



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
TRABALHO DE CONCLUSÃO DE CURSO EM ENGENHARIA
QUÍMICA



MODELO PARA PREDIÇÃO DE UMIDADE EM SISTEMA DE DESIDRATAÇÃO DE GÁS NATURAL

Autor: Murilo de Almeida dos Santos

Orientador: Marcelo Farenzena

Porto Alegre, setembro de 2023

Autor: Murilo de Almeida dos Santos

MODELO PARA PREDIÇÃO DE UMIDADE EM SISTEMA DE DESIDRATAÇÃO DE GÁS NATURAL

*Trabalho de Conclusão de Curso apresentado à
COMGRAD/ENQ da Universidade Federal do Rio
Grande do Sul como parte dos requisitos para a
obtenção do título de Bacharel em Engenharia
Química*

Orientador: Marcelo Farenzena

Banca Examinadora:

Prof. Dr., Jorge Otávio Trierweiler, UFRGS

MSc, Thiago Anzai, Petrobras

Porto Alegre

2023

AGRADECIMENTOS

Aos meus pais, Mario e Sandra, agradeço pelo amor e apoio incondicionais. Mesmo nos momentos mais difíceis estavam sempre presentes, me ajudando e incentivando a perseguir os meus sonhos. Ao meu irmão, Gabriel, que com a sua amizade e conversas me ensina a olhar a vida sob outra perspectiva.

A minha namorada, Maria Fernanda, que sempre está disposta a me ajudar, pelo carinho e momentos divididos nessa jornada. Que tornou mais leve e descontraído o tempo de escrita e me ajuda a levar a vida com mais leveza.

Aos colegas da engenharia, pelos longos momentos de estudos em grupo, pelas risadas e momentos de descontração no vale e na cidade baixa. Por estarem junto comigo e pelo suporte mútuo ao longo desses cinco anos.

Aos professores da UFRGS, que fizeram parte da minha formação como profissional, por toda a dedicação prestada, fazendo jus à excelência reconhecida da universidade e do curso de engenharia química.

Ao meu orientador, Professor Farenzena, pela paciência, pelas sugestões e ideias na resolução deste trabalho, mostrando novas possibilidades dentro da engenharia química.

RESUMO

Analisadores virtuais são modelos matemáticos baseados em dados que utilizam medições de outras variáveis, de controle ou monitoramento de processo, para predição da variável de interesse de forma contínua. Estes tipos de analisadores são vantajosos, pelo baixo custo associado de obtenção e manutenção, sendo uma alternativa para a redução de custos em processos industriais. No processamento e extração de gás natural, existe a etapa da desidratação, que tem por objetivo a remoção da umidade presente no gás. Uma das metodologias utilizadas para a desidratação é a adsorção “TSA”. Uma das maneiras de otimizar este processo é definir as concentrações de água na saída do leito, para evitar a exposição do leito à altas temperaturas, e conseqüente diminuição de vida útil. Este trabalho tem por objetivo a modelagem de um analisador virtual para predição de umidade no sistema, utilizando as metodologias de regressão linear multivariável e redes neurais artificiais. Foi realizada a simulação dinâmica do sistema para obtenção da base de dados, observando a utilização de parâmetros de entrada com medições típicas deste processo. O modelo gerado a partir da regressão linear multivariável apresentou baixo coeficiente de determinação (R^2) de 0,7322 e erro relativo médio de 0,4041% em função da baixa correlação linear com a variável de interesse. Já o modelo não linear, redes neurais, foi desenvolvido e otimizado em seus hiperparâmetros, que representam um ajuste melhor dos dados, obtendo coeficiente de determinação (R^2) de 0,9727 e erro relativo médio de 0,1230%.

Palavras-chave: analisador virtual, redes neurais, modelagem de processo, predição de umidade

ABSTRACT

Soft sensors are mathematical models based on data that use measurements of other variables, control or process monitoring, to predict the variable of interest continuously. These types of analyzers are advantageous due to the low cost associated with obtaining and maintaining them, being an alternative for reducing costs in industrial processes. In the processing and extraction of natural gas there is a dehydration step, which aims to remove the moisture present in the gas. One of the methodologies used for dehydration is the “TSA” adsorption. One of the ways to optimize this process is to define the water concentrations at the bed outlet to avoid exposure of the bed to high temperatures and consequent decrease in useful life. The aim of this work is to model a virtual analyzer for predicting moisture in the system using multivariate linear regression methodologies and artificial neural networks. A dynamic simulation of the system was carried out to obtain the database, observing the use of input parameters with typical measurements of this process. The model generated from multivariate linear regression showed a low coefficient of determination (R^2) of 0.7322 and a mean relative error of 0.4041% due to the low linear correlation with the variable of interest. The non-linear model, neural networks, was developed and optimized in its hyperparameters, which represent a better fit of the data, obtaining a coefficient of determination (R^2) of 0.9727 and a mean relative error of 0.1230%.

Keywords: *soft sensor, neural networks, process modeling, moisture prediction*

LISTA DE FIGURAS

Figura 2.1 – Produção de Gás Natural no Brasil	4
Figura 2.2 – Representação gráfica MRL.....	6
Figura 2.3 – Rede neural <i>feedforward</i> de camada simples modelo Perceptron	7
Figura 2.4 - Arranjo de uma rede neural Perceptron de múltiplas camadas	8
Figura 2.5 – Estrutura básica de um analisador virtual	11
Figura 2.6 – Metodologia para desenvolvimento de um analisador virtual.....	12
Figura 3.1 – Unidade de TSA para desidratação do GN com 3 colunas utilizando peneiras moleculares como adsorventes	17
Figura 3.2 – Curva de temperatura típica para uma planta de adsorção operando em ciclo de regeneração.	17
Figura 3.3 – Condensador em sistema de regeneração TSA em unidade de processamento de gás natural	18
Figura 3.4 – Arranjo utilizado na simulação do sistema de resfriamento	20
Figura 4.1 – Correlação de Pearson entre os parâmetros da base de dados.	25
Figura 4.2 – <i>Boxplot</i> das variáveis disponíveis na base de dados.	25
Figura 4.3 – Histograma das variáveis disponíveis na base de dados.....	26
Figura 4.4 – Comparação entre os valores preditos pelo modelo de regressão linear e valores reais em percentual	27
Figura 4.5 – Comparação entre o erro relativo e a fração de água no modelo MRL. 27	
Figura 4.6 - Comparação entre os valores preditos pelo modelo de redes neurais e valores reais.....	31
Figura 4.7 – Comparação entre o erro relativo e a fração de água no modelo RNA. 31	
Figura 4.8 - Comparação do erro relativo entre os modelos e a fração de água.....	32

LISTA DE TABELAS

Tabela 3.1 – Medições tipicamente disponíveis em unidades TSA	19
Tabela 3.2 – Dados de entrada das correntes no simulador	20
Tabela 4.1 – Tabela de títulos	23
Tabela 4.2 – Análise preliminar da base de dados obtidos na simulação	24
Tabela 4.3 – Otimização da função de ativação utilizada para a modelagem utilizando redes neurais.....	28
Tabela 4.4 – Otimização do tamanho da amostra utilizada para a modelagem utilizando redes neurais.....	28
Tabela 4.5 – Otimização do número de neurônios utilizados na primeira camada oculta para a modelagem utilizando redes neurais	29
Tabela 4.6 – Otimização do número de neurônios utilizados na segunda camada oculta para a modelagem utilizando redes neurais.....	29
Tabela 4.7 – Resultados de coeficiente de determinação para modelos de analisadores virtuais obtidos na literatura.....	32

SUMÁRIO

1	Introdução	1
2	Revisão Bibliográfica	3
2.1	Processamento de gás natural	3
2.2	Modelagem matemática de sistemas	4
2.2.1	Regressão Linear Multivariável	5
2.2.2	Redes Neurais Artificiais	7
2.3	Analisadores Virtuais	9
2.4	Desenvolvimento de Analisadores Virtuais	11
2.4.1	Primeira inspeção de dados	12
2.4.2	Seleção de dados históricos	12
2.4.3	Pré-tratamento dos dados	13
2.4.3.1	Dados Faltantes	13
2.4.3.2	Outliers	14
2.4.4	Seleção, treinamento e validação do modelo	14
2.4.5	Manutenção do analisador virtual	14
3	Materiais e Métodos	16
3.1	Estudo de Caso – Unidade de Desidratação de Gás Natural	16
3.1.1	Processo TSA para desidratação de gás natural	16
3.2	Variáveis com Medição Disponível	19
3.3	Desenvolvimento do Analisador Virtual	19
3.3.1	Aquisição dos dados de processo	19
3.3.2	Pré-tratamento dos dados	21
3.3.3	Seleção, treinamento e validação do modelo	21
4	Resultados	23
4.1	Pré-tratamento dos dados	23
4.2	Modelo de Regressão Linear Multivariável (MRL)	26
4.3	Modelo de Redes Neurais Artificiais Multicamadas (ANN)	28
4.4	Comparação dos modelos desenvolvidos	32
5	Conclusões e Trabalhos Futuros	34
	REFERÊNCIAS	35

1 Introdução

O custo associado a indústrias de óleo e gás é muito alto, seja na hora do investimento inicial, seja na manutenção dos equipamentos e instrumentos instalados. O controle de processos neste segmento é muito importante e depende da confiabilidade da instrumentação. Porém, são recorrentes os casos de falta de medição em instalações industriais por avarias em sensores. Nessa condição, o uso de analisadores virtuais se torna uma opção viável para substituição ou redundância de instrumentos de campo. Podem ainda ser utilizados para o monitoramento contínuo de parâmetros, onde em comparação com instrumentos convencionais possuem baixo custo, ou quando não existe tecnologia para a medição da variável desejada (FACCHIN, 2005).

A descoberta e exploração do pré-sal permitiu a ampliação da produção de gás natural e petróleo no Brasil. Aliado a políticas de incentivo governamentais sobre mudança de matriz energética, o gás natural vem ganhando um papel cada vez mais importante como fonte de energia industrial. Porém, os desafios relacionados ao processamento deste combustível têm se tornado ponto de atenção. A desidratação é uma etapa indispensável na produção, onde ocorre a redução do teor de água presente no gás, evitando a formação de hidratos e prevenindo a incrustação em linhas e equipamentos (CAVALCANTE JUNIOR, 2023).

O processo Temperature Swing Adsorption (TSA) é um processo largamente utilizado na etapa de desidratação de gás natural. Este processo consiste na passagem da mistura gasosa pressurizada por um leito fixo, recheado com um material adsorvente, que possui função de aprisionar em sua superfície o contaminante, removendo-o da mistura. O gás contendo o principal contaminante, a água, passa por um resfriador que promove a condensação da água e é direcionado a um vaso de flash, onde a fase líquida é removida do processo (SANTOS, 2016).

Este processo ocorre de forma cíclica, em etapas de adsorção e dessorção. A dessorção é a fase determinante do processo, onde ocorre o aquecimento do leito até a temperatura necessária para a remoção da água. A otimização nesta etapa, então, possibilita ganhos no processo. O leito, composto de peneiras moleculares, pode

sofrer redução de sua vida útil quando exposto desnecessariamente a altas temperaturas (SANTOS, 2016).

Nesse contexto, o principal objetivo deste trabalho é o desenvolvimento de um modelo capaz de prever a umidade do gás na saída de leitos fixos de unidades TSA, utilizando medições inline, típicas deste processo. Dessa forma, a unidade tem a possibilidade de evitar o consumo de gás previamente especificado desnecessariamente, diminuir o consumo de energia no permutador de aquecimento, prolongar a vida útil das colunas e, ainda, reduzir a duração dos ciclos de regeneração.

Os objetivos específicos do trabalho contemplam:

- I. Modelagem dinâmica de um sistema de resfriamento de gás natural, presente na etapa de dessorção de uma unidade TSA, para geração da base de dados;
- II. Comparação das metodologias de desenvolvimento de analisadores virtuais: Regressão linear multivariável e Redes Neurais Artificiais;
- III. Comparação de diferentes hiperparâmetros da rede neural, para obter a otimização do modelo.

Este trabalho está estruturado em 5 capítulos: o Capítulo 1 traz as motivações e objetivos do trabalho; o Capítulo 2 apresenta o referencial bibliográfico, contendo tópicos essenciais sobre processamento de gás natural, modelagem matemática de sistemas (utilizando as técnicas de regressão linear multivariável e redes neurais artificiais) e analisadores virtuais, descrevendo-os e revisando sobre a metodologia de desenvolvimento e manutenção; o Capítulo 3 aborda o caso de estudo, o qual descreve a unidade que será baseada para o desenvolvimento do modelo, e também a metodologia empregada nas etapas de desenvolvimento do analisador e obtenção da base de dados; o Capítulo 4 apresenta os resultados e discussões sobre o tratamento dos dados e modelagem, otimização e validação dos modelos. Por fim, o Capítulo 5 encerra o presente trabalho com as conclusões decorrentes da análise dos resultados e apresenta ideias de trabalhos futuros.

2 Revisão Bibliográfica

Este capítulo tem por objetivo fazer uma contextualização do tema de estudo e apresentar a fundamentação teórica para o desenvolvimento deste trabalho. Além disso, discute algumas metodologias que serão utilizadas

2.1 Processamento de gás natural

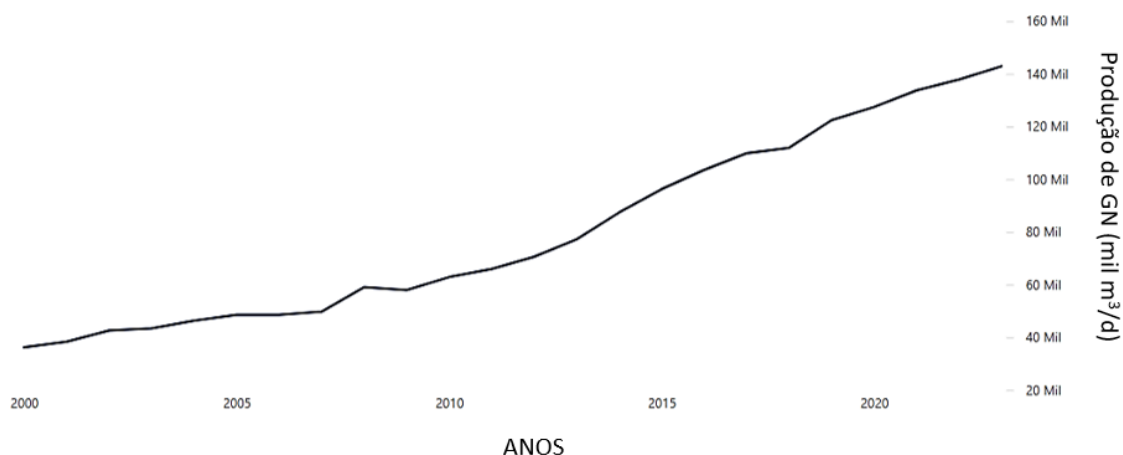
O gás natural pode ser definido como a porção de petróleo que existe na fase gasosa ou em solução no óleo, nas condições de reservatório e que permanece na fase gasosa em condições atmosféricas de pressão e temperatura (VAZ *et al.*, 2008). O gás natural é uma fonte de energia muito versátil, que pode ser utilizado como combustível para fornecimento de calor, geração de eletricidade e de força motriz. É utilizado como matéria-prima nas indústrias siderúrgica, química, petroquímica e de fertilizantes. Na área de transportes, serve como substituto do óleo diesel, gasolina e álcool (GONZAGA, 2014).

O avanço tecnológico que ocorreu desde o final do século XX, trouxe muitos benefícios que melhoraram a vida das pessoas. Ao mesmo tempo, trouxe novos desafios para os governantes em relação ao futuro do planeta. Pauta de grandes reuniões ao redor do mundo, as mudanças climáticas se tornaram um problema emergente.

Definições como a Conferência Mundial das Nações Unidas (RIO-92), o Protocolo de Quioto (1997) e mais recentemente o Acordo de Paris (COP-21), assumiram um papel importante na redução da emissão de gases de efeito estufa (GEE), através da redução do consumo de combustíveis e mudança de matriz energética.

Neste cenário, o gás natural vem ganhando força em relação às fontes de energia concorrentes, como carvão e petróleo (VAZ *et al.*, 2008). No Brasil, o consumo também é incentivado pela descoberta de novos poços de petróleo no pré-sal, que possuem grandes quantidades de gás natural. A demanda industrial e a melhoria em tecnologias utilizadas para a extração em unidades de processamento possibilitaram o aumento da produção desse combustível no Brasil, como mostra a Figura 2.1.

Figura 2.1 – Produção de Gás Natural no Brasil.



Fonte: Painéis Dinâmicos de Produção de Petróleo e Gás Natural ANP

Para atender às especificações de qualidade da Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP), o gás natural precisa passar por diversos processos de desidratação e remoção de componentes ácidos antes de ser utilizado como fonte de energia. Esta remoção evita a formação de hidratos e a corrosão em tubulações e equipamentos de processamento e distribuição. O processo ainda aumenta o poder calorífico do gás natural (ROCHA, 2015).

2.2 Modelagem matemática de sistemas

Modelos matemáticos são representações de sistemas ou fenômenos, formulados em termos de equações ou relações matemáticas. Descrevem as correspondências entre variáveis e fornecem uma estrutura para entender, analisar e prever o comportamento do sistema. Os modelos estão sujeitos ao limite de sua concepção, como a precisão dos dados utilizados, as faixas de funcionamento e o nível de profundidade dos fenômenos envolvidos. FACCHIN (2005) subdivide os modelos em grupos: fenomenológicos, análogos, empíricos e semi-empíricos.

Modelos fenomenológicos são baseados em equações que possuem sua origem em princípios científicos básicos da física e química. Utilizam teorias fundamentadas e seguem os princípios de conservação de massa, energia e quantidade de movimento. Esse tipo de modelo possui alta demanda de esforço para obtenção, em sistemas complexos, porém uma maior capacidade extrapolativa. Essa metodologia pode ser chamada de modelos caixa branca (FACCHIN, 2005).

Modelos empíricos são equações baseadas exclusivamente em dados de sistemas reais, sem fundamentação teórica explícita ou conhecimentos detalhados dos mecanismos presentes no fenômeno de estudo. São relações entre variáveis de entrada do sistema resultando na variável de estudo. Esse tipo de metodologia também pode ser definido como modelo caixa preta (KADELEC *et al.*, 2009).

Modelos semi-empíricos (ou caixa cinza) são baseados nos princípios fenomenológicos, mas alguns parâmetros são substituídos por modelos empíricos sedimentados em dados experimentais (FACCHIN, 2005).

De acordo com KADELEC *et al.* (2009), as principais técnicas utilizadas para o desenvolvimento de modelos empíricos e semi-empíricos são as de Regressão Linear Multivariável (MRL), com adição da técnica de Mínimos Quadrados Parciais (PLS), Regressão por Componentes Principais (PCR) e Redes Neurais Artificiais (ANN).

2.2.1 Regressão Linear Multivariável

A técnica de Regressão Linear Multivariável é um método estatístico usado para modelar a relação de variáveis independentes (ou de entrada), para prever uma variável de resposta (ou dependente) de um determinado sistema. Essa técnica é muito versátil e amplamente utilizada para analisar o efeito concomitante de diferentes mudanças nas variáveis de entrada no resultado obtido (ZIMMER; ANZANELLO, 2014).

GELADI e KOWALSKI (1986) estabelecem essa situação em notação matricial, sendo \mathbf{x} um conjunto de variáveis de entrada, composto por m colunas e n linhas, e \mathbf{y} um vetor contendo a variável de resposta, com uma coluna de n linhas. Esta relação linear entre as variáveis do sistema pode ser descrita matematicamente como:

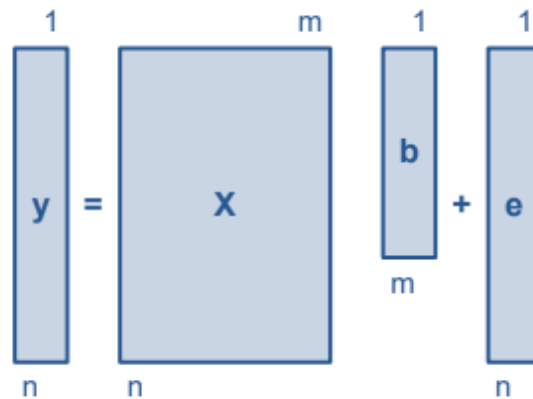
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_m x_m + \varepsilon \quad (1)$$

onde ε é o resíduo (ou erro), β_m são os coeficientes lineares correspondentes as variáveis independentes x_m . Essa correlação é aplicável para $n=1$, para situações onde $n > 1$ utiliza-se notação matricial, i.e.:

$$y = X * \beta + \varepsilon \quad (2)$$

Para um melhor entendimento, a equação 2 pode ser representada graficamente através da Figura 2.2.

Figura 2.2 – Representação gráfica MRL.



Fonte: Adaptado GELADI e KOWALSKI (1986).

GELADI e KOWALSKI (1986) dissertam sobre três possíveis situações que podem ocorrer na utilização da técnica MRL, em função do número de amostras n e do número de variáveis independentes m :

- $m > n$: Neste cenário existem mais variáveis independentes do que amostras, gerando a possibilidade de infinitos números de soluções para o vetor β , situação que deve ser evitada.
- $m = n$: Número de amostras exatamente igual ao número de variáveis independentes, fornece uma única solução para β . A ocorrência desse cenário é improvável em problemas reais.
- $m < n$: Existem mais amostras do que variáveis, com esse tipo de situação não é possível encontrar uma solução exata para β .

Existem diversas técnicas para a solução do problema apresentado em $m < n$. Pode-se destacar o Método de Mínimos Quadrados, amplamente difundido. — A minimização deste erro se dá com:

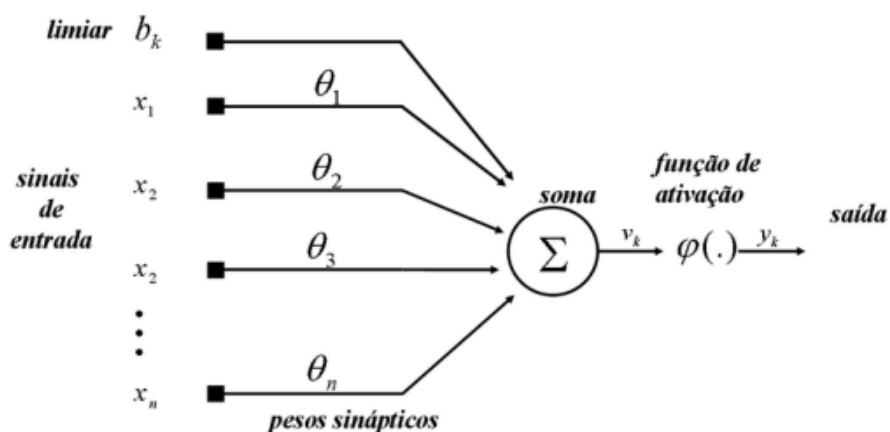
$$\beta = (X^T * X)^{-1} * X^T * y \quad (3)$$

2.2.2 Redes Neurais Artificiais

A técnica de redes neurais artificiais são modelos computacionais não-lineares inspirados no sistema nervoso de organismos vivos. Consiste em conjuntos de unidades de processamento denominados neurônios artificiais, que são interconectados por um número de conexões de maneira similar às sinapses biológicas. Este modelo possui aprendizado adaptativo, onde a rede neural ajusta seus parâmetros internos com base na experiência, permitindo a melhoria contínua de seu desempenho (SILVA et al. 2010).

Além disso, as redes neurais podem ter vários arranjos, chamados de arquiteturas, como por exemplo *feedforward* de camada simples ou *feedforward* de camadas múltiplas. O modelo Perceptron é um exemplo típico de arranjo *feedforward* de camada simples, que utiliza função de ativação que trabalha em conjunto com a função de decisão, amplamente utilizado nas mais diversas arquiteturas. Como ilustra a Figura 2.3, esse arranjo é constituído de uma única camada de um único neurônio podendo possuir n sinais de entrada, representados por b_k (BIAS) x_n , ponderados individualmente por pesos sinápticos, θ_n , adicionados o limiar de ativação e passando pela função de ativação, ϕ , resultando em apenas uma única saída, a variável de interesse (BOMFIM et al. 2020).

Figura 2.3 – Rede neural *feedforward* de camada simples modelo Perceptron.



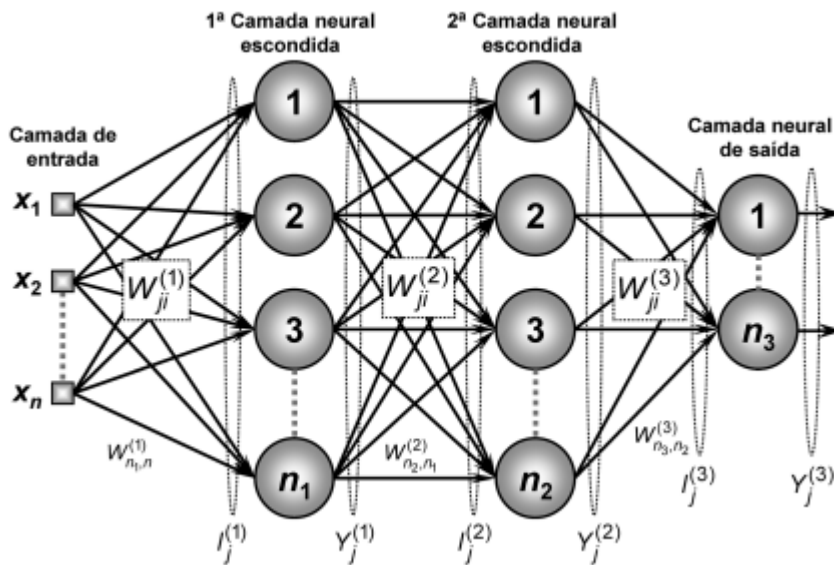
Fonte: BOMFIM, et al. (2020).

Essa arquitetura possui algumas limitações. Nos casos de problemas não linearmente separáveis, há ocorrência de não convergência. Dessa forma, as redes Perceptron multicamadas surgiram com a capacidade de resolução de problemas mais complexos e possuem uma ampla gama de aplicabilidade, como aproximação

de funções, reconhecimento de padrões, identificação e controle de processos, previsão de séries temporais e otimização de sistemas (BELISÁRIO, 2020).

Dessa forma, essa arquitetura de redes neurais será utilizada para o desenvolvimento do modelo neste trabalho. O princípio de funcionamento é similar ao descrito previamente para o modelo de camada simples, porém neste modelo os sinais de entrada são propagados de camada a camada no sentido da camada de saída, como pode ser verificado na Figura 2.4.

Figura 2.4 - Arranjo de uma rede neural Perceptron de múltiplas camadas.



Fonte: SILVA et al. (2010)

Neste modelo, $w_{ji}^{(L)}$ indicam o valor do peso sináptico conectando o j -ésimo neurônio da camada (L) ao i -ésimo neurônio da camada ($L-1$) e $Y_j^{(L)}$ corresponde à saída do j -ésimo neurônio da camada L . Inicialmente, as observações de treinamento são alimentadas diretamente à primeira camada de neurônios ocultos. Para cada neurônio (ou unidade), é calculada uma entrada ponderada, $I_j^{(L)}$, dada pela combinação linear entre valores das entradas e seus pesos correspondentes, adicionado ao limiar $b_j^{(L)}$, que funciona como limite para variar a atividade do neurônio, conforme equação 4 (SILVA et al. 2010).

$$I_j^{(L)} = \sum_i (w_{ji}^{(L)} x_i) + b_j^{(L)} \quad (4)$$

Sobre a entrada ponderada $I_j^{(L)}$ é aplicada uma função de ativação, com o objetivo realizar uma transformação não linear nos valores de entrada, conforme apresentado na equação 5.

$$Y_j^{(L)} = g(I_j^{(L)}) \quad (5)$$

Para fins de comparação de modelo, foi utilizado o coeficiente de determinação (R^2), métrica estatística de avaliação entre dados preditos com os dados reais. Quanto mais próximo de 1, melhor o ajuste dos dados. Pode ser descrita conforme a equação 6.

$$R^2 = 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(\hat{y}_i - \bar{y}_i)^2} \quad (6)$$

onde \hat{y}_i são os valores gerados pelo modelo, y_i os dados originais utilizados para elaboração do modelo e \bar{y}_i o valor médio do conjunto de dados originais.

Para isso, é necessário observar a necessidade da separação da amostra original em etapas de treino e de teste para que os modelos possam ser avaliados. Essa separação é realizada segregando uma parcela da base de dados utilizada, de 10 a 40%, onde a maior fração será utilizada para o treinamento, onde ocorre o aprendizado de máquina, e a menor na validação dos dados, para seja possível confirmar o ajuste do modelo sem interferência com os dados treinados como feito por LINHARES (2010).

Outro ponto importante na concepção do arranjo é a decisão do número de neurônios em cada camada oculta e o número de camadas necessárias para descrever o sistema de forma apropriada é empírica. Dessa forma, começa-se com arranjos simples e aumentando o grau de complexidade conforme a necessidade do modelo (SOUZA *et al.*, 2010).

2.3 Analisadores Virtuais

Analisadores virtuais são algoritmos matemáticos que utilizam correlações entre variáveis de processo e medidas online para a obtenção de resultados em tempo real,

de variáveis que apresentam dificuldades de aquisição de forma contínua ou que não podem ser medidas por sensores (FORTUNA *et al.*, 2005).

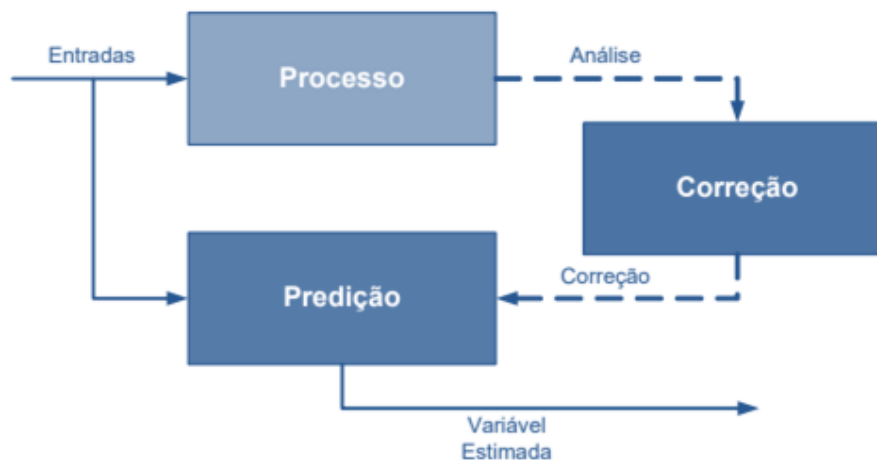
Esse tipo de estratégia de monitoramento e obtenção de dados apresenta diversas vantagens em sua utilização:

- Opção de baixo custo quando comparada a instrumentação online;
- Permite a obtenção de dados em tempo real, eliminando o tempo de atrasos de instrumentação e/ou análises laboratoriais;
- Facilidade de criação do analisador e implementação para o monitoramento;
- Aumento de confiabilidade de sistemas de controle: quando utilizado de forma concomitante à sensores físicos, pode promover identificação de falhas em instrumentação de campo;
- Possibilidade de prever variáveis que não poderiam ser avaliadas pelo modo convencional, em função de limitações tecnológicas.

A Figura 2.5 denota os componentes básicos de um analisador virtual, onde uma ou mais variáveis medidas são utilizadas como dados de entrada para o algoritmo de predição que gera uma estimativa para a variável desejada. O modelo conta ainda com um sistema de adaptação com medições para validação e manutenção do analisador frente a novos dados (FACCHIN, 2005).

As entradas do algoritmo são variáveis de processo que se relacionam com a variável desejada no equacionamento matemático do analisador. Dessa forma, o analisador recebe estas informações para realizar a predição da variável de interesse. Como em todos processos, ocorrem mudanças naturais com o passar da campanha, como variações de temperaturas ou pressões, o modelo pode ser reajustado em seus parâmetros com base em análises da variável a qual deseja-se substituir para apresentar como saída dados coerentes com a unidade (FACCHIN, 2005)

Figura 2.5 – Estrutura básica de um analisador virtual.



Fonte: FACCHIN (2005)

Analisadores virtuais são normalmente empregados em processos de alta complexidade, como indústrias de processos químicos e petroquímicos, onde existe uma grande quantidade de dados históricos disponíveis. Essa quantidade permite que sejam desenvolvidos de forma a atender as especificações de precisão. Esses modelos são, de maneira geral, puramente baseados em dados, caixa preta, ou de maneira semi-empírica, caixa cinza. (KADLEC *et al.*, 2009)

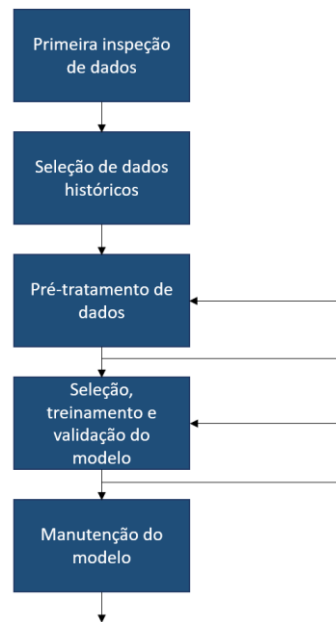
A utilização deste tipo de tecnologia em plantas industriais vem sendo estudada por diversos pesquisadores, de forma a suprir a necessidade de obtenção de dados que, por diversas razões, encontram-se indisponíveis. Na literatura, observa-se inúmeros trabalhos sobre o desenvolvimento e utilização de analisadores virtuais. Por exemplo, a predição do índice de fluidez (MI) em uma planta de produção de polietileno de alta densidade (MOHR, 2004), predição do ponto final de ebulição da nafta de craqueamento (OLIVEIRA, 2010) ou estimar as frações molares de contaminantes em GLP em unidade de processamento de gás natural (LINHARES, 2010).

2.4 Desenvolvimento de Analisadores Virtuais

Existem diversas metodologias empregadas na literatura para o desenvolvimento de analisadores virtuais, que envolvem desde a seleção da técnica de modelagem à técnica de correção a ser aplicada no analisador desenvolvido. Uma vez que este trabalho tem por objetivo o desenvolvimento de um modelo baseado em dados,

KADLEC et al. (2009) descrevem a estrutura básica para o desenvolvimento destes modelos que será utilizado neste trabalho, conforme Figura 2.6.

Figura 2.6 – Metodologia para desenvolvimento de um analisador virtual.



Fonte: Adaptado de KADLEC *et al.* (2009)

2.4.1 Primeira inspeção de dados

O ponto de partida para o desenvolvimento de um analisador virtual é uma inspeção preliminar da estrutura dos dados que estão disponíveis a serem utilizados. Verificam-se quaisquer problemas óbvios que possam trazer desvios muito grandes para o modelo. Por exemplo, valores de vazão negativos ou temperaturas muito fora do range de operação devem ser eliminados da base de dados. Neste ponto, deve-se ater à variável desejada se essa apresenta um comportamento possível de predição, como a variação ser maior que o erro intrínseco dos medidores das variáveis independentes (KADLEC et al. 2009).

2.4.2 Seleção de dados históricos

Para a modelagem do analisador, em grande parte dos casos serão utilizados dados correspondentes a estados estacionários do processo. Dessa forma, deve-se realizar a identificação dos estados estacionários do processo e selecioná-los para o treinamento e avaliação do modelo a ser desenvolvido. Para processos em batelada, visto que o processo ocorre em constante estado transiente, deve se observar a

utilização de dados de lotes confiáveis e representativos do processo (KADLEC *et al.*, 2009).

2.4.3 Pré-tratamento dos dados

A ideia por trás deste passo é a melhoria dos dados de entrada utilizados para a modelagem do analisador, para que o modelo desenvolvido seja mais efetivo em seu processamento. Uma ação típica para este passo é a utilização da técnica de Análise de Componentes Principais (PCA) ou Regressão dos componentes principais (PCR). Como apontado na Figura 2.6, é possível verificar que esta etapa apresenta um ponto crítico deste processo, onde os dados devem ser constantemente trabalhados para que o modelo continue válido e apresentando respostas coerentes. Alguns dos problemas recorrentes desta etapa são a presença de dados faltantes e a detecção e substituição de *outliers* (KADLEC *et al.*, 2009).

2.4.3.1 Dados Faltantes

Dados faltantes podem ser definidos como a amostra, ou conjunto de amostras, onde as variáveis (medições) têm um valor diferente do valor real. Essas variáveis afetadas geralmente possuem valores como infinito, zero ou outro valor constante (CHEN e CHEN, 2000).

Esses valores podem possuir diversas origens. Plantas de processamento possuem grande número de instrumentos ativos, realizando medidas diversas e armazenando os dados processados. Dessa forma, é inerente ao funcionamento que em um dado momento ocorra falha em alguns sensores. Seja devido a problemas de funcionamento do instrumento, na transmissão do sinal gerado ou no armazenamento e acesso ao banco de dados.

KADLEC *et al.* (2009) afirmam que a maior parte das técnicas de modelagem utilizadas não possuem mecanismos para “filtrar” dados faltantes. Dessa forma, surge a necessidade da implementação de estratégias de substituição destes valores. Essas estratégias podem ocorrer de duas formas. A primeira com a exclusão dos dados naquele dado instante ou, da segunda maneira, utilizando médias, mediana ou moda da variável para evitar grandes interferências. Para situações complexas também podem ser utilizados algoritmos potentes ou regressões para a utilização dos dados, porém cada caso deve ser avaliado (SCHEFFER, 2002).

2.4.3.2 *Outliers*

Outliers são valores medidos por sensores que desviam dos intervalos típicos ou significativos dos valores medidos. De acordo com QIN (1997), pode se separar em dois tipos. Óbvios, onde os valores extrapolam as limitações físicas ou tecnológicas de um dado equipamento - esses dados devem ser observados na concepção -, e os não-óbvios, aqueles que são mais difíceis de identificar pois não violam nenhuma limitação, porém não refletem os valores reais para a variável medida.

2.4.4 *Seleção, treinamento e validação do modelo*

Essa é a fase crítica do desenvolvimento do analisador virtual. A seleção das variáveis secundárias afeta diretamente a performance do analisador. Em modelos caixa-preta, existem diversas técnicas para a determinação das variáveis mais efetivas, como por exemplo algoritmos de busca aleatória, busca exaustiva e outras combinações para chegar ao melhor resultado. Porém, nos modelos caixa-cinza o conhecimento do processo e correlações teóricas intrínsecas ao sistema podem proporcionar o desenvolvimento de um sensor que possua melhor desempenho (FACCHIN, 2005).

KADLEC *et al.* (2009) afirmam que não existe um modelo ideal ou uma classificação para determinar a melhor técnica a ser utilizada. A escolha depende do processo, desempenho requerido e da experiência do desenvolvedor. A concepção inicia com a utilização de técnicas mais simples e aumentando a complexidade dos métodos perante a necessidade de melhor performance.

2.4.5 *Manutenção do analisador virtual*

A manutenção do modelo deve-se ao fato de algumas condições de operação ou características do processo variarem ao longo do tempo. Dessa forma, o analisado passa a prever valores com diferença entre o valor real mais acentuada. Para que o modelo siga eficaz, é interessante que periodicamente seja adaptado e revisado à situação atual, de forma que a confiabilidade dos valores de saída se mantenham fiéis ao processo (FACCHIN, 2005).

As mudanças inerentes do processo podem vir de diversas fontes, como por exemplo desgaste físico em instrumentos ou equipamentos, pureza de matérias-

primas ou desativação catalítica, que impactam diretamente os parâmetros de processo com o passar do tempo (KADLEC *et al.*, 2009).

A técnica de correção para estes modelos mais utilizada é a adição de um *bias* contendo a diferença entre a variável real e a variável de saída do modelo. Esse método apresenta alta efetividade para medições do valor real frequentes. No caso de análises laboratoriais, onde a frequência é baixa, indica-se a utilização de técnicas mais avançadas como Mínimos Quadrados Recursivos ou Filtro de Kalman Estendido, onde os parâmetros do modelo são alterados assim que os novos resultados reais são obtidos (FACCHIN, 2005).

3 Materiais e Métodos

Nesta seção, será apresentado o estudo de caso que será utilizado para o desenvolvimento do trabalho, e as variáveis tipicamente disponíveis para este sistema. Também será discutida a metodologia utilizada para obtenção dos dados utilizados para a modelagem do analisador virtual, o pré-tratamento da base de dados e o método de obtenção dos modelos de analisador virtual de umidade.

3.1 Estudo de Caso – Unidade de Desidratação de Gás Natural

Trata-se de uma unidade de *Temperature Swing Adsorption* (TSA), utilizando peneiras moleculares como adsorventes, para a remoção de umidade do gás natural em uma plataforma *off-shore* (DE MARCO, 2019).

3.1.1 Processo TSA para desidratação de gás natural

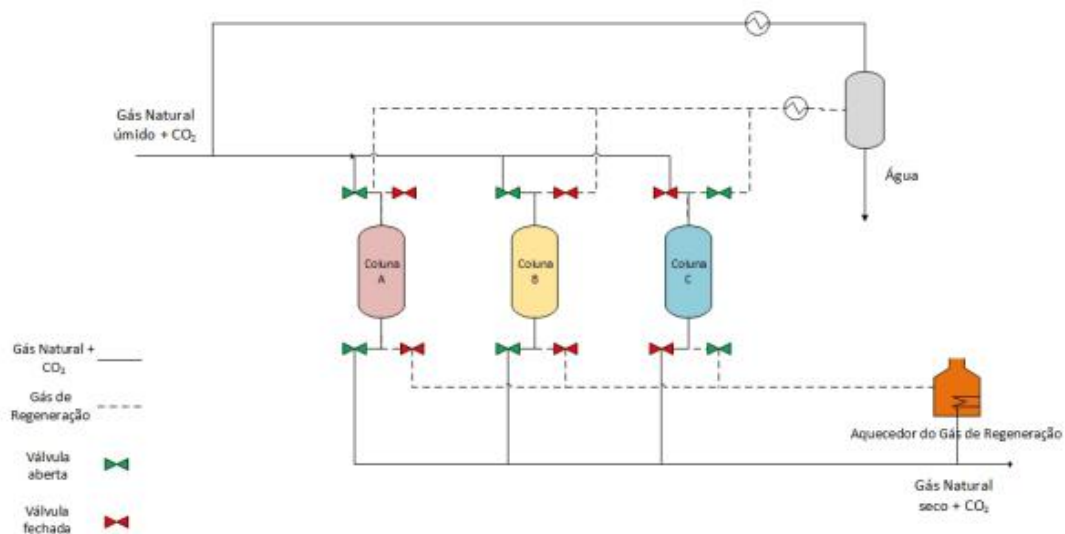
O método utilizado para a remoção de umidade em frações leves de petróleo que será abordado neste trabalho é o de *Temperature Swing Adsorption* (TSA), utilizando peneiras moleculares como adsorventes. Este processo é largamente utilizado para separação de misturas em unidade *offshore* de produção de gás natural (SANTOS, 2016).

Esse tipo de fracionamento pode ser dividido em duas etapas: a de adsorção, onde ocorre a remoção da fração indesejada a temperaturas baixas, neste caso a água, e o processo de dessorção ou regeneração, onde são removidos os contaminantes do material adsorvente a altas temperaturas. A etapa de regeneração é a etapa lenta e determinante deste processo, e a adsorção só recomeça após o resfriamento do leito à temperatura correta de trabalho (DE MARCO, 2019).

Dessa forma, a configuração para produção contínua nesse método é utilização de no mínimo duas colunas em paralelo. Enquanto uma adsorve na temperatura de trabalho, outra é regenerada a altas temperaturas (SANTOS, 2016).

No ciclo de processamento de gás natural, a remoção de umidade é essencial para a especificação do gás e vida útil da unidade produtiva. Dessa forma, a seção de desidratação assume um importante papel na produção de gás natural. A representação gráfica simplificada do processo pode ser verificada na Figura 3.1, para uma unidade com 3 colunas de adsorção.

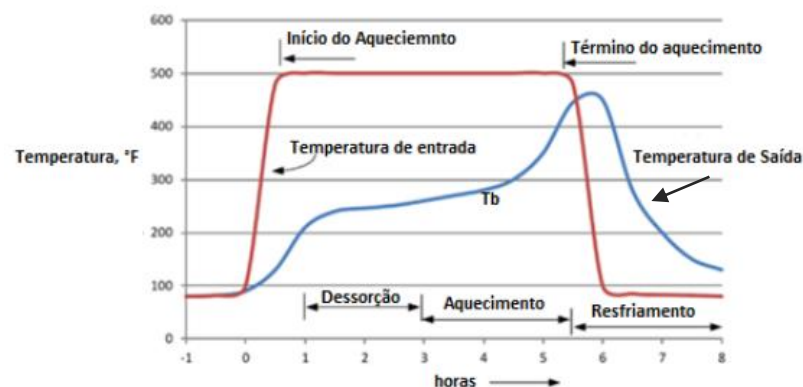
Figura 3.1 – Unidade de TSA para desidratação do GN com 3 colunas utilizando peneiras moleculares como adsorventes.



Fonte: DE MARCO (2019)

O processo conta com duas colunas operando a temperaturas em torno de 40°C, especificando o gás processado, enquanto uma parcela deste gás, cerca de 5 a 10% é aquecido e utilizado como gás seco para o ciclo de regeneração. Neste, o gás de regeneração passa em contra fluxo pela coluna e é resfriado por um permutador utilizando água de resfriamento. Após esse processo, é realizado um *flash* para remoção da água como fração líquida, e a fase gás retorna para a entrada, junto ao gás úmido (SANTOS, 2016). Na Figura 3.2, é possível verificar uma curva de temperatura típica do sistema.

Figura 3.2 – Curva de temperatura típica para uma planta de adsorção operando em ciclo de regeneração.



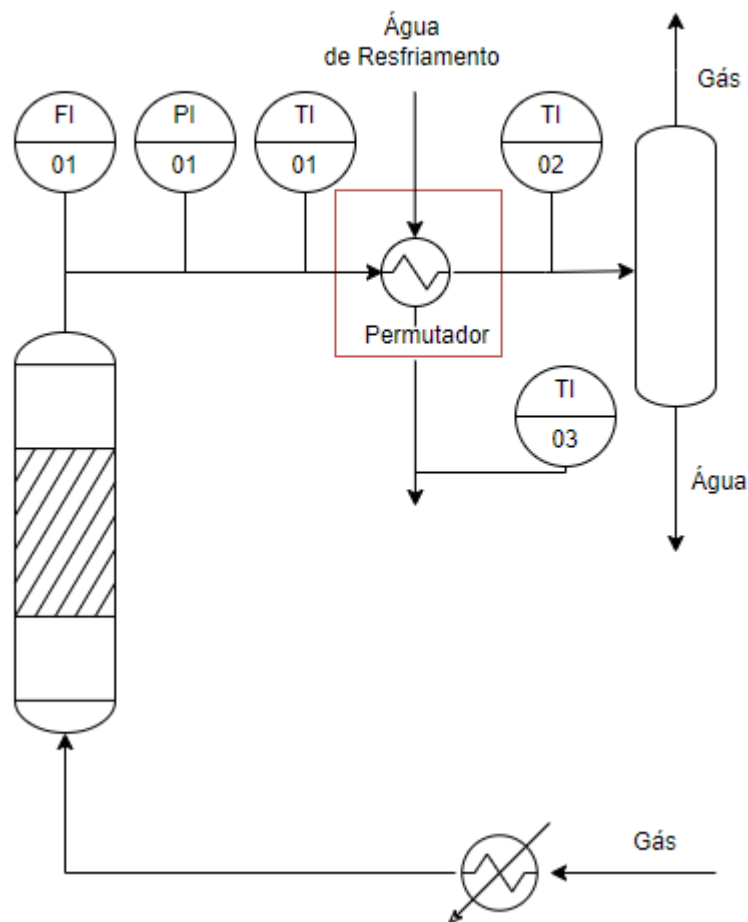
Fonte: MOKHATAB et al. (2015)

Este processo opera com tempos fixos para as etapas de adsorção e regeneração, o que ocasiona uma degradação dos leitos de forma mais rápida em função da

exposição do adsorvente a altas temperaturas de forma contínua por um período mais longo que o necessário. O término das etapas de regeneração e de adsorção de forma precipitada ou atrasada provoca a diminuição na vida útil do leito utilizado. Dessa forma é necessária a otimização deste processo (DE MARCO, 2019).

Além disso, nos processos de TSA para remoção de umidade em gás natural, em muitas unidades não é realizada a medição da umidade na saída dos leitos. Nesse sentido, a única forma de determinação do encerramento do ciclo é pautada por outras variáveis de processo. O permutador de resfriamento, no fim do ciclo de regeneração, é o equipamento que possui um número maior de medições que podem ser utilizadas para esta determinação como apontado na Figura 3.3.

Figura 3.3 – Condensador em sistema de regeneração TSA em unidade de processamento de gás natural .



Fonte: Autor

3.2 Variáveis com Medição Disponível

Considera-se para fins de estudo, que a unidade de processamento TSA de gás natural utilizada como estudo de caso possui as seguintes variáveis apresentadas na Tabela 3.1.

Tabela 3.1 – Medições tipicamente disponíveis em unidades TSA e apresentadas na Figura 3.3

Nº	TAG	VARIÁVEL
1	FI-01	Vazão do Gás Natural antes do permutador
2	PI-01	Pressão do Gás Natural antes do permutador
3	TI-01	Temperatura do Gás Natural antes do permutador
4	TI-02	Temperatura do Gás Natural depois do permutador
5	TI-03	Temperatura da Água de Resfriamento depois do permutador

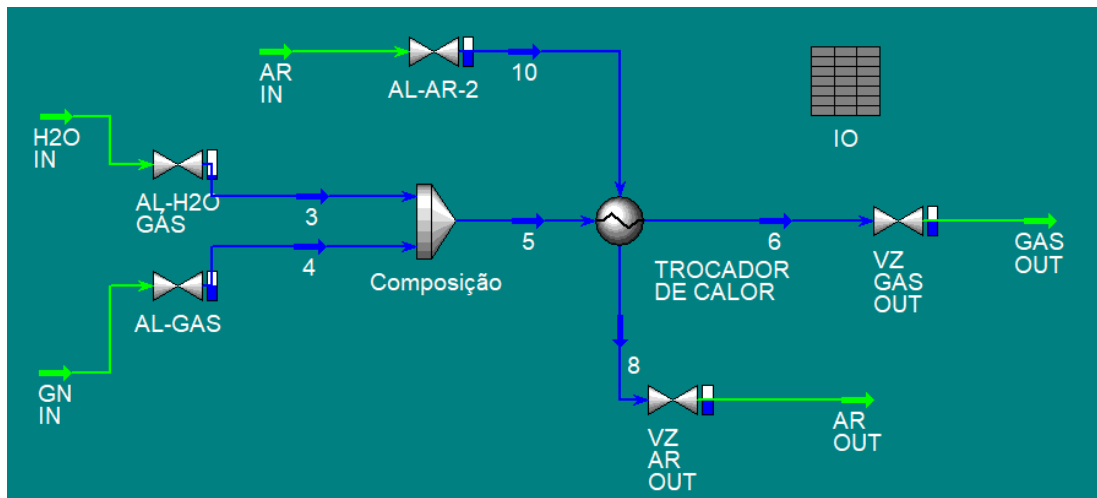
3.3 Desenvolvimento do Analisador Virtual

O desenvolvimento e consolidação dos dados utilizados no presente trabalho foram gerados na linguagem de programação Python 3.5, seguindo o procedimento sugerido por KADLEC *et al.* (2009) para o desenvolvimento dos modelos.

3.3.1 Aquisição dos dados de processo

A base de dados utilizada foi gerada a partir da modelagem do sistema de resfriamento de gás natural utilizado na etapa de regeneração (TSA), apresentada no tópico 3.1.1. Os cenários foram simulados utilizando o *UniSim Design R470*, um robusto software comercial que permite a criação de modelos estáticos e dinâmicos com alta acurácia para projetos de sistemas. Foi realizada uma modelagem dinâmica e o arranjo pode ser observado na Figura 3.4.– Arranjo utilizado na simulação do sistema de resfriamento

Figura 3.4 – Arranjo utilizado na simulação do sistema de resfriamento



Fonte: Autor

De forma a obter dados coerentes e próximos a dados de planta, foram utilizados dados típicos referentes a esse processo, encontrados na literatura. O modelo foi idealizado pensando em realizar variações de composição de água. Dessa forma, foi dividida a V_{gn} em GN IN contendo a parcela de gás e H2O IN contendo a parcela de água entrando. As frações são direcionadas a um misturador sem perda de carga controladas pelas válvulas nas respectivas correntes. O permutador utiliza água de resfriamento como fluido frio, representado na simulação como AR IN. Foi realizada a variação dos parâmetros de entrada de pressão, temperatura e composição do gás de entrada, cujos dados de entrada podem ser verificados na Tabela Tabela 3.2 – Dados de entrada das correntes no simulador. Foram avaliados 100.000 cenários de condições de processo.

Tabela 3.2 – Dados de entrada das correntes no simulador

CORRENTE	GN IN	H2O IN	AR IN
PRESSÃO (KPA)	4.000 – 4.900	4.000 – 4.900	500
TEMPERATURA (°C)	50-150 °C	50-150 °C	20 °C
COMPOSIÇÃO MÁSSICA	0,9 CH ₄ , 0,06 C ₂ H ₆ , 0,02 CO ₂ , 0,02 C ₃ H ₈	Água pura	Água pura

No trocador de calor, foi utilizada perda de carga de 70 kPa nos tubos e 10 kPa no casco. O misturador, utilizado como uma ferramenta para a simulação dinâmica, foi considerado como ideal, sem perda de carga associada.

A base de dados foi formada utilizando a simulação dinâmica do software e programação em Python para realizar as variações e armazenamento de dados, utilizando comunicação serial. O sistema partiu de um estado estacionário para evitar problemas decorrentes dessas variações transientes iniciais. Para variar a composição de água no do gás foi alterado o percentual de abertura da válvula AL-H₂O-GÁS até 40%. A pressão foi variada no intervalo de 4000 a 4900 kPa e a temperatura das correntes de entrada do gás de 50 a 150 °C, para simulação do aquecimento da coluna em processo de regeneração.

3.3.2 Pré-tratamento dos dados

Como descrito por KADELC *et al.* (2009), a etapa de pré-tratamento dos dados é importante para o melhor entendimento do conjunto de dados e verificação prévia de inconsistências que possam interferir na qualidade do modelo. A base de dados foi avaliada utilizando a biblioteca *Pandas 2.0.3* e *Statsmodels 0.14.0*, verificando a correlação linear entre as variáveis, dados faltantes e desvios entre os dados.

Em muitos casos, se faz necessário a utilização de normalizações na base de dados para uma melhoria de desempenho dos algoritmos de aprendizado de máquina. Essa normalização ocorre quando os dados de entrada possuem diferenças muito grandes na escala que estão.

A saída da simulação possuía como padrão o sistema internacional de unidades para os dados. Esses possuíam diferenças significativas de escala entre si, principalmente nos seguintes parâmetros: Fração de água (y_{H_2O}) (alterada para percentual), P_{GN} alterada para bar e V_{GN} (alterada para t/h). Dessa forma, todos os parâmetros possuíam a mesma escala, e desvios relacionados à normalização foram corrigidos.

3.3.3 Seleção, treinamento e validação do modelo

Os modelos de regressão linear e rede neural foram arquitetados utilizando a biblioteca *Scikitlearn 1.3.0*, biblioteca de código aberto muito utilizada para o

desenvolvimento de modelos de *machine learning*. Para a rede neural foi utilizado o modelo de redes neurais multicamadas *Perceptron*.

Foi realizada uma análise preliminar para identificar a melhor configuração dos modelos e utilizada com premissa inicial. O primeiro passo foi a divisão da base de dados em dados de treinamento e dados de teste dos modelos, utilizados para a validação dos arranjos propostos. O parâmetro de parcela utilizada para treino e validação foi um dos selecionados para otimização das redes neurais e foi utilizado o melhor valor para o método de regressão linear multivariável conforme LINHARES (2010).

Então, foram selecionados parâmetros de ajuste para a otimização do modelo, variando um dos parâmetros enquanto os demais eram fixados. Os parâmetros utilizados para a otimização foram: função de ativação, porcentagem de amostra utilizado para teste, número de neurônios na primeira camada e número de neurônios na segunda camada. Os modelos foram treinados utilizando 1000 iterações, podendo apresentar variação em até 100 iterações. Isso proporcionou valores mais assertivos para o modelo. Foram realizados três treinamentos por modelo, obtendo valores muito próximos, e selecionados o maior valor como resultado.

A validação do modelo foi realizada utilizando o parâmetro R^2 , conforme detalhado na seção 2.2.2, o qual avalia os desvios dos dados preditos pelo modelo em relação aos dados reais. Como os valores apresentaram baixa correlação linear em relação a variável de interesse não foi avaliada a utilização progressiva do número de variáveis, modelando com todas as variáveis disponíveis.

4 Resultados

Nesta seção serão apresentados os resultados obtidos na realização do presente trabalho. Foram discutidos o pré-tratamento de dados e apresentados os modelos obtidos com as respectivas otimizações.

A Tabela Tabela 4.1 apresenta as abreviações utilizadas para a denominação das variáveis da base de dados na seção de resultados, bem como as respectivas unidades.

Tabela 4.1 – Tabela de títulos

Abreviatura	Descrição	Unidade
Frac. H2O	Fração mássica de água presente no gás de entrada	%
Temp. GN out	Temperatura de saída do gás natural do permutador	°C
Temp. AR out	Temperatura de saída da água de resfriamento do permutador	°C
Pressão Ent.	Pressão do gás natural na entrada do permutador	bar
Temp. GN in	Temperatura de entrada do gás natural no permutador	°C
Vazão Ent.	Vazão mássica do gás natural na entrada do permutador	t/h

4.1 Pré-tratamento dos dados

Conforme descrito na seção de metodologia, a primeira etapa para o desenvolvimento do modelo do analisador virtual é a verificação da base de dados, para evitar desvios do modelo referente a base de dados. Dessa forma, a análise preliminar realizada pode ser visualizada na Tabela Tabela 4.2. A matriz é composta por 100.000 linhas de dados e não possui dados faltantes, já que foram obtidas através de um software de simulação.

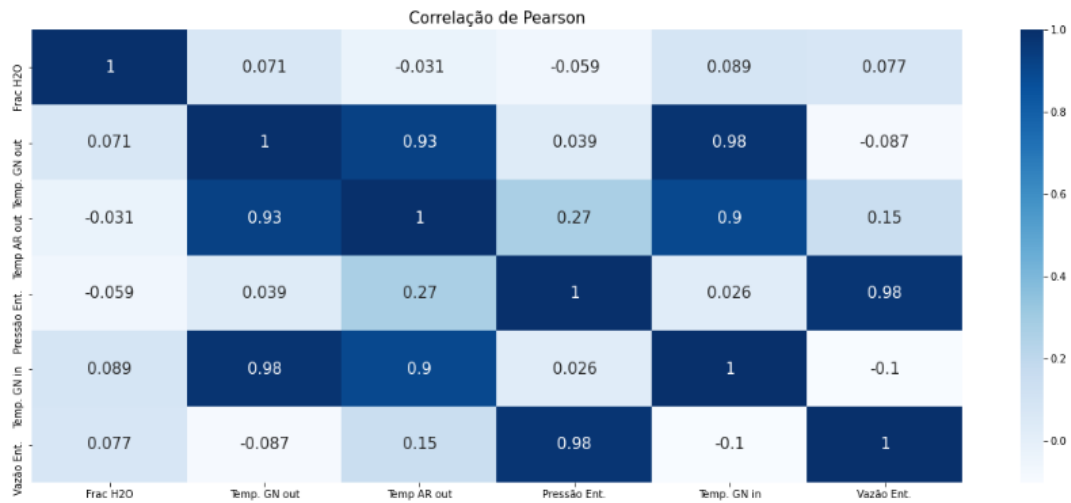
Tabela 4.2 – Análise preliminar da base de dados obtidos na simulação

	Frac H2O	Temp. GN out	Temp AR out	Pressão Ent.	Temp. GN in	Vazão Ent.
Quantidade	100.000	100.000	100.000	100.000	100.000	100.000
Média	21,92	83,09	25,28	44,13	87,04	25,54
Erro Padrão	10,29	20,47	1,76	2,72	21,30	6,84
Valor Mínimo	0,71	43,29	22,28	39,85	47,09	11,89
25%	13,52	65,80	23,73	41,78	69,11	19,96
50%	23,32	84,57	25,27	43,80	88,57	25,98
75%	30,82	101,05	26,80	46,51	105,48	31,26
Valor Máximo	38,78	118,47	28,46	48,65	142,55	39,88

Para realizar uma avaliação de como as variáveis estão associadas entre si, foi gerada uma matriz de correlação de Pearson, apresentada na Figura 4.1. Essa matriz avalia a relação linear entre duas variáveis contínuas. A primeira coluna apresenta as correlações entre as frações de água e demais variáveis utilizadas.

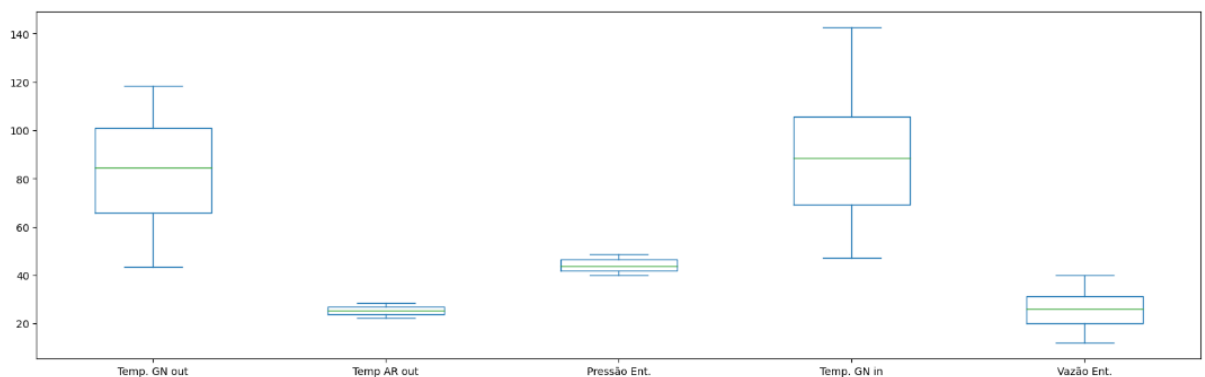
Observando os resultados obtidos, verifica-se na primeira coluna valores bem próximos a zero, o que representa baixa correlação linear das variáveis individualmente em relação a fração mássica de água, a variável de interesse. Dessa forma, é possível inferir que será necessário um modelo de alta complexidade para a realização da predição da fração mássica de água.

Figura 4.1 – Correlação de Pearson entre os parâmetros da base de dados.



Fonte: Autor

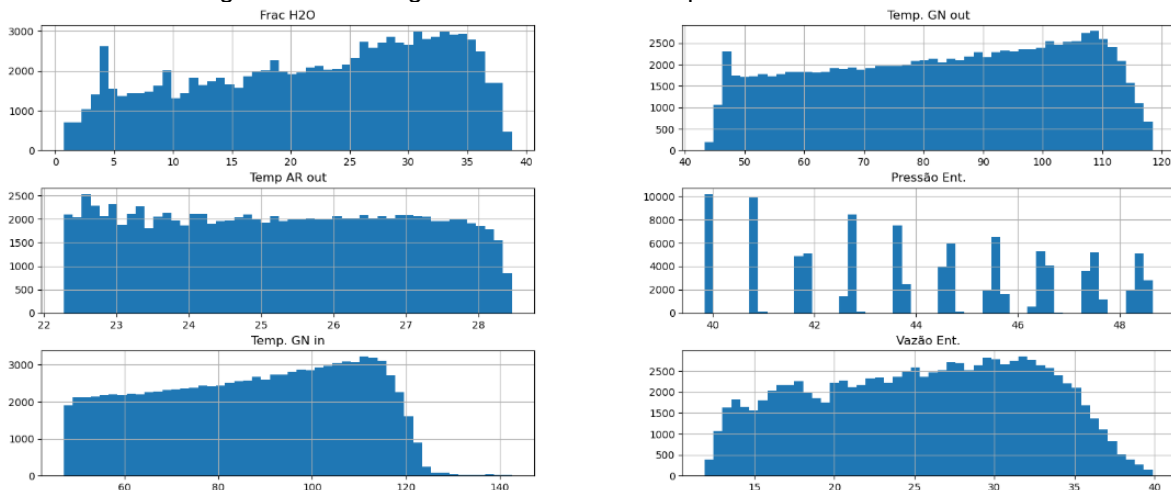
A Figura 4.2 fornece uma análise visual da dispersão dos dados presentes na base de dados e serve como parâmetro para remoção de valores atípicos. Nesta avaliação, a linha verde representa a mediana dos dados daquela variável, os retângulos representam o segundo e terceiro quartis (que representam 50%), dos dados obtidos e as linhas acima e abaixo representam os valores de variação tolerável para que os resultados sejam válidos. Como pode ser observado, as variáveis de temperatura do gás são as mais afetadas. Essa informação é coerente uma vez que ficam no intervalo proposto de avaliação do modelo, a curva de aquecimento da coluna. Usualmente essa etapa da análise resulta na remoção de 5% da amostra. Porém, como a fonte de dados é o simulador comercial, com modelagem caixa branca, assumiu-se que todo o espectro de valores representa valores válidos, uma vez que o sistema partiu de um estado estacionário.

Figura 4.2 – *Boxplot* das variáveis disponíveis na base de dados.

Fonte: Autor

A Figura 4.3 apresenta a distribuição e recorrência dos valores presentes na base de dados, separada pelas variáveis utilizadas no processo de modelagem do analisador virtual. Essa análise mostra a simetria do número de amostras por concentração de água para cada variável. Devido a boa distribuição espera-se valores confiáveis para a faixa observada.

Figura 4.3 – Histograma das variáveis disponíveis na base de dados.



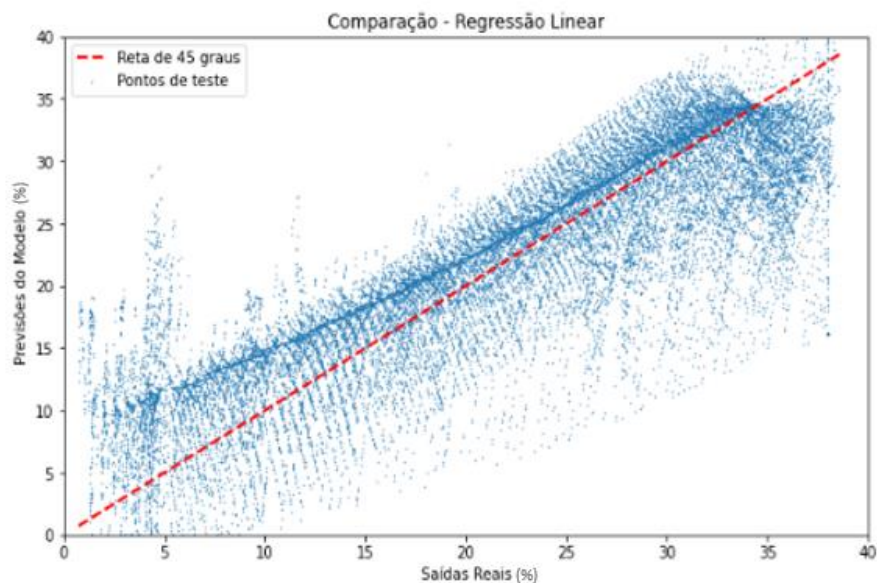
Fonte: Autor

4.2 Modelo de Regressão Linear Multivariável (MRL)

Como apontado na seção 4.1, era esperado um baixo valor de R^2 para o ajuste do modelo de regressão linear, devido aos baixos valores de correlação linear entre as variáveis independentes e a dependente (fração de água). Essa informação foi validada com o ajuste dos dados, onde foi obtido um coeficiente de determinação (R^2) de **0,7322** quando comparado os valores preditos utilizando as variáveis de entrada da fração de dados para teste com os valores esperados. Essa avaliação pode ser observada na Figura 4.4, onde os valores em azul representam o par de previsão pelo modelo e valor real. Idealmente, quanto mais próximo da reta vermelha com inclinação de 45° , mais bem ajustado está o modelo.

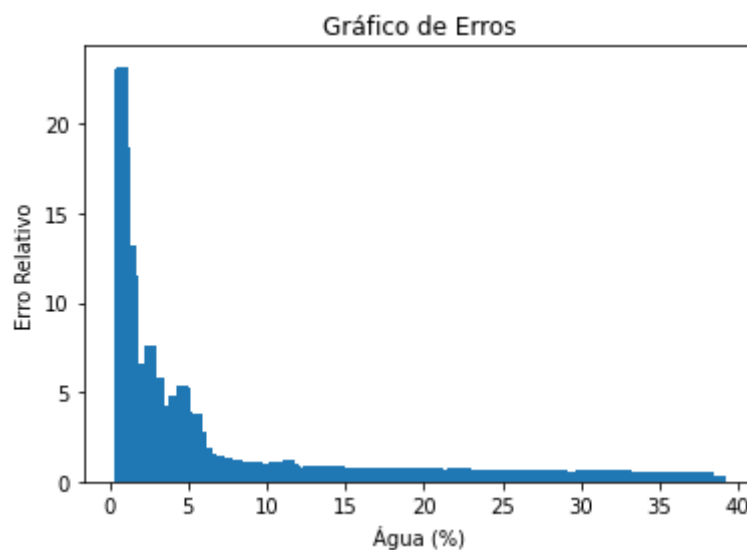
Para avaliar a magnitude do desvio, foi plotado um gráfico do valor de erro por amostra apresentado na Figura 4.5. O erro relativo médio foi de 0,4041% para o modelo de regressão linear, porém verifica-se um crescente desvio quando os percentuais de água vão em direção a 0.

Figura 4.4 – Comparação entre os valores preditos pelo modelo de regressão linear e valores reais em percentual.



Fonte: Autor

Figura 4.5 – Comparação entre o erro relativo e a fração de água no modelo MRL.



Fonte: Autor

O modelo de regressão linear multivariável apresentou um baixo valor de R^2 e com isso, se fez necessário o desenvolvimento de um analisador com um método mais robusto. O modelo ajustado pode ser verificado na equação 7.

$$\begin{aligned} \text{Frac. } H_2O = & 0,2949 \text{ Temp. GN out} - 4,9466 \text{ Temp. AR out} & (7) \\ & - 15,8843 \text{ Pressão Ent.} + 0,4043 \text{ Temp. GN in} \\ & + 6,6647 \text{ Vazão Ent.} \quad + 617,9953 \end{aligned}$$

4.3 Modelo de Redes Neurais Artificiais Multicamadas (ANN)

O primeiro hiperparâmetro otimizado foi o de função de ativação. Conforme a Tabela Tabela 4.3, o tipo sigmóide foi o que obteve o maior coeficiente de determinação, de 0,8410.

Tabela 4.3 – Otimização da função de ativação utilizada para a modelagem utilizando redes neurais.

Função de ativação	Tamanho da amostra	Neurônios 1 ^a Camada	Neurônios 2 ^a Camada	R ²
Linear	25%	5	0	0,7247
Sigmoide	25%	5	0	0,8410
Tangente Hiperbólica	25%	5	0	0,8370
Relu	25%	5	0	0,7853

Para a avaliação do tamanho da amostra foram avaliados valores de 10 a 40% e verificado que o tamanho da amostra de teste de 25% era o mais adequado para a elaboração do modelo, obtendo um coeficiente de determinação de 0,8410. Esta análise pode ser observada na Tabela 4.4.

Tabela 4.4 – Otimização do tamanho da amostra utilizada para a modelagem utilizando redes neurais.

Função de ativação	Tamanho da amostra	Neurônios 1 ^a Camada	Neurônios 2 ^a Camada	R ²
Sigmoide	10%	5	0	0,8051
Sigmoide	15%	5	0	0,8251
Sigmoide	25%	5	0	0,8410
Sigmoide	35%	5	0	0,7651
Sigmoide	40%	5	0	0,7202

Ao avaliar o número de neurônios na primeira camada, apresentado na Tabela 4.5, observa-se que com o aumento do número houve uma melhora significativa até um certo ponto e depois um decréscimo de eficiência. Este fenômeno é denominado *overfitting*. Esse efeito ocorre quando um modelo de aprendizado de máquina se ajusta muito bem aos dados de treino, porém, não obtém um bom desempenho ao prever os novos dados. Isso ocorre quando um modelo se torna excessivamente complexo em relação à quantidade de dados disponíveis.

É possível verificar que com o uso de 125 neurônios na primeira camada oculta é obtido o maior valor para o coeficiente de determinação, de 0,9517.

Tabela 4.5 – Otimização do número de neurônios utilizados na primeira camada oculta para a modelagem utilizando redes neurais.

Função de ativação	Tamanho da amostra	Neurônios 1 ^a Camada	Neurônios 2 ^a Camada	R ²
Sigmoide	25%	5	0	0,8410
Sigmoide	25%	25	0	0,9387
Sigmoide	25%	125	0	0,9517
Sigmoide	25%	200	0	0,9212
Sigmoide	25%	500	0	0,9483

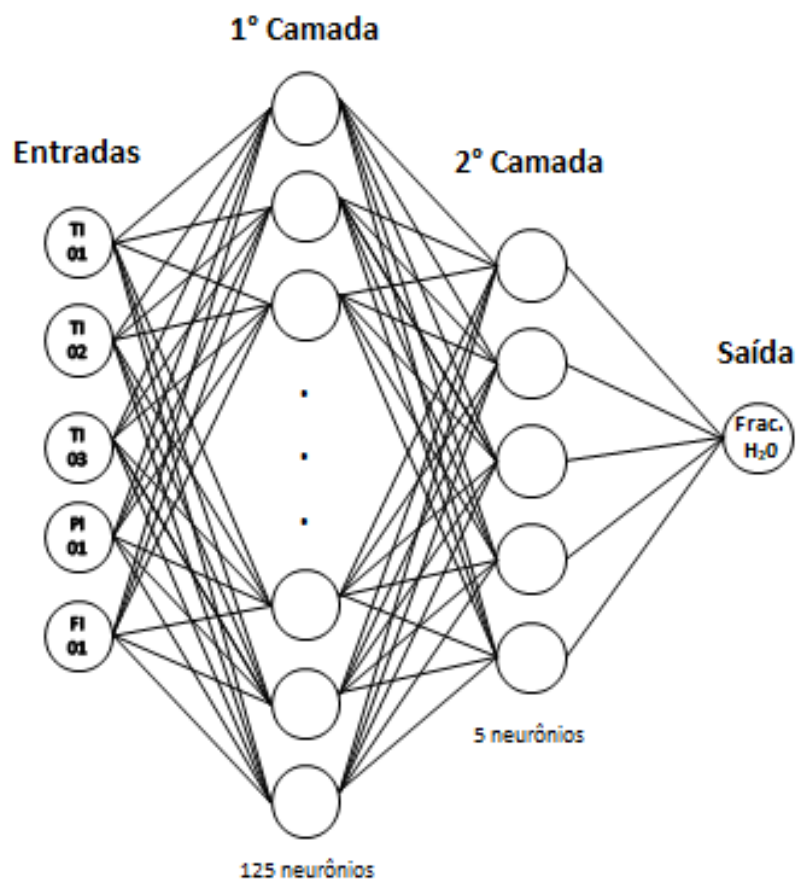
A Tabela Tabela 4.6 mostra a diminuição do ajuste pelo coeficiente de determinação ao aumentar o número de neurônios. Isso se deve ao fenômeno de *overfitting* e, verifica-se que, com o uso de 5 neurônios na segunda camada oculta o valor para o coeficiente de determinação é de 0,9727.

Tabela 4.6 – Otimização do número de neurônios utilizados na segunda camada oculta para a modelagem utilizando redes neurais.

Função de ativação	Tamanho da amostra	Neurônios 1 ^a Camada	Neurônios 2 ^a Camada	R ²
Sigmoide	25%	125	0	0,9517
Sigmoide	25%	125	5	0,9727
Sigmoide	25%	125	25	0,9587
Sigmoide	25%	125	125	0,9543

Dessa forma, o modelo arquitetado que possuiu o melhor valor utilizado para a comparação foi desenvolvido utilizando a função de ativação sigmóide, amostra de teste de 25%, 125 neurônios na primeira camada e 5 neurônios na segunda camada, possuindo um R^2 de 0,9727 pode ser observado na Figura 4.6. Esse valor indica uma boa capacidade do modelo não linear de prever resultados reais.

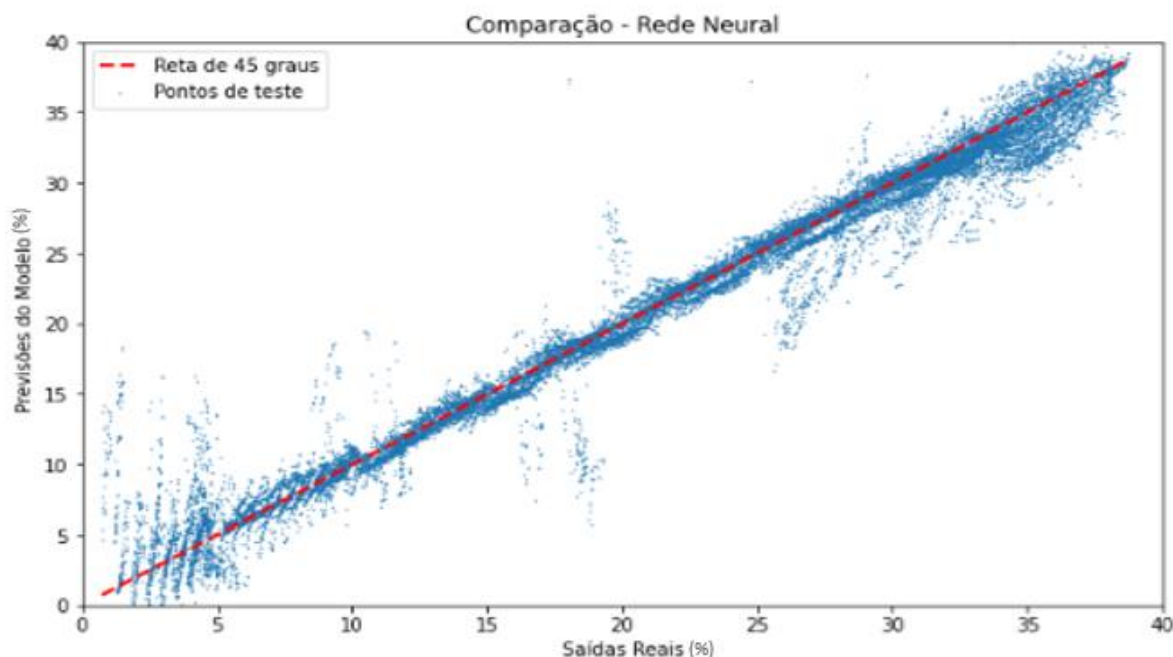
Figura 4.6 – Modelo de predição utilizando redes neurais feedforward



Fonte: Autor

De maneira análoga ao modelo de regressão linear, foi plotado um gráfico de valores preditos pelo modelo arquitetado de redes neurais artificiais e valores reais em percentual. A Figura 4.7 apresenta essa comparação.

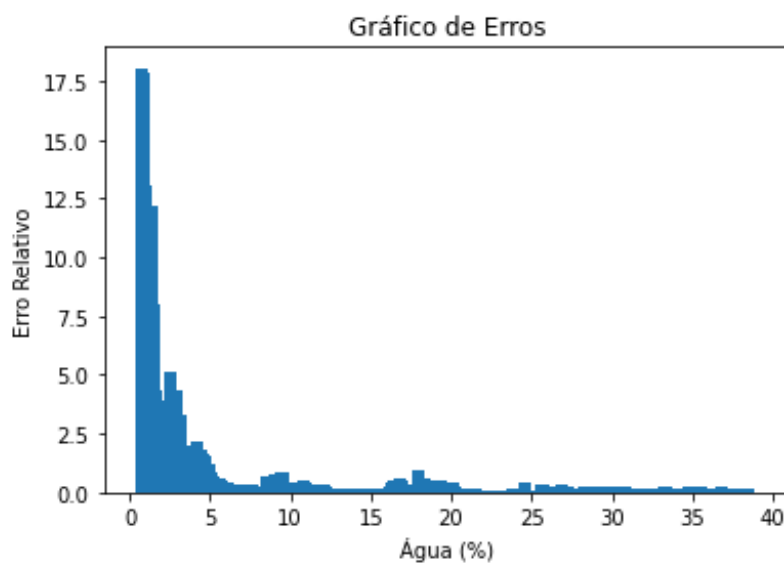
Figura 4.7 - Comparação entre os valores preditos pelo modelo de redes neurais e valores reais.



Fonte: Autor

Para efeitos de comparação foi avaliada a magnitude do desvio para o modelo de redes neurais apresentado na Figura 4.8. O valor do erro relativo médio para este modelo foi de 0,1230%. Verifica-se que o modelo tende a apresentar mais dispersão em valores próximos a zero, o que pode interferir na confiabilidade do analisador para baixas concentrações, que ocorrem no fim de cada ciclo.

Figura 4.8 – Comparação entre o erro relativo e a fração de água no modelo RNA.



Fonte: Autor

4.4 Comparação dos modelos desenvolvidos

Ao comparar os resultados obtidos pelos melhores modelos desenvolvidos na etapa de validação, verifica-se que o método MRL apresenta valores de R^2 de 0,7322 erro relativo médio de 0,4041%. Enquanto o modelo de redes neurais apresentou R^2 de 0,9727 e erro relativo médio e 0,1230%. É possível inferir que o modelo de redes neurais apresenta um melhor ajuste para o sistema proposto, obtendo uma assertividade maior em relação ao modelo de regressão linear.

A Figura 4.9 apresenta os módulos de erros relativos entre previsão dos modelos e valores reais. É possível observar que, ambos os modelos, apresentam boas aproximações para altas concentrações de água. Porém, conforme o percentual de água diminui, os modelos perdem sensibilidade, sendo que o modelo de redes neurais possui um desvio menor durante todo o *range* de avaliação com desvio menor que o modelo de regressão linear.

A Tabela 4.7 apresenta resultados obtidos por outros pesquisadores com os métodos de modelagem para diferentes aplicações para fins de comparação.

Figura 4.9 - Comparação do erro relativo entre os modelos e a fração de água.

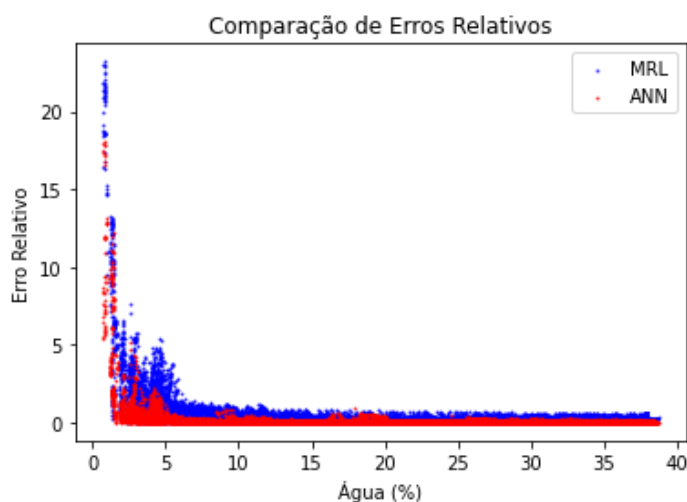


Tabela 4.7 – Resultados de coeficiente de determinação para modelos de analisadores virtuais obtidos na literatura.

Autores	Método	Aplicação	R2 obtido
ROGINA et al. (2011)	MRL	RVP da Nafta	0,48
ROGINA et al. (2011)	ANN	RVP da Nafta	0,85
KURNIAWAN et al. (2023)	ANN	Teor de O ₂ em combustão de caldeira	0,99

Percebe-se que os valores obtidos para o coeficiente de determinação apresentam valores coerentes. O método de MRL possui limitações quando as variáveis de entrada possuem um baixo valor de correlação linear com a variável de interesse. Essa limitação se torna mais acentuada em sistemas complexos, como o apresentado por ROGINA et al. (2011), que apresentou valores ainda mais baixos para predição de pressão de vapor REID da nafta utilizando esta técnica.

As redes neurais apresentam uma alternativa para a modelagem de sistemas complexos, tendo maior diversidade de arranjos e possibilitando a otimização para cada sistema, mesmo que não linear. Os valores obtidos na determinação de umidade em gás natural desenvolvidos neste trabalho apresentam R^2 superiores aos apresentados por ROGINA et al. (2011) em ambas técnicas. Porém, quando comparados com os valores de redes neurais para predição de teor de oxigênio em combustão em caldeiras apresentam um resultado maior para o coeficiente de determinação.

5 Conclusões e Trabalhos Futuros

No presente trabalho, foram desenvolvidos e otimizados dois modelos preditivos de umidade em coluna de desidratação de gás natural, utilizando os parâmetros de Temp. GN out, Temp. AR out, Pressão Ent., Temp. GN in, Vazão Ent. O primeiro, utilizando o método de regressão linear multivariada, obteve resultados de coeficiente de determinação de 0,7322 e erro relativo médio de 0,4041%. O segundo modelo foi desenvolvido utilizando a técnica de rede neural do tipo feedforward, apresentando coeficiente de determinação de 0,9727 e erro relativo médio de 0,1230%. O analisador baseado em redes neurais foi otimizado utilizando uma combinação de 25% da base de dados para validação, função de ativação do tipo sigmóide, 125 neurônios na primeira camada e 5 na segunda camada.

Além disso, foi realizada com sucesso a modelagem dinâmica do sistema de resfriamento de gás natural, utilizando dados típicos de processo TSA. Para a coleta e análise de dados, foi utilizada a linguagem Python, que possibilitou a obtenção da ampla base de dados necessária para a fabricação dos modelos de predição de umidade.

Ambos os modelos apresentaram resultados com baixo valor de erro associado para concentrações de água maiores que 5%. Perdendo, porém, confiabilidade nos dados quando muito próximos a zero. O modelo de regressão linear multivariada apresentou um maior desvio em função do baixo coeficiente linear entre as variáveis, enquanto o modelo de redes neurais por possuir um modelo não linear e aprendizado de máquina, apresentou valores de erro menores, comparados à ao modelo MRL, mesmo tendendo a zero.

Sugere-se para trabalhos futuros, a modelagem do analisador virtual utilizando dados reais de planta, verificando disponibilidade de mais variáveis de processo para melhorar a predição do modelo. Pode-se ainda, verificar a utilização de outros tipos de técnicas de modelagem como filtro de kalman ou *decision tree*.

REFERÊNCIAS

BELISÁRIO, A. B. **Análise de emissões em caldeiras de recuperação química de fábricas de celulose kraft: predição e análise de sensibilidade com redes neurais artificiais.** 2020. Dissertação (Mestrado em Engenharia Química) – UFMG, Universidade Federal de Minas Gerais, Belo Horizonte/MG.

BOMFIM, B. P.; BRATIFICH, R.; DA SILVA, M. M.; SILVA, H. G. **Previsão de consumo de energia utilizando rede neural com retardo de tempo (TDNN).** *Colloquium Exactarum*, 2020. Vol.12 N.4.

CAVALCANTE JUNIOR, C. J. **Modelagem e simulação do processo de desidratação de gás natural em peneiras moleculares do tipo 4A.** 2023. Dissertação (Mestrado em Engenharia Química) – USP, Escola Politécnica da Universidade de São Paulo, São Paulo/SP.

CHEN, J. M.; CHEN, B. S. **System parameter estimation with impt/output noisy data and missing measurements.** *IEEE transactions on Signal Processing*, 2000. v. 48, n. 6.

DE MARCO, L. M. **Determinação do tempo de vida útil remanescente em processos cíclicos.** 2019. Dissertação (Mestrado em Engenharia Química) – UFRGS, Universidade Federal do Rio Grande do Sul, Porto Alegre/RS.

DENN, M. **Process modelling.** New York: Longman, 1986.

GELADI, P.; KOWALSKI, B. R. **Partial least-squares regression - a tutorial.** *Analytica Chimica Acta*, 1986. v. 185.

FACCHIN, S. **Técnicas de análise multivariável aplicadas ao desenvolvimento de analisadores virtuais.** 2005. Dissertação (Mestrado em Engenharia Química) – UFRGS, Universidade Federal do Rio Grande do Sul, Porto Alegre/RS.

FORTUNA, L.; GRAZIANI, S.; XIBILIA, M. G. **Soft sensors for product quality monitoring in debutanizer distillation columns.** *Control engineering practice*, 2005. v. 13, n. 4.

GONZAGA, C. S. B. **Uma metodologia monte carlo para projeto de processamento offshore de gás natural.** 2014. Dissertação (Mestrado em Tecnologia de Processos Químicos e Bioquímicos) – UFRJ, Universidade Federal do Rio de Janeiro, Rio de Janeiro/RJ.

KADLEC, P.; GABRYS, B.; STRANDT, S. **Data-driven soft sensors in the process industry.** 2009. v. 33.

KURNIAWAN, E.D.; EFFENDY, N.; ARIF, A.; DWIANTORO, K.; MUDDIN, N. **Soft sensor for the prediction of oxigen content in boiler flue faz using neural networks and extrene gradiente boosting.** *Neural computing and Applications.* 2023.

LINHARES, L. L. S. **Sistema híbrido de inferência baseado em análise de componentes principais e redes neurais artificiais aplicado a plantas de processamento de gás natural.**

2010. Dissertação (Mestrado em Automação e Sistemas) – UFRN, Universidade Federal do Rio Grande do Norte, Natal/RN.

MOHR, T. **Desenvolvimento de um analisador virtual para uma planta de polietileno de alta densidade**. 2004. Dissertação (Mestrado em Engenharia Química) – UFRGS, Universidade Federal do Rio Grande do Sul, Porto Alegre/RS.

MOKHATAB S.; POE, W. A.; SPEIGHT, J. G. *Handbook of Natural Gas Transmission and Processing*. 3^a. Redigerad av Elsevier. GPP- Gulf Professional Publishing, 2015.

OLIVEIRA, L. C. **Desenvolvimento de analisadores virtuais e sua aplicação na predição do ponto final de ebulição da nafta de craqueamento**. 2010. Dissertação (Mestrado em Engenharia Química) – UFMG, Universidade Federal de Minas Gerais, Belo Horizonte/MG.

QIN, S. J. *Neural networks for intelligent sensors and control: practical issues and some solutions*. *Neural Systems for Control*. 1997. n. 4.

ROCHA, C. C. M. **Análise termodinâmica da captura e reinjeção de CO₂ no processamento de gás natural em uma instalação de produção de petróleo offshore**. 2015. Dissertação (Mestrado em Engenharia Mecânica) – PUC-Rio, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro/RJ.

ROGINA A.; SISCO, I.; MOHLER, I.; UJEVIC, Z; BOLF, N. *Soft sensor for continuous product quality estimation (in crude distillation unit)*. *Chemical Engineering Research and Design*. *IChemE*. 2011.

SANTOS, M. G. R. S. **Desidratação de gás natural com alto teor de CO₂ em peneira molecular: simulação e aspectos de comportamento de fase**. 2016. Dissertação (Mestrado em Tecnologia de Processos Químicos e Bioquímicos) – UFRJ, Universidade Federal do Rio de Janeiro, Rio de Janeiro/RJ.

SCHEFFER, J. *Dealing with missing data*. *Research Letter in the Information and Mathematical Sciences*. 2002. v. 3. n. 1.

SILVA, I. N. da; SPATTI, D. H.; FLAUZINO, R. A. **Redes neurais artificiais para engenharia e ciências aplicadas**. São Paulo: Artliber, v. 23, n. 5, 2010.

SOUZA, F. A. A.; ARAÚJO, R.; MENDES, J. *Review of soft sensor methods for regression applications*. *Chemometrics and Intelligent Laboratory Systems*, Elsevier, v. 152, p. 69–79, 2016.