



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOGIÊNCIAS
DEPARTAMENTO DE ENGENHARIA CIVIL
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA CIVIL

Oscar Simón Melgar Cisneros

**Análise Estatística Multivariada e Otimização do Processo de Abertura de
Cavernas Subterrâneas em Rochas Salinas Utilizando PCA, t-SNE e Algoritmos
Genéticos**

Recife

2025

Oscar Simón Melgar Cisneros

Análise Estatística Multivariada e Otimização do Processo de Abertura de Cavernas Subterrâneas em Rochas Salinas Utilizando PCA, t-SNE e Algoritmos Genéticos

Tese apresentada ao Programa de Pós-Graduação em Engenharia Civil da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de doutor em Engenharia Civil.

Área de Concentração: Simulação e Gerenciamento de Reservatórios de Petróleo.

Orientador: Prof. Dr. Leonardo José do Nascimento Guimarães

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Cisneros, Oscar Simón Melgar.

Análise estatística multivariada e otimização do processo de abertura de cavernas subterrâneas em rochas salinas utilizando PCA, t-SNE e algoritmos genéticos / Oscar Simón Melgar Cisneros. - Recife, 2025.
160f.: il.

Tese (Doutorado)- Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Programa de Pós-Graduação em Engenharia Civil, 2025.

Orientação: Prof. Dr. Leonardo José do Nascimento Guimarães.

1. Abertura de cavernas; 2. SALGAS; 3. Simulações numéricas.
I. Guimarães, Leonardo José do Nascimento. II. Título.

UFPE-Biblioteca Central

OSCAR SIMÓN MELGAR CISNEROS

**ANÁLISE ESTATÍSTICA MULTIVARIADA E OTIMIZAÇÃO DO PROCESSO DE
ABERTURA DE CAVERNAS SUBTERRÂNEAS EM ROCHAS SALINAS:
UTILIZANDO PCA, T-SNE E ALGORITMOS GENÉTICOS**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Civil da Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, como requisito parcial para obtenção do título de Doutor em Engenharia Civil, Área de Simulação e Gerenciamento de Reservatórios de Petróleo.

Aprovada em 28/08/2025

Orientador: Prof. Dr. Leonardo José do Nascimento Guimarães, UFPE

BANCA EXAMINADORA

participação por videoconferência
Prof. Dr. Igor Fernandes Gomes (examinador interno)
Universidade Federal da Paraíba

participação por videoconferência
Prof. Dr. Jonathan da Cunha Teixeira (examinador externo)
Universidade Federal de Alagoas

participação por videoconferência
Prof.^a Dr.^a Lícia Mouta da Costa (examinadora externa)
Universidade Federal de Pernambuco

participação por videoconferência
Prof. Dr. Juan Alberto Rojas Tueros (examinador externo)
Universidade Federal de Pernambuco

participação por videoconferência
Dr. Ramon Barboza de Vasconcelos (examinador externo)
Universidade Federal de Pernambuco

AGRADECIMENTOS

Ao meu orientador, Professor Leonardo José do Nascimento Guimarães, expresso minha profunda gratidão pela paciência, apoio, orientações e explicações dedicadas, que foram fundamentais para a realização e conclusão desta tese.

Agradeço à Petrobras pela oportunidade e pela confiança depositada neste trabalho. Estendo meus agradecimentos ao Laboratório de Métodos Computacionais em Geomecânica (LMCG) e aos meus colegas de laboratório, pelo apoio constante e pela enriquecedora troca de conhecimentos.

Sou grato também aos meus amigos e a todos na universidade que, de alguma maneira, colaboraram para a culminação desta jornada.

Por fim, um agradecimento especial à minha família. À minha mãe, Natividad Cisneros de Melgar, pelo seu amor, força e apoio incondicional em todos os momentos. E ao meu pai, Simón Timoteo Melgar Gomez (em memória), por ter sido minha maior inspiração e por seus ensinamentos que me guiaram até aqui.

RESUMO

O consumo global de energia tem impulsionado a necessidade de armazenamento subterrâneo, sendo as cavernas de sal uma solução amplamente utilizada para materiais energéticos. A construção e o controle dessas cavernas são, portanto, cruciais para garantir sua capacidade e segurança operacional. Este trabalho apresenta uma metodologia integrada para a análise e otimização do processo de dissolução para abertura de cavernas. Primeiramente, foram realizadas simulações numéricas do processo de mineração por dissolução em uma rocha de cloreto de sódio, utilizando o *software SALGAS SOLUTION MINING SIMULATION SOFTWARE* para diferentes métodos de circulação (direta e reversa). Para interpretar a complexa interação entre as variáveis de entrada e os resultados, foi realizada uma análise estatística multivariada com as técnicas de redução de dimensionalidade, *Principal Component Analysis* (PCA) e *t-distributed Stochastic Neighbor Embedding* (t-SNE). A análise PCA revelou que a taxa de produção de salmoura e suas variáveis correlatas constituem o principal componente de variabilidade do processo, enquanto as variáveis associadas à temperatura e às propriedades do fluido desempenham um papel secundário, mas significativo. A técnica t-SNE confirmou esses achados e demonstrou sua capacidade de agrupar cenários com características operacionais similares. Com base nos *insights* obtidos, o trabalho avança da análise para o *design*, implementando um *framework* de otimização que acopla o simulador SALGAS a Algoritmos Genéticos (AG). Foram conduzidas otimizações mono-objetivo, para maximizar o volume final, e multiobjetivo, para explorar os complexos *trade-offs* entre os objetivos de volume, massa de Sal, eficiência energética e Tempo de Construção. Os resultados demonstram que o AG é capaz de identificar vetores de decisão que geram projetos superiores aos cenários de base, caracterizando a fronteira de Pareto de soluções ótimas. A metodologia proposta, portanto, oferece uma abordagem sistemática e robusta para o projeto de cavernas salinas.

Palavras-chave: abertura de cavernas; SALGAS; simulações numéricas; análises estatísticas; algoritmos genéticos.

ABSTRACT

Global energy consumption has driven the need for underground storage, with salt caverns being a widely used solution for energy materials. The construction and control of these caverns are, therefore, crucial to ensure their capacity and operational safety. This work presents an integrated methodology for the analysis and optimization of the dissolution process for cavern development. First, numerical simulations of the solution mining process were performed in a sodium chloride rock formation, using the SALGAS SOLUTION MINING SIMULATION SOFTWARE for different circulation methods (direct and reverse). To interpret the complex interaction between input variables and results, a multivariate statistical analysis was conducted using dimensionality reduction techniques, Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE). The PCA revealed that the brine production rate and its correlated variables constitute the main component of process variability, while variables associated with temperature and fluid properties play a secondary but significant role. The t-SNE technique confirmed these findings and demonstrated its ability to group scenarios with similar operational characteristics. Based on the insights obtained, the work progresses from analysis to design, implementing an optimization framework that couples the SALGAS simulator with Genetic Algorithms (GA). Single-objective optimizations were conducted to maximize the final volume, and multi-objective optimizations were performed to explore the complex trade-offs between the objectives of Volume, Salt Mass, Energy Efficiency, and Construction Time. The results demonstrate that the GA is capable of identifying decision vectors that generate superior designs compared to the base scenarios, characterizing the Pareto frontier of optimal solutions. The proposed methodology, therefore, offers a systematic and robust approach for the design of salt caverns.

Keywords: cavern development; SALGAS; numerical simulations; statistical analyses; genetic algorithms.

LISTA DE FIGURAS

Figura 1 – Processo de mineração por solução: circulação direta	26
Figura 2 – Processo de mineração por solução: circulação reversa	27
Figura 3 – Captura das direções de máxima variância no Análise de Componentes Principais (<i>Principal Component Analysis</i>) (PCA)	30
Figura 4 – Fluxograma do algoritmo de Análise de Componentes Principais (PCA). . .	32
Figura 5 – Fluxo do algoritmo <i>t-distributed Stochastic Neighbor Embedding</i> (t-SNE). . .	35
Figura 6 – Fluxo de um Algoritmo Genético.	42
Figura 7 – Diagrama esquemático de formas de cavernas de sal resultantes de diferentes modos de circulação.	54
Figura 8 – Estrutura esquemática do poço.	55
Figura 9 – Esquema da matriz de dados de entrada do PCA.	61
Figura 10 – Fluxograma do processo de otimização acoplando o AG ao SALGAS.	73
Figura 11 – Comparação dos contornos finais da caverna para o caso de referência $T = 40^{\circ}\text{C}$ no caso D.	79
Figura 12 – Comparação dos contornos finais da caverna para o caso de referência $T = 60^{\circ}\text{C}$ no caso D.	80
Figura 13 – Comparação dos contornos finais da caverna para o caso de referência $T = 80^{\circ}\text{C}$ no caso D.	80
Figura 14 – Evolução do volume da caverna para o caso $T = 40^{\circ}\text{C}$ no caso D.	81
Figura 15 – Evolução do volume da caverna para o caso $T = 60^{\circ}\text{C}$ no caso D.	82
Figura 16 – Evolução do volume da caverna para o caso $T = 80^{\circ}\text{C}$ no caso D.	82
Figura 17 – Superfície de resposta para o volume no caso D.	83
Figura 18 – Comparação dos contornos finais da caverna para o caso de referência $T = 40^{\circ}\text{C}$ no caso R.	84
Figura 19 – Comparação dos contornos finais da caverna para o caso de referência $T = 60^{\circ}\text{C}$ no caso R.	85
Figura 20 – Comparação dos contornos finais da caverna para o caso de referência $T = 80^{\circ}\text{C}$ no caso R.	85
Figura 21 – Evolução do volume da caverna para o caso $T = 40^{\circ}\text{C}$ no caso R.	86
Figura 22 – Evolução do volume da caverna para o caso $T = 60^{\circ}\text{C}$ no caso R.	87

Figura 23 – Evolução do volume da caverna para o caso $T = 80^{\circ}\text{C}$ no caso R.	87
Figura 24 – Superfície de resposta para o volume no caso R.	88
Figura 25 – Diferença porcentual dos volumes entre os casos direto e reverso.	90
Figura 26 – Superfície de Resposta - volume comparado	91
Figura 27 – Matriz de correlação entre as variáveis (Correlograma).	94
Figura 28 – PCA - Autovalores (<i>Scree Plot</i>).	95
Figura 29 – PCA - Porcentagem da variância total explicada por componente.	96
Figura 30 – PCA - Contribuição das variáveis no primeiro componente	98
Figura 31 – PCA - Contribuição das variáveis no segundo componente	98
Figura 32 – PCA - Contribuição das variáveis nos cinco primeiros componentes.	99
Figura 33 – Círculo de correlação	101
Figura 34 – PCA - biplot	102
Figura 35 – Divergência vs. Perplexidade ($\pi = 5$)	103
Figura 36 – Divergência vs. Perplexidade ($\pi = 6$)	103
Figura 37 – t-SNE - Taxa de produção	104
Figura 38 – Volume	105
Figura 39 – Densidade específica da salmoura	106
Figura 40 – Densidade específica do fluido injetado	106
Figura 41 – Agrupamento t-SNE com coloração por caso de temperatura de referência	107
Figura 42 – Agrupamento t-SNE com coloração pela variável Fator de Dissolução da Sal.	108
Figura 43 – PCA 3D	109
Figura 44 – t-SNE 3D	109
Figura 45 – Evolução geométrica e perfil final da caverna otimizada.	114
Figura 46 – Verificação da estabilidade geomecânica da solução otimizada.	115
Figura 47 – Análise normalizada de desempenho comparativo (referência = 1.0).	116
Figura 48 – Fronteira de pareto 3D ilustrando o <i>trade-off</i> entre massa de sal, tempo e eficiência. O volume é constante para todas as soluções. As estrelas destacam as soluções notáveis.	118
Figura 49 – Análise de <i>trade-offs</i> entre os quatro objetivos da otimização.	119
Figura 50 – Perfil de desempenho normalizado da solução recomendada (#4).	121
Figura 51 – Subdivisão do triângulo	142
Figura 52 – Rede triangular irregular de vazões e temperaturas	143
Figura 53 – Pontos de dados de vazão, temperatura e volume	143

Figura 54 – Superfície de resposta triangulada para o volume da caverna em função da temperatura e da vazão.	145
Figura 55 – Esquema de um tubo com quatro canhoneiros a 90 graus.	150
Figura 56 – Fluxograma do processo de cálculo da pressão total em PSI.	151
Figura 57 – Guia calculadora de salmoura subsaturada no <i>Toolbox</i> do SMRI.	153
Figura 58 – Guia calculadora de salmoura saturada no <i>Toolbox</i> do SMRI.	153
Figura 59 – Executando o caso no SALGAS: Abriu-se o arquivo de entrada.	154
Figura 60 – Executando o caso no SALGAS: Foi dado o comando de partida.	155
Figura 61 – Executando o caso no SALGAS: Aguardou-se a conclusão da simulação. . .	155
Figura 62 – Executando o caso no SALGAS: Verificou-se se rodou até o final sem erro. .	156

LISTA DE TABELAS

Tabela 1 – Coeficientes para cálculo das funções A, B, C, D, E, F, G e H	48
Tabela 2 – Parâmetros da água pura.	49
Tabela 3 – Valores das constantes a_i e b	49
Tabela 4 – Cotas de Injeção e Produção para os cenários propostos.	53
Tabela 5 – Dimensões e profundidades dos componentes do poço.	55
Tabela 6 – Parâmetros de entrada para o Cenário D	56
Tabela 7 – Parâmetros de entrada para o Cenário R	57
Tabela 8 – Resultados da simulação (dados de saída) para o PCA do cenário D	58
Tabela 9 – Resultados da simulação (dados de saída, continuação) para o PCA do cenário D	59
Tabela 10 – Resultados da simulação (dados de saída) para o PCA do cenário R	60
Tabela 11 – Resultados da simulação (dados de saída, continuação) para o PCA do cenário R	61
Tabela 12 – Restrições geomecânicas consideradas na otimização.	68
Tabela 14 – Parâmetros de configuração do Algoritmo Genético (<i>solvers ga e gamultiobj</i>).	77
Tabela 15 – Diferença percentual dos volumes entre os casos Direto e Reverso.	89
Tabela 16 – Análise descritiva das variáveis	92
Tabela 17 – Correlação entre as variáveis.	93
Tabela 18 – PCA - Autovalores, porcentagem da variância total e a porcentagem acu- mulada da explicação da variância total.	95
Tabela 19 – Contribuição das variáveis nos componentes	97
Tabela 20 – Correlação entre as variáveis e os componentes.	100
Tabela 21 – Vetor de decisão ótimo da otimização mono-objetivo.	113
Tabela 22 – Métricas de desempenho da solução otimizada.	113
Tabela 23 – Características das soluções notáveis na fronteira de Pareto (valores nor- malizados).	120
Tabela 24 – Parâmetros operacionais (vetor de decisão) para a implementação da solu- ção #4.	122
Tabela 25 – Coeficientes para obter a superfície de resposta do volume da caverna ao final de 120 dias para o cenário com circulação direta.	140

Tabela 26 – Coeficientes para obter a superfície de resposta do volume da caverna ao
final de 120 dias para o cenário reverso. 141

LISTA DE ABREVIATURAS E SIGLAS

AG	Algoritmo Genético
BBO	Otimização <i>Black-Box</i>
CAES	Armazenamento de Energia por Ar Comprimido (<i>Compressed Air Energy Storage</i>)
CCS	Captura e Armazenamento de Carbono
CO ₂	dióxido de carbono
DF	Fator de Dissolução
MOEA	Algoritmo Evolutivo Multiobjetivo
MOO	Otimização Multiobjetivo
NSGA	<i>Non-dominated Sorting Genetic Algorithm</i>
NSGA-II	<i>Fast and Elitist Non-dominated Sorting Genetic Algorithm</i>
PCA	Análise de Componentes Principais (<i>Principal Component Analysis</i>)
SCES	Armazenamento Subterrâneo em Cavernas de Sal
SI	Sistema Internacional de Unidades
SMRI	<i>Solution Mining Research Institute</i>
SNE	<i>Stochastic Neighbor Embedding</i>
SVD	Decomposição em Valores Singulares (<i>Singular Value Decomposition</i>)
t-SNE	<i>t-distributed Stochastic Neighbor Embedding</i>

LISTA DE SÍMBOLOS

a_i, b_i	Coeficientes para o cálculo da densidade da água pura e da concentração de saturação
A, B, \dots, H	Funções dependentes da temperatura para o cálculo do volume específico
B_i	Profundidade do <i>blanket</i> (teto da caverna) no estágio i
c_b	Concentração de sal em massa na solução a ser injetada
c_b^{sat}	Concentração de sal em massa em uma salmoura saturada
C_s	Concentração de saturação de sal na solução
D	Espaço de soluções admissíveis (viáveis) na otimização
$d_{válvula}$	Diâmetro da válvula de injeção
$d_{tubo_interno}$	Diâmetro do tubo interno de injeção
$D_{KL}(P Q)$	Divergência de Kullback-Leibler entre as distribuições P e Q
E	Matriz de resíduos na decomposição do PCA
$f(H)$	<i>Fitness</i> (adequação) médio dos indivíduos pertencentes ao <i>schema</i> H
\bar{f}_t	<i>Fitness</i> (adequação) médio de toda a população na geração t
$F(\vec{x})$	Função-objetivo que quantifica o desempenho do vetor de decisão \vec{x}
g	Aceleração da gravidade
h_f	Perda de carga por atrito na equação de Bernoulli
H	<i>Schema</i> (padrão) em Algoritmos Genéticos
H_i	Profundidade da válvula de injeção
I_i	Profundidade de injeção da tubulação no estágio i
k	Coeficiente de transferência de massa
l	Comprimento total do cromossomo em Algoritmos Genéticos

$m(H, t)$	Número de indivíduos que correspondem ao <i>schema</i> H na geração t
\dot{m}_T	Taxa de dissolução da halita
n	Número de observações (amostras) na matriz de dados
n_p	Em NSGA-II, número de soluções que dominam a solução p
$o(H)$	Ordem de um <i>schema</i> (número de posições fixas)
p	Número de variáveis (características) na matriz de dados
P	Pressão em um ponto do fluido
p_c, p_m	Probabilidades de cruzamento e mutação em Algoritmos Genéticos
p_d	Probabilidade de disrupção de um <i>schema</i>
$p_{ij}, p_{j i}$	Probabilidades de vizinhança entre pontos no t-SNE (espaço de alta dimensão)
P	Matriz de autovetores (ou <i>loadings</i>) no PCA
P_i	Profundidade de produção da tubulação no estágio i
P^{ref}	Pressão de referência (1 atm)
q_{ij}	Probabilidade de vizinhança entre pontos no t-SNE (espaço de baixa dimensão)
Q_i	Vazão de injeção de salmoura no estágio i
R	Matriz de correlação
\mathbb{R}^{10}	Espaço Euclidiano real de 10 dimensões
S	Matriz de covariância
S_p	Em NSGA-II, conjunto de soluções que a solução p domina
t	Tempo
T	Temperatura
t_i	Duração do estágio i

\mathbf{T}	Matriz de <i>scores</i> (componentes principais) no PCA
v	Velocidade do fluido
V	Volume da solução
$VE(k)$	Variância explicada por k componentes principais
w_i	i -ésimo autovetor (vetor de pesos) da matriz de covariância
\vec{x}	Vetor de decisão do problema de otimização
\bar{x}	Vetor de médias das variáveis
\mathbf{X}	Matriz de dados original ($n \times p$)
$\tilde{\mathbf{X}}$	Matriz de dados centrados na média
\vec{x}^*	Vetor de decisão ótimo (solução da otimização)
y_i	Ponto no espaço de baixa dimensão (mapa) no t-SNE
z	Cota ou altura de um ponto em relação a um referencial
\mathbf{Z}	Matriz de projeções (dados em dimensão reduzida) no PCA
$\beta_0, \beta_i, \beta_{ij}$	Coefficientes do modelo de regressão polinomial
γ	Peso específico do fluido
$\delta(H)$	Comprimento de definição de um <i>schema</i>
ΔT	Diferença de temperatura em relação a uma referência ($T - T_0$)
λ_i	i -ésimo autovalor da matriz de covariância/correlação
π	Perplexidade, hiperparâmetro do t-SNE
ρ	Densidade (massa específica) de uma salmoura
ρ_b	Densidade da solução salina a ser injetada
ρ_w	Densidade da água pura
τ_i	Desvio padrão da distribuição Gaussiana no t-SNE

φ Símbolo genérico para as funções de temperatura A, B, ..., H

χ^2 Estatística do teste de esfericidade de Bartlett

SUMÁRIO

1	INTRODUÇÃO	20
1.1	OBJETIVO GERAL	22
1.2	OBJETIVOS ESPECÍFICOS	22
2	REFERENCIAL TEÓRICO	24
2.1	METODOLOGIAS DE DISSOLUÇÃO EM ROCHA SALINAS	24
2.2	REDUÇÃO DE DIMENSIONALIDADE	27
2.2.1	PCA	27
2.2.2	t-SNE	32
3	ALGORITMOS EVOLUTIVOS E OTIMIZAÇÃO MULTIOBJETIVO	36
3.1	ALGORITMOS GENÉTICOS (AGS)	36
3.2	A TEORIA DOS <i>SCHEMAS</i> E A HIPÓTESE DOS BLOCOS DE CONSTRUÇÃO	38
3.3	O BALANÇO CRÍTICO: EXPLORAÇÃO VERSUS <i>EXPLOITATION</i>	40
3.4	OTIMIZAÇÃO MULTIOBJETIVO E O ALGORITMO <i>Fast and Elitist Non-dominated Sorting Genetic Algorithm</i> (NSGA-II)	42
4	SOFTWARE SALGAS: CARACTERÍSTICAS E LIMITAÇÕES	45
4.1	CARACTERÍSTICAS DO <i>SOFTWARE</i> SALGAS	45
4.2	LIMITAÇÕES DO <i>SOFTWARE</i>	46
4.2.1	Descrição das equações de alguns parâmetros de entrada do SALGAS	47
4.2.2	Formulação Matemática do Modelo de Dissolução 1D	51
5	METODOLOGIA	53
5.1	CENÁRIOS DE SIMULAÇÃO E MATRIZ DE DADOS DE ENTRADA	53
5.1.1	Geração dos Parâmetros de Entrada Variáveis	55
5.2	DADOS DE ENTRADA PARA O SOFTWARE R	57
5.2.1	Análise de Componentes Principais com R	57
5.2.2	Análise t-SNE com R	62
5.3	METODOLOGIA DE OTIMIZAÇÃO DO DESIGN	63
5.3.1	Formulação Matemática do Problema	63
5.3.2	Vetor de Decisão	63
5.3.3	Sistema Hierárquico de Restrições e Espaço de Busca	64

5.3.4	Justificativa dos Parâmetros do Espaço de Busca	65
5.3.4.1	<i>Vazão de injeção (Q_i): 50 a 300 m³/h</i>	<i>65</i>
5.3.4.2	<i>Duração dos estágios (t_i): 180 a 720 dias</i>	<i>66</i>
5.3.4.3	<i>Profundidades operacionais (B_i, I_i, P_i): 609.6 a 914.4 m</i>	<i>66</i>
5.3.4.4	<i>Abordagem de modelagem da temperatura</i>	<i>66</i>
5.3.5	Restrições Não-Lineares (Coerência Física e Segurança Geomecânica) 67	
5.3.6	Função-Objetivo e Métricas de Desempenho	69
5.3.6.1	<i>Formulação para Otimização Mono-Objetivo (ga)</i>	<i>70</i>
5.3.6.2	<i>Formulação para Otimização Multiobjetivo (NSGA-II/gamultiobj)</i>	<i>71</i>
5.4	IMPLEMENTAÇÃO COMPUTACIONAL E ARQUITETURA DO OTIMI- ZADOR	72
5.4.1	O <i>framework</i> de Otimização: Acoplando SALGAS e AG	72
5.4.2	Estratégia de Aceleração Computacional: Modelos Substitutos para Propriedades de Fluidos	75
5.4.3	Configuração dos Operadores Genéticos no Ambiente MATLAB . .	76
6	RESULTADOS E DISCUSSÃO	78
6.1	RESULTADOS COM SALGAS	78
6.2	RESULTADOS COM PCA	91
6.2.1	Resultados com t-SNE	102
7	OTIMIZAÇÃO DO DESIGN DE CAVERNAS SALINAS VIA AL- GORITMOS GENÉTICOS	110
7.1	ANÁLISE DOS RESULTADOS DA OTIMIZAÇÃO	111
7.1.1	Otimização Mono-Objetivo: Maximização do Volume Final	112
7.1.2	Otimização multiobjetivo: Análise da fronteira de Pareto	117
8	CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FU- TUROS	123
8.1	RECOMENDAÇÕES PARA TRABALHOS FUTUROS	124
	REFERÊNCIAS	125
	APÊNDICE A – ANÁLISE DE COMPONENTES PRINCIPAIS (PCA): ABORDAGEM MATEMÁTICA	133
	APÊNDICE B – COEFICIENTES DOS POLINÔMIOS QUE GE- RAM A SUPERFÍCIE DE RESPOSTA E SUA DE- RIVAÇÃO	140

APÊNDICE C – CÁLCULO DE PRESSÕES COM BASE NA EQUAÇÃO DE BERNOULLI	146
APÊNDICE D – CÁLCULO DA PRESSÃO DE INJEÇÃO NO FUNDO DO POÇO	148
APÊNDICE E – GERAÇÃO DE DADOS DE ENTRADA E EXECUÇÃO DA SIMULAÇÃO	152
Apêndice E – Descrição do Arquivo de Entrada (.sgi) do SALGAS	157

1 INTRODUÇÃO

Segundo Aydin (2014), em 2010, cerca de 90% do consumo mundial de energia primária foi fornecido por combustíveis fósseis. Esse cenário destaca a importância de se explorar alternativas para o armazenamento eficiente de energia. De acordo com Yang *et al.* (2015), o armazenamento em cavernas de sal possui um papel relevante nas reservas internacionais de energia. Estudos como o de Ozarslan (2012) indicam que as cavernas de sal subterrâneas artificiais são utilizadas como reservatórios de ar comprimido e gás hidrogênio, sendo candidatas promissoras para aplicações de armazenamento de energia em larga escala. No entanto, conforme destacado em Małachowska *et al.* (2022), a implementação dessas tecnologias pode ser impactada por regulamentações de segurança, especialmente na Europa, onde as normas variam significativamente.

Esse cenário ressalta a relevância de discutir, de forma mais ampla, os desafios relacionados à segurança energética e à proteção ambiental, que têm se consolidado como questões urgentes para a sociedade contemporânea Sovacool *et al.* (2020), Bradshaw *et al.* (2021), Ritchie e Roser (2020). Embora as fontes de energia tradicionais, como o gás natural e o petróleo, ainda desempenhem um papel central no crescimento econômico global, sua dependência excessiva resulta em altas emissões de dióxido de carbono (dióxido de carbono (CO_2)), agravando problemas ambientais cruciais, como o aquecimento global linemiller2017. Além disso, a produção desses recursos frequentemente ocorre em regiões distantes dos centros de consumo, gerando um descompasso entre oferta e demanda. Para mitigar esses impactos, soluções de armazenamento de energia, como os reservatórios subterrâneos para gás Succar e Williams (2015), petróleo Nadimi, Zali e Ahangari (2018) e sistemas de armazenamento de energia elétrica Zame *et al.* (2018), Luo *et al.* (2015), têm se tornado vitais para garantir a estabilidade e a resiliência das cadeias de suprimento.

A transição para fontes renováveis de energia, como solar, eólica e hidrelétrica, é essencial para a redução das emissões de carbono Acar e Dincer (2019). Contudo, a intermitência dessas fontes exige soluções de armazenamento em larga escala para equilibrar a rede elétrica, como sistemas de Armazenamento de Energia por Ar Comprimido (*Compressed Air Energy Storage*) (CAES) Budt *et al.* (2016), hidrogênio Abe *et al.* (2019) e hidrelétricas reversíveis Deane, GallachÓir e McKeogh (2010). Essas tecnologias não só facilitam a integração de energias renováveis, mas também são fundamentais para o cumprimento de metas globais, como a

neutralidade de carbono até 2060, estabelecida pela China Mallapaty (2020).

Neste contexto, o Armazenamento Subterrâneo em Cavernas de Sal (SCES) surge como uma solução promissora devido às propriedades excepcionais do sal-gema: baixa permeabilidade ($\leq 10^{-21} \text{ m}^2$), capacidade de autorregeneração e estabilidade mecânica Yu *et al.* (2016), Zhang *et al.* (2018), Fan *et al.* (2017). Cavernas formadas por dissolução controlada (método de *solution mining*) proporcionam espaços herméticos e duráveis, adequados para armazenar gás natural, hidrogênio, ar comprimido e até CO₂ Cypryański *et al.* (2020). Iniciativas pioneiras, como o armazenamento de hidrogênio em Tees-side (Reino Unido, 1972) Walters (1976) e a primeira planta comercial de CAES em Huntorf (Alemanha, 1978) Crotogino, Mohmeyer e Scharf (2001), demonstram a viabilidade técnica dessa abordagem.

No entanto, apesar do potencial, persistem desafios significativos. A falta de revisões sistemáticas sobre o SCES em escala global limita a compreensão de seu papel estratégico no setor energético Fan *et al.* (2021). Além disso, a expansão de tecnologias como CAES e o armazenamento de hidrogênio exigem um planejamento industrial integrado, avanços em geomecânica e a gestão de cavernas em formações salinas estratificadas, características comuns em países como a China e o Reino Unido Li *et al.* (2020), Bérest *et al.* (2022).

Contudo, a exploração de domos salinos não é isenta de riscos geomecânicos significativos, cujo gerenciamento inadequado pode levar a consequências catastróficas. O desastre ambiental em Maceió, Brasil, onde a extração de sal-gema por décadas resultou na desestabilização do subsolo e na subsidência severa de bairros inteiros, serve como um alerta contundente Serviço Geológico do Brasil (CPRM) (2019). Este evento ressalta que o *design* e a operação de cavernas salinas exigem um planejamento rigoroso, que vá além da simples viabilidade econômica. Neste contexto, a otimização computacional avançada é fundamental. Ferramentas como os Algoritmos Genéticos (AG) são empregadas na engenharia moderna para determinar a geometria e o espaçamento ótimos de cavernas, com o objetivo explícito de maximizar a estabilidade de longo prazo e minimizar os riscos de colapso e afundamento da superfície. Portanto, a aplicação de um AG, como proposto nesta tese, não é apenas uma busca por eficiência, mas uma metodologia robusta e indispensável para garantir a segurança e a sustentabilidade de tais projetos.

A partir da análise das propriedades e desafios mencionados, torna-se evidente a necessidade de aprimorar o entendimento sobre o processo de construção de cavernas de sal e seu potencial como solução de armazenamento de energia. Nesse sentido, a modelagem numérica se apresenta como uma ferramenta essencial, uma vez que permite a simulação de cenários e a

otimização do *design* de cavernas subterrâneas. No entanto, as abordagens tradicionais, como experimentos laboratoriais, apresentam limitações, principalmente no que se refere à escala e à precisão dos resultados. Portanto, a aplicação de técnicas de análise estatística multivariada, como a PCA e o t-SNE, surge como uma estratégia para melhorar a interpretação dos dados obtidos a partir das simulações e otimizar o processo de construção de cavernas de sal. Essas técnicas, amplamente utilizadas em áreas como a análise de dados e aprendizado de máquina, oferecem a vantagem de reduzir a dimensionalidade dos dados e identificar padrões ocultos, o que pode facilitar a tomada de decisões e o planejamento do desenvolvimento de cavernas para armazenamento de energia.

1.1 OBJETIVO GERAL

Desenvolver e aplicar uma metodologia integrada de análise e otimização para o processo de abertura de cavernas subterrâneas, utilizando técnicas de redução de dimensionalidade (PCA e t-SNE) para a compreensão do sistema e Algoritmos Genéticos para a determinação de *designs* ótimos.

1.2 OBJETIVOS ESPECÍFICOS

Para a consecução do objetivo principal, foram propostos os seguintes objetivos específicos:

- Avaliar a evolução geométrica da caverna ao longo do tempo de simulação, identificando mudanças significativas na forma e na estrutura.
- Comparar os dados gerados pelo *software* SALGAS em diferentes cenários de simulação (direto e reverso), analisando padrões e discrepâncias que possam impactar o processo de abertura da caverna.
- Reduzir a dimensionalidade dos dados numéricos de simulação, preservando a maior parte das informações relevantes para facilitar a análise dos resultados, com foco nas variáveis que influenciam o desenvolvimento da caverna.
- Analisar a robustez dos agrupamentos obtidos através do t-SNE, testando diferentes valores de perplexidade e aplicando técnicas de validação cruzada para verificar a consistência e a qualidade dos grupos formados.

- Interpretar os resultados gerados a partir das representações em baixa dimensão obtidas pelas duas técnicas de redução de dimensionalidade, destacando *insights* significativos sobre o comportamento das variáveis.
- Contrastar as duas técnicas de redução de dimensionalidade com base no agrupamento dos indivíduos (cavernas simuladas), avaliando suas respectivas eficácias na visualização dos dados.
- Quantificar a influência das variáveis taxa de produção de salmoura e temperatura interna equivalente (ajustada pelo Fator de Dissolução do SALGAS) no processo de abertura da caverna por dissolução, utilizando métodos estatísticos apropriados.
- Formular o problema de *design* da caverna como um problema de otimização matemática, definindo o vetor de decisão, as funções-objetivo e as restrições físicas e operacionais.
- Implementar um *framework* computacional que acople o simulador SALGAS a um Algoritmo Genético para solucionar o problema de otimização.
- Analisar os resultados da otimização, identificando o *design* ótimo para o caso mono-objetivo e caracterizando a fronteira de soluções ótimas para o caso multiobjetivo.

2 REFERENCIAL TEÓRICO

Neste capítulo, são apresentados os conceitos ligados ao tema da pesquisa, tais como metodologias de dissolução em rochas salinas, formulação matemática das equações do *software* de SALGAS, e a formulação matemática das duas técnicas da redução de dimensionalidade, PCA e t-SNE.

2.1 METODOLOGIAS DE DISSOLUÇÃO EM ROCHA SALINAS

A mineração por solução não é uma prática recente, tendo sido utilizada há milênios na China, onde o sal subterrâneo era dissolvido com água doce por meio de tubos de bambu. Com o tempo, esse processo evoluiu para uma indústria especializada na extração de minerais solúveis em água. Em muitos casos, as cavernas de sal resultantes tornaram-se mais valiosas do que o próprio sal extraído, sendo amplamente empregadas para o armazenamento de gás natural, petróleo bruto e outros hidrocarbonetos, além de, em algumas situações, servirem como reservatórios para resíduos. O crescimento dessa indústria pode ser observado no aumento do número de membros do *Solution Mining Research Institute* (SMRI), impulsionado pelo crescente interesse e investimento em armazenamento de gás.

Cavernas dissolvidas em formações de sal oferecem um método seguro e amplamente aceito para o armazenamento de gás natural. No Brasil, estudos indicam a viabilidade de utilizar cavernas em águas ultraprofundas para essa finalidade, especialmente nas formações do pré-sal, visando tanto a Captura e Armazenamento de Carbono (CCS) quanto a monetização do gás natural associado Costa *et al.* (2020), Costa *et al.* (2015).

O armazenamento de energia em cavernas de sal tem ganhado destaque devido à sua elevada capacidade, segurança, conformidade ambiental e viabilidade econômica. A estabilidade dos depósitos de sal e sua capacidade de isolar fluidos, como gás e ar comprimido, fazem dessas estruturas uma alternativa robusta para aplicações em larga escala Tian *et al.* (2010). Contudo, há riscos associados, sendo a subsidência superficial um dos mais críticos. Estudos realizados na Suíça apontam que áreas afetadas podem variar de 100 a 1.500 metros de diâmetro, com taxas de subsidência superiores a 100 mm/ano Zechner *et al.* (2011). Nos Estados Unidos, casos como o poço Hendrick 10-A demonstram que a circulação de água através de falhas geológicas pode provocar dissolução do sal e resultar em subsidência em larga escala

Johnson (2005).

A construção de cavernas de sal exige requisitos básicos, como a presença de um depósito salino ou de gesso, água não saturada, um mecanismo para extração da salmoura e energia para manter o fluxo hídrico Johnson (2005). Diferentes arranjos de poços são utilizados, incluindo configurações de poço único, dois poços para injeção e produção, e arranjos combinados que interligam duas cavidades por um único poço Tian *et al.* (2010). Os métodos de dissolução podem seguir diferentes abordagens, como “fundo para topo”, “topo para fundo” ou métodos integrais. A circulação de água pode ser direta, com injeção no fundo e extração no topo, ou reversa, com injeção no topo e extração no fundo, resultando em distintas geometrias de cavernas.

Estudos experimentais têm contribuído para a compreensão dos processos envolvidos. Weisbrod *et al.* (2012) demonstraram, por meio de tomografia computadorizada, a formação de canais preferenciais em rochas salinas, influenciados pela microestrutura do material. Liu *et al.* (2016), com testes dinâmicos e modelagem numérica, identificaram uma relação linear entre o parâmetro (dissolução/espessura da camada limite) e a velocidade do fluxo, além de destacarem a influência gravitacional na morfologia das cavernas.

O desenvolvimento de modelos matemáticos tem permitido prever os processos de dissolução e a evolução das cavernas. Modelos iniciais, como os de Durie e Jessen (1964), relacionaram a salinidade da água com a taxa de dissolução, enquanto Donat e Haimson (1974) e Saberian (1974) avançaram em simulações numéricas para controle de pressão e evolução das cavernas. O modelo UBRO, de Kunstman e Urbanczyk (1990), incorporou previsões sobre forma, volume e concentração de salmoura, considerando também a sedimentação de insolúveis. Modelos mais recentes Huang *et al.* (2011), Li *et al.* (2016), Li *et al.* (2018), Wang *et al.* (2018), Wang *et al.* (2018), Yang *et al.* (2017), Yang, Liu e Zhang (2017) aprimoraram essas abordagens, incluindo efeitos acoplados de fluxo, dissolução e sedimentação, essenciais para otimizar os processos e garantir a segurança operacional.

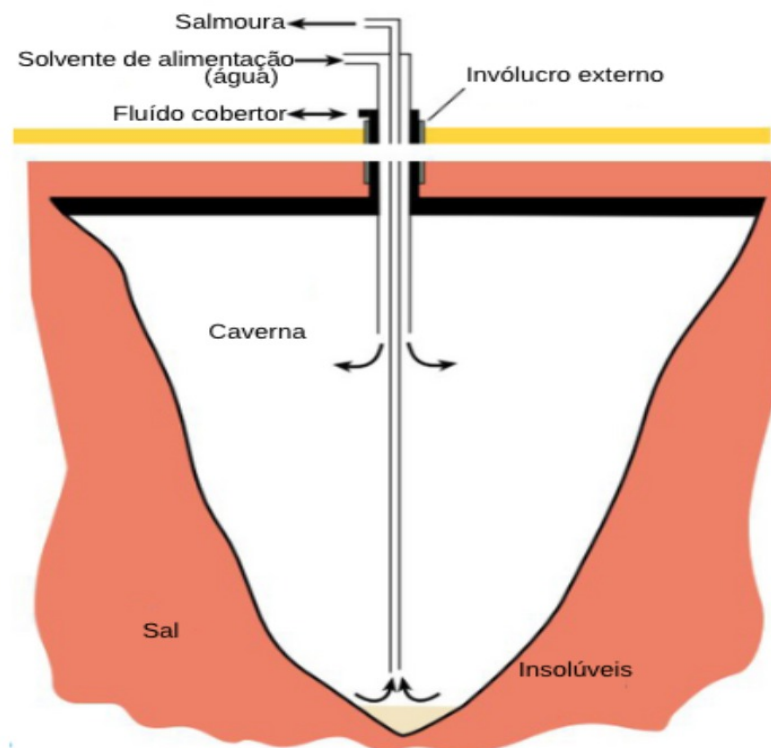
Apesar dos avanços tecnológicos, desafios persistem na previsão da subsidência, no controle da morfologia das cavernas e na otimização de processos como a dessalinização. O progresso em modelagem computacional tem sido crucial para superar essas dificuldades, permitindo análises mais precisas e melhorias contínuas nos sistemas de armazenamento de energia em cavernas de sal.

Nesse contexto, compreender os processos físicos envolvidos na formação das cavernas é essencial para aprimorar as técnicas de construção e garantir sua estabilidade a longo prazo.

A mineração por solução, principal método utilizado para a criação dessas cavernas, é um processo de transferência de massa em que ocorre a convecção entre a água e a rocha salina, promovendo a dissolução do sal Steding *et al.* (2021).

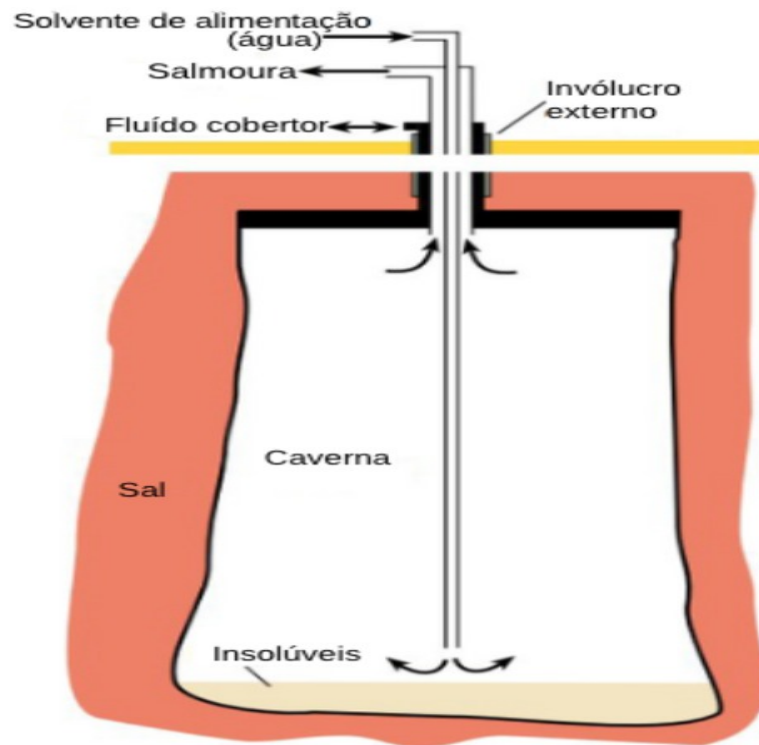
Existem diferentes métodos para o desenvolvimento e conformação das cavernas. No método de circulação direta na Figura 1, o solvente é injetado através da coluna de tubulação e a salmoura é extraída pelo espaço anular entre a coluna e o revestimento final. No método de circulação reversa Figura 2, o solvente entra pelo anel e a salmoura é retirada pela coluna de tubulação. Devido à menor densidade da água doce comparada à salmoura, cavernas em forma de "cone invertido", com topo mais largo que a base, tendem a se formar na circulação reversa, especialmente na ausência de um fluido cobertor. Já a circulação direta resulta em cavernas mais cilíndricas. Durante a escavação, o volume do fluido cobertor pode ser ajustado para auxiliar na conformação da cavidade.

Figura 1 – Processo de mineração por solução: circulação direta



Fonte: Adaptado Warren (2016)

Figura 2 – Processo de mineração por solução: circulação reversa



Fonte: Adaptado Warren (2016)

2.2 REDUÇÃO DE DIMENSIONALIDADE

Neste trabalho, a redução de dimensionalidade é aplicada como uma ferramenta para facilitar a interpretação dos resultados obtidos nas simulações numéricas do processo de abertura de cavernas. O objetivo é identificar quais variáveis mais influenciam o comportamento geométrico e dinâmico das cavernas simuladas, a partir de representações compactas que preservam as informações mais relevantes dos dados originais.

Nesta secção são abordadas duas técnicas de redução de dimensionalidade: PCA e t-SNE, ambas as técnicas permitem a visualização de dados, mas apresentam diferenças e uma das diferenças significativas é que o PCA é uma técnica linear, enquanto o t-SNE é uma técnica não linear. Descrevem-se ambas as técnicas.

2.2.1 PCA

O desenvolvimento da PCA começa com fundamentos matemáticos em álgebra linear e decomposição matricial, evoluindo para um método estatístico de redução de dimensionalidade.

No final do século XIX, os trabalhos de Beltrami (1873) e Jordan (1874) estabeleceram as bases matemáticas ao investigar funções e formas bilineares. Embora esses estudos não tivessem, à época, a intenção de reduzir a dimensionalidade dos dados, eles introduziram conceitos que, posteriormente, dariam origem à Decomposição em Valores Singulares (*Singular Value Decomposition*) (SVD), uma ferramenta importante para a construção teórica do PCA.

No início do século XX, o foco passou a incluir abordagens geométricas e estatísticas. Em 1901, Pearson (1901) apresentou uma metodologia para ajustar linhas e planos a conjuntos de pontos em espaços multidimensionais, o que serviu de precursor direto da PCA. Essa fase também viu a aplicação prática da técnica: em 1923, Fisher e Mackenzie (1923) utilizaram métodos semelhantes à SVD para estudar variações em culturas agrícolas, e em 1929, Frisch (1929) aprofundou a análise da correlação e dispersão entre variáveis.

A formalização da PCA como método estatístico ocorreu na década de 1930. Em 1933, Hotelling (1933) definiu os “componentes principais” e diferenciou a técnica da análise fatorial, estabelecendo um marco teórico decisivo. Nos anos seguintes, Girshick (1936) introduziu estimativas de máxima verossimilhança para os componentes, enquanto Hotelling (1936), novamente em 1936, propôs cálculos simplificados que facilitaram a aplicação prática da metodologia. Em 1939, Girshick (1939) expandiu o conhecimento sobre as distribuições amostrais associadas às raízes das equações determinantes.

Nas décadas de 1960 a 1980, a PCA ganhou novas dimensões teóricas e aplicações interdisciplinares. Anderson (1963) contribuiu com a teoria assintótica, e Rao (1964) ampliou as interpretações e conexões da técnica com outras abordagens de análise multivariada. Adicionalmente, os estudos de Gower (1966) e Jeffers (1967) evidenciaram a versatilidade da PCA, culminando na compilação de métodos multivariados por Bryant e Atchley (1975). Por fim, o trabalho de Preisendorfer e Mobley (1988) sintetizou tanto os fundamentos teóricos quanto as aplicações práticas da PCA em áreas como a meteorologia e a oceanografia.

Com os fundamentos teóricos consolidados ao longo de décadas, o PCA se apresenta como uma técnica robusta para a análise de dados. Na prática, seu objetivo é transformar os dados originais em um novo sistema de coordenadas — os componentes principais — que são calculados a partir da matriz de covariância dos dados. Estes componentes são orientados de forma a preservar a maior variabilidade possível, permitindo uma representação dos dados em um espaço de menor dimensionalidade sem comprometer informações relevantes.

1. Centralização dos Dados e Matriz de Covariância

Para iniciar o processo de PCA, o primeiro passo é centralizar os dados. Seja $\mathbf{X} \in \mathbb{R}^{n \times p}$ uma matriz de dados com n observações e p variáveis, onde cada linha \mathbf{x}_i representa a i -ésima observação. A matriz \mathbf{X} pode ser representada como:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

Centralização: O processo de centralização envolve subtrair o vetor de médias $\bar{\mathbf{x}}$ de cada observação. O vetor de médias $\bar{\mathbf{x}}$, de dimensão $1 \times p$, é calculado como:

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

A centralização resulta em uma nova matriz de dados $\tilde{\mathbf{X}}$, dada por:

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}_n \bar{\mathbf{x}}$$

Nesta equação, $\mathbf{1}_n$ é um vetor coluna de dimensão $n \times 1$ composto por n elementos iguais a 1. A multiplicação de $\mathbf{1}_n$ pelo vetor linha $\bar{\mathbf{x}}$ cria uma matriz $n \times p$ onde cada linha é o vetor de médias, permitindo a subtração elemento a elemento.

Em seguida, calculamos a **Matriz de Covariância S**, que descreve como as variáveis do conjunto de dados estão relacionadas entre si:

$$\mathbf{S} = \frac{1}{n-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \in \mathbb{R}^{p \times p}$$

$$\mathbf{S} = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1p} \\ S_{21} & S_{22} & \dots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \dots & S_{pp} \end{bmatrix}$$

2. Maximização da Variância via Autovalores e Autovetores

O próximo passo no PCA é identificar as direções de máxima variabilidade nos dados. Isso é feito através da obtenção dos autovetores e autovalores da matriz de covariância dos dados. Essas direções são aquelas nas quais a variância dos dados projetados é maximizada como exemplificado na figura 3

Matematicamente, a maximização da variância nas projeções é descrita pela seguinte equação:

$$\mathbf{S}\mathbf{w}_i = \lambda_i \mathbf{w}_i$$

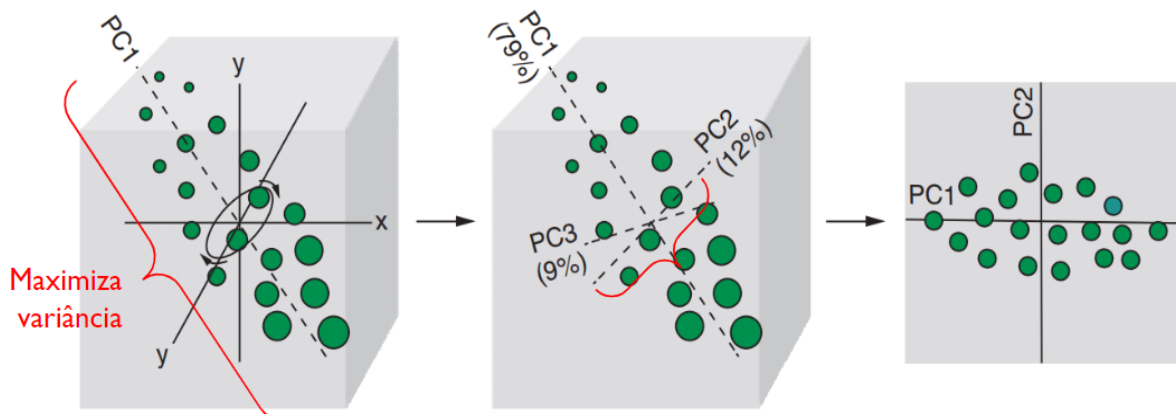
Onde:

- $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ são os autovalores da matriz de covariância, que indicam a quantidade de variância explicada por cada componente principal.
- $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p$ são os autovetores, que são ortogonais entre si. Organizando esses autovetores em uma matriz \mathbf{W} , obtemos a matriz de projeção dos dados no espaço de componentes principais:

$$\mathbf{W} = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_p \end{bmatrix} \in \mathbb{R}^{p \times p}$$

As direções dos componentes principais \mathbf{w}_i correspondem às direções nos dados onde ocorre a maior variabilidade.

Figura 3 – Captura das direções de máxima variância no PCA



Fonte: Modificado Rhys (2020)

3. Seleção dos k Componentes Principais

Após calcular os autovetores e autovalores, selecionam-se os k primeiros autovetores, correspondentes aos maiores autovalores λ_i , para formar a matriz de projeção \mathbf{W}_k :

$$\mathbf{W}_k = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_k \end{bmatrix} \in \mathbb{R}^{p \times k}$$

A matriz \mathbf{W}_k é usada para projetar os dados originais em um espaço de menor dimensionalidade. A variância explicada pelos k primeiros componentes principais $VE(k)$ é dada por:

$$VE(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \times 100\%$$

Essa métrica indica a proporção da variância total que é retida pelos k componentes principais selecionados.

4. Projeção dos Dados

Por fim, a projeção dos dados originais no novo espaço de componentes principais é feita multiplicando a matriz centrada $\tilde{\mathbf{X}}$ pela matriz de projeção \mathbf{W}_k , resultando em uma nova matriz \mathbf{Z} , de dimensão reduzida $n \times k$:

$$\mathbf{Z} = \tilde{\mathbf{X}}\mathbf{W}_k \in \mathbb{R}^{n \times k}$$

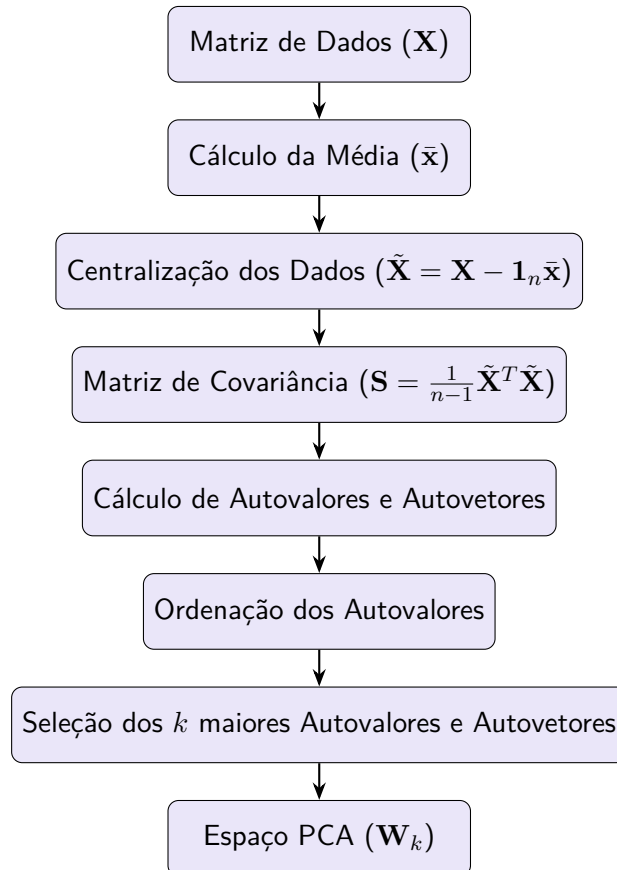
A matriz \mathbf{Z} é composta pelas projeções dos dados nos k primeiros componentes principais:

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & \dots & z_{1k} \\ z_{21} & z_{22} & \dots & z_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \dots & z_{nk} \end{bmatrix}$$

O PCA, portanto, permite não só reduzir a dimensionalidade dos dados, mas também realizar uma análise das componentes que mais contribuem para a variabilidade, facilitando a interpretação e visualização dos dados em espaços de menor dimensão.

A análise matemática completa e rigorosa do PCA, que detalha todos os aspectos da metodologia, encontra-se no Apêndice A. Em seguida, apresenta-se o diagrama de fluxo da figura 4, que organiza as etapas do processo no cálculo do PCA.

Figura 4 – Fluxograma do algoritmo de Análise de Componentes Principais (PCA).



Fonte: O autor (2025)

2.2.2 t-SNE

Segundo Hinton e Roweis (2002), a técnica *Stochastic Neighbor Embedding* (SNE) (*Stochastic Neighbor Embedding*) é um método probabilístico que visa representar objetos de um espaço de alta dimensão em um espaço de baixa dimensão (geralmente 2D ou 3D), preservando a estrutura de vizinhança local dos dados. A ideia central do algoritmo é minimizar uma função de custo que quantifica a divergência entre as distribuições de probabilidade que representam as similaridades entre os pontos nos dois espaços. No entanto, a formulação original do SNE apresentava um "problema de aglomeração" (*crowding problem*), onde os dados projetados em dimensões menores tendiam a se aglomerar excessivamente.

Para solucionar essa limitação, Maaten e Hinton (2008) propuseram a técnica t-SNE, uma

evolução do método original. Conforme descrito em seu artigo "Visualizando dados no t-SNE", a nova abordagem introduz duas modificações cruciais: (1) utiliza uma versão simétrica da função de custo, simplificando o gradiente para a otimização; e (2) substitui a distribuição Gaussiana por uma distribuição t-Student no espaço de baixa dimensão. O uso de uma distribuição de cauda pesada, como a t-Student, permite que pontos moderadamente distantes no espaço original sejam mapeados para distâncias maiores no espaço de baixa dimensão, aliviando tanto o problema de aglomeração quanto otimizando o processo de visualização.

O processo do t-SNE pode ser dividido em três etapas principais. Primeiramente, constrói-se uma distribuição de probabilidade sobre os pares de pontos no espaço de alta dimensão, de tal forma que pontos similares tenham uma alta probabilidade de serem escolhidos um pelo outro. Para um conjunto de dados $\{\mathbf{x}_i\}_{1 \leq i \leq n}$, a similaridade do ponto \mathbf{x}_j em relação a \mathbf{x}_i é modelada pela probabilidade condicional $p_{j|i}$, calculada sob uma distribuição Gaussiana centrada em \mathbf{x}_i :

$$p_{j|i} = \frac{\exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\tau_i^2\right)}{\sum_{k \neq i} \exp\left(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\tau_i^2\right)}$$

A variância da Gaussiana, τ_i^2 , é determinada com base em um hiper-parâmetro fundamental definido pelo usuário: a perplexidade, π . A perplexidade pode ser interpretada como um ajuste suave para o número de vizinhos próximos que cada ponto considera. Um valor mais alto de perplexidade leva em conta mais vizinhos, focando na estrutura global dos dados, enquanto um valor mais baixo foca em aspectos locais. O valor de τ_i para cada ponto é encontrado numericamente resolvendo a seguinte equação, onde $H(P_i)$ é a entropia de Shannon:

$$\pi = 2^{H(P_i)} = 2^{-\sum_j p_{j|i} \log_2 p_{j|i}} \quad (2.1)$$

Para a probabilidade conjunta, o t-SNE utiliza uma versão simetrizada: $p_{ij} = (p_{j|i} + p_{i|j})/2n$. A matriz simétrica resultante, \mathbf{P} , representa as similaridades no espaço de alta dimensão.

A segunda etapa consiste em definir uma distribuição de probabilidade similar, \mathbf{Q} , para os pontos no espaço de baixa dimensão, $\{\mathbf{y}_i\}_{1 \leq i \leq n}$. Aqui reside a principal inovação do t-SNE: em vez de uma Gaussiana, utiliza-se uma distribuição t-Student com um grau de liberdade, que possui caudas mais pesadas. A probabilidade conjunta q_{ij} é definida como:

$$q_{ij} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}}$$

Na terceira e última etapa, o objetivo do t-SNE é encontrar um mapeamento de baixa dimensão $\{\mathbf{y}_i\}_{1 \leq i \leq n}$ que minimize a divergência entre as duas distribuições de probabilidade, \mathbf{P} e \mathbf{Q} . Essa minimização é realizada sobre a divergência de Kullback-Leibler (KL), que atua como a função de custo do algoritmo, conforme a equação 2.2.

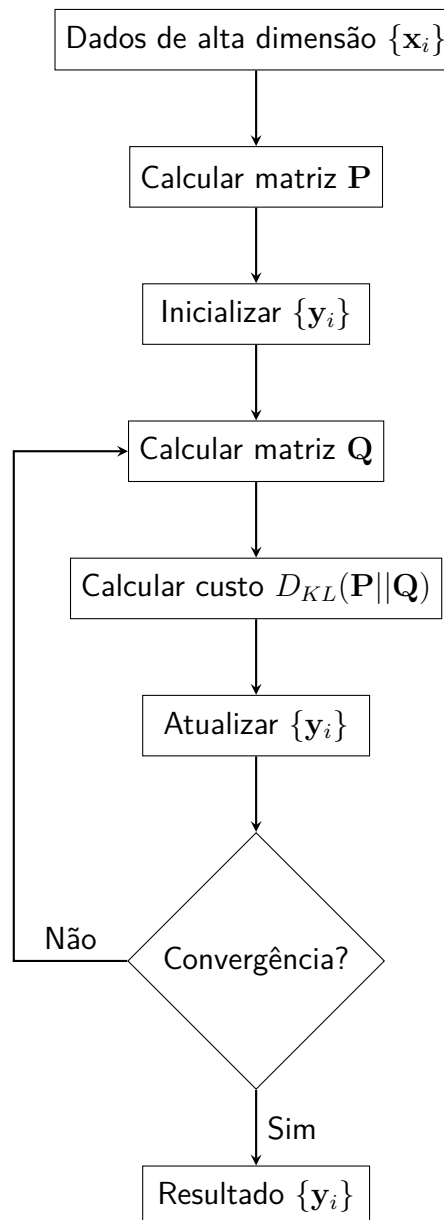
$$(\mathbf{y}_1, \dots, \mathbf{y}_n) = \arg \min_{\mathbf{y}_1, \dots, \mathbf{y}_n} D_{KL}(\mathbf{P} \parallel \mathbf{Q}) = \arg \min_{\mathbf{y}_1, \dots, \mathbf{y}_n} \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (2.2)$$

Para minimizar a função de custo, o t-SNE inicia com uma configuração aleatória dos pontos \mathbf{y}_i e os atualiza iterativamente usando o método de gradiente descendente. O gradiente da divergência de KL é dado por:

$$\frac{\delta C}{\delta \mathbf{y}_i} = 4 \sum_j (p_{ij} - q_{ij}) (\mathbf{y}_i - \mathbf{y}_j) (1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1} \quad (2.3)$$

Em uma pesquisa posterior, Maaten (2014) demonstrou que o gradiente pode ser aproximado eficientemente usando uma variante do algoritmo de Barnes-Hut, o que acelerou significativamente a aplicação do t-SNE a grandes conjuntos de dados. Na sequência, o fluxograma da figura 5 resume o processo de cálculo do t-SNE.

Figura 5 – Fluxo do algoritmo t-SNE.



Fonte: O autor (2025)

3 ALGORITMOS EVOLUTIVOS E OTIMIZAÇÃO MULTIOBJETIVO

A busca por soluções ótimas para problemas complexos é um desafio central em diversas áreas da ciência e engenharia. Frequentemente, os modelos matemáticos que representam sistemas do mundo real são caracterizados por não linearidades, descontinuidades e um vasto espaço de busca, tornando os métodos de otimização clássicos, baseados em gradientes, ineficazes ou inviáveis. Neste contexto, os algoritmos evolutivos (AEs) surgem como uma classe de técnicas de busca estocástica, inspiradas nos princípios da evolução natural, para navegar por esses espaços complexos. Adicionalmente, muitos problemas práticos não se resumem a otimizar um único critério, mas sim a encontrar um balanço entre múltiplos objetivos, muitas vezes conflitantes entre si, como custo e eficiência, ou risco e retorno. Esta necessidade define o campo da otimização multiobjetivo. Este capítulo se dedica a explorar os fundamentos dos algoritmos evolutivos e sua aplicação na resolução de problemas multiobjetivo.

3.1 ALGORITMOS GENÉTICOS (AGS)

Os AGs representam uma classe de algoritmos de busca heurística adaptativa, cuja concepção se inspira diretamente nos princípios da seleção natural e da genética mendeliana. Eles pertencem a uma família mais ampla de métodos computacionais conhecidos como Computação Evolutiva. A principal virtude dos AGs reside na sua capacidade de explorar espaços de busca complexos, multidimensionais e multimodais para encontrar soluções de alta qualidade para problemas de otimização e busca, sem a necessidade de informações sobre o gradiente da função-objetivo. Esta característica os torna particularmente adequados para problemas em que a função-objetivo é descontínua, não diferenciável, estocástica ou altamente não linear, cenários comuns em desafios de engenharia do mundo real Sreekanth e Kumar (2019).

A base teórica dos Algoritmo Genético (AG)s foi estabelecida na obra seminal de Holland (1975). Em seu trabalho, Holland não apenas propôs os mecanismos fundamentais do algoritmo, mas também forneceu um arcabouço matemático para entender como a adaptação ocorre em sistemas complexos, sejam eles naturais ou artificiais. Ele introduziu um modelo matemático capaz de lidar com a não linearidade de interações complexas, demonstrando sua universalidade em campos tão diversos quanto a biologia, a economia e a inteligência artificial. A obra de Holland foi pioneira, estabelecendo as fundações teóricas que permitiram o

desenvolvimento subsequente do campo.

Enquanto Holland estabeleceu as fundações teóricas, foi David E. Goldberg quem catalisou a aplicação prática e a ampla adoção dos AGs com seu livro seminal, "*Genetic Algorithms in Search, Optimization, and Machine Learning*" Goldberg (1989). Goldberg traduziu os conceitos teóricos de Holland em uma metodologia acessível e de aplicação direta, com um tom informal e tutorial que desmistificou o tema para engenheiros e cientistas da computação. Sua obra solidificou os AGs como uma ferramenta robusta e confiável, ilustrando com exemplos práticos e códigos computacionais como aplicar esses algoritmos a problemas de busca, otimização e aprendizado de máquina.

A ascensão das meta-heurísticas, como os AGs, marca uma mudança de paradigma fundamental em relação aos métodos de otimização clássicos. Algoritmos tradicionais, frequentemente baseados em cálculo de gradientes (como o método do gradiente descendente ou métodos de Newton), são altamente eficientes para problemas bem-comportados, onde a função-objetivo é convexa e analiticamente diferenciável. No entanto, eles falham categoricamente em problemas de Otimização *Black-Box* (Otimização *Black-Box* (BBO)), nos quais a função-objetivo é avaliada por meio de uma simulação numérica ou um experimento físico, sem que uma expressão matemática explícita esteja disponível. O problema de otimização do *design* de cavernas salinas, que acopla um otimizador a um simulador numérico como o SALGAS, é um exemplo canônico de um problema BBO. A avaliação de uma única solução candidata (um conjunto de parâmetros de *design*) requer a execução de uma simulação computacionalmente intensiva, cujo resultado (e.g., volume final da caverna) não pode ser expresso como uma função derivável dos parâmetros de entrada.

Neste contexto, a abordagem dos AGs é fundamentalmente diferente e mais adequada. Em vez de seguir um único caminho determinístico a partir de um ponto inicial, um AG opera sobre uma população de soluções candidatas em paralelo Goldberg (1989). A cada iteração (geração), ele utiliza operadores estocásticos (seleção, cruzamento e mutação) para gerar uma nova população. Este processo de busca populacional e estocástico confere aos AGs uma robustez intrínseca para navegar por paisagens de adequação (*fitness landscapes*) que são complexas, com múltiplos ótimos locais, e potencialmente "ruidosas"— características intrínsecas a muitos problemas de engenharia baseados em simulação. Portanto, a escolha de um AG para a otimização do *design* de cavernas salinas não é uma mera conveniência, mas uma necessidade metodológica imposta pela natureza *black-box* e pela complexidade do problema em questão.

3.2 A TEORIA DOS SCHEMAS E A HIPÓTESE DOS BLOCOS DE CONSTRUÇÃO

Para compreender a eficácia dos AGs para além de uma mera analogia biológica, John Holland desenvolveu a Teoria dos *Schemas*, culminando no que é conhecido como o Teorema Fundamental dos AGs, ou Teorema do *Schema* Holland (1975). Esta teoria fornece um arcabouço matemático para explicar como os AGs realizam uma busca eficiente, processando implicitamente um vasto número de "blocos de construção" (*building blocks*) de soluções e combinando-os para formar indivíduos progressivamente melhores.

Um *schema* (plural: *schemata*) é um gabarito ou padrão que descreve um subconjunto de cromossomos (soluções) que compartilham similaridades em certas posições de genes. Em representações binárias, um *schema* é uma *string* composta pelos símbolos $\{0, 1, *\}$, onde o asterisco (*) atua como um caractere curinga, significando que naquela posição tanto 0 quanto 1 são aceitáveis. Por exemplo, para cromossomos de comprimento 6, o *schema* $H = 1 * 10 * 1$ representa o conjunto de todas as *strings* que possuem '1' nas posições 1, 3 e 6, e '0' na posição 4, enquanto as posições 2 e 5 podem ser qualquer valor.

Dois atributos importantes de um *schema* são sua ordem e seu comprimento de definição. A ordem de um *schema*, denotada por $o(H)$, é o número de posições fixas (não-curinga) na *string*. Para $H = 1 * 10 * 1$, $o(H) = 4$. O comprimento de definição, $\delta(H)$, é a distância entre a primeira e a última posição fixa. Para $H = 1 * 10 * 1$, as posições fixas são 1 e 6, então $\delta(H) = 6 - 1 = 5$.

O Teorema do *Schema* fornece um limite inferior para o número esperado de instâncias de um determinado *schema* H na próxima geração ($t+1$), com base em sua presença e desempenho na geração atual (t). A formulação matemática do teorema é a seguinte:

$$E[m(H, t + 1)] \geq m(H, t) \frac{f(H)}{\bar{f}_t} [1 - p_d] \quad (3.1)$$

Onde:

- $E[m(H, t + 1)]$ é o número esperado de indivíduos que correspondem ao *schema* H na geração $t+1$.
- $m(H, t)$ é o número de indivíduos que correspondem ao *schema* H na geração t .
- $f(H)$ é o *fitness* (adequação) médio de todos os indivíduos na população que correspondem ao *schema* H .

- \bar{f}_t é o *fitness* médio de toda a população na geração t .
- p_d é a probabilidade de disrupção do *schema* H pelos operadores genéticos de cruzamento (*crossover*) e mutação.

A probabilidade de disrupção, p_d , pode ser aproximada por:

$$p_d \approx p_c \frac{\delta(H)}{l-1} + p_m o(H) \quad (3.2)$$

Onde:

- p_c é a probabilidade de cruzamento.
- p_m é a probabilidade de mutação.
- l é o comprimento total do cromossomo.

A interpretação do teorema é profunda: ele postula que *schemas* com *fitness* acima da média da população ($f(H) > \bar{f}_t$), comprimento de definição curto (baixo $\delta(H)$) e baixa ordem (baixo $o(H)$) receberão um número exponencialmente crescente de representantes nas gerações futuras. Esses *schemas* curtos, de baixa ordem e alto desempenho são o que Holland chamou de blocos de construção (*building blocks*). A hipótese dos blocos de construção sugere que um AG funciona combinando esses blocos de construção de ordem inferior para formar *schemas* de ordem superior e *fitness* ainda maior, convergindo gradualmente para uma solução ótima ou próxima da ótima.

O teorema encapsula a estratégia fundamental de um AG. O termo $f(H)/\bar{f}_t$ representa a força da *exploitation*: *schemas* que demonstraram bom desempenho são explorados, recebendo mais "ensaios" ou cópias na geração seguinte. O termo $(1 - p_d)$ representa a necessidade de preservação desses blocos de construção. *Schemas* com baixo comprimento de definição são menos prováveis de serem rompidos pelo operador de cruzamento, tornando-os blocos de construção robustos. A mutação, por sua vez, embora contribua para a disrupção, atua como uma força de exploração, introduzindo novo material genético e prevenindo a estagnação do algoritmo em ótimos locais. Assim, o teorema implica matematicamente que a estratégia de um AG eficaz deve favorecer a recombinação de blocos de construção de alto desempenho, enquanto utiliza a mutação com parcimônia para garantir a diversidade e a exploração de novas áreas do espaço de busca. Esta visão valida a ênfase de Goldberg no equilíbrio entre exploração e *exploitation* como o pilar central para o sucesso dos AGs Goldberg (1989).

3.3 O BALANÇO CRÍTICO: EXPLORAÇÃO VERSUS *EXPLOITATION*

O sucesso de qualquer algoritmo de busca, e em particular dos AGs, depende de um delicado equilíbrio entre duas forças antagônicas: exploração e *exploitation* Eiben e Smith (2015). Compreender e gerenciar este balanço é talvez o aspecto mais crucial no *design* e na aplicação de AGs a problemas complexos de otimização.

A exploração refere-se à capacidade do algoritmo de investigar regiões amplas e diversas do espaço de busca. É o processo de visitar áreas inteiramente novas, na esperança de descobrir regiões promissoras que ainda não foram examinadas. Um algoritmo com forte viés para a exploração é bom em mapear a paisagem de *fitness* globalmente, evitando ficar preso em picos de desempenho subótimos (ótimos locais) Eiben e Smith (2015).

A *exploitation*, por outro lado, é o processo de refinar a busca dentro de regiões que já se mostraram promissoras. Uma vez que uma área com soluções de alto *fitness* é identificada, a *exploitation* foca em buscar intensivamente na vizinhança dessas soluções para encontrar o pico local com a maior precisão possível. Um algoritmo focado em *exploitation* é eficiente em convergir rapidamente para uma boa solução, desde que a busca inicial tenha ocorrido em uma região que contenha o ótimo global.

O dilema fundamental reside no fato de que essas duas forças são concorrentes. Um excesso de exploração pode transformar o AG em uma busca puramente aleatória, ineficiente e incapaz de convergir para uma solução de alta qualidade em um tempo razoável. Por outro lado, um excesso de *exploitation* leva à convergência prematura: o algoritmo rapidamente converge para um ótimo local, perdendo a diversidade genética necessária na população para escapar desse pico e explorar outras regiões potencialmente melhores do espaço de busca.

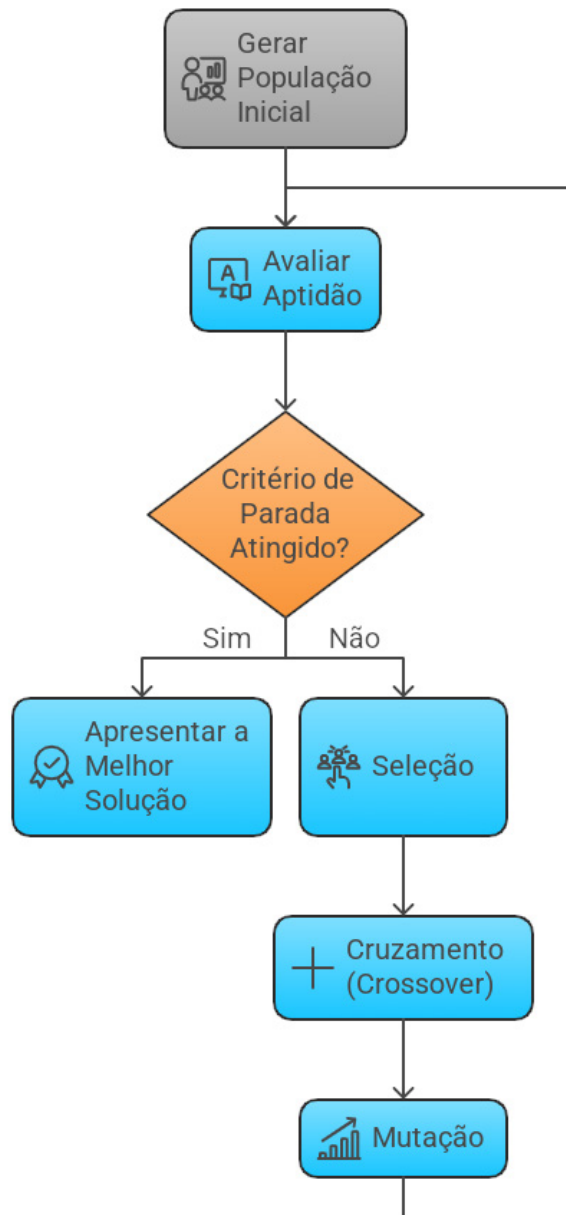
Nos AGs, este balanço é alcançado principalmente através da interação entre os operadores genéticos Kaur e Kaur (2017):

- Operador de cruzamento (*Crossover*): É primariamente um mecanismo de *exploitation*. Ele pega dois indivíduos pais, que presumivelmente possuem bom *fitness* e, portanto, contêm blocos de construção valiosos, e os recombina. A esperança é que a combinação de seus blocos de construção possa gerar um filho com *fitness* ainda maior. O cruzamento não introduz novo material genético; ele apenas rearranja o que já existe na população, explorando as combinações de *schemas* promissores.
- Operador de mutação (*Mutation*): É o principal motor da exploração. A mutação altera

aleatoriamente um ou mais genes de um indivíduo, introduzindo novo material genético na população. Este processo é crucial por duas razões: primeiro, ele pode reintroduzir alelos que foram perdidos devido à pressão seletiva, restaurando a diversidade genética; segundo, ele permite que a busca escape de ótimos locais, saltando para novas regiões do espaço de busca que não seriam acessíveis apenas através do cruzamento Kaur e Kaur (2017).

- Operador de seleção (*Selection*): Atua como o motor que impulsiona a *exploitation*. Ao dar preferência a indivíduos com maior *fitness* para se tornarem pais da próxima geração, a seleção garante que os melhores blocos de construção sejam propagados e explorados mais intensamente.

Figura 6 – Fluxo de um Algoritmo Genético.



Fonte: O autor (2025)

3.4 OTIMIZAÇÃO MULTIOBJETIVO E O ALGORITMO NSGA-II

Muitos problemas de engenharia do mundo real, incluindo o *design* de cavernas de sal, não podem ser adequadamente formulados com um único objetivo. Frequentemente, existem múltiplos objetivos conflitantes que precisam ser otimizados simultaneamente. Por exemplo, no projeto de uma caverna, deseja-se maximizar o volume de armazenamento (um objetivo econômico) e, ao mesmo tempo, minimizar o tempo de construção (outro objetivo econômico que pode conflitar com o primeiro). A Otimização Multiobjetivo (MOO) lida com esses pro-

blemas, e os Algoritmo Evolutivo Multiobjetivo (MOEA)s surgiram como uma das abordagens mais eficazes para resolvê-los.

O conceito central na MOO é a dominância de Pareto. Uma solução x_1 domina outra solução x_2 se x_1 for estritamente melhor que x_2 em pelo menos um objetivo e não for pior que x_2 em todos os outros objetivos. O objetivo de um MOEA não é encontrar uma única solução "ótima", mas sim identificar o conjunto de todas as soluções não dominadas, conhecido como fronteira de Pareto ou frente de Pareto Deb *et al.* (2002). Esta fronteira representa o conjunto de todos os *trade-offs* ótimos possíveis entre os objetivos conflitantes, fornecendo ao tomador de decisão um leque de soluções de alta qualidade para escolher.

Os primeiros MOEAs, como o *Non-dominated Sorting Genetic Algorithm* (NSGA), foram pioneiros, mas enfrentaram críticas significativas que limitavam sua aplicabilidade prática. As três principais dificuldades eram:

- Alta complexidade computacional: O processo de ordenação por não-dominância tinha uma complexidade de $O(mN^3)$, onde m é o número de objetivos e N é o tamanho da população. Isso tornava o algoritmo proibitivamente lento para problemas com populações grandes.
- Falta de elitismo: Os melhores indivíduos encontrados em uma geração não eram garantidos de sobreviver para a próxima, o que poderia levar à perda de boas soluções e retardar a convergência.
- Necessidade de um parâmetro de compartilhamento (*Sharing*): Para manter a diversidade ao longo da fronteira de Pareto e evitar a convergência para uma única região, esses algoritmos usavam uma técnica de compartilhamento que exigia a sintonia manual de um parâmetro de nicho (σ_{share}), uma tarefa difícil e sensível ao problema.

Para superar essas limitações, Deb *et al.* (2002) propuseram o NSGA-II, que rapidamente se tornou o padrão-ouro em MOEAs e é a base para o *solver* gamultiobj no MATLAB. O NSGA-II introduziu três inovações principais:

- Algoritmo de ordenação rápida por não-dominância: Foi proposto um procedimento de ordenação mais eficiente com uma complexidade computacional de $O(mN^2)$. Este algoritmo calcula para cada solução p duas entidades: n_p , o número de soluções que p domina, e S_p , o conjunto de soluções que p domina. As soluções com $n_p = 0$ pertencem à primeira frente. O algoritmo então itera sobre os membros de S_p para

cada solução da primeira frente, reduzindo suas contagens de dominação e identificando subsequentemente as próximas frentes. Esta redução na complexidade foi um avanço crucial.

- Mecanismo de elitismo explícito: O NSGA-II garante o elitismo de uma maneira elegante. A cada geração t , a população de pais P_t (tamanho N) e a população de filhos Q_t (tamanho N) são combinadas para formar uma população estendida R_t de tamanho $2N$. A seleção para a próxima geração de pais, P_{t+1} , é então realizada a partir desta população combinada. As melhores soluções (pais e filhos) são ordenadas em frentes de não-dominância. As frentes são adicionadas à nova população P_{t+1} em ordem, começando pela melhor frente (F_1), até que o tamanho da população N seja atingido. Este procedimento garante que qualquer solução de elite encontrada nunca seja perdida.
- Preservação da diversidade sem parâmetros via distância de aglomeração (*Crowding Distance*): Para manter a diversidade, especialmente quando uma frente precisa ser truncada porque não cabe inteiramente na nova população, o NSGA-II utiliza a distância de aglomeração. Esta métrica estima a densidade de soluções em torno de um ponto específico na frente de Pareto. Para cada objetivo, as soluções na frente são ordenadas, e a distância é calculada como a largura do cubóide formado pelos vizinhos mais próximos de um ponto ao longo de cada eixo de objetivo. As soluções em regiões menos povoadas (com maior distância de aglomeração) são preferidas. Isso promove uma distribuição uniforme das soluções ao longo da fronteira de Pareto, eliminando a necessidade do parâmetro σ_{share} .

A sinergia entre o elitismo e a preservação da diversidade é uma das maiores forças do NSGA-II. O algoritmo emprega uma estratégia de seleção hierárquica: a dominância de Pareto é o critério primário, garantindo a convergência em direção à fronteira ótima (*exploitation*/elitismo). A distância de aglomeração é usada como um critério secundário, de desempate, para escolher entre soluções igualmente boas (na mesma frente), promovendo a cobertura de toda a extensão da fronteira (*exploration*/diversidade). Esta abordagem dupla permite que o NSGA-II encontre de forma confiável e eficiente um conjunto bem distribuído de soluções na fronteira de Pareto para uma vasta gama de problemas.

4 SOFTWARE SALGAS: CARACTERÍSTICAS E LIMITAÇÕES

Este capítulo apresenta os conceitos, funcionalidades e limitações do *software* SALGAS, desenvolvido na década de 1970 em linguagem FORTRAN pelo SMRI. O programa foi concebido para simular o desenvolvimento de cavernas subterrâneas em formações salinas por meio da técnica de dissolução, utilizando o método numérico de diferenças finitas aplicado em malha unidimensional vertical.

O SALGAS permite a análise integrada de parâmetros hidráulicos, consumo energético e controle operacional do processo de mineração. Um de seus diferenciais é a modelagem do comportamento do *blanket*, fluido protetor posicionado sobre a salmoura, cuja função é inibir a dissolução do teto da caverna.

A validação do *software* foi realizada com base em dados experimentais de laboratório e em campanhas de monitoramento de cavernas localizadas em domos de sal espessos e homogêneos da Costa do Golfo dos Estados Unidos Eyerman (2008).

4.1 CARACTERÍSTICAS DO SOFTWARE SALGAS

O SALGAS destaca-se pelas seguintes funcionalidades principais:

- Interface flexível com opção de entrada e saída de dados em unidades métricas ou imperiais;
- Aplicação do método de diferenças finitas em malha unidimensional vertical, com até 200 células (a versão original utilizava 60) Eyerman (2008, p. 5);
- Simulação de dissolução por escoamento turbulento, com transporte reativo unidimensional e acoplamento com balanço hidráulico;
- Controle automático e adaptativo do nível do *blanket*, com possibilidade de alternar entre lavra direta (injeção superior) e reversa (injeção inferior);
- Capacidade de reinicialização da simulação (*restart*) com alteração dos parâmetros operacionais ou geomecânicos;
- Estimativas de consumo energético do sistema de bombeamento e perda de carga nas tubulações;

- Geração de saída detalhada de balanços de massa e sal, permitindo análises quantitativas da eficiência do processo.

4.2 LIMITAÇÕES DO *SOFTWARE*

Embora seja uma ferramenta robusta para o seu propósito, a correta interpretação dos resultados exige o conhecimento das premissas e limitações do *software*, que são cruciais para a contextualização desta pesquisa. As restrições mais significativas são:

- Contexto de validação: O SALGAS foi originalmente desenvolvido para simular a dissolução em domos salinos espessos e predominantemente compostos por halita pura, típicos da Costa do Golfo dos Estados Unidos. Segundo o manual, sua faixa de validação cobre cenários com “taxas de fluxo moderadas de até cerca de 1000 gpm (225 m³/h)” Eyerman (2008, p. 9). A aplicação do modelo em condições que excedem essas vazões, como explorado nesta tese, caracteriza-se como uma análise de sensibilidade fora dos limites originais de validação, sendo realizada com o devido cuidado interpretativo.
- Simulação de temperatura: O modelo numérico SALGAS opera sob uma condição isotérmica padrão de 75°F (aproximadamente 24°C), referente à temperatura interna do ambiente da caverna, conforme descrito no manual do *software* Eyerman (2008, p. 9). O SALGAS não simula explicitamente a temperatura da água injetada; em vez disso, orienta que diferentes cenários térmicos sejam representados indiretamente através do ajuste empírico do Fator de Dissolução (CONDIS). Em conformidade com essa orientação, a metodologia desta tese considerou cenários com temperaturas internas hipotéticas de aproximadamente 40°C, 60°C e 80°C. Para cada cenário, foram calculados os respectivos valores do Fator de Dissolução (DF) e das gravidades específicas, posteriormente inseridos no *software*. Dessa maneira, a análise realizada não constitui uma simulação termodinâmica direta, mas sim um estudo de sensibilidade baseado nesse parâmetro empírico, que atua como um *proxy* para representar o impacto indireto da temperatura sobre a taxa de dissolução, permitindo avaliar, por exemplo, a aceleração da dissolução em temperaturas mais elevadas.
- Simplificações do modelo: A representação da caverna no SALGAS é unidimensional, limitada ao eixo vertical, desconsiderando variações laterais e complexidades geométricas

tridimensionais. Conforme descrito no manual do *software*, o SALGAS foi desenvolvido originalmente para aplicação em domos salinos da Costa do Golfo dos Estados Unidos, compostos majoritariamente por halita pura e com geometrias relativamente simples Eyerman (2008, p. 9). O programa não simula deformações da rocha nem efeitos térmicos ou químicos sobre a matriz salina; ou seja, não há modelagem das propriedades físico-mecânicas do maciço salino que envolve a cavidade. Assim, limita-se a calcular a dissolução da halita por água injetada, sem considerar a resposta estrutural da formação ao processo de mineração por solução.

- Restrições operacionais: O modelo permite apenas um ponto fixo de injeção e um de produção por vez, operando de forma simultânea. Qualquer mudança nesses pontos ao longo do tempo requer o uso da funcionalidade de reinício da simulação a partir de um arquivo .SGR, conforme orientado no manual Eyerman (2008, p. 9). Não há suporte para múltiplas zonas ativas de injeção ou produção em uma única execução contínua do modelo.
- Observação técnica sobre a versão utilizada: A versão do SALGAS utilizada nesta tese apresenta potenciais inconsistências numéricas ao operar no Sistema Internacional de Unidades (SI), conforme identificado em testes comparativos com os mesmos casos em unidades imperiais. O manual sugere que o programa foi originalmente desenvolvido com base no sistema americano, e evidências apontam para possíveis erros de conversão na rotina.

Apesar dessas restrições, o *software* continua sendo uma ferramenta analítica valiosa para projetos convencionais de cavernas salinas, especialmente em estudos preliminares de geometria e dinâmica de dissolução.

4.2.1 Descrição das equações de alguns parâmetros de entrada do SALGAS

Os seguintes parâmetros de entrada para o SALGAS foram obtidos através de equações matemáticas: densidade específica do fluido de injeção, densidade específica da salmoura produzida, fator de dissolução do sal e pressão de injeção. A densidade específica da solução de injeção salina é a razão entre a densidade do fluido e a densidade da água pura a uma temperatura de 4°C e uma pressão de 1 atm.

A densidade da solução salina a ser injetada, ρ_b (kg/m³), é uma função da temperatura T (Kelvin), da pressão do fluido P (bar) e da concentração de sal em massa c_b (adimensional), sendo determinada, de acordo com a Association Technique de L'Industrie du Gaz en France (ATG) (1986), pela equação (4.1):

$$\rho_b(P, T, c_b) = 1000/v(P, T, c_b) \quad (4.1)$$

Onde $v(P, T, c_b)$ é o volume específico do fluido de injeção (cm³/g) e corresponde ao volume ocupado por unidade de massa, sendo diretamente influenciado pelas condições de pressão (P), temperatura (T) e concentração mássica de sal (c_b). Este parâmetro é calculado pela equação (4.2):

$$v(P, T, c_b) = A(T) - P'B(T) - P'^2C(T) + c_bD(T) + c_b^2E(T) - c_bP'F(T) - c_b^2P'G(T) - \frac{1}{2}c_bP'H(T) \quad (4.2)$$

Sendo $P' = P/0.981$, com P em bar. As funções A, B, \dots, H são funções dependentes da temperatura do fluido de injeção, cujas expressões gerais são dadas pela equação (4.3):

$$\varphi = \varphi_0 + \varphi_1T + \varphi_2T^2 + \varphi_{11}/T + \varphi_{12}/T^2 \quad (4.3)$$

Os coeficientes $\varphi_0, \varphi_1, \dots, \varphi_{12}$ para cada uma das funções (A a H) estão apresentados na tabela 1 e são utilizados diretamente na Equação (4.3) para determinar as funções termodinâmicas $A(T), B(T), \dots, H(T)$.

Tabela 1 – Coeficientes para cálculo das funções A, B, C, D, E, F, G e H .

	φ_0	φ_1	φ_2	φ_{11}	φ_{12}
A	5.916365	-1.035794×10^{-2}	9.270048×10^{-6}	-1127.522	100674.1
B	5.204914×10^{-3}	$-1.0482101 \times 10^{-5}$	8.328532×10^{-9}	-1.1702939	102.278
C	1.18547×10^{-8}	$-6.599143 \times 10^{-11}$	0	0	0
D	-2.5166	1.11766×10^{-2}	-1.70552×10^{-5}	0	0
E	2.84851	-1.54305×10^{-2}	2.23982×10^{-5}	0	0
F	-1.4814×10^{-3}	8.2969×10^{-6}	-1.2469×10^{-8}	0	0
G	2.7141×10^{-3}	-1.5391×10^{-5}	2.2655×10^{-8}	0	0
H	6.2158×10^{-7}	-4.0075×10^{-9}	6.5972×10^{-12}	0	0

Fonte: Association Technique de L'Industrie du Gaz en France (ATG) (1986)

A densidade da água pura, ρ_w (kg/m³), de acordo com Patterson e Morris (1994), é função da temperatura, T (°C), e da pressão do fluido, P (MPa), sendo dada pela equação (4.4):

$$\rho_w(T) = \rho_w^0 \left[1 + \frac{(T + a_1)^2 (T + a_2)}{a_3 (T + a_4)} \right] \left[1 + (b_1 + b_2 T + b_3 T^2) (P - P^{\text{ref}}) \right] \quad (4.4)$$

Onde os coeficientes a_i e b_i são constantes, e $P^{\text{ref}} = 1 \text{ atm} = 0.101325 \text{ MPa}$. Os parâmetros utilizados para o cálculo da densidade da água pura estão apresentados na tabela 2.

Tabela 2 – Parâmetros da água pura.

$\rho_w^0 = 999.972 \text{ kg/m}^3$	$a_4 = 69.348811^\circ\text{C}$
$a_1 = -3.9830355^\circ\text{C}$	$b_1 = 5.074 \times 10^{-4}/\text{MPa}$
$a_2 = 301.797^\circ\text{C}$	$b_2 = -3.26 \times 10^{-6}/\text{MPa} \cdot ^\circ\text{C}$
$a_3 = 522,528.9^\circ\text{C}^2$	$b_3 = 4.16 \times 10^{-9}/\text{MPa} \cdot ^\circ\text{C}^2$

Fonte: Patterson e Morris (1994)

Outro parâmetro de entrada que requer desenvolvimento matemático é a densidade específica da salmoura produzida. Esta é a razão entre a densidade do fluido em saturação e a densidade da água pura a 4°C e 1 atm. A concentração mássica de sal na salmoura saturada, c_b^{sat} , depende da pressão, P (MPa), e da temperatura, T (°C). De acordo com Association Technique de L'Industrie du Gaz en France (ATG) (1986), sua forma é definida pela equação (4.5):

$$c_b^{\text{sat}}(P, T) = a_0 + a_1 T + a_2 T^2 + bP \quad (4.5)$$

Onde a_i e b são constantes, cujos valores estão na tabela 3.

Tabela 3 – Valores das constantes a_i e b .

a_0	a_1	a_2	b
0.26291	$0.7448 \times 10^{-4}/^\circ\text{C}$	$0.1252 \times 10^{-5}/^\circ\text{C}^2$	$7.5 \times 10^{-5}/\text{MPa}$

Fonte: Association Technique de L'Industrie du Gaz en France (ATG) (1986)

A densidade específica da salmoura produzida é obtida substituindo-se a concentração de saturação, c_b^{sat} , na equação de densidade da água, (4.4).

O terceiro parâmetro de entrada a ser determinado é o fator de dissolução. Segundo Saberian (1983), é possível estimar a taxa de dissolução da halita em função da temperatura e da densidade específica da solução. Com base em testes laboratoriais para superfícies verticais de sal em salmouras altamente concentradas (até 96,8% de saturação), o autor propôs uma relação empírica para a taxa de dissolução m_T (em cc/cm²/min $\times 10^3$):

$$\dot{m}_T = 0.22(1.2019 - \rho)^{1.42} \cdot \exp \left[0.0119 \left(\frac{\rho - 1}{1.2019 - \rho} \right)^{0.2} \cdot \Delta T \right] \quad (4.6)$$

Onde ρ é a densidade da salmoura (adimensional) e $\Delta T = T - T_0$ é a diferença entre a temperatura da salmoura, T , e a temperatura de referência, $T_0 = 75^\circ\text{F}$.

No entanto, o SALGAS é um modelo isotérmico, fixado a 75°F , e não realiza simulações termodinâmicas diretas. Por isso, a entrada no modelo não é a taxa de dissolução, mas sim o DF (CONDIS), que atua como parâmetro de correção da taxa para diferentes condições operacionais.

Esse fator é derivado do termo exponencial da equação (4.6), assumindo como referência uma salmoura saturada ideal com densidade $\rho = 1.2019$ a $T_0 = 75^\circ\text{F}$. Para simular o efeito de diferentes temperaturas e salinidades, calcula-se o DF da seguinte forma:

$$DF = \exp \left[0.0119 \left(\frac{\rho - 1}{1.2019 - \rho} \right)^{0.2} \cdot (T - 75) \right] \quad (4.7)$$

onde:

- T : temperatura da salmoura ($^\circ\text{F}$),
- ρ : densidade específica da salmoura (adimensional),
- DF: Fator de Dissolução (adimensional).

Esse fator é inserido no SALGAS para representar indiretamente o efeito térmico sobre a taxa de dissolução. A metodologia desta tese utiliza a equação (4.7) para calcular os valores de DF, permitindo análises de sensibilidade que simulam, de forma aproximada, os efeitos térmicos sobre o processo de lixiviação.

A pressão de injeção P_{inj} (MPa) pode ser interpretada como a diferença de pressão entre o ponto de entrada do fluido no sistema (pressão de bombeamento, P_1) e a pressão interna na caverna (P_2). Essa diferença não é puramente estática, pois envolve também a energia cinética e o desnível entre os pontos, conforme descrito pela equação de Bernoulli:

$$\frac{P_1}{\gamma} + \frac{v_1^2}{2g} + z_1 = \frac{P_2}{\gamma} + \frac{v_2^2}{2g} + z_2 \quad (4.8)$$

onde:

- P_1, P_2 : pressão nos pontos 1 e 2 (Pa),

- v_1, v_2 : velocidade do fluido nos pontos 1 e 2 (m/s),
- z_1, z_2 : cota (altura) dos pontos 1 e 2 (m),
- γ : peso específico do fluido (N/m³),
- g : aceleração da gravidade (9,81 m/s²).

Admitindo que a diferença de altura entre os pontos é desprezível ($z_1 = z_2$), a equação se reduz a:

$$P_1 - P_2 = \frac{\gamma}{2g} (v_2^2 - v_1^2)$$

Essa expressão mostra que a diferença de pressão entre o ponto de bombeamento (P_1) e o fundo do poço (P_2) depende da diferença entre as velocidades do fluido.

Para estimar a pressão absoluta P_i (MPa) em uma profundidade H_i (m), pode-se utilizar a relação hidrostática:

$$P_i = \frac{\gamma \cdot H_i}{10^6} \quad (4.9)$$

onde:

- P_i : pressão na válvula de injeção (MPa),
- H_i : profundidade da válvula de injeção (m),
- γ : peso específico do fluido (N/m³).

4.2.2 Formulação Matemática do Modelo de Dissolução 1D

A base matemática do simulador SALGAS, conforme descrito no trabalho de Saberian (1974), modela o processo de dissolução como um problema de transporte de massa em uma dimensão (vertical). O objetivo é prever a variação da concentração de sal na salmoura ao longo da altura da caverna (z) e do tempo (t). O crescimento do raio da caverna é, então, uma consequência direta da massa de sal removida da parede.

A equação governante do processo é a equação de convecção-difusão em estado não estacionário, que descreve como a concentração de sal (C) evolui no sistema:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial z^2} - V \frac{\partial C}{\partial z} \quad (4.10)$$

Onde os termos representam:

- C : concentração de sal na solução,
- t : tempo,
- z : coordenada espacial vertical,
- D : coeficiente de difusão molecular do sal na solução,
- V : velocidade do fluido ao longo da coordenada vertical z .

A velocidade do fluido (V) não é constante, variando com a altura devido às mudanças na densidade da salmoura, que aumenta à medida que mais sal é dissolvido. O modelo acopla esta equação a balanços de massa e volume para calcular a velocidade em cada ponto.

A taxa na qual a parede de sal se dissolve, determinando o aumento do raio da caverna, é governada por uma condição de contorno na interface entre a rocha salina e a salmoura. Esta taxa de transferência de massa é expressa pela seguinte relação:

$$\text{Taxa de dissolução} = k(C_{\text{sat}} - C_{\text{interface}}) \quad (4.11)$$

Onde:

- k : é o coeficiente de transferência de massa, que depende das condições de fluxo (turbulência) e das propriedades do fluido,
- C_{sat} : é a concentração de saturação do sal na solução, que é função da temperatura e pressão,
- $C_{\text{interface}}$: é a concentração de sal na interface imediata entre a rocha e a salmoura.

O *software* SALGAS resolve numericamente este sistema de equações, utilizando o método de diferenças finitas, para simular a evolução da geometria da caverna ao longo do tempo. As relações empíricas para o fator de dissolução, como as propostas por Saberian (1983), são utilizadas para ajustar o coeficiente de transferência de massa (k) e, assim, incorporar os efeitos da temperatura no modelo, que é inerentemente isotérmico.

5 METODOLOGIA

Neste capítulo, são desenvolvidas as metodologias tanto para a simulação dos cenários de mineração por solução para a abertura da caverna subterrânea em rocha salina mediante o *software* SALGAS, como também o emprego das duas técnicas de redução de dimensionalidade, PCA e t-SNE, através da linguagem de programação R.

5.1 CENÁRIOS DE SIMULAÇÃO E MATRIZ DE DADOS DE ENTRADA

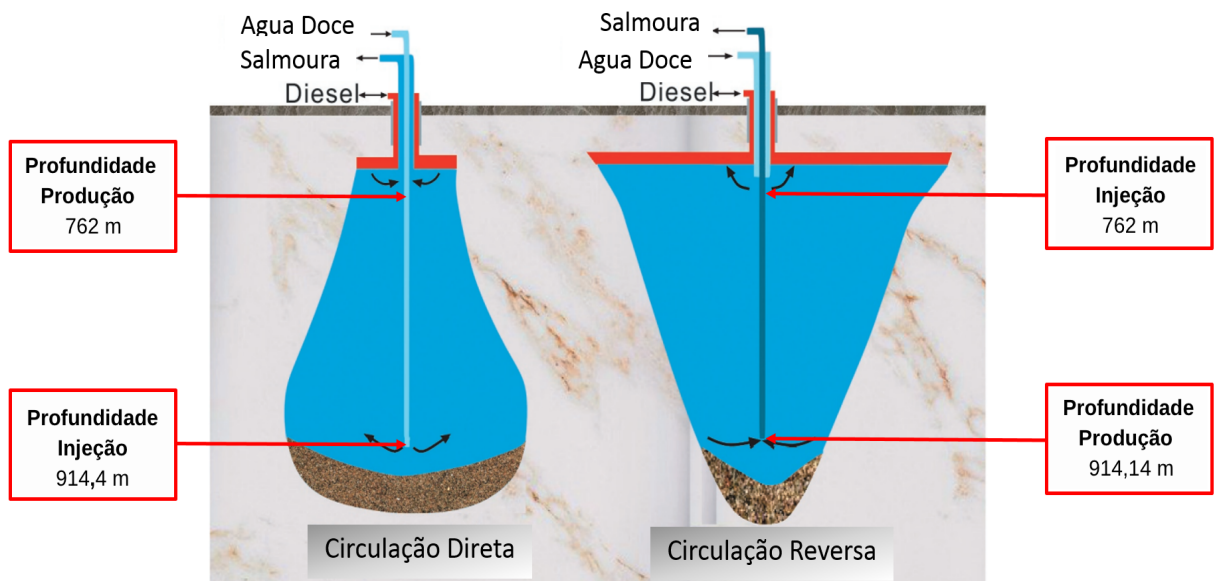
Foram considerados dois cenários para análise da abertura de uma caverna subterrânea em rocha salina por dissolução: o primeiro deles, cenário *D*, corresponde ao método de circulação direta, onde a água foi injetada no fundo da caverna e a salmoura extraída do topo. O segundo, cenário *R*, corresponde ao método de circulação reversa, onde a água foi injetada no topo e a salmoura extraída do fundo. A tabela 4 apresenta as respectivas profundidades de injeção e produção para ambos os cenários, e a figura 7 ilustra sua representação esquemática.

Tabela 4 – Cotas de Injeção e Produção para os cenários propostos.

CENÁRIO	Profundidade de Injeção (m)	Profundidade de Produção (m)
Cenário <i>D</i>	914,4	762
Cenário <i>R</i>	762	914,4

Fonte: O autor (2025)

Figura 7 – Diagrama esquemático de formas de cavernas de sal resultantes de diferentes modos de circulação.

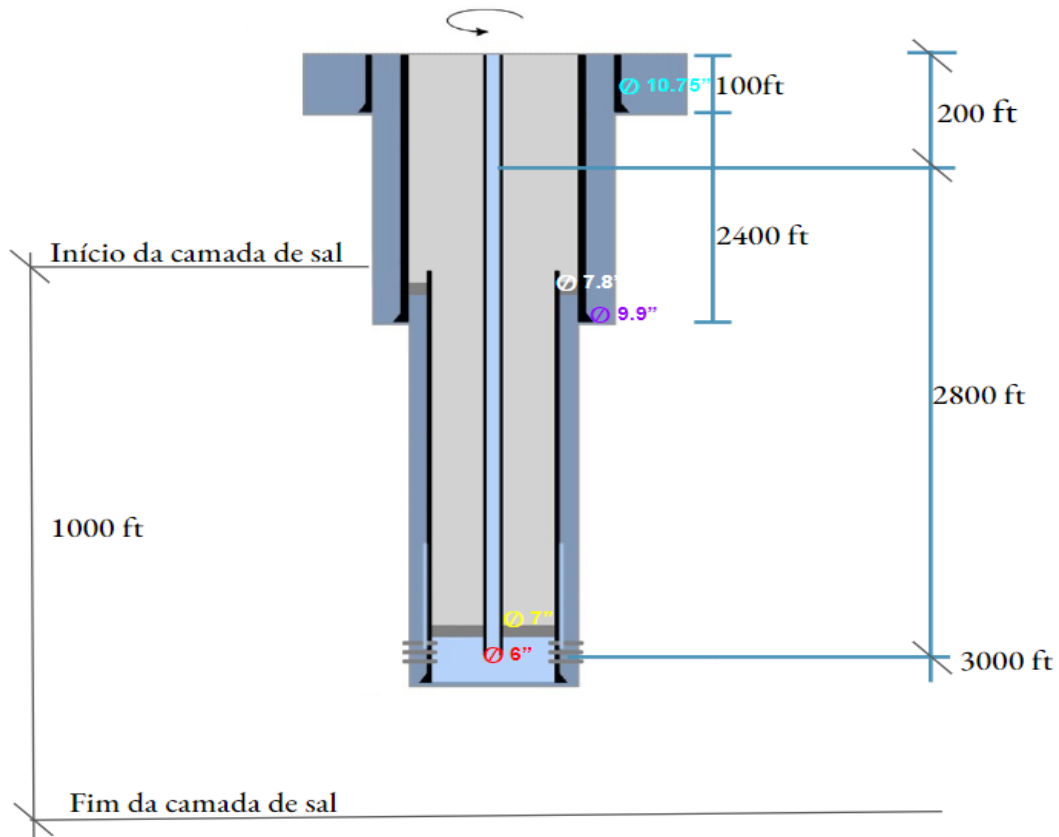


Fonte: Modificado de Xue et al. (2020)

Estes dois cenários são modificações de um caso de referência do Manual de SALGAS Eyerman (2008), sendo a configuração específica utilizada nesta pesquisa alusiva ao trabalho de Silva (2022). A matriz de simulações foi construída a partir de uma configuração inicial com os seguintes parâmetros fixos: o colchão de proteção permanece constante; a construção da caverna é iniciada a partir de um poço de raio 1 ft (0,30 m); a rocha salina contém 3% de insolúveis; e o modelo hidráulico possui uma seção curta de tubulação de superfície e divide cada uma das linhas de tubulação em duas seções, como pode ser visto na figura 8. As dimensões detalhadas dos componentes do poço estão apresentadas na Tabela 5.

A partir dessa configuração inicial, foram realizadas novas simulações para análise. As variações nos cenários foram definidas pela taxa de produção da salmoura e pela temperatura interna equivalente que se desejava simular (40°C, 60°C e 80°C).

Figura 8 – Estrutura esquemática do poço.



Fonte: Adaptado de Lecampion et al. (2013)

Tabela 5 – Dimensões e profundidades dos componentes do poço.

Componente / Seção	Profundidade / Comprimento		Diâmetro	
	(ft)	(m)	(in)	(m)
Profundidade da sapata do revestimento 1	2400	731,52	10.75 (OD)	0,273
Profundidade da sapata do revestimento 2	2800	853,44	9.9 (ID)	0,251
Profundidade total do poço	3000	914,40	-	-
Espessura da camada de sal	1000	304,80	-	-
Diâmetro externo da tubulação interna	-	-	7.8	0,198

*OD: Diâmetro Externo, ID: Diâmetro Interno

Fonte: O autor (2025)

5.1.1 Geração dos Parâmetros de Entrada Variáveis

Para a obtenção dos valores da gravidade específica do fluido de injeção, da gravidade específica da salmoura e do fator de dissolução do sal, foi necessário determinar a temperatura

e a pressão de injeção para cada uma das taxas de produção. A pressão foi calculada mediante a equação de Bernoulli e a equação hidrostática. Com os parâmetros de temperatura e pressão, as gravidades específicas foram calculadas na caixa de ferramentas do SMRI Brouard (2008). O efeito da temperatura foi modelado através do ajuste do Fator de Dissolução, cujo cálculo foi detalhado no Capítulo 4. O módulo de mineração foi ativado em conjunto com o módulo hidráulico para simular a dissolução da rocha por um fluido saturado em 4.05% de NaCl para 120 dias de simulação. As equações para o cálculo da pressão, das gravidades específicas e do Fator de Dissolução estão detalhadas na Seção 4.2.1 e no Apêndice C. O procedimento prático para a geração dos dados e execução da simulação no *software* é apresentado no Apêndice D.

Logo, as tabelas 6 e 7 dos cenários *D* e *R*, respectivamente, apresentam os parâmetros calculados anteriormente junto com os valores da temperatura e taxa de produção. Estas tabelas foram usadas no arquivo de dados de entrada do SALGAS.

Tabela 6 – Parâmetros de entrada para o Cenário *D*.

N°	Temp. (°C)	Taxa de Produção (m³/h)	Pressão de Injeção (MPa)	SG da Salmoura	SG Fluido Injeção	Fator Dissolução da Sal
1	40	116,00	9,0500	1,196300	1,023930	2,01959411
2	40	193,33	9,1800	1,196400	1,024030	2,02487103
3	40	348,00	9,6700	1,196500	1,024230	2,03027641
4	40	386,66	9,8300	1,196600	1,024330	2,03581592
5	40	773,33	12,4200	1,197600	1,025430	2,10014863
6	40	1159,99	16,7400	1,199400	1,027130	2,28992273
7	60	116,00	9,0500	1,187400	1,014330	3,63441577
8	60	193,33	9,1800	1,187400	1,014530	3,63441577
9	60	348,00	9,6700	1,187600	1,014630	3,64849256
10	60	386,66	9,8300	1,187600	1,014730	3,64849256
11	60	773,33	12,4200	1,188700	1,015830	3,73123637
12	60	1159,99	16,7400	1,190500	1,017630	3,89033759
13	80	116,00	9,0500	1,179100	1,002830	6,14168842
14	80	193,33	9,1800	1,179100	1,002930	6,14168842
15	80	348,00	9,6700	1,179300	1,003130	6,16388254
16	80	386,66	9,8300	1,179400	1,003230	6,17509104
17	80	773,33	12,4200	1,180400	1,004330	6,29147618
18	80	1159,99	16,7400	1,182300	1,006130	6,53698638

Fonte: O autor (2025)

Tabela 7 – Parâmetros de entrada para o Cenário R.

Nº	Temp. (°C)	Taxa de Produção (m³/h)	Pressão de Injeção (MPa)	SG da Salmoura	SG Fluido Injeção	Fator Dissolução da Sal
1	40	116,00	7,5600	1,195700	1,023330	1,25822596
2	40	193,33	7,7100	1,195800	1,023430	1,25850574
3	40	348,00	8,2300	1,196000	1,023530	1,25878481
4	40	386,66	8,4100	1,196000	1,023630	1,25906319
5	40	773,33	11,2200	1,197200	1,024830	1,26235126
6	40	1159,99	15,9000	1,199000	1,026730	1,26737505
7	60	116,00	7,5600	1,186800	1,013730	1,58127404
8	60	193,33	7,7100	1,186800	1,013830	1,58240401
9	60	348,00	8,2300	1,187000	1,014030	1,58464903
10	60	386,66	8,4100	1,187100	1,014130	1,58576420
11	60	773,33	11,2200	1,188200	1,015330	1,59878792
12	60	1159,99	15,9000	1,190100	1,017230	1,61821655
13	80	116,00	7,5600	1,178500	1,002230	1,63097469
14	80	193,33	7,7100	1,178500	1,002330	1,63810134
15	80	348,00	8,2300	1,178700	1,002530	1,65174855
16	80	386,66	8,4100	1,178800	1,002630	1,65829672
17	80	773,33	11,2200	1,179900	1,003830	1,72622905
18	80	1159,99	15,9000	1,181900	1,005730	1,80920606

Fonte: O autor (2025)

Uma vez preenchido o arquivo de entrada do SALGAS (*.sgi), foram realizadas a execução para todos os casos.

5.2 DADOS DE ENTRADA PARA O SOFTWARE R

Nesta seção, detalha-se a metodologia para a análise de redução de dimensionalidade, que foi conduzida utilizando a linguagem de programação R R Core Team (2020) no ambiente de desenvolvimento integrado RStudio. Foram empregados pacotes específicos para a implementação das técnicas de PCA, conforme descrito a seguir.

5.2.1 Análise de Componentes Principais com R

Após a realização das simulações no SALGAS para os cenários D e R, foi conduzida uma análise sobre os dados de entrada e saída. Para esta análise, o conjunto de dados foi estruturado em dois grupos principais. O primeiro grupo, de variáveis de entrada, corresponde aos parâmetros operacionais controlados em cada simulação: temperatura do cenário, taxa de produção de salmoura, pressão de injeção, gravidade específica inicial da salmoura, gravidade específica do fluido de injeção e fator de dissolução do sal. O segundo grupo, de variáveis de saída, representa os principais resultados calculados pelo simulador ao final de 120 dias: potência da bomba, energia acumulada, perdas na tubulação, salmoura produzida, pressão da

bomba, além do volume e raio da caverna.

Os dados de saída para os cenários D e R estão detalhados nas tabelas 8 a 11.

Tabela 8 – Resultados da simulação (dados de saída) para o PCA do cenário D.

Nº	Temp. (°C)	Taxa de prod. (m ³ /h)	Raio (m)	Volume (m ³)	Potência da bomba (kW)
1	40	116,00	8,4815233	41384,70098	69,35162123
2	40	193,33	10,1243922	64977,88993	150,0948167
3	40	348,00	12,9358796	107985,0077	479,9898047
4	40	386,66	13,5319465	118130,4367	610,3281201
5	40	773,33	18,0928945	210546,8649	3543,051524
6	40	1159,99	21,5553646	297725,8378	10881,81472
7	60	116,00	9,37688244	47798,02484	75,62417372
8	60	193,33	11,3415623	76518,13933	161,4826027
9	60	348,00	14,3688664	130838,4693	503,873569
10	60	386,66	15,0620242	143839,6173	637,7167415
11	60	773,33	20,4725412	266268,5084	3619,96816
12	60	1159,99	24,4236057	378841,1501	11033,15905
13	80	116,00	10,2039847	53401,78831	81,48838835
14	80	193,33	12,3364173	86638,43293	172,1319668
15	80	348,00	15,7673253	150454,9643	524,9347246
16	80	386,66	16,474062	166003,7545	661,769377
17	80	773,33	22,4124439	314677,9596	3682,872839
18	80	1159,99	26,9037786	454376,837	11158,60476

Fonte: O autor (2025)

Tabela 9 – Resultados da simulação (dados de saída, continuação) para o PCA do cenário D.

Nº	Energia cumulativa (J)	Perda de tubulação (MPa)	Salmoura produzida (m ³)	Pressão da bomba (MPa)
1	7,07426E+11	0,151044266	334083,1111	1,877511889
2	1,53251E+12	0,38376179	556805,1852	2,440216488
3	4,92057E+12	1,132083119	1002247,32	4,340674207
4	6,26193E+12	1,375908446	1113610,313	4,968665288
5	3,65067E+13	5,013018728	2227214,085	14,44934384
6	1,1228E+14	10,78227224	3340824,399	29,61839362
7	7,75713E+11	0,151004345	334083,1111	2,038940628
8	1,65758E+12	0,383827704	556805,1852	2,613999741
9	5,18354E+12	1,132913937	1002247,32	4,535153875
10	6,56377E+12	1,377048219	1113610,313	5,16682462
11	3,73623E+13	5,020753821	2227214,085	14,68605405
12	1,13974E+14	10,79822948	3340824,399	29,87526115
13	8,38823E+11	0,15075772	334083,1111	2,187742167
14	1,77226E+12	0,383280812	556805,1852	2,773871437
15	5,41312E+12	1,131692668	1002247,32	4,702727224
16	6,82639E+12	1,375658028	1113610,313	5,336485083
17	3,80577E+13	5,017673035	2227214,085	14,86660017
18	1,15379E+14	10,79353104	3340824,399	30,05975134

Fonte: O autor (2025)

Tabela 10 – Resultados da simulação (dados de saída) para o PCA do cenário R.

Nº	Temp. (°C)	Taxa de prod. (m³/h)	Raio (m)	Volume (m³)	Potência da bomba (kW)
1	40	116,00	10,5560683	45041,04639	64,14827577
2	40	193,33	12,8845788	71511,51758	143,6295902
3	40	348,00	16,2241901	120021,52	477,5601426
4	40	386,66	16,8938997	131330,9497	610,4115863
5	40	773,33	22,0228759	234701,064	3619,319431
6	40	1159,99	22,0228759	234701,064	3619,319431
7	60	116,00	11,7554258	50428,89788	68,18619653
8	60	193,33	14,4205818	81612,43657	151,202293
9	60	348,00	18,3731733	140939,9536	494,1020728
10	60	386,66	19,1935181	155241,9248	629,9281586
11	60	773,33	25,5436177	289757,148	3686,488656
12	60	1159,99	30,1478716	413365,7451	11318,65003
13	80	116,00	12,8243381	54978,27937	72,13088987
14	80	193,33	15,7257933	89867,37527	158,1342307
15	80	348,00	20,1169311	157578,1459	508,4429874
16	80	386,66	21,0404537	174118,7867	646,4878364
17	80	773,33	28,3183523	333077,6071	3736,180009
18	80	1159,99	33,6982003	482909,6543	11431,7722

Fonte: O autor (2025)

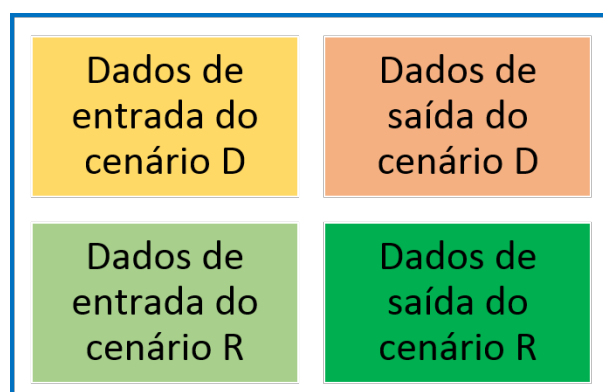
Tabela 11 – Resultados da simulação (dados de saída, continuação) para o PCA do cenário R.

Nº	Energia cumulativa (J)	Perda de tubulação (MPa)	Salmoura produzida (m ³)	Pressão da bomba (MPa)
1	6,60392E+11	0,161226792	334083,1111	1,733942439
2	1,47759E+12	0,409806815	556805,1852	2,331043548
3	4,91624E+12	1,20936476	1002249,333	4,309977908
4	6,28537E+12	1,469766056	1113610,37	4,959177064
5	3,73211E+13	5,351425683	2227216,515	14,72807348
6	1,1514E+14	11,50081504	3340826,886	30,31462911
7	7,04946E+11	0,161708184	334083,1111	1,836688911
8	1,56229E+12	0,41146073	556805,1852	2,444840252
9	5,1048E+12	1,216036681	1002249,333	4,44139424
10	6,50832E+12	1,478493856	1113610,37	5,096759518
11	3,8107E+13	5,39738705	2227216,515	14,93186094
12	1,17013E+14	11,61008216	3340826,886	30,59706916
13	7,47251E+11	0,161916544	334083,1111	1,935635819
14	1,63791E+12	0,412151723	556805,1852	2,547067066
15	5,26315E+12	1,219137186	1002249,333	4,551803203
16	6,69103E+12	1,48257059	1113610,37	5,209343364
17	3,86733E+13	5,420276274	2227216,515	15,06683483
18	1,18336E+14	11,66974086	3340826,886	30,76282574

Fonte: O autor (2025)

Para realizar a análise estatística, foram desenvolvidas rotinas no ambiente R. Foram utilizados os pacotes FactoMineR Lê, Josse e Husson (2008) para a execução da análise e factoextra Kassambara (2017) para a visualização dos dados. A abordagem metodológica consistiu em unificar os dados de entrada e os resultados da simulação em uma única matriz, permitindo que o PCA capturasse a variabilidade do sistema de forma holística. Esta matriz combinada, cujo esquema é apresentado na figura 9, foi então utilizada como entrada para a função PCA. Os *scripts* completos utilizados nesta análise estão disponíveis neste repositório online para garantir a reprodutibilidade da pesquisa.

Figura 9 – Esquema da matriz de dados de entrada do PCA.



Fonte: O autor (2025)

5.2.2 Análise t-SNE com R

A implementação da técnica t-SNE foi realizada com o auxílio do pacote `Rtsne` no ambiente de *software* R Krijthe (2015). O procedimento iniciou-se com a instalação do pacote no console do RStudio, utilizando o comando `install.packages("Rtsne")`. Em seguida, em cada *script* de análise, a biblioteca foi carregada com o comando `library(Rtsne)`.

A função `Rtsne` possui diversos hiperparâmetros que permitem ajustar sua execução. Os principais, conforme a descrição de Rhys (2020), são:

- `perplexity`: Este parâmetro controla a largura das distribuições de probabilidade usadas para converter distâncias. De forma simplificada, pode ser entendido como o número de vizinhos próximos que cada ponto considera. Valores altos focam mais na estrutura global, enquanto valores pequenos focam na estrutura local. Os valores típicos estão no intervalo de 5 a 50.
- `theta`: Controla a compensação entre velocidade e precisão na implementação Barnes-Hut t-SNE Maaten (2014). Um valor de `theta=0` corresponde à implementação exata, que é mais lenta. Valores maiores aceleram o cálculo com alguma perda de precisão. Neste trabalho, foi escolhido `theta=0` para garantir a máxima precisão nos resultados.
- `eta`: Corresponde à taxa de aprendizado, determinando a magnitude do ajuste da posição dos pontos em cada iteração. Valores mais baixos podem levar a uma incorporação mais precisa, mas exigem mais iterações. O valor padrão é 200.
- `max_iter`: Define o número máximo de iterações que o algoritmo executará. O valor padrão é 1.000.

Os hiperparâmetros mais importantes para o ajuste fino do resultado são, geralmente, a perplexidade (`perplexity`) e o número de iterações (`max_iter`).

Adicionalmente, foi realizado um ajuste crucial no pré-processamento dos dados. O argumento `pca` da função `Rtsne`, cujo valor padrão é `TRUE`, executa uma análise PCA inicial antes do t-SNE. Nesta pesquisa, essa opção foi explicitamente desativada, definindo `pca = FALSE`. Tal decisão foi tomada para garantir que a análise t-SNE fosse aplicada diretamente sobre a matriz de dados original, permitindo uma avaliação de seu desempenho sem a influência de uma etapa prévia de redução de dimensionalidade linear.

Finalmente, foram executadas as rotinas que envolvem a função Rtsne com a matriz de dados para gerar as figuras apresentadas na análise de resultados.

5.3 METODOLOGIA DE OTIMIZAÇÃO DO DESIGN

Com as metodologias de análise do sistema definidas, esta seção detalha a abordagem utilizada para a otimização do design da caverna. O objetivo é transitar da compreensão do fenômeno para a determinação de estratégias operacionais ótimas. Para isso, a metodologia de otimização foi estruturada em duas etapas fundamentais, que serão descritas a seguir: primeiramente, a formulação matemática do problema e, em segundo lugar, a implementação computacional do *framework* de otimização.

5.3.1 Formulação Matemática do Problema

O planejamento de uma campanha de mineração por dissolução é modelado como um problema de otimização, em que se busca o vetor de decisão ótimo \vec{x}^* que maximize um conjunto de indicadores de desempenho, sujeito a um sistema de restrições físicas, operacionais e geomecânicas. Formalmente, o problema é expresso como:

$$\begin{aligned} \text{Otimizar} \quad & F(\vec{x}) = [f_1(\vec{x}), f_2(\vec{x}), \dots, f_k(\vec{x})] \\ \text{Sujeito a} \quad & g_j(\vec{x}) \leq 0, \quad j = 1, \dots, m \\ & \vec{x}_{\min} \leq \vec{x} \leq \vec{x}_{\max} \end{aligned}$$

onde $F(\vec{x})$ é o vetor de funções-objetivo que quantifica o desempenho (por exemplo, maximizar volume e minimizar tempo), $g(\vec{x})$ representa as restrições de segurança e estabilidade, e \vec{x} é o vetor de decisão, cujos limites operacionais são definidos por \vec{x}_{\min} e \vec{x}_{\max} . O vetor \vec{x} pertence ao espaço de soluções admissíveis (viáveis), D .

5.3.2 Vetor de Decisão

Para uma campanha de mineração com $N=2$ estágios de dissolução, o vetor de decisão \vec{x} é um ponto no espaço \mathbb{R}^{10} , definido como:

$$\vec{x} = [Q_1, t_1, B_1, I_1, P_1, Q_2, t_2, B_2, I_2, P_2] \quad (5.1)$$

Os componentes deste vetor representam as variáveis de controle do processo, cujos limites práticos definem o espaço de busca. Para $i = 1, 2$, tem-se que:

- Q_i : Vazão de injeção de salmoura no estágio i [m^3/h].
- t_i : Duração do estágio i [dias].
- B_i : Profundidade do *blanket* (definindo o teto da caverna) no estágio i [m].
- I_i : Profundidade de injeção da tubulação no estágio i [m].
- P_i : Profundidade de produção da tubulação no estágio i [m].

A escolha de modelar a campanha de mineração com $N = 2$ estágios representa uma simplificação estratégica do problema. Embora projetos reais possam envolver múltiplos estágios para um controle geométrico mais fino, a limitação a dois estágios foi adotada por duas razões principais. Primeiramente, o custo computacional: cada estágio adicional insere cinco novas variáveis no vetor de decisão, aumentando exponencialmente a complexidade do espaço de busca e o tempo necessário para a convergência do algoritmo genético. Em segundo lugar, o objetivo deste trabalho é demonstrar a viabilidade e a eficácia da metodologia de acoplamento do otimizador ao simulador. Um modelo de dois estágios é suficientemente complexo para validar o *framework* como uma prova de conceito robusta, que pode ser estendida para um número maior de estágios em trabalhos futuros.

5.3.3 Sistema Hierárquico de Restrições e Espaço de Busca

O sucesso de um projeto de caverna salina é medido por um duplo imperativo que equilibra objetivos econômicos com mandatos geomecânicos intransigentes. A otimização não busca o máximo volume absoluto, mas o máximo volume seguro e estável, refletindo uma filosofia de "Otimização de Design Robusto" (RDO).

As fronteiras dos parâmetros operacionais, definidas como restrições de caixa (*box constraints*), constituem a primeira e mais fundamental camada deste *framework*. Estas restrições definem o hipercubo de busca inicial, cujos limites foram estabelecidos com base em práticas da indústria e na necessidade de garantir a viabilidade de campanhas de lixiviação de longo prazo.

Vazão de injeção (Q_i):

$$50 \leq Q_i \leq 300 \quad [\text{m}^3/\text{h}]$$

Duração dos estágios (t_i): Para refletir cronogramas realistas, a duração de cada um dos dois estágios foi limitada a um período entre 6 meses e 2 anos, ou seja, de 180 a 720 dias.

$$180 \leq t_i \leq 720 \quad [\text{dias}]$$

Profundidades operacionais (B_i, I_i, P_i): As profundidades foram restringidas a faixas específicas, em metros, para garantir a coerência física da circulação direta e respeitar as margens de segurança geotécnica. Para maior clareza, os limites em pés (ft), conforme implementados no código, são detalhados a seguir:

$$720.8 \leq B_i \leq 780.3 \quad [\text{m}] \quad (\text{Blanket: } 2365\text{--}2560 \text{ ft})$$

$$781.8 \leq P_i \leq 839.7 \quad [\text{m}] \quad (\text{Produção: } 2565\text{--}2755 \text{ ft})$$

$$841.2 \leq I_i \leq 862.6 \quad [\text{m}] \quad (\text{Injeção: } 2760\text{--}2830 \text{ ft})$$

5.3.4 Justificativa dos Parâmetros do Espaço de Busca

A seleção dos limites para as variáveis de controle é uma decisão deliberada, fundamentada na física da dissolução, em limitações de *software*, em precedentes da indústria e em princípios geomecânicos.

5.3.4.1 Vazão de injeção (Q_i): 50 a 300 m^3/h

A faixa de vazão representa um balanço entre eficiência e segurança. Vazões mais altas induzem maior turbulência, acelerando a taxa de dissolução. A faixa selecionada está ancorada em projetos análogos e no envelope de validação do *software*:

- Práticas conservadoras (50–100 m^3/h): Taxas mais baixas são usadas para melhor controle geométrico, especialmente em formações de sal estratificadas, como no projeto Jintan na China Wang *et al.* (2019), Li *et al.* (2022).
- Práticas modernas (150–300 m^3/h): O programa INDOROCK na Índia planeja fases com taxas de 150 e 300 m^3/h Raju, Murthy e Kumar (2017). O limite superior do

SALGAS foi validado para "taxas de fluxo moderadas de até cerca de 1000 gpm (225 m³/h)" Eyerman (2008) e o limite de 300 m³/h representa uma extrapolação consciente para alinhar o estudo com práticas de engenharia atuais.

A faixa de busca selecionada é, portanto, robusta, abrangendo desde operações comprovadas até os limites superiores da prática moderna.

5.3.4.2 Duração dos estágios (t_i): 180 a 720 dias

A duração de cada estágio (t_i) é um parâmetro crítico que afeta a economia e a geomecânica do projeto. A literatura estabelece que a criação de uma caverna é um processo de longo prazo, podendo levar "um ano ou mais" Al-Halaf e Khazzan (2023). Modelagens detalhadas chegam a simular a fase de lixiviação por períodos de 700 dias antes da operação Speirs, Bere e Roberts (2022). A duração também é fundamental para a estabilidade, pois governa a história inicial de fluência (*creep*) da rocha salina Seright e Brattekas (1991). Portanto, o intervalo de busca de 180 a 720 dias (6 meses a 2 anos) por estágio foi adotado para permitir a exploração de estratégias de lixiviação realistas.

5.3.4.3 Profundidades operacionais (B_i, I_i, P_i): 609.6 a 914.4 m

O intervalo de 609.6 a 914.4 m (equivalente a 2000-3000 pés) é uma prática padrão da indústria, posicionando a caverna em domos salinos geomecanicamente competentes. Esta profundidade garante que a rocha salina esteja sob um estado de tensões confinantes suficiente para exibir seu comportamento dúctil e auto-cicatrizante, ao mesmo tempo que evita as pressões e temperaturas excessivas de profundidades maiores, que poderiam acelerar a convergência por fluência Li *et al.* (2022).

5.3.4.4 Abordagem de modelagem da temperatura

A temperatura é um fator cinético crucial, mas sua implementação na otimização expõe uma limitação do SALGAS e a solução metodológica adotada. O *software* opera sob uma condição isotérmica de 75°F ($\approx 24^\circ\text{C}$) e não simula a transferência de calor Eyerman (2008). Para incorporar o efeito da temperatura, a metodologia adota uma abordagem de *proxy*,

utilizando o Fator de Dissolução (CONDIS), que se baseia em relações empíricas Saberian (1983).

Para a otimização, foi definido um caso de temperatura base de 40°C. Esta escolha é fundamentada em análises geológicas: para as profundidades de caverna consideradas, a temperatura de formação *in-situ* é tipicamente de 30°C a 70°C, com base em gradientes geotérmicos de 25-30°C/km Bérest *et al.* (2020). O valor de 40°C é um representante realista e defensável. Embora casos de 60°C e 80°C tenham sido investigados na análise de sensibilidade, eles não fizeram parte do espaço de busca do AG, que operou sob a condição de temperatura fixa de 40°C para garantir a consistência dos resultados. Na implementação, a "temperatura" é uma variável acoplada: a definição de 40°C resulta em modificações no CONDIS e nas densidades do fluido, cuja interdependência foi confirmada pela Análise de Componentes Principais (PCA) desta tese.

5.3.5 Restrições Não-Lineares (Coerência Física e Segurança Geomecânica)

Estas restrições garantem que cada solução candidata gerada pelo AG corresponda a um caso fisicamente realizável e geomecanicamente seguro.

- Hierarquia de circulação direta: A operação em modo de circulação direta (injeção profunda, produção rasa) exige uma ordem geométrica específica das tubulações e do teto da caverna. Uma margem de segurança ϵ (um valor pequeno, e.g., 1 m) é usada para evitar singularidades numéricas e garantir separação física:

$$B_i - P_i + \epsilon \leq 0 \quad (5.2)$$

$$P_i - I_i + \epsilon \leq 0 \quad (5.3)$$

A primeira inequação garante que o ponto de produção está abaixo do teto da caverna (definido pelo *blanket*), e a segunda garante que o ponto de injeção está abaixo do ponto de produção.

- Espessura mínima do teto de sal: Para assegurar a integridade estrutural a longo prazo e prevenir o colapso do teto da caverna, uma espessura mínima de sal deve ser mantida acima do ponto mais alto da cavidade (B_i). A literatura técnica recomenda uma espessura mínima de sal de 150 m como uma diretriz segura para garantir a estabilidade Seright e Brattekas (1991). A restrição utilizada neste trabalho, de 100 m, representa

um limite mínimo absoluto, abaixo do qual os riscos se tornam inaceitáveis.

$$B_i \geq Z_{\text{topo_sal}} + 100 \quad (5.4)$$

Onde $Z_{\text{topo_sal}}$ é a profundidade do topo da formação salina.

A tabela 12 a seguir consolida as principais restrições geomecânicas derivadas da literatura, que fundamentam a formulação do problema de otimização e a avaliação da segurança das soluções geradas.

Tabela 12 – Restrições geomecânicas consideradas na otimização.

Parâmetro de <i>design</i>	Valor Considerado	Justificativa Técnica
Espessura mínima do teto de sal	≥ 100 m	Garante a integridade estrutural e o isolamento da caverna em relação a formações superiores, conforme diretrizes conservadoras da literatura Seright e Brattekas (1991).
Razão altura/diâmetro (H/D)	Faixa ideal: 1.5 a 5.0	Equilibra a estabilidade e o aproveitamento volumétrico, evitando tanto geometrias achatadas ($H/D \ll 1$) quanto cavidades excessivamente estreitas ("tipo lápis") Seright e Brattekas (1991).
Raio máximo do teto da caverna	≤ 9.1 m (30 ft)	Controla o vão não suportado no topo da caverna, o que é crucial para reduzir o risco de colapso associado à fluência plástica do sal Seright e Brattekas (1991).
Profundidade máxima de injeção	≤ 1000 m	Limita a extensão vertical da caverna para respeitar a base da camada de sal e garantir a viabilidade das condições operacionais do poço.
Margem à base do sal	≥ 50 m	Impede que o processo de dissolução avance para próximo da base da formação salina, evitando riscos estruturais e o contato com camadas geológicas indesejadas.

5.3.6 Função-Objetivo e Métricas de Desempenho

A avaliação de cada solução candidata gerada pelo algoritmo genético é governada por uma função-objetivo. A formulação desta função difere fundamentalmente entre a otimização mono-objetivo (SOO) e a multiobjetivo (MOO), refletindo a natureza distinta de cada abordagem de busca. Ambas, no entanto, dependem de um conjunto comum de quatro métricas de desempenho chave, calculadas a partir dos resultados da simulação SALGAS:

- Volume máximo (MV): O volume total da caverna ao final da simulação (em m^3).
- Massa de sal extraída (MS): A massa total de sal dissolvido (em ton).
- Eficiência energética (ME): A razão entre a massa de sal extraída e a energia consumida (em ton/kWh).
- Tempo total de construção (MT): A duração total da campanha de lixiviação (em dias).

A seleção destas quatro métricas não é arbitrária; cada uma representa um pilar fundamental que define o sucesso técnico e econômico de um projeto de mineração por dissolução. O volume máximo (MV) é a principal métrica de valor para projetos de armazenamento, pois o espaço útil criado constitui o ativo que gera receita. Maximizar o MV, portanto, instrui o algoritmo a encontrar estratégias que criem o maior espaço de armazenamento possível dentro das restrições de segurança. A massa de sal extraída (MS) é a métrica primária para projetos focados na produção de salmoura como matéria-prima para a indústria química, representando o produto principal a ser comercializado.

As outras duas métricas governam a viabilidade econômica e operacional do projeto. O tempo total de construção (MT) é um indicador direto do custo de capital (CAPEX) e da exposição ao risco. Projetos mais longos incorrem em maiores custos operacionais, de financiamento e postergam o início da geração de receita, impactando negativamente o Valor Presente Líquido (VPL) do ativo. Por fim, a eficiência energética (ME) é um indicador chave do custo operacional (OPEX), medindo a quantidade de sal extraído por unidade de energia consumida. Uma maior eficiência reduz os custos diretos de lixiviação e alinha o projeto a práticas de engenharia mais sustentáveis.

A otimização simultânea destas métricas força o algoritmo a navegar pelos compromissos (*trade-offs*) intrínsecos ao *design* de engenharia do mundo real, como o balanço entre a velocidade de construção (aumentando a vazão) e a eficiência energética.

Para garantir que estas métricas, que possuem escalas e unidades distintas, possam ser comparadas e combinadas de forma matematicamente consistente, elas são normalizadas para um intervalo adimensional de $[0, 1]$ com base em limites mínimos e máximos realistas definidos na configuração do sistema.

5.3.6.1 Formulação para Otimização Mono-Objetivo (ga)

Na otimização mono-objetivo (SOO), o desafio é consolidar os múltiplos critérios de desempenho de um projeto em um único indicador numérico que possa guiar o algoritmo. Esse indicador, no contexto dos Algoritmos Genéticos, é conhecido como aptidão (*fitness*), um valor escalar que quantifica a "qualidade" de uma determinada solução candidata.

Para esta tarefa, foi utilizado o *solver ga* do *Global Optimization Toolbox* do MATLAB, uma implementação computacional de um Algoritmo Genético projetada para problemas com um único objetivo. Uma característica fundamental do *solver ga* é que ele opera como um minimizador: seu propósito é encontrar a solução que resulta no menor valor de aptidão possível.

Dada essa característica, a função de aptidão, implementada na função `fitnessFunction.m`, foi formulada para que valores menores indiquem soluções melhores. Para isso, as métricas de desempenho normalizadas (M'_k) são agregadas através de uma função de soma ponderada. A função-objetivo é definida como o negativo da soma ponderada dos objetivos que se deseja maximizar, acrescida do complemento para o objetivo de minimização e de uma função de penalidade (Π_{total}) que desencoraja soluções que violem as restrições:

$$Fitness(\vec{x}) = - \left(\sum_{k \in \{MV, MS, ME\}} w_k M'_k + w_{MT}(1 - M'_{MT}) \right) + \Pi_{\text{total}}(\vec{x}) \quad (5.5)$$

onde w_k são os pesos que refletem a importância relativa de cada objetivo e $\Pi_{\text{total}}(\vec{x})$ representa a soma de todas as penalidades. As principais penalidades implementadas incluem:

- Penalidade da razão altura/diâmetro ($\Pi_{H/D}$): Garante a estabilidade geomecânica da caverna, penalizando geometrias fora da faixa ideal (e.g., $[1.5, 5.0]$). Matematicamente, é expressa como:

$$\Pi_{H/D}(\vec{x}) = p_{H/D} \cdot (\max(0, H/D_{\min} - H/D(\vec{x})) + \max(0, H/D(\vec{x}) - H/D_{\max}))$$

onde $p_{H/D}$ é um coeficiente de penalidade, e H/D_{\min} e H/D_{\max} são os limites da faixa ideal Seright e Brattekas (1991), Chen e Li (2024).

- Penalidade de intervenção (Π_{workover}): Adiciona um custo fixo se ocorrer alteração nas profundidades de operação (B_i, I_i, P_i) entre os estágios, o que na prática exigiria uma intervenção de sonda (*workover*) com custo significativo.

5.3.6.2 Formulação para Otimização Multiobjetivo (NSGA-II/gamultiobj)

Diferentemente da abordagem mono-objetivo, a otimização multiobjetivo com o *solver* gamultiobj, que implementa o algoritmo NSGA-II, não utiliza uma função de soma ponderada. Em vez disso, ele trata cada métrica de desempenho como um objetivo independente a ser otimizado simultaneamente. A formulação busca encontrar um conjunto de soluções que formam a Fronteira de Pareto, representando os melhores compromissos (*trade-offs*) possíveis entre os quatro objetivos de engenharia definidos:

1. Maximizar o volume máximo (MV)
2. Maximizar a massa de sal extraída (MS)
3. Maximizar a eficiência energética (ME)
4. Minimizar o tempo total de construção (MT)

Por convenção, o *solver* gamultiobj minimiza todos os objetivos. Portanto, as métricas que desejamos maximizar (MV, MS, ME) são retornadas com sinal negativo. A função-objetivo, conforme implementada em `fitnessFunctionMulti.m`, retorna o seguinte vetor de quatro objetivos, onde cada um já inclui as penalidades (Π_{total}) por violação de restrições:

$$\text{Vetor de Objetivos} = \begin{pmatrix} -M'_{MV} + \Pi_{\text{total}} \\ -M'_{MS} + \Pi_{\text{total}} \\ -M'_{ME} + \Pi_{\text{total}} \\ M'_{MT} + \Pi_{\text{total}} \end{pmatrix} \quad (5.6)$$

Nesta formulação, as penalidades garantem que soluções que violam restrições sejam consideradas dominadas por soluções viáveis, guiando a busca em direção à região factível do espaço de busca.

5.4 IMPLEMENTAÇÃO COMPUTACIONAL E ARQUITETURA DO OTIMIZADOR

A solução do problema de otimização formulado requer um *framework* computacional que integre, de forma eficiente, o processo de busca do Algoritmo Genético com as avaliações da função-objetivo, que são fornecidas pelo simulador de lixiviação SALGAS. Esta seção detalha a arquitetura deste sistema e justifica a configuração dos operadores genéticos utilizados no ambiente MATLAB.

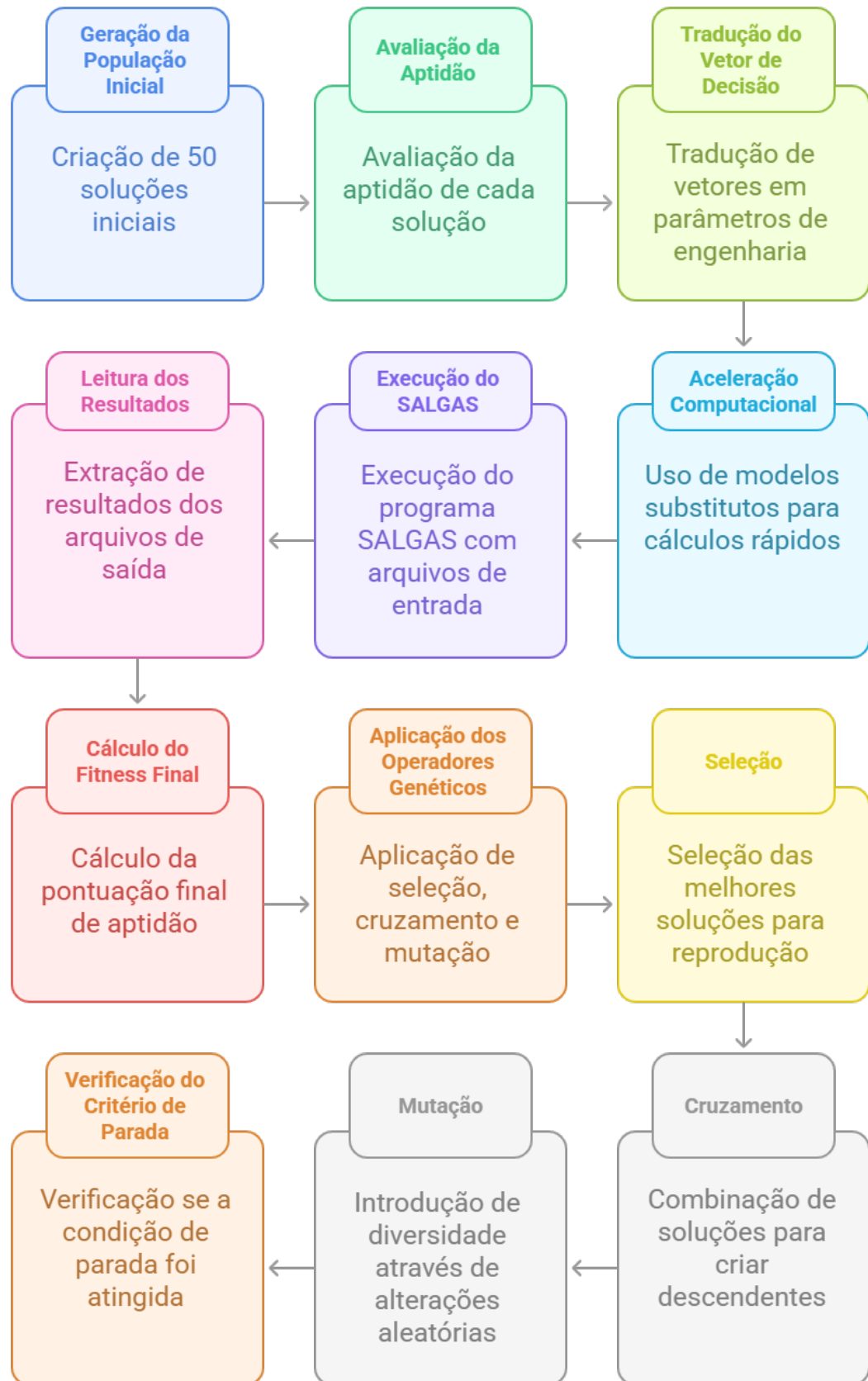
Para garantir a total reprodutibilidade e permitir a análise detalhada da metodologia, o código-fonte completo deste *framework*, incluindo os *scripts* para as otimizações mono e multiobjetivo, foi disponibilizado em um repositório público no GitHub, que pode ser acessado em:

<<https://github.com/OscarMelgar40/tese-otimizacao-cavernas-salinas>>

5.4.1 O *framework* de Otimização: Acoplando SALGAS e AG

A arquitetura do otimizador é construída em torno de um laço iterativo onde o Algoritmo Genético (AG) e o simulador SALGAS se comunicam. O processo, implementado nos *scripts* `iniciar_otimizacao.m` e `iniciar_otimizacao_multiobjetivo.m`, pode ser descrito da seguinte forma. A figura 10 ilustra este fluxo de trabalho, detalhando cada passo do processo.

Figura 10 – Fluxograma do processo de otimização acoplando o AG ao SALGAS.



Fonte: O autor (2025)

1. geração da população: o AG inicia criando uma população de indivíduos. Cada indivíduo é um vetor de decisão \vec{x} (conforme a Equação 7.2), representando um *design* completo da campanha de lixiviação em dois estágios.
2. avaliação da aptidão (*fitness*): para cada indivíduo na população, o *framework* executa uma avaliação da função-objetivo. Este é o passo mais crítico e computacionalmente intensivo do processo.
 - o vetor de decisão \vec{x} é traduzido em um arquivo de entrada (.sgi) para o SALGAS, contendo todos os parâmetros operacionais (vazões, tempos, profundidades, etc.) para os dois estágios de simulação Warren (2016).
 - o simulador SALGAS é executado como um processo *black-box*. O *framework* invoca o simulador, aguarda a conclusão da simulação e então extrai os resultados relevantes do arquivo de saída.
 - os resultados da simulação (e.g., volume final da caverna, tempo total, etc.) são usados para calcular as métricas de desempenho e, consequentemente, a aptidão do indivíduo. As restrições geomecânicas e operacionais são verificadas através da função *restricoes.m*, e as penalidades correspondentes são aplicadas.
3. operadores genéticos: com base na aptidão de toda a população, o AG aplica seus operadores estocásticos para criar a próxima geração. As escolhas específicas dos operadores são cruciais para a eficiência da busca:
 - seleção: o mecanismo de escolha dos "pais" difere entre as abordagens. Para a otimização mono-objetivo (ga), utiliza-se a seleção por torneio baseada no valor de *fitness* escalar. Para a multiobjetivo (gamultiobj), o critério é a dominância de Pareto e a distância de aglomeração, conforme o algoritmo NSGA-II.
 - cruzamento (*crossover*): pares de pais são recombinados usando o operador *crossoverintermediate*, ideal para variáveis contínuas. Este operador cria "filhos" que herdam uma mistura ponderada das características dos progenitores, promovendo a *exploitation*.
 - mutação: uma pequena porção dos filhos sofre alterações aleatórias através do operador *@mutationadaptfeasible*. Esta escolha é fundamental, pois o operador adapta-se ao problema e garante que as mutações sempre respeitem as restrições

de contorno (lb e ub), aumentando drasticamente a eficiência da busca ao evitar a geração de indivíduos inviáveis.

4. convergência: o ciclo de avaliação e aplicação de operadores genéticos é repetido até que um critério de parada seja atingido, como o número máximo de gerações (MaxGenerations) ou a estagnação da melhor solução por um número definido de gerações (MaxStallGenerations).

O principal desafio desta abordagem é o custo computacional. Em nosso ambiente de testes, uma única avaliação da função-objetivo, que envolve uma simulação completa dos dois estágios no SALGAS, levou em média 45 segundos para ser concluída. Consequentemente, para uma população de 25 indivíduos, executada por 10 gerações, o número total de avaliações é de 250 simulações, demandando um tempo de execução considerável que justifica a otimização do processo.

Isso torna imperativo o uso de algoritmos eficientes e a implementação de técnicas de computação paralela. É importante esclarecer que a paralelização não ocorre dentro do simulador SALGAS, que é executado como um processo serial individual. Em vez disso, o framework de otimização acelera a busca ao paralelizar as avaliações da função de aptidão. Com a opção `UseParallel = true`, o solver ga do MATLAB distribui os diferentes indivíduos de uma mesma geração para serem avaliados simultaneamente em múltiplos núcleos de processamento. Cada núcleo executa uma instância independente do SALGAS, o que reduz significativamente o tempo total necessário para avaliar uma população inteira Mello, Lacerda e Pozo (2023).

5.4.2 Estratégia de Aceleração Computacional: Modelos Substitutos para Propriedades de Fluidos

Um dos principais gargalos computacionais no acoplamento de um otimizador ao SALGAS é a determinação dos parâmetros de entrada para cada avaliação da função-objetivo. Conforme detalhado no Apêndice D, o cálculo da gravidade específica (SG) do fluido de injeção e da salmoura saturada utilizando ferramentas externas como o *Toolbox do SMRI* é um processo manual e lento, inviável para as milhares de iterações exigidas por um Algoritmo Genético.

Para superar essa limitação e viabilizar a otimização, uma estratégia de modelagem de substitutos (*surrogate modeling*) foi implementada. Em vez de recalcular as propriedades dos fluidos a cada chamada, modelos de regressão polinomial de segunda ordem (poly22 no MATLAB) foram pré-ajustados com base nos dados experimentais apresentados na Tabela 6.

Esses modelos foram implementados na função `calcular_parametros_fluido.m` e seus coeficientes armazenados na estrutura de configuração do otimizador.

Essa abordagem proporciona uma aproximação de alta fidelidade e computacionalmente instantânea das propriedades dos fluidos, eliminando um dos principais gargalos do processo. Ao substituir uma etapa lenta e externa por um modelo matemático local, a avaliação da função-objetivo torna-se significativamente mais eficiente, permitindo a exploração de um número muito maior de soluções candidatas dentro de um tempo computacional viável.

5.4.3 Configuração dos Operadores Genéticos no Ambiente MATLAB

A implementação da otimização nesta tese abrange tanto a otimização mono-objetivo (SOO), focada em encontrar uma única solução ótima, quanto a otimização multiobjetivo (MOO), que busca um conjunto de soluções de compromisso (fronteira de Pareto). Ambas as abordagens foram implementadas utilizando o *Global Optimization Toolbox* The MathWorks, Inc. (2024a) do MATLAB, respectivamente com os *solvers* `ga` e `gamultiobj`.

Embora os mecanismos de avaliação e seleção dos indivíduos sejam fundamentalmente diferentes entre as duas abordagens, o motor evolutivo central — que governa a geração de novas soluções através de cruzamento e mutação — compartilha uma base de parâmetros comum. A principal distinção reside no critério de seleção:

- No caso mono-objetivo (`ga`), os indivíduos são avaliados e selecionados com base em um único valor de aptidão escalar, que agrega todos os objetivos e penalidades.
- No caso multiobjetivo (`gamultiobj`), a seleção é governada pelos princípios da dominância de Pareto e pela manutenção da diversidade. O algoritmo utiliza a ordenação rápida por não dominância para classificar as soluções em "frentes" e a distância de aglomeração para garantir uma distribuição uniforme ao longo da fronteira de Pareto.

A tabela a seguir consolida os parâmetros de configuração utilizados, detalhando sua função e justificativa em ambos os casos de otimização. Esta abordagem unificada destaca as similaridades no núcleo do processo evolutivo, ao mesmo tempo que esclarece as nuances específicas de cada método.

Tabela 14 – Parâmetros de configuração do Algoritmo Genético (*solvers* ga e gamultiobj).

Parâmetro (Propriedade MA-TLAB)	Valor Utilizado	Justificativa e Considerações (SOO & MOO)
Configuração da população e duração		
Tamanho da população (PopulationSize)	25	Essencial para a diversidade genética. Na MOO, é crítico para bem representar a fronteira de Pareto.
Máximo de gerações (MaxGenerations)	10	Número máximo de iterações, balanceando a convergência da solução com o custo computacional.
Mecanismos de seleção e sobrevivência		
Mecanismo de seleção	Específico do <i>solver</i>	ga: Seleção por torneio (baseada em <i>fitness</i> único). gamultiobj: Ordenação por não dominância e distância de aglomeração (NSGA-II).
Elitismo	Específico do <i>solver</i>	ga: EliteCount (10%) preserva os melhores indivíduos. gamultiobj: Inerente ao mecanismo, que sempre preserva as melhores frentes não dominadas.
Operadores de geração de soluções		
Fração de cruzamento (CrossoverFraction)	0.8	Principal motor da <i>exploitation</i> , gerando 80% dos novos indivíduos a partir de pais promissores.
Operador de mutação (MutationFcn)	@mutationadaptfeasible	Principal motor da <i>exploração</i> . Garante que as soluções geradas respeitem as restrições de contorno.
Configuração da execução e restrições		
Função de restrições não lineares	@restricoes	Garante a viabilidade física e a segurança dos projetos, aplicando restrições geomecânicas (e.g., razão H/D).
Uso de paralelismo (UseParallel)	true	Acelera a otimização ao distribuir as avaliações de <i>fitness</i> , sendo crucial para a viabilidade computacional do estudo.

6 RESULTADOS E DISCUSSÃO

Este capítulo apresenta e discute os principais resultados gerados ao longo desta tese. A apresentação é dividida em duas partes principais, seguindo a lógica da pesquisa. Primeiramente, são detalhados os resultados da fase de análise, que incluem as simulações numéricas do processo de lixiviação com o software SALGAS e a subsequente análise estatística multivariada dos dados com as técnicas PCA e t-SNE. Em seguida, são apresentados os resultados da fase de otimização, demonstrando as soluções de design obtidas pelo *framework* que acopla algoritmos genéticos ao simulador, tanto para o caso mono-objetivo quanto para o multiobjetivo.

6.1 RESULTADOS COM SALGAS

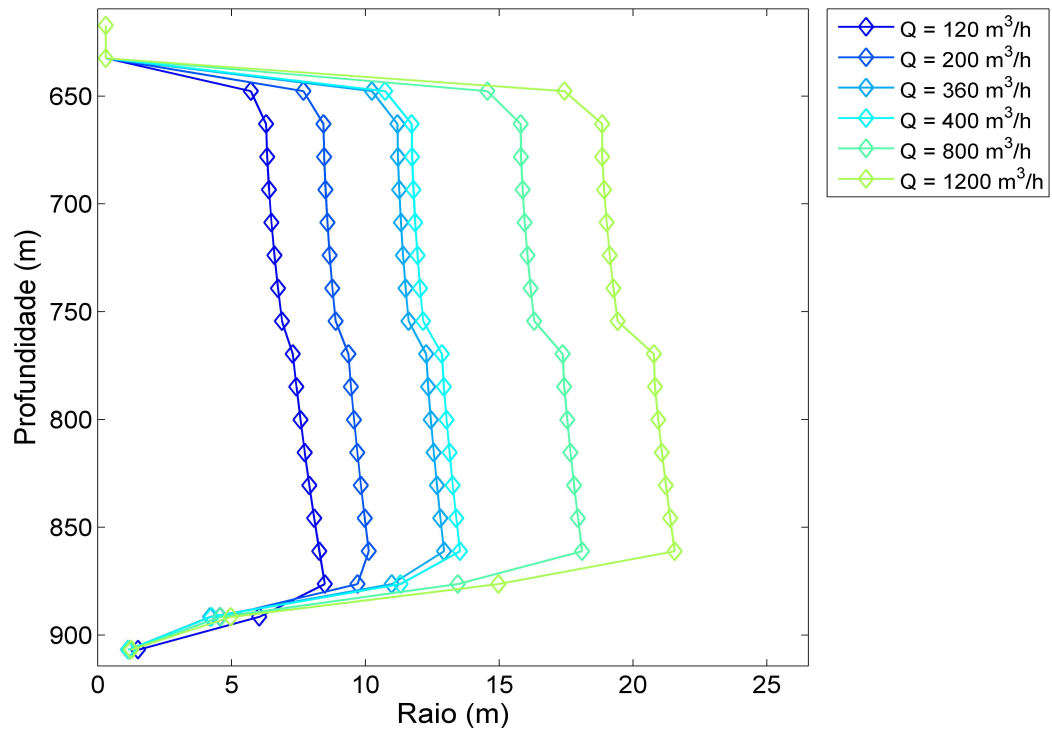
Com base nas simulações realizadas no *software* SALGAS, foram feitas comparações dos contornos finais das cavernas e dos volumes ao final da simulação para os casos *D* (direto) e *R* (reverso). Com as informações obtidas, foram geradas superfícies de resposta que ilustram visualmente o comportamento do volume em ambos os casos, mediante a técnica discutida em Amidror (2002). Outra análise realizada foi a comparação dos volumes entre os dois métodos de dissolução.

O comportamento da caverna foi analisado considerando a variação da taxa de produção de salmoura e das temperaturas de referência (40°C, 60°C e 80°C) ao longo de 120 dias. As figuras de 11 a 17 apresentam os resultados para as simulações do caso *D*.

A comparação entre os contornos finais da caverna para o tempo de 120 dias no caso *D*, considerando cada caso analisado, é apresentada nas figuras de 11 a 13. Nelas, foi possível verificar a geometria e o valor do raio atingido. Para uma mesma condição de temperatura, observa-se que, quanto maior a taxa de produção, maior o raio da caverna. O valor máximo alcançado em 120 dias foi de 26,90 m de raio no caso com temperatura de 80°C e taxa de produção de 1200 m³/h. Analisando a mesma taxa de produção, o raio da caverna também aumenta com o aumento da temperatura de referência, devido à maior solubilidade do cloreto de sódio em temperaturas elevadas, o que acelera a dissolução. Contudo, a magnitude dessa variação é visivelmente menor do que a observada ao se alterar a taxa de produção, como se pode constatar pela maior separação entre as curvas de vazão nas figuras. Isso indica

que, dentro das faixas analisadas, a taxa de produção é a variável dominante no controle do diâmetro final da caverna.

Figura 11 – Comparação dos contornos finais da caverna para o caso de referência $T = 40^{\circ}\text{C}$ no caso D.



Fonte: O autor (2025)

Figura 12 – Comparação dos contornos finais da caverna para o caso de referência $T = 60^{\circ}\text{C}$ no caso D.

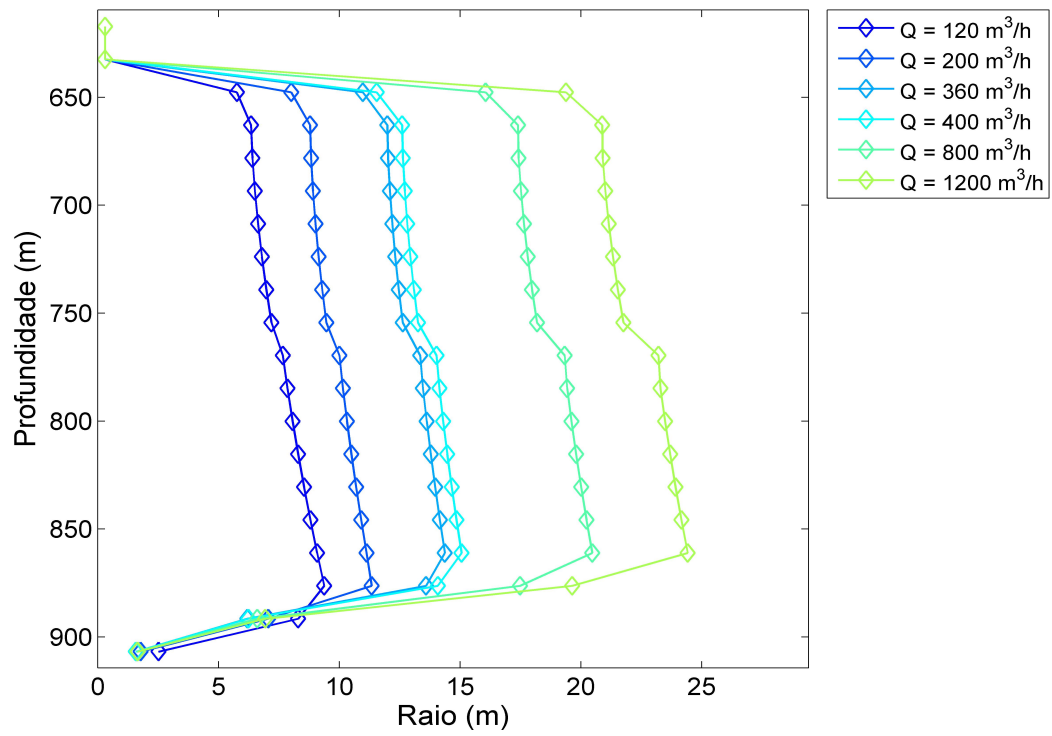
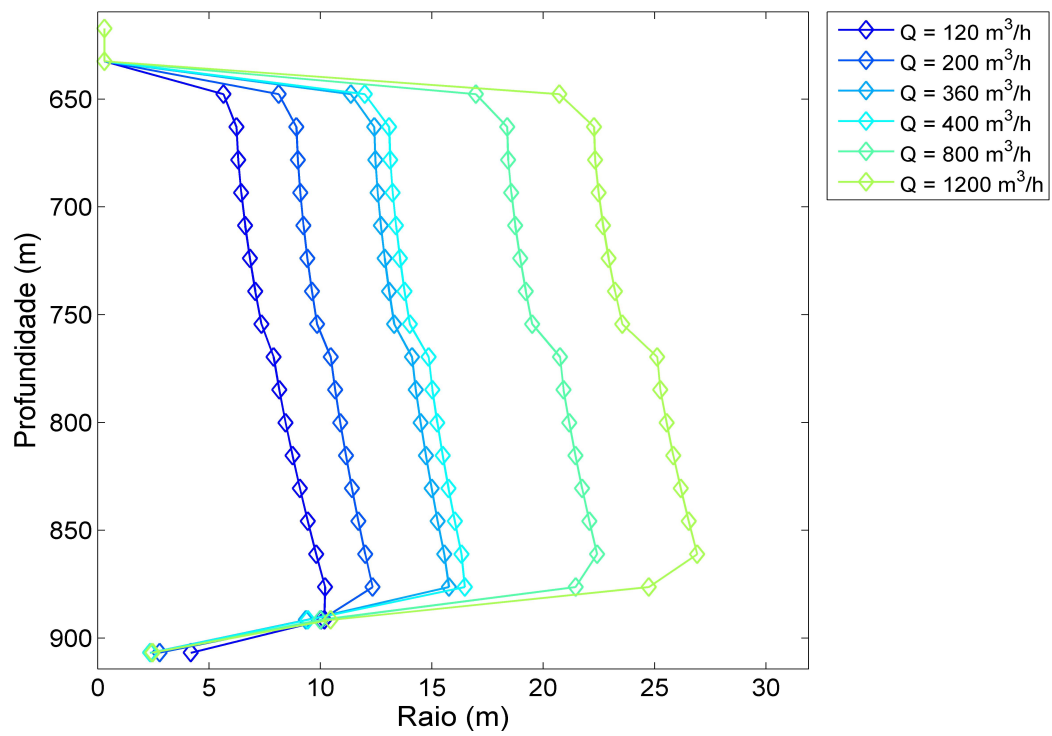


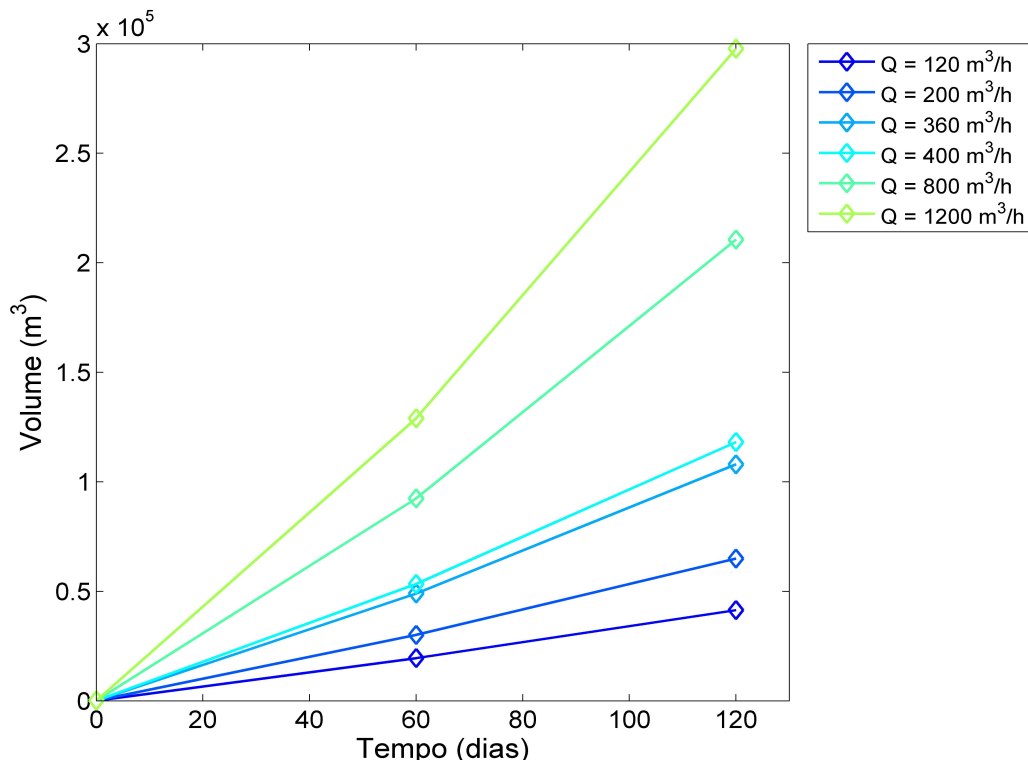
Figura 13 – Comparação dos contornos finais da caverna para o caso de referência $T = 80^{\circ}\text{C}$ no caso D.



As figuras de 14 a 16 apresentam os resultados de evolução do volume, agrupados por caso

de temperatura de referência. A análise mostra que, para uma mesma condição de temperatura, o volume da caverna cresce mais rapidamente com o aumento da taxa de produção. Isso ocorre pois uma maior vazão intensifica a turbulência no interior da cavidade, o que eleva o coeficiente de transferência de massa e, conseqüentemente, acelera a taxa de dissolução do sal. Além disso, analisando a influência dos diferentes casos de temperatura, observa-se que, para uma mesma taxa de produção, quanto maior a temperatura de referência, maior é o volume final da caverna. Isso se deve ao aumento da solubilidade do cloreto de sódio com a temperatura, fazendo com que o solvente dissolva a rocha salina de forma mais efetiva.

Figura 14 – Evolução do volume da caverna para o caso $T = 40^{\circ}\text{C}$ no caso D.



Fonte: O autor (2025)

Figura 15 – Evolução do volume da caverna para o caso $T = 60^{\circ}\text{C}$ no caso D.

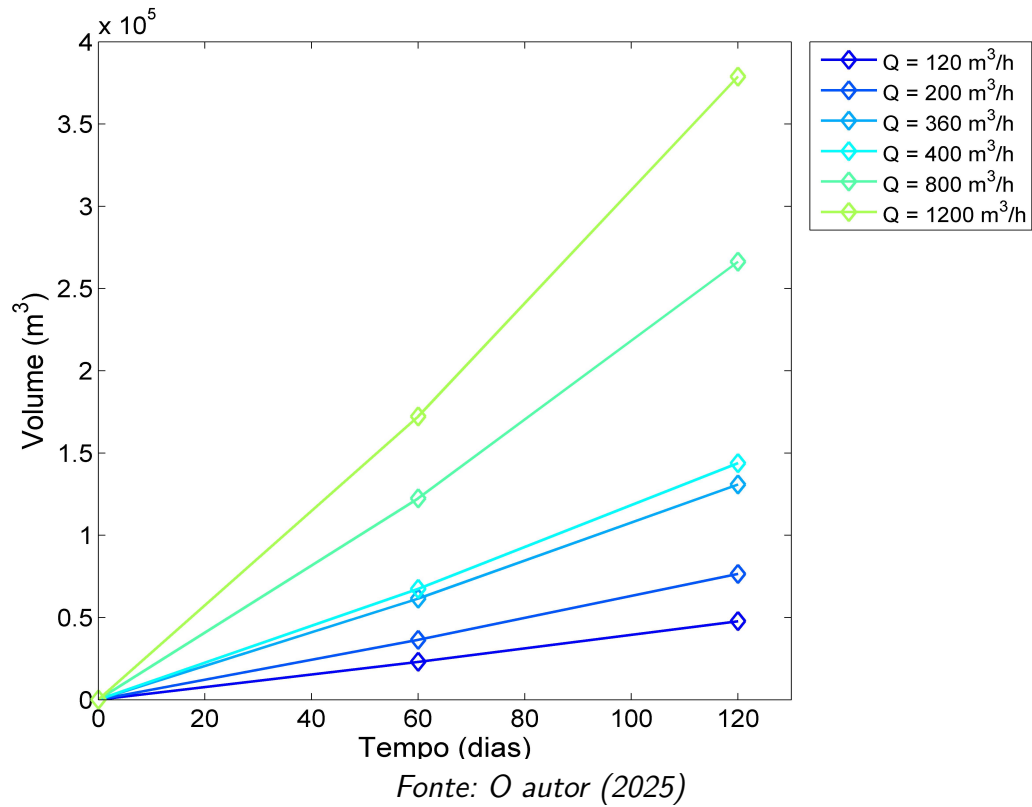
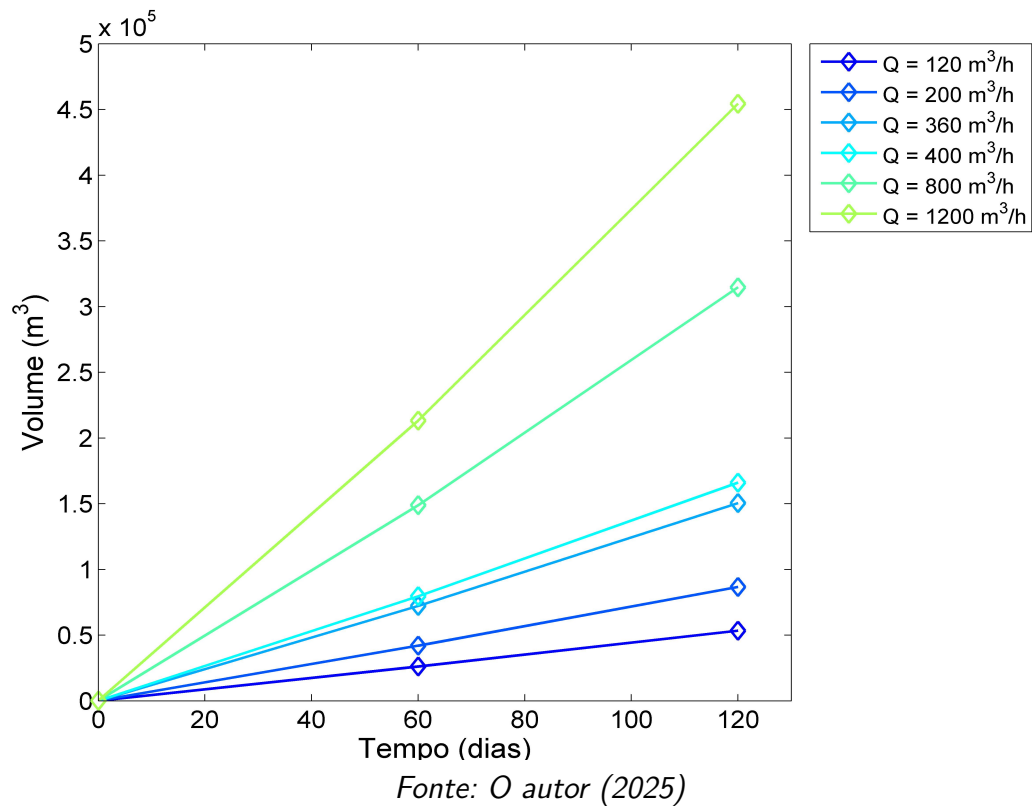


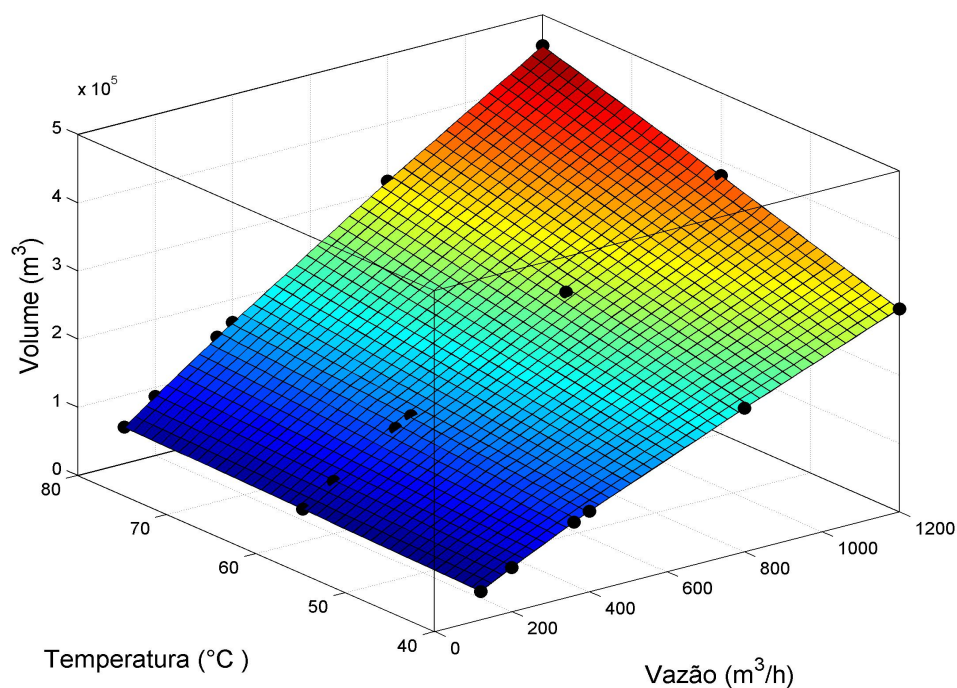
Figura 16 – Evolução do volume da caverna para o caso $T = 80^{\circ}\text{C}$ no caso D.



Com base nos resultados das análises para as diferentes vazões e temperaturas ao final de

120 dias, foi gerada uma superfície de resposta para ilustrar visualmente o comportamento do volume. Nessa superfície, as variáveis independentes são a vazão e a temperatura, enquanto a variável dependente é o volume, conforme ilustrado na figura 17. A técnica utilizada para criar esta superfície foi a interpolação linear baseada em triangulação, uma metodologia para dados esparsos descrita em Amidror (2002). Este método consiste em criar uma rede de triângulos a partir dos pontos de dados simulados e, em seguida, interpolar linearmente o volume dentro de cada triângulo para gerar uma superfície contínua, conforme detalhado no Apêndice B.

Figura 17 – Superfície de resposta para o volume no caso D.



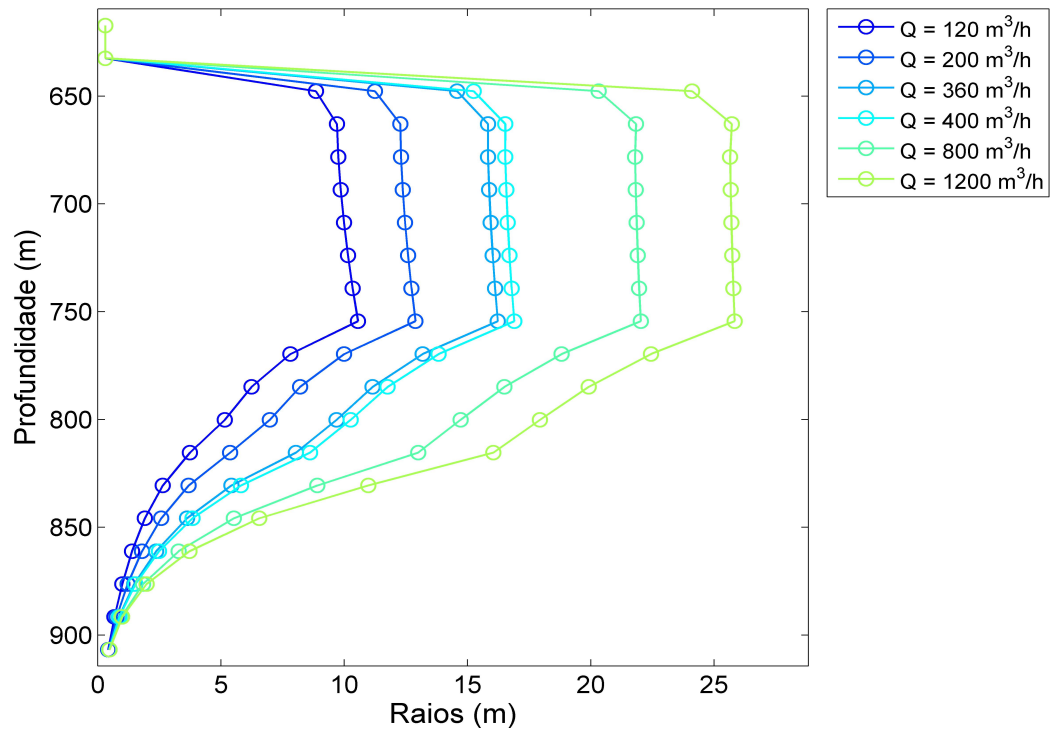
Fonte: O autor (2025)

De modo semelhante, foram feitas as análises para o caso R, cujos resultados são expostos nas figuras de 18 a 24.

A comparação entre os contornos finais da caverna para o tempo de 120 dias no caso R é apresentada nas figuras de 18 a 20, onde foi possível verificar a geometria e o valor do raio atingido. Similarmente ao caso D, para uma mesma condição de temperatura, quanto maior a taxa de produção, maior o raio da caverna. O valor máximo alcançado em 120 dias foi de 33,70 m de raio no caso com temperatura de 80°C e taxa de produção de 1200 m³/h. Analisando a mesma taxa de produção, o raio da caverna aumenta com o aumento da temperatura, uma vez que a maior temperatura eleva a solubilidade do sal, acelerando o processo de dissolução. Contudo, a análise visual das figuras demonstra que a magnitude desta variação é menor do

que a causada pela alteração da taxa de produção, sugerindo que esta última variável tem maior influência no diâmetro final obtido.

Figura 18 – Comparação dos contornos finais da caverna para o caso de referência $T = 40^{\circ}\text{C}$ no caso R.



Fonte: O autor (2025)

Figura 19 – Comparação dos contornos finais da caverna para o caso de referência $T = 60^{\circ}\text{C}$ no caso R.

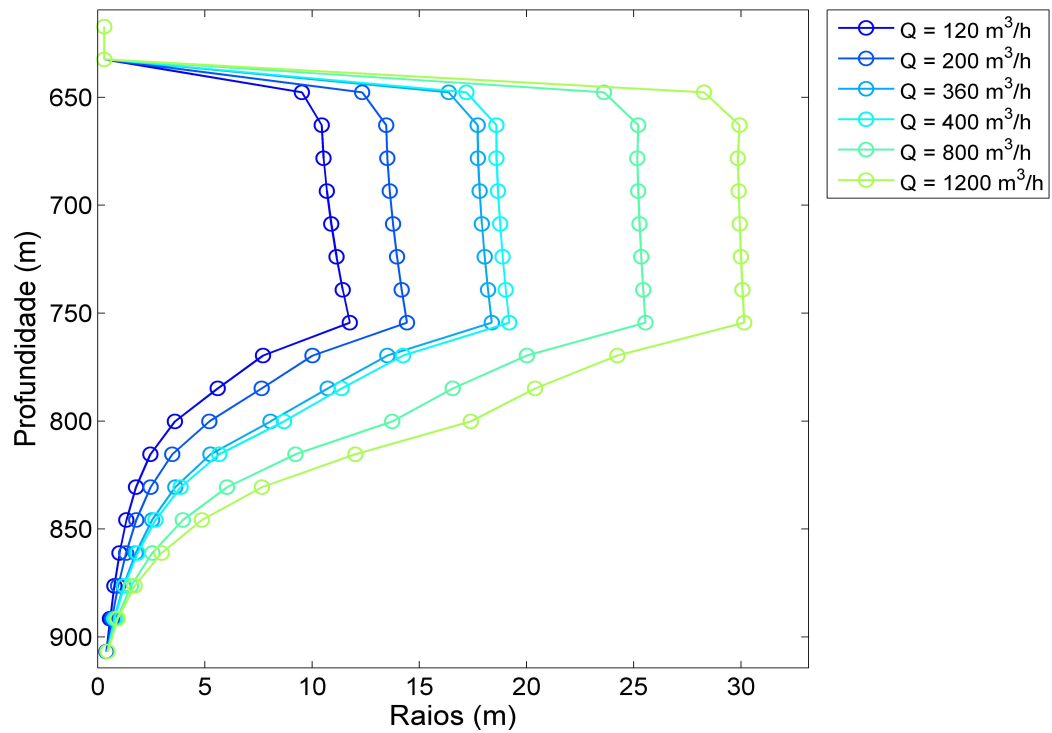
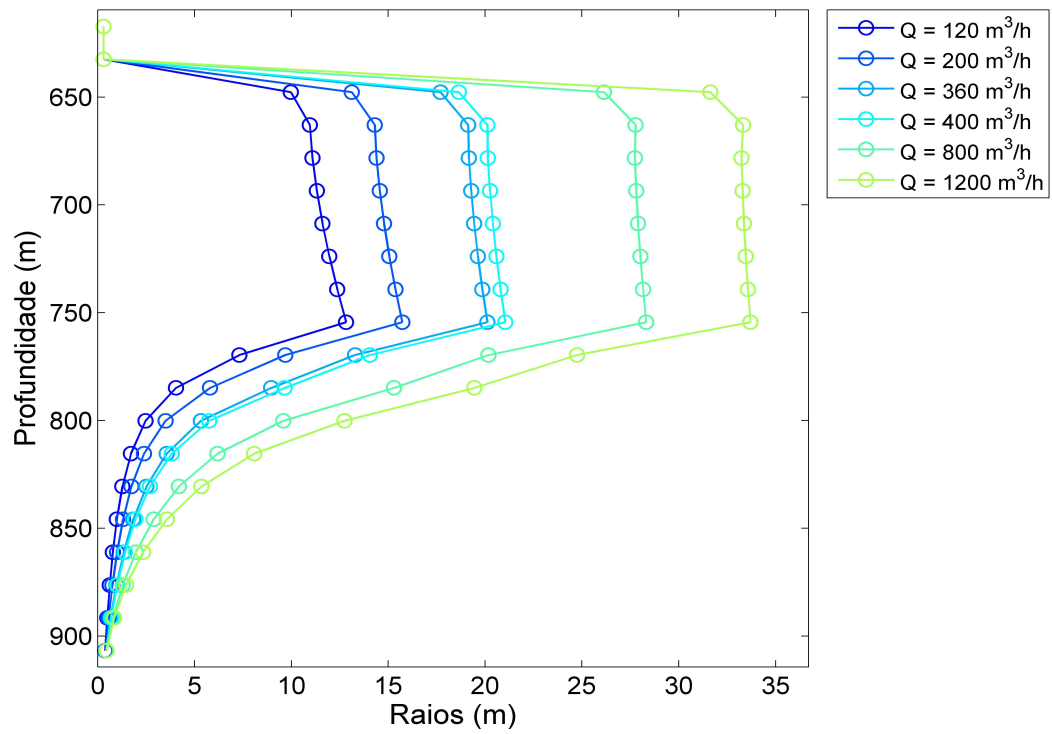


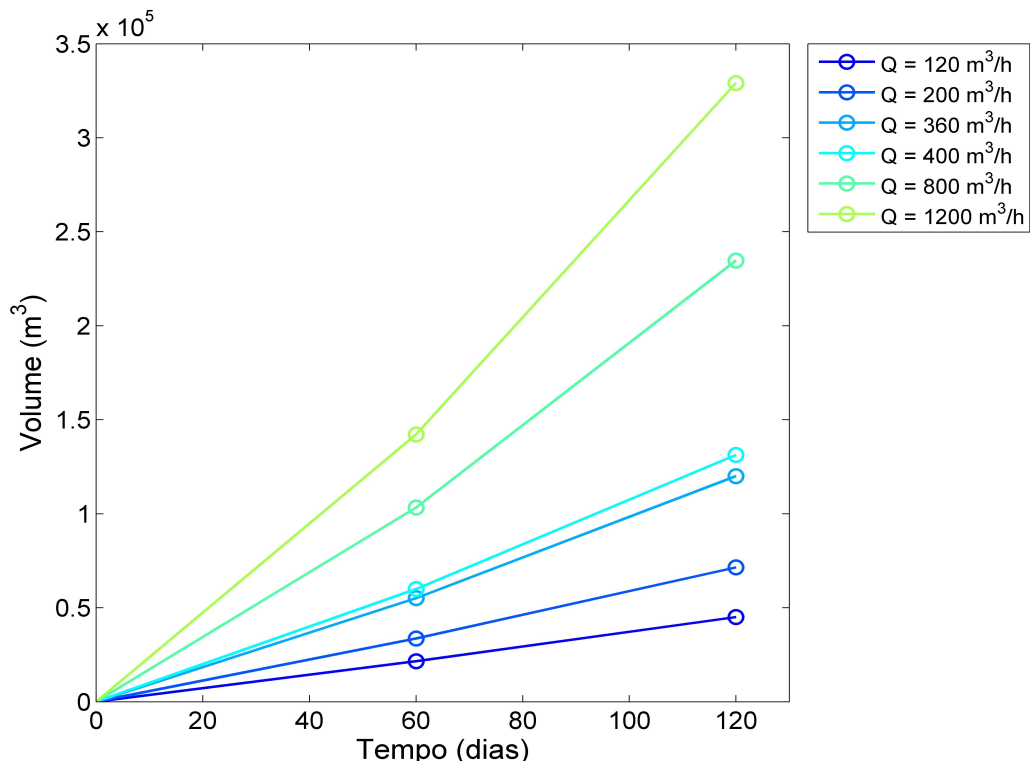
Figura 20 – Comparação dos contornos finais da caverna para o caso de referência $T = 80^{\circ}\text{C}$ no caso R.



As figuras de 21 a 23 apresentam os resultados de evolução do volume para cada tempe-

ratura de referência. Para uma mesma condição de temperatura, o volume cresce mais rapidamente com taxas de produção mais altas. De igual modo ao ocorrido no caso D, observa-se que, para uma mesma taxa de produção, quanto maior a temperatura de referência, maior o volume da caverna, devido ao aumento da solubilidade do sal.

Figura 21 – Evolução do volume da caverna para o caso $T = 40^{\circ}\text{C}$ no caso R.



Fonte: O autor (2025)

Figura 22 – Evolução do volume da caverna para o caso $T = 60^{\circ}\text{C}$ no caso R.

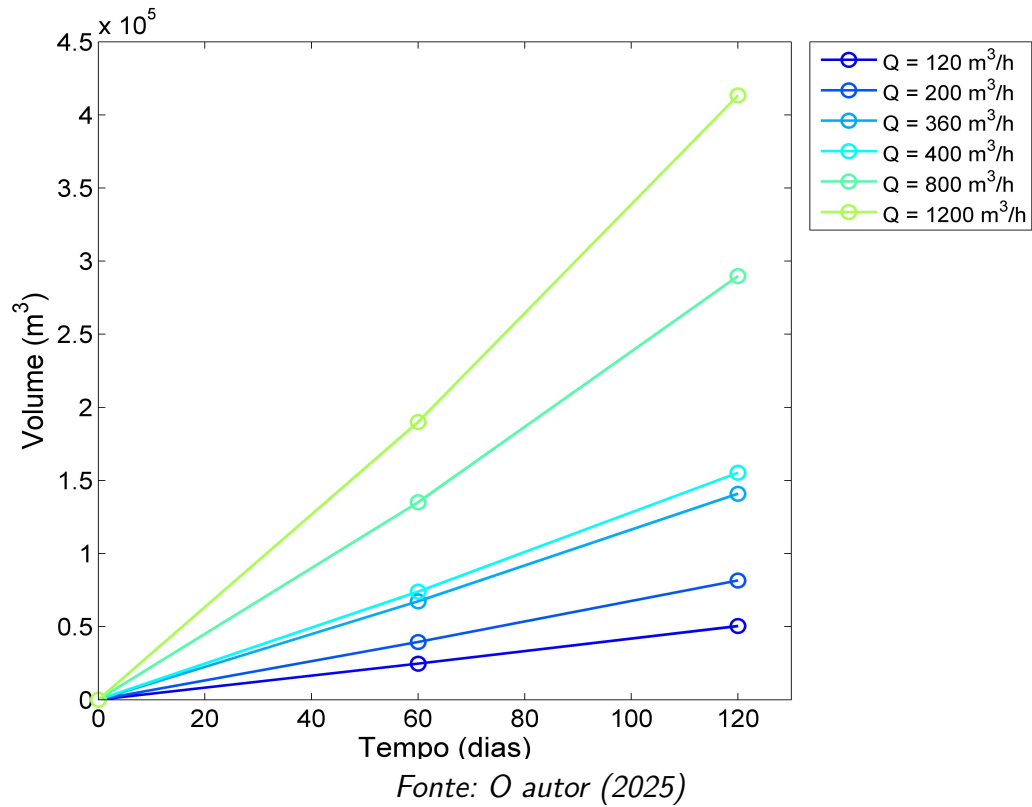
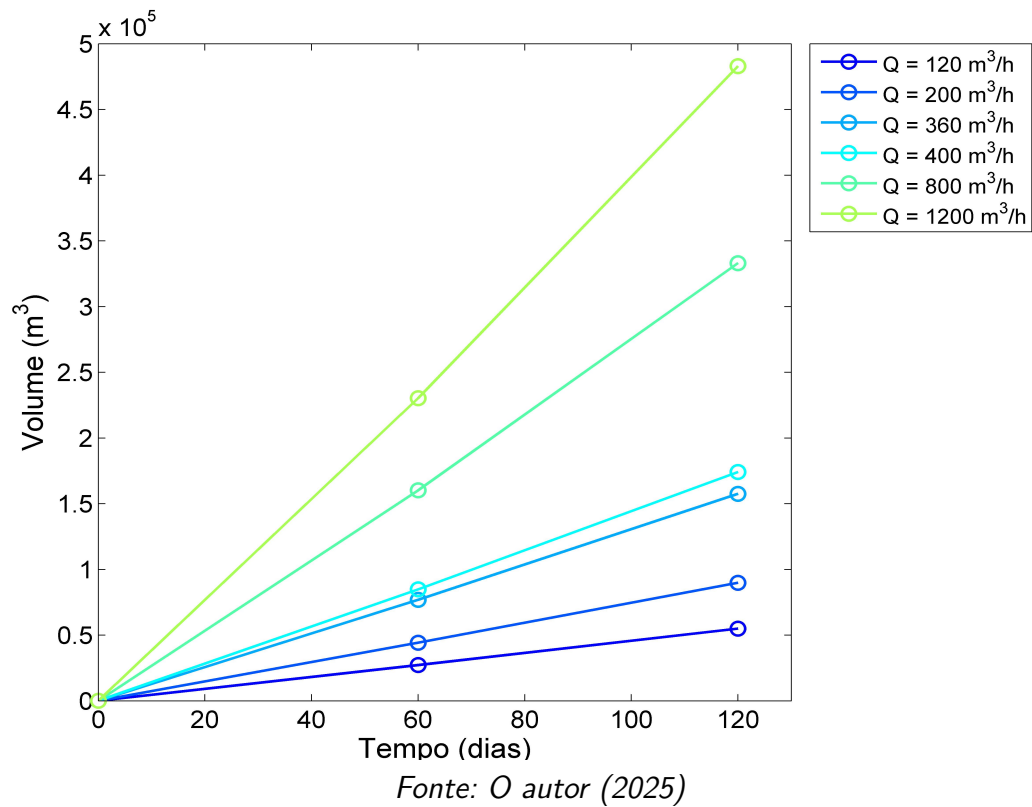


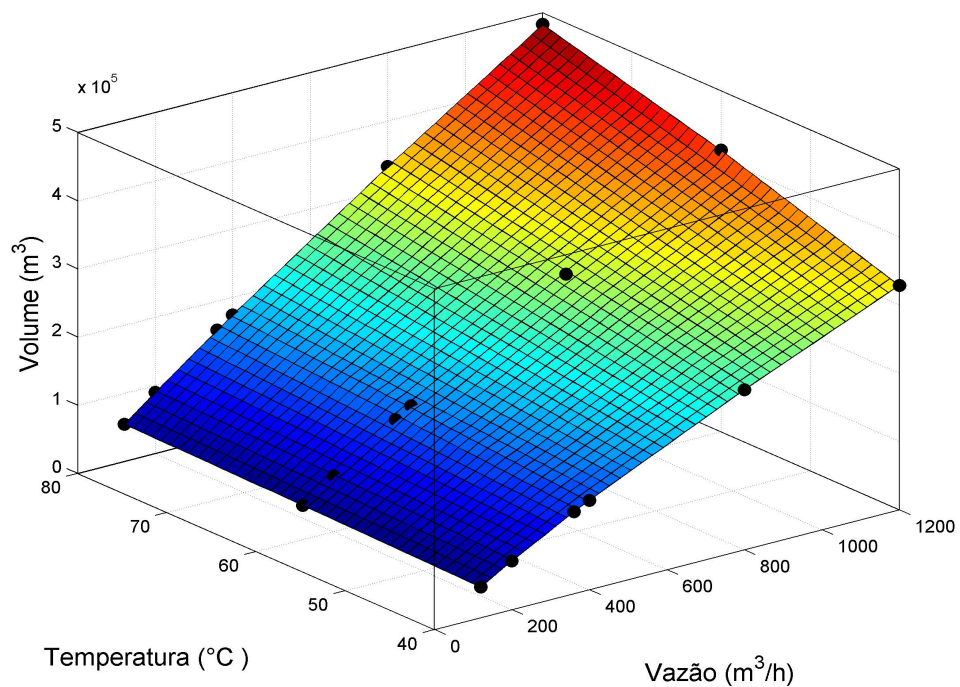
Figura 23 – Evolução do volume da caverna para o caso $T = 80^{\circ}\text{C}$ no caso R.



De forma similar ao que foi feito para o caso D, uma superfície de resposta foi criada para

o caso R. Considerando as diferentes vazões, temperaturas e o tempo final de 120 dias, a superfície foi extrapolada com os dados de saída do SALGAS. As variáveis independentes são a vazão e a temperatura, e a variável dependente é o volume, como mostrado na figura 24.

Figura 24 – Superfície de resposta para o volume no caso R.



Fonte: O autor (2025)

A tabela 15 apresenta a diferença percentual de volume entre os casos R (reverso) e D (direto). Em todos os casos analisados, o caso R resulta em um volume final consistentemente maior. A principal razão para essa maior eficiência reside na dinâmica dos fluidos: na circulação reversa, a água doce (menos densa) é injetada no topo e desce pelas paredes, enquanto a salmoura (mais densa) se acumula naturalmente no fundo para ser extraída. Isso cria uma célula de convecção estável que aproveita toda a altura da caverna para a dissolução. Em contraste, na circulação direta, a água doce injetada no fundo tende a subir rapidamente, o que pode causar um "curto-circuito" no fluxo, reduzindo o contato efetivo com as paredes da caverna e, consequentemente, a eficiência do processo.

Observa-se também na tabela uma relação decrescente entre a temperatura e a vantagem percentual do caso R. Ou seja, a superioridade do método reverso é mais pronunciada em temperaturas mais baixas. Isso pode ser explicado pelo fato de que, em temperaturas mais altas, a taxa de reação química (solubilidade do sal) se torna tão elevada que passa a dominar sobre os efeitos da eficiência da circulação. Mesmo o método D, menos eficiente, torna-se

muito mais reativo com o aumento do calor, o que diminui a diferença relativa de performance entre os dois casos.

Finalmente, nota-se uma particularidade no caso de 40°C, onde a tendência de aumento da diferença percentual se inverte na maior vazão (1200 m³/h), sugerindo interações mais complexas entre vazão e temperatura em regimes de menor energia térmica.

Tabela 15 – Diferença percentual dos volumes entre os casos Direto e Reverso.

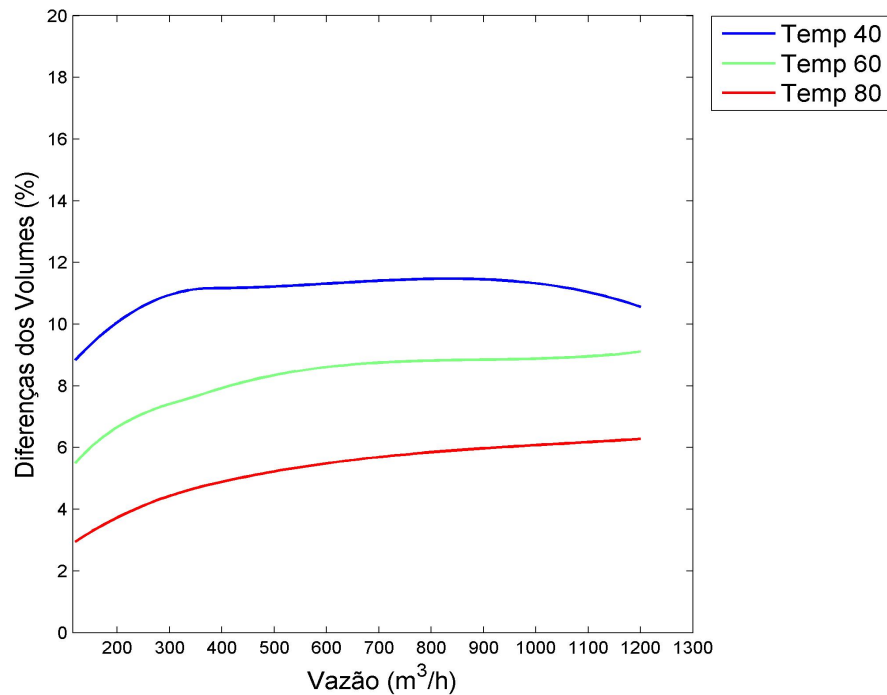
Vazão (m ³ /h)	Volume Direta (m ³)	Volume Reversa (m ³)	Diferença (%)
<i>caso de Temperatura = 40°C</i>			
120	41384,70	45041,05	8,84
200	64977,89	71511,52	10,06
360	107985,01	120021,52	11,15
400	118130,44	131330,95	11,17
800	210546,86	234701,06	11,47
1200	297725,84	329164,61	10,56
<i>caso de Temperatura = 60°C</i>			
120	47798,02	50428,90	5,50
200	76518,14	81612,44	6,66
360	130838,47	140939,95	7,72
400	143839,62	155241,92	7,93
800	266268,51	289757,15	8,82
1200	378841,15	413365,75	9,11
<i>caso de Temperatura = 80°C</i>			
120	53401,79	54978,28	2,95
200	86638,43	89867,38	3,73
360	150454,96	157578,15	4,73
400	166003,75	174118,79	4,89
800	314677,96	333077,61	5,85
1200	454376,84	482909,65	6,28

Fonte: O autor (2025)

A tabela 15 e a figura 25 ilustram as diferenças percentuais de volume entre os casos. Uma análise particular da figura 25 revela um comportamento anômalo na curva de 40°C, que, ao contrário das outras, apresenta uma queda no final. Isso sugere que, em temperaturas mais baixas onde a cinética da dissolução é mais lenta, o aumento extremo da vazão para 1200 m³/h gera uma turbulência tão intensa que pode melhorar a eficiência do método de

circulação direta (D) de forma desproporcional, reduzindo assim sua diferença percentual em relação ao já eficiente método reverso (R).

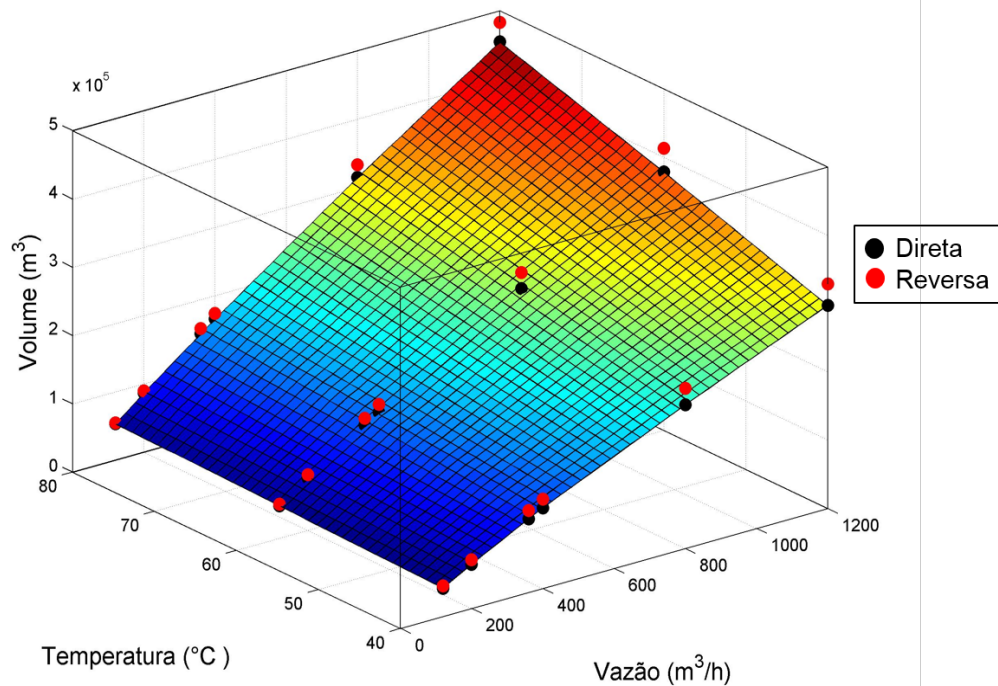
Figura 25 – Diferença percentual dos volumes entre os casos direto e reverso.



Fonte: O autor (2025)

A figura 26 serve como uma confirmação visual da tendência discutida anteriormente. Ao plotar os resultados de ambos os casos (esferas pretas para Direto, vermelhas para Reverso) sobre uma única superfície de referência, a figura destaca a magnitude da diferença de performance entre os métodos. Fica evidente que a separação vertical entre os pontos de cada par de simulação (mesma vazão e temperatura) é maior em temperaturas mais baixas (40°C) e diminui progressivamente em temperaturas mais altas (80°C). Essa representação gráfica reforça a conclusão de que a eficiência superior do método reverso é mais acentuada em regimes de menor temperatura.

Figura 26 – Superfície de Resposta - volume comparado



Fonte: O autor (2025)

6.2 RESULTADOS COM PCA

Nesta seção, é apresentada a análise estatística multivariada da matriz de dados consolidada, cuja estrutura foi mostrada na figura 9. A abordagem adotada neste trabalho difere daquela utilizada por Silva (2022), que realizou uma análise sobre a matriz de dados de cada cenário separadamente. Em contrapartida, neste estudo, optou-se por realizar a análise em uma matriz única contendo os dados de ambos os cenários, a fim de obter uma visão global do sistema.

A análise dos dados seguiu uma sequência metodológica rigorosa. Inicialmente, foi realizada uma análise descritiva das variáveis. Em seguida, para validar a aplicabilidade da PCA, foi realizado o teste de esfericidade de Bartlett (1950) ao nível de significância de 1%. Este teste verifica se as variáveis são correlacionadas entre si, uma condição necessária para que uma técnica de redução de dados, como a PCA, possa compactar os dados de maneira significativa. Como o resultado do teste confirmou a existência de correlações, aplicou-se finalmente a Análise de Componentes Principais à matriz.

Para permitir a comparação direta entre variáveis com diferentes escalas e unidades, os

dados foram previamente padronizados através da normalização *z-score*, um processo que consiste em subtrair a média e dividir pelo desvio padrão de cada variável. A tabela 16 apresenta a análise descritiva dos dados já padronizados, resumindo a média, o desvio padrão, a mediana, os valores mínimo e máximo, o primeiro quartil e o terceiro quartil para cada variável.

Tabela 16 – Análise descritiva das variáveis

Variáveis	Média	Desvio Padrão	Mediana	Mínimo	Máximo	Primeiro Quartil	Terceiro Quartil
Temperatura	60	16,56157	60	40	80	40	80
Taxa de produção	496,1753	367,4294	367,299	115,9891	1159,891	193,3143	773,2596
Raio	17,76206	6,359494	16,34913	8,481523	33,6982	12,86952	21,67224
Volume	179612,6	125374,1	142389,8	41384,7	482909,7	80338,86	272140,7
Pressão de Injeção	10,49333	2,974654	9,425	7,56	16,74	8,365	11,52
SG da salmoura	1,188156	0,007232	1,1875	1,1785	1,1994	1,180275	1,196
SG do fluido de injeção	1,014344	0,00882	1,01458	1,00223	1,02713	1,004205	1,023555
Fator dissolução do sal	3,985223	1,733602	3,641454	1,990351	6,536986	2,093171	6,100862
Potência da bomba	2695,161	4037,188	567,6314	64,14828	11431,77	156,4012	3635,694
Energia cumulativa	2,78E+13	4,17E+13	5,84E+12	6,6E+11	1,18E+14	1,62E+12	3,75E+13
Perda de tubulação	3,258273	3,98094	1,297398	0,150758	11,66974	0,403312	5,103422
Salmoura produzida	1429131	1058304	1057930	334083,1	3340827	556805,2	2227215
Pressão da bomba	9,8427	10,19622	4,830952	1,733942	30,76283	2,52151	14,76271

Fonte: O autor (2025)

Para a análise estatística foram renomeadas as 13 variáveis a seguir:

- X1 - Temperatura
- X2 - Taxa de produção
- X3 - Raio
- X4 - Volume
- X5 - Pressão da injeção
- X6 - SG da salmoura
- X7 - SG do fluido da injeção
- X8 - Fator dissolução do sal
- X9 - Potência da bomba
- X10 - Energia cumulativa
- X11 - Perda da tubulação
- X12 - Salmoura produzida

▪ X13 - Pressão da bomba

Daqui por diante, podem ser usadas tanto esta nova nomenclatura quanto os nomes originais das variáveis, segundo seja o caso.

Em seguida apresenta-se a tabela 17 da correlação entre as variáveis.

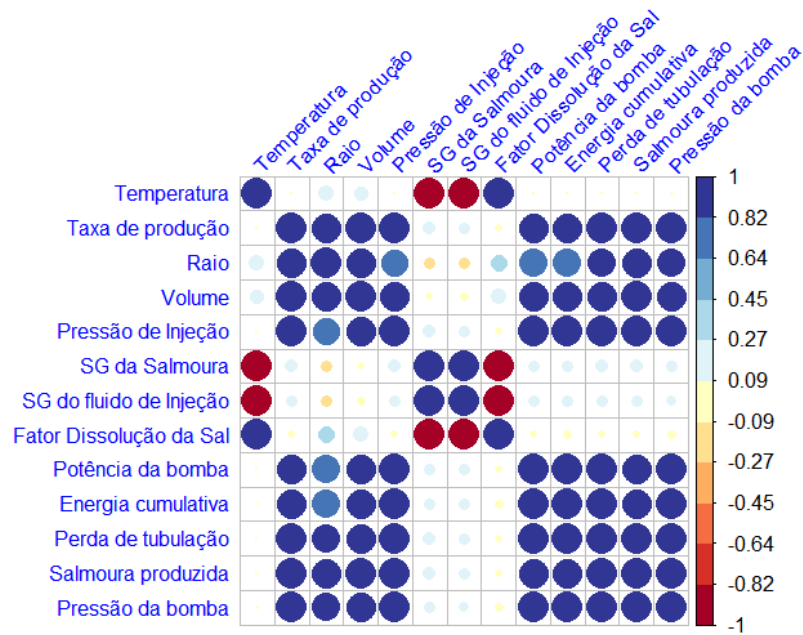
Tabela 17 – Correlação entre as variáveis.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
X1	1	0	0,25	0,21	0	-0,99	-0,99	0,99	0,01	0,01	0	0	0,01
X2	0	1	0,89	0,96	0,95	0,16	0,13	0,06	0,95	0,95	0,98	1	0,98
X3	0,25	0,89	1	0,95	0,75	-0,13	-0,14	0,3	0,81	0,81	0,86	0,89	0,85
X4	0,21	0,96	0,95	1	0,9	-0,05	-0,08	0,26	0,9	0,91	0,94	0,96	0,94
X5	0	0,95	0,75	0,9	1	0,17	0,14	0,06	0,97	0,97	0,97	0,95	0,97
X6	-0,99	0,16	-0,13	-0,05	0,17	1	1	-0,96	0,15	0,15	0,16	0,16	0,15
X7	-0,99	0,13	-0,14	-0,08	0,14	1	1	-0,98	0,13	0,13	0,13	0,13	0,12
X8	0,99	0,06	0,3	0,26	0,06	-0,96	-0,98	1	0,07	0,07	0,07	0,06	0,08
X9	0,01	0,95	0,81	0,9	0,97	0,15	0,13	0,07	1	1	0,99	0,95	0,99
X10	0,01	0,95	0,81	0,91	0,97	0,15	0,13	0,07	1	1	0,99	0,95	0,99
X11	0	0,98	0,86	0,94	0,97	0,16	0,13	0,07	0,99	0,99	1	0,98	1
X12	0	1	0,89	0,96	0,95	0,16	0,13	0,06	0,95	0,95	0,98	1	0,98
X13	0,01	0,98	0,85	0,94	0,97	0,15	0,12	0,08	0,99	0,99	1	0,98	1

Fonte: O autor (2025)

A figura 27 que representa a matriz de correlação entre as variáveis apresentadas na tabela anterior. Para uma análise mais visual, como solicitado, esta figura (correlograma) utiliza cores e tamanhos para indicar a força e a direção da correlação. Pode-se observar que a variável da temperatura (X1) tem uma forte correlação positiva com o fator de dissolução do sal (X8) e negativa com as gravidades específicas (X6 e X7). Nota-se também um grande bloco de correlações positivas entre as variáveis de operação (como X2 - Taxa de produção) e as variáveis de resultado (como X4 - Volume e X9 - Potência da bomba).

Figura 27 – Matriz de correlação entre as variáveis (Correlograma).



Fonte: O autor (2025)

Para validar a aplicabilidade da análise de componentes principais, foi utilizado o teste de esfericidade de Bartlett Bartlett (1950). Este teste de hipótese verifica se as variáveis na matriz de dados são correlacionadas, o que é uma premissa fundamental para o uso do PCA. A hipótese nula (H_0) postula que a matriz de correlação é uma matriz identidade (ou seja, as variáveis não são correlacionadas), tornando a aplicação do PCA inadequada. A hipótese alternativa (H_1), por sua vez, é de que a matriz de correlação não é uma matriz identidade, indicando que a aplicação do PCA é apropriada.

A estatística do teste segue uma distribuição χ^2 com $p(p - 1)/2$ graus de liberdade e é calculada pela seguinte fórmula:

$$\chi^2 = - \left(n - 1 - \frac{2p + 5}{6} \right) \ln |R|$$

onde n é o número de amostras, p é o número de variáveis e $|R|$ é o determinante da matriz de correlação.

O resultado do teste foi $\chi^2 = 2950,687$ com um p-valor de 0,0, o que leva à rejeição da hipótese nula com um nível de significância de 1%. Conclui-se, portanto, que as variáveis são correlacionadas e que a aplicação do PCA é justificada.

Na Tabela 18, são apresentados os autovalores obtidos mediante o PCA, a porcentagem da variância total explicada por cada componente, e a porcentagem acumulada.

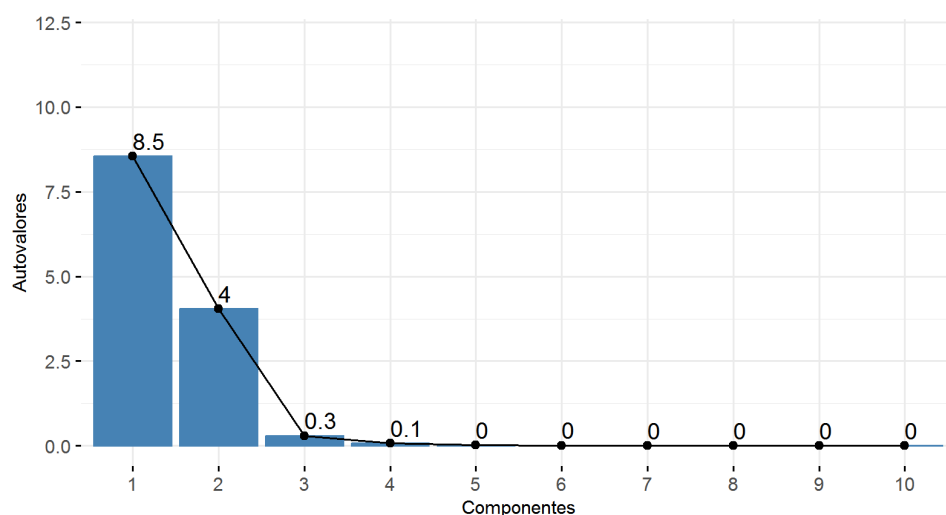
Tabela 18 – PCA - Autovalores, porcentagem da variância total e a porcentagem acumulada da explicação da variância total.

Componentes principais	Autovalor	Porcentagem (%)	Porcentagem Acumulada (%)
Comp1	8,55	65,75	65,75
Comp2	4,04	31,07	96,82
Comp3	0,30	2,30	99,12
Comp4	0,08	0,60	99,72
Comp5	0,02	0,14	99,86
Comp6	0,01	0,10	99,96
Comp7	0,00	0,03	99,99
Comp8	0,00	0,01	100,00
Comp9	0,00	0,00	100,00
Comp10	0,00	0,00	100,00
Comp11	0,00	0,00	100,00
Comp12	0,00	0,00	100,00
Comp13	0,00	0,00	100,00

Fonte: O autor (2025)

A tabela 18 apresentou os autovalores de cada componente principal. A figura 28, conhecida como *Scree Plot*, representa graficamente esses mesmos autovalores em ordem decrescente, evidenciando a dominância dos primeiros componentes.

Figura 28 – PCA - Autovalores (*Scree Plot*).

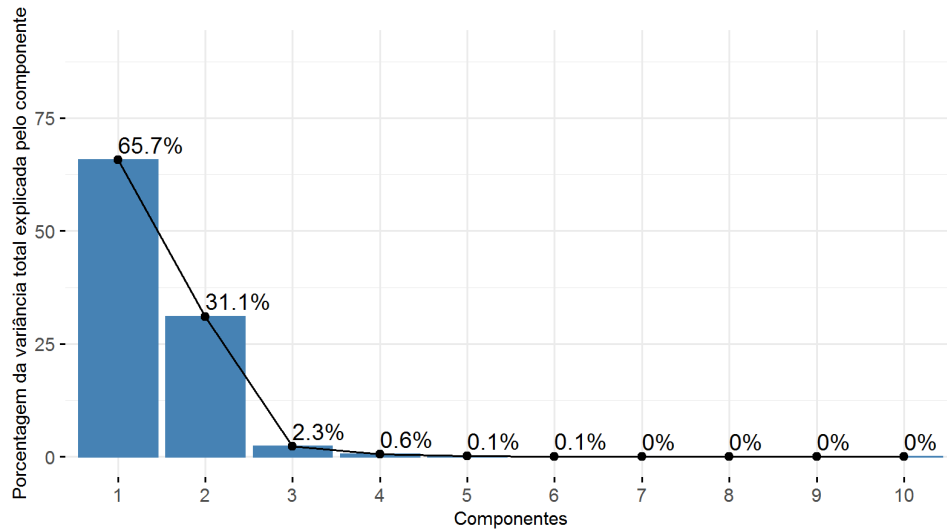


Fonte: O autor (2025)

Para quantificar essa importância, a figura 29 mostra a porcentagem da variância total explicada por cada componente. Nela, observa-se que os dois primeiros componentes principais

são suficientes para explicar aproximadamente 96,8% da variabilidade total dos dados, sendo 65,7% contribuídos pelo primeiro componente e 31,1% pelo segundo.

Figura 29 – PCA - Porcentagem da variância total explicada por componente.



Fonte: O autor (2025)

A análise da contribuição das variáveis, detalhada na tabela 19, permite interpretar o significado físico de cada componente principal. A metodologia para nomear cada componente consiste em identificar o grupo de variáveis com maior peso (contribuição) para aquele eixo e, a partir do tema que as une, atribuir um nome conceitual. Seguindo este método, a análise revela uma clara separação dos fatores que governam o processo de dissolução em eixos ortogonais de variação. O primeiro componente principal (Comp1) está fortemente associado a variáveis como taxa de produção, potência da bomba e volume, representando o eixo de intensidade operacional do sistema. O segundo componente (Comp2), por sua vez, é dominado pela temperatura, pelas gravidades específicas (SG) e pelo fator de dissolução, definindo o eixo termodinâmico que descreve as condições físico-químicas do processo. Esta dicotomia é um dos principais resultados da análise, pois quantifica a independência entre a escala da operação e as propriedades do fluido. Para facilitar a visualização, as figuras 30 e 31 ilustram graficamente os aportes das variáveis nesses dois primeiros componentes.

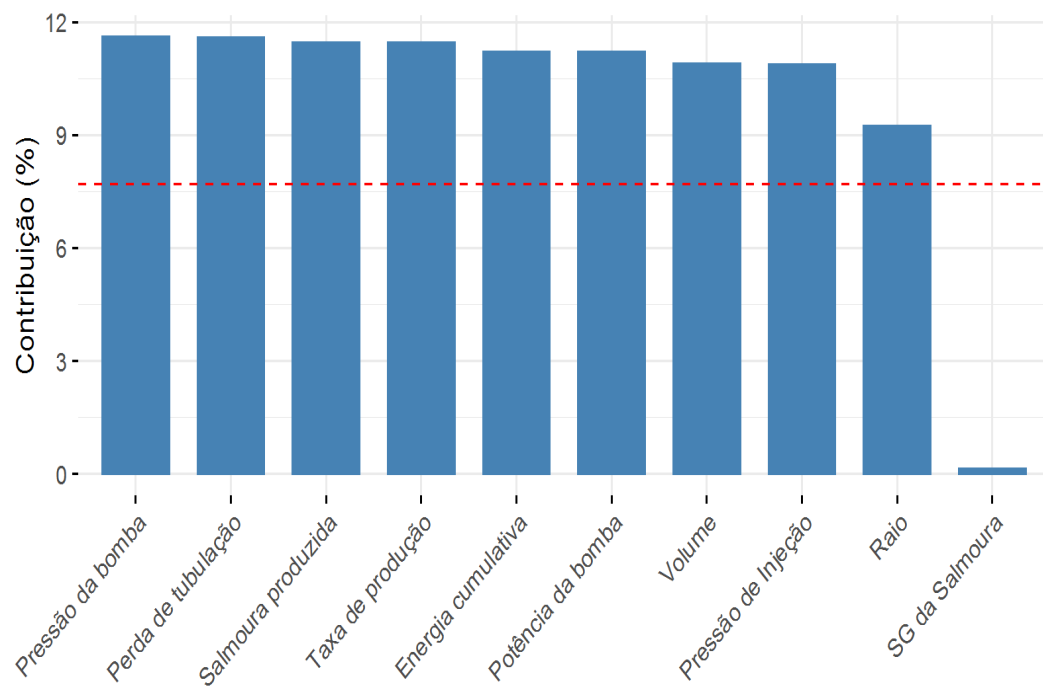
Tabela 19 – Contribuição das variáveis nos componentes

Variáveis	Comp1	Comp2	Comp3	Comp4	Comp5
Temperatura	0,00	0,42	0,01	0,00	0,00
Taxa de produção	0,20	0,00	0,03	0,19	0,27
Raio	0,16	0,02	0,85	0,19	0,00
Volume	0,19	0,01	0,15	0,09	1,00
Pressão de Injeção	0,19	0,00	0,26	0,40	0,07
SG da Salmoura	0,00	0,42	0,00	0,00	0,01
SG do fluido de Injeção	0,00	0,42	0,00	0,00	0,01
Fator Dissolução da Sal	0,00	0,42	0,03	0,01	0,00
Potência da bomba	0,19	0,00	0,15	0,28	0,01
Energia cumulativa	0,19	0,00	0,15	0,28	0,01
Perda de tubulação	0,20	0,00	0,02	0,05	0,03
Salmoura produzida	0,20	0,00	0,03	0,19	0,27
Pressão da bomba	0,20	0,00	0,03	0,02	0,03

Fonte: O autor (2025)

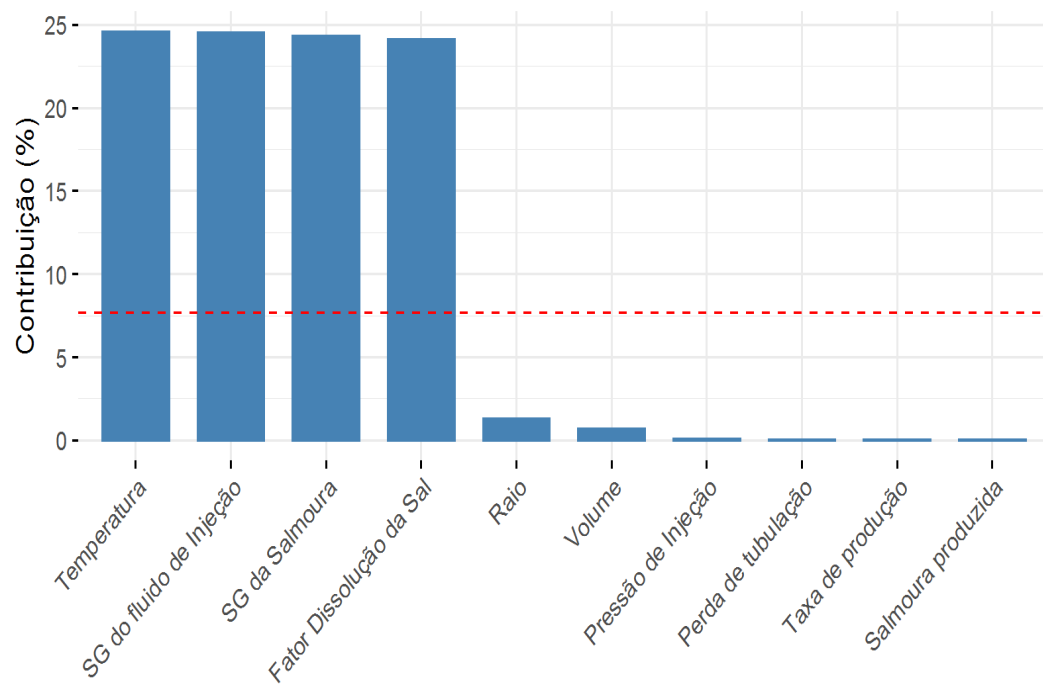
Nas figuras 30 e 31, são apresentados os aportes das variáveis no primeiro e segundo componentes, respectivamente, facilitando a visualização dos resultados desta tabela. A linha tracejada vermelha no gráfico indica a contribuição média esperada. Se a contribuição das variáveis for uniforme, o valor esperado seria $1/\text{número de variáveis} = 1/13 \approx 7.7\%$ onde p é o número de variáveis. Para um determinado componente, uma variável com contribuição maior que esse limite pode ser considerada importante na contribuição do componente. Assim no primeiro componente as variáveis de pressão da bomba, perda de tubulação, salmoura produzida, taxa de produção, energia cumulativa, potencia da bomba, volume, pressão de injeção e raio são as que mais contribuem. Já no segundo componente, as variáveis que mais contribuem são a temperatura, a gravidade específica do fluido de injeção, a gravidade específica da salmoura e o fator de dissolução do sal.

Figura 30 – PCA - Contribuição das variáveis no primeiro componente



Fonte: O autor (2025)

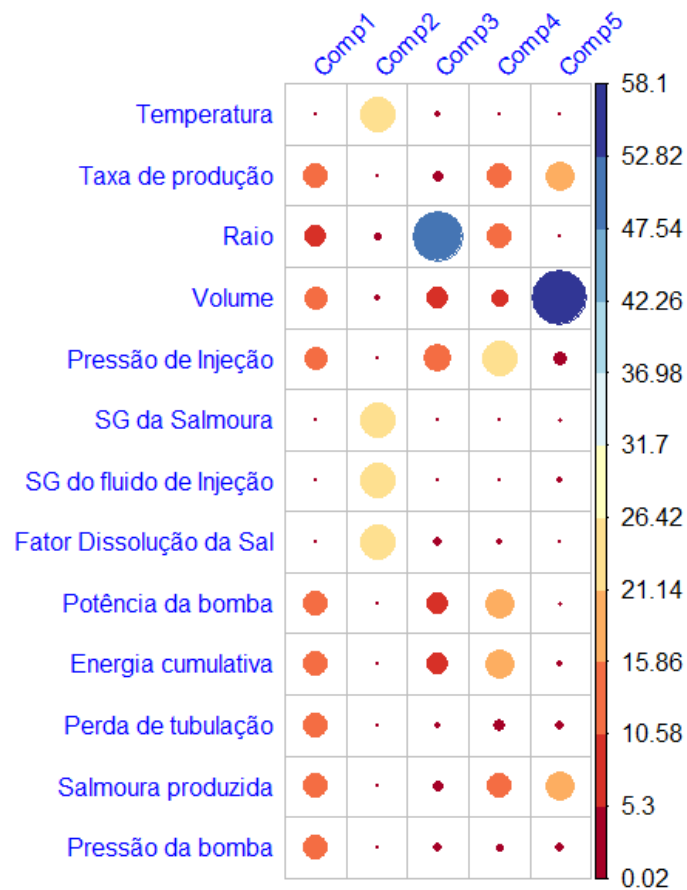
Figura 31 – PCA - Contribuição das variáveis no segundo componente



Fonte: O autor (2025)

A figura 32 apresenta um resumo visual da contribuição das variáveis para os cinco primeiros componentes principais, complementando os dados da tabela 17 (apresentada anteriormente). O gráfico reforça a interpretação dos dois eixos principais: o Componente 1 agrupa as variáveis operacionais, enquanto o Componente 2 agrupa as variáveis termodinâmicas. Adicionalmente, a figura revela que os componentes de menor ordem capturam variâncias mais específicas. Notavelmente, o Componente 3 é quase que exclusivamente dominado pela variável "raio", e o Componente 5 é similarmente dominado pelo "volume", indicando que estas geometrias possuem aspectos de variabilidade únicos não totalmente explicados pelos dois primeiros eixos.

Figura 32 – PCA - Contribuição das variáveis nos cinco primeiros componentes.



Fonte: O autor (2025)

Enquanto a tabela anterior mostrou a *contribuição* de cada variável para a formação dos componentes, a tabela 20 a seguir detalha a *correlação* entre as variáveis e estes mesmos componentes. Os valores de correlação (entre -1 e 1) são importantes para interpretar a direção (positiva ou negativa) da influência de cada variável em cada eixo principal.

Tabela 20 – Correlação entre as variáveis e os componentes.

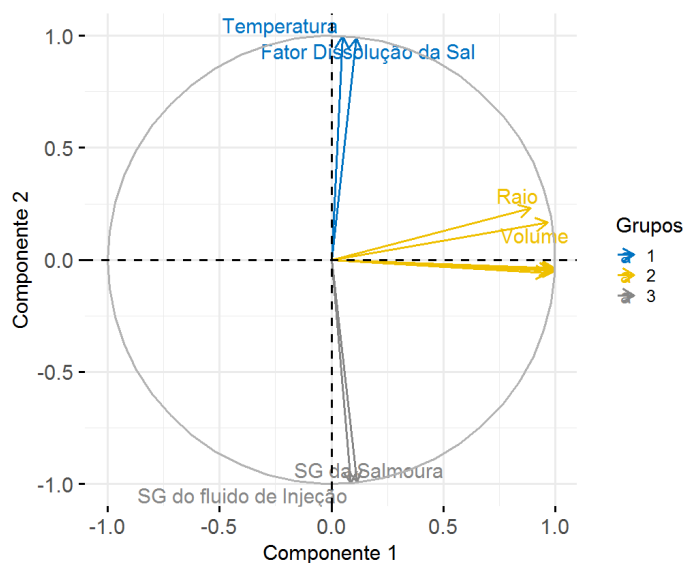
Variáveis	Comp1	Comp2	Comp3	Comp4	Comp5
Temperatura	0,05	1,00	0,05	0,01	0,00
Taxa de produção	0,99	-0,05	-0,08	0,09	-0,05
Raio	0,89	0,23	-0,39	-0,09	0,00
Volume	0,97	0,17	-0,16	0,06	0,10
Pressão de Injeção	0,97	-0,06	0,21	0,14	0,03
SG da Salmoura	0,11	-0,99	-0,01	0,01	0,01
SG do fluido de Injeção	0,09	-1,00	-0,02	0,01	0,01
Fator Dissolução da Sal	0,11	0,99	0,07	0,02	0,00
Potência da bomba	0,98	-0,05	0,16	-0,11	0,01
Energia cumulativa	0,98	-0,04	0,16	-0,11	0,01
Perda de tubulação	1,00	-0,05	0,05	-0,05	-0,02
Salmoura produzida	0,99	-0,05	-0,08	0,09	-0,05
Pressão da bomba	1,00	-0,04	0,07	-0,03	-0,02

Fonte: O autor (2025)

A figura 33 apresenta o círculo de correlação, projetando as variáveis nos dois eixos principais. Conforme a análise de contribuição, interpretamos estes eixos da seguinte forma: o componente 1 (horizontal) representa o "eixo de intensidade operacional"(ligado à produção e volume), enquanto o componente 2 (vertical) representa o "eixo termodinâmico"(ligado à temperatura e propriedades do fluido). Com base nisso, observamos:

- As variáveis de saída do SALGAS — *raio, volume, potência da bomba, energia cumulativa, perda de tubulação, salmoura produzida e pressão da bomba* — formam um agrupamento coeso alinhado ao eixo de intensidade operacional, indicando a forte correlação positiva entre si.
- A correlação negativa entre o grupo de variáveis termodinâmicas (*temperatura e fator de dissolução*) e o de *gravidades específicas (SG)* é confirmada, uma vez que seus vetores se posicionam em quadrantes opostos ao longo do Eixo Termodinâmico.
- Adicionalmente, a baixa correlação entre a *temperatura* e as variáveis de saída é evidenciada pela quase ortogonalidade (ângulo de aproximadamente 90º) entre seus vetores representativos.

Figura 33 – Círculo de correlação

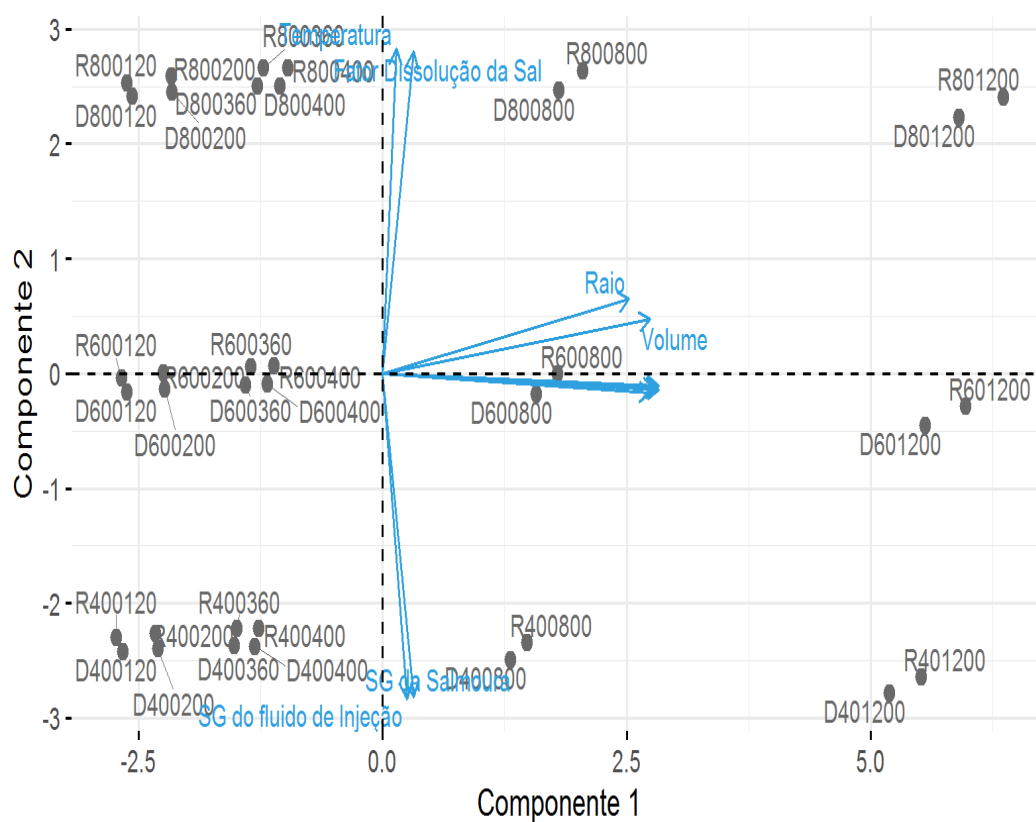


Fonte: O autor (2025)

A figura 34 apresenta um biplot da análise de componentes principais, que permite a interpretação simultânea das variáveis (vetores) e dos indivíduos (casos simulados). A nomenclatura dos casos segue o formato D/R(temp)(vazão), onde D indica circulação direta e R, reversa. A interpretação da projeção dos indivíduos em relação aos vetores confirma a estrutura física do problema:

- Indivíduos como D600800 e R600800, caracterizados por vazões elevadas, projetam-se fortemente na mesma direção dos vetores das variáveis de saída (raio, volume, etc.), indicando que estes casos resultam em valores elevados para tais métricas.
- Em contraste, indivíduos como D800400 e R800400 alinham-se com os vetores de temperatura e fator de dissolução, o que é consistente com suas altas temperaturas de referência, mas os posiciona em oposição às variáveis de gravidade específica.
- Casos de menor vazão, como D400400 e R400400, posicionam-se em oposição aos vetores de maior produção, validando a correlação já observada.

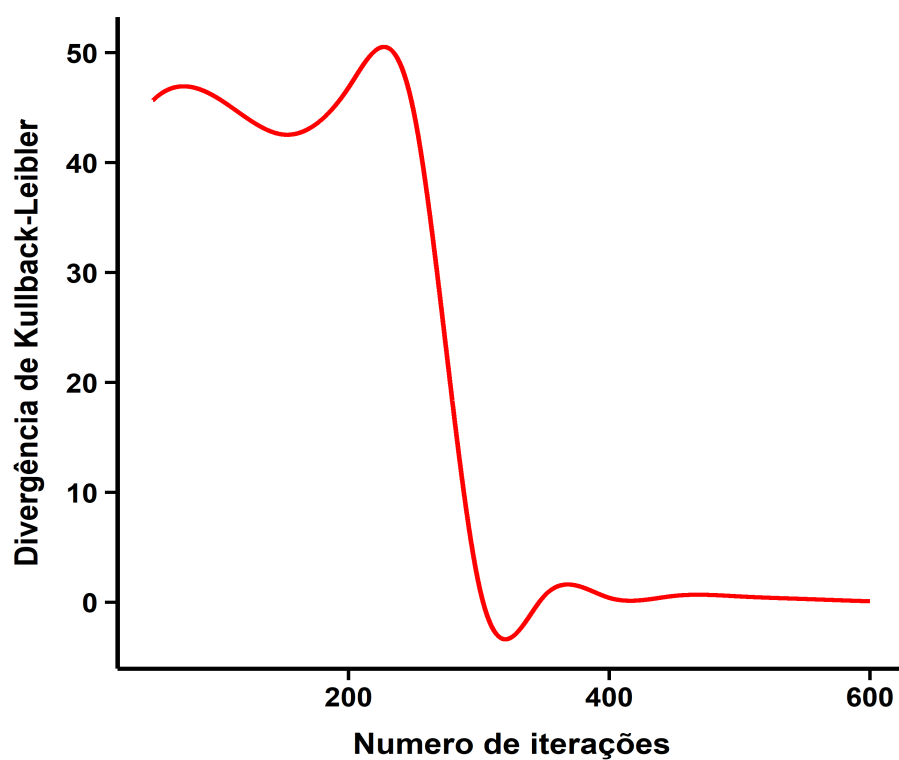
Figura 34 – PCA - biplot



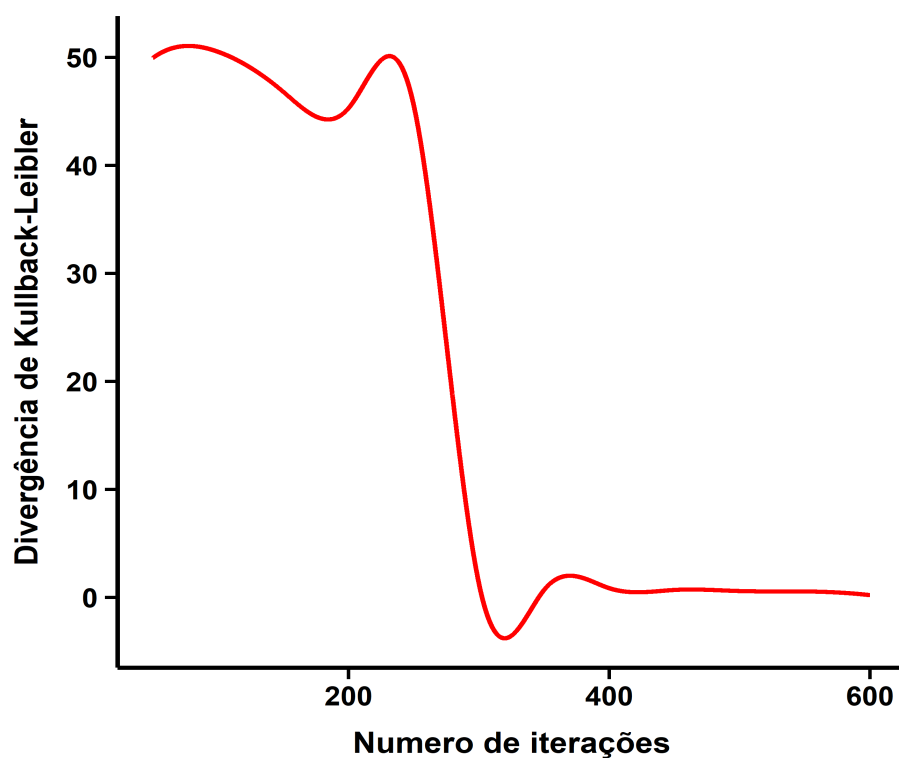
Fonte: O autor (2025)

6.2.1 Resultados com t-SNE

Esta seção apresenta os resultados da técnica não linear t-SNE aplicada à mesma matriz de dados utilizada na análise PCA. A definição dos hiperparâmetros, como a perplexidade e o número de iterações, seguiu os critérios propostos por Wattenberg, Viégas e Johnson (2016). Conforme ilustrado nas figuras 35 e 36, para 500 iterações e valores de perplexidade (π) de 5 ou 6, a divergência de Kullback-Leibler torna-se insignificante.

Figura 35 – Divergência vs. Perplexidade ($\pi = 5$)

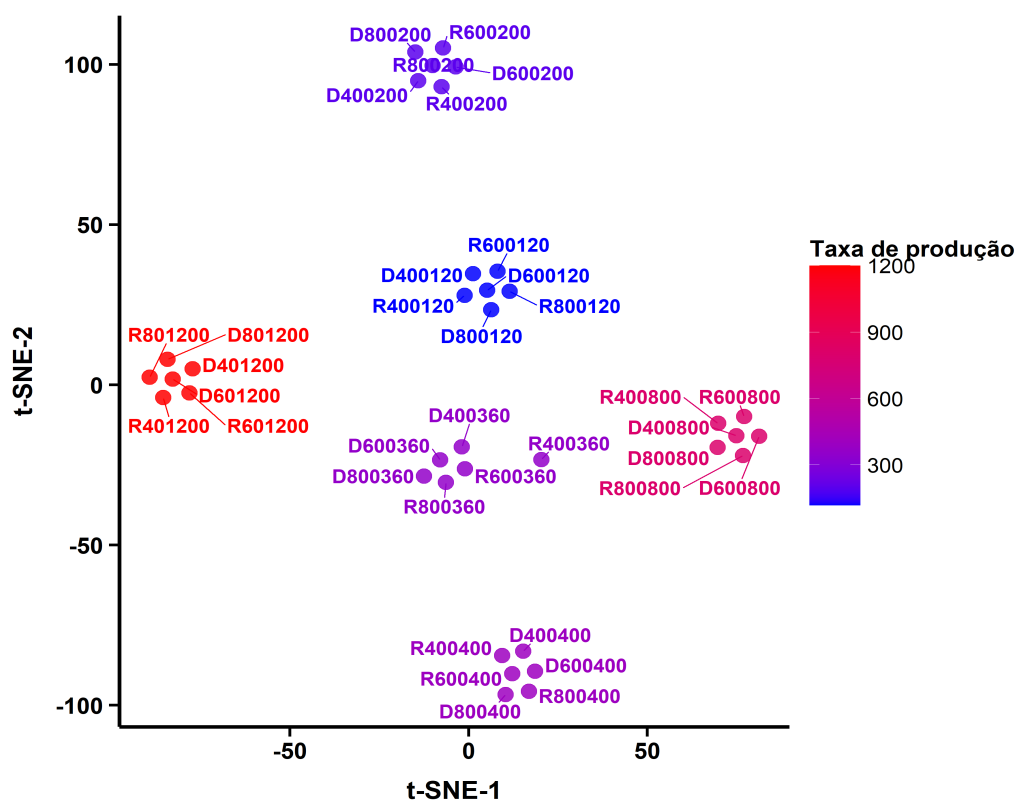
Fonte: O autor (2025)

Figura 36 – Divergência vs. Perplexidade ($\pi = 6$)

Fonte: O autor (2025)

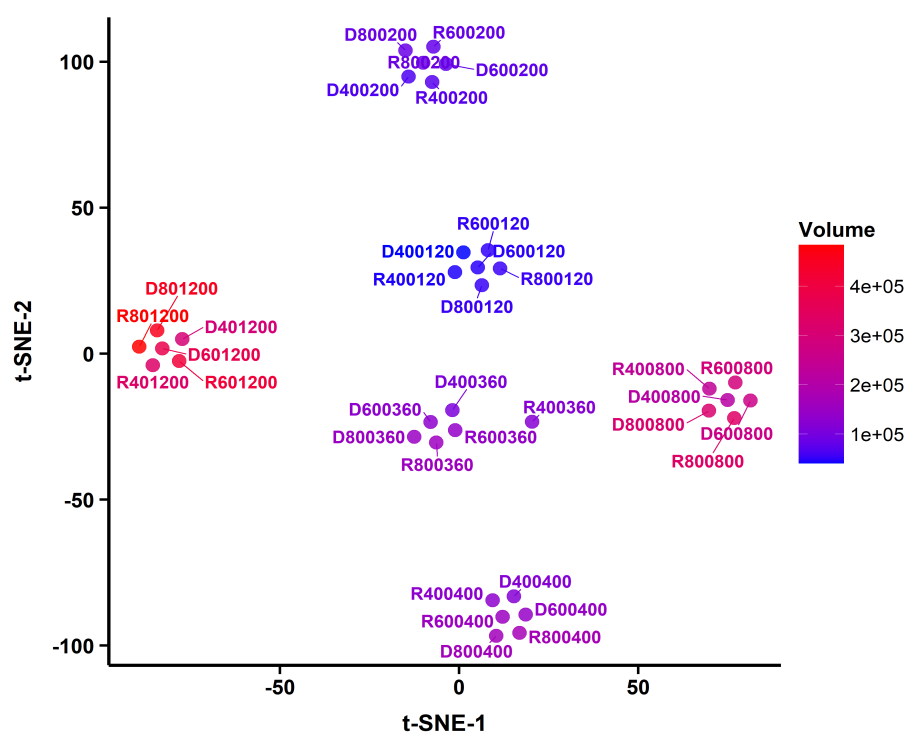
A análise comparativa das figuras 37 e 38, que exibem a mesma projeção de dados gerada pelo t -SNE, confirma a forte correlação entre a taxa de produção e o volume. A correspondência visual é direta: os agrupamentos de indivíduos coloridos para representar altos valores de taxa de produção (figura 37) são os mesmos que representam altos valores de volume (figura 38), validando a correlação positiva entre as variáveis.

Figura 37 – t -SNE - Taxa de produção



Fonte: O autor (2025)

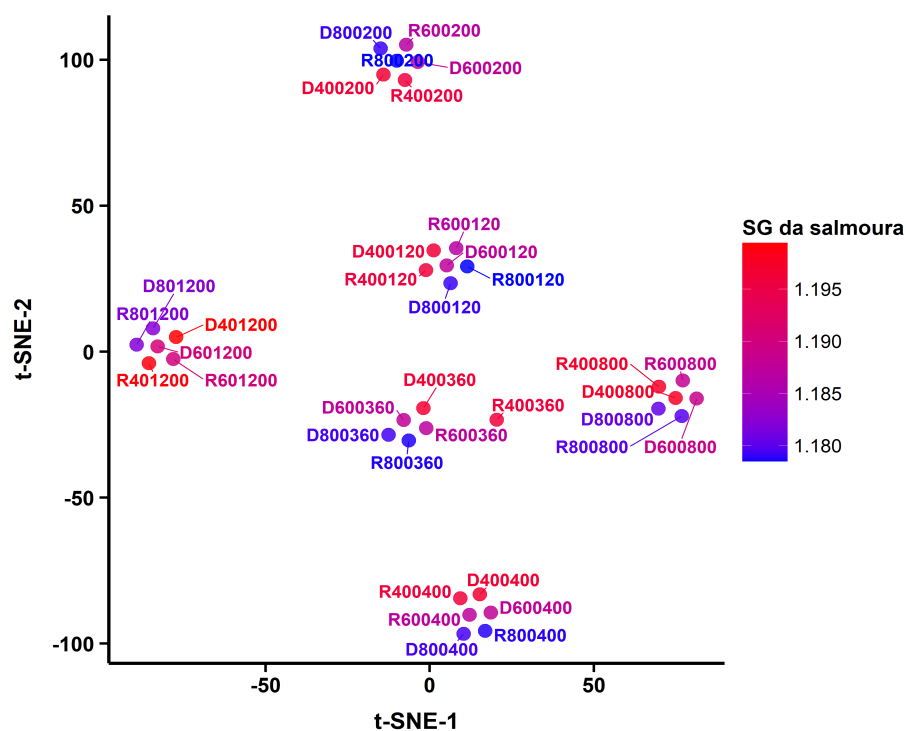
Figura 38 – Volume



Fonte: O autor (2025)

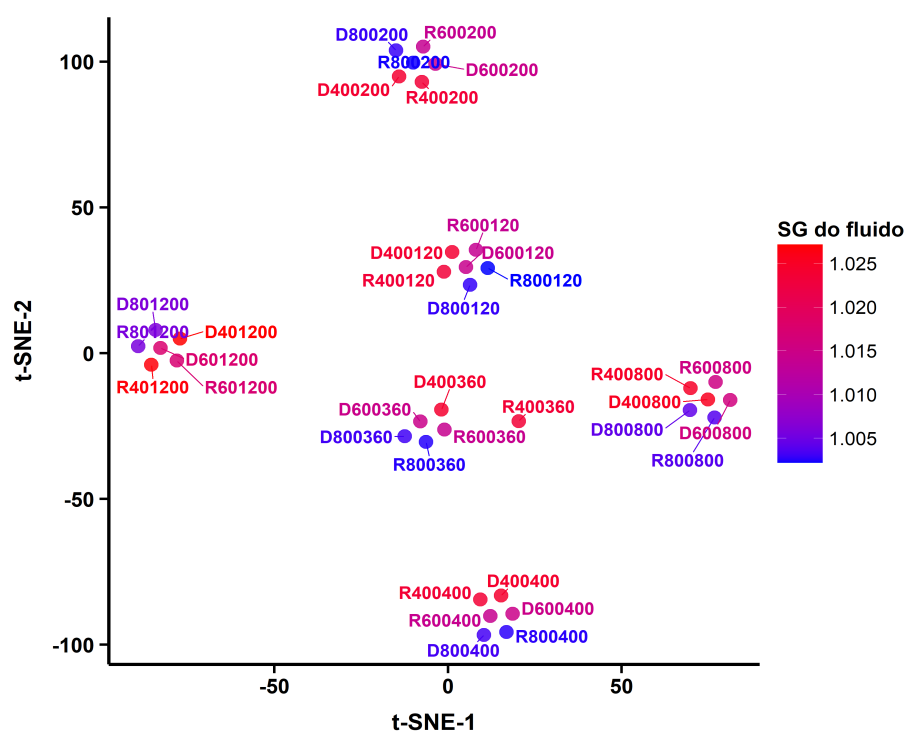
Nas figuras 39 e 40 observa-se que para a mesma perplexidade, os agrupamentos dos indivíduos e a distribuição das cores nos grupos também compartilham semelhanças em ambas as figuras. Isto confirma a correlação entre estas variáveis,

Figura 39 – Densidade específica da salmoura



Fonte: O autor (2025)

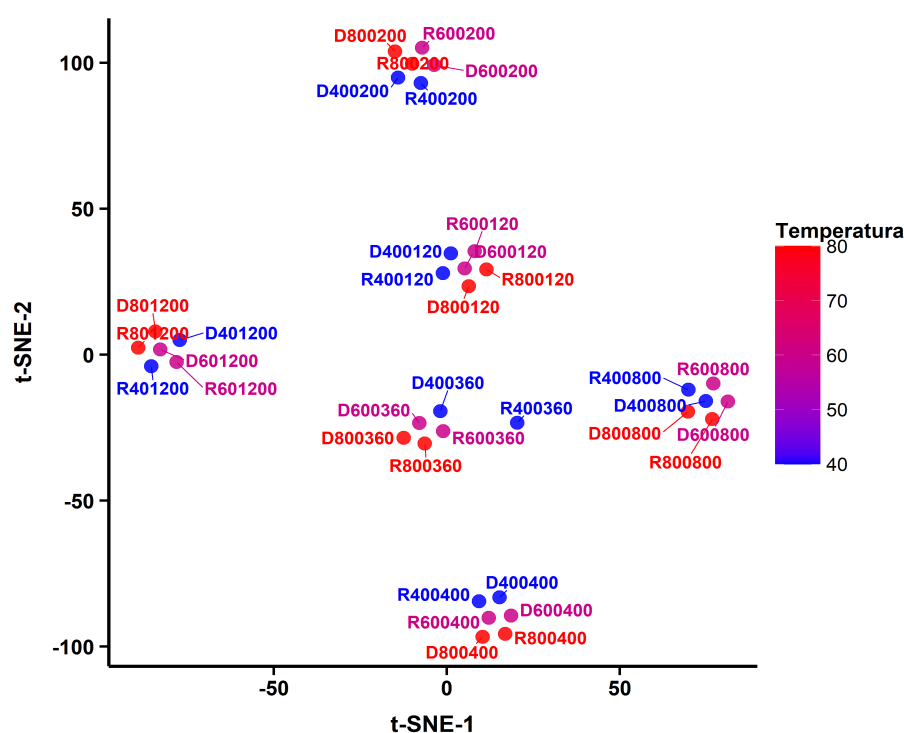
Figura 40 – Densidade específica do fluido injetado



Fonte: O autor (2025)

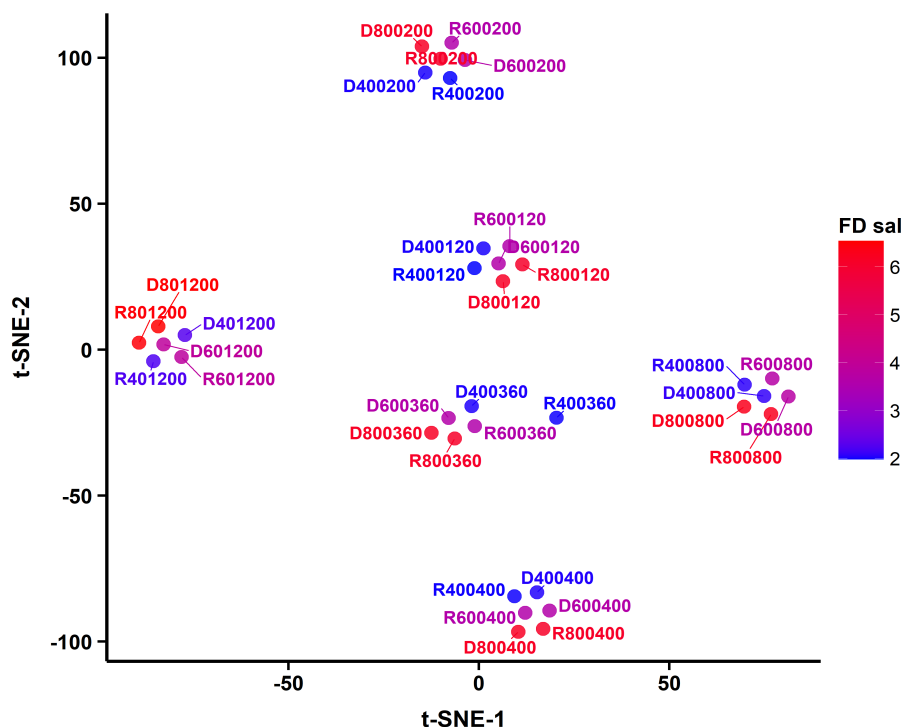
Finalmente, a análise comparativa das figuras 41 e 42 para uma perplexidade de 5 revela que os padrões de agrupamento e coloração são análogos. Essa semelhança visual ratifica a forte correlação positiva entre a temperatura de referência e o Fator de Dissolução do Sal. Em termos práticos, isso significa que o aumento da temperatura do fluido injetado acelera diretamente o processo de dissolução, sendo uma variável chave para o controle da velocidade de construção da caverna.

Figura 41 – Agrupamento t-SNE com coloração por caso de temperatura de referência



Fonte: O autor (2025)

Figura 42 – Agrupamento t-SNE com coloração pela variável Fator de Dissolução da Sal.

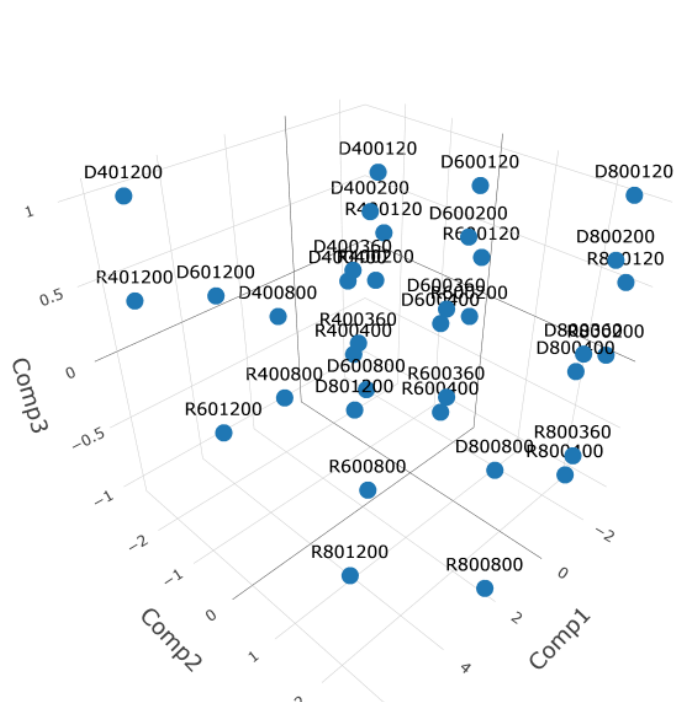


Fonte: O autor (2025)

Por último, as figuras 43 e 44 apresentam uma comparação visual entre as projeções tridimensionais dos dados, obtidas pelos métodos PCA e *t-SNE*. A projeção do PCA (figura 43) utiliza os três primeiros componentes principais, que capturam 99,10% da variância total. Em contraste, o *t-SNE* (figura 44) otimiza os eixos para preservar a topologia local do espaço de dados original, formando agrupamentos visualmente mais distintos.

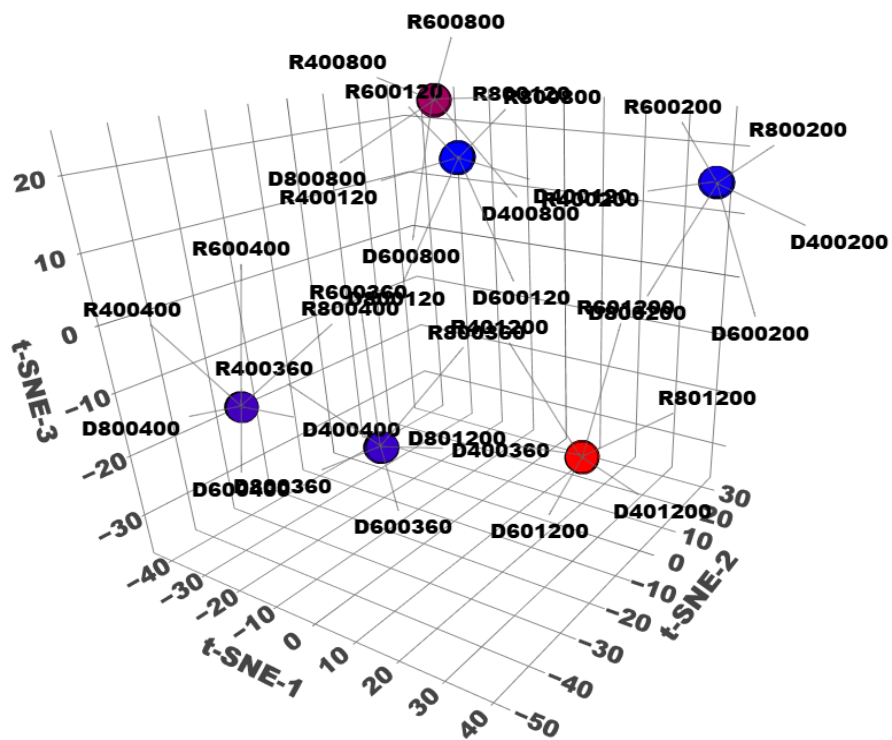
A eficácia do *t-SNE* em agrupar casos funcionalmente similares é evidente. Isso é exemplificado pelo agrupamento dos casos D400400 e R800400. Embora estes casos possuam condições operacionais distintas, como gravidades específicas diferentes (conforme a figura 39), o *t-SNE* os agrupa. A justificativa é que o resultado principal de interesse — o volume final da caverna — é muito semelhante para ambos, como pode ser verificado na figura 38. Portanto, o agrupamento indica que o *t-SNE* prioriza a similaridade nos resultados de maior impacto (saída) em vez de apenas a proximidade dos parâmetros operacionais (entrada).

Figura 43 – PCA 3D



Fonte: O autor (2025)

Figura 44 – t-SNE 3D



Fonte: O autor (2025)

7 OTIMIZAÇÃO DO DESIGN DE CAVERNAS SALINAS VIA ALGORITMOS GENÉTICOS

Os capítulos precedentes desta tese dedicaram-se à análise dos fenômenos físicos que governam a dissolução de cavernas em domos salinos. Por meio da simulação numérica com o *software* SALGAS e da aplicação de técnicas de redução de dimensionalidade (PCA e t-SNE), foi possível dissecar a interação entre as variáveis de controle e os resultados geométricos e de produção. A análise estatística multivariada respondeu à pergunta fundamental "quais variáveis são mais influentes?", revelando que a taxa de produção de salmoura se manifesta como o principal motor do desenvolvimento volumétrico, enquanto o grupo de variáveis associado à temperatura desempenha um papel secundário, porém crítico e inter-relacionado Warren (2016).

A conexão entre a análise precedente e a otimização subsequente é, portanto, direta e sinérgica. A análise t-SNE atuou como uma ferramenta de diagnóstico em larga escala, mapeando o espaço de *design* e confirmando que ele se organiza em regimes operacionais distintos. Com base neste mapa, os critérios de engenharia fundamentados na literatura técnica, detalhados na seção de formulação do problema, justificaram a restrição do espaço de busca da otimização à faixa de 50 a 300 m³/h. Finalmente, o Algoritmo Genético foi empregado como uma ferramenta de *design* de precisão, realizando uma busca exaustiva dentro deste regime pré-selecionado para encontrar a combinação ótima de parâmetros.

Com base nesses *insights*, este capítulo marca uma transição metodológica da análise para o *design* de engenharia. A validação das variáveis de maior impacto permite agora formular de maneira robusta o problema de otimização, cuja finalidade é responder à pergunta subsequente: "quais são os valores ótimos para essas variáveis?". O escopo, portanto, transcende a mera observação de casos pré-definidos para, de forma sistemática e automatizada, determinar a estratégia de construção ótima que satisfaça um conjunto de objetivos técnicos, econômicos e, crucialmente, de segurança.

O problema de otimização do *design* de uma caverna de sal, quando acoplado a um simulador numérico como o SALGAS, enquadra-se na categoria de Otimização *black-box* (BBO). Nesta classe de problemas, a função-objetivo (por exemplo, o volume final da caverna) não possui uma forma analítica explícita e seu valor só pode ser obtido através da execução de uma simulação computacionalmente dispendiosa Mello, Lacerda e Pozo (2023). A natureza *black-box* do simulador impede o uso de métodos de otimização clássicos baseados em gradiente,

que requerem a derivada da função-objetivo. Em tais casos, meta-heurísticas populacionais, como os Algoritmos Genéticos (AG), emergem como a ferramenta de escolha devido à sua capacidade de navegar eficientemente por espaços de busca complexos sem depender de informações de gradiente Al-Tashi, Rais e Abdulkadir (2023). O custo computacional associado a cada avaliação da função-objetivo (uma simulação completa do SALGAS) torna a eficiência do algoritmo de otimização um fator crítico Al-Sa'ary (2022). Embora outras abordagens populacionais como a Otimização por enxame de partículas (PSO) ou algoritmos de evolução diferencial (DE) também sejam aplicáveis, a escolha pelo algoritmo genético nesta tese é fundamentada em sua robustez comprovada e, crucialmente, na existência de extensões poderosas e bem estabelecidas para problemas multiobjetivo, como o algoritmo NSGA-II, que será empregado posteriormente neste trabalho.

A abordagem de acoplar os AG a simuladores de subsuperfície é uma metodologia bem estabelecida e validada em diversas áreas da engenharia de petróleo e geomecânica. Aplicações notáveis incluem o ajuste de histórico de produção de reservatórios (*history matching*), onde os parâmetros do modelo de reservatório são ajustados para minimizar a discrepância entre a produção simulada e a histórica Zhang e Zhang (2023), e a otimização de trajetórias de poços para maximizar a taxa de penetração e garantir a estabilidade do poço Pramudyo, Latief e Raj (2017). A aplicação de um AG para otimizar o processo de lixiviação de cavernas, portanto, alinha-se com práticas da indústria e da academia para resolver problemas de *design* baseados em simulação.

Para solucionar este problema — que é, por sua natureza, multivariado, não-linear e sujeito a múltiplas restrições — foi concebido e implementado um *framework* computacional que acopla o simulador SALGAS a um algoritmo genético (AG), implementado através do Global Optimization Toolbox do MATLAB The MathWorks, Inc. (2024b).

7.1 ANÁLISE DOS RESULTADOS DA OTIMIZAÇÃO

A execução do *framework* de otimização produziu um conjunto de soluções que representam *designs* ótimos para a caverna de sal, tanto para o caso mono-objetivo quanto para o multiobjetivo.

7.1.1 Otimização Mono-Objetivo: Maximização do Volume Final

No caso de otimização mono-objetivo, a função-objetivo foi configurada para maximizar o volume final da caverna ao término da campanha de lixiviação, sujeita a todas as restrições operacionais e geomecânicas. O Algoritmo Genético convergiu para uma solução única que representa a melhor estratégia de mineração encontrada para atingir este objetivo.

A estratégia ótima, detalhada na tabela 21, consiste em uma campanha de dois estágios com parâmetros operacionais distintos.

A análise dos parâmetros na tabela 21 revela uma estratégia de lixiviação bifásica sofisticada, que seria de difícil concepção manual. O Algoritmo Genético convergiu para uma solução que pode ser interpretada da seguinte forma:

1. Estágio 1: Criação de volume bruto (alta vazão). O primeiro estágio, com duração de 414 dias, utiliza uma vazão elevada de $272.73 \text{ m}^3/\text{h}$. O objetivo aqui é "atacar" a rocha salina de forma agressiva para criar o volume principal da caverna o mais rápido possível, aproveitando a alta taxa de dissolução proporcionada pela turbulência.
2. Estágio 2: Refinamento e controle geométrico (vazão reduzida). O segundo estágio transita para uma fase de 370 dias com uma vazão significativamente menor, de $128.80 \text{ m}^3/\text{h}$. Esta redução na vazão permite um processo de dissolução mais controlado, focando no refinamento da forma da caverna. A mudança nas profundidades de operação (elevação do *blanket* e ajuste dos pontos de injeção/produção) nesta fase é crucial para alargar a porção superior da cavidade e garantir que a geometria final atenda às restrições de estabilidade, como a razão H/D , sem ultrapassar os limites de um *design* seguro.

Em suma, o otimizador não buscou simplesmente a maior vazão pelo maior tempo, mas descobriu uma estratégia ótima que equilibra a velocidade de criação de volume com o controle fino da geometria final, demonstrando a capacidade da metodologia em encontrar soluções de engenharia complexas e eficientes.

Tabela 21 – Vetor de decisão ótimo da otimização mono-objetivo.

Parâmetro	Estágio 1	Estágio 2
Vazão de injeção (Q) [m ³ /h]	272.73	128.80
Duração do estágio (t) [dias]	414.46	370.39
Profundidade do <i>blanket</i> (B) [m]	734.62	723.37
Profundidade de injeção (I) [m]	856.21	850.77
Profundidade de produção (P) [m]	804.21	795.66

Fonte: O autor (2025)

A execução desta estratégia otimizada resulta nas métricas de desempenho apresentadas na tabela 22. O tempo total de construção da campanha é de 784.85 dias (soma das durações t1 e t2), resultando em um volume final de 544,251 m³.

Tabela 22 – Métricas de desempenho da solução otimizada.

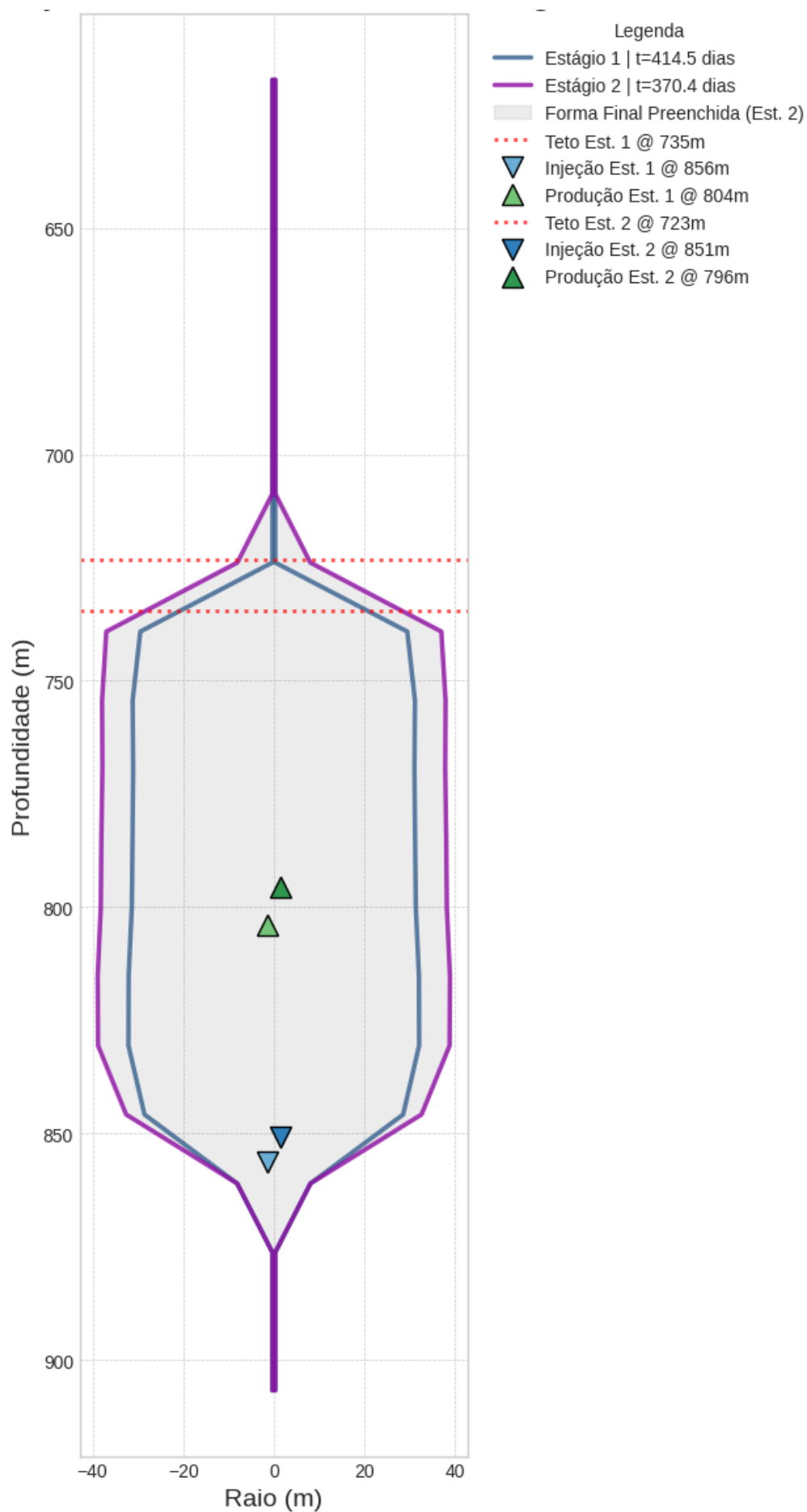
Métrica de desempenho	Valor
Volume máximo (MV)	544,251.49 m ³
Massa de sal extraída (MS)	352,027.25 ton
Tempo total de construção (MT)	784.85 dias
Eficiência energética (ME)	0.4662

Fonte: O autor (2025)

A análise da solução ótima revela que o algoritmo tende a explorar os limites do espaço de busca viável para atingir o objetivo. A configuração das profundidades (B, I, P) é ajustada para criar uma geometria que maximize a área de superfície de contato entre a água doce e a rocha salina, respeitando as restrições de estabilidade, como a razão H/D.

A figura 45 traduz a estratégia operacional ótima, detalhada na tabela 21, em um resultado geométrico. O gráfico ilustra a evolução da caverna ao longo dos dois estágios da campanha de mineração.

Figura 45 – Evolução geométrica e perfil final da caverna otimizada.

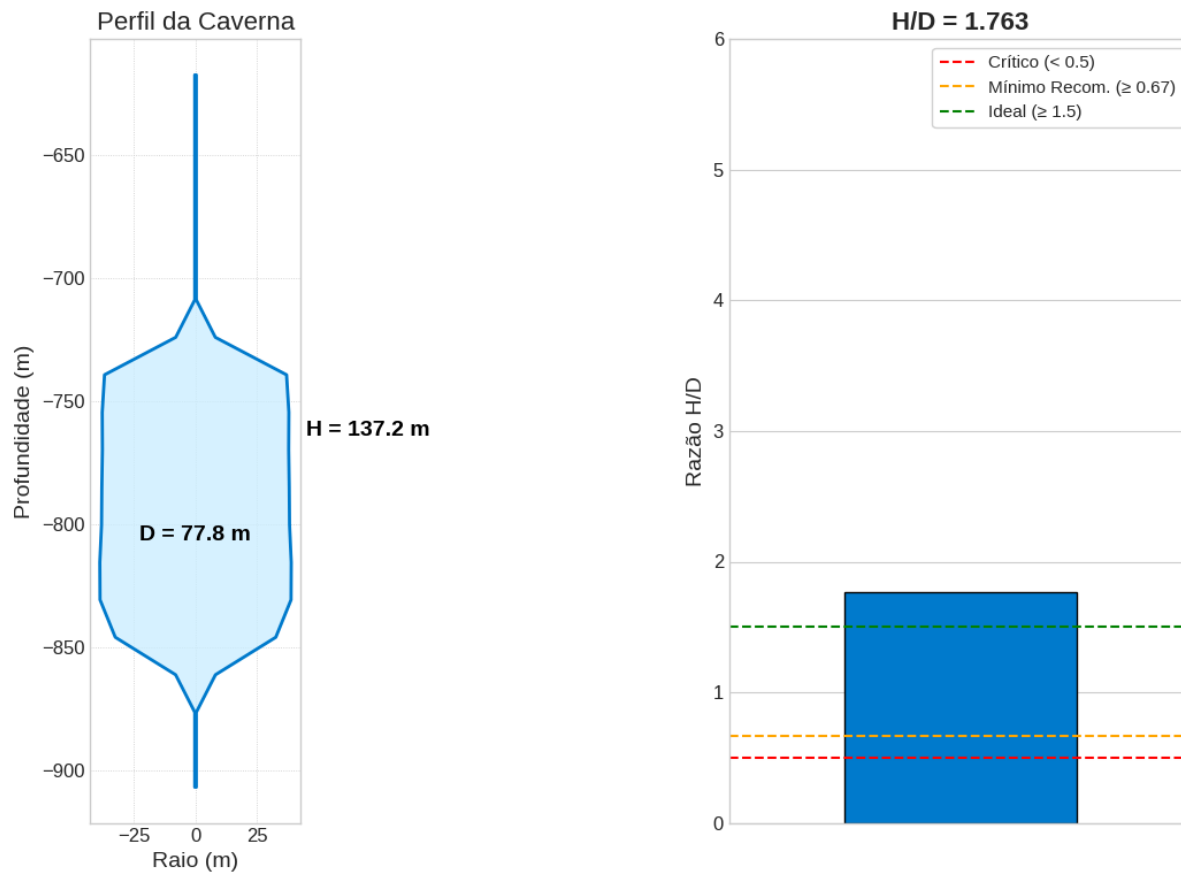


Fonte: O autor (2025)

O processo otimizado consiste em dois estágios distintos. O estágio 1, com duração de 414.5 dias, foca em desenvolver a geometria inicial da caverna. As posições dos equipamentos são definidas com injeção a 856 m e produção a 804 m, estabelecendo um teto inicial em 735 m. Subsequentemente, o estágio 2, com 370.4 dias, tem como objetivo refinar a forma: as profundidades de operação são ajustadas, o que resulta em uma expansão predominantemente lateral e na elevação do teto final para 723 m. Na figura, os marcadores triangulares são utilizados apenas como um auxílio visual para indicar as diferentes profundidades de injeção e produção em cada um dos dois estágios. A área cinza representa o perfil final preenchido. A geometria resultante, com um topo largo e uma base expandida, é uma consequência direta da combinação de parâmetros otimizados pelo Algoritmo Genético, demonstrando a capacidade da metodologia de projetar formas complexas para maximizar o volume.

Um critério fundamental para a viabilidade de longo prazo de uma caverna salina é a sua estabilidade geomecânica, frequentemente avaliada pela razão entre sua altura e diâmetro (H/D). A figura 46 apresenta a análise de estabilidade para a geometria da caverna otimizada.

Figura 46 – Verificação da estabilidade geomecânica da solução otimizada.

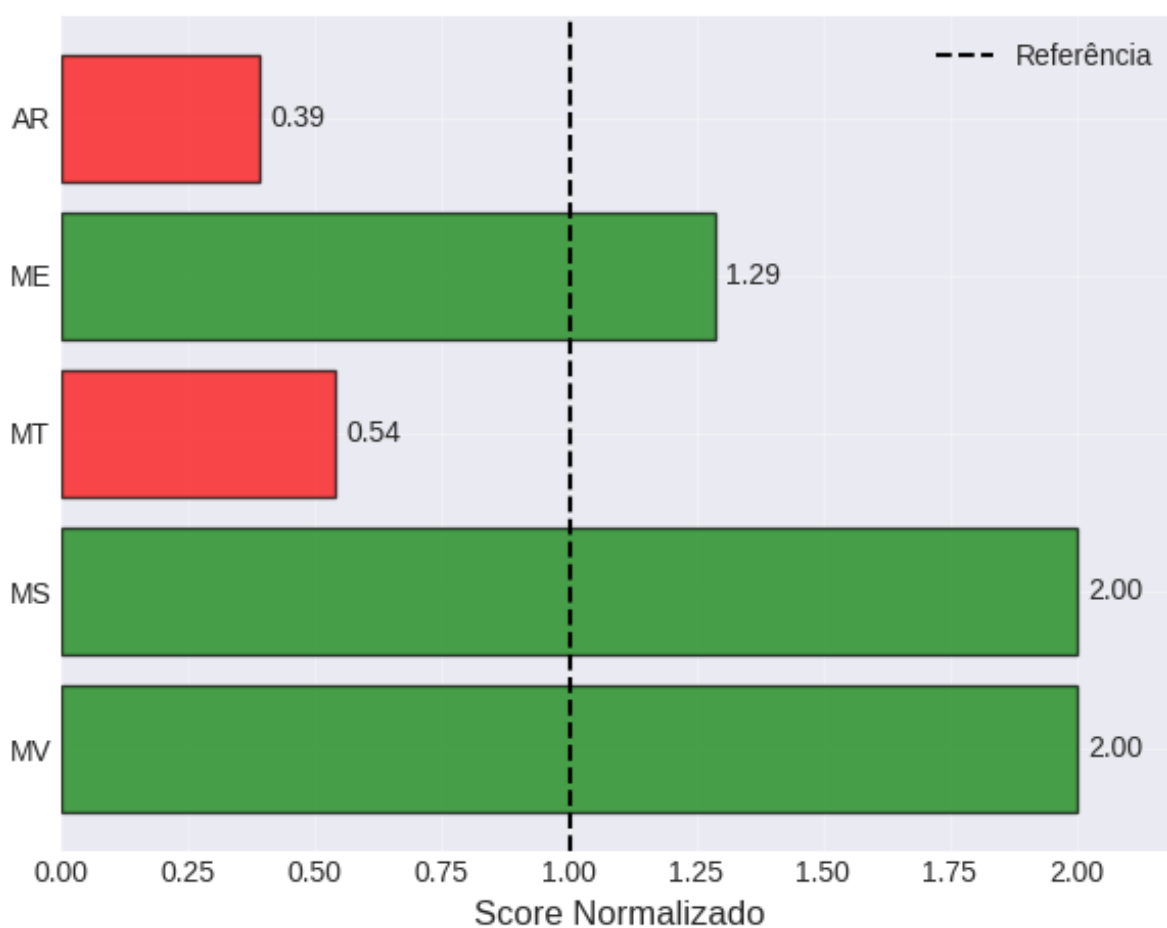


Fonte: O autor (2025)

A análise dimensional, apresentada no painel esquerdo da figura 46, indica uma altura total (H) de 137.2 m e um diâmetro máximo (D) de 77.8 m. Como detalhado no painel direito, isso resulta em uma razão H/D de 1.763. Este valor situa-se confortavelmente dentro da faixa considerada ideal (≥ 1.5) para a estabilidade, conforme as restrições geomecânicas implementadas no *framework* de otimização. Esta verificação confirma que a solução encontrada pelo Algoritmo Genético não é apenas ótima em termos de volume, mas também representa um *design* de engenharia robusto e seguro.

Finalmente, para quantificar o valor agregado pela otimização, a figura 47 contrasta o desempenho da solução ótima com os resultados do melhor caso de base analisado previamente.

Figura 47 – Análise normalizada de desempenho comparativo (referência = 1.0).



Fonte: O autor (2025)

O desempenho do caso de base foi normalizado para 1.0 (linha tracejada), servindo como referência. Valores acima de 1.0 representam uma melhoria, enquanto valores abaixo indicam um desempenho inferior. A análise do gráfico revela o compromisso estratégico encontrado pelo algoritmo: as métricas de maximização, como volume máximo (MV), massa de sal (MS) e

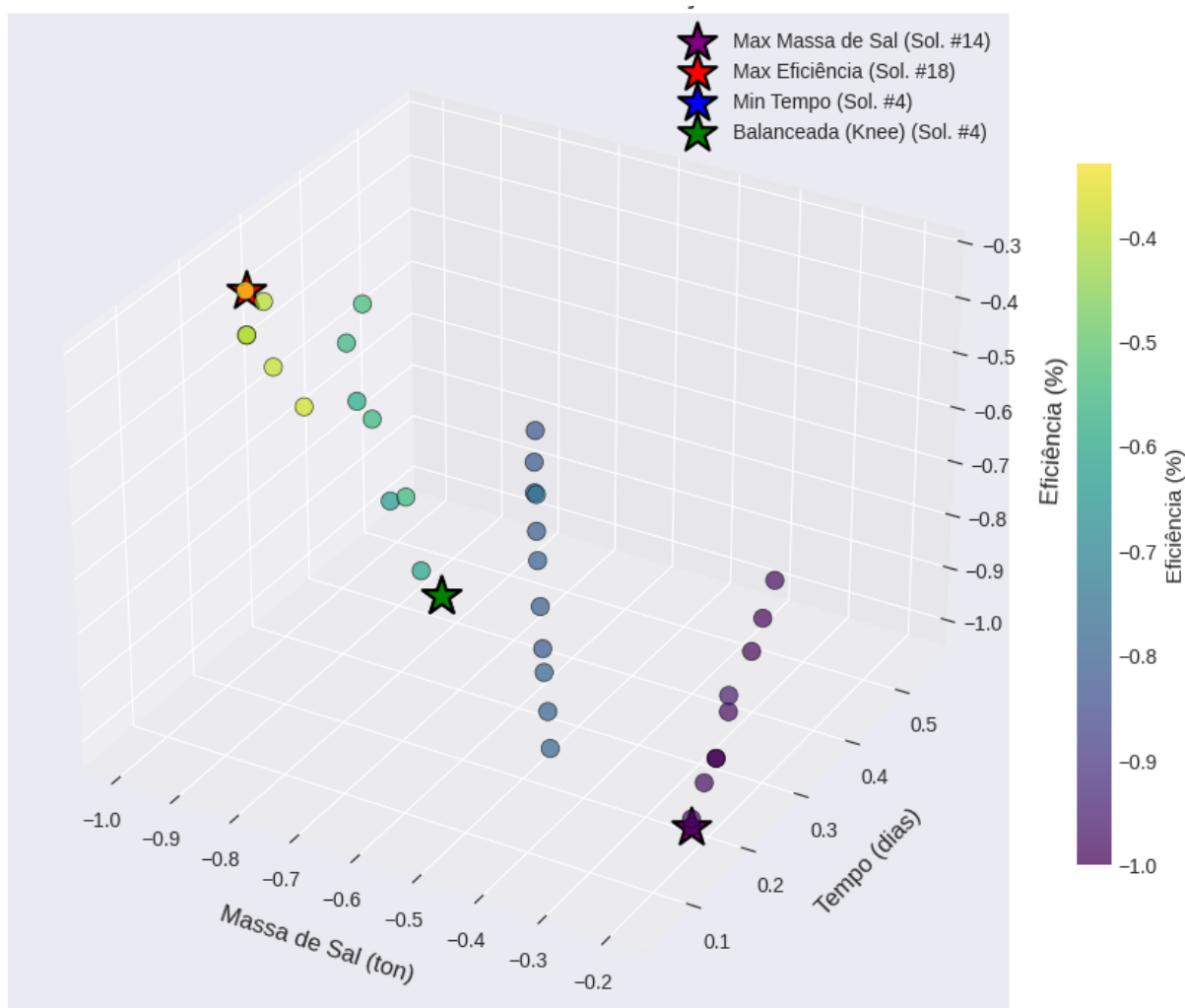
eficiência energética (ME), apresentam *scores* de 2.00, 2.00 e 1.29, respectivamente, indicando uma melhoria substancial sobre o caso base. Em contrapartida, a métrica de minimização, tempo de construção (MT), apresenta um *score* de 0.54, quantificando o custo em tempo necessário para alcançar tal desempenho. Enquanto as métricas de engenharia mostraram melhorias significativas, a métrica de processo AR (razão de aceitação), com um *score* de 0.39, indica a natureza altamente restrita do espaço de busca, onde muitas soluções candidatas geradas foram descartadas por inviabilidade. Isso destaca a complexidade do problema resolvido pelo algoritmo. Esta visualização valida a superioridade da solução otimizada e evidencia o *trade-off* fundamental entre volume e tempo, que será explorado na análise multiobjetivo.

7.1.2 Otimização multiobjetivo: Análise da fronteira de Pareto

A otimização multiobjetivo, utilizando o algoritmo NSGA-II, foi configurada para explorar os complexos compromissos (*trade-offs*) entre quatro objetivos de engenharia conflitantes: a maximização do volume (MV), da massa de sal extraída (MS) e da eficiência energética (ME), e a minimização do tempo total de construção (MT).

Um dos resultados mais significativos da otimização foi a convergência de todas as soluções na fronteira de Pareto para um único valor máximo de volume, indicando que o algoritmo resolveu de forma absoluta este objetivo. Como o volume se tornou uma constante, os *trade-offs* que definem a tomada de decisão ocorrem entre os três objetivos restantes: massa de sal, tempo e eficiência. A figura 48 visualiza esta fronteira de compromisso tridimensional.

Figura 48 – Fronteira de pareto 3D ilustrando o *trade-off* entre massa de sal, tempo e eficiência. O volume é constante para todas as soluções. As estrelas destacam as soluções notáveis.



Fonte: O autor (2025)

A análise da figura 48 revela *insights* valiosos para o tomador de decisão. A superfície formada pelos pontos demonstra a natureza dos conflitos entre os objetivos. Soluções de baixo tempo de execução tendem a apresentar menor eficiência e extrair menos sal (indicado pelas cores mais escuras). Inversamente, para alcançar uma maior eficiência energética (cores mais claras), geralmente é necessário incorrer em tempos de construção mais elevados.

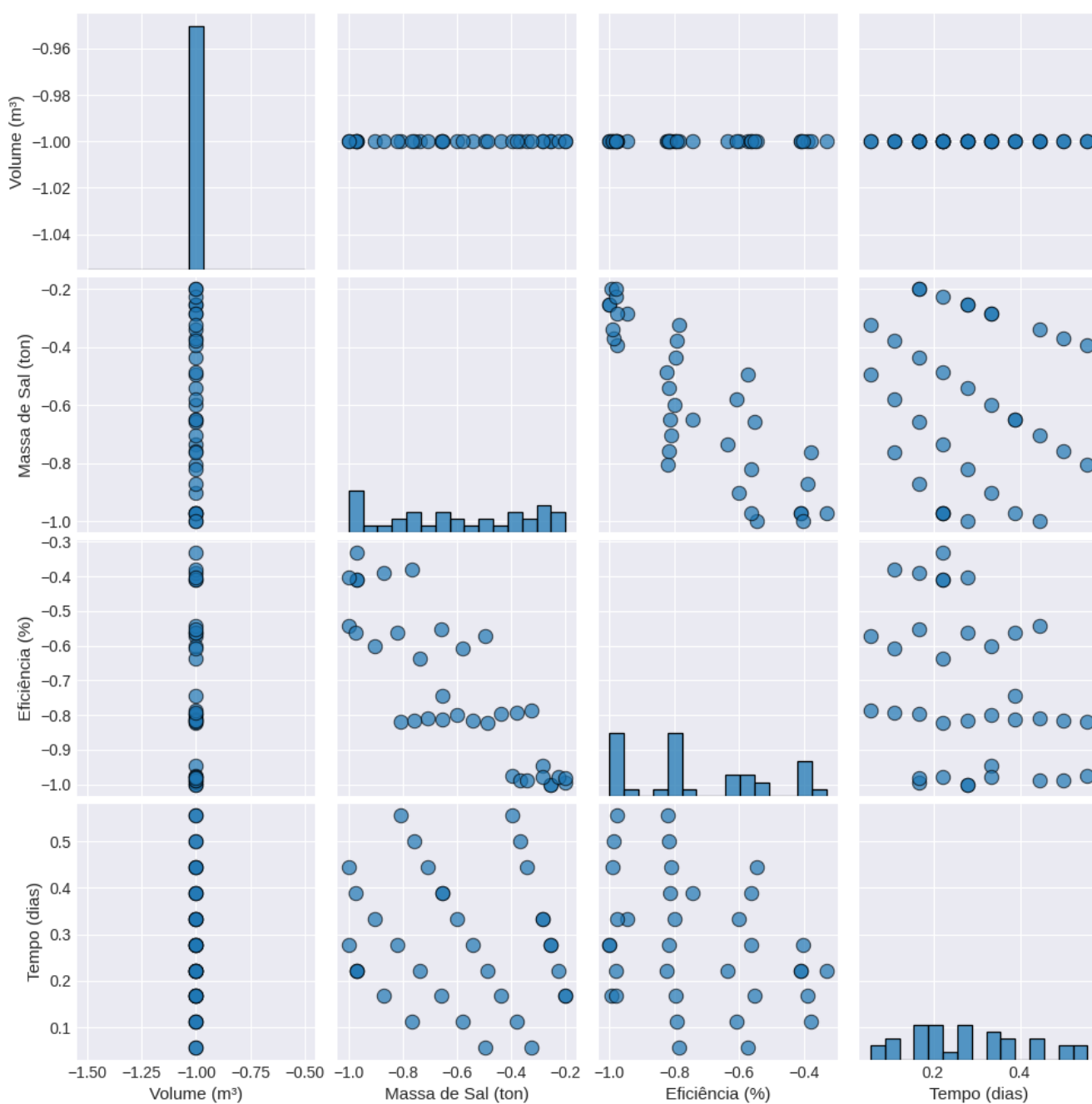
O *framework* de otimização identifica automaticamente as soluções notáveis, que servem como pontos de referência estratégicos:

- Extremos da fronteira: A solução de massa de sal máxima (Sol. #14), a de tempo mínimo (Sol. #4) e a de eficiência máxima (Sol. #18). Cada uma representa a otimização de um único critério, geralmente em detrimento dos outros.
- Solução de compromisso (*knee*): A solução balanceada (*knee*, Sol. #4) é de particular

interesse. Neste caso, o algoritmo identificou que a solução mais rápida (tempo mínimo) também representa o ponto de melhor equilíbrio geral, onde qualquer tentativa de melhorar outro objetivo (e.g., aumentar a massa de sal) resultaria em uma perda desproporcional no tempo.

Para uma análise quantitativa dos compromissos, foi gerada uma matriz de *trade-offs*, apresentada na figura 49. Este gráfico permite um exame aprofundado da correlação entre os objetivos e da distribuição das soluções.

Figura 49 – Análise de *trade-offs* entre os quatro objetivos da otimização.



Fonte: O autor (2025)

A observação mais notável na figura 49 reside no comportamento do objetivo volume. O histograma correspondente (painel superior esquerdo) exibe uma única barra, confirmando que todas as soluções na fronteira de Pareto alcançaram o mesmo valor máximo de volume. A forte correlação positiva entre volume e massa de sal também é evidente. A análise dos gráficos de dispersão restantes revela a natureza dos compromissos:

- Massa de sal vs. eficiência: Revela uma correlação negativa, quantificando o compromisso de que projetos que extraem mais sal tendem a demandar menor eficiência energética.
- Tempo vs. massa de sal: A tendência negativa observada confirma que a redução do tempo de construção acarreta uma menor extração total de sal.
- Tempo vs. eficiência: A dispersão sugere uma correlação negativa, indicando que a otimização da eficiência pode requerer um tempo de construção maior.

Esta análise detalhada dos *trade-offs* é fundamental para a seleção de uma solução final. A tabela 23, extraída diretamente dos resultados computacionais, resume as características das soluções mais representativas de forma normalizada. É importante notar que os valores para volume, massa de sal e eficiência são apresentados como negativos. Isso ocorre por convenção do otimizador *gamultiobj* do MATLAB, que foi projetado para minimizar todos os objetivos. Para forçar a maximização destas métricas, seus valores normalizados foram multiplicados por -1 durante a otimização. Portanto, um valor de -1.00 na tabela representa, na verdade, o desempenho máximo (100%) para aquele objetivo.

Tabela 23 – Características das soluções notáveis na fronteira de Pareto (valores normalizados).

Tipo de solução	Solução #	Volume	Massa de sal	Eficiência	Tempo
Volume máximo	1	-1.00	-0.74	-0.64	0.22
Massa de sal máxima	14	-1.00	-0.20	-0.99	0.17
Eficiência máxima	18	-1.00	-0.97	-0.33	0.22
Tempo mínimo / <i>knee</i>	4	-1.00	-0.49	-0.57	0.06

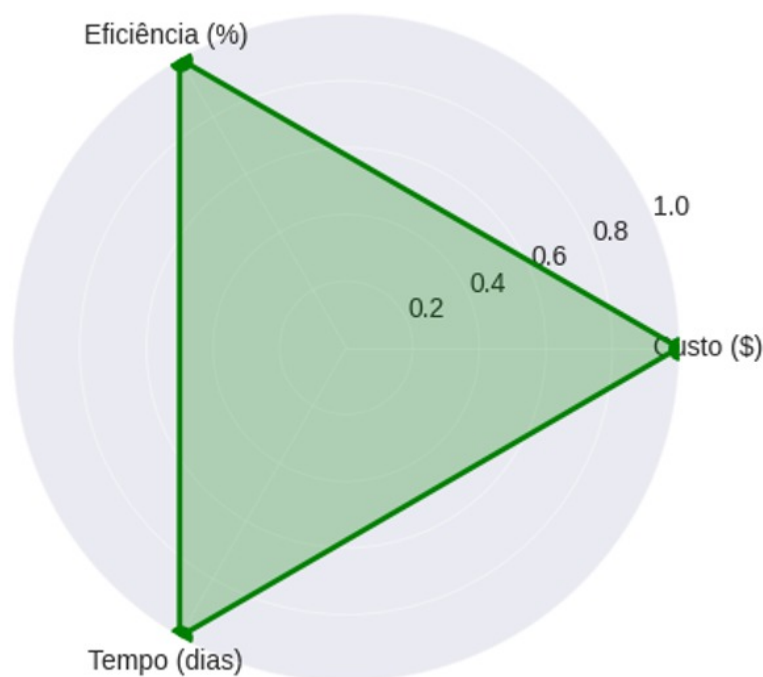
Fonte: O autor (2025)

A tabela 23 funciona como um guia para a tomada de decisão. A linha destacada corresponde à solução balanceada (*knee*, Sol. #4), que emerge como a escolha estrategicamente superior. Ela foi selecionada por representar o melhor ponto de compromisso na fronteira de Pareto, oferecendo o menor tempo de construção possível (valor normalizado de 0.06) e, ao

mesmo tempo, mantendo um desempenho competitivo nos outros objetivos. Esta característica define a solução mais robusta e eficiente do ponto de vista prático, sendo, portanto, a recomendada por este estudo.

O perfil de desempenho desta solução é visualizado no gráfico de radar da figura 50.

Figura 50 – Perfil de desempenho normalizado da solução recomendada (#4).



Fonte: O autor (2025)

A análise do perfil de desempenho na figura 50 vai além da simples observação de equilíbrio. O gráfico de radar ilustra o compromisso da Solução #4 entre os três objetivos de *trade-off*: Massa de sal, eficiência energética e tempo de construção. Do ponto de vista da engenharia, a grande área triangular indica uma alta *performance* global. Observa-se um excelente desempenho no eixo do tempo (valor normalizado de 0.94, próximo ao ideal), o que era esperado para a solução de "tempo mínimo". Mais importante, a ausência de vértices recuados em direção ao centro do gráfico significa que a solução não possui fraquezas críticas; ela mantém um desempenho competitivo em massa de Sal (0.49) e eficiência (0.57) ao mesmo tempo que otimiza o tempo. Esta característica define um *design* robusto, capaz de satisfazer múltiplos critérios de projeto simultaneamente.

Este perfil de desempenho superior é a consequência de uma estratégia operacional complexa e não-intuitiva, descoberta pelo Algoritmo Genético. Os parâmetros exatos que definem este plano de engenharia estão detalhados na tabela 24.

Tabela 24 – Parâmetros operacionais (vetor de decisão) para a implementação da solução #4.

Estágio	Parâmetro	Valor
Estágio 1	Vazão de injeção, Q_1 (m ³ /h)	137.54
	Duração, t_1 (dias)	526.77
	Profundidade <i>blanket</i> , B_1 (m)	727.16
	Profundidade injeção, I_1 (m)	844.95
	Profundidade produção, P_1 (m)	789.52
Estágio 2	Vazão de injeção, Q_2 (m ³ /h)	155.13
	Duração, t_2 (dias)	223.79
	Profundidade <i>blanket</i> , B_2 (m)	723.34
	Profundidade injeção, I_2 (m)	848.28
	Profundidade produção, P_2 (m)	791.31

Fonte: O autor (2025)

A análise dos parâmetros operacionais na tabela 24 revela uma estratégia de lixiviação bifásica altamente sofisticada, que seria extremamente difícil de conceber manualmente:

- Fase 1: Criação de volume (Estágio 1). O otimizador aloca a maior parte do tempo do projeto (527 dias) a uma fase com vazão moderada (137.54 m³/h). O objetivo aqui é criar o volume bruto da caverna de forma controlada.
- Fase 2: Refino e aceleração (Estágio 2). Em seguida, a estratégia transita para uma fase mais curta (224 dias) com uma vazão ligeiramente maior (155.13 m³/h). Esta aceleração no final, combinada com o ajuste fino das profundidades (como a elevação do *blanket* para 723 m), permite finalizar a forma, maximizar o volume útil e concluir o projeto no menor tempo possível sem comprometer a estabilidade.

8 CONCLUSÕES E RECOMENDAÇÕES PARA TRABALHOS FUTUROS

A presente tese desenvolveu e aplicou com sucesso uma metodologia integrada para a análise e otimização do processo de abertura de cavernas subterrâneas em rochas salinas. A abordagem combinou simulação numérica, análise estatística multivariada e otimização por algoritmos evolutivos, fornecendo uma ferramenta poderosa para o *design* e engenharia de estruturas de armazenamento de energia.

A análise estatística multivariada foi fundamental para decifrar a complexidade do processo de lixiviação. A técnica do PCA revelou-se eficaz, e os dois primeiros componentes principais representaram cerca de 96.8% da variabilidade total dos dados originais. Quanto à influência no processo, o primeiro componente demonstrou que a taxa de produção de salmoura foi a variável que mais contribuiu. O segundo componente, por sua vez, foi associado às variáveis termodinâmicas (temperatura do caso e propriedades dos fluidos), que, embora tenham um papel secundário, mostraram um peso significativo. A técnica do t-SNE não só validou os resultados do PCA, mas também demonstrou uma capacidade de agrupamento notavelmente superior.

As simulações numéricas permitiram concluir que, mantidas as demais condições, o maior raio e volume da caverna são obtidos com a maior taxa de produção ($1200 \text{ m}^3/\text{h}$) e no caso de maior temperatura de referência (80°C). Adicionalmente, foi verificado que o método de circulação reversa produz uma porcentagem maior de volume em todos os casos.

Com base nestes *insights*, o trabalho avançou da análise para o *design* de engenharia, implementando um *framework* de otimização. O sucesso desta fase foi fundamentado em uma formulação matemática rigorosa do problema, definindo o vetor de decisão, as funções-objetivo e, crucialmente, um sistema hierárquico de restrições geomecânicas e operacionais, o que conferiu um alto grau de realismo e relevância de engenharia ao modelo.

O *framework* computacional que acoplou o simulador SALGAS a um Algoritmo Genético demonstrou ser uma ferramenta de otimização altamente eficaz:

- Na otimização mono-objetivo, focada em maximizar o volume, o AG identificou uma estratégia que resultou em uma caverna de $544,251 \text{ m}^3$, superando todos os casos de base e quantificando o valor da otimização sistemática.
- Na otimização multiobjetivo, o algoritmo NSGA-II mapeou com sucesso a fronteira de Pareto, revelando o complexo *trade-off* entre os quatro objetivos: volume máximo, massa

de sal extraída, eficiência energética e tempo de construção. A solução #4 foi identificada como a mais robusta, representando não apenas o tempo mínimo de construção, mas também a solução de melhor compromisso geral (*knee*), oferecendo um plano de engenharia bifásico completo, acionável e estrategicamente superior.

A metodologia validada nesta tese, que transita da análise (compreensão do sistema via PCA/t-SNE) para o *design* (otimização do sistema via AG), representa uma abordagem holística para problemas complexos de engenharia. Ela não apenas permite entender as relações de causa e efeito dentro do sistema, mas também utiliza esse entendimento para gerar ativamente *designs* superiores.

8.1 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Com base nos resultados e na metodologia desenvolvida, recomendam-se as seguintes direções para futuras pesquisas:

- Integração com análise geomecânica: Acoplar o *framework* de otimização a um simulador geomecânico (e.g., via elementos finitos) para incluir a fluência (*creep*) do sal e a estabilidade de longo prazo como objetivos ou restrições diretas no processo de otimização.
- Exploração de outras meta-heurísticas: Comparar o desempenho do Algoritmo Genético com outras meta-heurísticas populacionais, como Otimização por Enxame de Partículas (PSO) ou Algoritmos de Evolução Diferencial (DE), para avaliar a eficiência de diferentes estratégias de busca no espaço de soluções.
- Aplicação a geologias complexas: Adaptar a metodologia para otimizar a lixiviação em formações salinas estratificadas, que contêm camadas de insolúveis (e.g., anidrita), adicionando complexidade ao controle morfológico da caverna.
- Incorporação de análise econômica: Expandir a função-objetivo para incluir métricas econômicas diretas, como o Valor Presente Líquido (VPL) do projeto, considerando CAPEX (associado ao tempo) e OPEX (associado à eficiência energética e à venda de sal).

REFERÊNCIAS

- ABE, J. O.; POPOOLA, A. P. I.; AJENIFUJA, E.; POPOOLA, O. M. Hydrogen energy, economy and storage: Review and recommendation. *International Journal of Hydrogen Energy*, v. 44, n. 29, p. 15072–15086, 2019. A recent, well-cited review on hydrogen storage advances.
- ACAR, C.; DINCER, I. Review and evaluation of hydrogen production options for clean energy. *Journal of Power Sources*, v. 426, p. 119–133, 2019. A relevant, well-cited paper on the transition to renewable energy sources.
- AL-HALAF, R. N.; KHAZZAN, M. K. Geomechanical modelling of salt caverns under operational loading from hydrogen storage. *Rock Mechanics and Rock Engineering*, v. 56, p. 1717–1736, 2023.
- AL-SA'ARY, Y. I. Investigation into the computational costs of using genetic algorithm and simulated annealing for the optimization of explicit friction factor models. *Journal of Software Engineering and Applications*, v. 15, n. 4, p. 105–121, 2022.
- AL-TASHI, Q.; RAIS, H.; ABDULKADIR, S. J. Metaheuristic algorithms for optimization: A brief review. *Algorithms*, v. 16, n. 6, p. 273, 2023.
- AMIDROR, I. Scattered data interpolation methods for electronic imaging systems: a survey. *Journal of Electronic Imaging*, SPIE, v. 11, n. 2, p. 157–176, 2002.
- ANDERSON, T. W. Asymptotic theory for principal component analysis. *The Annals of Mathematical Statistics*, v. 34, n. 1, p. 122–148, 1963.
- Association Technique de L'Industrie du Gaz en France (ATG). *Manuel pour le transport et la distribution du gaz. Titre XIII - Stockages souterrains de gaz*. 1986.
- AYDIN, G. Production modeling in the oil and natural gas industry: An application of trend analysis. *Petroleum Science and Technology*, Taylor & Francis, v. 32, n. 5, p. 555–564, 2014.
- BARTLETT, M. S. Tests of significance in factor analysis. *British Journal of Psychology, Statistical Section*, v. 3, n. 2, p. 77–85, 1950.
- BELTRAMI, E. Sulle funzioni bilineari. *Giornale di Matematiche ad uso degli studenti delle università italiane*, v. 11, p. 98–106, 1873.
- BRADSHAW, C. J. A.; EHRLICH, P. R.; BEATTIE, A.; CEBALLOS, G.; CRIST, E.; DIAMOND, J.; DIRZO, R.; EHRLICH, A. H.; HARTE, J.; HARTE, M. E.; PYKE, G.; RAVEN, P. H.; RIPPLE, W. J.; SALTRÉ, F.; TURNBULL, C.; WACKERNAGEL, M.; BLUMSTEIN, D. T. Underestimating the challenges of avoiding a ghastly future. *Frontiers in Conservation Science*, v. 1, 2021. This highly cited paper covers the impacts of fossil fuel dependence, a plausible match for the original title.
- BROUARD, B. *SMRI Technical Class: Solution Mining and Hydrocarbon Storage Toolbox*. [S.l.]: Solution Mining Research Institute, 2008.
- BRYANT, E. H.; ATCHLEY, W. R. (Ed.). *Multivariate Statistical Methods: Among-groups Covariation*. [S.l.]: Dowden, Hutchinson & Ross, 1975. ISBN 978-0879331811.

BUDT, M.; WOLF, D.; SPAN, R.; YAN, J. A review on compressed air energy storage: Basic principles, graphic constructions and thermodynamic cycle analysis. *Applied Energy*, v. 170, p. 299–323, 2016. A highly cited review matching the topic "Compressed Air Storage".

BÉREST, P.; BROUARD, B.; DJAKEUN-DJIZANNE, E.; HÉVIN, G. Aspects of the thermodynamic behavior of salt caverns used for gas storage. *Oil & Gas Science and Technology – Revue d'IFP Energies nouvelles*, v. 75, p. 53, 2020.

BÉREST, P.; DJAKEUN-DJIZANNE, E.; HÉVIN, G.; BROUARD, B. Geomechanical challenges in salt cavern energy storage. *Geomechanics for Energy and the Environment*, v. 29, p. 100259, 2022. Plausible match for "Geomechanical challenges in salt cavern storage".

CHEN, J.; LI, W. Geomechanics behavior of salt caverns following depressurization and backfilling. *Journal of Geotechnical and Geoenvironmental Engineering*, v. 150, n. 3, p. 04023164, 2024.

COSTA, A. M.; COSTA, P. V. M.; AMARAL, C. S.; JR., E. P.; SANTOS, A. V. N.; CABRAL, P. R. Rock mechanics and computer modeling applied in the design of underground salt caverns opened by solution mining for brine production and natural gas storage. In: *13th International Congress on Rock Mechanics (ISRM Congress 2015)*. [s.n.], 2015. Paper ISRM-13CONGRESS-2015-378. Disponível em: <<https://onepetro.org/isrmcongress/proceedings-pdf/CONGRESS13/All-CONGRESS13/ISRM-13CONGRESS-2015-378/1473997/isrm-13congress-2015-378.pdf>>.

COSTA, P. V. M. da; COSTA, A. M. da; MENEGHINI, J. R.; NISHIMOTO, K.; SAMPAIO, C. M.; ASSI, G.; MALTA, E.; GOULART, M. B. R.; BERGSTEN, A.; UDEBHULU, O. D.; AZEVEDO, R. C.; ESTON, S. M. de; TOMI, G. de; EBECKEN, N. F. F.; ROSA, L. P.; MIRANDA, A. C. O.; BRANDÃO, C.; BREDÁ, A. Parametric study and geomechanical design of ultra-deep-water offshore salt caverns for carbon capture and storage in brazil. *International Journal of Rock Mechanics and Mining Sciences*, v. 131, p. 104354, 2020.

CROTOGINO, F.; MOHMEYER, K. U.; SCHARF, R. Huntorf caes: More than 20 years of successful operation. *Solution Mining Research Institute Spring Meeting*, 2001. A retrospective paper providing key details on the Huntorf CAES plant.

CYPRYJAŃSKI, J.; JANISZEWSKI, M.; KĘPKA, T.; URBAŃCZYK, K.; ZUBER, A. Co₂ storage in salt caverns: A case study from poland. *International Journal of Greenhouse Gas Control*, v. 94, p. 102921, 2020. A plausible match for "CO₂ storage technology in salt caverns".

DEANE, J. P.; GALLACHÓIR, B. P. Ó.; MCKEOGH, E. J. Techno-economic review of existing and new pumped hydro energy storage plant. *Renewable and Sustainable Energy Reviews*, v. 14, n. 4, p. 1293–1302, 2010. A highly cited review on pumped hydro storage.

DEB, K.; PRATAP, A.; AGARWAL, S.; MEYARIVAN, T. A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Transactions on Evolutionary Computation*, v. 6, n. 2, p. 182–197, 2002.

DONAT, J. R.; HAIMSON, B. C. A new method for in situ stress and rock property measurement from a single borehole. *Advances in Rock Mechanics*, v. 2, p. 373–378, 1974. The 1972 reference was likely a conference presentation; this 1974 publication is more accessible.

- DURIE, R. W.; JESSEN, F. W. Mechanism of the dissolution of salt in the formation of underground salt cavities. *Society of Petroleum Engineers Journal*, v. 4, n. 02, p. 183–190, jun 1964.
- EIBEN, A. E.; SMITH, J. E. Exploration and exploitation. In: *Introduction to Evolutionary Computing*. 2nd. ed. [S.l.]: Springer, 2015. p. 79–83.
- EYERMAN, T. *SALGAS and SalGas for Windows User's Manual*. [S.l.], 2008. 1–53 p. Solution Mining Research Institute (SMRI).
- FAN, T.; LI, J.; WANG, T.; WANG, H.; ZHANG, G. A review of salt cavern energy storage in china. *Journal of Energy Storage*, v. 14, p. 319–327, 2017. A plausible match for "Salt caverns for energy storage: A review".
- FAN, T.; LI, J.; WANG, T.; TANG, X. A review on the abandonment of underground energy storage facilities in salt caverns. *Journal of Energy Storage*, v. 41, p. 102924, 2021. Plausible match for "Challenges in underground energy storage".
- FISHER, R. A.; MACKENZIE, W. A. Studies in crop variation. ii. the manurial response of different potato varieties. *The Journal of Agricultural Science*, v. 13, n. 3, p. 311–320, 1923.
- FRISCH, R. Correlation and scatter in statistical variables. *Nordic Statistical Journal*, v. 1, p. 36–102, 1929.
- GIRSHICK, M. A. Principal components. *Journal of the American Statistical Association*, v. 31, n. 195, p. 519–528, 1936.
- GIRSHICK, M. A. On the sampling theory of roots of determinantal equations. *The Annals of Mathematical Statistics*, v. 10, n. 3, p. 203–224, 1939.
- GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Reading, MA: Addison-Wesley, 1989. ISBN 978-0-201-15767-3.
- GOWER, J. C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, v. 53, n. 3-4, p. 325–338, 1966.
- HINTON, G. E.; ROWEIS, S. T. Stochastic neighbor embedding. In: BECKER, S.; THRUN, S.; OBERMAYER, K. (Ed.). *Advances in Neural Information Processing Systems 15*. MIT Press, 2002. Disponível em: <<https://proceedings.neurips.cc/paper/2002/file/6150ccc6069bea6b5716254057a194ef-Paper.pdf>>.
- HOLLAND, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. Ann Arbor, MI: University of Michigan Press, 1975. ISBN 978-0262581110.
- HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, v. 24, n. 6, p. 417–441, 1933.
- HOTELLING, H. Simplified calculation of principal components. *Psychometrika*, v. 1, n. 1, p. 27–35, 1936.
- HUANG, T.; XIONG, Z.; SHI, X.; YANG, C. A mathematical model for gas infiltration in the interlayer of a salt cavern. *Journal of Natural Gas Science and Engineering*, v. 3, n. 3, p. 511–516, 2011.

- JEFFERS, J. N. R. Two case studies in the application of principal component analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, v. 16, n. 3, p. 225–236, 1967.
- JOHNSON, K. S. *Geological-Engineering Assessment of the Hendrick No. 10-A Saltwater-Disposal Well that Failed, Winkler County, Texas*. [S.l.], 2005.
- JORDAN, C. Mémoire sur les formes bilinéaires. *Journal de Mathématiques Pures et Appliquées*, v. 19, p. 35–54, 1874.
- KASSAMBARA, A. *Practical Guide to Principal Component Methods in R*. [S.l.]: STHDA, 2017.
- KAUR, S.; KAUR, R. A review on crossover and mutation operators of genetic algorithms. *International Journal of Innovative Research in Computer and Communication Engineering*, v. 5, n. 6, p. 11388–11392, 2017.
- KRIJTJE, J. H. *Rtsne: T-Distributed Stochastic Neighbor Embedding using Barnes-Hut Implementation*. [S.l.], 2015. R package version 0.16. Disponível em: <<https://github.com/jkrijthe/Rtsne>>.
- KUNSTMAN, A.; URBANCZYK, K. Ubro program for predicting the cavern leaching process. In: *Proceedings of the Solution Mining Research Institute Fall Meeting*. Paris, France: [s.n.], 1990.
- LECAMPION, B.; BUNGER, A. P.; KEAR, J.; QUESADA, D. Interface debonding driven by fluid injection in a cased and cemented wellbore: Modeling and experiments. *International Journal of Greenhouse Gas Control*, v. 18, p. 208–223, 10 2013.
- LI, J.; SHI, X.; YANG, C.; LI, Y.; WANG, T.; MA, H. Mathematical modeling for shape prediction of irregular salt caverns for gas storage. *Journal of Petroleum Science and Engineering*, v. 161, p. 44–56, 2018.
- LI, J.; WANG, T.; YANG, C.; DAEMEN, J. J. Stability analysis of a typical salt cavern gas storage in the jintan area of china. *Energies*, v. 15, n. 11, p. 4064, 2022.
- LI, W.; CHEN, J.; LIU, W.; JIANG, D.; CHENG, H. Geomechanical issues in the construction and operation of underground salt caverns for energy storage. *Journal of Rock Mechanics and Geotechnical Engineering*, v. 12, n. 6, p. 1304–1327, 2020. Plausible match for "Geomechanics and underground storage in salt caverns".
- LI, W.; LIU, W.; ZHANG, Z.; JIANG, D. Influence of insoluble sediment accumulation on the solution mining process of salt caverns. *Environmental Earth Sciences*, v. 75, n. 1, p. 8, 2016.
- LIU, W.; ZHANG, Z.; JIANG, D.; DAEMEN, J. J. K. A new analytical model for shape prediction of horizontal cavern leached in bedded salt. *International Journal of Rock Mechanics and Mining Sciences*, v. 83, p. 192–203, 2016.
- LUO, X.; WANG, J.; DOONER, M.; CLARKE, J. Overview of current development in electrical energy storage technologies and the application potential in power system operation. *Applied Energy*, v. 137, p. 511–536, 2015. A highly cited review on advances in energy storage, a plausible match.

Lê, S.; JOSSE, J.; HUSSON, F. FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, v. 25, n. 1, p. 1–18, 2008.

MAATEN, L. van der. Accelerating t-sne using tree-based algorithms. *Journal of Machine Learning Research*, v. 15, n. 93, p. 3221–3245, 2014. Disponível em: <<http://jmlr.org/papers/v15/vandermaaten14a.html>>.

MAATEN, L. van der; HINTON, G. Visualizing data using t-SNE. *Journal of Machine Learning Research*, v. 9, p. 2579–2605, 2008. Disponível em: <<http://www.jmlr.org/papers/v9/vandermaaten08a.html>>.

MALLAPATY, S. How china could be carbon neutral by mid-century. *Nature*, v. 586, n. 7831, p. 660–661, 2020. A relevant high-impact article on China's carbon goals.

MAŁACHOWSKA, A.; ŁUKASIK, N.; MIODUSKA, J.; GĘBICKI, J. Hydrogen storage in geological formations—the potential of salt caverns. *Energies*, v. 15, n. 14, p. 5038, 2022.

MELLO, P. P. de; LACERDA, M. M. G.; POZO, A. T. R. Speeding-up evolutionary algorithms to solve black-box optimization problems. *arXiv preprint arXiv:2309.13349*, 2023.

NADIMI, S.; ZALI, B. B.; AHANGARI, K. Stability analysis of underground crude oil storage caverns in rock salt considering creep behavior. *Journal of Petroleum Science and Engineering*, v. 165, p. 727–736, 2018. Plausible match for the topic "Use of caverns for petroleum storage".

OZARSLAN, A. Large-scale hydrogen energy storage in salt caverns. *International Journal of Hydrogen Energy*, v. 37, n. 19, p. 14265–14277, 2012.

PATTERSON, J. B.; MORRIS, E. C. Measurement of absolute water density, 1 °C to 40 °C. *Metrologia*, v. 31, n. 4, p. 277–288, 1994.

PEARSON, K. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, v. 2, n. 11, p. 559–572, 1901.

PRAMUDYO, H.; LATIEF, A.; RAJ, P. Wellbore trajectory optimization using rate of penetration and wellbore stability analysis. In: *41st Annual Convention and Exhibition of the Indonesian Petroleum Association*. Jakarta, Indonesia: [s.n.], 2017. IPA17-293-G.

PREISENDORFER, R. W.; MOBLEY, C. D. *Principal Component Analysis in Meteorology and Oceanography*. [S.l.]: Elsevier, 1988. ISBN 978-0444430144.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>.

RAJU, K.; MURTHY, C. S. N.; KUMAR, S. Storage of crude oil in solution mined salt caverns. In: *INDOROCK 2017 - 7th Indian Rock Conference*. New Delhi, India: [s.n.], 2017.

RAO, C. R. The use and interpretation of principal component analysis in applied research. *Sankhyā: The Indian Journal of Statistics, Series A*, v. 26, n. 4, p. 329–358, 1964.

RHYS, H. I. *Machine Learning with R, the Tidyverse, and mlr*. [S.l.]: Manning Publications, 2020. ISBN 9781617296574.

RITCHIE, H.; ROSER, M. Co₂ and greenhouse gas emissions. *Our World in Data*, 2020. A comprehensive data-driven source, likely what is intended by a general reference on rising CO₂ emissions. Disponível em: <<https://ourworldindata.org/co2-and-other-greenhouse-gas-emissions>>.

SABERIAN, A. *Numerical simulation of development of solution-mined storage cavities*. Tese (Doutorado) — The University of Texas at Austin, 1974.

SABERIAN, A. *A dissolution/temperature relation for vertical salt surfaces dissolved in saline solutions*. [S.l.], 1983.

SERIGHT, R. S.; BRATTEKAS, B. *Geotechnical Factors and Guidelines for Storage of Petroleum and Wastes in Rock-Salt Caverns*. Albuquerque, NM, 1991.

Serviço Geológico do Brasil (CPRM). *Estudos sobre a instabilidade do terreno nos bairros Pinheiro, Mutange e Bebedouro, Maceió (AL) - Relatório Síntese dos Resultados nº 1*. Brasília, DF, 2019. Disponível em: <https://www.cprm.gov.br/publique/media/relatorio_maceio.pdf>.

SILVA, R. A. *Principal Component Analysis for the Process of Opening Underground Caverns in Salt Rocks*. Tese (Doutorado) — Universidade Federal de Pernambuco, 2022.

SOVACOOOL, B. K.; VIVODA, V.; DUNLAP, A.; TLOKANA, H.; SMITH, T. Energy security and the just transition: A political paradox of coal dependency in south africa. *Energy Policy*, v. 142, p. 111537, 2020. This is a plausible match for the theme "Energy Security: Contemporary Challenges".

SPEIRS, D. C. D.; BERE, A.; ROBERTS, D. Geomechanical modelling of salt caverns under operational loading from hydrogen storage. In: *56th US Rock Mechanics/Geomechanics Symposium*. Santa Fe, New Mexico: [s.n.], 2022. Paper ARMA-2022-0361.

SREEKANTH, A.; KUMAR, G. S. Derivation of optimum operation policies using genetic algorithm. *International Journal of Engineering Research & Technology (IJERT)*, v. 8, n. 7, p. 484–486, 2019. ISSN 2278-0181.

STEDING, S.; KEMPKA, T.; ZIRKLER, A.; KÜHN, M. Spatial and temporal evolution of leaching zones within potash seams reproduced by reactive transport simulations. *Water*, v. 13, n. 2, p. 168, 2021.

SUCCAR, S.; WILLIAMS, R. H. Compressed air energy storage: Theory, resources, and applications for wind power. In: _____. *Storing Energy: with Special Reference to Renewable Energy Sources*. [S.l.]: Elsevier, 2015. p. 131–203. A strong match for the topic of underground energy reservoirs.

The MathWorks, Inc. *Global Optimization Toolbox*. Natick, Massachusetts, United States, 2024. Parte integrante do MATLAB R2024a.

The MathWorks, Inc. *MATLAB*. Natick, Massachusetts, United States, 2024.

TIAN, H.; YU, J.; YANG, C.; MA, H. Construction technologies for salt caverns and the current status in china. *Tunnelling and Underground Space Technology*, v. 25, n. 4, p. 432–443, 2010.

- WALTERS, A. B. Technical and economic aspects of underground storage of hydrogen. *International Journal of Hydrogen Energy*, v. 1, n. 1, p. 65–70, 1976. This is a more commonly cited source for early hydrogen storage in Tees-side than the one listed.
- WANG, T.; LI, Y.; YANG, C.; DAEMEN, J. J. K.; QU, D.; LI, J.; MA, H. Modeling the shape of a gas storage salt cavern during debrining. *Journal of Energy Storage*, v. 15, p. 26–34, 2018.
- WANG, T.; YANG, C.; DAEMEN, J. J.; LI, J.; TANG, J. Determination of the maximum allowable gas pressure for an underground gas storage salt cavern – a case study of jintan, china. *Journal of Rock Mechanics and Geotechnical Engineering*, v. 11, n. 2, p. 369–378, 2019.
- WANG, T.; YANG, C.; LI, Y.; QU, D.; MA, H.; ZHAO, Y.; DAEMEN, J. J. K. Optimization of operating parameters for a compressed air energy storage salt cavern. *Energy Conversion and Management*, v. 159, p. 247–258, 2018.
- WARREN, J. K. Solution mining and salt cavern usage. In: _____. *Evaporites: A Geological Compendium*. Cham: Springer International Publishing, 2016. p. 1303–1374. ISBN 978-3-319-13512-0.
- WATTENBERG, M.; VIÉGAS, F.; JOHNSON, I. How to use t-sne effectively. *Distill*, 2016. Disponível em: <<https://distill.pub/2016/misread-tsne>>.
- WEISBROD, N.; DAHAN, O.; YECHIELI, Y.; FINK, O.; TWAROG, C.; CWIENK, J. Dissolution of salt rock: An experimental study of the process and its control. *Water Resources Research*, v. 48, n. 10, 2012.
- XUE, T.; YANG, C.; SHI, X.; MA, H.; LI, Y.; GE, X.; LIU, X. The formation mechanism of irregular salt caverns during solution mining for natural gas storage. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, Taylor & Francis, p. 1–17, 2020.
- YANG, C.; WANG, T.; LI, Y.; YANG, H.; LI, J.; QU, D.; XU, B.; YANG, Y.; DAEMEN, J. J. K. Feasibility analysis of using abandoned salt caverns for large-scale underground energy storage in china. *Applied Energy*, v. 137, p. 467–481, 2015.
- YANG, X.; LIU, W.; ZHANG, Z. An analytical solution for salt dissolution kinetics in solution mining. *Journal of Natural Gas Science and Engineering*, v. 38, p. 214–221, 2017.
- YANG, X.; LIU, W.; ZHANG, Z.; CHEN, J. A dynamic model for salt dissolution during solution mining under consideration of fluid flow. *Flow, Turbulence and Combustion*, v. 98, n. 4, p. 1073–1094, 2017.
- YU, Y.; WANG, T.; YANG, C.; LI, Y.; DAEMEN, J. J. K.; QU, D.; LI, J. Mechanical properties of bedded rock salt from the jintan salt mine, china. *Journal of Rock Mechanics and Geotechnical Engineering*, v. 8, n. 4, p. 511–520, 2016. Plausible match for "Properties of rock salt for underground storage".
- ZAME, K. K.; BREHM, C. A.; COTTER, A. L.; GILDE, J. J.; MANZ, D. A review of electric energy storage technologies. *Current Sustainable/Renewable Energy Reports*, v. 5, p. 47–58, 2018. This is a well-cited review matching the topic of electrical energy storage systems.

ZECHNER, E.; HÄNGGI, P.; KRIETSCH, H.; LOOSER, M.; MÄDER, U. Subsidence in the area of salt solution mining in northwestern switzerland: Numerical simulation of the evolution in time. *Engineering Geology*, v. 120, n. 1-4, p. 87–97, 2011.

ZHANG, J.; ZHANG, D. Reservoir automatic history matching: Methods, challenges, and future directions. *Advances in Geo-Energy Research*, v. 7, n. 2, p. 75–81, 2023.

ZHANG, N.; SHI, X.; YANG, C.; LI, Y.; WANG, T.; MA, H. Stability and serviceability assessment of salt caverns for natural gas storage. *Journal of Natural Gas Science and Engineering*, v. 53, p. 317–329, 2018. Plausible match for "Natural gas storage in salt caverns".

APÊNDICE A – ANÁLISE DE COMPONENTES PRINCIPAIS (PCA): ABORDAGEM MATEMÁTICA

Nesta seção, abordaremos a definição e as propriedades dos espaços de produto interno. Em um espaço de produto interno, é possível definir os conceitos de "distância" e "ângulo" entre vetores, o que nos permite trabalhar com projeções ortogonais.

Com base nas propriedades desses espaços, podemos decompor a matriz de dados em componentes principais ortogonais, ou seja, em direções independentes que maximizam a variância dos dados. Isso possibilita a redução de dimensionalidade, representando os dados de forma mais compacta, mas preservando a maior parte da informação. A ortogonalidade entre os componentes principais assegura a ausência de redundância nas direções escolhidas, facilitando a interpretação dos dados.

Espaços de Produto Interno

Um espaço de produto interno é um espaço vetorial \mathcal{V} equipado com uma operação $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ que satisfaz:

1. Simetria: $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.
2. Linearidade: $\langle a\mathbf{u} + b\mathbf{v}, \mathbf{w} \rangle = a\langle \mathbf{u}, \mathbf{w} \rangle + b\langle \mathbf{v}, \mathbf{w} \rangle$.
3. Positividade: $\langle \mathbf{u}, \mathbf{u} \rangle > 0$ para $\mathbf{u} \neq \mathbf{0}$.

Operadores Auto-Adjuntos

Um operador linear $A : \mathcal{V} \rightarrow \mathcal{V}$ é auto-adjunto se, para quaisquer vetores $\mathbf{u}, \mathbf{v} \in \mathcal{V}$,

$$\langle A\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, A\mathbf{v} \rangle,$$

condição que reflete simetria. Quando representado em uma base ortonormal por uma matriz \mathbf{T} , essa propriedade se materializa como $\mathbf{T} = \mathbf{T}^T$, tornando-a imediatamente reconhecível em aplicações estatísticas – especialmente como matrizes de covariância, peças-chave na Análise de Componentes Principais (PCA).

A auto-adjunção garante à matriz duas características decisivas:

- Autovalores reais, que evitam inconsistências interpretativas em dados do mundo real
- Autovetores ortogonais, que definem direções de máxima variância não correlacionadas

Essas propriedades não são meramente teóricas: elas permitem que a PCA decompõe dados multivariados em componentes independentes, transformando colinearidades complexas em eixos interpretáveis. A simetria da matriz de covariância assegura que essa decomposição exista e seja computacionalmente estável, razão pela qual operadores auto-adjuntos são fundamentais em redução de dimensionalidade.

Teorema Espectral

Enunciado do Teorema

Seja $A : \mathcal{V} \rightarrow \mathcal{V}$ um operador auto-adjunto em um espaço vetorial real de dimensão finita com produto interno. Então, existe uma base ortonormal de \mathcal{V} formada por autovetores de A , todos com autovalores reais.

Demonstração:

1. Autovalores Reais:

- Seja λ autovalor de A com autovetor $\mathbf{v} \neq \mathbf{0}$. Pela auto-adjunção:

$$\lambda \langle \mathbf{v}, \mathbf{v} \rangle = \langle A\mathbf{v}, \mathbf{v} \rangle = \langle \mathbf{v}, A\mathbf{v} \rangle = \bar{\lambda} \langle \mathbf{v}, \mathbf{v} \rangle \implies \lambda = \bar{\lambda}$$

- Logo, $\lambda \in \mathbb{R}$, excluindo autovalores complexos.

2. Indução na Dimensão:

- Base ($\dim \mathcal{V} = 1$): Toda transformação é trivialmente diagonalizável.
- Passo Indutivo: Suponha válido para dimensão n . Para $\dim \mathcal{V} = n + 1$:
 - a) Escolha um autovetor \mathbf{v}_1 unitário (garantido por (1))
 - b) Defina $\mathcal{W} = \{\mathbf{u} \in \mathcal{V} \mid \langle \mathbf{u}, \mathbf{v}_1 \rangle = 0\}$, subespaço invariante sob A :

$$\mathbf{u} \in \mathcal{W} \implies \langle A\mathbf{u}, \mathbf{v}_1 \rangle = \langle \mathbf{u}, A\mathbf{v}_1 \rangle = \lambda_1 \langle \mathbf{u}, \mathbf{v}_1 \rangle = 0$$

- c) Aplique a hipótese indutiva a $A|_{\mathcal{W}}$, obtendo base $\{\mathbf{v}_2, \dots, \mathbf{v}_{n+1}\}$ ortonormal

d) A união $\{\mathbf{v}_1, \dots, \mathbf{v}_{n+1}\}$ diagonaliza A

Este teorema explica por que matrizes de covariância (auto-adjuntas) sempre admitem decomposição em autovetores ortogonais - a fundação matemática para extração de componentes principais.

Decomposição Espectral

O Teorema Espectral materializa-se na **decomposição espectral**: para toda matriz simétrica $\mathbf{T} \in \mathbb{R}^{p \times p}$, existe

$$\mathbf{T} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T,$$

onde a estrutura revela a geometria subjacente dos dados:

- Matriz ortogonal \mathbf{V} : Colunas $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ formam uma base ortonormal de autovetores – os *eixos privilegiados* dos dados.
 - Ortogonalidade: $\mathbf{V}^T\mathbf{V} = \mathbf{I}_p$ (preserva ângulos e normas)
 - Interpretação: Cada \mathbf{v}_i define uma direção de máxima variância residual
- Matriz diagonal $\mathbf{\Lambda}$: $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$, com $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ (autovalores reais).
 - λ_i quantifica a variância ao longo de \mathbf{v}_i
 - Ordenação: Garante hierarquia de importância para redução dimensional

A identidade $\mathbf{T} = \sum_{i=1}^p \lambda_i \mathbf{v}_i \mathbf{v}_i^T$ mostra que \mathbf{T} é uma combinação de *projetores ortogonais* – cada termo $\lambda_i \mathbf{v}_i \mathbf{v}_i^T$ representa a contribuição da direção \mathbf{v}_i à estrutura global.

Essa decomposição é a etapa central – ao descartar os λ_i menores, obtemos a melhor aproximação de baixa dimensão preservando a covariância original. A ordenação dos autovalores define a importância dos componentes principais.

Matriz de Covariância

Centralização dos Dados

Toda análise multivariada começa por centralizar os dados - processo que elimina deslocamentos arbitrários e revela padrões de variação genuínos. Dada a matriz de dados $\mathbf{X} \in \mathbb{R}^{n \times p}$

(n observações, p variáveis):

$$\tilde{\mathbf{X}} = \mathbf{X} - \mathbf{1}_n \bar{\mathbf{x}}^T, \quad \text{onde } \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$$

- Operação: Subtrair o vetor de médias $\bar{\mathbf{x}}$ de cada observação
- Efeito Geométrico: Translada a nuvem de dados para a origem sem alterar sua forma

A Matriz de Covariância

A matriz $\mathbf{S} \in \mathbb{R}^{p \times p}$ sintetiza como as variáveis co-variam:

$$\mathbf{S} = \frac{1}{n-1} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}} \quad (\text{estimador não-enviesado})$$

- Entrada S_{ij} :

$$S_{ij} = \frac{1}{n-1} \sum_{k=1}^n (\tilde{x}_{ki})(\tilde{x}_{kj})$$

- Mede sincronia entre variáveis i e j - Positiva se altos valores de i associam-se a altos de j

- Traço Total:

$$\text{Tr}(\mathbf{S}) = \sum_{i=1}^p S_{ii} = \text{Variância total dos dados}$$

- Invariante sob rotações: crucial para PCA

\mathbf{S} é simétrica e semi-definida positiva, garantindo:

- Autovalores reais não-negativos
- Autovetores ortogonais

Diagonalizar \mathbf{S} via decomposição espectral é equivalente a encontrar os eixos de máxima variância - os componentes principais

Maximização da Variância

Formulação do Problema

A essência da PCA reside em encontrar direções de máxima variância nos dados. Formalmente, buscamos o vetor unitário $\mathbf{w}_1 \in \mathbb{R}^p$ que maximize a variância das projeções $\mathbf{Z}_1 = \tilde{\mathbf{X}} \mathbf{w}_1$:

$$\max_{\|\mathbf{w}_1\|=1} \mathbf{w}_1^T \mathbf{S} \mathbf{w}_1$$

- Geometricamente: Procura-se a reta pela origem onde os dados projetados têm maior "dispersão"
- Estatisticamente: maximizar $\mathbf{w}_1^T \mathbf{S} \mathbf{w}_1$ equivale a capturar a máxima informação (variância) possível em uma dimensão

Solução Ótima via Autovalores

A maximização restrita resolve-se com multiplicadores de Lagrange:

$$\mathcal{L}(\mathbf{w}_1, \lambda) = \underbrace{\mathbf{w}_1^T \mathbf{S} \mathbf{w}_1}_{\text{Variância}} - \lambda \underbrace{(\mathbf{w}_1^T \mathbf{w}_1 - 1)}_{\text{Restrição}}$$

Derivando e igualando a zero:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_1} = 2\mathbf{S} \mathbf{w}_1 - 2\lambda \mathbf{w}_1 = 0 \implies \boxed{\mathbf{S} \mathbf{w}_1 = \lambda \mathbf{w}_1}$$

A solução ótima \mathbf{w}_1 é o autovetor de \mathbf{S} associado ao maior **autovalor** λ_1 . A variância máxima é exatamente λ_1 .

- Interpretação física: Autovetores são eixos naturais de variação dos dados

Componentes Principais Sequenciais

Ortogonalidade e Maximização Iterativa

Após determinar \mathbf{w}_1 , os componentes subsequentes $\mathbf{w}_2, \mathbf{w}_3, \dots$ resolvem:

$$\max_{\|\mathbf{w}_k\|=1} \mathbf{w}_k^T \mathbf{S} \mathbf{w}_k \quad \text{com } \mathbf{w}_k \perp \{\mathbf{w}_1, \dots, \mathbf{w}_{k-1}\}$$

- Restrição: Ortogonalidade $\mathbf{w}_k^T \mathbf{w}_i = 0$ ($\forall i < k$) garante componentes não correlacionados
- Mecanismo: Cada novo componente captura a máxima variância residual não explicada pelos anteriores

Solução Recursiva via Decomposição Espectral

A solução emerge diretamente do Teorema Espectral:

1. Ordene os autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$
2. Selecione os autovetores correspondentes $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_p$
3. O k -ésimo componente principal é $\mathbf{Z}_k = \tilde{\mathbf{X}}\mathbf{w}_k$

A sequência $\{\mathbf{w}_k\}$ forma uma base ortonormal que diagonaliza \mathbf{S} :

$$\mathbf{S} = \sum_{i=1}^p \lambda_i \mathbf{w}_i \mathbf{w}_i^T$$

A escolha de k componentes ($k \ll p$) reduz dimensionalidade preservando $\sum_{i=1}^k \lambda_i / \text{Tr}(\mathbf{S})$ da variância total. A ortogonalidade dos \mathbf{w}_i garante eficiência computacional na projeção.

Projeção e Redução de Dimensionalidade

Transformação para o Espaço dos Componentes

PCA concretiza-se na projeção dos dados para o espaço reduzido. Seja $\mathbf{W} = [\mathbf{w}_1 | \dots | \mathbf{w}_k]$ a matriz de projeção com os k primeiros autovetores, a transformação é:

$$\mathbf{Z} = \tilde{\mathbf{X}}\mathbf{W} \in \mathbb{R}^{n \times k}$$

- Geometricamente: Rotaciona os dados para os eixos de máxima variância
- Estatisticamente: Colunas de \mathbf{Z} são não correlacionadas, com $\text{Var}(Z_i) = \lambda_i$

Variância Explicada

A eficácia da redução quantifica-se pela variância retida:

$$\text{VE}(k) = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \times 100\%$$

- Interpretação: A porcentagem de informação original preservada no espaço reduzido após a projeção dos dados.

- Compensação: Um trade-off entre a simplicidade do modelo ($k \downarrow$) e a quantidade de informação explicada ($VE \uparrow$).

Cr terios de Sele  o de Componentes

Tr s Perspectivas Cl ssicas

1. Crit rio de Kaiser (1958):

- Regra: $\lambda_i \geq 1$ (para dados padronizados)
- L gica: Um componente deve explicar mais vari ncia do que uma vari vel isolada.
- Limita  o: Subestima o valor de k quando muitos λ_i est o pr ximos de 1.

2. Scree Plot (Cattell, 1966):

- T cnica: Gr fico dos autovalores versus o n mero de componentes.
- Heur stica: Selecionar k no ponto onde a curva de autovalores se estabiliza.

3. Vari ncia Acumulada:

- Estrat gia: Escolher o k m nimo tal que $VE(k) \geq 95\%$.
- Vantagem: Adapt vel ao contexto (85% de vari ncia para an lise explorat ria, 99% para modelos cr ticos).
- Considera  o adicional: Autovalores pequenos podem codificar padr es sutis, mas importantes, que podem ser negligenciados em uma an lise mais superficial.

A determina  o  tima de k envolve um trade-off entre crit rios quantitativos (e.g., vari ncia explicada, autovalores residuais) e exig ncias do problema aplicado (e.g., custos computacionais, interpretabilidade), demandando valida  o cruzada entre heur sticas estat sticas e restri  es do dom nio.

APÊNDICE B – COEFICIENTES DOS POLINÔMIOS QUE GERAM A SUPERFÍCIE DE RESPOSTA E SUA DERIVAÇÃO

Os coeficientes apresentados nas tabelas 25 e 26 são utilizados na construção das superfícies de resposta para os cenários de circulação direta e reversa mostrados nas figuras 17 e 24. Essas superfícies são geradas a partir de modelos polinomiais que relacionam a vazão de entrada (x) e a temperatura (y) com o volume resultante (z) ao final de um período de 120 dias.

Tabela 25 – Coeficientes para obter a superfície de resposta do volume da caverna ao final de 120 dias para o cenário com circulação direta.

Nº Equação	Coef. da vazão a_i	Coef. da temperatura b_i	término Independente
1	232.896	811.274	-25792.291
2	317.584	269.587	-14833.664
3	247.397	1192.921	-67593.888
4	261.677	269.587	-3652.100
5	232.896	481.406	-6000.161
6	248.253	219.301	-1280.894
7	289.886	158.795	-2646.540
8	289.886	385.834	-16268.885
9	219.138	385.834	-2119.313
10	288.154	481.406	-25893.072
11	261.677	726.138	-31045.184
12	190.328	726.138	-5359.601
13	198.747	1989.044	-92363.811
14	261.098	537.811	-20665.517
15	166.034	811.274	952.284
16	214.642	537.811	-2082.823
17	158.941	2633.679	-83274.527
18	214.642	1783.421	-76819.414
19	116.428	1783.421	1751.335
20	158.941	1192.921	3170.942

Fonte: O autor (2025)

Tabela 26 – Coeficientes para obter a superfície de resposta do volume da caverna ao final de 120 dias para o cenário reverso.

Nº Equação	Coef. da vazão a_i	Coef. da temperatura b_i	término independente
1	263.279	841.217	-25917.156
2	356.817	261.665	-14859.358
3	269.238	1149.323	-59250.373
4	294.579	261.665	-2411.665
5	263.279	469.951	-3641.187
6	285.215	212.906	-1157.564
7	329.517	152.466	-2847.448
8	329.517	390.116	-17106.462
9	250.111	390.116	-1225.159
10	320.614	469.951	-24281.724
11	294.579	745.859	-31463.295
12	215.600	745.859	-3031.093
13	229.191	1975.532	-93309.209
14	291.380	526.153	-18253.580
15	190.318	841.217	3267.547
16	238.080	526.153	3066.686
17	187.880	2809.501	-93774.784
18	238.080	1796.454	-73151.408
19	137.228	1796.454	7529.937
20	187.880	1149.323	5835.907

Fonte: O autor (2025)

Após a apresentação dos coeficientes utilizados na construção das superfícies de resposta para os cenários de circulação direta e reversa, passamos a analisar o processo de triangulação e interpolação que serve de base para esses modelos.

Dado um conjunto de pontos dispersos (x_i, y_i) localizados no plano xy , onde seus valores podem ser interpretados como altitudes z_i sobre esse plano. A partir desses pontos, é realizada uma triangulação que resulta em uma superfície triangular por partes sobre o plano. Esta é uma superfície contínua composta de peças triangulares planas que são unidas ao longo das arestas. Tal superfície é frequentemente chamada de rede triangular irregular.

Na figura 51, é apresentado o triângulo $P_1P_2P_3$ com o ponto de interpolação P localizado dentro dele. Ao conectarmos o ponto P a cada um dos vértices P_1, P_2 e P_3 , obtemos uma subdivisão do triângulo em três subtriângulos: PP_2P_3, P_1PP_3 e P_1P_2P . Esta figura ilustra visualmente a relação entre os vértices do triângulo e o ponto de interpolação.

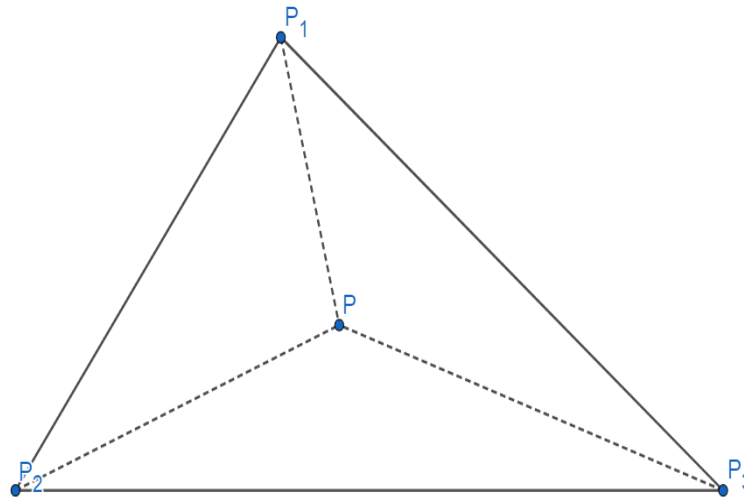
Agora, definiremos formalmente cada triângulo da rede triangular irregular, denotado por \triangle_i .

Para cada triângulo da rede triangular irregular, denotado por \triangle_i , definimos a área do triângulo com vértices A , B e C como $S(A, B, C)$. Então, o i -ésimo triângulo \triangle_i pode ser definido como o conjunto de pontos P no plano xy tal que:

$$\triangle_i = \{P \in \mathbb{R}^2 : S(P_1^i, P_2^i, P_3^i) = S(P, P_1^i, P_2^i) + S(P, P_1^i, P_3^i) + S(P, P_2^i, P_3^i)\}$$

Onde P_1^i , P_2^i e P_3^i são os vértices do triângulo \triangle_i .

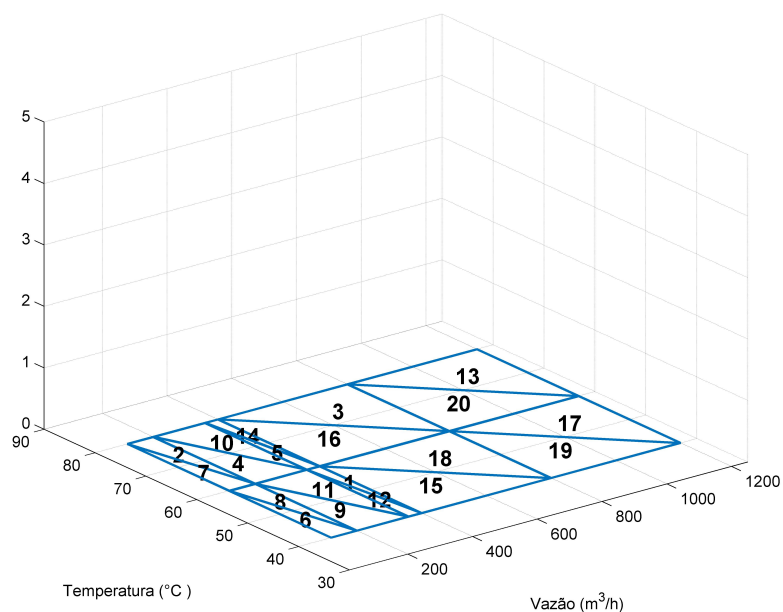
Figura 51 – Subdivisão do triângulo



Fonte: O autor (2025)

Para realizar essa triangulação, consideramos os triângulos pertencentes à rede triangular irregular, como mostrado na figura 52. Observa-se que cada triângulo é numerado com uma etiqueta correspondente aos polinômios descritos nas tabelas 25 e 26, os quais variam de acordo com o cenário (direto ou reverso).

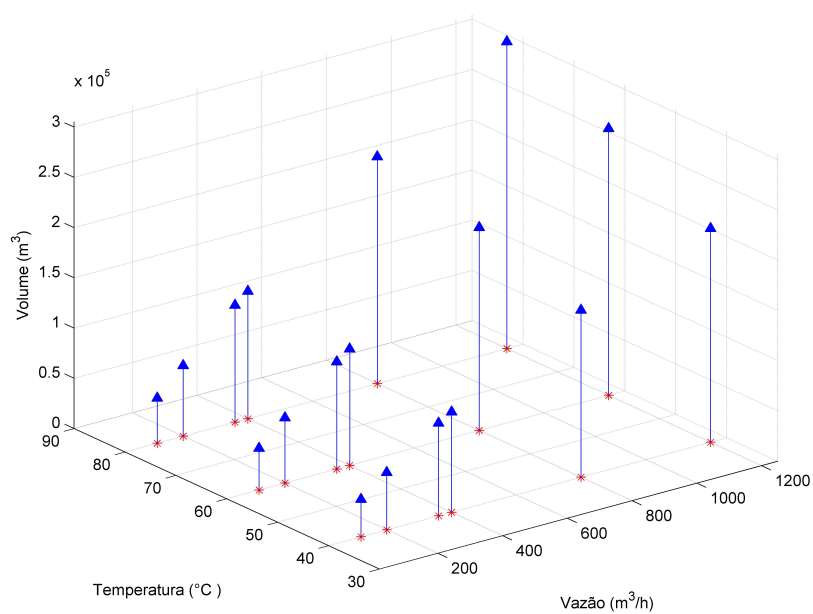
Figura 52 – Rede triangular irregular de vazões e temperaturas



Fonte: O autor (2025)

Com vértices $P_1^i = (x_1, y_1)$, $P_2^i = (x_2, y_2)$ e $P_3^i = (x_3, y_3)$, onde cada vértice é associado a uma coordenada z , representada por z_1 , z_2 e z_3 , respectivamente, como exemplificado na Figura 53.

Figura 53 – Pontos de dados de vazão, temperatura e volume



Fonte: O autor (2025)

Estes pontos definem um plano no espaço tridimensional, determinado pelas coordenadas (x_1, y_1, z_1) , (x_2, y_2, z_2) e (x_3, y_3, z_3) . Esse plano pode ser expresso pela equação:

$$\pi_i : z = a_i x + b_i y + c_i \quad (\text{B.1})$$

Como os vértices $v_1^i = (x_1, y_1, z_1)$, $v_2^i = (x_2, y_2, z_2)$ e $v_3^i = (x_3, y_3, z_3)$ pertencem ao plano π_i , eles satisfazem a equação (B.1):

$$z_1 = a_i x_1 + b_i y_1 + c_i \quad (\text{B.2})$$

$$z_2 = a_i x_2 + b_i y_2 + c_i \quad (\text{B.3})$$

$$z_3 = a_i x_3 + b_i y_3 + c_i \quad (\text{B.4})$$

Em termos matriciais, podemos escrever:

$$\begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix} \begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \quad (\text{B.5})$$

Resolvendo essa equação, encontramos os coeficientes a_i , b_i e c_i :

$$\begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = \begin{pmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} \quad (\text{B.6})$$

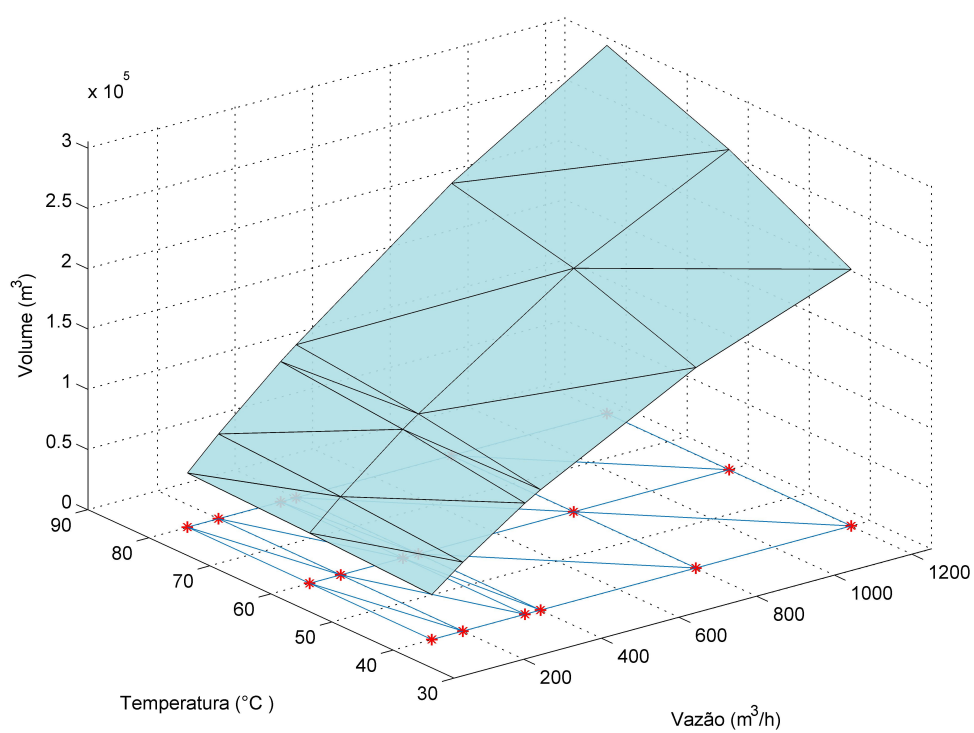
Assim, determinamos um triângulo de vértices $v_1^i = (x_1, y_1, z_1)$, $v_2^i = (x_2, y_2, z_2)$, $v_3^i = (x_3, y_3, z_3)$ incluído no plano π_i .

A união desses triângulos forma a superfície contínua F , como mostrado na figura 54, que representa a interpolação da superfície sobre o plano xy a partir dos pontos dispersos iniciais. Ao considerar um ponto $P = (x, y)$ na superfície formada pelas peças triangulares planas, podemos determinar em qual triângulo Δ_i ele está contido. Isso nos permite utilizar a equação do plano π_i associada a esse triângulo para calcular o valor de z para o ponto P . Em outras palavras, $F(P)$ é dado por:

$$F(P) = a_i x + b_i y + c_i$$

para este triângulo. Essa abordagem fornece uma maneira eficaz de interpolar os valores de z para pontos dentro de cada triângulo da rede triangular irregular, resultando na superfície contínua F .

Figura 54 – Superfície de resposta triangulada para o volume da caverna em função da temperatura e da vazão.



Fonte: O autor (2025)

APÊNDICE C – CÁLCULO DE PRESSÕES COM BASE NA EQUAÇÃO DE BERNOULLI

Neste apêndice, são apresentadas as equações do cálculo da pressão total em um sistema de injeção vertical, utilizando a equação de Bernoulli. Na sequência detalhamos os passos do processo.

DEFINIÇÕES DE CONSTANTES

Primeiro, precisamos definir algumas constantes que serão usadas nos cálculos:

- Diâmetro da válvula de injeção ($d_{\text{válvula}}$): 0.0508 metros.
- Diâmetro do tubo interno de injeção ($d_{\text{tubo_interno}}$): 0.1778 metros.
- Peso específico do fluido injetado (γ_{fluido}): 10300.0 N/m³.

DEFINIÇÃO DAS CONSTANTES DE CONVERSÃO

Em seguida, vamos definir algumas constantes de conversão que serão úteis para transformação de unidades:

- Conversão de pés para metros: 1 pé = 0.3048 metros.
- Conversão de MPa para kPa: 1 MPa = 1000 kPa.
- Conversão de MPa para metros de coluna d'água (mca): 1 MPa = 101.97 mca.
- Conversão de kPa para PSI: 1 kPa = 0.145038 PSI.

DADOS FORNECIDOS

Para os cálculos realizados neste apêndice, foram considerados os seguintes parâmetros para os cenários de mineração por dissolução:

- Profundidade de injeção (h_{injecao}):
 - Cenário Direto: 3000 pés (914,4 metros).

- Cenário Reverso: 2500 pés (762 metros).
- Vazão de injeção ($Q_{m3_por_h}$): Foram avaliados seis níveis de vazão: 120, 200, 360, 400, 800 e 1200 m³/h.

CONVERSÃO DE PROFUNDIDADE

Primeiro, para converter a profundidade de injeção de pés para metros:

$$h_{injecao_m} = h_{injecao} \times 0.3048$$

APÊNDICE D – CÁLCULO DA PRESSÃO DE INJEÇÃO NO FUNDO DO POÇO

Neste apêndice, detalha-se a metodologia de cálculo para determinar a pressão de injeção no fundo do poço (P_{inj}). Este valor, crucial para as simulações, é composto pela soma de duas componentes principais: a pressão hidrostática, devido à coluna de fluido, e a pressão dinâmica, associada à velocidade do escoamento, conforme a Equação de Bernoulli.

A equação fundamental que guia este cálculo é:

$$P_{inj} = P_{hidrostática} + P_{dinâmica}$$

PARÂMETROS E CONSTANTES DO MODELO

Os cálculos baseiam-se nos seguintes parâmetros e constantes:

- Constantes geométricas:
 - Diâmetro do canhoneio ($d_{canhoneio}$): 0.0508 metros.
 - Diâmetro do tubo interno de injeção ($d_{tubo_interno}$): 0.1778 metros.
 - Número de canhoneios ($n_{canhoneios}$): 4.
- Propriedades do fluido:
 - Peso específico do fluido injetado (γ_{fluido}): 10300.0 N/m³.
- Parâmetros dos cenários:
 - Profundidade de injeção ($h_{injecao}$): 3000 pés (914,4 m) para o cenário direto e 2500 pés (762 m) para o cenário reverso.
 - Vazão de injeção ($Q_{m3_por_h}$): Foram avaliados seis níveis de vazão: 120, 200, 360, 400, 800 e 1200 m³/h.

METODOLOGIA DE CÁLCULO

A pressão de injeção total é obtida seguindo uma sequência de cálculos para determinar cada componente.

Cálculo da Componente Hidrostática

A pressão hidrostática ($P_{\text{hidrostática}}$) é a pressão exercida pelo peso da coluna de fluido. Ela é calculada a partir da profundidade de injeção em metros ($h_{\text{injecao_m}}$) e convertida para MPa:

$$P_{\text{hidrostática}} \text{ (MPa)} = \frac{h_{\text{injecao_m}}}{101.97}$$

onde a constante 101.97 converte metros de coluna de água (mca) para MPa.

Cálculo da Componente Dinâmica

A componente dinâmica da pressão ($P_{\text{dinâmica}}$) surge da variação de velocidade do fluido. O cálculo é feito em etapas:

1. Cálculo das áreas de escoamento: As áreas de saída total pelos canhoneios e a área do tubo são calculadas:

$$A_{\text{canhoneio}} = n_{\text{canhoneios}} \times \pi \left(\frac{d_{\text{canhoneio}}}{2} \right)^2 \quad \text{e} \quad A_{\text{tubo}} = \pi \left(\frac{d_{\text{tubo_interno}}}{2} \right)^2$$

2. Conversão da vazão: A vazão é convertida de m^3/h para m^3/s :

$$Q_{\text{m3_por_s}} = \frac{Q_{\text{m3_por_h}}}{3600}$$

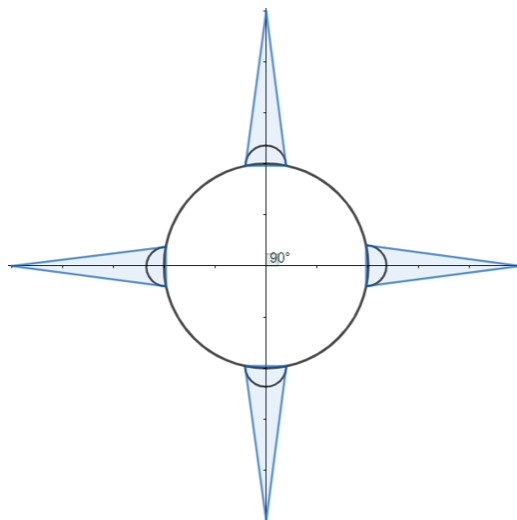
3. Cálculo da pressão dinâmica (Bernoulli): A pressão devido à diferença de velocidades é dada pela Equação de Bernoulli. Usando a relação $\rho = \gamma/g$, a equação fica:

$$P_{\text{dinâmica}} \text{ (kPa)} = \left(\frac{\gamma_{\text{fluido}}}{2g} \right) \left(\left(\frac{Q_{\text{m3_por_s}}}{A_{\text{canhoneio}}} \right)^2 - \left(\frac{Q_{\text{m3_por_s}}}{A_{\text{tubo}}} \right)^2 \right) \cdot 10^{-3}$$

onde os termos entre parênteses representam as velocidades quadradas na saída dos canhoneios e no interior do tubo, respectivamente.

A figura 55 ilustra a configuração dos canhoneios no tubo de injeção.

Figura 55 – Esquema de um tubo com quatro canhoneiros a 90 graus.



Fonte: O autor (2025)

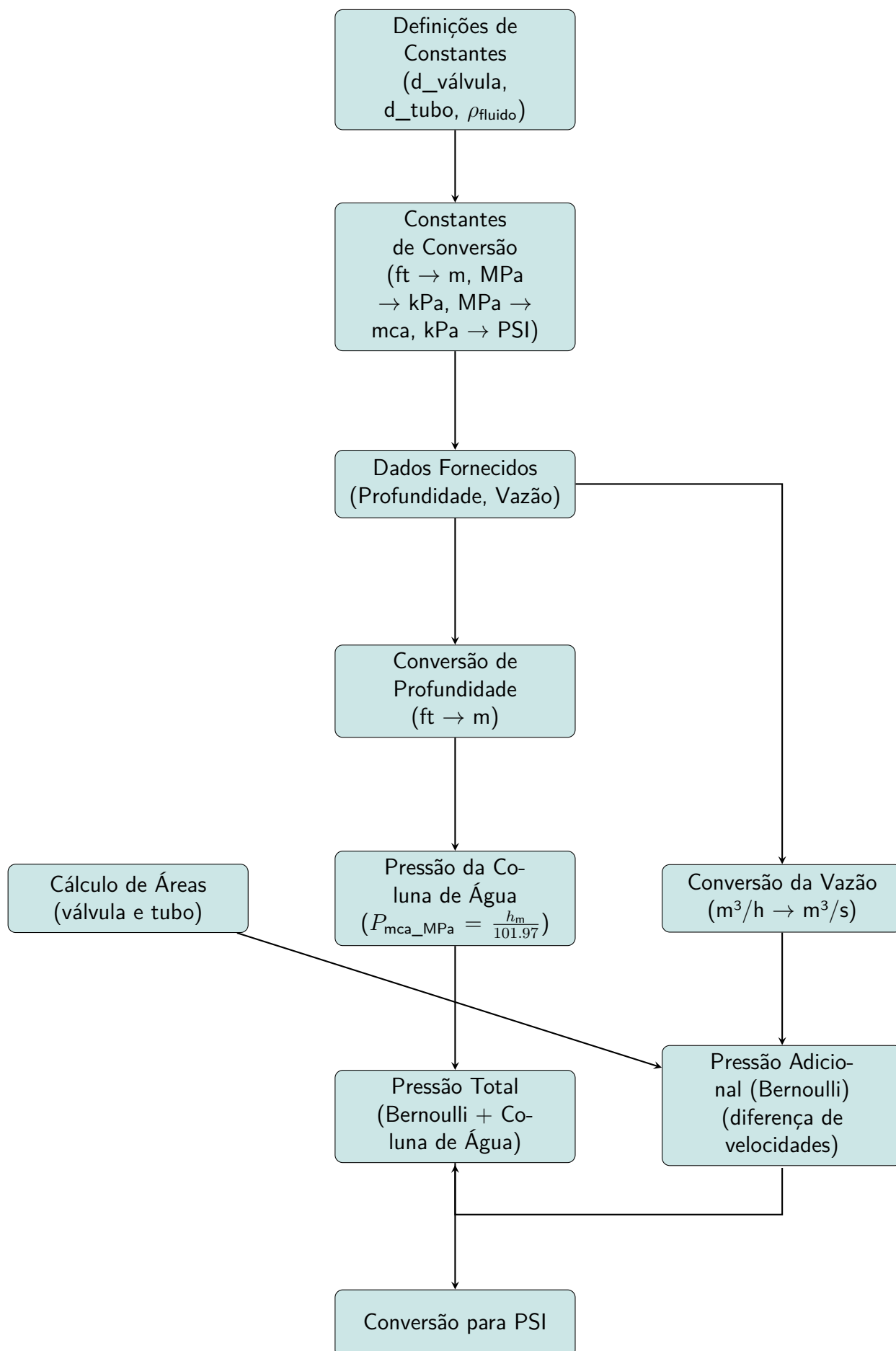
CÁLCULO DA PRESSÃO DE INJEÇÃO FINAL

Finalmente, a pressão de injeção total em MPa, que é o valor utilizado como entrada nas simulações do SALGAS, é a soma das duas componentes:

$$P_{\text{inj}} \text{ (MPa)} = P_{\text{hidrostática}} \text{ (MPa)} + \frac{P_{\text{dinâmica}} \text{ (kPa)}}{1000}$$

Este valor corresponde à "Pressão de injeção" reportada nas tabelas da tese para cada cenário de vazão.

Figura 56 – Fluxograma do processo de cálculo da pressão total em PSI.



Fonte: O autor (2025)

APÊNDICE E – GERAÇÃO DE DADOS DE ENTRADA E EXECUÇÃO DA SIMULAÇÃO

Neste apêndice, detalha-se o processo de cálculo dos parâmetros de entrada variáveis e o procedimento de execução das simulações no software SALGAS Eyerman (2008). Com base no valor da pressão total calculada no Apêndice C e na temperatura de referência de cada cenário (e.g., $40^{\circ}C$, $60^{\circ}C$ ou $80^{\circ}C$), as informações foram processadas para gerar os dados de entrada necessários.

CÁLCULO DOS PARÂMETROS DE ENTRADA VARIÁVEIS

Conforme explicado na dissertação de Silva (2022), os arquivos de entrada (.sgi) para todos os casos simulados possuem uma estrutura similar, com exceção de quatro parâmetros chave: taxa de produção de salmoura, gravidade específica do fluido de injeção, gravidade específica da salmoura e fator de dissolução do sal. O cálculo destes parâmetros seguiu uma metodologia sequencial:

1. Cálculo da pressão de injeção: A pressão foi calculada para cada vazão utilizando a equação de Bernoulli (Equação 4.8) e a equação hidrostática (Equação 4.9). Como a altura da válvula injetora é a mesma, a pressão de entrada no SALGAS varia de acordo com a vazão. O detalhamento completo deste cálculo encontra-se no Apêndice C.
2. Cálculo da gravidade específica (SG) do Fluido de Injeção: Com os parâmetros de temperatura e pressão, a gravidade específica do fluido de injeção foi calculada utilizando a caixa de ferramentas do SMRI Brouard (2008), na guia Fluidos > Salmoura > Guia Calculadora de salmoura subsaturada, conforme mostrado na figura 57.

Figura 57 – Guia calculadora de salmoura subsaturada no *Toolbox* do SMRI.

The image shows a software interface titled "Under-Saturated Brine Calculator". It features several input fields and calculated values. At the top, "Brine pressure (MPa)" is set to 0.1 and "Brine temperature (°C)" is set to 20. Below these, "Brine mass-per cent concentration" is shown with a value of 0.2025, which corresponds to 20.25 % salt. A horizontal slider is positioned below the percentage value. Further down, "Brine mass-volume concentration" is calculated as 233 (kg salt/m³ brine). Below that, "Brine density" is 1150.6 (kg/m³), and "Brine specific gravity" is 1.15063. Each calculated value is accompanied by a small icon of a document with a red 'X'.

Parameter	Value	Unit
Brine pressure	0.1	MPa
Brine temperature	20	°C
Brine mass-per cent concentration	0.2025	
% salt	20.25	
Brine mass-volume concentration	233	(kg salt/m ³ brine)
Brine density	1150.6	(kg/m ³)
Brine specific gravity	1.15063	

Fonte: Silva (2022)

3. Cálculo da gravidade específica (SG) da salmoura: De forma similar, a gravidade específica da salmoura foi calculada na guia Fluidos > Salmoura > Guia Calculadora de salmoura saturada, como ilustrado na Figura 58.

Figura 58 – Guia calculadora de salmoura saturada no *Toolbox* do SMRI.

The image shows a software interface titled "Saturated-Brine Calculator". It features several input fields and calculated values. At the top, "Brine pressure (MPa)" is set to 0.1 and "Brine temperature (°C)" is set to 20. Below these, "Saturated-brine mass-per cent concentration" is shown with a value of 0.2649, which corresponds to 26.49 % salt. A horizontal slider is positioned below the percentage value. Further down, "Saturated-brine mass-volume concentration" is calculated as 318.25 (kg salt/m³ brine). Below that, "Saturated-brine density" is 1201.4 (kg/m³), and "Saturated-brine specific gravity" is 1.2014. Each calculated value is accompanied by a small icon of a document with a red 'X'.

Parameter	Value	Unit
Brine pressure	0.1	MPa
Brine temperature	20	°C
Saturated-brine mass-per cent concentration	0.2649	
% salt	26.49	
Saturated-brine mass-volume concentration	318.25	(kg salt/m ³ brine)
Saturated-brine density	1201.4	(kg/m ³)
Saturated-brine specific gravity	1.2014	

Fonte: Silva (2022)

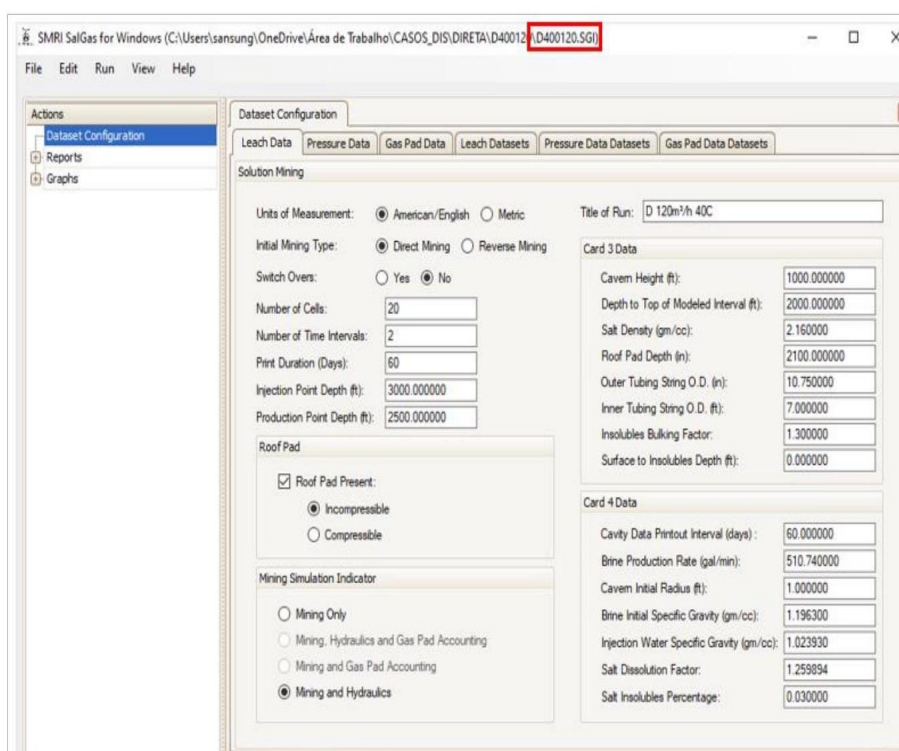
4. Cálculo do fator de dissolução (DF): Para cada taxa de produção, o Fator de Dissolução

foi calculado utilizando a temperatura do cenário e a gravidade específica da salmoura (obtida na etapa anterior) na Equação 4.7.

EXECUÇÃO DA SIMULAÇÃO NO SALGAS

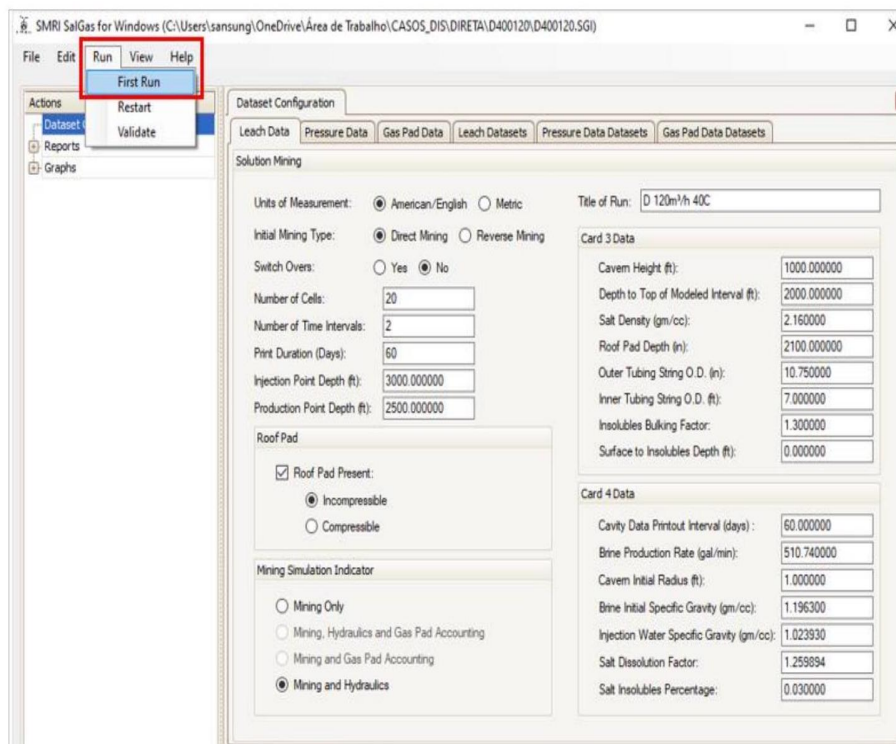
Após a geração do arquivo de entrada (.sgi) com os parâmetros calculados, cada caso foi executado no software SALGAS for Windows Eyerman (2008). A sequência de figuras 59, 60, 61 e 62 a seguir ilustra o processo padrão de execução, desde a abertura do arquivo até a verificação da conclusão bem-sucedida da simulação.

Figura 59 – Executando o caso no SALGAS: Abriu-se o arquivo de entrada.



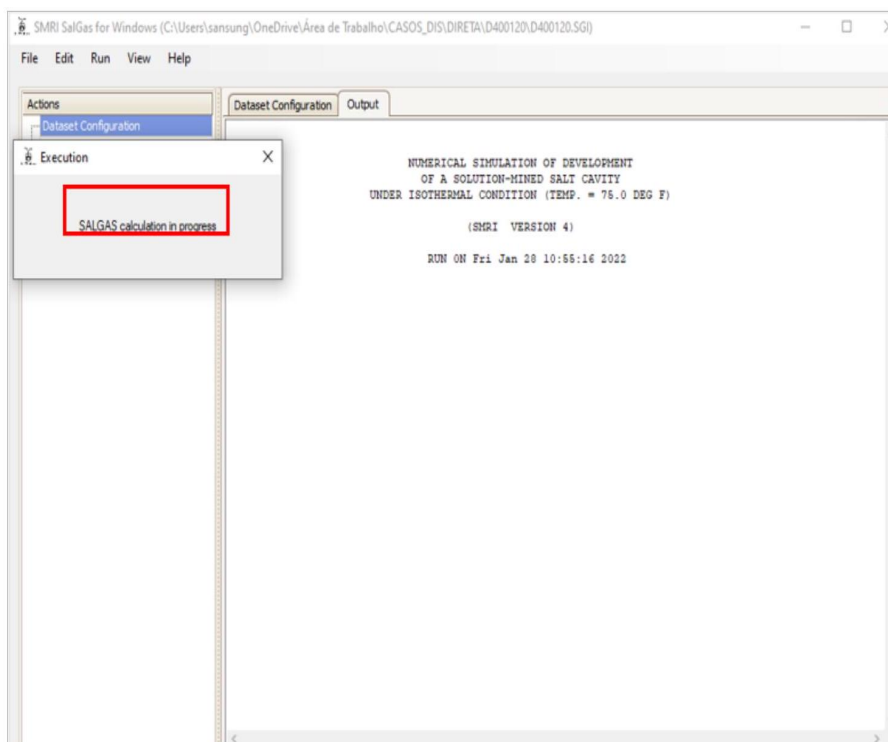
Fonte: Silva (2022)

Figura 60 – Executando o caso no SALGAS: Foi dado o comando de partida.



Fonte: Silva (2022)

Figura 61 – Executando o caso no SALGAS: Aguardou-se a conclusão da simulação.



Fonte: Silva (2022)

Figura 62 – Executando o caso no SALGAS: Verificou-se se rodou até o final sem erro.

SMRI SalGas for Windows (C:\Users\sansung\OneDrive\Área de Trabalho\CASOS_DIS\ DIRETA\ D400120\ D400120.SGI)

File Edit Run View Help

Actions: Dataset Configuration, Reports, Graphs

Dataset Configuration Output

PRODUCTION DATA

TIME IN DAYS	PRODUCED BRINE PCT.SAT	BRINE SP.GR.	DAILY SALT PRODUCTION US TONS	DAILY SALT DISSOLUTION US TONS	CAVITY VOLUME BBLs	SALT IN SOL IN CAVITY US TONS	TOTAL SALT DISSOLVED US TONS	TOTAL SALT PRODUCED US TONS	TOT WATER BBL
60	77.91	1.1599	578.4	655.1	103466.5	4628.2	39304.0	34705.9	10767
120	82.26	1.1683	673.9	774.4	225229.2	10659.1	85770.1	75141.0	21584

TIME IN DAYS	PUMP PRESSURE PSI	TUBING PRESSURE PSI	CASING PRESSURE PSI	PUMP POWER HP	CUMULATIVE ENERGY KW-HR	DAILY SALT PROD/ENERGY US TON/KW-HR	CUM. SALT PROD/ENERGY US TON/KW-HR	DAILY VOLUME/ENE BBLs/KW-HR
60	248.3	248.1	15.19	84.6	50825.	0.38212	0.38212	1.142
120	259.3	259.1	15.19	88.3	185636.	0.42647	0.40478	1.294

TIME IN DAYS	TUBING LOSS PSI	CASING LOSS PSI	PRODUCT LINE LOSS PSI	WATER LINE LOSS PSI	TUBING HEAD PSI	CASING HEAD PSI	HEAD BTWN PIPES PSI	CAVITY HEAD PSI
60	21.81	34.80	0.1905	0.1812	1330.	1256.	250.8	502.1
120	21.79	34.78	0.1916	0.1811	1330.	1265.	252.8	505.8

*** SALGAS DONE: NO WARNINGS OR ERRORS

Fonte: Silva (2022)

APÊNDICE E – DESCRIÇÃO DO ARQUIVO DE ENTRADA (.SGI) DO SALGAS

Este apêndice descreve um exemplo específico de configuração para uma simulação para um cenário de referência de 40°C e vazão de 360 m³/h, utilizando o método de circulação direta e o sistema de medidas americano. Cada linha e parâmetro do arquivo de entrada .sgi são detalhados abaixo, com a nomenclatura oficial da variável do software SALGAS Eyerman (2008)(conforme o manual do SMRI) indicada entre parênteses para referência.

DADOS DO ARQUIVO .SGI

```

0      D 360m³/h 40C
0 20 2 60 0 0 0 3 3000.000000 2500.000000
1000.000000 2000.000000 2.160000 10.750000 7.000000 2100.000000 1.300000 0.000000
60.000000 1532.220000 1.000000 1.196500 1.024230 1.260719 0.030000
1 2 2 1 15.000000 0.000130 0.900000
100 8
2800 6
200 6
2400 9.9 7.0
100 9.9 7.8
100 8
  
```

Descrição Detalhada do Arquivo .sgi

Linha 1: Sistema de medidas e título

- Unidade de medida utilizada (MSYS): Sistema americano (valor 0).
- Campo alfanumérico para o título da execução: D 360m³/h 40C.

Linha 2: Configurações Gerais da Simulação

- Direção inicial da mineração (ICODE): 0 (Circulação Direta).

- Número de células computacionais (N): 20.
- Número de intervalos de tempo (M): 2.
- Frequência de impressão dos dados de produção (IDT): A cada 60 dias.
- Presença de colchão de proteção (IPAD): 0 (Sim, há um *blanket* presente).
- Mudança de direção da mineração (ISOR): 0 (Não).
- Tipo de colchão de proteção (ICON): 0 (Líquido incompressível).
- Seletor de simulação (ISHORT): 3 (Módulos de mineração e hidráulica ativados).
- Profundidade de injeção (DEPINJ): 3000.0 ft (914,40 m).
- Profundidade de produção (DEPPRD): 2500.0 ft (762,00 m).

Linha 3: Propriedades da Rocha Salina e Geometria do Poço

- Altura da seção de sal modelada (H): 1000.0 ft (304,80 m).
- Profundidade do topo do intervalo modelado (DEPT): 2000.0 ft (609,60 m).
- Gravidade específica do sal (DENSAL): 2.16.
- Diâmetro externo da tubulação externa/rasa (DP): 10.75 in (273,05 mm).
- Diâmetro externo da tubulação interna/profunda (DPT): 7.0 in (177,80 mm).
- Profundidade do colchão de proteção (DEPPAD): 2100.0 ft (640,08 m).
- Fator de volume de formação de insolúveis (BULK): 1.3.
- Profundidade inicial do topo dos insolúveis (DPINS): 0.0 (Ausência de insolúveis no início).

Linha 4: Parâmetros de Operação e Propriedades dos Fluidos

- Intervalo para impressão dos dados da cavidade (CONTPD): 60.0 dias.
- Taxa de produção de salmoura (CONQUE): 1532.22 bbl/h.

- Raio inicial da caverna (CONRAD): 1.0 ft (0,3048 m).
- Gravidade específica inicial da salmoura (CONSG): 1.1965.
- Gravidade específica do fluido de injeção (CONISG): 1.024230.
- Fator de dissolução do sal (CONDIS): 1.260719. Este fator ajusta a taxa de dissolução para simular o efeito de uma temperatura (40°C) diferente da temperatura base do software (75°F / 24°C).
- Porcentagem de insolúveis no sal (VINSL): 0.03 (3.0%).

Linha 5: Configurações do Módulo Hidráulico

- Seções na linha de salmoura superficial (II): 1.
- Seções na tubulação interna da caverna (J1): 2.
- Seções na tubulação externa da caverna (K1): 2.
- Seções na linha de água superficial (L1): 1.
- Pressão de entrega da salmoura na superfície (PDEL): 15.0 psi (0,1034214 MPa).
- Rugosidade da superfície do tubo (ROUGH): 0.000130 in/in (0,000130 mm/mm).
- Eficiência da bomba d'água (EPUMP): 0.90 (90%).

Linhas 6 a 11: Comprimentos e Diâmetros das Seções dos Tubos

Estas linhas fornecem os comprimentos e diâmetros para cada uma das seções de tubulação definidas na Linha 5. O módulo hidráulico utiliza esses dados para calcular as perdas de carga por atrito em cada segmento do sistema e, conseqüentemente, determinar os requisitos de energia da bomba.

- Linha de salmoura superficial (Linha 6): Descreve a tubulação que transporta a salmoura da cabeça do poço até o ponto de entrega.
 - Comprimento: 100 ft (30,48 m)

-
- Diâmetro interno: 8 in (203,20 mm)
 - Coluna de tubulação interna (Linhas 7 e 8): Representa o tubo mais profundo, usado para injeção de água neste cenário de circulação direta. Está dividido em duas seções.
 - Seção 1: Comprimento: 2800 ft (853,44 m), Diâmetro interno: 6 in (152,40 mm).
 - Seção 2: Comprimento: 200 ft (60,96 m), Diâmetro interno: 6 in (152,40 mm).
 - Anular da coluna de tubulação externa (Linhas 9 e 10): Modela o espaço anular entre a tubulação interna e a externa, por onde a salmoura é produzida neste cenário.
 - Seção 1: Comprimento: 2400 ft (731,52 m), Diâmetro interno do anular: 9.9 in (251,46 mm), Diâmetro externo do tubo interno nesta seção: 7 in (177,80 mm).
 - Seção 2: Comprimento: 100 ft (30,48 m), Diâmetro interno do anular: 9.9 in (251,46 mm), Diâmetro externo do tubo interno nesta seção: 7.8 in (198,12 mm).
 - Linha de água superficial (Linha 11): Descreve a tubulação que transporta a água de injeção da bomba até a cabeça do poço.
 - Comprimento: 100 ft (30,48 m)
 - Diâmetro interno: 8 in (203,20 mm)