

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
ESCOLA DE QUÍMICA

Lucas Anelli de Paula



OTIMIZAÇÃO DE UMA PLANTA DE GÁS NATURAL USANDO
META-MODELOS

RIO DE JANEIRO

2024

Lucas Anelli de Paula

Otimização de uma planta de gás natural usando Meta-Modelos

Trabalho de Conclusão de Curso apresentado à Escola de Química da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do grau de Engenheiro Químico.

Orientadores: Roymel Rodriguez Carpio
Argimiro Resende Secchi

Rio de Janeiro

2024

CIP - Catalogação na Publicação

A578o Anelli de Paula, Lucas
Otimização de uma planta de gás natural usando
Meta-Modelos / Lucas Anelli de Paula. -- Rio de
Janeiro, 2024.
70 f.

Orientador: Argimiro Resende Secchi.
Coorientador: Roymel Rodriguez Carpio.
Trabalho de conclusão de curso (graduação) -
Universidade Federal do Rio de Janeiro, Escola de
Química, Bacharel em Engenharia Química, 2024.

1. Otimização. 2. Gás Natural. 3. Aspen Hysys. I.
Resende Secchi, Argimiro, orient. II. Rodriguez
Carpio, Roymel, coorient. III. Título.

Lucas Anelli de Paula

Otimização de uma planta de gás natural usando meta-modelos

Trabalho de Conclusão de Curso apresentado à
Escola de Química da Universidade Federal do
Rio de Janeiro, como parte dos requisitos
necessários à obtenção do grau Engenheiro
Químico.

Aprovado em 06 de fevereiro de 2024.

Argimiro Resende Secchi, D. Sc., Universidade Federal do Rio de Janeiro

Roymel Rodriguez Carpio, D. Sc., Universidade Federal do Rio de Janeiro

Heloísa Lajas Sanches Fernandes, D. Sc., Universidade Federal do Rio de Janeiro

Lizandro de Sousa Santos, D. Sc., Universidade Federal Fluminense

Rio de Janeiro
2024

AGRADECIMENTOS

Agradeço aos meus pais, Lilian e Luis Claudio, por todo o zelo durante minha criação e por acreditarem em mim desde minha infância. Agradeço também à minha companheira Carol, por me dar apoio durante minha jornada e me motivar a dar o meu melhor em tudo o que faço um dia após o outro.

Gostaria de agradecer aos professores Roymel e Argimiro, por me orientarem com excelência durante a realização deste trabalho, instigando meu pensamento crítico a cada encontro, me fazendo crescer intelectualmente e despertando meu interesse pelos desafios propostos.

Agradeço também a todos os integrantes do corpo discente da Escola de Química da UFRJ, por me proporcionarem um lugar onde eu pudesse me encontrar profissionalmente e ao mesmo tempo me sentir acolhido.

Por fim, gostaria de agradecer pelo meu falecido cachorro Link, que sempre me recebeu em casa com alegria e assumiu o papel de um irmão para mim nos momentos mais difíceis.

*“Eu tenho uma história repleta de falhas e fracassos em minha vida.
E é exatamente por isso que sou um sucesso”*

Michael Jordan

RESUMO

Anelli de Paula, Lucas. Otimização de uma planta de gás natural usando meta-modelos. Rio de Janeiro, 2024. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) - Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2024.

Sendo precursor de diversos insumos, como combustíveis de baixa emissão, gás de síntese, etano e propano, o gás natural é uma matéria-prima cujo consumo e produção global tem crescido nos últimos anos. Após extração, o gás é fracionado em uma Unidade de Processamento de Gás Natural (UPGN), produzindo correntes com maior valor de mercado, como GNV, GLP e C5⁺. O objetivo deste trabalho consistiu na otimização de uma Unidade de Fracionamento de Líquidos (UFL) contida dentro de uma UPGN, através do uso de meta-modelos. A partir de uma simulação da planta em software Aspen Hysys, foi proposta a construção de meta-modelos representativos em linguagem Python para previsão das variáveis dependentes do sistema e aplicação dos meta-modelos na otimização do lucro da UFL, considerando três cenários econômicos distintos. Os meta-modelos foram treinados e testados a partir de dados coletados da simulação, usando os métodos de Regressão Linear Múltipla, *Random Forest*, *Gradient Boosting*, *XGBoost* e Kriging para ajuste das variáveis. O meta-modelo que apresentou melhor desempenho fez uso do método Kriging, alcançando valor de R² acima de 99% na previsão da variável referente ao lucro obtido e das variáveis de restrições de qualidade dos produtos, seguido do meta-modelo ajustado por *Gradient Boosting*. Portanto, esses dois meta-modelos foram selecionados para uso na otimização. A otimização consistiu em maximizar a função objetivo que considera o lucro penalizado pelo desvio das restrições de qualidade dos produtos. Após verificar que o problema não possuía múltiplos ótimos locais, foram avaliadas três metodologias de otimização: Assistido por meta-modelo Kriging, assistido por meta-modelo *Gradient Boosting* e otimização direta na simulação no Aspen Hysys, para três cenários econômicos propostos. Em todos os cenários, foi observado que a otimização com meta-modelo Kriging levou a resultados de função objetivo melhores ou equiparáveis aos obtidos na otimização com simulação direta. Os resultados indicam ganhos de pelo menos 3% no lucro da UFL através das condições de operação encontradas na otimização assistida por meta-modelo Kriging.

Palavras-chave: Otimização. Gás natural. Aspen Hysys.

ABSTRACT

Anelli de Paula, Lucas. **Optimization of a natural gas plant using metamodels**. Rio de Janeiro, 2023. Trabalho de Conclusão de Curso (Graduação em Engenharia Química) - Escola de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2022.

As a precursor to various inputs, such as low-emission fuels, synthesis gas, ethane and propane, natural gas is a raw material whose global consumption and production has grown over the recent years. After its extraction, the gas is fractionated in a Natural Gas Processing Unit (NGPU), producing streams with greater market value, such as CNG, LPG and C_5^+ . The aim of this work was to optimize a Liquid Fractionation Unit (LFU) contained within a NGPU, using meta-models. Based on a simulation of the plant using Aspen Hysys software, a proposal was made to build representative meta-models in Python to predict the system's dependent variables and apply the meta-models to optimize the LFU's profit, considering three different economic scenarios. The meta-models were trained and tested using simulation data, using Multiple Linear Regression, Random Forest, Gradient Boosting, XGBoost and Kriging methods to adjust the variables. The meta-model that performed the best used the Kriging method, achieving an R^2 value of over 99% when predicting the variable related to plant profit and product quality restriction variables, followed by the meta-model adjusted by Gradient Boosting. These two meta-models were therefore selected for use in the optimization. The optimization consisted of maximizing the objective function that considers the profit penalized by the deviation from the product quality restrictions. After verifying that the problem did not have multiple local optima, three optimization methodologies were evaluated: assisted by Kriging meta-model, assisted by Gradient Boosting meta-model and direct optimization in the Aspen Hysys simulation, for three proposed economic scenarios. In all scenarios, it was observed that optimization with the Kriging meta-model led to better or comparable objective function results to those obtained with direct simulation optimization. The results indicate gains of at least 3% in LFU's profit through the operating conditions found in the Kriging meta-model-assisted optimization.

Keywords: Optimization. Natural Gas. Aspen Hysys.

LISTA DE ILUSTRAÇÕES

Figura 1 - Produção e consumo globais de gás natural. Fonte: ANP (2023)	18
Figura 2 - Preços internacionais do gás natural. Fonte: globalpetrolprices.com (2023)	19
Figura 3 - Dados de mercado de gás no Brasil. Fonte: ANP (2023).....	20
Figura 4 – Representação típica de um separador trifásico de óleo.....	21
Figura 5 - Diagrama de blocos de uma UPGN. Adaptado de SOUZA (2022)	22
Figura 6 - Diagrama de blocos que ilustra o treinamento e o teste de um modelo	25
Figura 7- Diagrama de blocos do funcionamento de um meta-modelo	27
Figura 8 - Exemplo de uma estrutura de árvore de decisão. Fonte: JAMES et al. (2013).....	29
Figura 9 - Exemplo de problema com múltiplos ótimos locais	36
Figura 10 - Exemplo de aplicação do método de Powell	38
Figura 11 - Simulação da UFL em Hysys	42
Figura 12 - Simulação da mistura das correntes de Gás de Venda	44
Figura 13 - Correlação da variável C2 em GLP com as variáveis RR, T_1 e T_2	51
Figura 14 - Correlação da variável C5 em GLP com as variáveis RR, T_1 e T_2	52
Figura 15 - Correlação da variável C2/C5 em GLP com as variáveis RR, T_1 e T_2	53
Figura 16 - Correlação da variável PVR em C5+ com as variáveis RR, T_1 e T_2	54
Figura 17 - Correlação da variável lucro no cenário 1 com as variáveis RR, T_1 e T_2	55
Figura 18 - Correlação da variável lucro no cenário 2 com as variáveis RR, T_1 e T_2	56
Figura 19 - Correlação da variável lucro no cenário 3 com as variáveis RR, T_1 e T_2	57
Figura 20 - Relevância da função objetivo no cenário 1.....	60
Figura 21 - Relevância da função objetivo no cenário 2.....	61
Figura 22 - Relevância da função objetivo no cenário 3.....	61

LISTA DE TABELAS

Tabela 1 - Condições das correntes de entrada da simulação da planta de gás natural	41
Tabela 2 - Cenários de mercado considerados para a otimização	45
Tabela 3 - Variáveis consideradas ao sistema	45
Tabela 4 - Intervalo usado na análise preliminar de variáveis	46
Tabela 5 - Normalização das variáveis do problema.....	47
Tabela 6 - Limites definidos da região de busca	58
Tabela 7 - Desempenho dos meta-modelos analisados	59
Tabela 8 - Resultados da otimização para o cenário econômico 1	62
Tabela 9 – Resultados de desempenho da otimização para o cenário econômico 1	63
Tabela 10 - Resultados da otimização para o cenário econômico 2	64
Tabela 11- Resultados de desempenho da otimização para o cenário econômico 2	64
Tabela 12 - Resultados da otimização para o cenário econômico 3	65
Tabela 13 - Resultados de desempenho da otimização para o cenário econômico 3	66
Tabela 14 - Condições operacionais de maximização do lucro encontradas na otimização	68

LISTA DE ABREVIATURAS E SIGLAS

GNV	Gás Natural Veicular
GLP	Gás Liquefeito do Petróleo
C5 ⁺	Gasolina natural
GEE	Gases de Efeito Estufa
ANP	Agência Nacional do Petróleo Gás Natural e Biocombustível
UPGN	Unidade de Processamento de Gás Natural
GV	Gás de Venda
LGN	Líquido de Gás Natural
UTG	Unidade de Tratamento de Gás
UFL	Unidade de Fracionamento de Líquidos
GR	Gás de Reciclo
HHV	Higher Heating Value
LHV	Lower Heating Value
GHV	Gross Heating Value
LHS	Latin Hypercube Sampling
RLM	Regressão Linear Múltipla
MSE	Mean Squared Error
SLSQP	Sequential Least Squares Programming
PSO	Particle Swarm Optimization
PVR	Pressão de Vapor Reid
T ₁	Temperatura de fundo da coluna desetanizadora
T ₂	Temperatura de fundo da coluna desbutanizadora
RR	Razão de refluxo da coluna desbutanizadora

LISTA DE SÍMBOLOS

$\Delta H_{H_2O,vap}$	Entalpia de evaporação da água (kJ kg^{-1})
$n_{H_2O,out}$	Número de moles de água vaporizada (mol)
$n_{fuel,in}$	Número de moles de combustível queimado (mol)
$MW_{H_2O,out}$	Peso molecular da água (g mol^{-1})
$MW_{fuel,in}$	Peso molecular do combustível (g mol^{-1})
V_{norm}	Valor normalizado da observação
V	Valor da observação
V_{LI}	Limite inferior definido para a observação
V_{LS}	Limite superior definido para a observação
\hat{y}	Valor estimado da variável dependente pelo modelo
$\hat{\beta}_0$	Coefficiente linear do modelo RLM
$\hat{\beta}_i$	Coefficiente angular do modelo RLM para a variável independente i
x_i	Valor da variável independente i
p	Número de variáveis independentes
h_m	Erro residual
y	Valor da variável dependente
$\hat{\mu}$	Termo do método Kriging responsável por capturar a tendência geral dos dados em uma regressão
z	Termo do método Kriging responsável pelo ajuste fino dos resultados
β	Matriz composta por parâmetros obtidos por regressão no modelo Kriging
F	Matriz transposta de um conjunto de funções escolhidas pelo modelo Kriging
r	Vetor de correlação entre um ponto e pontos de projeto do modelo Kriging
R	Matriz correlação dos pontos de projeto do modelo Kriging
Y	Matriz de valores da variável dependente
n	Número de valores no espaço amostral
\bar{y}	Média dos valores de variáveis dependentes
R^2	Métrica de R^2 do modelo testado
x_i^{k+1}	Posição da partícula i na iteração k+1 no método de PSO
x_i^k	Posição da partícula i na iteração k no método de PSO
v_i^{k+1}	Velocidade da partícula i na iteração k+1 no método PSO

w	Constante de contribuição inercial no método PSO
v_i^k	Velocidade da partícula i na iteração k no método PSO
φ_p	Constante de contribuição do fator cognitivo no método PSO
r_p	Número decimal randomicamente gerado para o fator cognitivo no método PSO
p_i	Melhor posição encontrada pela partícula i no método PSO
φ_g	Constante de contribuição do fator social no método PSO
r_g	Número decimal randomicamente gerado para o fator social no método PSO
g_i	Melhor posição encontrada pelo enxame no método PSO
F_i	Vazão do componente i (kg h ⁻¹)
LHV_i	Lower Heating Value da corrente i (kJ kg ⁻¹)
c_i	Custo de venda da corrente i (R\$ MMbtu ⁻¹)
G	Constante de conversão de unidade de energia (9,48e-007 MMbtu kJ ⁻¹)
c_e	Custo por energia (R\$ kWh ⁻¹)
E_i	Consumo energético do consumidor i (kW)
L	Lucro da planta (R\$ h ⁻¹)
R	Receita da planta (R\$ h ⁻¹)
$Opex$	Custo por operação da planta (R\$ h ⁻¹)
obj	Valor da função objetivo
P	Valor da penalidade aplicada ao lucro
$Desv_k$	Desvio da variável de especificação k com relação à sua especificação
$V_{k,spec}$	Valor de especificação da variável k
\hat{f}_{bag}	Estimativa final do modelo após <i>Bagging</i>
B	Número de árvores de decisões treinadas por banco de dados gerado por <i>Bootstrapping</i>
\hat{f}^{*b}	Resultado do b-ésimo modelo obtido por <i>Bootstrapping</i>

SUMÁRIO

1.	INTRODUÇÃO	15
1.1	CONTEXTUALIZAÇÃO	15
1.2	OBJETIVOS	17
1.3	ESTRUTURA DO TRABALHO	17
2.	REVISÃO BIBLIOGRÁFICA	18
2.1	ANÁLISE DE MERCADO DO GÁS NATURAL.....	18
2.2	PROCESSAMENTO DE GÁS NATURAL	20
2.3	MODELAGEM MATEMÁTICA	24
2.3.1	Preparo dos dados	25
2.3.2	Meta-Modelos	26
2.3.3	Regressão Linear Múltipla (RLM)	27
2.3.4	Árvore de decisão	28
2.3.5	Random Forest.....	29
2.3.6	Gradient Boosting.....	30
2.3.7	Kriging.....	32
2.3.8	Validação do modelo	33
2.4	OTIMIZAÇÃO.....	35
2.4.1	Análise prévia do problema	35
2.4.2	Nelder-Mead	36
2.4.3	Powell.....	37
2.4.4	SLSQP	39
2.4.5	PSO	39
3.	METODOLOGIA.....	41
3.1	SIMULAÇÃO DO PROCESSO	41
3.2	ANÁLISE DE VARIÁVEIS	46
3.3	ELABORAÇÃO DE META-MODELOS	46
3.4	PRÉ-AVALIAÇÃO DO PROBLEMA	48
3.5	OTIMIZAÇÃO DO PROCESSO.....	50
4.	RESULTADOS E DISCUSSÃO	51
4.1	ANÁLISE DE VARIÁVEIS	51
4.2	ELABORAÇÃO DE META-MODELOS	58
4.3	PRÉ-AVALIAÇÃO DO PROBLEMA	60

4.4 OTIMIZAÇÃO DA PLANTA	62
4.4.1 Resultados da otimização cenário 1	62
4.4.2 Resultados da otimização cenário 2	64
4.4.3 Resultados da otimização cenário 3	65
4.4.4 Resultados gerais da otimização.....	67
5. CONCLUSÃO.....	69
6. REFERÊNCIAS	71
7. APÊNDICES	74
7.1 APÊNDICE I – PARÂMETROS UTILIZADOS NO MODELO DE RLM.....	74
7.2 APÊNDICE II - PARÂMETROS UTILIZADOS NO MODELO DE RANDOM FOREST	75
7.3 APÊNDICE III – PARÂMETROS UTILIZADOS NO MODELO DE XGBOOSTING .	76
7.4 APÊNDICE IV – PARÂMETROS UTILIZADOS NO MODELO DE GRADIENT BOOSTING.....	77
7.5 APÊNDICE V – PARÂMETROS UTILIZADOS NO MODELO KRIGING (SCIKIT LEARN)	78
7.6 APÊNDICE VI – PARÂMETROS UTILIZADOS NO MODELO KRIGING (CARPIO, 2019).....	79

1. INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

O gás natural é um combustível fóssil na forma de gás rico em hidrocarbonetos leves, podendo ser encontrado em reservas naturais próprias ou em campos de petróleo e carvão. É considerado um combustível mais ambientalmente aceito quando comparado com outras fontes fósseis (carvão e óleo) (FARAMAWY et al., 2016). O gás é constituído em sua maior parte por metano, contendo também outros compostos orgânicos - como etano, propano e butano -, gases leves - como dióxido de carbono e nitrogênio - e contaminantes - como mercúrio e ácido sulfídrico.

O gás natural convencional é extraído de reservas comuns na sua forma associada ou não-associada. O gás natural associado é encontrado em reservas de petróleo, e em seu estado natural está dissolvido no óleo, enquanto o gás não-associado pode ser encontrado na sua forma livre em reservas próprias do gás ou em reservas de petróleo, acumulado na fase superior. No Brasil, do total de gás produzido em 2022, o gás associado representou 87,8% ou 44,2 bilhões de m³, cujo volume de produção em relação a 2021 subiu 6,1% (ANP, 2023). Após a separação do gás natural do petróleo, é necessário dar ao gás um destino dentre os três: uso comercial, reinjeção no poço ou queima na plataforma. Uma vez que a queima total em *flares* não é mais uma opção devido ao impacto ambiental, a parte do gás que não pode ser reinjetada no reservatório precisa ser enviada para unidades *onshore* para ser comercializado, caso contrário a produção de petróleo pode ser comprometida (SOUZA, 2022).

O gás natural é de grande relevância no mercado mundial e possui dois principais usos: como fonte de energia ou como insumo base para a indústria química. O gás natural pode ser tratado e fracionado para obter correntes com alto valor energético, como o Gás Natural Veicular (GNV), o Gás Liquefeito do Petróleo (GLP) e a gasolina natural (C5⁺), que possuem maior valor de mercado do que o gás extraído. Considerando uma menor emissão de poluentes que a gasolina do petróleo, em um cenário global de transição energética o uso do gás natural como combustível transicional pode viabilizar a migração gradual para fontes mais sustentáveis (GÜRSAN & DE GOOYERT, 2021). Além de seu potencial energético, o gás natural também é matéria-prima para importantes *building blocks* da indústria química,

como o gás de síntese, etano, propano e butano, componentes fundamentais na produção de outros insumos, como amônia, metanol e polímeros.

Considerando a importância do gás natural para a produção de uma variedade de insumos necessários à manutenção da sociedade, é de interesse estratégico a aplicação de técnicas computacionais para operar uma planta de processamento gás natural em suas condições ótimas, com o intuito de maximizar o lucro da planta e aprimorar sua capacidade de atender ao mercado de gás. No entanto, frente à multiplicidade de variáveis, a maximização do lucro de uma unidade de gás mostra-se um problema de natureza complexa, que requer o uso de métodos sofisticados para sua solução (SOUZA, 2022).

Softwares de simulação de processos são modelos robustos e baseados em fundamentos físico-químicos, que têm sido aplicados nos últimos anos para simular o comportamento de processos industriais. Tem se popularizado na indústria o conceito de *Digital Twin*, onde toda uma planta ou até uma unidade fabril tem seu processo digitalmente duplicado dentro de um *software* de simulação. Uma grande vantagem da aplicação da simulação decorre da capacidade de visualização de resultados próximos aos do processo real em diferentes condições operacionais, dispensando a necessidade de intervenção na planta. Simuladores de processo podem ser usados para solucionar o problema de otimização do lucro de uma planta, mas vale ressaltar que também podem levar a uma grande quantidade de horas despendidas para atingir o ótimo global (ALI, 2022). Isso se deve, principalmente, à complexidade envolvida no cálculo dos parâmetros do simulador a cada iteração.

Por meio de diferentes métodos de *Statistical Learning*, área que integra conceitos estatísticos a técnicas de *Machine Learning*, é possível construir modelos simplificados de um sistema, capazes de aprender com um banco de dados e então prever resultados para novos valores de entrada. A partir disso, introduz-se o conceito de meta-modelo, que se refere a um modelo cujo banco de dados utilizado para seu aprendizado é originado de outro modelo. Dessa maneira, através de dados coletados de um simulador de processo (modelo complexo), é possível treinar um meta-modelo simplificado capaz de prever o resultado do simulador. Caso a acurácia do modelo seja suficientemente elevada, pode-se afirmar que se trata de um meta-modelo representativo e então a resposta entregue pelo meta-modelo é próxima da resposta do simulador. Portanto, um meta-modelo representativo pode ser usado no lugar do simulador durante a resolução de uma otimização de um processo, com o intuito de alcançar o lucro ótimo da planta de maneira mais eficiente computacionalmente.

1.2 OBJETIVOS

O presente trabalho possui como objetivos gerais:

1. Desenvolver e avaliar meta-modelos que auxiliam na previsão de variáveis para otimização de uma planta de gás natural;
2. Realizar a otimização de uma planta de gás natural com o uso de meta-modelos.

1.3 ESTRUTURA DO TRABALHO

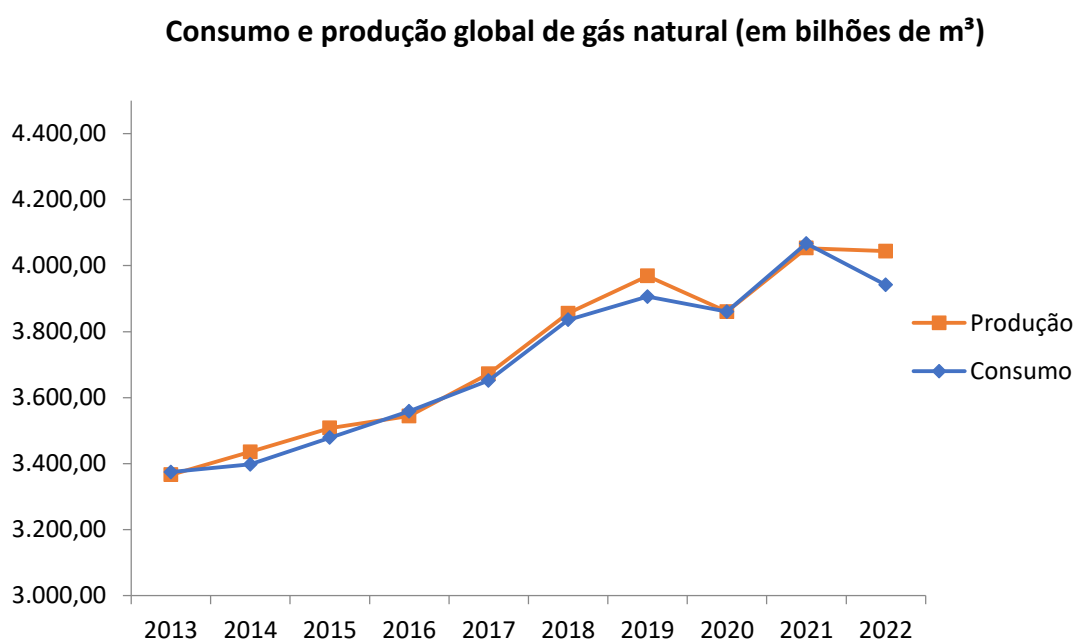
O trabalho é dividido em cinco capítulos. Inicia-se pela introdução no primeiro capítulo. Em seguida, o segundo capítulo apresenta uma revisão bibliográfica que aborda os principais conceitos utilizados no trabalho. No terceiro capítulo, a metodologia adotada para a realização do trabalho proposto é detalhada. No quarto capítulo, os resultados obtidos são apresentados e discutidos. O quinto capítulo consiste na conclusão do trabalho. No sexto capítulo do trabalho são deixadas as referências utilizadas e no sétimo os apêndices, onde são expostos os parâmetros adotados para cada método de modelagem usado.

2. REVISÃO BIBLIOGRÁFICA

2.1 ANÁLISE DE MERCADO DO GÁS NATURAL

Baseado na classificação proposta em 1970 por Charles H. Kline, o gás natural é um produto químico comercialmente classificado como *commodity*, baseado no seu elevado volume de produção mundial e baixa diferenciação. O gráfico da Figura 1 apresenta o consumo e a produção global de gás natural para um período entre 2013 e 2022.

Figura 1 - Produção e consumo globais de gás natural. Fonte: ANP (2023)

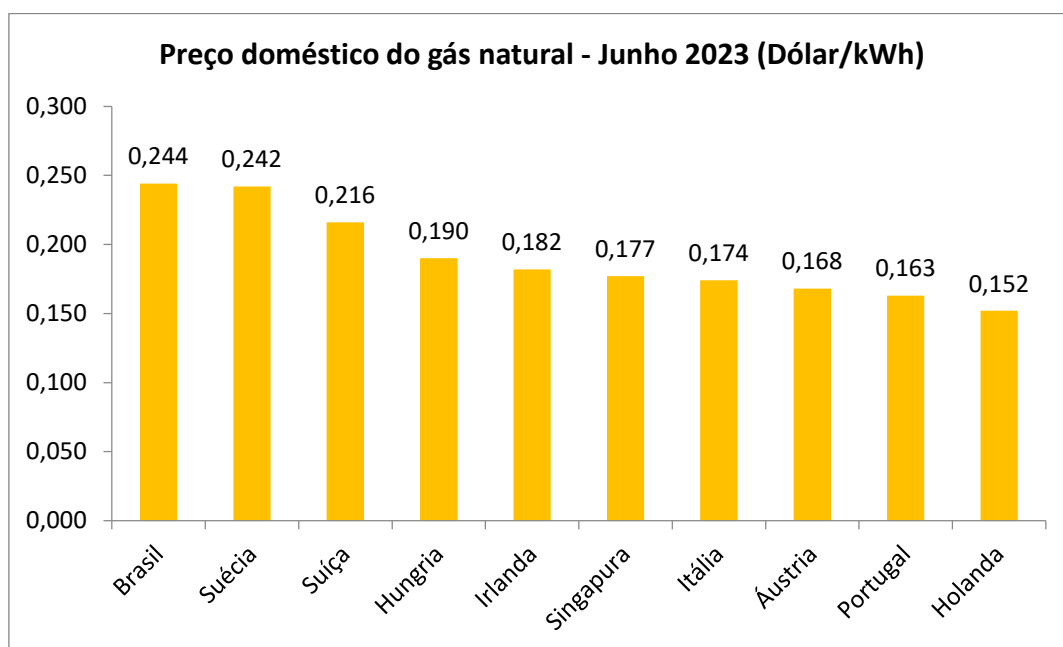


Nos últimos dez anos, o consumo mundial de gás natural aumentou em 17% e houve também um aumento de 20% na produção durante o mesmo período. Queda no consumo mundial de gás natural pode ser observada exclusivamente em dois momentos: No ano 2020, que pode ser justificado pela redução no consumo causada pela pandemia e no ano de 2022, quando ocorreu um aumento de preços do insumo causado pelo conflito envolvendo Rússia e Ucrânia. Tais eventos atípicos afetaram diretamente esse mercado, conforme discutido em CHEN et al. (2023). Num panorama geral, a Figura 1 indica que o mercado do gás natural global está em uma tendência de crescimento ao longo prazo.

A Figura 2 demonstra os preços domésticos para o gás natural praticados em junho de 2023 nos dez países onde o gás natural é mais caro. Percebe-se que o Brasil foi o país que apresentou o maior preço doméstico do gás natural nesse momento recente. Uma das causas

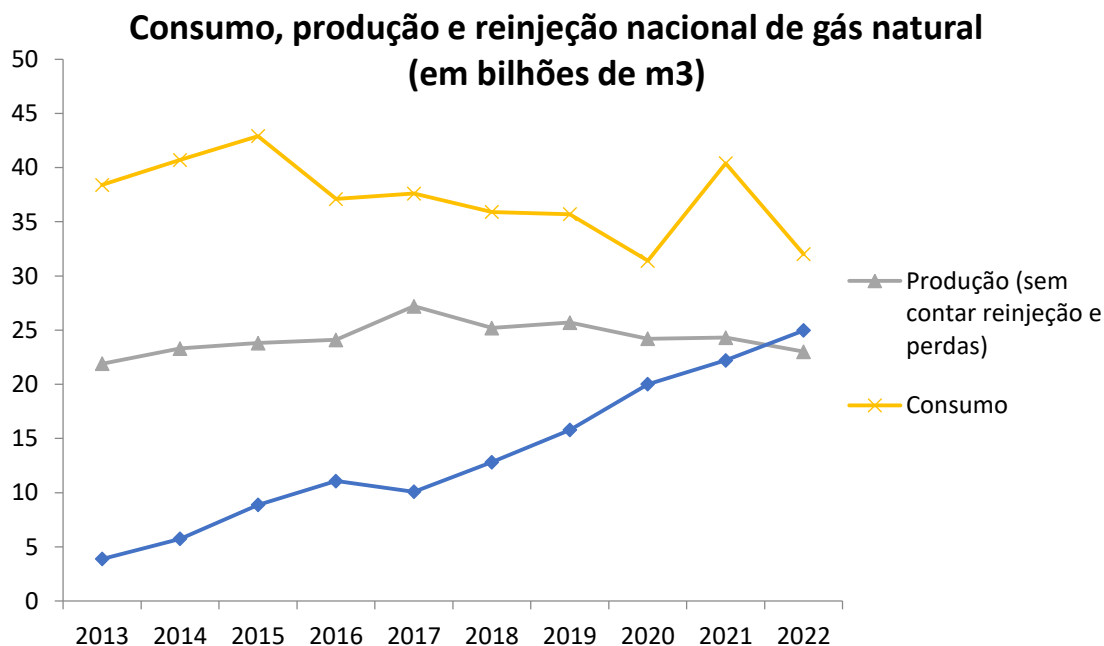
do custo elevado é a baixa oferta. Cerca de 30% do gás natural que o país consome é importado, pois grande parte do gás produzido é reinjetado nas plataformas (WOLKE, 2018). Conforme discutido em CNI (2019), o preço elevado do gás natural produzido no Brasil está relacionado ao custo associado na adequação desse gás às especificações nacionais de mercado. Uma vez que o gás natural produzido na maioria das reservas brasileiras apresenta teor de metano inferior à especificação de mercado (atualmente 88%, pela resolução ANP 16 de 2008), torna-se mais caro para refinar esse gás na planta industrial. Consequentemente, a demanda por combustíveis produzidos a partir do gás natural é inibida no território nacional.

Figura 2 - Preços internacionais do gás natural. Fonte: globalpetrolprices.com (2023)



A Figura 3 mostra os dados de consumo, produção e reinjeção de gás natural no Brasil para um período entre 2013 e 2022. Nesse período de dez anos, o consumo desse insumo caiu em 17% no Brasil, e a produção cresceu 5% no mesmo período. Uma queda de cerca de 20% no consumo nacional do gás natural de 2021 para 2022 pode ser justificada pela retomada das hidrelétricas, após um período de seca, o que levou a uma redução na queima do gás em termoeletricas (IEG, 2022). Além disso, mesmo com um consumo acima da quantidade de gás produzida, a quantidade de gás natural que é reinjetado nos poços aumentou em 12% de 2021 para 2022, chegando a representar mais de 50% de todo gás natural produzido em solo nacional.

Figura 3 - Dados de mercado de gás no Brasil. Fonte: ANP (2023)

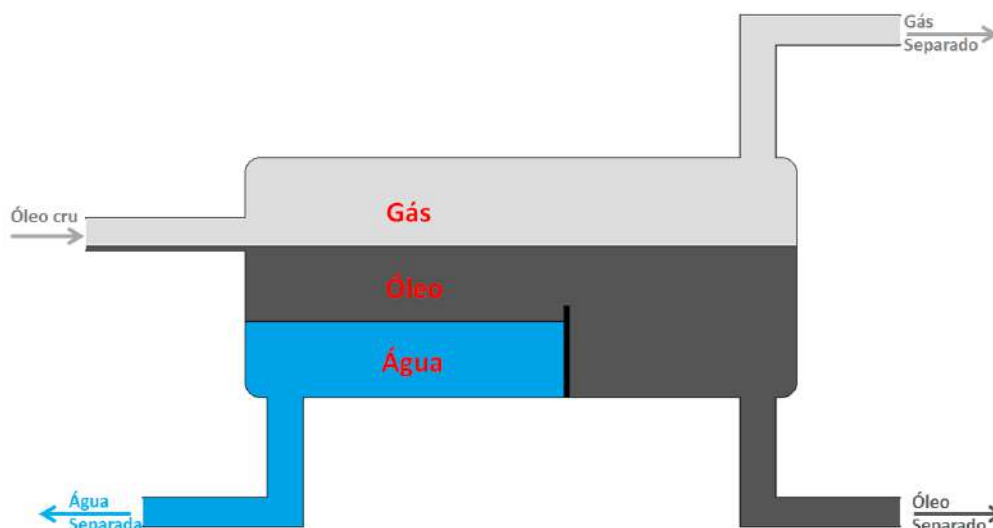


Nos últimos trinta anos, o uso do gás natural cresceu significativamente, alcançando 24,2% da energia primária total e 21% das emissões de GEEs. Já no Brasil, a participação do gás natural ainda corresponde a apenas 12,2% da matriz energética brasileira (GUTIERREZ, 2022). Dessa forma, é perceptível como a relevância dessa *commodity* apresenta um ritmo crescente no cenário global, com expectativa forte de crescimento. Enquanto no cenário nacional, a baixa oferta inibe a demanda do mercado por esse insumo e leva a preços elevados a padrões internacionais.

2.2 PROCESSAMENTO DE GÁS NATURAL

Durante o processamento primário, ainda na plataforma, o gás natural associado é separado do óleo através de um separador trifásico, conhecido como vaso gravitacional, onde o gás, por ser um fluido menos denso, é inicialmente separado dos líquidos pela ação da gravidade (BIGDOLI, 2018). A Figura 4 ilustra a separação do gás por um separador trifásico tipicamente usado em plataformas. O gás separado é enviado para a unidade de compressão da plataforma e, então, os teores de contaminantes (principalmente água e CO₂) devem ser reduzidos a níveis aceitáveis antes da transferência do gás por gasodutos até a unidade de processamento ou reinjeção deste no poço (BIGDOLI, 2018).

Figura 4 – Representação típica de um separador trifásico de óleo



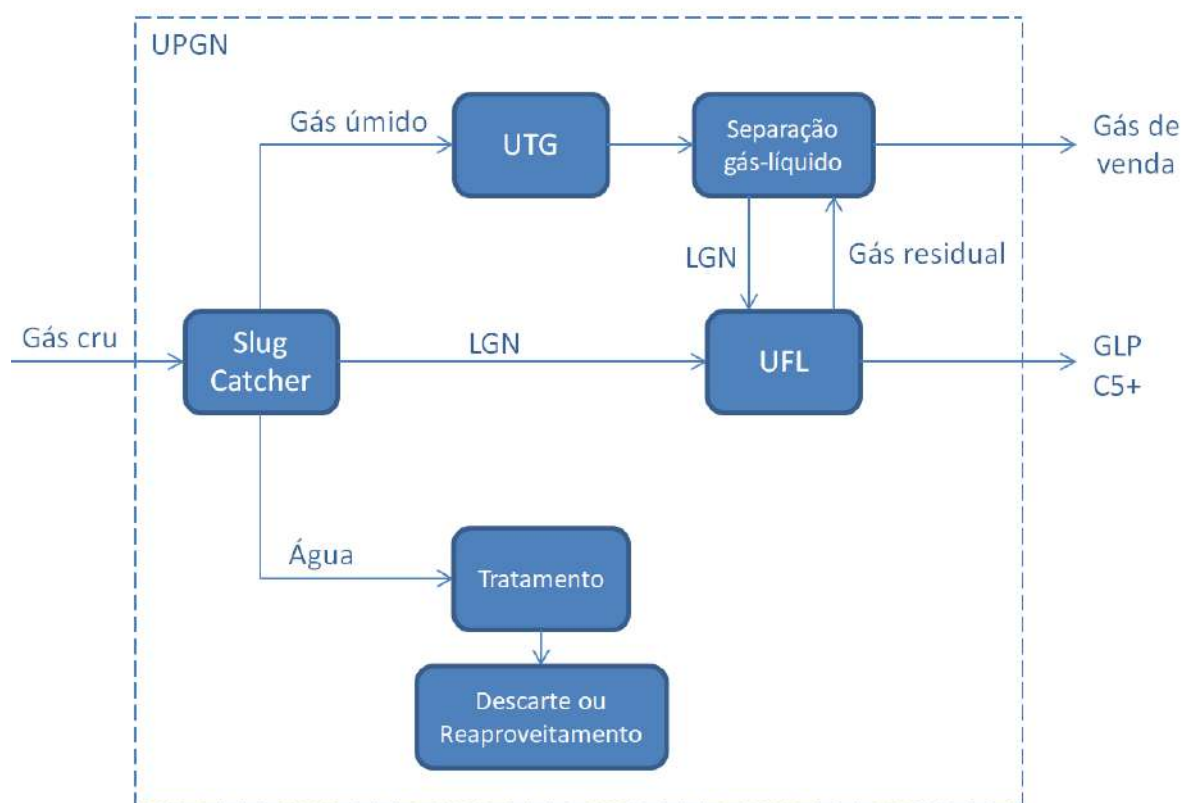
Após sua transferência, o gás recém-chegado na Unidade de Produção de Gás Natural (UPGN), chamado gás cru, é submetido a uma série de processos para fracionamento em correntes de produtos de maior valor de mercado. Sendo as principais correntes de produto: O Gás de Venda (GV), o Líquido de Gás Natural (LGN), o GLP e o $C5^+$. De forma singular, cada UPGN tem suas particularidades, desempenhos, capacidades máxima e mínima, modos de operação e produtos resultantes. Como essas unidades estão localizadas no meio da cadeia produtiva, é inerente que elas tenham que lidar com variações nas correntes de entrada, como mudanças em vazões, impurezas, composição, pressão e teor de contaminantes (SOUZA, 2022).

De modo a definir o tipo de planta de processamento de gás e as diferentes unidades de tratamento, os produtos a serem obtidos e suas especificações devem ser fixadas, onde a quantidade e o tipo de impurezas na alimentação determinam as etapas de tratamento necessárias (MOKHATAB et al., 2012). A Figura 5 mostra um diagrama de blocos típico de uma UPGN.

Conforme citado por SOUZA (2022), o *Slug Catcher* é o primeiro separador em que o gás natural cru é processado ao chegar à UPGN e é de onde serão direcionadas as correntes principais da unidade. No *Slug Catcher*, o gás cru é separado em três frações: Gases de hidrocarbonetos, chamado de gás úmido, líquido de hidrocarbonetos condensados, ou LGN, e água residual. A corrente de gás úmido é levada para uma Unidade de Tratamento de Gás (UTG) e então para uma unidade de separação gás-líquido, enquanto a corrente de LGN é

direcionada para a Unidade de Fracionamento de Líquidos (UFL). A fase aquosa será tratada para o descarte ou reaproveitamento na unidade.

Figura 5 - Diagrama de blocos de uma UPGN. Adaptado de SOUZA (2022)



O gás cru deve ser pré-tratado para remoção de gases ácidos (CO_2 e H_2S), água e mercúrio para produção de um gás purificado adequado para recuperação líquida a temperaturas criogênicas (MOKHATAB et al., 2012). Desse modo, a UTG pode conter as seguintes unidades de tratamento:

- Unidade de desidratação - Remoção de água para teores menores de 10 ppmv;
- Unidade de absorção com aminas - Remoção de CO_2 para teores abaixo de 1% e remoção de H_2S para teores abaixo de 4 ppmv;
- Unidade de absorção com solventes - Remoção de mercúrio para teores inferiores a $0,01 \mu\text{g}/\text{Nm}^3$.

Remoção de nitrogênio pode ser necessária em casos em que tal condicionamento leva à redução no volume de transporte e aumento do valor energético do gás. Destilação criogênica é o método mais econômico para grandes quantidades, porém a baixas vazões de

gás o uso de membranas e *Pressure Swing Adsorption* são tecnologias viáveis economicamente (MOKHATAB et al., 2012).

Conforme mostrado na Figura 5, após remoção dos contaminantes na UTG, o gás úmido é fracionado em GV e LGN por meio de uma coluna de destilação, chamada coluna desmetanizadora. A coluna desmetanizadora é dimensionada para separação de frações leves no topo, contendo pelo menos 88% de seu conteúdo molar em metano, pela resolução ANP 16 de 2008. Dessa seção de fracionamento, obtemos como produto de topo o GV, ou gás seco, e como produto de fundo o LGN, que poderá então ser direcionado para a UFL, para seu reaproveitamento.

A UFL é uma instalação industrial que objetiva separar o LGN, obtido em unidades anteriores, em correntes contendo etano, propano, GLP e C5⁺ (ANP, 2023). Na UFL encontram-se duas principais colunas de destilação: Uma coluna desmetanizadora, que separa compostos leves até C2 na corrente de topo, o gás de reciclo (GR), e compostos mais pesados que C2 na corrente de fundo, e uma coluna desbutanizadora, que recebe a corrente de fundo da desmetanizadora e separa em uma corrente de compostos leves até C4 no topo, que será condicionada e comercialmente vendida como GLP, e uma corrente de fundo contendo compostos mais pesados que o C4, conhecida por C5⁺, que também possui valor comercial e pode ser misturada à gasolina para especificação, reprocessada ou adicionada à corrente do petróleo.

Na avaliação econômica de uma UPGN, é importante quantificar o poder calorífico dos produtos fracionados para avaliar a receita obtida. Há três formas de expressar o valor calorífico de combustíveis: *Higher Heating Value* (HHV), *Lower Heating Value* (LHV) e *Gross Heating Value* (GHV). HHV é o calor liberado ao se reverter todos os produtos da combustão de volta à temperatura pré-combustão e condensando vapores de água produzidos, se referindo ao máximo de calor que poderia ser produzido pela queima do combustível. O GHV em reações de combustão se refere ao calor necessário para vaporizar a água presente no combustível antes da combustão. Já o LHV é obtido a partir da diferença entre o HHV e o calor de vaporização da água (ISLAM & DINCER, 2018). A correlação entre LHV e HHV é descrita pela Equação 1, citada em ISLAM & DINCER (2018).

$$LHV = HHV - (\Delta H_{vap}) \left(\frac{n_{H_2O,out}}{n_{fuel,in}} \right) \left(\frac{MW_{H_2O,out}}{MW_{fuel,in}} \right) \quad (1)$$

Em que:

- $\Delta H_{H_2O,vap}$ – Entalpia de vaporização da água;
- $n_{H_2O,out}$ – Número de moles de água vaporizada;
- $n_{fuel,in}$ – Número de moles de combustível queimado;
- $MW_{H_2O,out}$ – Massa molar da água;
- $MW_{fuel,in}$ – Massa molar do combustível.

Portanto, o LHV é a menor quantidade de calor que um combustível pode liberar por unidade de volume durante sua queima. Assim, ter conhecimento do valor de LHV dos gases produzidos torna-se relevante para estimar os lucros recebíveis dentro de uma UPGN.

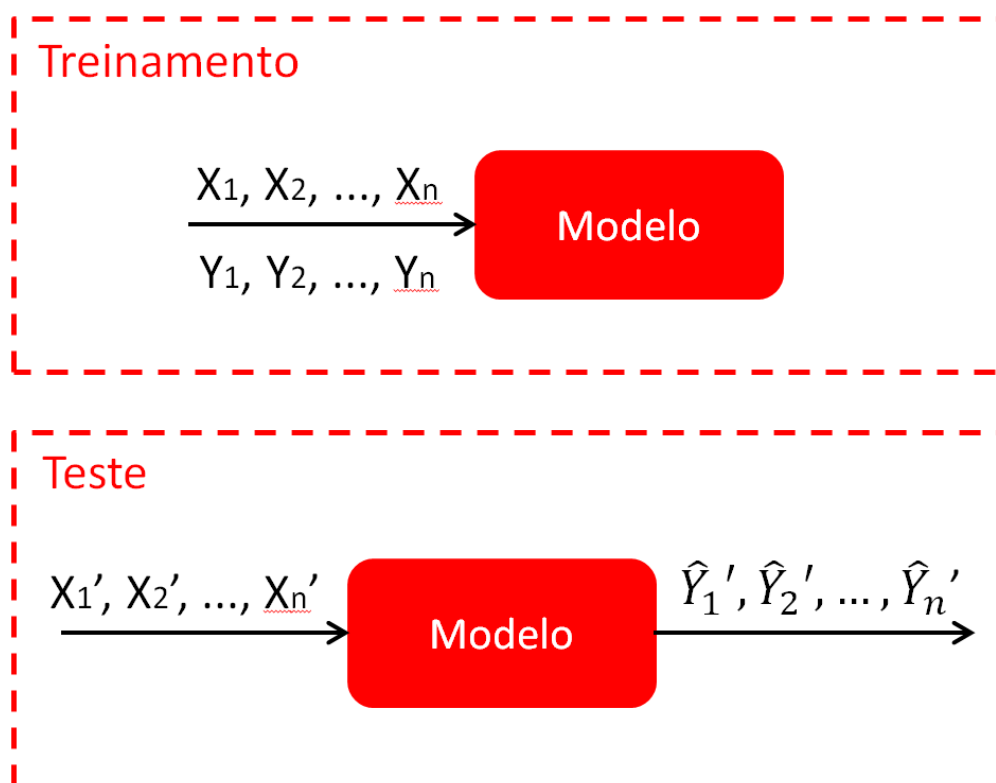
A capacidade de uma UPGN em variar suas condições operacionais, como vazões de alimentação, temperatura, pressão e razão de refluxo de colunas, a confere com uma ampla flexibilidade de produção, de modo que uma mudança das condições operacionais irá afetar principalmente os valores de LHV, vazão e composição dos produtos obtidos, variáveis que impactam diretamente no lucro da unidade. Dessa maneira, a busca por uma configuração que maximize o lucro por operação da planta torna-se um desafio complexo e requer o uso de ferramentas computacionais para previsão e otimização das variáveis atreladas ao lucro para se chegar à resolução desse problema.

2.3 MODELAGEM MATEMÁTICA

Modelos matemáticos podem ser interpretados como funções que recebem dados de entrada, valores de variáveis independentes, e retornam com dados de saída, valores para as variáveis dependentes. A construção de modelos paramétricos envolve uma abordagem de dois passos: Primeiro, o modelo assume uma função que descreve a relação entre as variáveis independentes com a variável dependente e, depois, é feito um procedimento de ajuste do modelo com dados de treinamento (JAMES et al., 2013).

Conforme apresentado no diagrama superior da Figura 6, a etapa de treinamento de um modelo consiste em inserir dados existentes do sistema no modelo, com o intuito que o algoritmo de modelagem ajuste esses dados em sua função. A etapa de teste, ilustrada no diagrama inferior da Figura 6, se baseia em usar esse modelo para previsão de novos dados do sistema. Dessa forma, elaborar um modelo matemático que represente uma variável de um sistema é particularmente útil para previsão daquela variável com base em novos dados.

Figura 6 - Diagrama de blocos que ilustra o treinamento e o teste de um modelo



2.3.1 Preparo dos dados

Antes da construção de um modelo matemático, é importante preparar as variáveis que serão utilizadas no treinamento, chamadas observações. É recomendável que as observações fornecidas sejam representativas ao sistema, ou seja, possuam valores bem distribuídos e na ausência de viés dentro do intervalo definido. Faz-se uso de métodos de amostragem durante a etapa de coleta de dados, com objetivo de se obter um banco de dados mais representativo ao sistema proposto. Dentre os métodos de amostragem, o *Latin Hypercube Sampling* (LHS) baseia-se na estratificação dos experimentos gerados em uma figura geométrica de hipercubo latino, para gerar amostras quase aleatórias em uma distribuição multidimensional. Uma vantagem do LHS surge quando a resposta é dominada por uma pequena quantidade de componentes de entrada. Nesse caso, o método garante que cada componente seja representado de forma completamente estratificada, sem levar em consideração quais componentes mostrem maior importância (MCKAY et al., 1979).

Após a coleta dos dados, é de interesse realizar o preparo dessas observações antes de usá-las para construção do modelo. Dois processos são comumente realizados para condicionar dados: Remoção de valores atípicos, ou *outliers*, e normalização dos valores. A

presença de pontos que destoam muito dos demais, chamados *outliers*, aumenta a variância das observações e torna mais difícil o ajuste do modelo, por isso é importante remover essas observações da amostragem, uma vez identificadas. A irregularidade nos valores das variáveis dos dados de entrada pode trazer um viés ao modelo construído. Por isso, é comum realizar a normalização das observações, limitando seus valores a um intervalo definido – normalmente entre zero e um. Uma forma de normalização comumente aplicada é a min-max, que utiliza os limites superior e inferior da variável normalizada, aplicando a fórmula descrita na Equação 2.

$$V_{norm} = \frac{V - V_{LI}}{V_{LS} - V_{LI}} \quad (2)$$

Sendo:

- V_{norm} – Valor normalizado da observação;
- V – Valor da observação;
- V_{LI} – Limite inferior definido para a observação;
- V_{LS} – Limite superior definido para a observação.

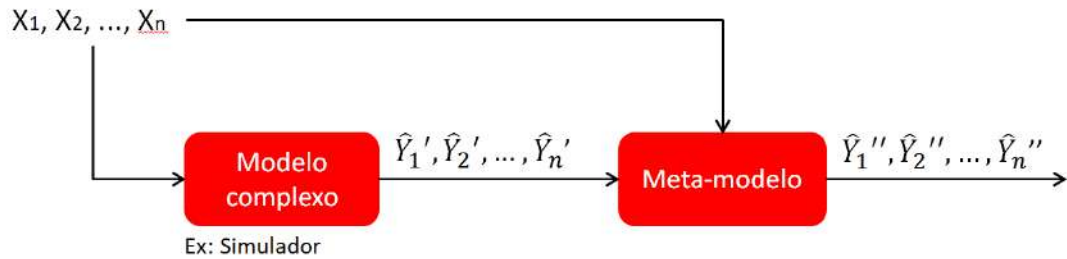
2.3.2 Meta-Modelos

Metamodelagem se refere à elaboração de um modelo simplificado de um modelo mais complexo. Tal conjunto de técnicas é amplamente utilizado quando o modelo complexo é inaplicável ou de difícil aplicação no contexto proposto (NEVEUX et al., 2022). Dessa forma, a partir de uma quantidade suficiente de dados de entrada produzidos pelo modelo complexo, é possível treinar um meta-modelo capaz de entregar resultados para o sistema suficientemente próximos àqueles observados pelo modelo complexo, desde que o meta-modelo construído apresente uma boa representatividade desse modelo.

A Figura 7 ilustra o procedimento de elaboração de um meta-modelo a partir de um modelo complexo e um banco de dados de entrada. Primeiro, o banco de dados de entrada é usado para obter um banco de dados de saídas a partir do modelo complexo. Depois, os dados de entrada e saída são divididos: uma parte será usada no treinamento e outra parte no teste do meta-modelo. Durante o treinamento, o meta-modelo ajusta os dados recebidos para fornecer estimativas similares ao resultado do modelo complexo, enquanto no teste a acurácia obtida vai dizer o quão próximas estão essas estimativas do esperado. Caso as estimativas do meta-

modelo sejam suficientemente próximas ao do modelo complexo, pode-se afirmar que temos um meta-modelo representativo.

Figura 7- Diagrama de blocos do funcionamento de um meta-modelo



Softwares de simulação de processos geralmente fazem uso de modelos fenomenológicos complexos capazes de representar mais fielmente o comportamento de um sistema real para uma ampla faixa de valores. Para aplicações onde há alta demanda por eficiência computacional, ao exemplo da solução de um problema de otimização, o uso de modelo complexo pode levar a um esforço maior (NEVEUX et al., 2022). Nesses casos, torna-se útil o treinamento de meta-modelos capazes de representar os resultados do simulador, melhorando a eficiência desse processo. Ali (2022) reportou as vantagens no emprego de meta-modelos para resolução de problemas de otimização com simulador Aspen Hysys, citando a minimização do custo computacional realizado durante as iterações da otimização.

2.3.3 Regressão Linear Múltipla (RLM)

Dentre as diferentes metodologias conhecidas na literatura do *Statistical Learning* para treinamento de modelos preditivos contendo múltiplas variáveis, a mais simples e interpretável é a RLM, que adapta o conceito da regressão linear simples para comportar múltiplas variáveis. Na RLM, o modelo se ajusta às observações por meio da Equação 3, utilizando o método dos mínimos quadrados para estimar os coeficientes $\hat{\beta}$, e a partir da mesma equação é capaz de prever o novo valor da variável com base em novos dados de entrada.

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^p \hat{\beta}_i x_i \quad (3)$$

Sendo:

- \hat{y} – Variável dependente estimada;
- $\hat{\beta}_0$ – Coeficiente linear;
- $\hat{\beta}_i$ – Coeficiente angular da variável i ;
- x_i – Variável independente i ;
- p – Número de variáveis independentes.

Em contraponto à sua simplicidade, a RLM pode não apresentar uma boa acurácia em comparação a outras metodologias para sistemas não lineares. O modelo de regressão linear assume que há uma relação linear entre variáveis dependentes e independentes. Se essa relação se afaste de ser linear, então se pode afirmar que as conclusões tiradas do modelo ajustado são suspeitas (JAMES et al, 2013). Nesses casos, metodologias mais robustas e flexíveis são preferíveis para obter um melhor ajuste do modelo, ao exemplo de métodos baseados em árvore de decisão.

2.3.4 Árvore de decisão

A árvore de decisão define uma série de ramificações a critério de testes lógicos para transformar as variáveis independentes observadas em valores de variáveis dependentes. O conjunto de ramificações de testes lógicos definidos recebe o nome de árvore de decisão, onde cada teste é uma ramificação e os valores finais são as raízes da árvore. Quando a resposta prevista ao final do método é qualitativa, temos uma árvore de classificação. Já quando a resposta prevista é quantitativa, temos uma árvore de regressão.

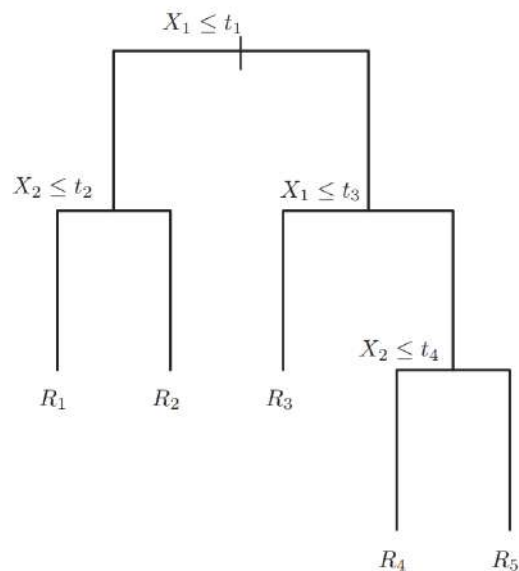
Um esquema de árvore de decisão é mostrado na Figura 7, na qual X_1 e X_2 são variáveis independentes que são testadas a cada ramo em inequações contendo os valores t_1 , t_2 , t_3 e t_4 até alcançarem um valor de variável dependente, dentre R_1 , R_2 , R_3 , R_4 ou R_5 . Durante a regressão, a variável independente começa do topo da árvore e se ramifica até alcançar um resultado na base da árvore.

De forma geral, a construção de uma árvore de decisão para problemas de regressão é realizada em duas etapas (JAMES et al., 2013):

1. O espaço de valores possíveis para as variáveis independentes é dividido em uma quantidade definida não sobressalente de regiões;

2. Para cada observação que se adequa a uma região específica, é atribuído o resultado de tal região, que é a média da resposta dos valores para a observação na região.

Figura 8 - Exemplo de uma estrutura de árvore de decisão. Fonte: JAMES et al. (2013)



O método de regressão por árvore de decisões permite uma maior flexibilidade quando comparado ao método de RLM, uma vez que o espaço de soluções não assume um formato linear. Por tal motivo, árvores de decisões são usadas como base para outros métodos de modelagem do *Statistical Learning*.

2.3.5 Random Forest

O método de *Random Forest* parte da premissa de elaborar um conjunto de árvores de decisão para previsão da variável dependente. Melhorias significativas na acurácia de previsão são obtidas a partir da construção de um conjunto de árvores de decisão e conduzindo uma votação pelo resultado mais popular dentre elas, num procedimento nomeado de *Random Forest* (BREIMAN, 2001). O método utiliza as técnicas de *Bootstrapping* e *Bagging* para obtenção de um modelo de previsão robusto e de maior acurácia (JOSSO et al., 2023).

A técnica de *Bootstrapping* consiste na criação de um novo banco de dados de treinamento a partir de uma amostragem com reposição do banco de dados inicial. Essa

técnica pode ser aplicada a qualquer método de modelagem como forma de reduzir a variância do modelo sem aumentar o viés. O uso do *Bootstrapping* em *Random Forests* é interessante, uma vez que permite reduzir a sensibilidade do modelo, levando a resultados mais homogêneos.

A técnica estatística de *Bagging* é utilizada no método *Random Forest* e tem como premissa retornar a resposta final do modelo como o valor médio das respostas fornecidas por todos os modelos de árvores de decisão treinados por um banco de dados gerado por *Bootstrapping*. Através do uso da média dos resultados dos modelos, calculada pela Equação 4, pode-se obter uma resposta com menor variância (quando comparado ao uso de uma única árvore de decisão), obtendo assim um modelo mais robusto. *Bagging* demonstra fornecer aprimoramento significativo em acurácia ao usar combinações de centenas ou milhares de árvores em um único procedimento (JAMES et al., 2013).

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x) \quad (4)$$

Sendo:

- \hat{f}_{bag} – Estimativa final do modelo após *Bagging*;
- B – Número de árvores de decisões treinadas por banco de dados gerado por *Bootstrapping*;
- \hat{f}^{*b} – Resultado do b-ésimo modelo obtido por *Bootstrapping*.

2.3.6 Gradient Boosting

O *Gradient Boosting*, assim como o *Random Forest*, é um método de elaboração de modelos baseado no aprimoramento do conceito da árvore de decisão com uso de múltiplas árvores. A diferença desse método para o anterior é que este tem uma abordagem de aprimoramento gradativo do modelo inicial de árvore a partir da contribuição de outras árvores, ao invés do uso de múltiplas árvores paralelamente. Por meio da técnica de *Boosting*, o algoritmo do método reduz gradativamente o erro residual do modelo de árvore de decisão, calculado pela Equação 5, e assim eleva sua acurácia.

$$h_m(x) = y - F_m(x) \quad (5)$$

Considerando:

- $h_m(x)$ – Erro residual da observação x ;
- y – Real valor da variável dependente;
- $F_m(x)$ – Previsão do modelo para x .

A estratégia de *Boosting*, ao contrário do *Bagging*, dispensa o uso de *Bootstrapping* para amostragem das árvores de decisão. A ideia é reduzir o erro residual da primeira árvore de decisão, calculado pela Equação 5, através da incorporação gradual da contribuição de novas árvores por meio da operação descrita na Equação 6. Uma vez que os modelos são construídos sequencialmente, cada nova árvore é ajustada com informações da árvore anterior. Após alcançar o número de passos definido, o *Boosting* então termina o ajuste do modelo, resultando na estimativa final descrita na Equação 7.

$$\hat{f}(x) = \hat{f}(x) + \lambda \hat{f}^b(x) \quad (6)$$

$$\hat{f}(x) = \sum_{b=1}^B \lambda \hat{f}^b(x) \quad (7)$$

Sendo:

- $\hat{f}(x)$ – Estimativa final do modelo treinado por *Gradient Boosting*;
- B – Parâmetro referente ao número de árvores;
- λ – Parâmetro de encolhimento do *Boosting*;
- $\hat{f}^b(x)$ – Resultado do b -ésimo modelo obtido por *Boosting*.

Por meio da metodologia descrita, o método é capaz de aprender de forma lenta e gradual com o perfil dos dados do sistema proposto e assim gerar um ajuste robusto. O parâmetro de encolhimento (λ), presente nas Equações 6 e 7, precisa ser ajustado adequadamente, de modo a permitir que a taxa de aprendizado do modelo seja suficiente para capturar a tendência dos dados. O *Gradient Boosting* é proeminentemente empregado juntamente com árvores de decisão, sendo capaz de superar até modelos de *Random Forest* para alguns conjuntos de dados (BAI & CHANDRA, 2023).

2.3.7 Kriging

O *Kriging*, também conhecido como Processo Gaussiano, é um método robusto para treinamento de modelos de regressão que se diferencia por se basear no conceito geoestatístico de estimação de valores dos dados de saída a partir da posição das observações nas proximidades conhecidas. *Kriging* é uma metodologia usada em problemas de otimização “black-box” (onde a forma da função de origem dos dados não é conhecida) devido à sua eficiência computacional, sua habilidade na modelagem matemática não linear e vantagem de prover um intervalo de confiança para cada predição (BOUKOUVALA & IERAPETRITOU, 2012). A diferença de outros métodos para o *Kriging* é que este faz o mínimo de suposições sobre a forma da função a qual ele visa representar, o que justifica a eficácia desse método (FORRESTER & KEANE, 2009). Matematicamente, o modelo é baseado na Equação 8.

$$\hat{y}(x) = \hat{\mu}(x) + z(x) \quad (8)$$

Em que:

- $\hat{y}(x)$ – Valor previsto da variável dependente;
- $\hat{\mu}(x)$ – Termo do modelo responsável por capturar a tendência geral dos dados em uma regressão, descrito pela Equação 9;
- $z(x)$ – Termo aleatório do modelo responsável pelo ajuste fino dos resultados, descrito pela Equação 10.

$$\hat{\mu}(\beta, x) = f(x)^T \beta \quad (9)$$

Sendo:

- β – Um vetor composto por parâmetros obtidos via regressão;
- $f(x)$ – Vetor contendo um conjunto de funções escolhidas no método.

$$z(x) = r^T(x)R^{-1}(Y - F\beta) \quad (10)$$

Sendo:

- $r(x)$ – Vetor de correlação entre os pontos x e pontos de projeto definidos no modelo

- R – Matriz de correlação dos pontos de projeto,
- Y – Vetor de valores da variável dependente,
- F – Matriz transposta de um conjunto de funções escolhidas pelo modelo e
- β – Vetor de parâmetros obtidos via regressão.

Acoplando as Equações 9 e 10 na Equação 8, chega-se à expressão geral que o método de *Kriging* usa para estimar o resultado de uma nova observação i a partir da Equação 11.

$$\hat{y}(x_i) = f(x_i)^T \beta + r^T(x_i) R^{-1} (Y - F \beta) \quad (11)$$

Então, para obter a predição da resposta em um novo valor de x somente é necessário computar os vetores $f(x_i)$ e $r(x_i)$ e realizar duas simples operações de multiplicação (CARPIO, 2019). Devido ao custo de estimar os parâmetros no *Kriging*, o método torna-se mais útil quando a função a qual ele deseja representar é particularmente intensiva computacionalmente (FORRESTER & KEANE, 2009).

2.3.8 Validação do modelo

Após treinamento do modelo por meio de um dado método, é necessário testá-lo com a inserção de novos valores de dados de entrada, de modo a medir a acurácia do mesmo e avaliar sua representatividade ao sistema. Com os valores previstos para as variáveis dependentes do modelo e de posse dos resultados reais das mesmas observações, métricas estabelecidas podem ser calculadas para quantificar a acurácia do modelo. Há duas principais métricas usadas para avaliar a acurácia de um modelo: O MSE e o R^2 .

No cenário da regressão, a métrica mais comumente usada é o erro quadrático médio (do inglês MSE), definido pela Equação 12 (JAMES et al., 2013).

$$MSE = \frac{1}{n} \sum_{i=1}^n [y_i - \hat{f}(x_i)]^2 \quad (12)$$

Em que:

- n – Número do espaço amostral fornecido;
- y_i – Valor real da variável dependente;
- $\hat{f}(x_i)$ – Previsão da variável dependente para aqueles valores de entrada (x_i).

Por mais que o valor de MSE seja uma métrica usual para medir o erro de um modelo, ele é de difícil interpretação, uma vez que seu valor depende do valor da variável ao quadrado. Portanto, uma métrica de fácil interpretação para medição do erro é o valor de R^2 , calculado pela Equação 13.

$$R^2 = 1 - \frac{\sum_{i=1}^n [y_i - \hat{f}(x_i)]^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

Em que:

- n – Número do espaço amostral fornecido;
- y_i – Valor real da variável dependente para a observação i ;
- $\hat{f}(x_i)$ – Previsão da variável dependente do modelo para a variável de entrada x_i ;
- \bar{y} – Média das variáveis dependentes.

Dessa maneira, a métrica R^2 expressa o quanto um modelo consegue representar o comportamento real da variável dependente com relação às variáveis independentes. O valor de R^2 varia entre 0 e 1, tendendo ao valor 1 quando o modelo se aproxima de uma boa representação.

No entanto, essa métrica na forma descrita pela Equação 13 é mais apropriada a modelos de uma única variável independente, uma vez que a inclusão de múltiplas variáveis independentes é capaz inflar esse valor e passar um valor impreciso de acurácia. Portanto, a métrica de R^2 deve ser ajustada por meio da Equação 14 para reduzir a distorção causada pelo aumento no número de variáveis, assim trazendo maior segurança para avaliação de modelos de múltiplas variáveis (James et al., 2013).

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n-1)}{n-p-1} \quad (14)$$

Sendo:

- R^2_{adj} – Valor de R^2 ajustado;
- R^2 – Valor de R^2 obtido pela Equação 13;
- n – Número de observações no espaço amostral;
- p – Número de variáveis presentes no modelo.

2.4 OTIMIZAÇÃO

Uma otimização se baseia na seleção da solução de melhor resultado dentre uma série de soluções candidatas, onde a qualidade das soluções é quantificada por uma função objetivo, que deve ser minimizada ou maximizada (EL-HALWAGI, 2006). Dentro de um problema de otimização, devem ser estabelecidos os limites inferiores e superiores de cada variável independente usada, constituindo a fronteira da região de busca do problema. Também pode haver restrições para a otimização, ou seja, considerando a região de busca estabelecida, há condições definidas que o ponto de ótimo deve satisfazer.

Problemas de otimização podem ser univariáveis ou multivariáveis e os métodos para resolução desses problemas podem ser analíticos ou numéricos, diretos ou indiretos. Os métodos analíticos dependem de condições específicas relacionadas à forma da função objetivo para serem utilizados e, portanto, são de difícil aplicação para cenários onde a forma da função objetiva é desconhecida. Métodos numéricos promovem uma busca da solução na região de busca por tentativas, chamadas de iterações, calculando a função objetivo, e às vezes suas derivadas, para diferentes valores de variáveis independentes (PERLINGEIRO, 2005).

Os métodos diretos utilizam apenas o valor da função objetivo calculado a cada iteração, enquanto métodos indiretos utilizam, como informação adicional, o valor da derivada primeira da função objetivo nesses mesmos pontos e em alguns casos o valor da derivada segunda. Uma vez que usam a informação das derivadas, métodos indiretos convergem para um número menor de tentativas, porém o tempo despendido para o cálculo dessas derivadas, quando realizado de forma aproximada por perturbação, pode tornar tais métodos mais lentos (PERLINGEIRO, 2005).

2.4.1 Análise prévia do problema

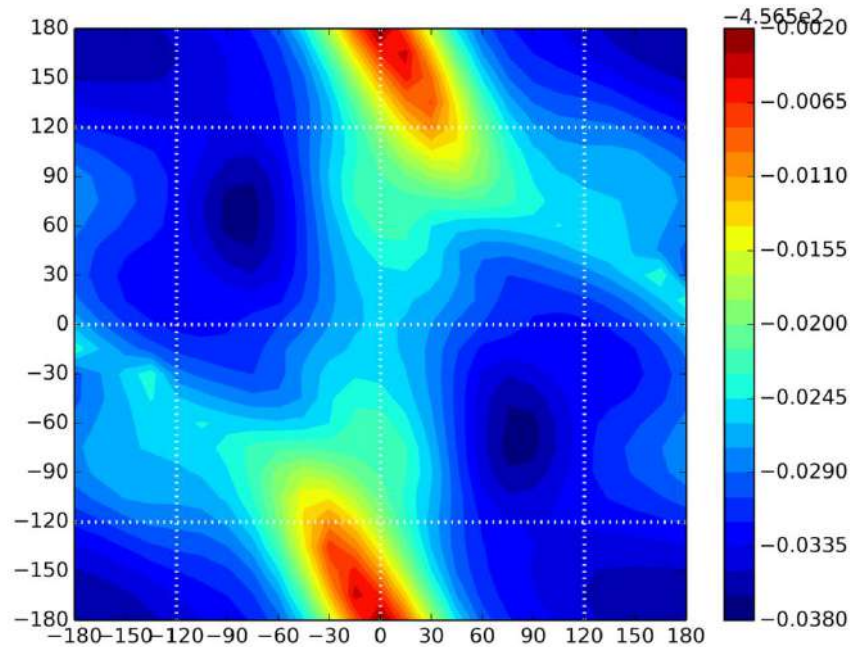
Os métodos numéricos mais comuns baseiam-se na suposição da unimodalidade da função objetivo, ou seja, terminam quando encontram um ponto extremo. Se a função for multimodal, esse extremo pode ser apenas local (PERLINGEIRO, 2005). Nesse prisma, antes da escolha de um método para resolução da otimização, é recomendada a realização de uma análise exploratória da região de busca para descobrir se o problema possui mais de uma região em que pode ser encontrada a solução ótima (multimodalidade). Uma análise exploratória da região de busca consiste em coletar uma quantidade representativa de pontos e

então gerar um gráfico de relevo que mostre a distribuição da função objetivo em relação aos valores de variáveis independentes. No caso de mais de duas variáveis independentes, torna-se necessária a avaliação de múltiplos gráficos, de modo a cobrir todas as combinações de variáveis do problema. A quantidade de gráficos (G) necessária para realizar a análise exploratória da região de busca é dada pela combinação sem repetição de n variáveis em duas dimensões, expressa pela Equação 15.

$$G = \frac{n!}{2(n-2)} \quad (15)$$

Durante a análise exploratória de um problema de otimização, a observação de mais de uma região de ótimo local é um indicativo de que alguns métodos baseados em convergência local podem não ser capazes de alcançar o resultado ótimo global da função objetivo. Portanto, em casos como o ilustrado na Figura 8, devem ser empregados métodos de otimização global para encontrar o ponto de ótimo na região de busca.

Figura 9 - Exemplo de problema com múltiplos ótimos locais



2.4.2 Nelder-Mead

Um exemplo de método direto utilizado na resolução de problemas de otimização é o Nelder-Mead, descrito em NELDER & MEAD (1965). O método consiste na busca de um

mínimo local em uma região a partir da convergência de um *simplex*, uma forma geométrica de $n+1$ vértices sendo n o número de variáveis independentes do problema. Para melhor compreensão do conceito, um *simplex* em duas dimensões se refere a um triângulo e um *simplex* em três dimensões seria um tetraedro. O método consiste na deformação desse poliedro a cada iteração, visando convergi-lo a um ponto ótimo por meio de um algoritmo baseado quatro operações algébricas: Reflexão, expansão, contração e redução.

O algoritmo se inicia a partir da operação de reflexão do *simplex*, onde o vértice de maior resultado na função objetivo é refletido com relação ao centroide do poliedro. Dependendo do resultado da reflexão, a próxima operação realizada sobre o *simplex* será uma das seguintes:

- Se o ponto refletido for o menor valor entre os vértices, ocorre a operação de expansão do poliedro com relação ao centroide;
- Se o ponto refletido estiver entre o pior e o segundo pior vértice, ocorre a operação de contração do poliedro com relação ao centroide;
- Se o ponto refletido se tornar pior que o ponto original, o poliedro sofre uma redução de tamanho com base no melhor vértice.

Uma vez que o desvio de cada vértice em relação ao centroide convergir a um valor menor do que uma tolerância estabelecida, o algoritmo se encerra e o menor valor dentre os vértices é informado como sendo o ótimo da função.

O método algorítmico de Nelder-Mead é um método direto e por isso pode necessitar mais iterações para atingir a convergência, devido à ausência de informação da derivada da função. No entanto, tais métodos se mostram vantajosos por permitirem a realização de otimização acoplada a simulações complexas, onde a obtenção da informação da derivada é de elevado esforço computacional (SOUZA, 2022).

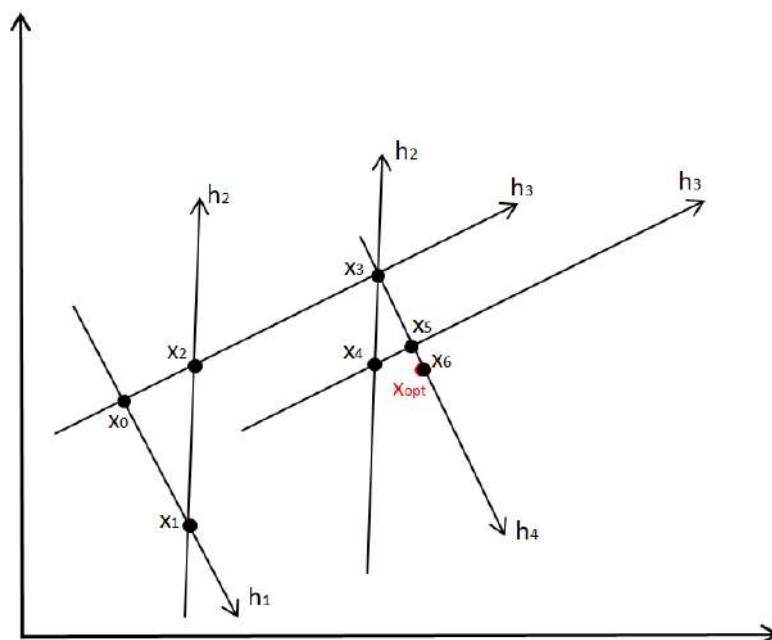
2.4.3 Powell

Outro método direto de otimização é o Powell, proposto em 1984 pelo autor de mesmo nome. Dado um problema de otimização de n variáveis, a partir de uma estimativa inicial x_0 e de um conjunto de n vetores direcionais $\{h_1, h_2, \dots, h_n\}$, o método de Powell realiza um algoritmo de busca vetorial por um mínimo local dentro da região de busca. O procedimento iterativo do algoritmo consiste nas seguintes etapas:

1. Começando em x_0 , a função é avaliada ao longo do primeiro vetor h_1 . O valor de mínimo encontrado na avaliação é definido como o ponto de partida para a próxima busca x_1 ;
2. Começando em x_1 , o processo anterior se repete para o vetor h_2 , e assim sucessivamente para os demais vetores passados ao método. O valor de mínimo após avaliação do último vetor h_n será o ponto de partida da próxima iteração x_n ;
3. Define-se um novo vetor h_{n+1} , baseado no escalar da posição final x_n com relação à estimativa inicial x_0 daquela iteração;
4. O conjunto de n vetores é alterado para a próxima iteração, de modo que o vetor h_i que mais contribuiu para a nova posição x_n é substituído pelo vetor h_{n+1} ;
5. Repetição do processo iterativo, agora usando o novo conjunto de n vetores para realização da busca. Se a diferença entre a avaliação final de uma iteração e da iteração anterior for inferior a uma tolerância definida, o método é encerrado e o último mínimo é definido como sendo o valor ótimo.

A resolução de um problema de otimização para duas variáveis ($n = 2$) pelo método de Powell é ilustrado na Figura 9, partindo de um ponto inicial x_0 utilizando os vetores h_1 e h_2 na busca bidirecional da primeira iteração e buscando alcançar um ponto ótimo desconhecido x_{opt} (em vermelho).

Figura 10 - Exemplo de aplicação do método de Powell



2.4.4 SLSQP

O *Sequential Least Squares Programming* (SLSQP) é um exemplo de método de otimização indireto, proposto por SCHITTKOWSKI em 1982. O método é capaz de minimizar localmente a função objetivo a partir da resolução de um subproblema de programação quadrática. O algoritmo do método é descrito por KRAFT (1988).

Uma vantagem no uso do SLSQP é que esse método de otimização é baseado no gradiente da função objetivo, levando a uma maior eficiência na convergência do problema graças ao uso da informação da derivada. Algoritmos de otimização que dispensam o uso do gradiente são por vezes mais fácil de implementar ao problema, porém são tipicamente menos eficientes que métodos baseados no uso do gradiente (AKBA et al., 2023).

2.4.5 PSO

Para a otimização de um problema com múltiplas regiões de ótimos locais, é necessário utilizar métodos de busca global, capazes de analisar de forma ampla a região de busca e retornar o máximo ou mínimo global. Tais métodos resultam em um maior esforço computacional, quando comparados a métodos de otimização local, porém são capazes de convergir para a solução global do problema.

O *Particle Swarm Optimization*, PSO (KENNEDY & EBERHART, 1995), é um algoritmo não determinístico amplamente usado para otimização global, assim como os Algoritmos Genéticos, GAs (HOLLAND, 1992), são métodos baseados em comportamento populacional (MORAES et al., 2014). O PSO parte de um número definido de partículas pontuais espalhadas na região de busca, chamado enxame, as quais irão procurar em conjunto o melhor resultado. Esse método foi inspirado no comportamento social de grupos de peixes e aves, onde cada integrante se move à procura do mesmo objetivo, compartilhando informações entre os elementos do enxame para uma melhor eficiência de busca.

A posição futura de cada partícula se baseia na velocidade inicial daquela partícula, em informações coletadas por experiências anteriores da mesma partícula e também em informações coletadas em experiências anteriores do grupo como um todo. A cada iteração do método PSO, a posição de cada partícula é atualizada com base Equação 16.

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (16)$$

Sendo:

- x_i^{k+1} – Nova posição, que será assumida pela partícula i na iteração k ;
- x_i^k – Posição da partícula i na posição k ;
- v_i^{k+1} – Nova velocidade que será assumida pela partícula i . Calculada através da Equação 17.

$$v_i^{k+1} = [w v_i^k] + [\varphi_p r_p (p_i - x_i^k)] + [\varphi_g r_g (g_i - x_i^k)] \quad (17)$$

Em que:

- w – Constante de contribuição inercial;
- v_i^k – Velocidade da partícula i na iteração k ;
- φ_p – Constante de contribuição do fator cognitivo;
- r_p – Número aleatoriamente gerado;
- p_i – Melhor posição encontrada pela partícula i ;
- φ_g – Constante de contribuição do fator social;
- r_g – Número aleatoriamente gerado;
- g_i – Melhor posição encontrada pelo enxame.

A Equação 17 apresenta uma interpretação intuitiva baseado no conceito do método. O primeiro termo, $[w v_i^k]$, representa a influência da velocidade anterior da partícula (v_i^k) sobre a nova velocidade, simulando o efeito de inércia. O segundo termo, $[\varphi_p r_p (p_i - x_i^k)]$, representa a influência do fator cognitivo daquela partícula na sua nova velocidade. O fator cognitivo é representado pela distância atual da partícula com relação à melhor posição que aquela partícula já encontrou. Por fim, o terceiro termo, $[\varphi_g r_g (g_i - x_i^k)]$, representa a influência do fator social daquela partícula na sua nova velocidade. Sendo o fator social representado pela distância da partícula com relação à melhor posição que todo o enxame já encontrou. Sendo o peso de cada contribuição definido pelos fatores w , φ_p e φ_g .

A partir do algoritmo descrito, o PSO é capaz de convergir para o ótimo global da região de busca, com uma elevada probabilidade. Apesar de normalmente demandar um grande número de avaliações da função objetivo, o PSO se mostra um método robusto para resolver problemas de otimização global (MORAES et al., 2014).

3. METODOLOGIA

3.1 SIMULAÇÃO DO PROCESSO

A simulação foi realizada em software Aspen Hysys versão 14, sobre uma planta de UFL presente em uma UPGN. A simulação utilizou o pacote termodinâmico Peng-Robinson, que é adequado para descrever misturas de hidrocarbonetos, para estimar propriedades termodinâmicas dos fluidos. A composição das duas correntes de entrada, representadas na Figura 10, é mostrada na Tabela 1.

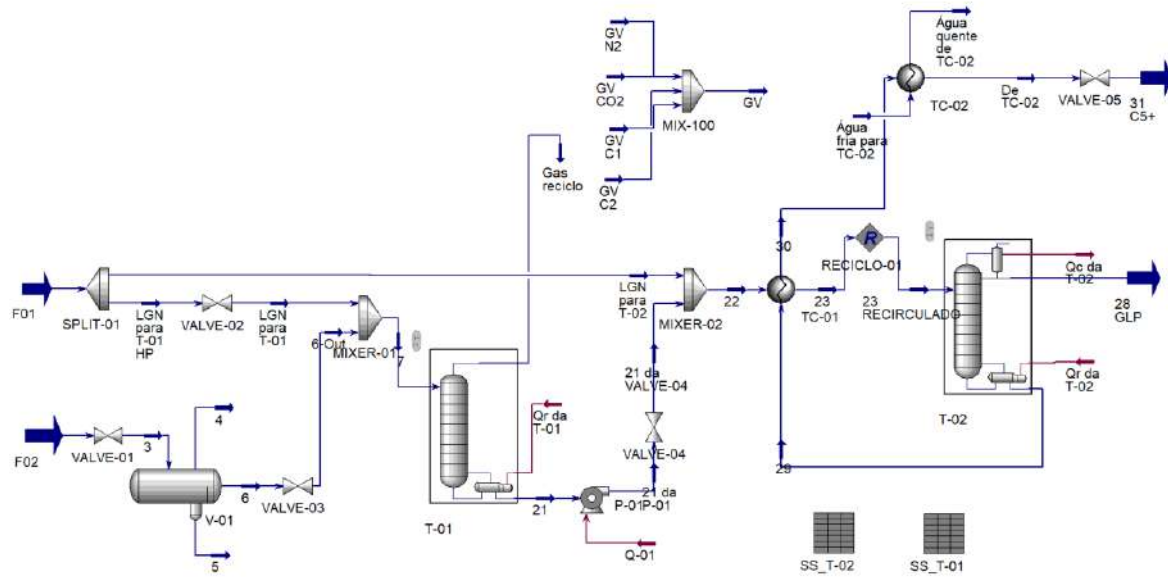
Tabela 1 - Condições das correntes de entrada da simulação da planta de gás natural

Componente	Vazão mássica em F01 (kg/h)	Vazão mássica em F02 (kg/h)
N ₂	-	16
H ₂ O	-	3
CO ₂	27	177
CH ₄	1	2017
C ₂ H ₆	3571	1604
C ₃ H ₈	3427	3030
i-C ₄ H ₁₀	762	1083
n-C ₄ H ₁₀	1340	2503
i-C ₅ H ₁₂	273	932
n- C ₅ H ₁₂	302	1266
C ₆ H ₁₄	145	1298
C ₇ H ₁₆	39	1227
C ₈ H ₁₈	5	669
C ₉ H ₂₀	2	316
C ₁₀ H ₂₂	-	119

A Figura 10 mostra a simulação da unidade de fracionamento de líquidos realizada no software Hysys, contemplada com os seguintes elementos:

- Duas torres de destilação: Uma torre desetanizadora (T-01) e outra desbutanizadora (T-02);
- Um vaso de separação trifásico (V-01);
- Uma bomba (P-01);
- Dois trocadores de calor;
- Dois elementos de corrente: Um separador e um misturador;
- Cinco válvulas;
- Vinte e quatro correntes.

Figura 11 - Simulação da UFL em Hysys



A T-01 recebe uma corrente de LGN proveniente da mistura da corrente F02 com parte da corrente F01, produzindo uma corrente de topo de GR rica em frações mais leves que propano e uma corrente de fundo de LGN contendo principalmente frações mais pesadas que etano. A corrente de fundo da coluna T-01 é misturada com parte da corrente F01 para alimentação da T-02, que irá produzir GLP como corrente de topo, produto concentrado em propano e butano, e de fundo uma corrente de C5⁺, com as frações mais pesadas que butano.

Da corrente de topo da T-01, foi estimada a parte de GR que seria transformada em GV na UPGN. As vazões mássicas de CO₂, N₂, CH₄ e C₂H₆ presentes no GR foram dispostas na tabela “SS_T-01” (Figura 11), na simulação. Foi estabelecido que o teor de metano para o GV é de no mínimo 80% da vazão total. Dessa forma, a vazão limitante de etano nessa corrente pode ser descrita pela Equação 19, derivada da Equação 18.

$$F_{C1} = 0,8 (F_{C1} + F_{C2} + F_{CO_2} + F_{N_2}) \quad (18)$$

$$F_{C2lim} = \frac{F_{C1} - 0,8 (F_{C1} + F_{CO_2} + F_{N_2})}{0,8} \quad (19)$$

Em que:

- F_{C2lim} - Vazão mássica limitante de etano no GR (em kg/h);
- F_{C1} - Vazão mássica de metano no GR (em kg/h);

- F_{CO_2} - Vazão mássica de dióxido de carbono no GR (em kg/h);
- F_{N_2} - Vazão mássica de nitrogênio no GR (em kg/h).

Na tabela “SS_T-01”, foi estabelecido que a vazão de etano a ser incluída no GV seria definida pela Equação 20.

$$F_{C_2}^{GV} = \min(F_{C_2}^{GR}, F_{C_2lim}^{GR}) \quad (20)$$

Sendo:

- $F_{C_2}^{GV}$ – Vazão mássica de etano no GV;
- $F_{C_2}^{GR}$ – Vazão mássica total de etano no GR;
- $F_{C_2lim}^{GR}$ – Vazão mássica limite de etano no GR.

Dessa maneira, a não ser que a quantidade de etano no gás de reciclo ultrapasse do limite estabelecido, é suposto que na UPGN todo o etano do gás de reciclo vá para o gás de venda. Caso ultrapasse o valor limitante, é suposto que na planta de separação gás-líquido da UPGN esse GR seja convertido em GV pela remoção do teor necessário de etano da corrente por meio de uma coluna desmetanizadora. Em seguida, as vazões individuais de cada componente na tabela foram exportadas para uma corrente pura de seus respectivos componentes e misturadas para obter-se o GV, incluído na simulação pelo sistema mostrado na Figura 12. Esse processo é importante para a obtenção do valor de LHV do GV, que será usado no cálculo da receita da unidade.

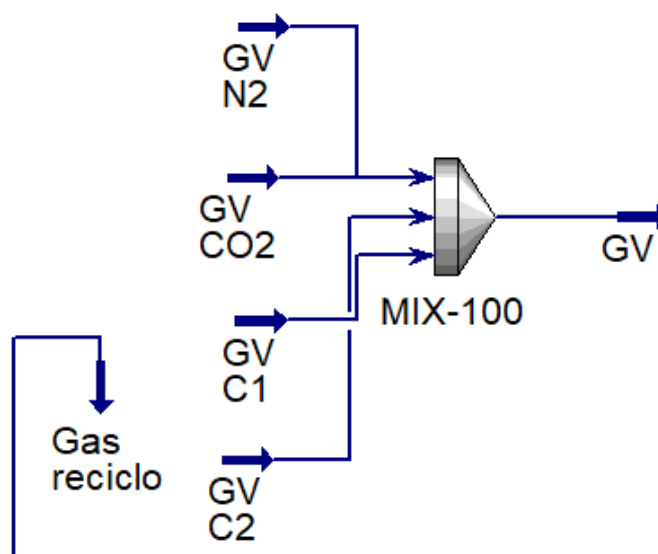
Os valores de vazão mássica e LHV das correntes de GV, GLP e C5⁺ foram importados para a tabela “SS_T-02” presente na simulação, para realização do cálculo da receita obtida nas condições da planta. Na tabela “SS_T-02”, o valor de receita foi calculado através da Equação 21.

$$Receita = \sum_i (F_i LHV_i c_i G) \quad (21)$$

Sendo:

- F_i – Vazão mássica da corrente i (em kg/h);
- LHV_i – *Lower Heating Value* da corrente i (em kJ/kg);
- c_i – Custo de venda da corrente i (em R\$/MMbtu);
- G – Constante de conversão de unidade de energia (igual a 9,48e-007 MMBtu/kJ).

Figura 12 - Simulação da mistura das correntes de Gás de Venda



Além da receita, foi calculado o custo operacional da planta a partir do consumo energético total e do custo por energia estimado. Como o consumo energético varia também com as condições de operação da planta, esse valor é obtido na simulação pela soma dos consumos de cada elemento consumidor de energia na planta. O valor do custo da operação para a simulação realizada foi inserido na tabela “SS_T-02” e foi calculado através da Equação 22.

$$Opex = c_e \sum_i E_i \quad (22)$$

Sendo:

- $Opex$ – Custo por operação (em R\$/h);
- c_e – Custo da energia (valor estimado em 0,5 R\$/kWh);
- E_i – Consumo energético do elemento consumidor i (em kW).

Os seguintes consumidores energéticos foram considerados na simulação da planta:

- Refervedor da coluna desetanizadora T-01;
- Refervedor da coluna desbutanizadora T-02;
- Bomba P-01;
- Condensador da coluna desbutanizadora T-02.

A partir do custo de operação e da receita, o lucro por operação pode ser calculado com base na Equação 23.

$$Lucro = Receita - Opex \quad (23)$$

Dessa maneira, o valor de lucro foi diretamente calculado na tabela “SS_T-02”, considerando três cenários de mercado diferentes, variando o preço de venda de cada produto da planta em cada um deles. A Tabela 2 mostra os valores de custo considerados para os produtos em cada cenário econômico.

Tabela 2 - Cenários de mercado considerados para a otimização

Corrente	Cenário 1 (MMbtu/m³)	Cenário 2 (MMbtu/m³)	Cenário 3 (MMbtu/m³)
GLP	40	50	30
C5 ⁺	30	40	50
GV	50	30	40

Para a simulação em questão, nas colunas da Tabela 3 foram inseridas as variáveis consideradas, a sua função como variável do problema, seus respectivos valores base e, caso seja uma variável de especificação, os valores da especificação de qualidade. Os valores base considerados nas variáveis foram obtidos da simulação inicial da planta de gás natural e foram utilizados para normalização dos dados utilizados nas próximas etapas do trabalho.

Tabela 3 - Variáveis consideradas ao sistema

Variável	Função	Valor base	Especificação
Temperatura de fundo de T-01 (T1)	Variável independente	74 ⁰ C	-
Temperatura de fundo de T-02 (T2)	Variável independente	165 ⁰ C	-
Razão de refluxo de T-02 (RR)	Variável independente	1,05	-
Fração volumétrica de etano em GLP	Variável dependente	6,5% v/v	< 14% v/v
Fração volumétrica de pentano em GLP	Variável dependente	0,95% v/v	< 2% v/v
Razão de C2/C5 em GLP	Variável dependente	6,8	< 16
PVR em C5 ⁺	Variável dependente	69,27 kPa	< 85 kPa
Lucro calculado para o cenário 1	Variável dependente	25.428,49 R\$/h	-
Lucro calculado para o cenário 2	Variável dependente	31.355,23 R\$/h	-
Lucro calculado para o cenário 3	Variável dependente	26.098,38 R\$/h	-

As variáveis independentes foram inseridas nas tabelas “SS_T-01” e “SS_T-02” e tiveram seus valores exportados para os respectivos elementos na simulação. Dessa forma, a

partir da mudança dos valores nas células das tabelas da simulação, era possível alterar as condições de operação da planta. Enquanto isso, os valores das variáveis dependentes eram importados para a tabela “SS_T-02”, para facilitar a coleta e análise dessas variáveis.

3.2 ANÁLISE DE VARIÁVEIS

Antes de seguir para a resolução do problema, é necessário compreender a correlação entre as variáveis independentes e as variáveis dependentes do sistema, para então definir os limites da região de busca na otimização respeitando os limites estabelecidos. Considerando que algumas variáveis agem como restrições do problema da otimização, os limites da região de busca devem ser impostos de forma adequada para aumentar a chance de obter pontos ótimos dentro das especificações passadas para as correntes de produtos, evitando o excesso de pontos fora de especificação. As especificações estabelecidas para o problema podem ser encontradas na Tabela 3.

Para isso, 200 valores das variáveis independentes foram gerados para um amplo intervalo de valores, mostrado na Tabela 4. Em seguida, os gráficos para cada variável dependente foram criados, com as variáveis independentes no eixo x e uma variável dependente no eixo y . Um algoritmo na linguagem Python foi utilizado para estabelecer uma interface com o Hysys, ler ou alterar valores das células contidas nas tabelas “SS_T-01” e “SS_T-02” e então gerar os pontos dos gráficos. Assim, o código em Python testou diferentes valores das variáveis independentes e salvou os resultados das variáveis dependentes obtidas para construção dos gráficos. As correlações foram analisadas e, baseado nas tendências de cada variável, os limites da otimização foram estabelecidos.

Tabela 4 - Intervalo usado na análise preliminar de variáveis

Variável independente	Mínimo	Máximo
T_1	50	100
T_2	150	200
RR	0,1	5,0

3.3 ELABORAÇÃO DE META-MODELOS

Usando os intervalos definidos na etapa de análise das variáveis, foram treinados diferentes meta-modelos para ajuste dos dados coletados a partir da simulação. Esses meta-

modelos foram depois testados e avaliados quanto suas métricas de MSE e R^2 . Os métodos para construção dos meta-modelos utilizadas foram: RLM, *Random Forest*, *XGBoost*, *Gradient Boosting* e *Kriging*. O problema em questão envolvia a predição de sete variáveis dependentes a partir de três variáveis independentes, as mesmas que estão representadas na Tabela 3.

Através do método de amostragem de LHS, foram gerados 450 valores para as variáveis independentes nos limites estabelecidos, em que 150 observações foram usadas para treinamento dos modelos construídos e 300 para teste destes modelos. Um algoritmo em Python inseria as variáveis independentes no Hysys, coletava os resultados das variáveis dependentes e as armazenava em um banco de dados. Em seguida, todas as variáveis do banco de dados eram normalizadas e exportadas para um arquivo Excel.

Durante a normalização, cada variável utilizou intervalos de normalização próprios, baseado nos limites inferiores e superiores que os valores de tal variável poderiam assumir. Os limites superiores e inferiores das variáveis independentes foram definidos durante a análise de variáveis, portanto estes foram os limites usados na normalização. No caso das variáveis dependentes de especificações, observou-se que um limite superior de 100 e um limite inferior de zero compreende um intervalo satisfatório de normalização. No caso das variáveis dependentes relacionadas ao lucro nos diferentes cenários econômicos, onde não são conhecidos os valores de máximo e mínimo da variável, foi estabelecido que o valor base normalizado adotasse o valor de 1,0. Portanto, a normalização das variáveis dos modelos seguiu a metodologia descrita na Tabela 5.

Tabela 5 - Normalização das variáveis do problema

Variáveis	V_{norm}
Variáveis independentes	$\frac{V - V_{LI}}{V_{LS} - V_{LI}}$
Variáveis dependentes de especificações	$\frac{V}{100}$
Variáveis dependentes de lucro	$\frac{V}{V_{base}}$

Considerando que cada meta-modelo é capaz de prever uma variável dependente, os dados normalizados foram usados para treinar sete meta-modelos, usando cinco métodos diferentes, levando a um total de 35 meta-modelos. Após o treinamento, os 35 meta-modelos

tiveram sua acurácia testada na previsão dos valores de variáveis dependentes do banco de dados normalizado de teste.

O código utilizado para criação dos meta-modelos foi retirado da biblioteca “*Scikit Learn*”, com exceção dos meta-modelos de *XGBoosting* e *Kriging*. O código do método de *XGBoosting* utilizado foi importado da biblioteca “*xgboost*”. Para o método de *Kriging*, foram testados dois códigos diferentes: um disponível na biblioteca “*Scikit Learn*” e outro foi uma versão do método implementada por CARPIO (2019), adaptada para a linguagem Python. Os parâmetros utilizados em cada método de modelagem são apresentados nos apêndices do trabalho.

A acurácia dos meta-modelos foi avaliada pelas métricas de R^2 e MSE comparativamente entre os meta-modelos que previam a mesma variável dependente. Ao final, os dois meta-modelos que apresentaram melhor acurácia para a maioria das variáveis foram escolhidos para serem usados na etapa de otimização.

3.4 PRÉ-AVALIAÇÃO DO PROBLEMA

Buscando avaliar se o problema de otimização apresentava múltiplos ótimos locais, foi realizado um algoritmo em Python para gerar os gráficos de relevo da análise exploratória. A análise consistiu em gerar três gráficos de relevo para cada variável de função objetivo, utilizando combinações de duas variáveis independentes nos eixos bidimensionais. O problema de otimização consistiu em minimizar a função objetivo definida na Equação 24.

$$obj_{i,j} = -(L_{i,j} - P_j) \quad (24)$$

Em que:

- $obj_{i,j}$ – Função objetivo no cenário econômico i na condição de operação j ;
- $L_{i,j}$ – Lucro referente ao cenário econômico i na condição de operação j ;
- P_j – Penalidade aplicada quando a condição de operação j não atende às especificações em uma das variáveis de especificação. Calculada a partir da Equação 25.

$$P_j = \sum_k F_k \max(0, Desv_{k,j}) \quad (25)$$

Em que:

- F_k – Constante para adequar o valor da penalidade (ajustado por tentativas);
- $Desv_{k,j}$ – Desvio da variável de especificação k com relação à sua especificação na condição de operação j . Calculada pela Equação 26.

$$Desv_{k,j} = \frac{V_{k,j} - V_{k,spec}}{V_{k,spec}} \quad (26)$$

Sendo:

- $V_{k,j}$ – Valor da variável k na da condição de operação j ;
- $V_{k,spec}$ – Valor de especificação da variável k , retirado da Tabela 3.

Dessa forma, se o valor da variável k for inferior ao valor da especificação, o desvio resulta em um valor negativo e, pelo teste lógico $\max(0, Desv_{k,j})$, tal variável não contribuirá para o aumento da penalidade aplicada ao lucro. No entanto, se esse desvio for positivo, essa variável irá contribuir para a penalidade aplicada e, portanto, aumentará o valor da função objetivo, dificultando a convergência do método para a observação que não se adequa ao valor especificado.

Foi observado que o valor do desvio calculado na Equação 26 poderia enviesar o resultado da otimização. Um valor excessivo de desvio para um valor da variável dependente próximo, porém acima do valor da especificação, atrapalharia a convergência da otimização para o resultado ótimo, uma vez que a função objetivo se afastaria demais do limite da especificação. Enquanto isso, se o impacto desse desvio for insignificativo na função objetivo, a tendência da otimização é de convergir para um ótimo fora de especificação. Dessa maneira, o fator F_k na Equação 25 foi proposto e ajustado por tentativa e erro durante o desenvolvimento do algoritmo de otimização, de modo que o impacto do desvio de cada variável na função objetivo fosse apropriado.

Para realizar a análise prévia, foi proposto um algoritmo para gerar 30 valores igualmente divididos ao longo da região de busca para duas variáveis independentes, mantendo a terceira variável constante em um valor médio, e calcular a função objetivo para cada ponto nesse espaço amostral, composto por 900 pontos ao todo. Considerando três cenários econômicos e três variáveis independentes, foram gerados nove gráficos de relevo para essa análise.

Caso em algum deles fosse observado duas ou mais regiões de mínimo local na função objetivo, seria necessário o uso de métodos de otimização global. No entanto, como será abordado no Capítulo 4, não foi o caso, e a otimização será então realizada com métodos de otimização local, o que justifica a metodologia que será abordada na Seção 3.5.

3.5 OTIMIZAÇÃO DO PROCESSO

O problema de otimização consistiu na minimização da função objetivo descrita na Equação 24 para três cenários econômicos distintos. Foram usados três métodos de otimização acoplados a simulação direta ou a meta-modelo. Os três métodos de otimização empregados foram: Nelder-Mead, Powell e SLSQP.

Para resolução do problema de otimização foi utilizada a sub-rotina “optimize.minimize” da biblioteca “*Scipy*”, disponível em Python. Nesta sub-rotina, é passada uma função, um método de otimização, um ponto inicial, a tolerância para atingir convergência e os limites da otimização. Com esses dados, a sub-rotina é capaz de realizar o processo iterativo de busca por um mínimo local da função, utilizando o método informado, partindo do ponto inicial e respeitando os limites da região de busca definidos.

As três funções objetivo usadas, referentes aos três cenários econômicos, são as descritas pela Equação 24. Foi adotado um ponto inicial referente ao ponto médio da região de busca, usando os valores do cenário base das variáveis independentes mostrados na Tabela 3. Os limites da otimização foram definidos após a etapa de análise de variáveis. A tolerância adotada para convergência dos métodos foi de 10^{-6} .

As funções passadas recebem as variáveis independentes e retornam com o resultado da função objetivo. Para otimizações assistidas por meta-modelo, os meta-modelos eram treinados uma única vez e os valores das variáveis dependentes eram obtidos através do teste daquele meta-modelo para as variáveis independentes informadas. Para otimizações que faziam uso da simulação direta, estabelecia-se uma interface Hysys-Python e as variáveis independentes eram inseridas diretamente na simulação, enquanto os resultados das variáveis dependentes eram coletados. Após se obter os valores dessas variáveis, a função objetivo era calculada e retornada para o otimizador.

Uma vez que o método tivesse convergido para um ótimo local, o terminal do programa retornava com o valor de ótimo alcançado, as condições operacionais que levaram a esse ponto e o tempo que cada otimização levou. Essas informações foram coletadas em uma tabela Excel para análise dos resultados.

4. RESULTADOS E DISCUSSÃO

4.1 ANÁLISE DE VARIÁVEIS

Durante a análise da correlação entre as variáveis dependentes e independentes, foram gerados os gráficos apresentados nas Figuras 13 a 19. No eixo vertical do gráfico estão representadas as variáveis dependentes do problema em suas respectivas unidades, enquanto o eixo horizontal representa os valores das variáveis independentes. Cada figura possui três gráficos empilhados, referentes a cada variável independente presente no eixo horizontal. O primeiro gráfico em azul é referente ao comportamento da variável com relação à variável razão de refluxo da coluna T-02 (RR), enquanto o segundo gráfico em vermelho é com relação à variável temperatura de fundo da T-01 (T_1) e o terceiro gráfico em verde demonstra o comportamento com relação à variável temperatura de fundo da T-02 (T_2).

Figura 13 - Correlação da variável C2 em GLP com as variáveis RR, T_1 e T_2

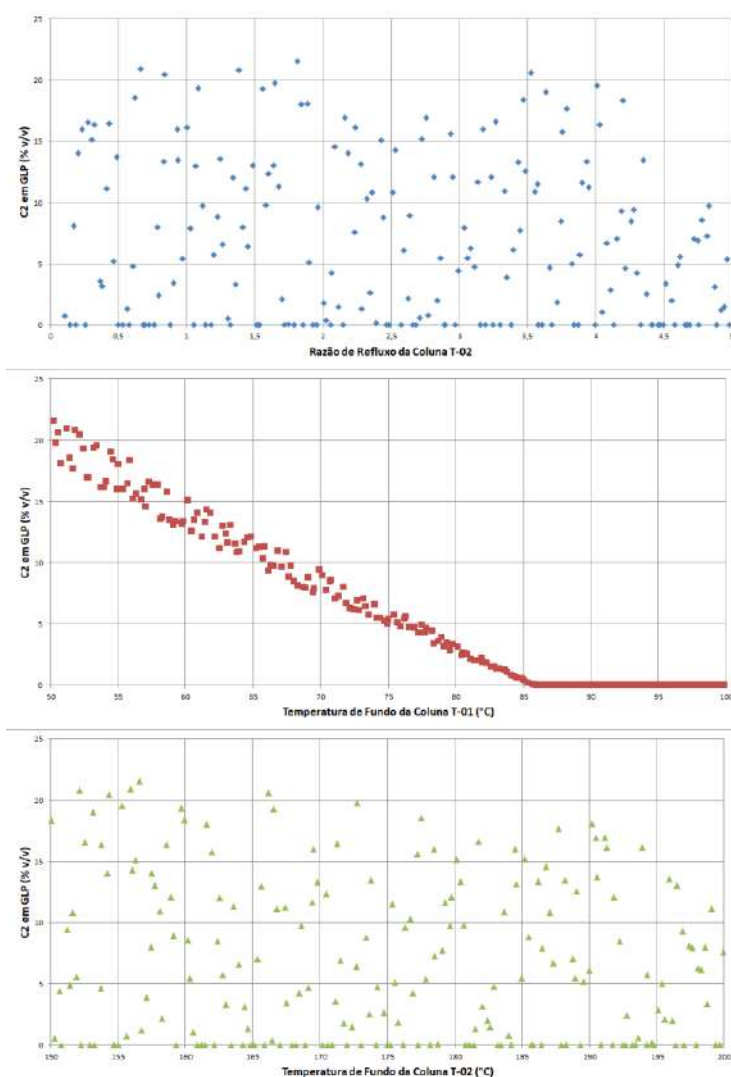


Figura 14 - Correlação da variável C5 em GLP com as variáveis RR, T_1 e T_2

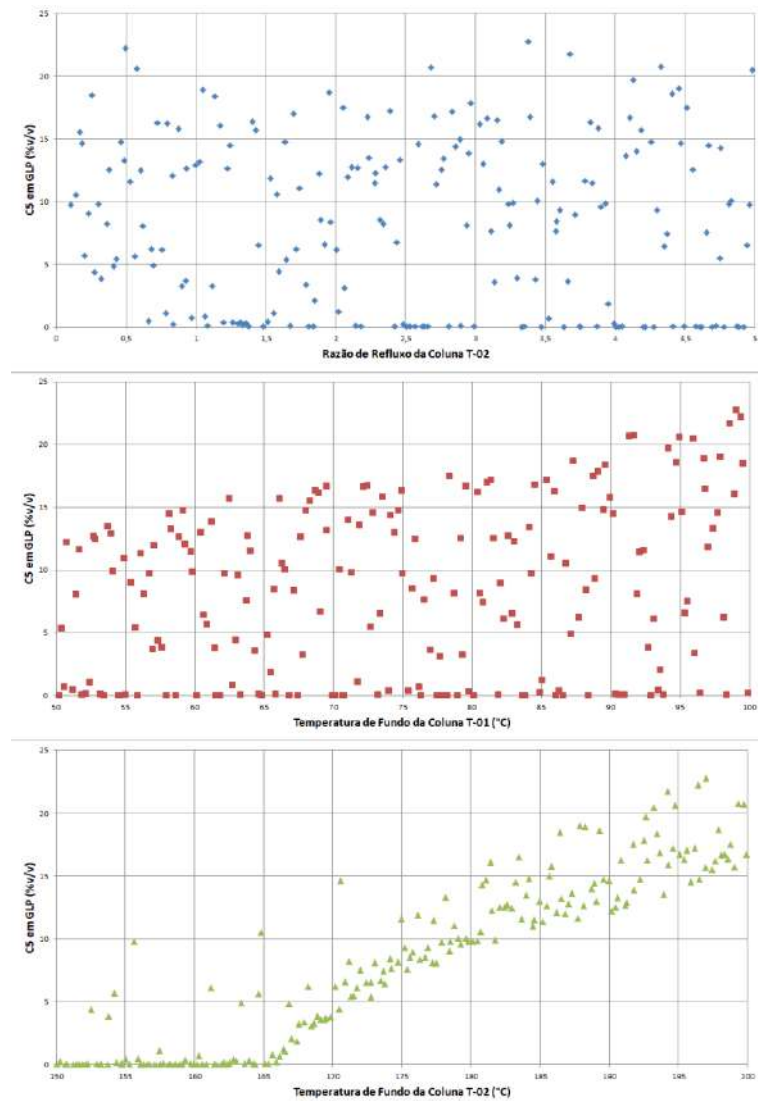


Figura 15 - Correlação da variável C2/C5 em GLP com as variáveis RR, T₁ e T₂

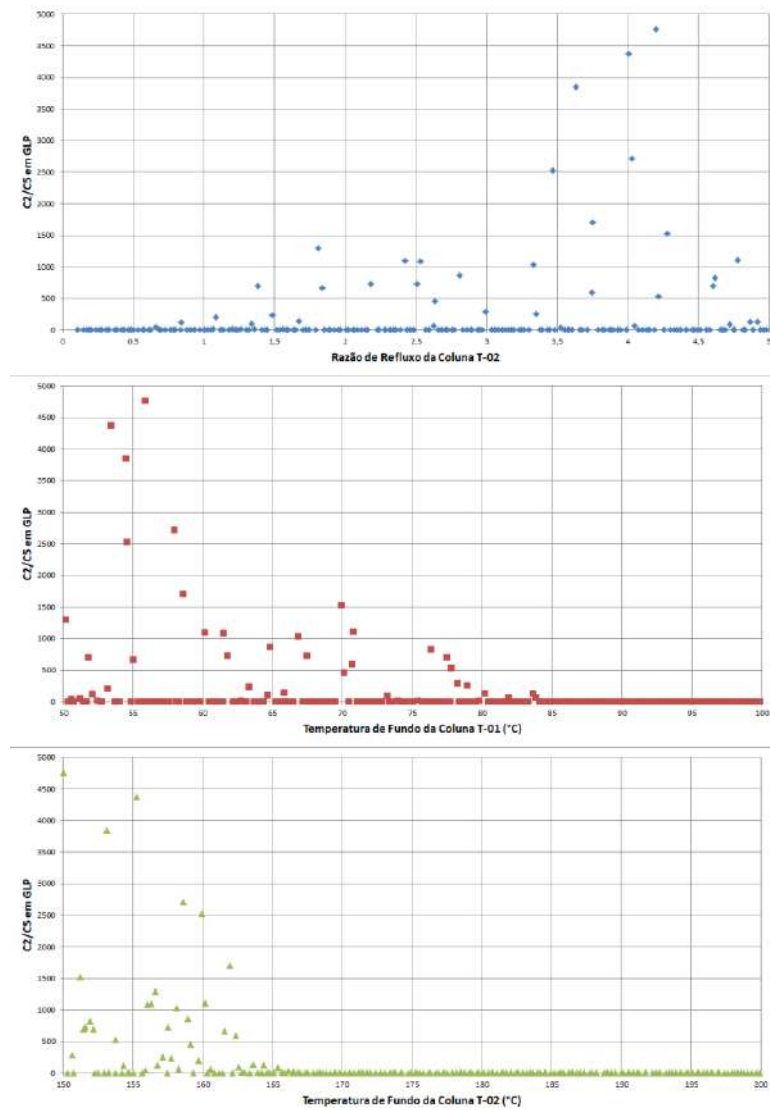
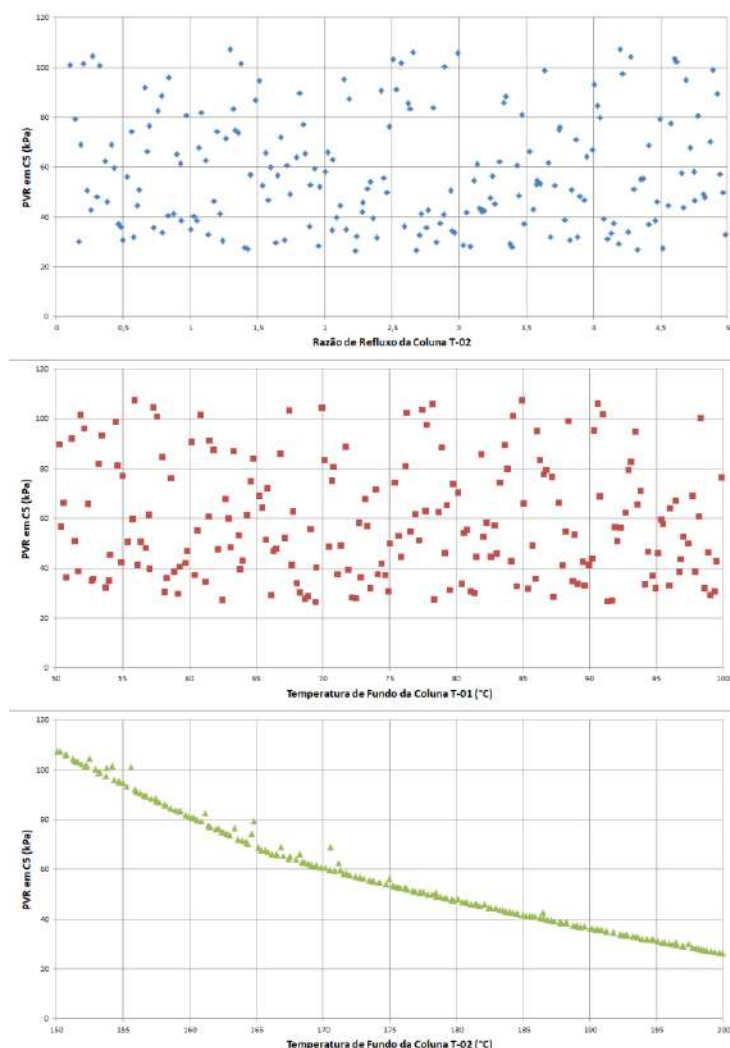


Figura 16 - Correlação da variável PVR em C5+ com as variáveis RR, T_1 e T_2



Pela análise da Figura 13, pode-se perceber que o teor de C2 em GLP possui uma correlação inversa à variável T_1 , portanto um aumento em T_1 provoca uma redução na variável dependente. Esse comportamento é de se esperar para essa variável, uma vez que aumentar a temperatura de fundo da coluna T-01 seria responsável por aumentar o teor de etano na corrente de topo, reduzindo a quantidade disponível na T-02 e, assim, na corrente de GLP. Como é desejável que esse teor não ultrapasse 14% v/v, a fronteira da região de busca deve ter um limite inferior para T_1 de modo que evite a presença de valores fora de especificação.

Observando a Figura 14, fica perceptível que a variável teor de C5 em GLP aumenta para uma elevação da variável independente T_2 . O que também é uma correlação esperada, uma vez que o aumento da temperatura de fundo da T-02 tende a contribuir para o aparecimento de componentes menos voláteis, como o pentano, na corrente de topo da coluna.

Considerando que o teor de C5 em GLP deve ser menor que 2% v/v, foi definido um limite superior para T_2 respeitando essa especificação da variável dependente.

Pela análise da Figura 15, percebe-se que a razão C2/C5 apresenta uma maior frequência de valores atípicos quanto menor for T_1 e T_2 . Como mostrado na Tabela 3, essa variável dependente deve conter valor inferior a 16. No entanto, devido à aleatoriedade de posição desses pontos, não foi possível estabelecer um limite claro para filtrar essas observações da região de busca.

Já na Figura 16, observa-se que a PVR da corrente de $C5^+$ aumenta quanto menor for o valor de T_2 . É interessante notar que um aumento na temperatura de fundo da T-02 levaria à volatilização dos componentes da corrente de fundo, começando pelos componentes de maior valor de PVR. Assim, faz sentido que o aumento de T_2 contribua para a redução da PVR total da corrente de $C5^+$. Portanto, estabeleceu-se um limite inferior para T_2 de modo a impedir a presença de valores de PVR acima de 85 kPa, conforme especificado para essa corrente.

Figura 17 - Correlação da variável lucro no cenário 1 com as variáveis RR, T1 e T2

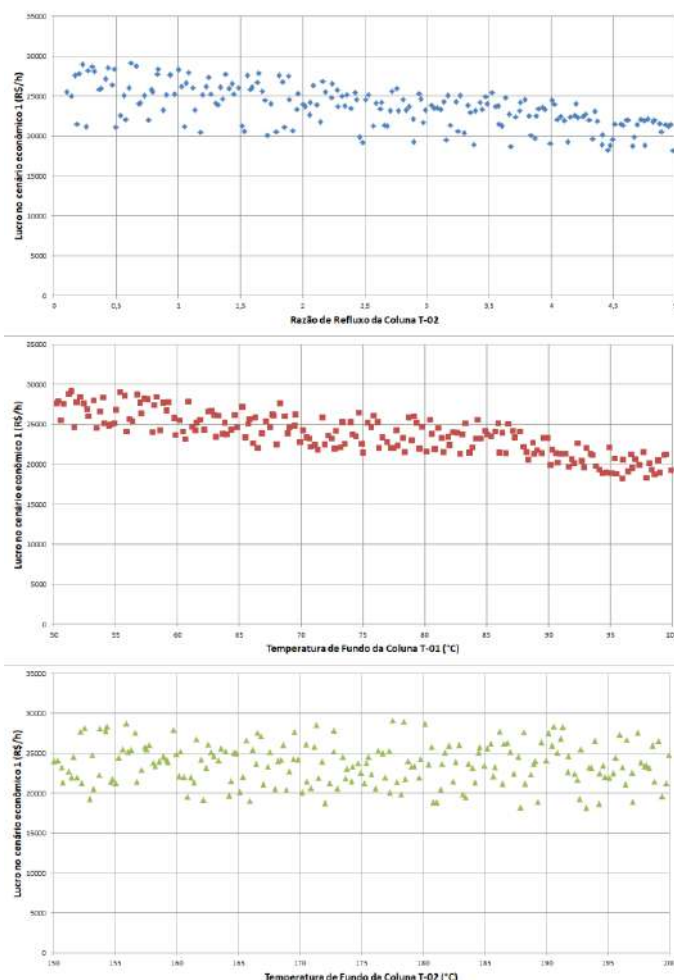


Figura 18 - Correlação da variável lucro no cenário 2 com as variáveis RR, T1 e T2

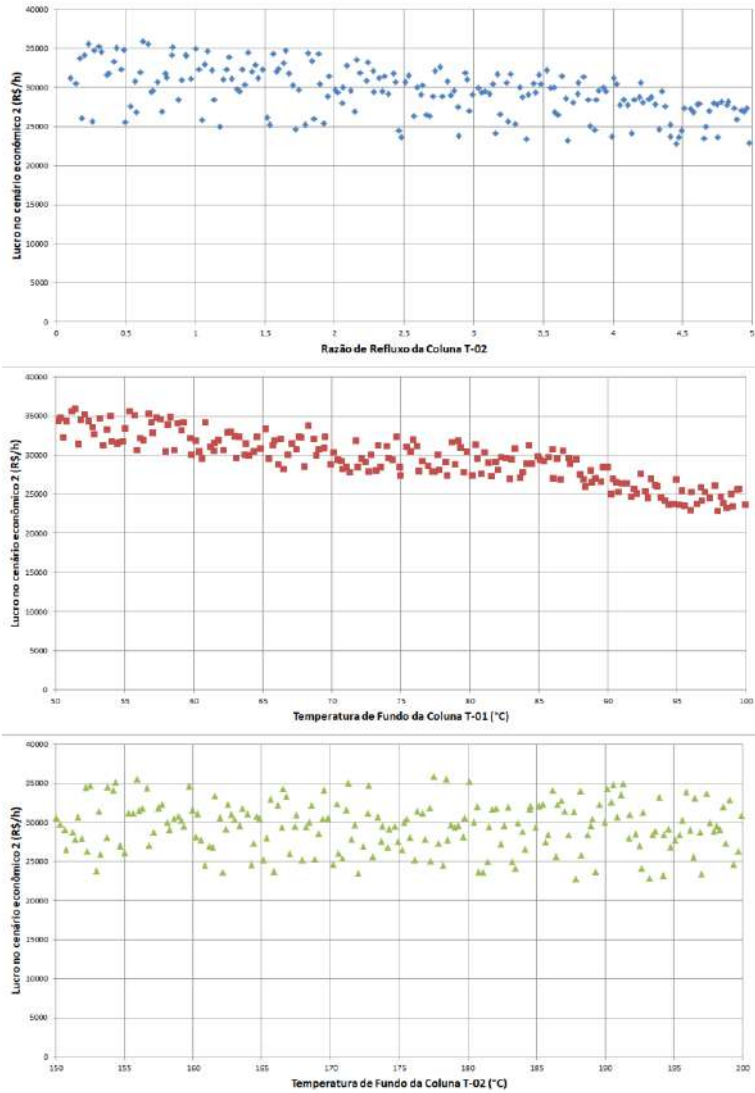
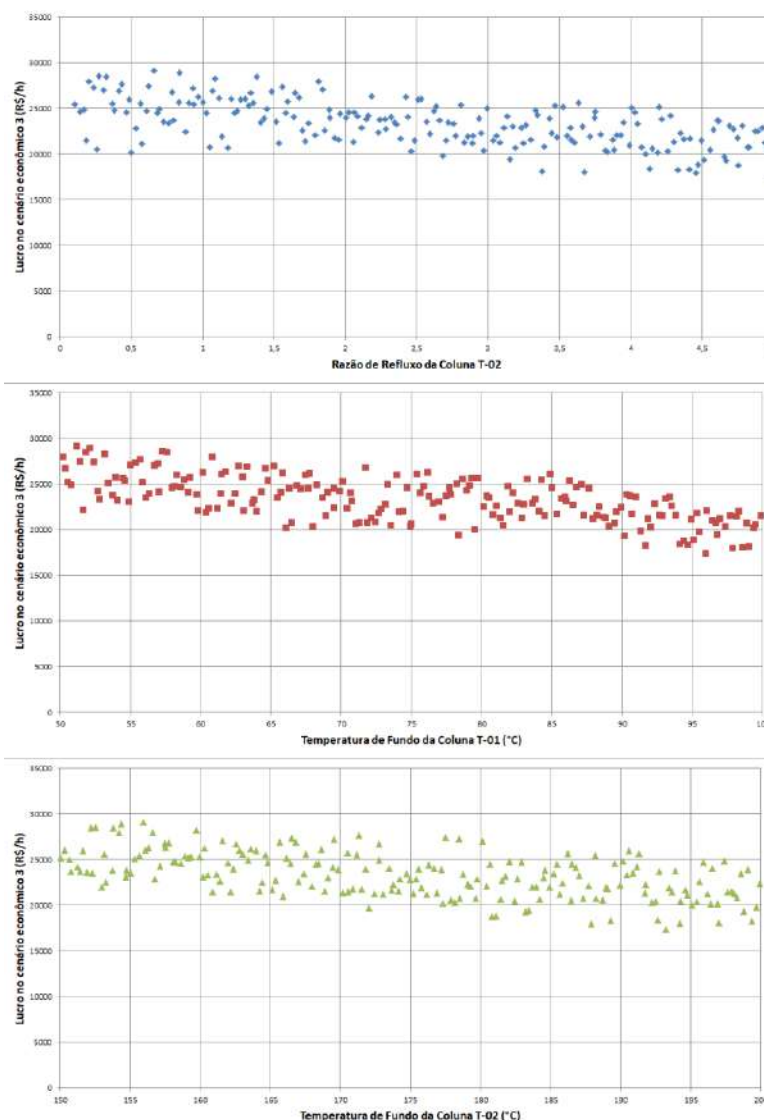


Figura 19 - Correlação da variável lucro no cenário 3 com as variáveis RR, T1 e T2



Analisando os gráficos das Figuras 17, 18 e 19, infere-se que haveria maior possibilidade de encontrar valores maiores de lucro para os três cenários em valores menores de T_1 e RR. Além disso, a Figura 19 também sugeriria valores maiores de lucro no terceiro cenário econômico para menores valores de T_2 . Visto que a região de busca estabelecida deve primeiramente atender às especificações, essa análise não foi tomada em conta quando foram definidos os limites da otimização. No entanto, é importante analisar os gráficos de correlação das variáveis de lucro, pois estes podem dar indícios do que esperar para os resultados da otimização.

Após análise dos gráficos, os limites superiores e inferiores foram definidos para as variáveis independentes do problema e estão representados na Tabela 6. Tais limites serão

utilizados para construção dos meta-modelos e como fronteira da região de busca da otimização.

Tabela 6 - Limites definidos da região de busca

Variável independente	Limite inferior	Limite superior
T1	69°C	79°C
T2	160°C	170°C
RR	0,1	2,0

4.2 ELABORAÇÃO DE META-MODELOS

A partir da definição dos limites das variáveis independentes, dados da simulação foram coletados e usados no treinamento de meta-modelos a partir dos métodos de RLM, *Random Forest*, *XGBoost*, *Gradient Boosting* e *Kriging*. Os parâmetros utilizados são expostos no capítulo de apêndices do trabalho. Os meta-modelos treinados foram testados pelas métricas de R^2 e MSE e os resultados encontram-se na Tabela 7.

Com base na Tabela 7, é possível afirmar que o meta-modelo que apresentou melhor desempenho foi construído pelo método de *Kriging*, uma vez que este alcançou valores de R^2 superiores e MSE inferiores aos demais meta-modelos para todas as variáveis consideradas. O segundo meta-modelo de melhor desempenho foi construído pelo método de *Gradient Boosting*, que obteve valores de R^2 relativamente altos para as variáveis de lucro, mantendo R^2 e MSE relativamente satisfatórios para as demais variáveis de especificação.

O alto desempenho do meta-modelo *Kriging* para previsão de dados da simulação é esperado, conforme reportado por BOUKOUVALA & IERAPETRITOU (2012). O desempenho do *Kriging* para esses dados está relacionado com o fato do método fazer o mínimo de suposições da forma da função que ele objetiva representar, trazendo assim ótimo ajuste para dados não lineares como os originados pelo simulador, conforme citado por FORRESTER & KEANE (2009). Por isso, é de se esperar que na otimização o meta-modelo *Kriging* se sobressaia na busca pelo lucro ótimo da planta.

Vale observar que os meta-modelos gerados por *Random Forest* e *XGBoosting* obtiveram resultados bem próximos ao meta-modelo gerado por *Gradient Boosting*, o que está de acordo com o esperado, uma vez que todos esses métodos são baseados em técnicas de aprimoramento do modelo base de árvore de decisão.

Tabela 7 - Desempenho dos meta-modelos analisados

Modelo	Variável	R2	MSE
Regressão Linear Múltipla	C2 de topo	99,70%	8,0E-07
	C5 de topo	72,67%	1,9E-04
	C2/C5 de topo	48,17%	4,7E-02
	PVR de fundo	95,48%	2,1E-04
	Lucro 1	99,50%	3,0E-06
	Lucro 2	99,55%	2,0E-06
	Lucro 3	86,82%	3,8E-05
Random Forest	C2 de topo	99,25%	1,9E-06
	C5 de topo	98,54%	1,0E-05
	C2/C5 de topo	83,10%	1,6E-02
	PVR de fundo	97,74%	9,8E-05
	Lucro 1	98,37%	9,3E-06
	Lucro 2	98,14%	8,3E-06
	Lucro 3	96,20%	1,2E-05
XGBoosting	C2 de topo	99,35%	1,7E-06
	C5 de topo	98,61%	9,6E-06
	C2/C5 de topo	77,88%	2,0E-02
	PVR de fundo	97,96%	9,4E-05
	Lucro 1	98,10%	1,1E-05
	Lucro 2	97,98%	8,8E-06
	Lucro 3	95,17%	1,4E-05
<i>Gradient Boosting</i>	C2 de topo	99,61%	1,0E-06
	C5 de topo	99,15%	5,9E-06
	C2/C5 de topo	81,24%	1,3E-02
	PVR de fundo	99,01%	4,5E-05
	Lucro 1	98,51%	8,8E-06
	Lucro 2	98,60%	6,1E-06
	Lucro 3	96,92%	8,9E-06
Kriging (Scikit Learn)	C2 de topo	99,14%	2,2E-06
	C5 de topo	99,70%	2,1E-06
	C2/C5 de topo	99,35%	5,9E-04
	PVR de fundo	99,55%	2,1E-05
	Lucro 1	98,90%	6,5E-06
	Lucro 2	98,85%	5,0E-06
	Lucro 3	98,26%	5,0E-06
Kriging (CARPIO, 2019)	C2 de topo	99,96%	1,2E-07
	C5 de topo	99,89%	7,8E-07
	C2/C5 de topo	99,26%	6,8E-04
	PVR de fundo	99,97%	1,2E-06
	Lucro 1	99,97%	1,5E-07
	Lucro 2	99,98%	1,0E-07
	Lucro 3	99,81%	5,5E-07

Quando comparado aos demais meta-modelos, o meta-modelo de RLM apresentou resultados inferiores de R^2 para as variáveis dependentes referentes ao teor de C5 em GLP, razão de C2/C5 na corrente de GLP e lucro no terceiro cenário. Com base no que foi discutido na Seção 2.3, era esperado que o modelo linear apresentasse baixa acurácia para variáveis de comportamento não linear. Portanto, supõe-se que essas três variáveis possuem um comportamento não linear, o que justificaria os valores observados de R^2 serem menores para essas variáveis em comparação a outras em todos os meta-modelos estudados.

4.3 PRÉ-AVALIAÇÃO DO PROBLEMA

Os gráficos de relevo das Figuras 20, 21 e 22 foram gerados com o objetivo de conferir a presença de múltiplos ótimos locais na região de busca da otimização para quaisquer dimensões do problema. Baseado nos gráficos apresentados, não foram identificados múltiplos ótimos locais para o problema de otimização. Dessa forma, é de se esperar que a otimização possa ser solucionada por meio de métodos de otimização local.

Figura 20 - Relevo da função objetivo no cenário 1

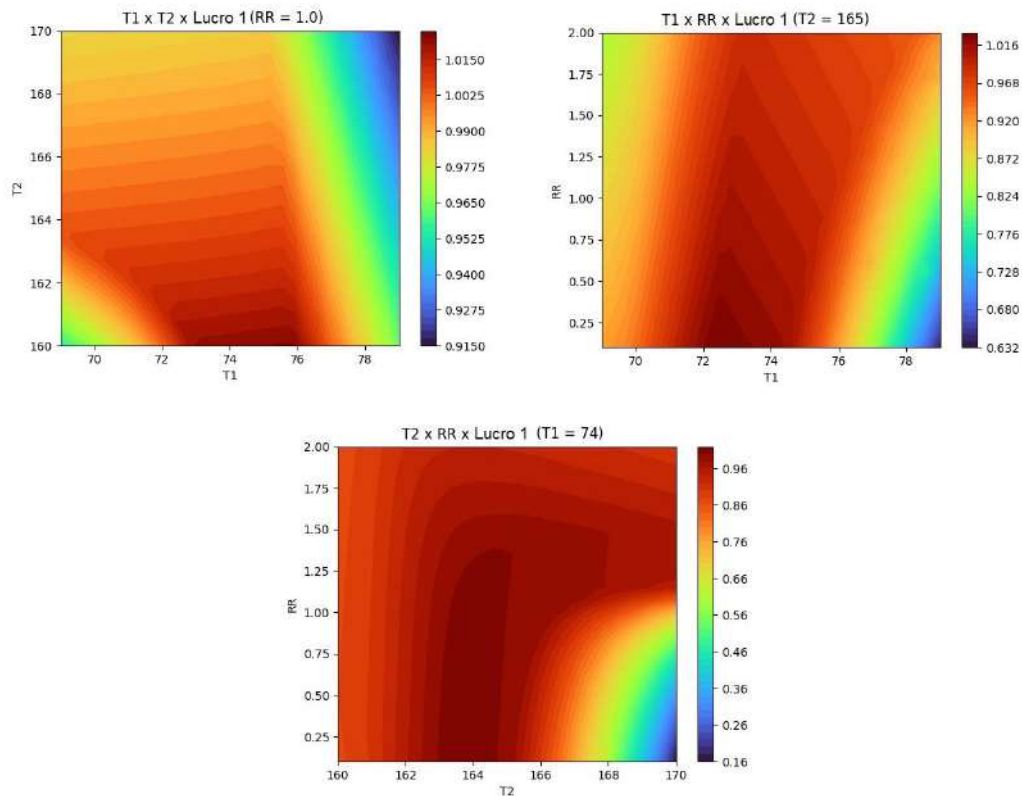


Figura 21 - Relevo da função objetivo no cenário 2

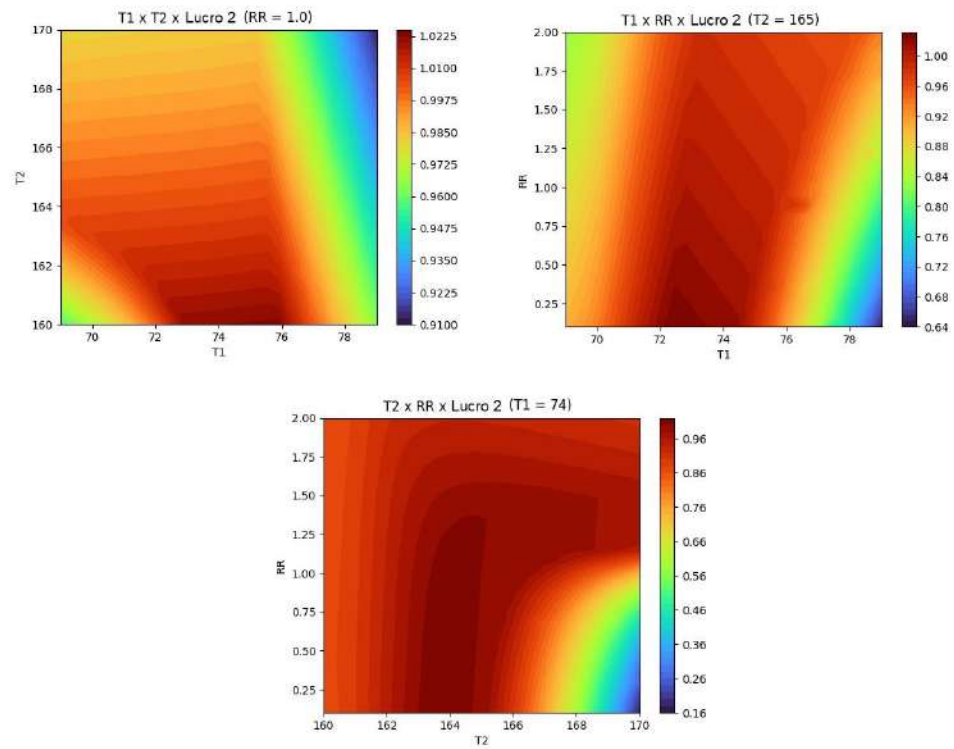
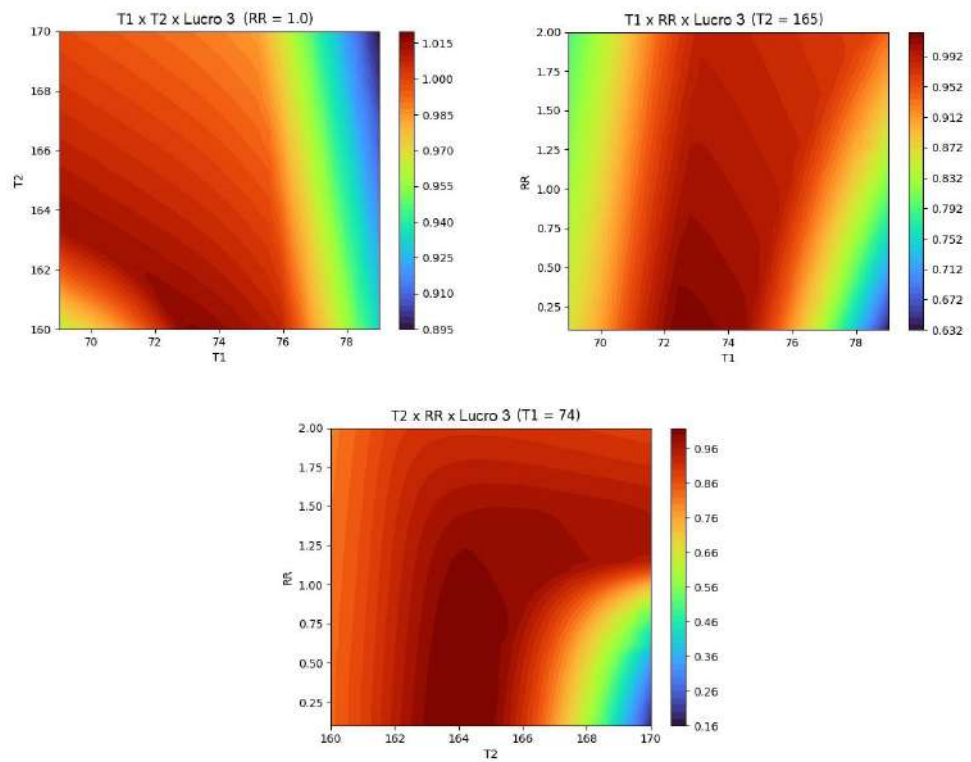


Figura 22 - Relevo da função objetivo no cenário 3



4.4 OTIMIZAÇÃO DA PLANTA

A otimização da planta foi realizada usando algoritmo em Python, com base na minimização da função objetivo da Equação 24 para três cenários econômicos, usando três métodos de otimização, conforme descrito na Seção 3.5. Foi adotada uma tolerância de 10^{-6} na otimização para todos os métodos, com exceção ao Nelder-Mead acoplado na simulação, que foi incapaz de convergir para tal tolerância dentro de um limite de 10.000 iterações. Portanto, para o Nelder-Mead acoplado a simulador, a tolerância foi reduzida para 10^{-4} , de modo a atingir a convergência em uma quantidade viável de iterações.

Os resultados para as 27 otimizações realizadas estão representados nas Tabelas 8, 10 e 12, contendo: o lucro ótimo previsto por cada método (melhor função objetivo encontrada), o lucro ótimo real (resultado do simulador nas condições operacionais ótimas), o erro atribuído à previsão dos meta-modelos, o indicativo de se o cenário ótimo encontrado atende às restrições da Tabela 3 e o ganho percentual com relação ao cenário base na Tabela 3, referente aos cenários econômicos 1, 2 e 3. Nas Tabelas 9, 11 e 13 foram inseridos os resultados referentes à eficiência de cada método de otimização, mostrando o número de iterações realizadas, o número de avaliações da função objetivo, o tempo de otimização e a tolerância adotada por cada método, para os três cenários econômicos.

4.4.1 Resultados da otimização cenário 1

Tabela 8 - Resultados da otimização para o cenário econômico 1

Método aplicado		Lucro previsto (R\$/h)	Lucro real (R\$/h)	Erro (R\$/h)	Espec.	Ganho percentual (%)
Cenário 1	Nelder-Mead com Gradient Boost	25774,02	25727,42	-46,60	OK	1,2
	Nelder-Mead com Kriging	26309,34	26296,51	-12,83	OK	3,4
	Nelder-Mead com Simulação	26159,78	26159,78	-	OK	2,9
	Powell com Gradient Boost	26129,53	26073,08	-56,45	OK	2,5
	Powell com Kriging	25977,08	25982,63	5,55	OK	2,2
	Powell com Simulação	26262,95	26262,95	-	OK	3,3
	SLSQP com Gradient Boost	26193,16	26170,24	-22,92	OK	2,9
	SLSQP com Kriging	26292,76	26291,60	-1,16	OK	3,4
	SLSQP com Simulação	26246,40	26246,40	-	OK	3,2

Tabela 9 – Resultados de desempenho da otimização para o cenário econômico 1

Método aplicado		Número de iterações	Avaliações da função objetivo	Tempo de opt (s)	Tolerância
Cenário 1	Nelder-Mead com Gradient Boost	31	103	0,3	1e-6
	Nelder-Mead com Kriging	196	348	17,7	1e-6
	Nelder-Mead com Simulação*	126	235	12,7	1e-4
	Powell com Gradient Boost	3	247	0,7	1e-6
	Powell com Kriging	2	202	10,2	1e-6
	Powell com Simulação	7	708	45,5	1e-6
	SLSQP com Gradient Boost	8	69	0,2	1e-6
	SLSQP com Kriging	8	68	3,4	1e-6
	SLSQP com Simulação	9	78	5,3	1e-6

*O método de Nelder-Mead com simulação não convergiu em até 10000 iterações para uma tolerância de 1e-6. Portanto, optou-se por reduzir a tolerância desse método para 1e-4.

Analisando a Tabela 8, referente aos resultados da otimização no primeiro cenário econômico, percebe-se que o método que alcançou melhor lucro da planta foi o Nelder-Mead assistido por meta-modelo *Kriging*, com ganho percentual de 3,4% com relação ao cenário-base. O método utilizando SLSQP com *Kriging*, por sua vez, apresentou resultados próximos ao melhor observado, por uma diferença de 5,37 R\$/h. Também foi observado que os métodos de otimização assistidos por meta-modelo *Gradient Boosting* apresentaram maior erro e menores ganhos, o que era de se esperar uma vez que o meta-modelo treinado por *Gradient Boosting* mostrou uma acurácia inferior ao meta-modelo treinado por *Kriging* na Seção 4.2 de resultados. Para todas as otimizações do primeiro cenário, foi confirmado que as especificações das correntes de produto da planta, mostradas na Tabela 3, foram respeitadas.

Pela análise da Tabela 9, pode-se verificar que o método de Nelder-Mead necessitou uma maior quantidade de iterações para atingir a convergência, enquanto o método de Powell fez uso de um maior número de avaliações da função objetivo. O método SLSQP se mostrou ser o mais eficiente durante a otimização do cenário 2. Tal resultado era esperado, levando em consideração de que os métodos diretos de otimização (Nelder-Mead e Powell) não utilizam a informação da derivada da função objetivo diferentemente do método indireto (SLSQP), consequentemente levando a uma menor eficiência para alcançar o valor de ótimo.

Ainda se baseando na Tabela 9, pode-se perceber também que os métodos assistidos por *Kriging* convergiram em menor tempo e com maior eficiência do que seus respectivos acoplados à simulação. Portanto, para esse cenário, o meta-modelo *Kriging* trouxe uma

otimização mais eficiente do que o método acoplado ao simulador. Os métodos que usaram *Gradient Boosting* foram os que convergiram mais rapidamente, porém como os resultados desses métodos foram inferiores aos demais, não é possível afirmar que a rápida convergência é um indicativo de eficiência do método.

4.4.2 Resultados da otimização cenário 2

Tabela 10 - Resultados da otimização para o cenário econômico 2

Método aplicado		Lucro previsto (R\$/h)	Lucro real (R\$/h)	Erro (R\$/h)	Espec.	Ganho percentual (%)
Cenário 2	Nelder-Mead com Gradient Boost	31813,82	31696,17	-117,65	OK	1,1
	Nelder-Mead com Kriging	32384,51	32373,15	-11,37	OK	3,2
	Nelder-Mead com Simulação	32369,46	32369,46	-	OK	3,2
	Powell com Gradient Boost	32178,95	32142,11	-36,84	OK	2,5
	Powell com Kriging	32043,74	32043,94	0,20	OK	2,2
	Powell com Simulação	32320,23	32320,23	-	OK	3,1
	SLSQP com Gradient Boost	31964,12	31865,21	-98,91	OK	1,6
	SLSQP com Kriging	32352,18	32355,22	3,04	OK	3,2
	SLSQP com Simulação	32306,85	32306,85	-	OK	3,0

Tabela 11- Resultados de desempenho da otimização para o cenário econômico 2

Método aplicado		Número de iterações	Avaliações da função objetivo	Tempo de opt (s)	Tolerância
Cenário 2	Nelder-Mead com Gradient Boost	29	105	0,3	1e-6
	Nelder-Mead com Kriging	291	522	26,4	1e-6
	Nelder-Mead com Simulação*	66	139	7,7	1e-4
	Powell com Gradient Boost	2	198	0,5	1e-6
	Powell com Kriging	2	244	12,4	1e-6
	Powell com Simulação	7	770	46,1	1e-6
	SLSQP com Gradient Boost	5	48	0,1	1e-6
	SLSQP com Kriging	11	98	5,0	1e-6
	SLSQP com Simulação	8	78	5,7	1e-6

*O método de Nelder-Mead com simulação não convergiu em até 10000 iterações para uma tolerância de 1e-6. Portanto, optou-se por reduzir a tolerância desse método para 1e-4.

Analisando a Tabela 10, referente aos resultados da otimização no segundo cenário econômico, percebe-se que o método que alcançou a melhor função objetivo foi novamente o Nelder-Mead utilizando meta-modelo *Kriging*, com ganho percentual de 3,2% com relação ao cenário base. O método utilizando SLSQP com *Kriging* também apresentou resultados próximos ao melhor observado, com uma diferença de 17,93 R\$/h. Vale ressaltar que, assim como no primeiro cenário, os métodos assistidos por *Gradient Boosting* mostraram maior erro e menores ganhos. Novamente, as especificações de corrente foram respeitadas para todas as otimizações do segundo cenário.

Como no cenário anterior, a Tabela 11 indica uma tendência dos métodos baseados em Nelder-Mead em utilizar mais iterações para atingir a convergência e os métodos baseados em Powell utilizarem maior número de avaliações da função objetivo. Portanto, novamente os métodos que utilizaram SLSQP se mostraram mais eficientes.

Além disso, para o segundo cenário econômico, os métodos assistidos por meta-modelo *Kriging* convergiram em menor tempo do que seus respectivos acoplados na simulação. No entanto, o método de SLSQP assistido por *Kriging* utilizou maior número de iterações e avaliações da função objetivo do que o SLSQP acoplado à simulação. Ainda sim, uma vez que o tempo de otimização foi inferior em todos os casos, pode-se afirmar que o emprego do meta-modelo *Kriging* trouxe ganhos de eficiência na otimização.

4.4.3 Resultados da otimização cenário 3

Tabela 12 - Resultados da otimização para o cenário econômico 3

Método aplicado		Lucro previsto (R\$/h)	Lucro real (R\$/h)	Erro (R\$/h)	Espec.	Ganho percentual (%)
Cenário 3	Nelder-Mead com Gradient Boost	26282,12	26081,01	- 201,12	OK	-0,1
	Nelder-Mead com Kriging	26907,00	26941,50	34,50	OK	3,2
	Nelder-Mead com Simulação	26960,74	26960,74	-	OK	3,3
	Powell com Gradient Boost	26581,95	26709,34	127,39	OK	2,3
	Powell com Kriging	26911,50	26942,40	30,90	OK	3,2
	Powell com Simulação	26948,23	26948,23	-	OK	3,3
	SLSQP com Gradient Boost	26384,54	26326,51	-58,03	OK	0,9
	SLSQP com Kriging	26906,44	26920,07	13,62	OK	3,2
	SLSQP com Simulação	26950,94	26950,94	-	OK	3,3

Tabela 13 - Resultados de desempenho da otimização para o cenário econômico 3

Método aplicado		Número de iterações	Avaliações da função objetivo	Tempo de opt (s)	Tolerância
Cenário 3	Nelder-Mead com Gradient Boost	24	89	0,2	1e-6
	Nelder-Mead com Kriging	131	249	12,6	1e-6
	Nelder-Mead com Simulação*	68	124	7,8	1e-4
	Powell com Gradient Boost	4	355	0,9	1e-6
	Powell com Kriging	4	674	34,0	1e-6
	Powell com Simulação	4	470	29,2	1e-6
	SLSQP com Gradient Boost	3	32	0,1	1e-6
	SLSQP com Kriging	37	184	9,3	1e-6
	SLSQP com Simulação	12	108	7,4	1e-6

*O método de Nelder-Mead com simulação não convergiu em até 10000 iterações para uma tolerância de 1e-6. Portanto, optou-se por reduzir a tolerância desse método para 1e-4.

Analizando a Tabela 12, referente aos resultados da otimização no terceiro cenário econômico, percebe-se que o método que alcançou a melhor função objetivo foi o Nelder-Mead acoplado à simulação, com ganho percentual de 3,3% com relação ao cenário-base. Os métodos assistidos por *Gradient Boosting* mostraram resultados inferiores também nesse cenário e todas as especificações de qualidade nas correntes de produto da planta foram respeitadas. É interessante notar que exclusivamente no terceiro cenário os métodos acoplados à simulação se sobressaíram aos métodos assistidos por meta-modelo *Kriging*. Isso pode ser justificado pela acurácia do meta-modelo *Kriging* para a variável de lucro do cenário 3 ser inferior à acurácia do meta-modelo para o lucro dos cenários 1 e 2, conforme pode ser visto na Tabela 7. O que está de acordo com a discussão levantada durante a análise do meta-modelo de RLM, aonde se chegou à conclusão de que o lucro para o terceiro cenário é uma variável de elevada variância, o que afeta na acurácia dos meta-modelos, repercutindo nos resultados da otimização. Ainda assim, os resultados dos métodos assistidos por *Kriging* foram próximos aos da simulação, com diferença no ganho percentual de apenas 0,1%, indicando que a acurácia do meta-modelo foi elevada, porém não o suficiente para sobressair à simulação no terceiro cenário.

A tendência de que o método de Nelder-Mead utiliza mais iterações para atingir a convergência e o método de Powell requer um maior número de avaliações da função objetivo para convergir se confirma novamente pelos resultados da Tabela 13. Novamente o método SLSQP se mostrou mais eficiente do que os demais. E os métodos assistidos por *Gradient*

Boosting também convergiram mais rapidamente para o terceiro cenário, porém como os resultados desses métodos foram inferiores aos demais, não se pode afirmar que esse fato se deve à eficiência do método.

Ao contrário dos cenários anteriores, a Tabela 13 indica que no terceiro cenário os métodos acoplados à simulação levaram menos tempo e foram mais eficientes do que os métodos que utilizaram meta-modelo *Kriging*, à exceção do método Nelder-Mead, que quando acoplado à simulação requer mais de 10000 iterações para convergir para a tolerância estabelecida. Esse fato corrobora com a hipótese de que a acurácia do meta-modelo treinado por *Kriging* não foi suficiente para se sobressair à otimização acoplada à simulação nesse cenário econômico.

4.4.4 Resultados gerais da otimização

Baseado nas otimizações realizadas, o método que levou às melhores condições de operação nos cenários econômicos 1 e 2 foi o Nelder-Mead assistido por meta-modelo *Kriging* e o método que obteve a melhor condição de operação no terceiro cenário foi o Nelder-Mead acoplado à simulação. No entanto, os métodos que usaram SLSQP mostraram resultados próximos aos métodos que usaram Nelder-Mead, com menor tempo de otimização e necessitando menos iterações e avaliações da função objetivo. A maior eficiência do SLSQP em relação ao Nelder-Mead e o Powell está de acordo com o esperado, uma vez que métodos indiretos utilizam a informação da derivada da função, conforme discutido na Seção 2.4.

Pode-se inferir também que os métodos assistidos por meta-modelo *Gradient Boosting* obtiveram resultados inferiores aos demais, apresentando também elevado valor de erro, o que pode ser justificado pela relativa baixa acurácia desse modelo quando comparado ao meta-modelo *Kriging*. A partir dessa comparação fica perceptível que a acurácia do meta-modelo usado é um fator decisivo para o sucesso da otimização assistida. Os métodos assistidos por meta-modelo *Kriging* se sobressaíram aos métodos acoplados ao simulador nos cenários econômicos 1 e 2, e mostraram resultados próximos àqueles obtidos pelos métodos acoplados ao simulador no cenário 3, indicando que o meta-modelo *Kriging* possui acurácia suficiente para trazer vantagens na otimização dos cenários 1 e 2 e apresenta uma boa acurácia para utilização no cenário 3.

As melhores condições operacionais encontradas para cada cenário econômico foram representadas na Tabela 14.

Tabela 14 - Condições operacionais de maximização do lucro encontradas na otimização

Cenário	Método aplicado	T1 (°C)	T2 (°C)	RR	C2 em GLP (% v/v)	C5 em GLP (% v/v)	C2/C5 em GLP	PVR em C5+ (kPa)
1	Nelder-Mead com Kriging	69,0	161,7	0,67	9,37	1,97	4,75	78,32
2	Nelder-Mead com Kriging	69,0	161,7	0,67	9,37	1,97	4,75	78,32
3	Nelder-Mead com Simulação	69,0	160,0	0,65	9,43	1,98	4,76	82,60

Considerando que o primeiro cenário é aquele em que a corrente de maior custo é a de GV, por ser a corrente de topo da T-01 seria intuitivo pensar que a condição ótima estaria a valores elevados de T_1 . No entanto, os resultados da Tabela 14 e a correlação do gráfico da Figura 17 para essa variável indicam que a condição ótima está quando a variável T_1 está no seu valor mínimo dentro da região de busca. Isso pode ser justificado pelo fato de que o teor de etano na corrente de GV está limitado pela restrição da Equação 20, implementada no simulador. De tal forma, um aumento da temperatura da coluna para acima de 69°C não contribuiria para um aumento da vazão de GV, apenas de etano em gás residual que não seria convertido por não atender à restrição de metano na corrente final. Enquanto isso, as condições de T_2 e RR foram otimizadas para maximizar o lucro da planta atendendo ainda as especificações das correntes de produtos.

O segundo cenário econômico é referente ao cenário onde o GLP é a corrente de maior preço de venda. Portanto, faz sentido que a temperatura T_1 seja mínima, uma vez que essa condição promove maior vazão na corrente de fundo de T-01, que alimenta a coluna T-02. Por outro lado, era de se esperar que a temperatura T_2 fosse elevada, visando maximizar a vazão de GLP da coluna T-02. No entanto, deve-se lembrar de que um aumento da temperatura T_2 promove um aumento da concentração de pentano em GLP. Dessa forma, é provável que essas condições ótimas de T_2 e RR no cenário 2 estavam atreladas à restrição de C5 em GLP, uma vez que a concentração desse componente em GLP estava próxima da restrição de 2,0%v/v.

O cenário 3 era referente ao cenário econômico onde a corrente de maior preço de venda é o C5⁺. Pelos mesmos motivos do segundo cenário, faz sentido que a temperatura T_1 assuma o valor mínimo possível. Também faz sentido que a temperatura T_2 assuma o valor mínimo, uma vez que tal condição maximiza a vazão da corrente de fundo. Dessa forma, a variável RR foi ajustada para maximizar a vazão de C5⁺, respeitando as restrições do problema.

5. CONCLUSÃO

O trabalho proposto objetivou a representação do comportamento de uma planta de processamento de gás natural simulada em software Hysys por meio de meta-modelos e o emprego dos melhores meta-modelos obtidos para assistir na otimização do lucro dessa planta. O meta-modelo *Kriging* desenvolvido a partir de dados do simulador foi capaz de ajustar todas as variáveis dependentes da planta, apresentando valores de R^2 acima de 99%. A aplicação do meta-modelo *Kriging* para assistência no problema de otimização permitiu encontrar condições de operação para a planta que gerassem mais lucro do que a otimização acoplada ao simulador, com ganhos de pelo menos 3% com relação ao cenário base.

Um estudo de correlação das variáveis contidas no sistema foi conduzido para compreender a relação entre as variáveis dependentes e independentes do sistema e estabelecer os limites da região de busca com base nas restrições propostas. Foi encontrado que um intervalo de busca de 0,1 a 2,0 para a razão de refluxo da coluna desbutanizadora, 69°C a 79°C para a temperatura de fundo da coluna desetanizadora e 160°C a 170°C para a temperatura de fundo da coluna desbutanizadora seria o ideal para uso como limites da otimização da planta.

Dentre os métodos de RLM, *Random Forest*, *Gradient Boosting*, *XGBoosting* e *Kriging*, o meta-modelo que obteve melhor resultado para as métricas de teste foi treinado com base no método de *Kriging*, seguido do meta-modelo treinado por *Gradient Boosting*. Portanto, esses dois meta-modelos foram escolhidos para assistir na otimização da planta.

Pela observação de nove gráficos de relevo durante a pré-avaliação do problema de otimização, foi concluído que não há a ocorrência de múltiplos ótimos locais em nenhum dos gráficos e, portanto, a resolução do problema de otimização pôde ser realizada por métodos de otimização local.

Considerando três métodos de otimização local, Nelder-Mead, Powell e SLSQP, e utilizando três métodos de obtenção da variável lucro, por meio da resposta de meta-modelo *Kriging* ou *Gradient Boosting* ou acoplado à simulação da planta, a metodologia que trouxe melhores resultados de lucro da planta para o problema proposto foi a que utilizava Nelder-Mead com meta-modelo *Kriging*. Tal metodologia se sobressaiu em dois dos três cenários econômicos propostos, com resultados próximos do ótimo no terceiro cenário econômico. No entanto, o método usando SLSQP assistido por *Kriging* mostrou obter resultados próximos daqueles obtidos por Nelder-Mead para uma maior eficiência de otimização.

As melhores condições de operação da planta encontradas para os cenários econômicos 1 e 2 foram as mesmas, com 69°C para o valor de temperatura de fundo da coluna desetanizadora, 161,7°C para a temperatura de fundo da coluna desbutanizadora e 0,67 na razão de refluxo da coluna desbutanizadora. Já no terceiro cenário econômico, a condição de operação ótima foi de 69°C na temperatura de fundo da coluna desetanizadora, 160°C na temperatura de fundo da coluna desbutanizadora e 0,65 na razão de refluxo na coluna desbutanizadora. Com essas condições, o lucro operacional teve um aumento percentual de 3,4% para o primeiro cenário, 3,2% para o segundo cenário e 3,3% para o terceiro cenário, quando comparado com as condições base propostas à planta.

Portanto, pode-se afirmar que os objetivos estabelecidos para o trabalho foram concluídos com êxito. Além disso, os conceitos adquiridos durante o desenvolvimento da metodologia e na análise dos resultados do trabalho podem servir como base para realização de outros estudos semelhantes focados no acréscimo de lucros em outros processos industriais.

6. REFERÊNCIAS

- AKBA, T.; BAKER, D.; MENGÜÇ, M. P. **Gradient-based optimization of micro-scale pressurized volumetric receiver geometry and flow rate**. Renewable energy, vol. 203, p. 741-752, 2023.
- ALI, W. **Optimal operational analysis of metamodel based single mixed refrigerant cryogenic process for floating liquefied natural gas plant technology**. Results in Engineering, v. 16, 100744, 2022.
- ANP, 2023. **Anuário Estatístico Brasileiro do petróleo, gás natural e biocombustíveis 2023**. Disponível em: <https://www.gov.br/anp/pt-br/centrais-de-conteudo/publicacoes/anuario-estatistico/anuario-estatistico-2023>. Acesso em: 27 jan. 2024.
- BAI, G.; CHANDRA, R. **Gradient boosting Bayesian neural networks via Langevin MCMC**. Neurocomputing, v. 558, p. 126726, 2023.
- BREIMAN, L. **Random Forests**. Machine Learning, 45, p. 5-32, 2001.
- BIGDOLI, A. A. **“Simulation and Optimization of Primary Oil and Gas Processing Plant of FPSO Operating in Pre-Salt Oil Field”**. Tese de Doutorado, USP, 2018.
- BOUKOUVALA, F.; IERAPETRITOU, M. G. **Feasibility analysis of black-box processes using an adaptative sampling Kriging-based method**. Computers & Chemical Engineering, vol. 36, p. 358-368, 2012.
- CARPIO, R. R. **Otimização do processo de produção de etanol de segunda geração em um sistema integrado à primeira geração**. Tese de Doutorado, UFRJ, 2019.
- CHEN, S.; BOUTESKA, A.; SHARIF, T.; ABEDIN, M. Z. **The Russian-Ukraine war and energy Market volatility: A novel application of the volatility ratio in the context of natural gas**. Resources Policy, vol. 85, part A, 2023.
- CNI. **Especificação do gás natural: Oportunidades e experiência internacional**. Confederação Nacional da Indústria – Brasília. 2019.
- SOUZA, T. E. G. **Modeling and Economic Optimization of an Industrial Site for Natural Gas Processing**. Dissertação de Mestrado, UFRJ, 2022.
- EL-HALWAGI, M. **Process Systems Engineering – Overview of optimization**. vol. 7, p. 285-314, 2006.
- FARAMAWY, S.; ZAKI, T; SAKR, A.A.-E. **Natural gas origin, composition, and processing: A review**. Journal of Natural Gas Science and Engineering, 2016. p. 34–54.
- FORRESTER, A. I. J.; KEANE, A. J. **Recent advances in surrogate-based optimization**. Progress in Aerospace Sciences 45, 2009, p. 50–79.
- GUTIERREZ, M. B. G. P. S. **O setor de gás natural no Brasil : uma comparação com os países da OCDE**. Texto para Discussão, p. 1–30, 27 jun. 2022. Disponível em:

https://repositorio.ipea.gov.br/bitstream/11058/11226/1/td_2777.pdf. Acesso em: 27 jan. 2024.

GÜRSAN, C.; DE GOOYERT, V. **The systemic impact of a transition fuel: Does natural gas help or hinder the energy transition?** *Renewable and Sustainable Energy Reviews*, 138, 2021.

HOLLAND, J. H. **Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence**. MIT Press Cambridge, MA, USA, 1 edition, 1992.

IEG. **Gas Market Report, Q1-2023**. Disponível em: <https://iea.blob.core.windows.net/assets/c6ca64dc-240d-4a7c-b327-e1799201b98f/GasMarketReportQ12023.pdf>. Acesso em: 27 jan. 2024.

ISLAM, S.; DINCER, I. **4.1 The Role of Energy Conversion. Em: Comprehensive Energy Systems**. [s.l.] Elsevier, 2018. p. 1–39.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning with Applications in R**. New York: Springer Science+Business Media, 2013.

JOSSO, P.; HALL, A.; WILLIAMS, C.; BAS, T. L.; LUSTY, P.; MURTON, B. **Application of random-forest machine learning algorithm for mineral predictive mapping of Fe-Mn crusts in the World Ocean**. *Ore Geology Reviews*, v. 162, p. 105671, nov. 2023.

KENNEDY, J.; EBERHART, R. **Particle Swarm Optimization**. *IEEE*, pp. 1942–1948, 1995.

MCKAY, M. D.; BECKMAN, R. J.; CONOVER, W. J. **A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code**. *Technometrics* vol.21, no.2, 1979.

MOKHATAB, S.; POE, W. A.; MAK, J. Y. **Handbook of Natural Gas Transmission and Processing**. Elsevier Inc, 2012.

MORAES, A. O.; MITRE, J. F.; LAGE, P. L.; SECCHI, A. R. **A robust parallel algorithm of the particle swarm optimization method for large dimensional engineering problems**. *Applied Mathematical Modelling*, v. 39, n. 14, pp. 4223–4241, 2014.

NELDER, J. A.; MEAD, R. **A Simplex Method for Function Minimization**. *The Computer Journal*, v. 7, n. 4, pp. 308–313, 1965. ISSN: 0010-4620.

NEVEUX, T.; ZANNANE, S.; DELAGE, T. **Metamodeling of chemical engineering unit operations using Kriging and prediction error estimation**. *Computer Aided Chemical Engineering*, 51, p. 535-540, 2022.

PROCESSONLINE.COM.AU. **Using electrical control valve actuators on oil and gas separators**. 2014. Report. Disponível em: <https://www.processonline.com.au/content/process-control-systems/article/using-electrical-control-valve-actuators-on-oil-and-gas-separators-75422284>. Acesso em: 27 jan. 2024.

WOLKE, V. **Preço do gás natural do Brasil é três vezes maior do que o dos Estados Unidos**. Portal da Indústria, 26 jun. 2018. Disponível em:

<https://noticias.portaldaindustria.com.br/noticias/infraestrutura/preco-do-gas-natural-do-brasil-e-tres-vezes-maior-do-que-o-dos-estados-unidos/>. Acesso em 27 jan. 2024.

7. APÊNDICES

7.1 APÊNDICE I – PARÂMETROS UTILIZADOS NO MODELO DE RLM

Para o método de RLM, foi utilizada a classe “`sklearn.linear_model.Ridge()`” da biblioteca “Scikit Learn”. Os parâmetros utilizados foram:

- `alpha = 0.001`
- `fit_intercept = True`
- `copy_X = True`
- `max_iter = None`
- `tol = 1e-4`
- `solver = ‘auto’`
- `positive = ‘False’`
- `random_state = None`

A descrição de cada parâmetro pode ser encontrada no website da documentação do código: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.Ridge.html

7.2 APÊNDICE II - PARÂMETROS UTILIZADOS NO MODELO DE RANDOM FOREST

Para o método de *Random Forest*, foi utilizada a classe “`sklearn.ensemble.RandomForestRegressor()`” da biblioteca “Scikit Learn”. Os parâmetros utilizados foram:

- `n_estimators = 100`
- `criterion = “squared_error”`
- `max_depth = None`
- `min_samples_split = 2`
- `min_samples_leaf = 1`
- `min_weight_fraction_leaf = 0.0`
- `max_features = 1.0`
- `max_leaf_nodes = None`
- `min_impurity_decrease = 0.0`
- `bootstrap = True`
- `oob_score = False`
- `n_jobs = None`
- `random_state = None`
- `verbose = 0`
- `warm_start = False`
- `ccp_alpha = 0.0`
- `max_samples = None`
- `monotonic_cst = None`

A descrição de cada parâmetro pode ser encontrada no website da documentação do código:

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>

7.3 APÊNDICE III – PARÂMETROS UTILIZADOS NO MODELO DE XGBOOSTING

Para o método de *XGBoosting*, foi utilizada a classe “`xgboost.XGBRegressor()`” da biblioteca “`xgboost`”.

Todos os parâmetros utilizados no método assumiram os valores padronizados do código.

A descrição de cada parâmetro pode ser encontrada no website da documentação do código:
https://xgboost.readthedocs.io/en/latest/python/python_api.html

7.4 APÊNDICE IV – PARÂMETROS UTILIZADOS NO MODELO DE GRADIENT BOOSTING

Para o método de *Gradient Boosting*, foi utilizada a classe “`sklearn.ensemble.GradientBoostingRegressor()`” da biblioteca “Scikit Learn”. Os parâmetros utilizados foram:

- `loss = 'squared_error'`
- `learning_rate = 0.1`
- `n_estimators = 100`
- `subsample = 1.0`
- `criterion = 'friedman_mse'`
- `min_samples_split = 2`
- `min_samples_leaf = 1`
- `min_weight_fraction_leaf = 0.0`
- `max_depth = 3`
- `min_impurity_decrease = 0.0`
- `init = None`
- `random_state = None`
- `max_features = None`
- `alpha = 0.9`
- `verbose = 0`
- `max_leaf_nodes = None`
- `warm_start = False`
- `validation_fraction = 0.1`
- `n_iter_no_change = None`
- `tol = 1e-4`
- `ccp_alpha = 0.0`

A descrição de cada parâmetro pode ser encontrada no website da documentação do código:

<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html>

7.5 APÊNDICE V – PARÂMETROS UTILIZADOS NO MODELO KRIGING (SCIKIT LEARN)

Para o método de *Kriging*, foi utilizada a classe “`sklearn.gaussian_process.GaussianProcessRegressor()`” da biblioteca “Scikit Learn”. Os parâmetros utilizados foram:

- `kernel = 1 * sklearn.gaussian_process.kernels.RBF(length_scale=0.3, length_scale_bounds = (1e-2, 1e2))`
- `alpha = 1e-10`
- `optimizer = “fmin_l_bfgs_b”`
- `n_restarts_optimizer = 3`
- `normalize_y = False`
- `copy_X_train = True`
- `n_targets = None`
- `random_state = None`

A descrição de cada parâmetro pode ser encontrada no website da documentação do código:

https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.GaussianProcessRegressor.html

7.6 APÊNDICE VI – PARÂMETROS UTILIZADOS NO MODELO KRIGING (CARPIO, 2019)

A segunda versão utilizada do método de *Kriging* foi uma adaptação em Python do código utilizado em CARPIO, 2019.

O código do método contendo os parâmetros utilizados foi disponibilizado em um github, disponível em:

<https://github.com/Lucas-Anelli/TCC>