, ao longo de um horizonte finito de tempo, garantindo que as demandas dos clientes sejam atendidas dentro dos períodos permitidos (*flexible due dates*). Assim, a Tabela 1 destaca as diferenças deste trabalho em relação à literatura revisada. Em particular, destacamos como as principais contribuições deste artigo:

- Proposição de um modelo matemático para o problema com flexibilidade nos prazos de entrega.
- Análise do impacto dos atrasos nos planos de produção e distribuição, considerando vários cenários.

Tabela 1: Comparativo dos aspectos dos problemas considerados na literatura revisada.

	[3]	[6]	[7]	[10]	[11]	[12]	[13]	[14]	[18]	<i>Este trabalho</i>
Multiprodutos	✗	✗	✓	✓	✗	✗	✓	✓	✗	✓
Frota	Hom.	Hom.	⊖	⊖	Hom.	⊖	1 v	Het.	Hom.	Het.
Instâncias artificiais	✓	✓	✓	✓	✓	✗	✓	✓	✓	✓
Instâncias reais	✗	✗	✗	✗	✗	✓	✗	✓	✗	✗
Análise de sensibilidade	✓	✗	✓	✓	✓	✓	✗	✗	✗	✓
Considera os estoques dos clientes	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
Cientes atendidos por envios diretos	✗	✗	✓	✓	✗	✓	✗	✗	✗	✗
Cientes atendidos por roteamento	✓	✓	✗	✗	✓	✗	✓	✓	✓	✓
Permite entregas antecipadas	✗	✗	✗	✗	✗	✗	✗	✗	✓	✓
Considera <i>flexible due dates</i>	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓
Penalização por atraso	✓	✓	✗	✗	✓	✗	✗	✗	✗	✓

Fonte: Produzida pelos autores. ⊖: Não é explicitado no texto. 1 v: Apenas um veículo.

O restante deste documento está dividido da seguinte maneira. A Seção 2 apresenta o modelo matemático para o problema de produção, estoque e roteamento. Na Seção 3 é apresentado o ambiente computacional sobre o qual os testes foram realizados, bem como os resultados gerais obtidos por meio dos experimentos computacionais com auxílio do *solver* comercial Gurobi (v11.0.1). Análises de sensibilidade com relação à variação dos parâmetros de quantidade de períodos permitidos para o atraso e do parâmetro penalizador também são apresentadas. Por fim, análises de sensibilidade com relação à variação dos parâmetros de capacidade veicular e prazos de entrega para uma instância são realizadas na Seção 4. A Seção 5 finaliza este trabalho com as conclusões, bem como algumas possibilidades para pesquisas futuras.

2 MODELO MATEMÁTICO PARA O PROBLEMA DE PRODUÇÃO E ROTEAMENTO COM PERMISSÃO DE ATRASOS

O modelo proposto neste artigo foi baseado no trabalho de [19], entretanto, as principais fontes para a incorporação da permissão de atrasos foram [3], [11], [6] e [10]. Sendo assim, formalmente, sejam $V, N, P, T \in \mathbb{N}^*$ as quantidades de veículos, clientes, categorias de produtos (distintos) e horizonte de planejamento e $\mathcal{V} = \{1, \dots, V\}$, $\mathcal{N} = \{0, \dots, N\}$, $\mathcal{P} = \{1, \dots, P\}$ e $\mathcal{T} = \{1, \dots, T\}$ os conjuntos de veículos, locais, produtos e períodos, com $N = 0$ representando a fábrica. $A = \{(i, k) \in \mathcal{N}^2; i, k \in \mathcal{N} \text{ com } i \neq k\}$ o conjunto de arcos ligando os locais i e k .

Os demais parâmetros para o modelo matemático com flexibilização nas entregas são os mesmos definidos em [19] e estão resumidos na Tabela 2. Além disso, os parâmetros R_i e Π_{pi} são adicionados para indicar a quantidade de períodos permitidos atrasar no cliente i e a penalização por atraso na entrega do produto p do cliente i , respectivamente.

Em relação às variáveis, é adicionada uma variável q_{pit}^0 para contabilizar a quantidade do estoque inicial de p usado para atender o local i no período t . Além disso, a variável q_{pvi} de [19] é redefinida para $q_{pvi}^{t'}$, a qual representa a quantidade do produto p entregue pelo veículo v ao cliente i no período t' para satisfazer a demanda de t . O modelo é definido a seguir:

Tabela 2: Parâmetros e variáveis utilizados na modelagem.

Parâmetros:

V	Quantidade de veículos
N	Quantidade de clientes
P	Quantidade de produtos
T	Horizonte de planejamento
$\mathcal{A}$	$= \{(i, j) \in \mathcal{N}^2; i, j \in \mathcal{N} \text{ com } i \neq j\}$ o conjunto de arcos ligando os locais i e j
B	Capacidade de produção por período
b_p	Quantidade de tempo necessário para produzir cada unidade de p
c_p	Custo unitário de produção do item p
s_p	Custo de preparação do produto p
M	Número suficientemente grande
U_{pi}	Capacidade máxima de estoque para o item p no local i
I_{pi0}	Quantidade do produto p em estoque no local i no período $t = 0$
h_{pi}	Custo unitário de estoque do produto p no local i
C_v	Capacidade máxima de itens no veículo v
f_v	Custo fixo de uso do veículo v no período
a_{vik}	Custo de percorrer o arco (i, k) com o veículo v
d_{pit}	Quantidade do produto p demandada no local i no período t
R_i	Períodos de atraso permitidos no cliente i
Π_{pi}	Taxa de penalização por atraso do produto p no cliente i

Variáveis:

x_{pt}	Quantidade do produto p produzido no período t
y_{pt}	$\begin{cases} 1 & \text{se o produto } p \text{ é produzido no período } t \\ 0 & \text{caso contrário} \end{cases}$
I_{pit}	Quantidade do produto p em estoque no local i no fim do período t
z_{vikt}	$\begin{cases} 1 & \text{se o veículo } v \text{ percorre o arco } (i,k) \text{ no período } t \\ 0 & \text{caso contrário} \end{cases}$
r_{pvikt}	Quantidade do produto p transportado no arco (i,k) no período t com o veículo v
q_{pit}^0	Quantidade do estoque inicial de p usado para atender o local i no período t
$q_{pvi}^{t't}$	Quantidade do produto p entregue pelo veículo v ao cliente i no período t' para satisfazer a demanda de t .

Fonte: Produzida pelos autores.

$$\begin{aligned} \text{Min. } & \sum_{p=1}^P \sum_{t=1}^T (s_p y_{pt} + c_p x_{pt}) + \sum_{p=1}^P \sum_{i=0}^N \sum_{t=1}^T h_{pi} I_{pit} + \sum_{v=1}^V \sum_{i=1}^N \sum_{t=1}^T f_v z_{v0it} \\ & + \sum_{v=1}^V \sum_{i=0}^N \sum_{k=0}^N \sum_{t=1}^T a_{vik} z_{vikt} + \sum_{p=1}^P \sum_{v=1}^V \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=t}^{\min\{T, t+R_i\}} q_{pvi}^{t't} \cdot (t' - t) \cdot \Pi_{pi} \end{aligned} \quad (2.1)$$

s.a.

$$x_{pt} + I_{p0,t-1} - \sum_{v=1}^V \sum_{i=1}^N \sum_{t'=\max\{1, t-R_i\}}^T q_{pvi}^{t't} = I_{p0t} \quad p = 1, \dots, P, t = 1, \dots, T \quad (2.2)$$

$$\sum_{v=1}^V \sum_{u=1}^t \sum_{t'=t+1}^T q_{pvi}^{ut'} + I_{pit0} - \sum_{s=1}^t q_{pis}^0 = I_{pit} \quad p = 1, \dots, P, i = 1, \dots, N, \quad (2.3)$$

$$I_{pit} \leq U_{pi} \quad p = 1, \dots, P, i = 0, \dots, N, \quad t = 1, \dots, T \quad (2.4)$$

$$\sum_{p=1}^P b_p x_{pt} \leq B \quad t = 1, \dots, T \quad (2.5)$$

$$x_{pt} \leq M y_{pt} \quad p = 1, \dots, P, t = 1, \dots, T \quad (2.6)$$

$$\sum_{t=1}^T q_{pit}^0 \leq I_{pi0} \quad p = 1, \dots, P, i = 1, \dots, N \quad (2.7)$$

$$I_{pi0} - \sum_{s=1}^T q_{pis}^0 = I_{piT} \quad p = 1, \dots, P, i = 1, \dots, N \quad (2.8)$$

$$q_{pit}^0 + \sum_{v=1}^V \sum_{t'=1}^{\min\{T, t+R_i\}} q_{pvi}^{t't} = d_{pit} \quad p = 1, \dots, P, i = 1, \dots, N, \quad t = 1, \dots, T \quad (2.9)$$

$$\sum_{v=1}^V \min\{T, t+R_i\} \sum_{t'=1}^{t'} q_{pvi}^{t't} + I_{pi0} - \sum_{s=1}^{t-1} q_{pis}^0 \geq d_{pit} \quad p = 1, \dots, P, i = 1, \dots, N, \quad (2.10)$$

$$\sum_{v=1}^V \min\{T, 1+R_i\} \sum_{t'=1}^{t'} q_{pvi}^{t'1} + I_{pi0} \geq d_{pi1} \quad p = 1, \dots, P, i = 1, \dots, N \quad (2.11)$$

$$\sum_{p=1}^P \sum_{v=1}^V \sum_{k=1}^N \sum_{t=1}^T r_{pv0kt} + \sum_{p=1}^P \sum_{k=1}^N I_{pi0} \leq \sum_{p=1}^P \sum_{k=1}^N \sum_{t=1}^T d_{pkt} \quad (2.12)$$

$$\sum_{\substack{i=0 \\ i \neq k}}^N z_{vikt} - \sum_{\substack{l=0 \\ l \neq k}}^N z_{vkl} = 0 \quad v = 1, \dots, V, k = 0, \dots, N, \quad (2.13)$$

$$\sum_{\substack{i=0 \\ i \neq k}}^N r_{pvikt} - \sum_{\substack{l=0 \\ l \neq k}}^N r_{pvkl} = \sum_{t'=\max\{1, t-R_k\}}^T q_{pvk}^{t't} \quad p = 1, \dots, P, v = 1, \dots, V, \quad (2.14)$$

$$\sum_{v=1}^V \sum_{i=1}^N r_{pv0it} - \sum_{v=1}^V \sum_{i=1}^N r_{pvit} = \sum_{v=1}^V \sum_{i=1}^N \sum_{t'=\max\{1, t-R_i\}}^T q_{pvi}^{t't} \quad p = 1, \dots, P, t = 1, \dots, T \quad (2.15)$$

$$\sum_{p=1}^P r_{pvikt} \leq C_v z_{vikt} \quad v = 1, \dots, V, i, k = 0, \dots, N, \quad (2.16)$$

$$\sum_{k=1}^N z_{v0kt} \leq 1 \quad v = 1, \dots, V, t = 1, \dots, T \quad (2.17)$$

$$\sum_{v=1}^V \sum_{\substack{i=0 \\ i \neq k}}^N z_{vikt} \leq 1 \quad k = 1, \dots, N, t = 1, \dots, T \quad (2.18)$$

$$y_{pt}, z_{vikt} \in \{0, 1\} \quad p = 1, \dots, P, v = 1, \dots, V, \quad (2.19)$$

$$x_{pt}, I_{pit}, r_{pvikt} \in \mathbb{Z}_+ \quad t = 1, \dots, T, i, k = 0, \dots, N, \text{ com } i \neq k, \quad (2.20)$$

$$q_{pvi}^{t't} \in \mathbb{Z}_+ \quad p = 1, \dots, P, v = 1, \dots, V, \quad (2.21)$$

$$q_{pit}^0 \in \mathbb{Z}_+ \quad i = 1, \dots, N, \quad (2.22)$$

Nesta abordagem, a penalização adicionada à função objetivo consiste em:

$$\begin{aligned}
 & \sum_{p=1}^P \sum_{v=1}^V \sum_{i=1}^N \sum_{t=1}^T \sum_{t'=t}^{\min\{T, t+R_i\}} q_{pvi}^{t'} \cdot (t' - t) \cdot \Pi_{pi} \\
 &= \sum_{p=1}^P \sum_{v=1}^V \sum_{i=1}^N \sum_{t=1}^T \left(q_{pvi}^{t,t} (t - t) + q_{pvi}^{t+1,t} (t + 1 - t) + q_{pvi}^{t+2,t} (t + 2 - t) \right. \\
 & \quad \left. + \dots + q_{pvi}^{\min\{T, t+R_i\}, t} (\min\{T, t+R_i\} - t) \right) \Pi_{pi} \\
 &= \sum_{p=1}^P \sum_{v=1}^V \sum_{i=1}^N \sum_{t=1}^T \left(q_{pvi}^{t+1,t} + 2q_{pvi}^{t+2,t} + \dots + (\min\{T, t+R_i\} - t) q_{pvi}^{\min\{T, t+R_i\}, t} \right) \Pi_{pi}. \quad (2.23)
 \end{aligned}$$

Ou seja, para cada período de atraso, será cobrada uma penalização diretamente proporcional às quantidades de produtos e períodos *atrasados* o que difere de [3], [12], [7], [6] e [10]. Desta forma, a função objetivo (2.1) visa a minimização dos custos de preparação, produção, estoque, transporte e penalização por entregas atrasadas. Além disso, como T , t e $t + R_i$ são parâmetros, isto é, são conhecidos, temos que adicionar essa proposta de penalização mantém a função objetivo linear.

Nas restrições (2.2), a parcela $\sum_{v=1}^V \sum_{i=1}^N \sum_{t'=\max\{1, t-R_i\}}^T q_{pvi}^{t'}$ significa a quantidade total entregue a todos os clientes por todos os veículos no período t para atender os períodos $\max\{1, t - R_i\}$, $\max\{1, t - R_i + 1\}$, $\max\{1, t - R_i + 2\}$, $\dots$, T , ou seja, representa as quantidades entregues para atender os períodos em atraso, presente e eventuais antecipações. Logo, as restrições (2.2) asseguram o balanço de estoque na fábrica. A Figura 1 ilustra essa restrição, bem como apresenta um exemplo de $\sum_{t'=\max\{1, t-R_i\}}^T q_{pvi}^{t'}$ quando $R_i = 3$ e $t > 3$.

Por sua vez, nas restrições (2.3), o termo $\sum_{v=1}^V \sum_{t=1}^t \sum_{t'=t+1}^T q_{pvi}^{t'}$ indica o total entregue por todos os veículos nos períodos de 1 a t para atender os períodos $t + 1, t + 2, \dots, T$, ou seja, significa as quantidades entregues para atender os próximos períodos em relação a t ($t + 1, t + 2, \dots, T$). Já o termo $I_{pi0} - \sum_{s=1}^t q_{pis}^0$ fornece a quantidade do estoque inicial restante no fim do período t . Portanto, essas restrições (2.3) estabelecem o estoque dos clientes do período 1 a $T - 1$.

As restrições (2.4) garantem que as quantidades em estoque não ultrapassem a capacidade de estocagem. As restrições (2.5) limitam a capacidade para produção em cada período t . (2.6) garantem que será incorrido o custo de preparação caso haja produção. As desigualdades (2.7) asseguram que as quantidades utilizadas do estoque inicial não sejam maiores que as quantidades iniciais disponíveis, enquanto (2.8) garantem que os estoques finais sejam ou zero, ou o que sobrou dos estoques iniciais. As restrições (2.9), (2.10) e (2.11) especificam que as quantidades entregues e estocadas satisfaçam às demandas, sendo que (2.10) e (2.11) funcionam como desigualdades válidas, enquanto (2.12) garantem que as quantidades transportadas limitem-se as demandas. As restrições (2.13) garantem que cada veículo retorne à fábrica ao final da rota.

Nas restrições (2.14) e (2.15) o somatório $\sum_{t'=\max\{1, t-R_i\}}^T q_{pvi}^{t'}$ indica as quantidades de cada produto p entregues pelo veículo v para atender os períodos em atraso, presente e eventuais

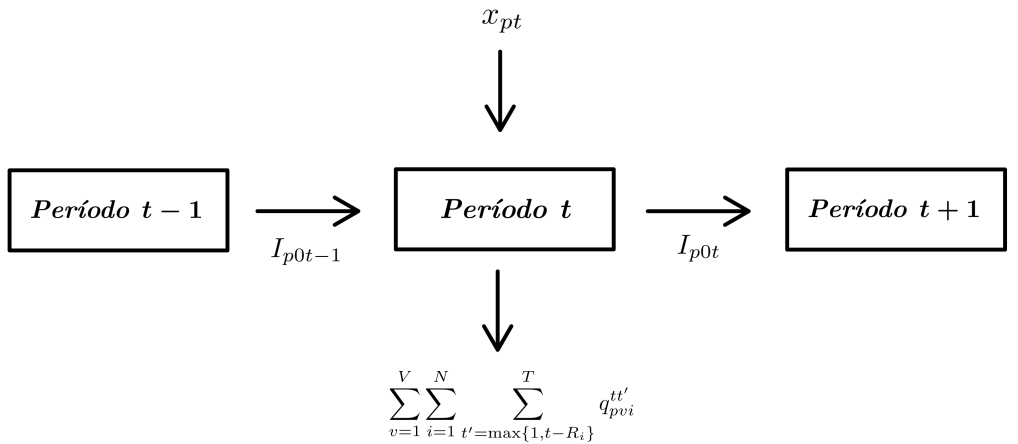


Ilustração de $\sum_{t'=\max\{1, t-R_i\}}^T q_{pvi}^{tt'}$ com $R_i = 3$:

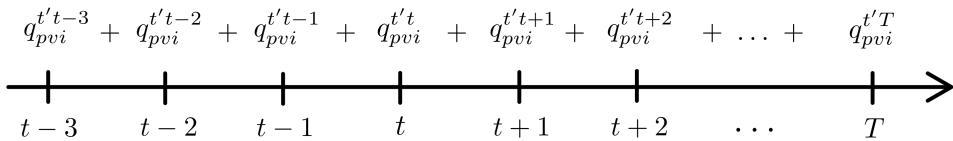


Figura 1: Ilustração da restrição (2.2) e exemplificação.

Fonte: Produzida pelos autores.

antecipações. Desta forma, estas duas restrições, uma vez que se tratam da extensão das restrições análogas apresentadas em [19], além de garantir a inexistência de subrotas, também asseguram a conservação do fluxo de produtos, isto é, especificam que apenas a demanda de certo local i é descarregada neste local. (2.16) determinam que as capacidades veiculares sejam respeitadas. As restrições (2.17) garantem que todo veículo partindo da fábrica vá para no máximo um local, e como não há subrotas, também asseguram que haverá no máximo uma rota. Enquanto (2.18) é quem impossibilita que cada cliente seja visitado por mais de um veículo em cada período. Por fim, (2.19) a (2.22) determinam os domínios das variáveis.

2.1 Quebras de simetrias em relação aos veículos

Conforme destacam [5], o problema integrado incorpora o dimensionamento de lotes com o problema de roteamento, dois outros problemas da literatura sabidamente NP-hard. Buscando auxiliar o método *branch-and-cut* do *solver* de otimização no desempenho, as restrições de quebras de simetria de rotas, (2.24) apresentadas em [2] foram adicionadas ao modelo. Neste contexto, rotas como $0-3-1-5-0$ e $0-5-1-3-0$ são simétricas, pois visitam os mesmos clientes, mas em ordem contrária.

$$z_{vi0t} \leq \sum_{j=1}^i z_{vj0t} \quad v = 1, \dots, V, \quad i = 1, \dots, N \quad t = 1, \dots, T \quad (2.24)$$

Neste caso, em cada período t tem-se que cada veículo v retorna à fábrica do cliente i apenas se este partiu da fábrica para algum cliente de índice no máximo até i . De forma equivalente, isso quer dizer que, caso o veículo v não tenha visitado um local de índice até i então ele não pode retornar à fábrica do local i . Essa é uma forma de quebra de simetria parcial de rotas que, frente às rotas simétricas, dá preferência pela que inicia no cliente de menor índice [2].

3 EXPERIMENTOS COMPUTACIONAIS

O modelo apresentado na Seção 2 é uma extensão da modelagem apresentada em [19], com o intuito de permitir atrasos nas entregas dos produtos aos clientes, porém com uma penalização pelo atraso. Ambos os modelos foram implementados e seus resultados analisados considerando a situação em que não se permite atraso, uma vez que os modelos são equivalentes.

A implementação computacional de ambos os modelos foi realizada com o auxílio da linguagem de modelagem algébrica JuMP (v1.22.1), desenvolvida especificamente para otimização matemática e incorporada à linguagem de programação Julia (v1.7.3), uma linguagem de alto nível, porém projetada para a alta performance desejada pela pesquisa científica. Tanto o JuMP quanto Julia são *open-source* e gratuitos. O *branch-and-cut* do *solver* comercial Gurobi (v11.0.1) foi o solucionador utilizado sendo que o único parâmetro modificado foi o MIPGap, que foi definido com precisão de $\text{MIPGap} = 10^{-8}$ (o *default* é 10^{-4}), o qual foi calculado da seguinte forma: $\text{GAP} = \frac{|UB - LB|}{|UB|}$ em que UB e LB referem-se aos limitantes primal e dual, respectivamente.

Os *hardwares* utilizados consistiram em processadores Intel Xeon E5-2680v2 de 2.8 GHz com 20 núcleos físicos e 40 núcleos considerando *Hyper-threading*. A memória RAM disponível totaliza 128 GB de memória DDR3 operando a 1866 MHz, porém, os *batches* têm 32 Gb de memória RAM.

As instâncias utilizadas foram geradas artificialmente a partir das especificações descritas no artigo [19]. A Tabela 3 descreve os domínios utilizados no gerador para a determinação dos parâmetros das instâncias.

Dois grupos de instâncias de frota heterogênea foram utilizados. A Tabela 4 descreve as configurações de tamanho do horizonte de planejamento, quantidades de produtos, quantidades de demandas, entre outros, de cada grupo testado.

Os experimentos computacionais foram realizados considerando os três seguintes testes:

1. Testes de desempenho;
2. Testes para investigação geral;
3. Testes para investigação particular.

Tabela 3: Parâmetros para geração das instâncias.

<i>Parâmetros presentes ambos os modelos</i>	
Demanda dos clientes	$d_{pit} \in [10, 100]$
Capacidade máxima de estoque nos clientes	$U_{pi} = \lfloor (\sum_t d_{pit} / T) \cdot \lambda \rfloor$ com $\lambda \in \{6, 9, 12, 15\}$
Estoque inicial na fábrica	$I_{p00} = 0$
Estoque inicial nos clientes	$I_{pit0} = \sum_{t=1}^{\tau} d_{pit}$, com $\tau \in \{1, 2, 3, 4\}$
Custo unitário de estocagem na fábrica	$h_{p0} \in \{3, 8\}$
Custo unitário de estocagem nos clientes	$h_{pi} \in \{n; n = 1, 2, \dots, 10\}$
Capacidade de produção por período	$B = \lfloor 3 \cdot (\sum_p \sum_i \sum_t d_{pit} / T) \rfloor$
Gasto de tempo para produção de cada item	$b_p = 1$
Custo de produção	$c_p = 10h_{p0}$
Custo de <i>setup</i>	$s_p = 100h_{p0}$
Coordenadas	$x_k \in [0, 500]$ e $y_k \in [0, 1000]$
Capacidade veicular - frota homogênea (quantidade de produtos)	$C = P \cdot \max\{U_{pi}; p \in \mathcal{P} \text{ e } i \in \mathcal{N}\}$
Capacidade veicular - frota heterogênea (quantidade de produtos)	$C_v = \left\lfloor C \cdot \frac{1}{v} \right\rfloor$ para $v = 1, \dots, V$
Custo de transporte - frota homogênea	$a_{ik} = \left\lfloor \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} + 0,5 \right\rfloor$
Custo de transporte frota heterogênea	$a_{ikv} = a_{ik} \cdot \frac{1}{v}$ para $v = 1, \dots, V$
Custo fixo de transporte	$f = (N + 1) \cdot \max_{\forall i,k} \{a_{ik}\}$
<i>Big M</i>	$M = \sum_p \sum_i \sum_t d_{pit}$
<i>Parâmetros presentes apenas no modelo com flexibilização</i>	
Períodos de atraso permitidos	$R_i \in \{0, 1, 3, 5\}$
Taxa de penalização por atraso	$\Pi_{pi} \in \{1,00 \cdot h_{pi}; 1,25 \cdot h_{pi}; 1,5 \cdot h_{pi}; 1,75 \cdot h_{pi}; 2,00 \cdot h_{pi}\}$

Fonte: Adaptado de [19]. Observação: As instâncias geradas consideraram a capacidade de estoque da fábrica como ilimitada, assim como feito em [19].

O intuito da primeira rodada de testes, os testes de desempenho, foi fazer o comparativo de performance do modelo estendido, não considerando a flexibilização dos prazos de entrega, com o modelo base descrito em [19]. No modelo estendido, é possível especificar quantos períodos de atraso são permitidos por meio do parâmetro R_i . Assim, ao fixar $R_i = 0$ para todos os clientes i , o valor funcional ótimo retornado para cada instância deve ser o mesmo que o resultado obtido com o modelo apresentado em [19]. Logo, a comparação de desempenho faz sentido nesse caso.

Tabela 4: Configurações de cada Grupo de instâncias.

	Grupo 1	Grupo 2
Quantidade de períodos	6	6
Quantidade de produtos	3	5
Quantidade de clientes	5	10
Quantidade de veículos	2	2
Quantidade de sementes	5	5
Demandas	Variável	Variável
Capacidade de produção	Constante	Constante
Capacidade de estoque na fábrica	Limitado	Limitado
Capacidade de estoque nos clientes	Constante	Constante
Estoque iniciais na fábrica	0	0
Estoque iniciais nos clientes	Variável	Variável
Capacidade dos veículos	Constante	Constante
Quantidade total de instâncias	20	19

Fonte: Retirado de [19].

A Seção 3.1 resume os resultados obtidos para essa rodada. Em ambos os casos, a quebra parcial de simetria de rotas (2.24) foi adicionada.

A segunda rodada de testes tem por objetivo a realização de análises agregadas gerais sobre como os custos de cada componente da função objetivo variaram a depender dos parâmetros especificados R_i (quantidade de períodos permitidos atrasar) e Π_{pi} (custo penalizador por atraso). A síntese dos resultados e análises obtidos são expostos na Seção 3.2.

Finalmente, uma análise de sensibilidade para uma instância é realizada na Seção 4. Cenários com diferentes configurações dos parâmetros R_i (permissão de atraso no cliente i) e C_v (capacidade do veículo v) são testados e alguns *insights* são apresentados.

3.1 Análises de desempenho

Ambos os modelos foram primeiro comparados em termos de suas relaxações lineares, sendo que, nas duas situações, o *solver* conseguiu obter as soluções ótimas dentro do tempo especificado de uma hora. Além disso, a Tabela 5 apresenta o número de variáveis geradas para os dois grupos de instâncias em ambos os modelos. Como era esperado, o número de variáveis no modelo proposto é superior ao do modelo clássico para o problema.

Em relação ao tempo de execução, tendências opostas foram identificadas. Mais precisamente, de modo geral, para as 20 instâncias do Grupo 1, um aumento médio no tempo de execução de 83,75% foi identificado, enquanto uma redução média de 44,56% foi observada para as 19 instâncias do segundo grupo. Isto posto, os resultados a seguir consideram o modelo sem a

Tabela 5: Número de variáveis do modelo relaxado de [19] e do modelo estendido.

	Número de variáveis	
	Grupo 1	Grupo 2
Modelo clássico de Rodrigues et al. [19]	972	8965
Modelo proposto	1512	12265
Taxa de aumento (%)	55,56	36,81

Fonte: Produzida pelos autores.

relaxação linear e mostram resultados relacionados ao número de nós, o tempo de execução, o GAP, os limitantes primais (UB) e duais (LB) para os dois modelos.

Ao considerar as instâncias do Grupo 1, o *branch-and-cut* do *solver* foi capaz de provar a otimalidade para ambos os modelos, como pode ser constatado na Tabela 6. Em média, o modelo sem considerar a flexibilização [19] levou 12,20 segundos para resolver as 20 instâncias do Grupo 1. Em contrapartida, ao considerar o modelo estendido, o *branch-and-cut* do *solver* precisou de 23,79 segundos, em média, sendo que em apenas uma instância (G1-I19) houve a redução no tempo para esse primeiro grupo. Para o Grupo 2, o *branch-and-cut* do *solver* estabeleceu a otimalidade em 17 instâncias para o modelo clássico, contra 13 para o modelo estendido com $R_i = 0$.

Em relação à quantidade de nós explorados, em média, o *solver*, para o modelo sem a flexibilização, explorou mais nós do que para o modelo estendido para o Grupo 1; o comportamento contrário ocorreu para o Grupo 2. Considerando apenas os casos em que o modelo proposto explorou mais nós, é possível notar que, para 9 instâncias do Grupo 1, houve um aumento médio de 92,22% e uma redução média de 49,80% para as 11 instâncias restantes. Em contrapartida, houve um aumento médio de 66,53% para 11 instâncias do Grupo 2 e uma redução média de 38,51% nas outras oito instâncias desse grupo.

No que diz respeito ao tempo de execução, houve um aumento médio de 123,64% para 19 instâncias do Grupo 1 e uma redução de 18,66% na instância restante, ao observar os resultados obtidos para o modelo estendido. Já no Grupo 2, desconsiderando também as seis instâncias em que a otimalidade não foi alcançada, nas 13 instâncias restantes, houve um aumento médio de 207,77% no tempo de execução na versão estendida.

Os resultados evidenciam que o modelo estendido, com a especificação do parâmetro de atraso igual a zero, em geral, levou a maiores tempos de execução para as instâncias em que a otimalidade foi atingida, bem como a uma redução na quantidade de soluções ótimas alcançadas dentro do tempo limite de uma hora: cerca de 97,87% das instâncias com o modelo descrito em [19], contra 87,18% das instâncias testadas com o modelo estendido. Tais resultados são coerentes com os aumentos nas quantidades de variáveis, de 972 e 8965 no modelo base para 1512 e 12265 no modelo estendido, para os grupos de instâncias 1 e 2 testados, conforme apresentado na Tabela 5.

Tabela 6: Modelo clássico [19] × Modelo modelo estendido.

Instância	Soluções dos modelos									
	Modelo descrito em [19]					Modelo modelo estendido				
	UB	LB	GAP (%)	Nº nós	Tempo (s)	UB	LB	GAP (%)	Nº nós	Tempo (s)
G1-I01	116866	116866	0	14174	5.30	116866	116866	0	8632	17.70
G1-I02	816976	816976	0	11361	4.15	816976	816976	0	6227	13.51
G1-I03	126387	126387	0	6267	3.35	126387	126387	0	1586	9.96
G1-I04	86603	86603	0	612	2.24	86603	86603	0	2347	3.52
G1-I05	159787	159787	0	3178	11.69	159787	159787	0	4453	30.11
G1-I06	1330057	1330057	0	4018	18.30	1330057	1330057	0	7221	20.27
G1-I07	165838	165838	0	6606	5.06	165838	165838	0	1585	8.92
G1-I08	138281	138281	0	4386	3.63	138281	138281	0	3699	5.84
G1-I09	299008	299008	0	3563	27.45	299008	299008	0	6586	38.27
G1-I10	2630368	2630368	0	11790	15.95	2630368	2630368	0	4148	18.32
G1-I11	310912	310912	0	32886	31.01	310912	310912	0	30312	68.64
G1-I12	273968	273968	0	20551	9.21	273968	273968	0	25543	33.38
G1-I13	227891.50	227891.50	0	5969	13.62	227891.50	227891.50	0	4321	18.69
G1-I14	1954901.50	1954901.50	0	3196	17.71	1954901.50	1954901.50	0	8240	26.76
G1-I15	238981	238981	0	10760	20.32	238981	238981	0	17198	46.29
G1-I16	206751	206751	0	23086	13.79	206751	206751	0	43038	54.65
G1-I17	169949.50	169949.50	0	5200	3.42	169949.50	169949.50	0	812	10.98
G1-I18	1413299.50	1413299.50	0	7342	6.06	1413299.50	1413299.50	0	168	8.40
G1-I19	179975.50	179975.50	0	6682	20.31	179975.50	179975.50	0	5697	16.52
G1-I20	150576.50	150576.50	0	17607	11.43	150576.50	150576.50	0	19973	25.10
Médias	-	-	0	9961.70	12.20	-	-	0	10089.30	23.79
G2-I01	681198	681198	0	4415	221.38	681198	681198	0	3469	996.83
G2-I02	5904528	5904528	0	3303	233.39	5904528	5904528	0	4233	845.38
G2-I03	700782	700782	0	4400	232.61	700782	700782	0	5279	537.38
G2-I04	608462	608462	0	67818	746.71	608462	608224.94	0.04	199081	3600.00
G2-I05	671890	671890	0	4718	249.31	671890	671890	0	7440	744.38
G2-I06	5883610	5883610	0	5349	365.17	5883610	5883610	0	9927	1113.90
G2-I07	692970	692970	0	47914	922.85	692970	689489.17	0.50	82792	3600.00
G2-I08	615650.50	606729.77	1.45	292636	3600.00	615654	605304.68	1.68	77780	3600.00
G2-I09	613001	613001	0	9854	295.46	613001	613001	0	4953	1248.48
G2-I10	5189231	5189231	0	4084	350.42	5189231	5189231	0	5615	966.64
G2-I11	634880.50	634880.50	0	4922	281.31	634880.50	634880.50	0	4406	527.96
G2-I12	540250	540250	0	121717	792.61	540250.00	539912.37	0.06	373722	3600.00
G2-I13	616398.50	616398.50	0	6199	319.45	616398.50	616398.50	0	4303	1266.83
G2-I14	5260848.50	5260848.5	0	3750	389.73	5260848.50	5260848.50	0	3921	1170.10
G2-I15	637091.50	637091.50	0	5907	285.11	637091.50	637091.50	0	7056	594.78
G2-I16	545740	545740	0	82355	1265.46	545740	544373.89	0.25	86943	3600.00
G2-I17	553486	553486	0	4393	311.93	553486	553486	0	3260	848.89
G2-I18	4722466	4722466	0	4205	258.52	4722466	4722466	0	3720	750.18
G2-I19	492667.50	486383.54	1.28	351308	3600.00	492739.50	485271.57	1.52	52430	3600.00
Médias	-	-	0.14	54170.90	774.82	-	-	0.21	49491.05	1748.02

Fonte: Produzida pelos autores. Nota: Tempo limite de 3600s.

Os resultados apresentados para as instâncias em que ambos os modelos obtiveram a otimalidade mostram que seus valores funcionais ótimos coincidiram, evidenciando que o modelo estendido consiste em uma extensão do modelo que não considera a permissão de entregas atrasadas, conforme descrito em [19]. As próximas análises apresentadas na Seção 3.2 investigam os impactos nas soluções quando $R_i > 0$.

3.2 Considerações gerais acerca dos custos

Buscando compreender os impactos que a permissão de entregas em atraso causa nos custos das soluções obtidas, esta Seção 3.2 é destinada à apresentação de algumas análises de como os custos são modificados com a variação dos parâmetros R_i e Π_{pi} . É importante destacar que, para cada valor do parâmetro R_i utilizado, este foi o mesmo para todos os clientes. Além disso, as análises aqui apresentadas se restringem ao segundo grupo de instâncias. Tal escolha se deve apenas por esse ser o grupo com as maiores instâncias testadas em termos de quantidades de produtos, clientes e veículos. Em particular, 17 das 19 instâncias obtiveram GAPs de até 0,89% para todas as combinações dos parâmetros R_i e Π_{pi} , sendo que, de forma geral, a consideração de todas as 19 instâncias não gerou impactos significativos nos gráficos gerados. Portanto, as análises a seguir consideram os resultados obtidos para as 19 instâncias deste segundo grupo. Observa-se que não foram utilizadas as 20 instâncias nesse grupo, pois os resultados de uma instância, G2-I19, foram descartados devido à incongruência com os resultados obtidos pelo *solver* devido a erros numéricos detectados durante os testes computacionais.

3.2.1 Variações nos custos relativos ao Grupo 2 de instâncias

A expressão (3.1) apresenta o cálculo utilizado para cada combinação de parâmetros Π_{pi} e R_i nas Figuras 2a, 2b, 3a e 3b, onde NI é o número de instâncias. Essas figuras apresentam as taxas de variações médias no valor da função objetivo, nos custos de *setup* mais (+) os custos de produção, nos custos de estoque e de transporte conforme variam os parâmetros Π_{pi} e R_i .

Note que a razão $Metrica_{R_i, \Pi_{pi}, j} / Metrica_{0, \Pi_{pi}, j}$ significa o valor obtido para a métrica avaliada ao permitir alguma flexibilização na entrega, dividido pelo valor obtido na mesma métrica, porém sem a permissão de atraso. Por exemplo, $estoque_{3, 1.75 \times h_{pi}, 7} / estoque_{0, 1.75 \times h_{pi}, 7}$ compara o quanto os custos de $estoque_{3, 1.75 \times h_{pi}, 7}$ da versão com até 3 períodos de atraso permitidos representam do valor obtido sem atraso ($R_i = 0$). Assim, valores positivos de (3.1) indicam aumento (percentual) do valor que está sendo analisado para o caso $R_i > 0$ em relação ao caso $R_i = 0$, ao passo que valores negativos indicam redução (percentual) nessa métrica para o caso R_i em relação ao caso $R_i = 0$.

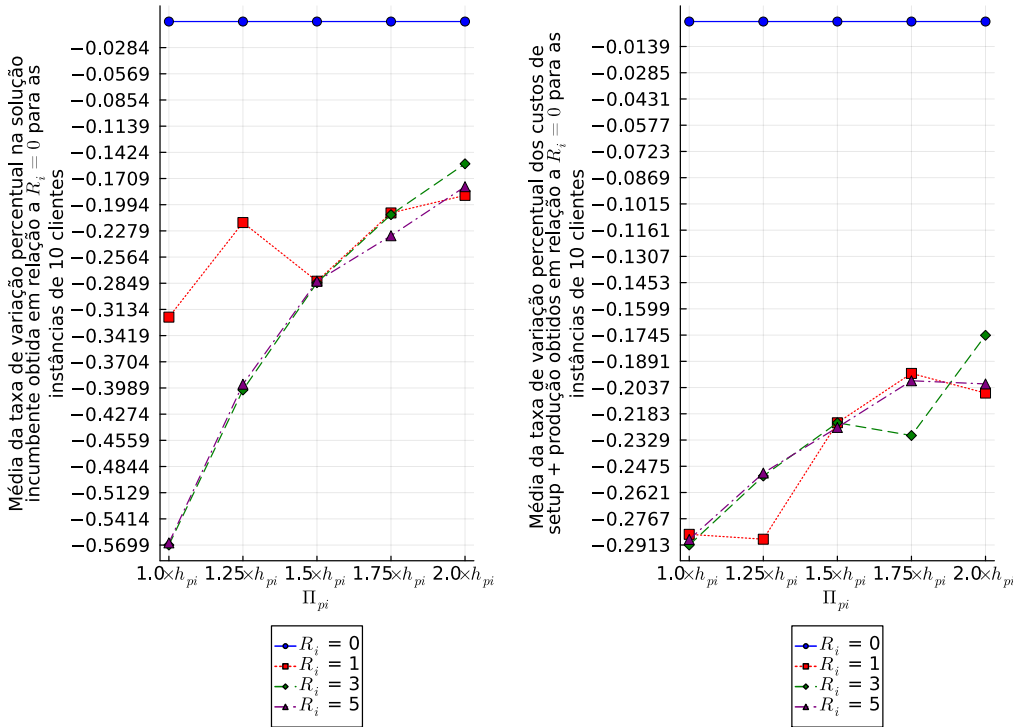
$$\frac{1}{NI} \sum_{j=1}^{NI} \left(\frac{Metrica_{R_i, \Pi_{pi}, j}}{Metrica_{0, \Pi_{pi}, j}} - 1 \right) \cdot 100 \quad (3.1)$$

com

- $Metrica_{R_i, \Pi_{pi}}$. Este valor pode assumir, dependendo da situação a ser analisada, um dos seguintes valores: valor funcional da solução ótima com a permissão de atraso R_i e fator penalizador Π_{pi} ; custos de *setup* mais custos de produção com a permissão de atraso R_i e fator penalizador Π_{pi} ; custos de estoque com a permissão de atraso R_i e fator penalizador Π_{pi} ; custos de transporte com a permissão de atraso R_i e fator penalizador Π_{pi} .
- $Metrica_{0, \Pi_{pi}}$. Este valor pode assumir um dos seguintes valores: valor funcional da solução ótima sem a permissão de atraso com fator penalizador Π_{pi} ; custos de *setup* mais o custo

da produção sem a permissão de atraso com fator penalizador Π_{pi} ; custos de estoque sem a permissão de atraso com fator penalizador Π_{pi} ; custos de transporte sem a permissão de atraso com fator penalizador Π_{pi} .

Quando o parâmetro Π_{pi} reduz, as Figuras 2a, 2b e 3b mostram uma tendência geral de redução nos custos, em relação a $R_i = 0$, nos custos totais, nos custos da componente de *setup* mais os custos de produção e da componente de transporte. As maiores reduções identificadas, respectivamente, foram de 0,5699 %, 0,2913 % e 16,1951 % ao permitir até três períodos de atraso. Em particular, uma tendência de redução na participação da componente de transporte no custo total ocorre com o aumento do parâmetro Π_{pi} até $1,50 \times h_{pi}$, com um comportamento não monótono a partir desse ponto. No que diz respeito à componente de estoque, com a Figura 3a pode-se notar que para algumas combinações de parâmetros, há aumentos nos custos de estocagem quando comparados aos valores obtidos para $R_i = 0$, podendo chegar a até 8,9650 % quando $\Pi_{pi} = 1,25 \times h_{pi}$ e $R_i = 1$.

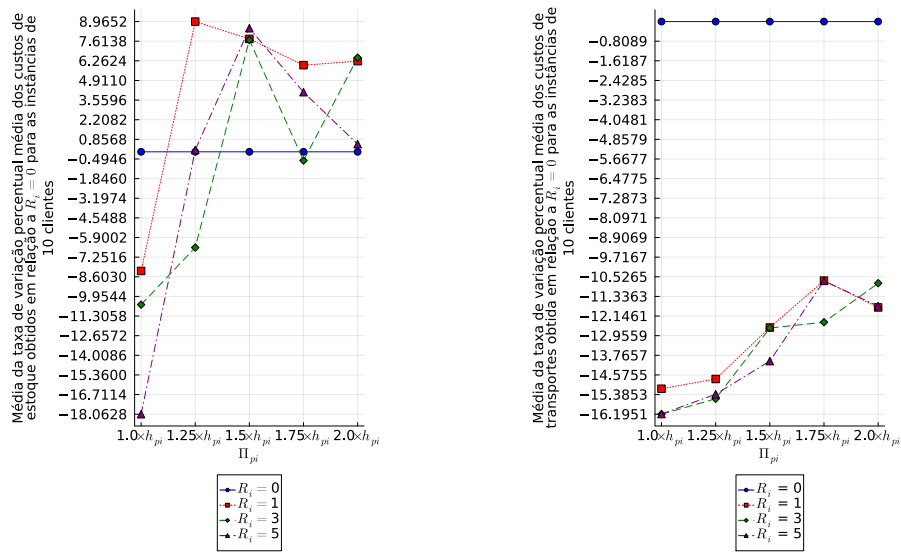


(a) Variação percentual na solução ótima.

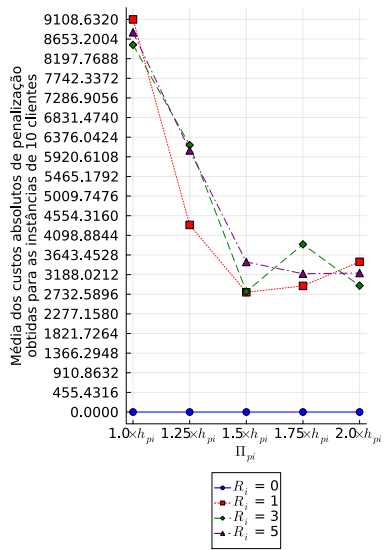
(b) Variação nos custos de *setup* + produção.

Figura 2: Média da taxa de variação percentual do valor obtido para R_i de cada métrica em relação ao valor obtido com $R_i = 0$ para o Grupo 2 de instâncias.

Fonte: Produzida pelos autores.



(a) Variação dos custos de estoque. (b) Variação dos custos de transportes.



(c) Média dos custos absolutos de penalização obtidos para as instâncias de 10 clientes.

Figura 3: Média da taxa de variação percentual do valor obtido para R_i de cada métrica em relação ao valor obtido com $R_i = 0$ para o Grupo 2 de instâncias.

Fonte: Produzida pelos autores.

Por fim, com a Figura 3c é possível notar uma clara tendência de redução na participação da penalidade nos custos totais ao aumentar o parâmetro Π_{pi} até $1,50 \times h_{pi}$. Porém, a partir deste valor, um comportamento não monótono ocorre.

A Tabela 7 apresenta a maior redução da função objetivo registrada e especifica, qual foi a combinação dos parâmetros Π_{pi} e R_i que levou a tais resultados. Também resume as maiores reduções obtidas para os componentes custos de *setup* e custos de produção, custos de transporte e de estoque da função objetivo.

Tabela 7: Resumo das reduções médias obtidos para cada componente da função objetivo ao permitir atrasos $R_i > 0$ em relação a $R_i = 0$ - Grupo 2 de instâncias.

Componente	Combinação de parâmetros	Redução (%)
Custos de Setup mais custos de Produção	$R_i = 3$ e $\Pi_{pi} = 1,0 \times h_{pi}$	0,2913
Custos de Transportes	$R_i = 3,5$ e $\Pi_{pi} = 1,0 \times h_{pi}$	16,1951
Custos de Estoque	$R_i = 5$ e $\Pi_{pi} = 1,0 \times h_{pi}$	18,0628
Custo total	$R_i = 3$ e $\Pi_{pi} = 1,0 \times h_{pi}$	0,5699

Fonte: Produzida pelos autores.

4 ANÁLISES DE SENSIBILIDADE PARTICULAR: IMPACTOS DA VARIAÇÃO DOS PARÂMETROS R_i E C_v

Nesta seção, será analisado como a flexibilidade de entrega e a capacidade veicular impactam os planos de produção e roteirização para uma instância, no caso, a G2-I01, escolhida de forma aleatória. 28 cenários foram obtidos a partir da variação do parâmetro de permissão de atraso (R_i) e capacidade veicular (C_v).

As variações das capacidades veiculares, C_v , foram especificadas de forma que $C_v \in \{0,50 \cdot C_v; 0,75 \cdot C_v; 1,00 \cdot C_v; 1,25 \cdot C_v; 1,50 \cdot C_v; 1,75 \cdot C_v; 2,00 \cdot C_v\}$. Para todas as variações, o *branch-and-cut* do *solver* provou otimalidade dentro do limite de tempo de 1 hora.

A Tabela 8 compila um resumo dos resultados de GAP, custo da solução incumbente, limitante inferior, quantidade de nós e tempo. As instâncias geradas pelas variações foram nomeadas por $V_{CapVeiculo}R_{atrasos}$, por exemplo, a instância representada por $V_{0,5C}R_0$ tem capacidade do veículo $0,50 \cdot C_v$ e $R_i = 0$, ou seja, não tem permissão de atraso, porém, a capacidade de veículos possui é metade do valor inicialmente gerado.

Os resultados na Tabela 8 mostram que, para cada capacidade veicular, ao variar a permissão de atraso, $R_i > 0$, os valores funcionais foram os mesmos. Ou seja, a permissão de até três ou de até cinco períodos de atraso leva a reduções nos custos iguais ao permitir apenas um período para entregas em atraso. Assim, apenas os resultados considerando $R_i = 0$ e $R_i = 1$ são apresentados. Em contrapartida, ao aumentar as capacidades veiculares, economias significativas podem ser obtidas, chegando a redução de até 1,39 % quando se olha o caso com $V_{0,5C}R_1$ e $V_{2,0C}R_1$. Embora,

nota-se que, para veículos com capacidades 1,25C, 1,50C e 1,75C, os mesmos valores tenham sido obtidos.

De modo geral, cenários com a mesma capacidade veicular, mas com $R_i > 0$, não tiveram alterações significativas a ponto de fazer com que houvesse produção em um período t que não havia antes, ou vice-versa. As Figuras (4b), (4c) e (4d) ilustram as diferenças (1) e (2) descritas anteriormente.

Tabela 8: Resumo dos resultados obtidos para cada combinação dos parâmetros C_v e R_i executada.

$V_{Cap}V_{eiculo}R_{atrasos}$	GAP (%)	UB	LB	Nós	Tempo
$V_{0.5C}R_0$	0,0000	686074	686074	9292	882,79
$V_{0.5C}R_1$	0,0000	680398	680398	14491	705,49
$V_{0.75C}R_0$	0,0000	683304	683304	2846	269,68
$V_{0.75C}R_1$	0,0000	673956	673956	5003	1082,24
$V_{1.0C}R_0$	0,0000	681198	681198	3469	989,21
$V_{1.0C}R_1$	0,0000	672507	672507	3027	294,35
$V_{1.25C}R_0$	0,0000	681198	681198	2967	688,71
$V_{1.25C}R_1$	0,0000	671283	671283	4708	1497,63
$V_{1.50C}R_0$	0,0000	681198	681198	4586	1632,30
$V_{1.50C}R_1$	0,0000	671283	671283	4216	1774,33
$V_{1.75C}R_0$	0,0000	681076	681076	4095	1427,46
$V_{1.75C}R_1$	0,0000	671283	671283	3990	1561,96
$V_{2.00C}R_0$	0,0000	680826	680826	4671	1277,54
$V_{2.00C}R_1$	0,0000	670911	670911	3598	1897,06

Fonte: Produzida pelos autores.

Por outro lado, os planos de produção obtidos para $R_i = 0$ são distintos em comparação aos cenários análogos de $R_i > 0$ de mesma capacidade veicular. Por exemplo, em todos os cenários de $R_i = 0$, o segundo período do horizonte foi utilizado para a realização de produção. Enquanto ao tolerar algum nível de atraso as soluções obtidas, em cada um dos casos, optou por iniciar a fabricação apenas no terceiro período do horizonte. As Figuras (4a), (4b), (4c) e (4d) exemplificam estas observações.

De forma geral, diferentemente do que ocorreu ao comparar os planos de produção obtidos quando $R_i = 1, 3$ e 5 , agora, ao comparar os planos obtidos para $R_i = 0$ com os gerados para $R_i = 1$ percebe-se a influência da variação deste parâmetro C_v . Mais precisamente, para os cenários até $1,50 \times C_v$ com $R_i = 0$, há produção em quatro dos seis períodos do horizonte, a saber, $t = 2, 4, 5$ e 6 . Por outro lado, para os casos com $R_i > 0$ a necessidade da utilização de quatro períodos ocorre apenas para o cenário com $0,50 \times C$, já que de $0,75 \times C$ a $1,75 \times C$ apenas três períodos são suficientes.

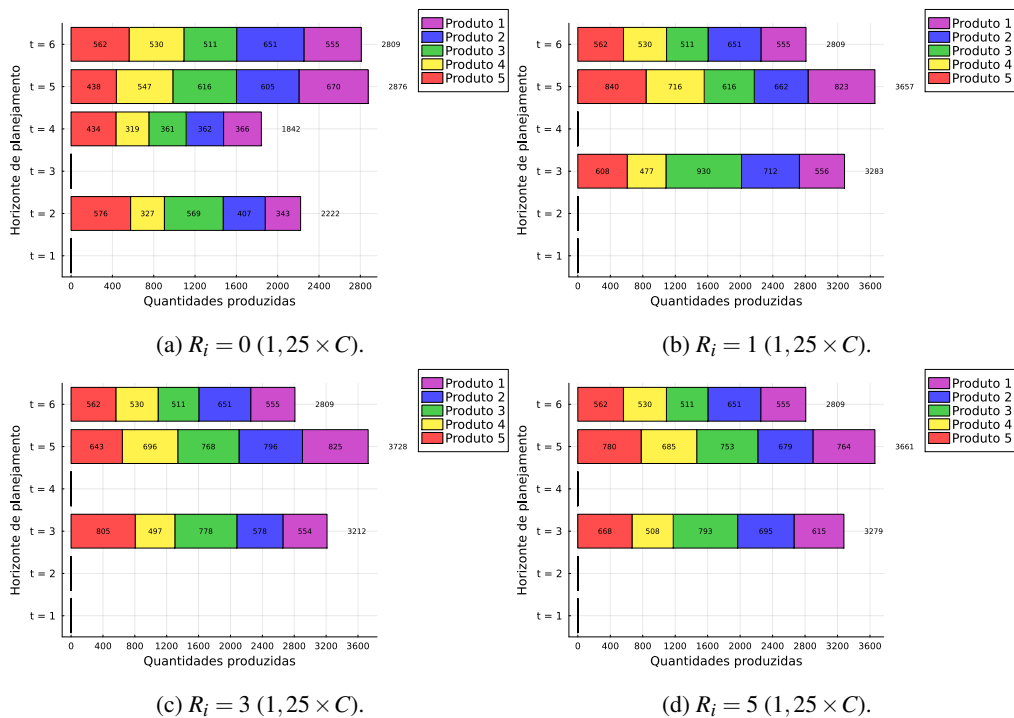


Figura 4: Planos de produção para a instância com $1,25 \times C$.

Fonte: Produzida pelos autores.

No que diz respeito aos planos de roteamento, ao permitir até um, três ou cinco períodos de atraso nas entregas, foi possível constatar que as rotas obtidas para os mesmos valores de capacidade veicular foram as mesmas. Mais especificamente, em todos esses casos, em cada período do horizonte de planejamento em que há a realização de entregas, apenas uma rota visitando todos os clientes foi utilizada. Portanto, visando a simplificação, apenas as rotas com obtidas com $R_i = 0$ e 1 são ilustradas nas Figuras 5 e 6. Também foi possível verificar que, mesmo ao não permitir atrasos, i.e., ao especificar $R_i = 0$, a estratégia de utilizar apenas um veículo para a realização das entregas se manteve. O que eventualmente diferiu foram: em quais períodos entregar e qual veículo utilizar (o de maior capacidade ou o de menor).

A permissão de atrasos $R_i > 0$, como pode ser constatado na Figura 6, além de possibilitar que o início dos atendimentos ocorra apenas no terceiro período, também permitiu o melhor aproveitamento das capacidades veiculares, já que houve redução no número de veículos utilizados ao longo do horizonte de planejamento, de quatro para três veículos.

Por fim, é importante destacar que dois *insights* gerenciais foram extraídos a partir da análise realizada. O primeiro foi observado nos cenários de $1,25C$. Em particular, ao especificar $R_i = 3$, apenas pedidos do segundo período foram entregues com atraso, enquanto, para $R_i = 1$ e $R_i = 5$, pedidos de dois períodos ($t = 2$ e $t = 5$) apresentaram atrasos. Assim, mais do que a existência

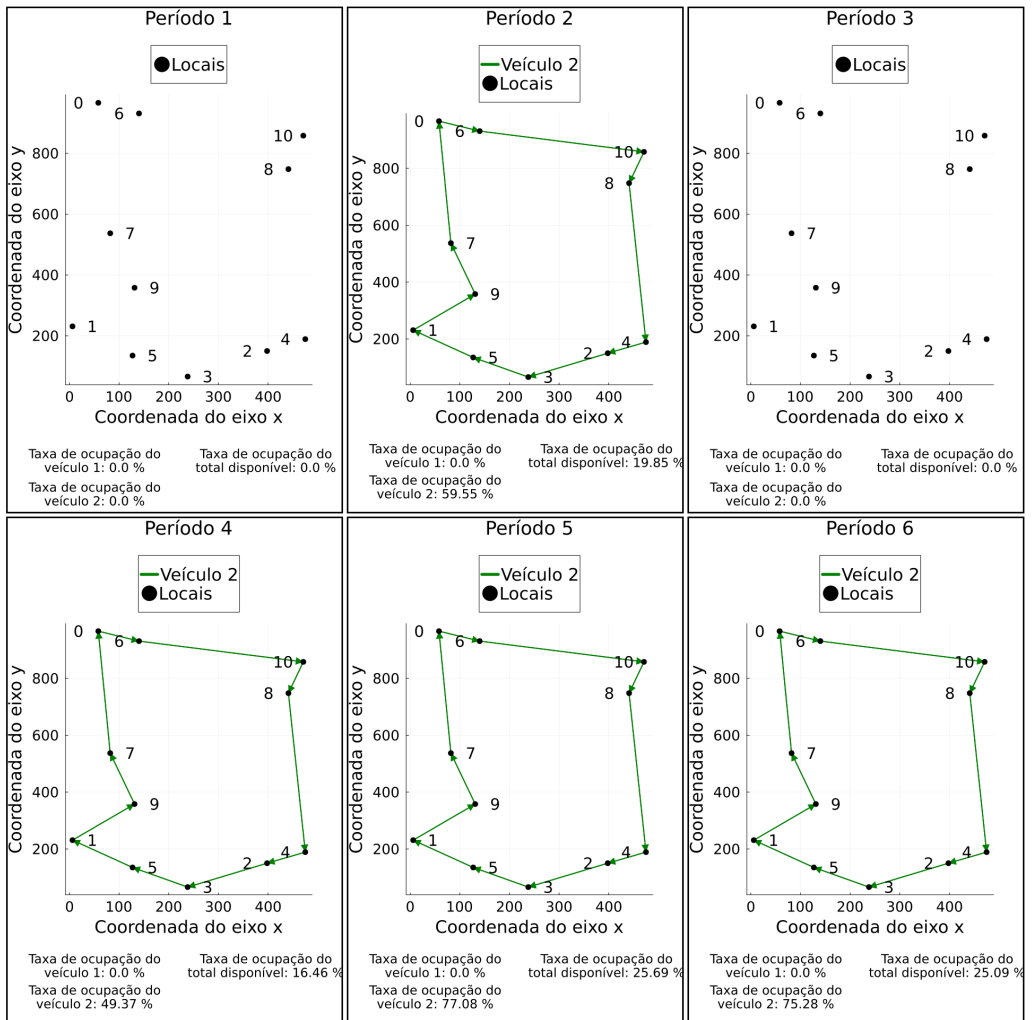


Figura 5: Plano de rotas para a instância - $R_i = 0$ e $1,25 \times C$.

Fonte: Produzida pelos autores.

de múltiplas soluções ótimas, essa característica distinta em relação à quantidade de períodos com demandas atrasadas no horizonte de planejamento é um aspecto relevante para o tomador de decisão, considerando que a frequência de atrasos pode impactar a imagem da empresa.

O segundo achado pode ser observado com os dados referentes a $0,75C$. Mais especificamente, mesmo permitindo até três períodos de atraso, as demandas dos períodos $t = 2, 4$ e 5 são entregues com até um período de atraso. Entretanto, ao especificar $R_i = 5$, além de ter pedidos de $t = 2, 4$ e 5 entregues em $t = 3, 5$ e 6 , também se verificou que parte dos pedidos do quarto período foi entregue no sexto. Assim, houve soluções ótimas de mesmo valor funcional, mas com características distintas em relação ao quão atrasado certos produtos foram entregues. Essa

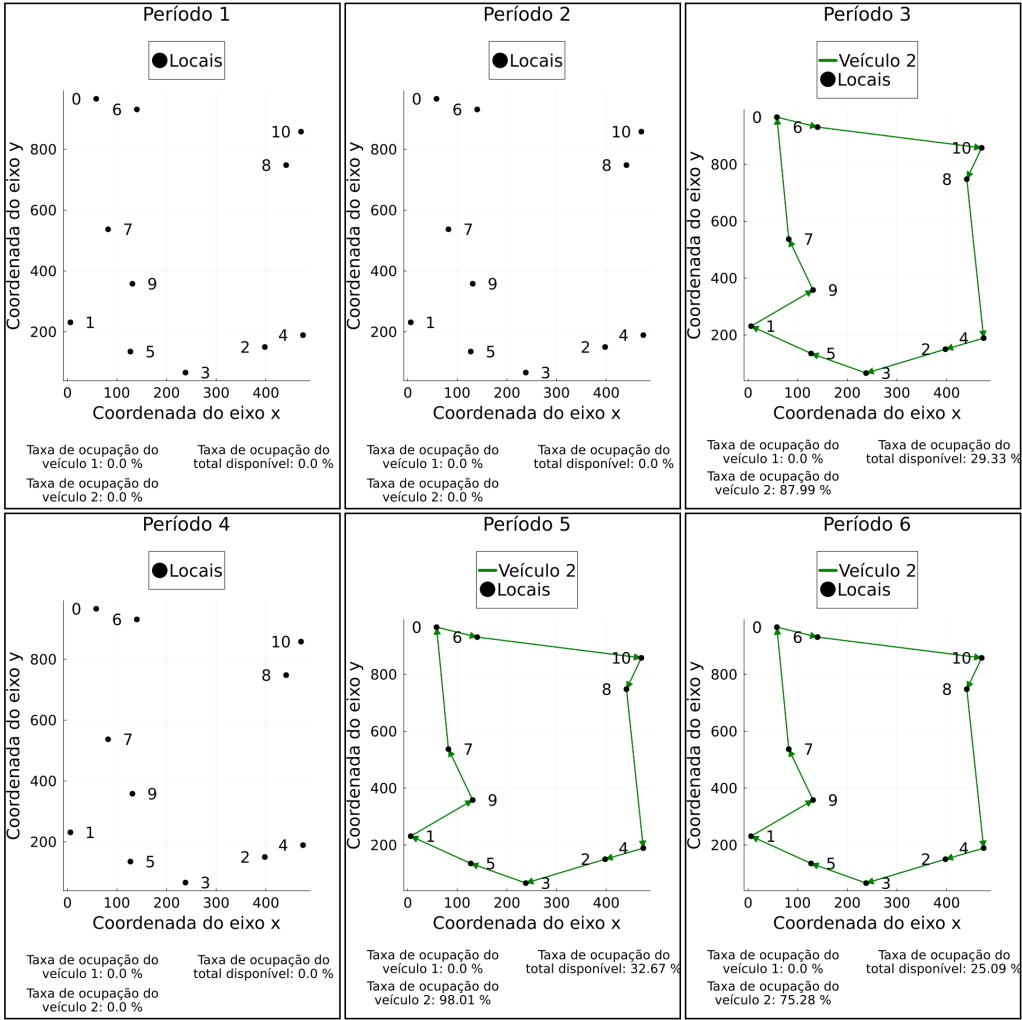


Figura 6: Plano de rotas para a instância - $R_i = 1$ e $1,25 \times C$ ($VarCR.Het$ 13).

Fonte: Produzida pelos autores.

informação pode ser relevante para o tomador de decisão, uma vez que o nível de atendimento da demanda é importante para a empresa.

Em resumo, ao variar os parâmetros C_v e R_i considerando uma instância, pode-se constatar que, ao não permitir a flexibilização das entregas, todos os cenários iniciaram a produção no segundo período do horizonte, enquanto, nos casos com $R_i > 0$, tal produção começa no terceiro período. Além disso, em todos os cenários, independentemente de $R_i = 0$ ou $R_i > 0$, apenas uma rota visitando todos os clientes foi utilizada, sendo que as mudanças identificadas nos planos de roteamento se limitaram a quais e quantos períodos utilizar para realizar as entregas, bem como a escolha do veículo a ser utilizado (o de maior capacidade ou o de menor).

O aumento da capacidade veicular mostrou uma tendência de redução no número de vezes em que as visitas são realizadas ao longo do horizonte. A possibilidade de atrasos, além de permitir o início das entregas um período após o caso sem flexibilização, também possibilitou eventuais reduções tanto no número de períodos em que a produção dos produtos é realizada quanto no número de entregas efetuadas ao longo do horizonte. Por fim, dado que todas as 28 soluções obtidas são ótimas, essa análise possibilitou identificar que flexibilizar os prazos de entrega, permitindo algum nível de atraso, de fato viabiliza a obtenção de soluções com melhores custos totais. Entretanto, o valor de permitir até três ou cinco períodos de atraso é o mesmo que permitir, no máximo, um período de atraso.

5 CONCLUSÕES

Este trabalho abordou o problema integrado de produção, estoque e roteamento, no qual se permite realizar entregas com períodos de atraso, mediante um custo penalizador proporcional às quantidades de períodos atrasados e às quantidades entregues em atraso. Um modelo de otimização linear inteira, visando à minimização dos custos de produção, de estoque, de roteamento e de penalização por atraso, foi proposto e implementado. Experimentos computacionais foram realizados com o objetivo de verificar o desempenho, e análises de sensibilidade foram conduzidas.

Experimentos iniciais permitiram a comparação das soluções obtidas pelo modelo, especificando a permissão de atrasos em zero, com um clássico da literatura. Em particular, o modelo apresentado em [19] obteve melhor desempenho quando se considera o tempo de execução.

Além disso, também foram conduzidos testes computacionais para analisar os impactos nos custos ao variar o parâmetro que determina quantos períodos são permitidos para atraso e o parâmetro responsável pela penalização por tais atrasos. Embora reduções próximas a 18,06% e 16,20% sejam alcançadas, respectivamente, nos custos relacionados aos estoques e transportes, o fato de as reduções nos custos totais ficarem, em média, restritas a 0,57% mostra que a componente produtiva é quem tem maior influência nas análises. Sendo assim, estudos visando análises de sensibilidade ao variar os custos relacionados a esta componente, bem como que considerem instâncias de maior porte, podem ser possibilidades para novos estudos.

Todavia, por mais que as economias médias obtidas fiquem em torno de 0,57%, o que pode representar um valor não desprezível dependendo dos dados reais, de modo geral, a permissão de entregas com algum grau de atraso no período em que elas são realizadas de fato levou a reduções nos custos totais. Como era de se esperar, as maiores economias são alcançadas quando o custo pelo atraso é menor.

A análise de sensibilidade particular permitiu identificar que a flexibilização dos períodos de entrega de fato pode conduzir a economias de até 1,39% no melhor dos cenários testados. Entretanto, apenas um período extra foi capaz de apresentar resultados tão bons quanto permitir até três ou cinco períodos. Tal análise também possibilitou a obtenção de alguns *insights* gerenciais que podem ser de grande utilidade para os tomadores de decisão.

Uma das principais limitações do modelo estendido é o fato de não levar em consideração a limitação do tempo de cada rota; ou seja, este modelo pode ser aplicado apenas em contextos em que as rotas podem ser realizadas dentro de um único período. Portanto, uma das possibilidades para pesquisas futuras é a incorporação de restrições visando à limitação da duração das rotas. Além disso, outra possibilidade é realizar uma análise de sensibilidade em relação ao impacto que a permissão de mais de uma viagem por período gera na solução, bem como nos custos de cada componente da função objetivo. Finalmente, um dos trabalhos futuros consiste em gerar novas instâncias e desenvolver abordagens de soluções heurísticas.

Agradecimentos

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e do CNPq Proc. 301967/2025-3. Também agradeço ao Laboratório de Otimização do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo – *campus* de São Carlos por me receberem e ao Centro de Ciências Matemáticas Aplicadas à Indústria (CeMEAI) financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (proc. 2013/07375-0) pelo acesso disponibilizado ao Cluster Euler.

Disponibilidade de dados

Os conjuntos de dados relacionados a este artigo estão disponíveis mediante solicitação ao autor correspondente.

Editor associado: Kelly Cristina Poldi

ABSTRACT. This paper investigates the integrated problem of production, inventory and routing with flexible delivery times, i.e., deliveries can occur in a set of periods, but subject to penalties. To deal with the problem, an integer linear optimization model is developed that allows flexibility in the delivery deadlines of the demanded items. A production environment is considered with a single factory, capable of producing and storing a finite quantity of distinct products that share production resources. There is a set of customers with their own inventories and a limited fleet of heterogeneous vehicles. Each vehicle performs at most one route per period. The objective of the problem is to determine the production, inventory and routing plan that minimizes costs over a finite horizon, meeting demands within the possible periods (*flexible due dates*). Analyses considering a set of instances with a heterogeneous fleet show that exploring flexibility in delivery times can significantly reduce costs, especially inventory costs. In addition, different scenarios with variations in vehicle capacity and delivery deadlines are evaluated.

Keywords: integrated production, inventory and routing problem; flexible due dates; integer optimization models; sensitivity analysis.

REFERÊNCIAS

- [1] Y. Adulyasak, J.F. Cordeau & R. Jans. The production routing problem: A review of formulations and solution algorithms. *Computers & Operations Research*, **55** (2015), 141–152.

- [2] C. Archetti, E. Fernández & D.L. Huerta-Muñoz. The Flexible Periodic Vehicle Routing Problem. *Computers & Operations Research*, **85** (2017), 58–70. doi:10.1016/j.cor.2017.03.008. URL <https://www.sciencedirect.com/science/article/pii/S0305054817300734>.
- [3] C. Archetti, O. Jabali & M.G. Speranza. Multi-period Vehicle Routing Problem with Due dates. *Computers & Operations Research*, **61** (2015), 122–134. URL <https://www.sciencedirect.com/science/article/pii/S0305054815000726>.
- [4] P. Chandra. A Dynamic Distribution Model with Warehouse and Customer Replenishment Requirements. *The Journal of the Operational Research Society*, **44**(7) (1993), 681–692. doi:10.2307/2584042.
- [5] P. Chandra & M.L. Fisher. Coordination of production and distribution planning. *European journal of operational research*, **72**(3) (1994), 503–517.
- [6] M. Darvish, C. Archetti, L.C. Coelho & M.G. Speranza. Flexible two-echelon location routing problem. *European Journal of Operational Research*, **277** (2019), 1124–1136. doi:10.1016/j.ejor.2019.04.002.
- [7] M. Darvish & L.C. Coelho. Sequential versus integrated optimization: Production, location, inventory control, and distribution. *European Journal of Operational Research*, **268**(1) (2018), 203–214. doi:10.1016/j.ejor.2018.01.028. URL <https://www.sciencedirect.com/science/article/pii/S0377221718300493>.
- [8] F. Fumero & C. Vercellis. Synchronized development of production, inventory, and distribution schedules. *Transportation science*, **33**(3) (1999), 330–340.
- [9] D. Hrabec, L.M. Hvattum & A. Hoff. The value of integrated planning for production, inventory, and routing decisions: A systematic review and meta-analysis. *International Journal of Production Economics*, **248** (2022), 108468.
- [10] M.P. Kidd, M. Darvish, L.C. Coelho & B. Gendron. A relax-and-restrict matheuristic for supply chain network design with facility location and customer due date flexibility. *Transportation Research Part E: Logistics and Transportation Review*, (2023). URL <https://www.sciencedirect.com/science/article/pii/S1366554523003587>.
- [11] H. Larrain, L.C. Coelho, C. Archetti & M.G. Speranza. Exact solution methods for the multi-period vehicle routing problem with due dates. *Computers & Operations Research*, **110** (2019), 148–158. doi:10.1016/j.cor.2019.05.026.
- [12] H.L. Maryam Darvish & L.C. Coelho. A dynamic multi-plant lot-sizing and distribution problem. *International Journal of Production Research*, **54**(22) (2016), 6707–6717. doi:10.1080/00207543.2016.1154623.
- [13] P.L. Miranda, R. Morabito & D. Ferreira. Optimization model for a production, inventory, distribution and routing problem in small furniture companies. *Top*, **26** (2018), 30–67.
- [14] P.L. Miranda, R. Morabito & D. Ferreira. Mixed integer formulations for a coupled lot-scheduling and vehicle routing problem in furniture settings. *INFOR: Information Systems and Operational Research*, **57**(4) (2019), 563–596.

- [15] S. Moons, K. Ramaekers, A. Caris & Y. Arda. Integrating production scheduling and vehicle routing decisions at the operational decision level: A review and discussion. *Computers & Industrial Engineering*, **104** (2017), 224–245.
- [16] N.A. Mostafa & A.B. Eltawil. The production-inventory-distribution-routing problem: An integrated formulation and solution framework. In “2015 International Conference on Industrial Engineering and Operations Management (IEOM)”. IEEE (2015), p. 1–10.
- [17] R.G. Ramesh Kumar, L. Ganapathy & M.K. Tiwari. Quantitative approaches for the integration of production and distribution planning in the supply chain: a systematic literature review. *International Journal of Production Research*, **58**(11) (2020), 3527–3553. doi:10.1080/00207543.2020.1762019.
- [18] B. Ramos, C. Alves & J. Valerio de Carvalho. An arc flow formulation to the multitrip production, inventory, distribution, and routing problem with time windows. *International Transactions in Operational Research*, **29**(1) (2022), 526–553.
- [19] L.F. Rodrigues, M.O.D. Santos & B. Almada-Lobo. A Memetic Algorithm for the multi-product Production Routing Problem. *Computers & Industrial Engineering*, (2023), 109388.
- [20] A.M. Sarmiento & R. Nagi. A review of integrated analysis of production–distribution systems. *IIE transactions*, **31**(11) (1999), 1061–1074.
- [21] D.Y. Wang, O. Grunder & A.E. Moudni. Integrated scheduling of production and distribution operations: a review. *International Journal of Industrial and Systems Engineering*, **19**(1) (2015), 94–122.

How to cite

A.A.N. Machado & M.O. Santos. O Problema Integrado de Produção e Roteamento com Flexibilidade nos Prazos de Entrega. *Trends in Computational and Applied Mathematics*, **27**(2026), e01861. doi: 10.5540/tcam.2026.027.e01861.

