

# Série dos Seminários de Acompanhamento à Pesquisa

**DEI**  
DEPARTAMENTO  
DE ENGENHARIA  
INDUSTRIAL

Número 07 | 05 2021

## Seleção de cenários com Weighted Set Covering

Autor(es):

Isabella Fischer Guindani Vieira



# Série dos Seminários de Acompanhamento à Pesquisa

Número 07 | 05 2021

## Seleção de cenários com Weighted Set Covering

Autor(es):

Isabella Fischer Guindani Vieira

### CRÉDITOS:

SISTEMA MAXWELL / LAMBDA  
<https://www.maxwell.vrac.puc-rio.br/>

**Organizadores:** Fernanda Baião / Soraida Aguilar

**Layout da Capa:** Aline Magalhães dos Santos



# Isabella Fischer

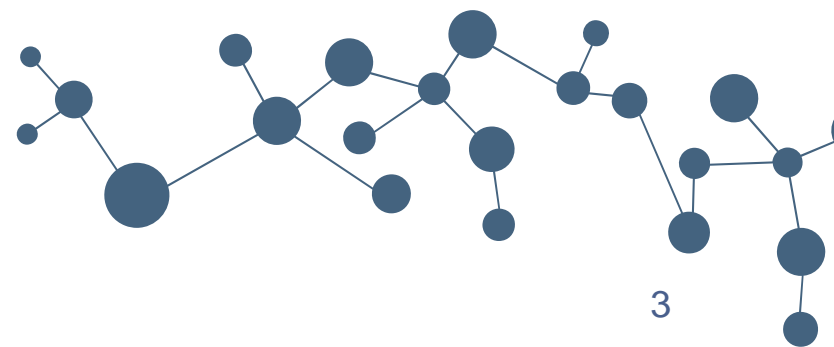
Mestrado

Áres de concentração: Transporte e Logística

Linha de Pesquisa: Planejamento de Sistemas de Transporte e Logística

5º período

Orientador: Rafael Martinelli



# Observação e identificação do problema

- Contexto geral: cálculo de Demanda Estocástica de ferramentas e materiais em empresa de grande porte do setor de energia.
- Planejamento determinístico com um horizonte extenso, grande número de tarefas e suscetível a diferentes fontes de incerteza.

# Observação e identificação do problema

Incertezas são calculadas e definidas a partir de dados históricos de cada tarefa e refletem em mudanças nos atributos:

- Data de início;
- Duração;
- Continuidade da sequência de tarefas.

# Observação e identificação do problema

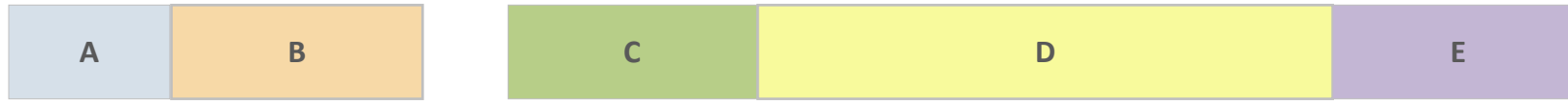
- Devido à confidencialidade dos dados históricos, as distribuições de probabilidade destas incertezas não são disponibilizadas a todos os setores, incluindo o setor responsável pelo cálculo de demanda de equipamentos.
- Para garantir a correta representação das distribuições, são utilizadas milhares de replicações do cronograma determinístico, aplicando valores sorteados na distribuição de cada tipo de incerteza, para cada atributo.

# Observação e identificação do problema

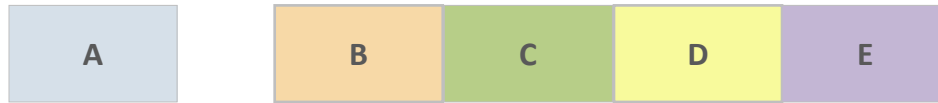
CRONOGRAMA  
DETERMINÍSTICO



CRONOGRAMA 1



CRONOGRAMA 2



CRONOGRAMA 3

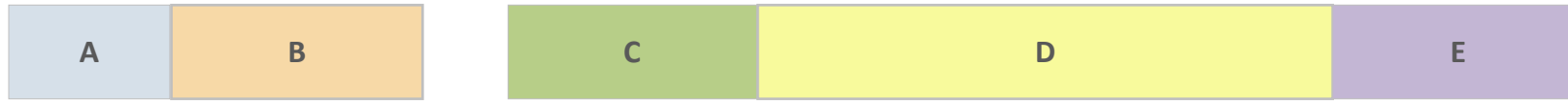


# Observação e identificação do problema

CRONOGRAMA  
DETERMINÍSTICO



CRONOGRAMA 1



CRONOGRAMA 2



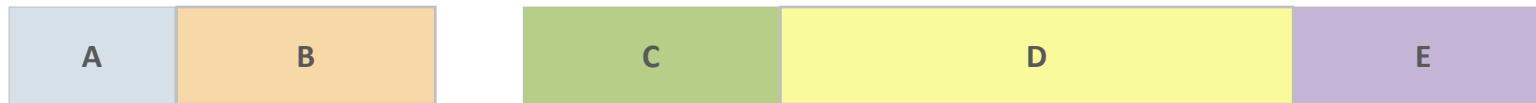
CRONOGRAMA 3



⋮

⋮

CRONOGRAMA 78



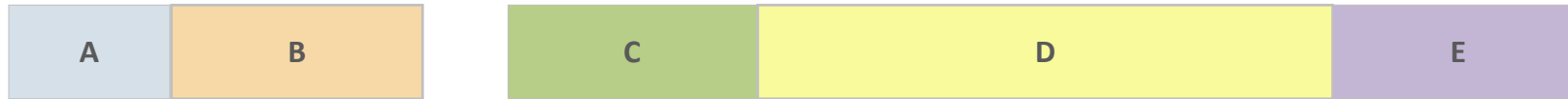


# Observação e identificação do problema

CRONOGRAMA  
DETERMINÍSTICO



CRONOGRAMA 1



CRONOGRAMA 2



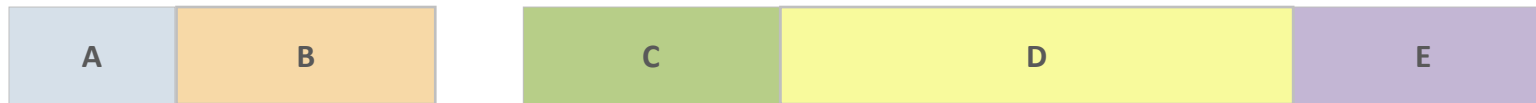
CRONOGRAMA 3



⋮

⋮

CRONOGRAMA 78



⋮

⋮

CRONOGRAMA 145

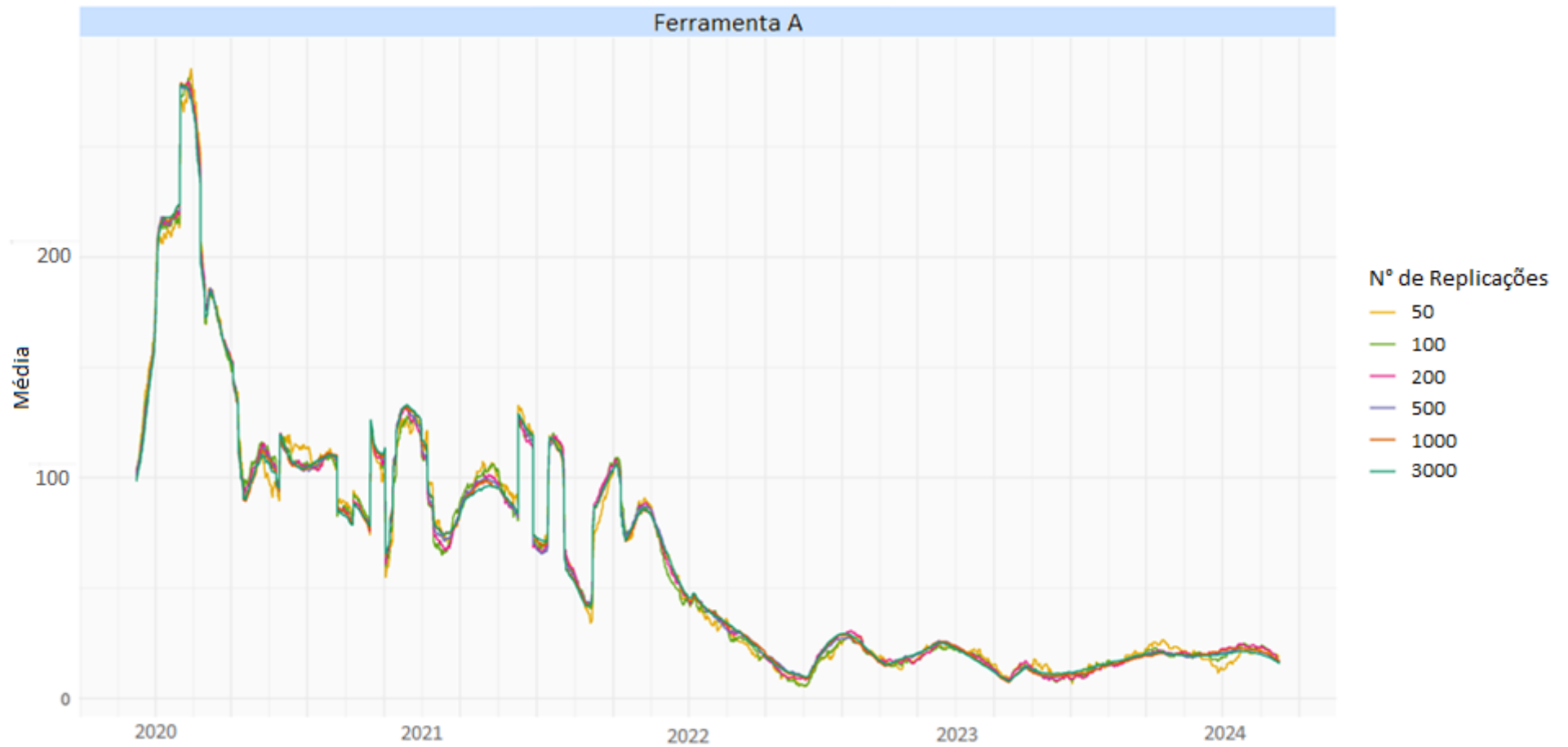


# Observação e identificação do problema

- Estes cenários semelhantes causam impacto na demanda que justifique todo este custo de tempo e processamento de dados?
- Para responder tal pergunta foi realizada uma análise de sensibilidade da demanda calculada ao variar o número de replicações:

3000 – 1000 – 500 – 200 – 100 – 50

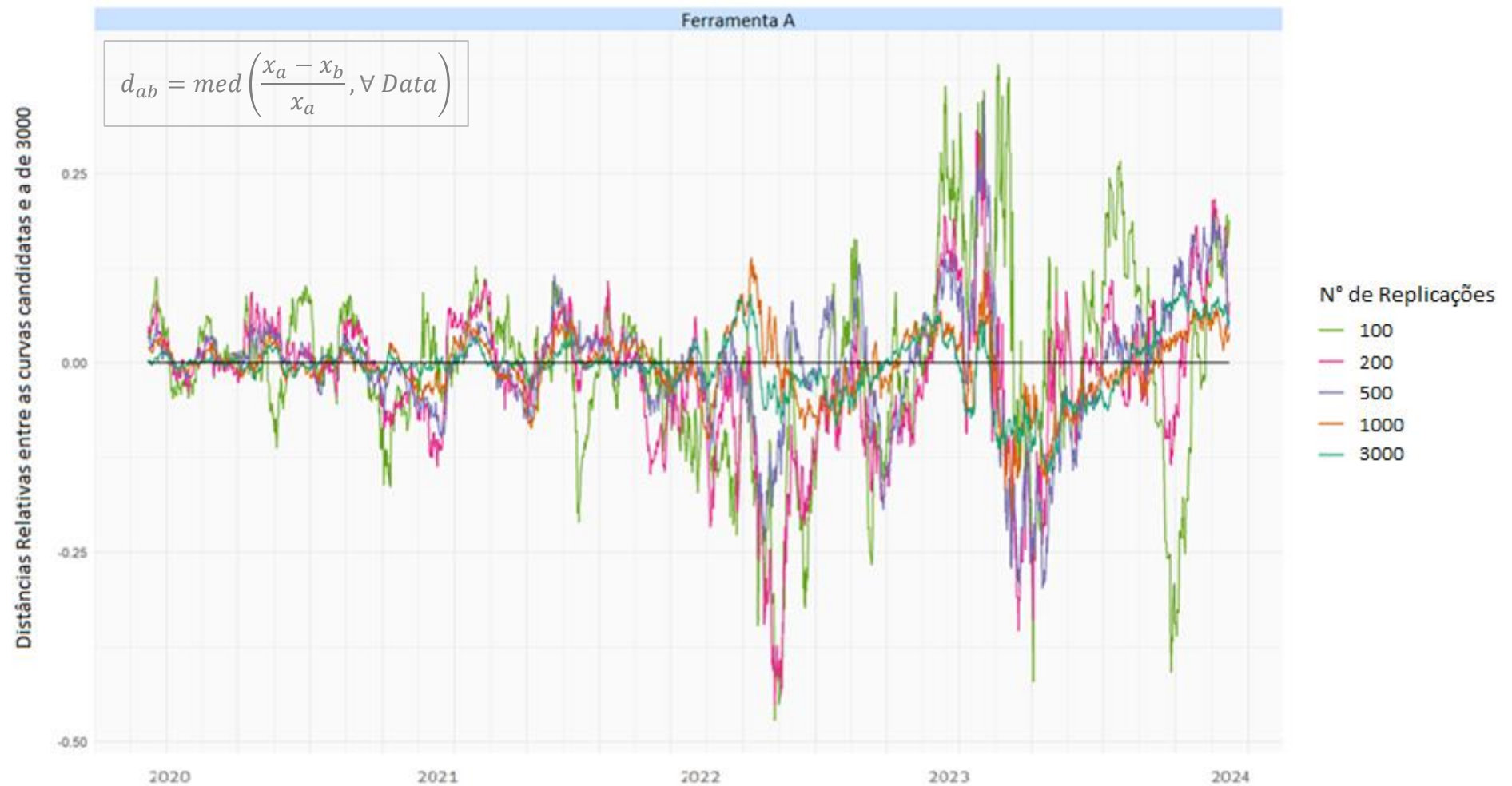
# Observação e identificação do problema



# Observação e identificação do problema



# Observação e identificação do problema



# Observação e identificação do problema

- Média das distâncias relativas para todo o horizonte – representa o quanto os indicadores de cada quantidade de iterações diferem, em média, da curva de 3000 replicações.

Nº Replicações	Média	P10	P50	P90	Desvio padrão	Desvio padrão/ Média	Variância
50	-2,3%	41,6%	-3,1%	-16,8%	-26,1%	-9,4%	-20,3%
100	-6,7%	24,9%	-3,2%	-7,7%	-18,7%	-3,9%	-16,9%
200	-3,4%	14,5%	-2,2%	-2,7%	-11,4%	-2,1%	-9,2%
500	-1,0%	5,0%	-0,9%	-0,6%	-4,5%	-0,7%	-1,2%
1000	-3,3%	3,2%	0,5%	-0,7%	-6,5%	-1,7%	-6,9%

# Observação e identificação do problema

- Média das distâncias relativas para todo o horizonte – representa o quanto os indicadores de cada quantidade de iterações diferem, em média, da curva de 3000 replicações.

Nº Replicações	Média	P10	P50	P90	Desvio padrão	Desvio padrão/ Média	Variância
50	-2,3%	41,6%	-3,1%	-16,8%	-26,1%	-9,4%	-20,3%
100	-6,7%	24,9%	-3,2%	-7,7%	-18,7%	-3,9%	-16,9%
200	-3,4%	14,5%	-2,2%	-2,7%	-11,4%	-2,1%	-9,2%
500	-1,0%	5,0%	-0,9%	-0,6%	-4,5%	-0,7%	-1,2%
1000	-3,3%	3,2%	0,5%	-0,7%	-6,5%	-1,7%	-6,9%

# Observação e identificação do problema

- Seleção aleatória: amostras de 500 iterações

$$d_{ab} = med\left(\frac{x_a - x_b}{x_a}, \forall Data\right)$$

Distâncias Relativas	1	2	3	4	5	6	7	8	9	10
1	0	-0,00108	-0,01281	-0,01007	-0,01546	-0,00560	-0,00480	-0,00194	-0,00377	0,00445
2	-0,00269	0	-0,01453	-0,01130	-0,01795	-0,00616	-0,00531	-0,00238	-0,00351	0,00374
3	0,00985	0,00977	0	0,00178	-0,00362	0,00544	0,00672	0,00918	0,00759	0,01601
4	0,00556	0,00574	-0,00537	0	-0,00710	0,00196	0,00279	0,00527	0,00397	0,01253
5	0,01098	0,01019	0,00018	0,00374	0	0,00641	0,00676	0,01053	0,00818	0,01703
6	0,00220	0,00320	-0,00943	-0,00580	-0,01226	0	-0,00011	0,00157	0,00065	0,00866
7	0,00080	0,00191	-0,01031	-0,00718	-0,01443	-0,00225	0	0,00093	0,00009	0,00791
8	-0,00228	-0,00111	-0,01383	-0,01056	-0,01626	-0,00657	-0,00496	0	-0,00332	0,00469
9	-0,00091	0,00101	-0,01217	-0,00864	-0,01554	-0,00421	-0,00261	-0,00012	0	0,00611
10	-0,00832	-0,00749	-0,01947	-0,01583	-0,02233	-0,01198	-0,01059	-0,00786	-0,00966	0



# Observação e identificação do problema

- Incertezas causam pouca variabilidade nos cronogramas criados
- Cenários ficam muito similares



# Definição do problema genérico

Grande Base de Dados

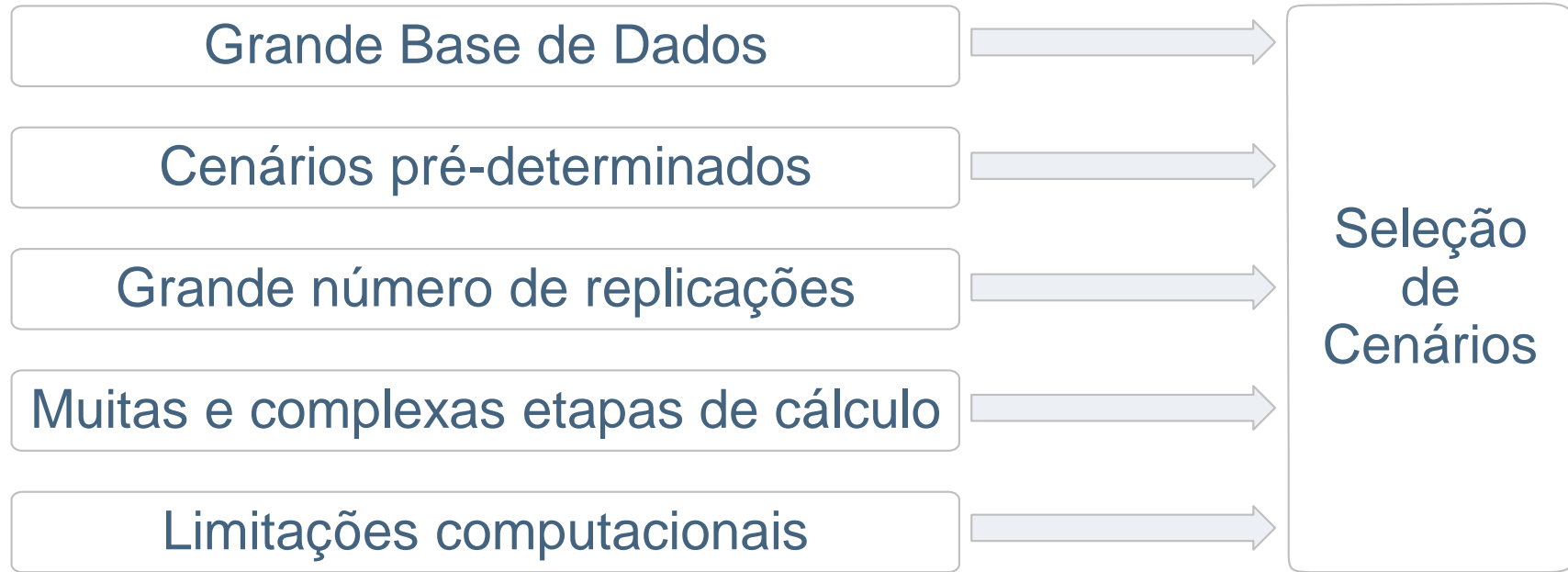
Cenários pré-determinados

Grande número de replicações

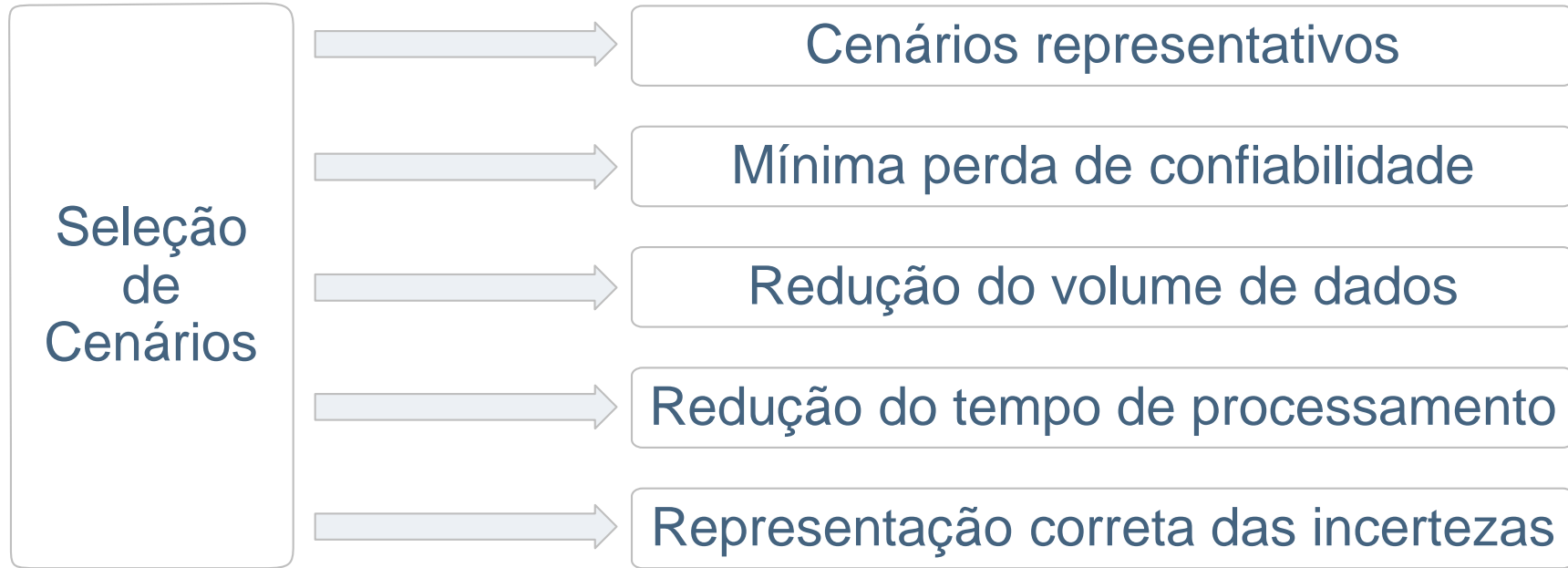
Muitas e complexas etapas de cálculo

Limitações computacionais

# Solução do problema genérico



# Solução do problema genérico



# Proposta de solução

## SELEÇÃO DE CENÁRIOS COM WEIGHTED SET COVERING

Cenários:

$$S = \{s_1, s_2, \dots, s_n\}, \text{ onde } s_d \in S$$

Distância:

$$f: S \times S \rightarrow \mathbb{R}^+$$

Valor de corte:

$$\rho \in \mathbb{R}^+$$

Arestas:

$$E = \{(s, s') | s \in S \wedge s' \in S\}$$

Critério de cobertura:

$$C_s = \{s\} \cup \{s' | (s, s') \in E \wedge f(s, s') \leq \rho\}$$

# Proposta de solução

## SELEÇÃO DE CENÁRIOS COM WEIGHTED SET COVERING

$$\min z = \sum_{s \in S} f(s, s_d) x_s$$

subject to

$$\sum_{s' \in C_s} x_{s'} \geq 1 \quad \forall s \in S$$

$$x_s \in \{0,1\} \quad \forall s \in S$$

# Proposta de solução

## Weighted Set Covering

- Modelo exato
- Heurísticas/Metaheurísticas

Modelo clássico da literatura: K-Medoids



# Resultados Preliminares

- Matriz de distâncias euclidianas entre as replicações (atributo: mudança na duração)
- Valor de corte: um cenário não pode ser coberto por outro se a distância entre eles for maior que 20%
- Heurística construtiva (first solution)
- Escolhe primeiro os sets com maior número de cenários ainda não cobertos e pára quando obtiver cobertura completa

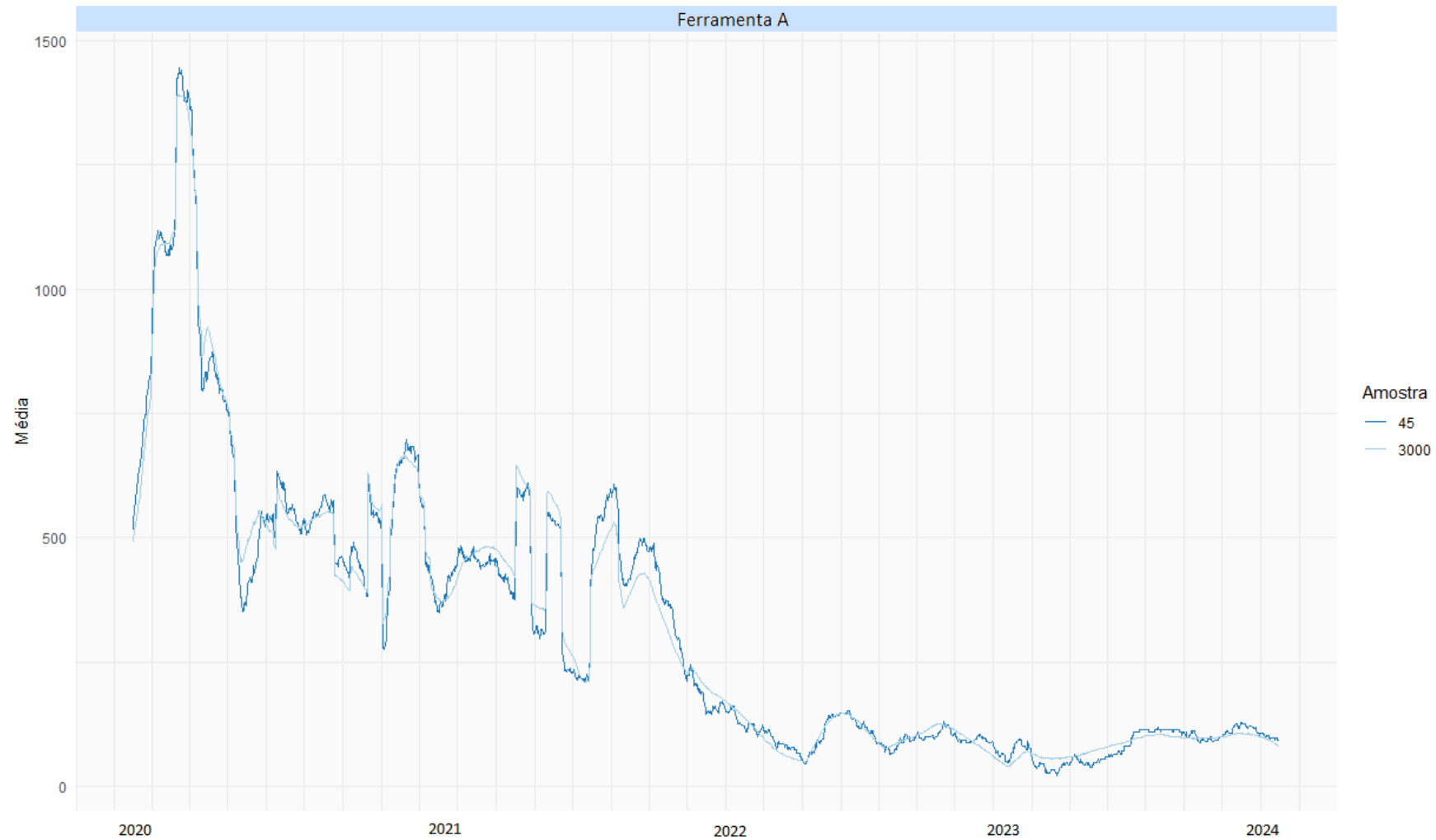


# Resultados Preliminares

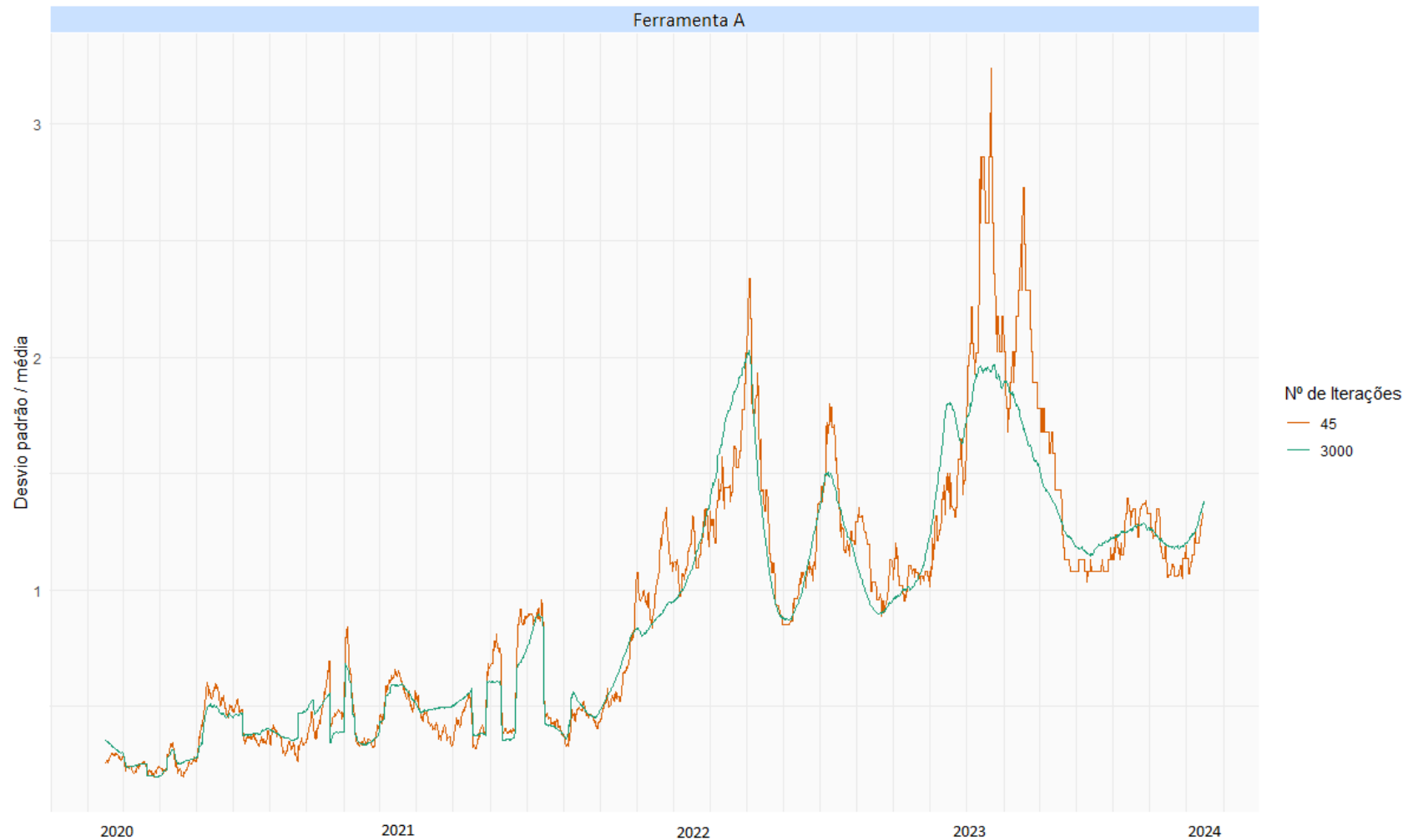
- Foram selecionados 45 cenários, cuja demanda calculada passou pela mesma análise feita anteriormente

Método	Nº Replicações	Média	P10	P50	P90	Desvio padrão	Desvio padrão/Média	Variância
Aleatório	50	-2,3%	41,6%	-3,1%	-16,8%	-26,1%	-9,4%	-20,3%
Aleatório	100	-6,7%	24,9%	-3,2%	-7,7%	-18,7%	-3,9%	-16,9%
Aleatório	200	-3,4%	14,5%	-2,2%	-2,7%	-11,4%	-2,1%	-9,2%
Aleatório	500	-1,0%	5,0%	-0,9%	-0,6%	-4,5%	-0,7%	-1,2%
Aleatório	1000	-3,3%	3,2%	0,5%	-0,7%	-6,5%	-1,7%	-6,9%
Set Cover	45	-1,8%	0,5%	-0,2%	3,2%	-0,7%	-2,3%	-0,4%

# Resultados Preliminares



# Resultados Preliminares



# Resultados Preliminares

Melhorias a serem feitas

- Metaheurísticas e local search
- Diferentes métodos de cálculo de distância entre cenários
- Análise de sensibilidade de valores de corte



Obrigada!

[isabellafischer@outlook.com](mailto:isabellafischer@outlook.com)

