

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
ESCOLA DE ENGENHARIA  
ENGENHARIA DE CONTROLE E AUTOMAÇÃO

**LEONARDO VEDUIM MARCUZZO - 00275631**

**DESENVOLVIMENTO DE MODELO  
NEURONAL PARA INFERÊNCIA DE  
ÍNDICE DE FLUIDEZ DE POLIETILENO  
EM REATOR SLURRY**

Porto Alegre  
2021

**LEONARDO VEDUIM MARCUZZO - 00275631**

**DESENVOLVIMENTO DE MODELO  
NEURONAL PARA INFERÊNCIA DE  
ÍNDICE DE FLUIDEZ DE POLIETILENO  
EM REATOR SLURRY**

Trabalho de Conclusão de Curso (TCC-CCA) apresentado à COMGRAD-CCA da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de *Bacharel em Engenharia de Controle e Automação*.

**ORIENTADOR:**

Prof. Dr. Pedro Rafael Bolognese Fernandes

Porto Alegre  
2021

**LEONARDO VEDUIM MARCUZZO - 00275631**

**DESENVOLVIMENTO DE MODELO  
NEURONAL PARA INFERÊNCIA DE  
ÍNDICE DE FLUIDEZ DE POLIETILENO  
EM REATOR SLURRY**

Este Trabalho de Conclusão de Curso foi julgado adequado para a obtenção dos créditos da Disciplina de TCC do curso *Engenharia de Controle e Automação* e aprovado em sua forma final pelo Orientador e pela Banca Examinadora.

Orientador: \_\_\_\_\_  
Prof. Dr. Pedro Rafael Bolognese Fernandes, UFRGS  
Doutor pela Universidade de Dortmund – Dortmund, Alemanha

Banca Examinadora:

Prof. Dr. Pedro Rafael Bolognese Fernandes, UFRGS  
Doutor pela Universidade de Dortmund – Dortmund, Alemanha

Prof. Dr. Marcelo Götz, UFRGS  
Doutor pela Universidade de Paderborn – Paderborn, Alemanha

Prof. Dr. Mário Roland Sobczyk Sobrinho, UFRGS  
Doutor pela Universidade Federal do Rio Grande do Sul – Porto Alegre, Brasil

---

Mário Roland Sobczyk Sobrinho  
Coordenador de Curso  
Engenharia de Controle e Automação

Porto Alegre, novembro de 2021.

## RESUMO

A modelagem de sistemas dinâmicos é uma área de interesse na indústria química, uma vez que modelos podem auxiliar no monitoramento e controle da produção. Atualmente, existem diversas técnicas que podem ser empregadas para modelar processos, cada uma com suas limitações e capacidade de representação. Uma estratégia que está em consonância com o avanço tecnológico promovido no ambiente industrial é a utilização de redes neurais, dada sua versatilidade de aplicação em diferentes problemas. Através do estudo de caso real de uma unidade de produção de polietileno, são propostos e analisados diferentes modelos neuronais para inferência de uma propriedade de qualidade das resinas. Visa-se, com o modelo, permitir a identificação de eventuais problemas no processo e realizar ações antecipadamente. Os resultados indicam que modelos desenvolvidos com redes neuronais convencionais (*feedforward*) não conseguem representar adequadamente a dinâmica do processo, ao passo que redes recorrentes possuem melhor capacidade de predição e podem, com as devidas considerações, ser empregadas para o problema em questão. Considerando a melhor estrutura de rede convencional obtida, obteve-se 0,338 para raiz do erro médio quadrático, ao passo que, para a recorrente, foi de 0,223. Também foi desenvolvido um modelo combinado, obtendo um valor de 0,208 para a melhor estrutura dentre as testadas.

**Palavras-chave:** Modelagem de processos, Indústria química, Redes neuronais, Polietileno.

## ABSTRACT

Modeling dynamic systems is an area of interest in the chemical industry, since models can assist in monitoring and controlling production. Currently, there are several techniques which can be employed to model processes, each with its own limitations and representational capabilities. One such strategy takes advantage of the versatility of neural networks to simulate complex systems, given their adaptability to different problems. This project proposes various different neural models for predicting the quality of polyethylene resins and evaluating their efficacy by comparing them to a real production unit. The objective of the model is to allow the identification of problems in the process and to carry out corrective action in advance. The results indicate that models developed with feedforward neural networks cannot adequately represent the dynamics of the process. However, recurrent neural networks have better predictive capacity and can, with the proper considerations, be employed for the problem at hand. Considering the best feedforward neural network obtained, the root mean square error was 0.338, while for the recurrent network it was 0.223. An ensemble model was also developed, obtaining a value of 0.208 for the best structure among those tested.

**Keywords:** Process modeling, Chemical industry, Neural networks, Polyethylene.

# SUMÁRIO

LISTA DE ILUSTRAÇÕES . . . . .	7
LISTA DE TABELAS . . . . .	8
LISTA DE ABREVIATURAS . . . . .	9
LISTA DE SÍMBOLOS . . . . .	10
<b>1 INTRODUÇÃO . . . . .</b>	<b>12</b>
<b>1.1 Motivação e objetivos . . . . .</b>	<b>12</b>
<b>1.2 Estrutura do trabalho . . . . .</b>	<b>13</b>
<b>2 REVISÃO BIBLIOGRÁFICA . . . . .</b>	<b>14</b>
<b>2.1 Polietileno . . . . .</b>	<b>14</b>
2.1.1 Propriedades poliméricas . . . . .	15
2.1.2 Processos produtivos . . . . .	15
<b>2.2 Redes Neurais . . . . .</b>	<b>16</b>
2.2.1 Redes Recorrentes . . . . .	17
2.2.2 Treinamento e Métricas de Validação . . . . .	19
<b>2.3 Inferência de propriedades de PE . . . . .</b>	<b>20</b>
<b>3 MATERIAIS E MÉTODOS . . . . .</b>	<b>22</b>
<b>3.1 Unidade Industrial . . . . .</b>	<b>22</b>
3.1.1 Características Gerais e Variáveis de Processo . . . . .	23
3.1.2 Estudo de Caso . . . . .	24
<b>3.2 Modelagem do Sistema . . . . .</b>	<b>24</b>
3.2.1 Coleta e Pré-tratamento de Dados . . . . .	24
3.2.2 Defasagem Temporal dos Dados . . . . .	26
3.2.3 Seleção de Variáveis . . . . .	27
3.2.4 Redes Neurais . . . . .	29
3.2.4.1 Estrutura da Rede Neuronal Convencional . . . . .	30
3.2.4.2 Estrutura da Rede Neuronal Recorrente . . . . .	30
<b>4 RESULTADOS . . . . .</b>	<b>31</b>
<b>4.1 Rede Convencional . . . . .</b>	<b>31</b>
4.1.1 Pré-processamento dos dados . . . . .	32
4.1.2 Inferência com a rede convencional . . . . .	32
<b>4.2 Rede Recorrente . . . . .</b>	<b>35</b>
4.2.1 Inferência com a rede recorrente . . . . .	35
<b>4.3 Rede Combinada . . . . .</b>	<b>37</b>

4.3.1	Inferência com a rede combinada . . . . .	38
5	CONCLUSÃO . . . . .	41
	REFERÊNCIAS . . . . .	43

## LISTA DE ILUSTRAÇÕES

1	Estrutura molecular do polietileno. . . . .	14
2	Representação esquemática de diferentes classes de polietileno. (a) PEAD; (b) PEBD; (c) PELBD. . . . .	15
3	Esquemático de uma unidade da rede neuronal. . . . .	17
4	Rede neuronal genérica. . . . .	17
5	Unidade de rede neuronal recorrente. . . . .	18
6	Segmento de rede neuronal recorrente genérica. . . . .	19
7	Esquemático da unidade industrial estudada. . . . .	22
8	Fluxograma de desenvolvimento de uma RN. . . . .	25
9	Correlação das variáveis de entrada do sistema com IF. . . . .	27
10	Gráfico de IF e relação entre concentração de hidrogênio e eteno. . . . .	28
11	Representação do <i>dropout</i> em uma rede neuronal convencional. . . . .	30
12	Gráfico de correlação e distribuição do erro - rede convencional. . . . .	33
13	Comparação entre saída real e predita - rede convencional. . . . .	33
14	Correlação para região de IF - rede convencional. . . . .	34
15	Comparação entre saída real e predita para região de IF - rede convencional. . . . .	34
16	Gráfico de correlação e distribuição do erro - rede recorrente. . . . .	36
17	Comparação entre saída real e predita - rede recorrente. . . . .	36
18	Problema de representação da rede recorrente. . . . .	37
19	Estrutura final da rede. . . . .	38
20	Gráfico de correlação e distribuição do erro - rede combinada. . . . .	39
21	Comparação entre saída real e predita - rede combinada. . . . .	39
22	Detalhamento da saída real e predita - rede combinada. . . . .	40



## LISTA DE TABELAS

1	Variáveis de processo do reator. . . . .	26
2	Ranking obtido para seleção de variáveis. . . . .	28
3	Testes com redes neuronais convencionais. . . . .	31
4	Testes com pré-processamento de dados. . . . .	32
5	Testes com redes neuronais recorrentes. . . . .	35
6	Testes com estruturas combinadas. . . . .	38

## LISTA DE ABREVIATURAS

PE	Polietileno
PEAD	Polietileno de Alta Densidade
PEBD	Polietileno de Baixa Densidade
PELBD	Polietileno Linear de Baixa Densidade
IF	Índice de Fluidez
CSTR	<i>Continuously Stirred-Tank Reactor</i>
RN	Redes Neurais
SVM	<i>Support Vector Machines</i>
RMSE	<i>Root Mean Square Error</i>

## LISTA DE SÍMBOLOS

$\Sigma$	Somatório
$f$	Função de ativação
$w$	Pesos da rede neuronal
$b$	Vieses da rede neuronal
$x$	Sinal de entrada
$y$	Sinal de saída
$s$	Estado de recorrência
$E$	Sinal de erro
$\eta$	Taxa de aprendizado
$n$	Número de amostras
$\hat{y}$	Saída predita
$\bar{y}$	Valor médio da saída
$R$	Coefficiente de determinação
$X_{norm}$	Vetor normalizado
$\sigma$	Desvio padrão
$L$	Nível do reator
$T$	Temperatura do reator
$P$	Pressão do reator
$P_{C_2}$	Pressão parcial de eteno
$[C_2]$	Concentração de eteno
$[H_2]$	Concentração de hidrogênio
$[C_4]$	Concentração de buteno
$F_{pi}$	Vazão de gás de topo (válvula $i$ )
$F_{H_2}$	Vazão de hidrogênio
$F_{C_2}$	Vazão de eteno
$F_{hex}$	Vazão de hexano

$F_{lm}$	Vazão de licor-mãe
$F_{cat}$	Vazão de catalisador
$F_{ativ}$	Vazão de ativador
$k$	Amostra
$\tau$	Tempo de residência

# 1 INTRODUÇÃO

Conforme exposto por Chikkali, Patel e Sivaram (2020), a demanda mundial por polietileno (PE) é crescente nos últimos anos, com projeção estimada de mercado de aproximadamente 143 bilhões de dólares em 2026. Ao passo que a produção é elevada e tende a crescer, intensifica-se o interesse no desenvolvimento de novos métodos produtivos, bem como novas estratégias para aprimoramento dos existentes, visando redução de custos e desperdícios.

Em termos do desenvolvimento de novas técnicas, Chun et al. (2021) retratam a importância da disponibilidade de dados nas indústrias, uma vez que é possível utilizar dados de processo para desenvolver modelos dos sistemas capazes de auxiliar no entendimento do processo e tomada de decisões. Chiang e Qin (2019) afirmam que a quantidade crescente de dados disponíveis auxilia no entendimento da cadeia produtiva como um todo e na formulação de novos métodos para avaliação de problemas existentes, tal como o uso do aprendizado de máquina (do inglês, *machine learning*) para realização de inferências sobre o processo, por exemplo. Essas abordagens geralmente são feitas em problemas que envolvem alguma restrição da unidade industrial e/ou potencial de melhora no processo produtivo, visando o produto comercial.

Uma vez que o foco da produção industrial é a comercialização, deve-se existir o controle das propriedades físico-químicas do produto, pois elas influenciam, por exemplo, nas características mecânicas do material, sendo determinantes para o tipo de produto a ser formado. Segundo Bryce (1996), para o polietileno, uma das propriedades mais importantes é o índice de fluidez, direcionando ações na operação da planta. Na unidade industrial estudada neste trabalho, a medição do índice de fluidez (IF) é realizada por meio de coleta manual e posterior análise em laboratório, sendo o tempo entre coleta de amostras de, geralmente, quatro horas, limitando a possibilidade de intervenções corretivas de operação, visto que uma dada amostra se refere a uma situação passada da planta.

## 1.1 Motivação e objetivos

Neste trabalho, objetiva-se o desenvolvimento de um modelo para inferência do índice de fluidez, uma das propriedades da resina de polietileno, em um processo petroquímico. Para tanto, são aplicadas diferentes estruturas de redes neurais na proposição de modelos, comparando-as em termos do erro em relação aos dados reais e analisando a predição frente às limitações das redes e complexidade do processo. Busca-se, então, a aplicação de conceitos de modelagem de sistemas para o desenvolvimento de um modelo que auxilie na antecipação de tomadas de decisões no ambiente industrial, agregando valor à cadeia produtiva e possibilitando melhor entendimento sobre o processo. Para tanto, são listados os objetivos específicos, que envolvem:

- analisar o processo e selecionar as variáveis mais relevantes;
- definir a estrutura de rede neuronal;
- realizar testes de validação, comparando a inferência com os dados reais da unidade industrial.

## **1.2 Estrutura do trabalho**

Na Seção 2 é apresentada a revisão bibliográfica do trabalho, abordando as propriedades e formas de produção do polímero estudado, bem como diferentes estruturas de redes neuronais e suas características. Ainda, são listados alguns trabalhos desenvolvidos na área de inferência de propriedades de resinas de polímero e as técnicas utilizadas.

Na Seção 3 são descritos os métodos utilizados para o desenvolvimento do modelo neuronal, sendo abordada, inicialmente, a explicação da unidade industrial avaliada e suas características de processo. Posteriormente, é descrita a modelagem do sistema, envolvendo as etapas de processamento dos dados, seleção de variáveis e estrutura das redes desenvolvidas.

A Seção 4 apresenta os testes realizados com os diferentes tipos de redes neuronais, variando alguns de seus parâmetros e avaliando métricas de desempenho. Por fim, a Seção 5 aborda uma avaliação final dos resultados, além de sugerir propostas para trabalhos futuros.

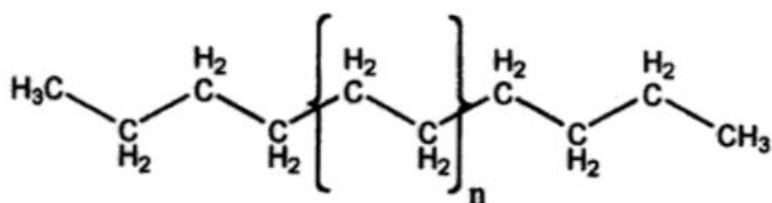
## 2 REVISÃO BIBLIOGRÁFICA

Esta seção busca apresentar, de maneira geral, uma revisão da literatura relacionada tanto ao conceito do problema estudado quanto ao desenvolvimento da solução. Dessa forma, abordam-se propriedades, características e formas de produção do polímero estudado, bem como as estruturas de redes neuronais utilizadas no trabalho.

### 2.1 Polietileno

Um dos tipos de resinas termoplásticas fabricadas industrialmente é o polietileno, sendo formado por meio da agregação de unidades de eteno, resultando em um hidrocarboneto de fórmula molecular  $C_{2n}H_{4n+2}$ , podendo possuir diferentes distribuições de peso molecular (PEACOCK, 2000). Em termos produtivos, o polietileno pode ser obtido por meio da nafta (do petróleo) ou por meio de fontes renováveis, tal como o etanol da cana-de-açúcar. A Figura 1 demonstra a estrutura molecular do polietileno, que consiste em uma cadeia interligada quimicamente de moléculas de eteno (chamado neste caso de monômero).

**Figura 1:** Estrutura molecular do polietileno.

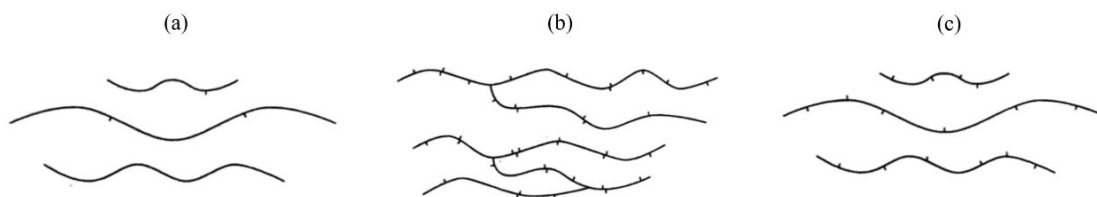


Fonte: Peacock (2000).

Com relação aos aspectos estruturais, uma molécula de polietileno pode possuir diferentes tamanhos e número de ramificações, impactando as propriedades do mesmo. Essas propriedades são alteradas principalmente pela adição de comonômeros na cadeia polimérica, que se ligam às estruturas-base do polímero, ramificando-as, ou pela proporção de hidrogênio frente ao eteno, dado que o hidrogênio funciona como um terminador de cadeia (PEACOCK, 2000). Considerando a diversidade de estruturas, existe uma divisão em classes pela faixa de densidade do polímero, de modo que as principais são o PE de alta densidade (PEAD), PE de baixa densidade (PEBD) e o PE linear de baixa densidade (PELBD), conforme demonstrado na Figura 2.

Em termos de aplicações, Awad et al. (2019) salientam a abrangência de usos para o polietileno, tais como garrafas plásticas, itens domésticos, brinquedos, tanques de combustíveis automotivos, filmes plásticos para uso em sacolas, dentre outros. Dada a

**Figura 2:** Representação esquemática de diferentes classes de polietileno. (a) PEAD; (b) PEBD; (c) PELBD.



Fonte: Peacock (2000).

gama de produtos que podem ser fabricados e a possibilidade de obtenção a partir de fontes renováveis, o PE continua sendo uma resina produzida em larga escala mundialmente.

### 2.1.1 Propriedades poliméricas

Para categorizar um polímero e avaliar sua qualidade, existem diversas propriedades que podem ser analisadas. Visto que a densidade é um fator decisivo para a subdivisão em classes de polímeros, conforme mencionado anteriormente, essa é uma das métricas mais utilizadas no ambiente produtivo para manutenção das características-chave do produto. No entanto, considerando o produto resultante do processo e suas aplicações comerciais, uma das propriedades de maior destaque é o índice de fluidez (BRYCE, 1996).

O índice de fluidez está associado à quantidade de polímero que escoar por um orifício em dado tempo quando submetido a uma certa temperatura (normalmente medido em gramas a cada dez minutos), sendo uma função das propriedades físicas do polímero. Esse parâmetro é importante para o processamento posterior do polímero, assemelhando-se, em escala reduzida, à extrusão realizada por outros setores para agregação de valor ao material (BRYCE, 1996).

Pelo fato de existirem diversas variáveis que impactam o processo e, por consequência, os indicadores de qualidade do produto, as empresas normalmente possuem diferentes classes de produtos, mais conhecidos como *grades*. Cada *grade* é produzido visando obter alguma característica física para o produto, de modo a permitir a comercialização para diferentes indústrias.

### 2.1.2 Processos produtivos

A produção de polietileno envolve a reação de interligação das moléculas de eteno, processo conhecido como polimerização. Industrialmente, essa reação pode ser feita de diferentes formas, sendo classificada pelas propriedades do sistema (tais como temperatura e pressão), além do meio reacional e do tipo de reator. Conforme apontado por Mainka (2018), a maior parte da produção é realizada por meio do processo de suspensão (também chamado de processo em lama - do inglês, *Slurry*).

O processo *Slurry* normalmente envolve dois reatores continuamente misturados (CSTR, *Continuously Stirred-Tank Reactor*), em configuração série ou paralela. O polietileno é produzido em um meio líquido de hidrocarboneto, de modo que a polimerização ocorre continuamente e o polímero permanece suspenso no meio reacional (MAINKA, 2018). Posteriormente, são empregados outros equipamentos e processos para que haja a separação do polímero sólido para comercialização.



## 2.2 Redes Neurais

Na indústria química, com o intuito de descrever os processos, modelos matemáticos são amplamente utilizados, tanto para predição de comportamento das variáveis de interesse como para compreensão das características das unidades industriais. Nesse contexto, existem duas classes de modelagem: fenomenológica, por meio do equacionamento físico-químico, e empírica, que não consideram as leis do processo diretamente na modelagem, sendo, portanto, baseada somente em dados (GABRYS; KADLEC; STRANDT, 2009).

A modelagem fenomenológica, baseada no equacionamento das leis que regem o sistema, busca representar matematicamente as relações envolvidas entre as variáveis e o impacto sobre a resposta em regime permanente e a dinâmica do processo. No entanto, em processos não-lineares, esses modelos são frequentemente complexos, necessitando conhecimento aprofundado das leis do sistema e capacidade computacional para solução das equações. Por outro lado, a modelagem empírica busca extrair as relações do sistema por meio dos dados, sem a necessidade de entendimento completo dos fenômenos envolvidos. Essa abordagem, embora consiga extrair informações significativas dos dados, ajustando-se a eles, possui as desvantagens de ser mais suscetível ao erro na extrapolação de resultados e potencial falta de representação de algumas dinâmicas do sistema. Contudo, o desenvolvimento crescente de novas estratégias de modelagem empírica já permite obter, com certa precisão, modelos satisfatórios para problemas da indústria química, mesmo considerando as não-linearidades inerentes de boa parte dos processos (CHIANG; QIN, 2019).

Segundo Chiang e Qin (2019), com a disponibilidade cada vez maior de dados, diversas indústrias passaram a utilizar essas informações para entendimento da sua cadeia produtiva e desenvolvimento de novas tecnologias, principalmente por meio do aprendizado de máquina. Essa área de interesse da inteligência artificial possui aplicações diversas, tais como predição de séries temporais via regressão, visão computacional e processamento de linguagem. Dentro desse ramo, um dos meios utilizados para alcançar tais objetivos são as redes neurais (RN), estruturas computacionais que surgiram com o objetivo de representar a capacidade de aprendizado do cérebro humano.

As redes neurais podem ser classificadas quanto ao aprendizado, sendo supervisionado, em que são fornecidas entradas e saídas para o treinamento, ou não supervisionado, aprendendo conforme contexto. No caso supervisionado, as redes possuem a capacidade de assimilar padrões e regras por meio de dados, com o objetivo de realizar predições conforme novos dados são fornecidos. Uma rede é composta por várias unidades individuais (neurônios) interconectadas por meio de camadas. Cada unidade possui uma sequência de entradas que, multiplicadas por pesos e somadas de um valor constante (mais conhecido como *bias*), passa por uma função de ativação, atribuindo a não-linearidade do dado de saída. A Figura 3 demonstra um esquemático de uma unidade.

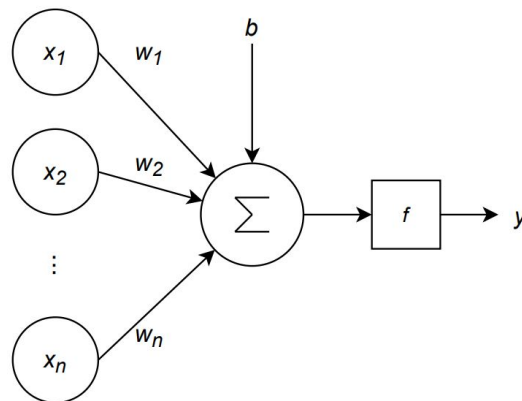
Assim, cada neurônio possuirá uma relação matemática com a saída, dada por

$$y = f\left(\sum(x_i \cdot w_i) + b\right), \quad (1)$$

em que  $x$  representa o sinal de entrada,  $w$  os pesos associados,  $b$  a constante,  $f$  a função de ativação e  $y$  a saída.

Para que essa estrutura seja utilizada para solução de problemas, é necessário que sejam utilizados conjuntos das unidades dadas pela Figura 3 em uma rede conectada, podendo ser variado o número de neurônios e camadas. A Figura 4 exemplifica uma rede neuronal, com

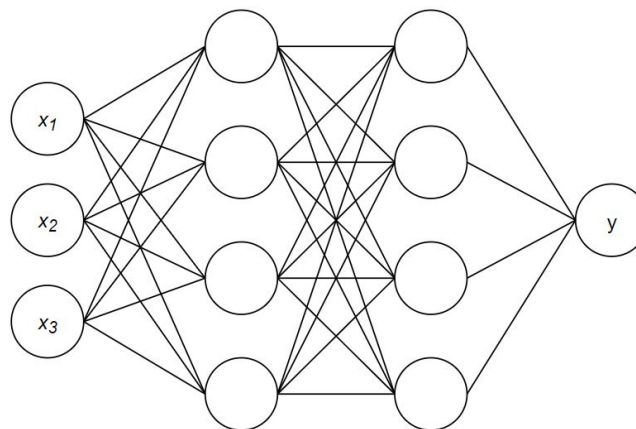
**Figura 3:** Esquemático de uma unidade da rede neuronal.



Fonte: Do autor.

duas camadas intermediárias, sendo composta uma estrutura com número de parâmetros associado à complexidade da rede.

**Figura 4:** Rede neuronal genérica.



Fonte: Do autor.

Com uma estrutura de rede neuronal com mais camadas, é constituída a chamada rede neuronal profunda. Em teoria, a complexidade da rede está relacionada à sua capacidade de predição, sendo uma relação diretamente proporcional (GABRYS; KADLEC; STRANDT, 2009). No entanto, segundo os mesmos autores, normalmente são utilizadas estruturas intermediárias, uma vez que redes complexas possuem menor poder de generalização, tendendo a modelar relações para variações pouco significativas dos dados, tais como erros de amostragem e ruído.

### 2.2.1 Redes Recorrentes

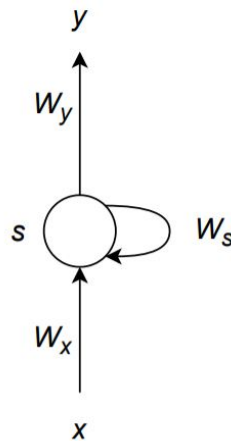
As redes apresentadas possuem como estrutura a sequência de camadas descrita anteriormente, com entradas, camadas intermediárias e saídas, sendo que o sinal é propagado apenas no sentido da entrada para a saída. Ou seja, esse tipo de estrutura (conhecida como convencional, direta ou *feedforward*) não apresenta realimentação, de modo que os dados de entrada impactam de forma imediata a saída (SHERSTINSKY, 2020).

Dessa forma, buscando representar processos em que a dinâmica é significativa, foram

desenvolvidos sistemas de predição sequenciais que levam em conta o histórico dos dados. Isso é possível por meio da aplicação de uma variável de estado passado no cálculo do estado atual, estabelecendo-se assim uma dependência do dado em um certo instante com os dados nos instantes anteriores, fornecendo certa memória ao modelo. Em termos de aplicações dessa estratégia, destacam-se o processamento de linguagem, predição do padrão de séries temporais e modelagem da dinâmica de processos (JHA; SAHOO; SIGNH, 2019).

Considerando a estrutura das redes neuronais, essa estratégia é representada por meio de redes recorrentes, sendo a recorrência responsável pela aplicação do estado anterior na rede. Esquemáticamente, uma unidade da rede recorrente pode ser representada pela Figura 5, em que, além do vetor de entradas  $x$  e a saída  $y$  (com as respectivas matrizes de pesos  $W_x$  e  $W_y$ ), tem-se o estado intermediário  $s$ , que realimenta a rede com o dado passado por meio de  $W_s$ . Dessa maneira, além da entrada e da saída, comuns à estrutura de rede neuronal convencional, existe um estado de recorrência, responsável por mapear o estado passado para a rede como um todo. Com relação aos pesos,  $W_x$  e  $W_y$  são equivalentes à rede convencional, ponderando o sinal para a próxima camada, enquanto que  $W_s$  multiplica o estado de recorrência, realimentando-o na rede.

**Figura 5:** Unidade de rede neuronal recorrente.



Fonte: Do autor.

Em termos matemáticos, a relação do estado de recorrência com os demais parâmetros do sistema é dada por

$$s_t = f(W_s \cdot s_{t-1} \oplus W_x \cdot x_t + b_h), \quad (2)$$

em que  $s_t$  representa o estado atual,  $s_{t-1}$  o estado no instante anterior,  $f$  a função de ativação e  $b_h$  um termo constante de viés do estado (*bias*). Assim, para compor o estado atual, é necessário realizar uma soma elemento a elemento do estado anterior com as entradas, sendo posteriormente adicionado o termo de viés e aplicada a função de ativação.

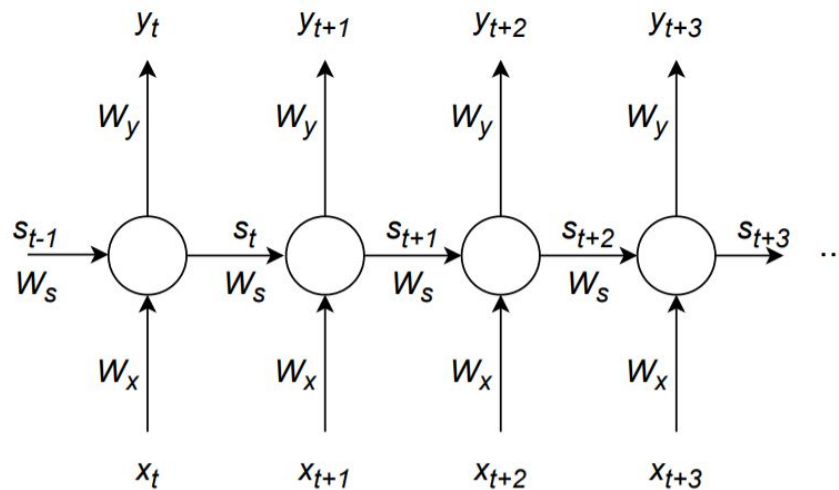
Já a saída do sistema pode ser representada por

$$y_t = f(W_y \cdot s_t + b_y), \quad (3)$$

sendo  $b_y$  o termo de viés da saída. Desse modo, a saída em um certo instante depende do estado naquele instante, que por sua vez depende das entradas e do estado anterior, constituindo-se, assim, a recorrência.

Da mesma forma que em uma rede convencional, é preciso utilizar várias unidades para compor uma estrutura que possibilite a resolução de problemas práticos. Então, agrupando-se as unidades da Figura 5, é composta a estrutura da Figura 6. Com a estrutura conectada de neurônios, o estado de recorrência é propagado sequencialmente pela rede, sendo possível realizar previsões com base em dados históricos.

**Figura 6:** Segmento de rede neuronal recorrente genérica.



Fonte: Adaptado de Jha, Sahoo e Signh (2019).

Conforme apontado por Sherstinsky (2020), com as redes recorrentes é possível, então, considerar o efeito de dados anteriores na modelagem de sistemas, viabilizando a avaliação da dinâmica dos processos envolvidos. Além disso, a recorrência torna a estrutura mais robusta a variações abruptas dos dados e discrepâncias provenientes dos sistemas de medição instalados, ao passo que atuam como filtros para que o erro não se propague para a saída no mesmo instante.

### 2.2.2 Treinamento e Métricas de Validação

Para que, a partir de uma estrutura definida, uma rede neuronal consiga realizar previsões de uma grandeza física, a mesma necessita ser treinada com um conjunto de dados representativo para o processo em questão. Visto que uma rede possui normalmente uma estrutura complexa, com funções de ativação em geral não-lineares encadeadas de forma implícita, é necessária uma estratégia de otimização para treinamento dos parâmetros da rede, minimizando uma função de custo.

Segundo Haykin e Leung (1991), no caso de redes neuronais, pode-se utilizar um algoritmo de propagação do erro pela rede, computando as derivadas parciais da diferença entre a saída real e a predita para atualização dos parâmetros, estratégia conhecida como retropropagação do erro (do inglês, *backpropagation*). Esse princípio é utilizado para o treinamento de qualquer tipo de rede, dado que indica a atualização dos pesos em função do erro associado ao seu sinal na resposta. Assim, a atualização dos pesos pode ser dada, de maneira simplificada, por

$$\Delta w_{ij} = -\eta \cdot \frac{\partial E}{\partial w_{ij}}, \quad (4)$$

em que  $E$  é o erro (diferença entre a saída real e a predita),  $w_{ij}$  o peso associado ao neurônio  $j$  da camada  $i$  e  $\eta$  um parâmetro de taxa de aprendizado. Assim, cada peso na rede é atualizado levando-se em conta a derivada parcial do erro associada àquele parâmetro em questão.

Com relação à avaliação dos resultados, conforme exposto por Ostertagova (2012), algumas das métricas mais comuns envolvendo métodos de regressão são a raiz do erro médio quadrático (*Root Mean Square Error*, RMSE) e o coeficiente de determinação. A primeira delas diz respeito à avaliação da raiz da diferença quadrática entre a saída real  $y$  e a saída predita  $\hat{y}$ , sendo ajustada pelo número de amostras  $n$ , dada por

$$RMSE = \sqrt{\frac{1}{n} \sum (\hat{y} - y)^2}. \quad (5)$$

Já o coeficiente de determinação, embora menos utilizado para avaliação dos resultados, é uma das métricas mais importantes na verificação do padrão de erro de uma regressão. Por meio dele, é possível avaliar a correlação linear entre as variáveis, podendo também ser representado graficamente, ao ser plotada a distribuição da saída predita frente à real (CHICCO; JURMAN; WARRENS, 2021). Em termos de equacionamento, o coeficiente de determinação é dado por

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

sendo  $\bar{y}$  o valor médio da saída na amostra selecionada. Por meio dessa relação, é obtido um valor entre zero e um, sendo que quanto maior o valor, maior a relação linear entre a saída real e a predita.

### 2.3 Inferência de propriedades de PE

Kim, Lee e Yeo (2008) elaboraram um analisador virtual (modelo de inferência) para índice de fluidez de resinas de polietileno de alta densidade, produzidas em reator *Slurry*, por meio da otimização de parâmetros de equações simplificadas obtidas em literatura. Posteriormente foi feita uma integração do modelo no processo produtivo para controle em malha fechada de variáveis que impactavam o IF, permitindo reduzir as perdas associadas à produção de PE em 48%. Já Lee et al. (2008), com o mesmo objetivo, também projetaram um analisador virtual, mas por meio da linearização de equações do sistema e uso de estimadores em espaço de estados.

Uma vez que a troca de *grade* produzido pelas unidades industriais é um dos períodos críticos do processo, alguns trabalhos foram desenvolvidos buscando modelar variáveis de qualidade com o intuito de permitir atuação mais rápida no processo durante esses períodos. Ahmed, Nazir e Yeo (2009), por meio da técnica de Regressão Parcial de Mínimos Quadrados (do inglês, *Partial Least Squares*), elaboraram um modelo empírico para predição de IF nas transições de produto. Já Kang, Kim e Park (2010) desenvolveram modelos para inferência de diferentes propriedades poliméricas (índice de fluidez, densidade e massa molar) para uso em momentos de troca de *grade*, reduzindo em 50% as perdas associadas à especificação do polímero.

Também foram desenvolvidos modelos em redes neuronais para inferência de propriedades mecânicas do polímero final formado. Alhindawi, Allaf e Altarazi (2019) realizaram testes com predição de propriedades físicas do produto, principalmente a tensão máxima,

comparando as técnicas de redes neurais e SVM (*Support Vector Machines*), para otimizar o processo visando obtenção de determinadas propriedades poliméricas. Hilal-Alnaqbi, Mourad e Yousef (2011) desenvolveram modelos de predição de curvas de tensão de resinas de polietileno e polipropileno por meio do uso de redes neurais.

Na mesma unidade industrial avaliada nesse trabalho, Mohr (2004) realizou estudos acerca das variáveis de processo, adaptando equacionamentos da literatura para o processo produtivo da planta de polietileno de alta densidade, obtido via processo *Slurry*. Assim, foi desenvolvido um modelo de inferência do índice de fluidez do polímero, por meio da estimação de parâmetros via Filtro de Kalman e Filtro de Kalman Estendido, obtendo um modelo contemplando diferentes situações operacionais da unidade.

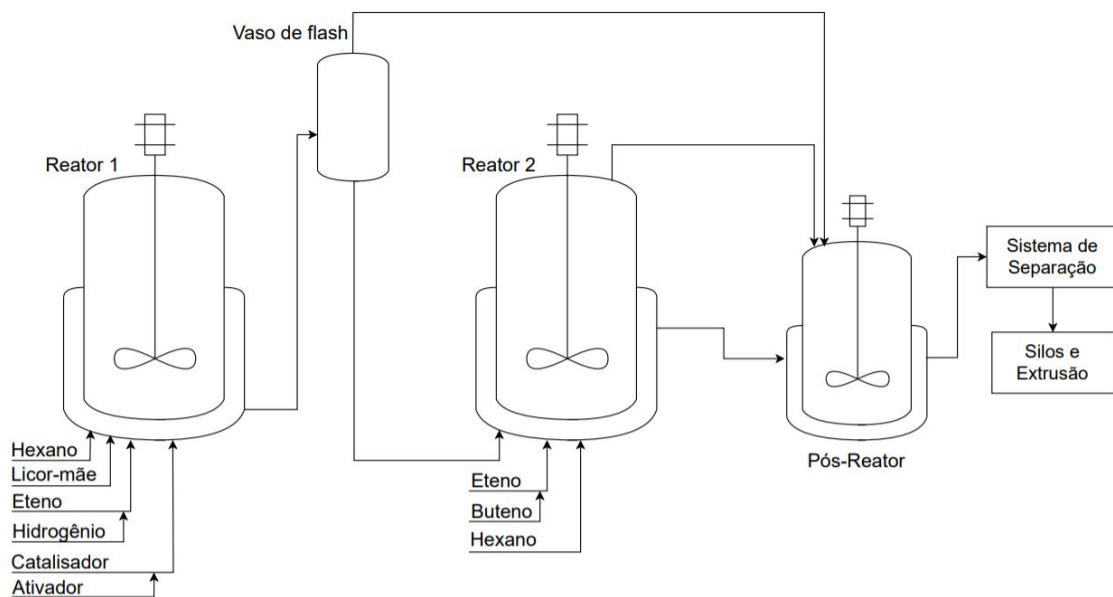
### 3 MATERIAIS E MÉTODOS

Esta seção busca apresentar os materiais e métodos utilizados no desenvolvimento e validação do trabalho. Para tanto, são descritas a unidade industrial estudada, a cadeia produtiva das resinas de polímero, além das etapas envolvidas na modelagem matemática do sistema (coleta de dados, seleção de variáveis e desenvolvimento da estrutura das redes neurais).

#### 3.1 Unidade Industrial

A unidade industrial que compõe o estudo de caso consiste no sistema de reação de uma planta de polietileno de alta densidade, com polimerização em suspensão. A unidade é composta por dois reatores CSTR em série, um vaso de *flash* e um pós reator, conforme mostrado na Figura 7. Após passar por esse conjunto de equipamentos, o produto é direcionado para um sistema de separação, para posterior adição em silos e extrusão.

**Figura 7:** Esquemático da unidade industrial estudada.



Fonte: Do autor.

O primeiro reator é alimentado pelo fundo com hexano puro e licor-mãe como solventes (meio de suspensão), sendo esse último uma corrente de solvente reciclado do

próprio processo. Também são adicionados pelo fundo, em fase gasosa, eteno (monômero da reação) e hidrogênio, responsável por regular o tamanho da cadeia e controlar propriedades-chave do polímero final, tais como massa molar e índice de fluidez. Para proporcionar a reação, são alimentados o catalisador, que é um sólido em suspensão de hexano, e um ativador, para remoção de impurezas e manutenção do catalisador ativo. Conforme o eteno é misturado no reator, o gás é distribuído em bolhas com solubilização parcial na fase líquida e se difunde até o catalisador, ocorrendo a polimerização. O tempo de residência, que consiste no tempo médio em que dado elemento fluido permanece no sistema, é de cerca de quatro horas para esse reator.

Após passar pelo primeiro reator, a mistura vai para um vaso vaporizador (*flash*), com a função de remover o excesso de hidrogênio, e posteriormente é direcionada para o segundo reator. Esse reator é alimentado com o efluente proveniente do primeiro reator, ou seja, além de polímero já formado, monômeros não reagidos. Assim, é feita uma alimentação pelo fundo de eteno, hexano e buteno, sendo o buteno responsável por ramificar a cadeia, reduzindo a massa molar e aumentando o índice de fluidez. Por fim, a mistura é enviada para o pós-reator, sendo posteriormente direcionada para o sistema de separação, onde o polímero formado é separado.

Vale destacar também que ambos os reatores possuem camisas para resfriamento, com circulação de água, uma vez que a reação é exotérmica. Além dessa circulação, existe controle de temperatura, por meio de um sistema de bombeamento da suspensão do reator para um trocador de calor, de modo que o sistema pode ser considerado isotérmico.

### 3.1.1 Características Gerais e Variáveis de Processo

Conforme exposto na Seção 2.1.1, as unidades industriais produzem polímeros com diferentes características, dependendo do *grade*, impactando nas propriedades físico-químicas do produto. Cada *grade* necessita de condições distintas, de modo que as variáveis de processo devem ser tais que mantenham a produção na faixa de qualidade. No entanto, mesmo com a unidade operando em forma de campanhas, produzindo alguns *grades* no mês para atender a demanda comercial, observam-se alterações na dinâmica das campanhas com o passar dos anos. Isso ocorre pelo fato de sistemas reais não serem, de fato, invariantes no tempo, de forma que os parâmetros de controle do processo precisam ser adaptados conforme o tempo de operação.

Para controle dessas variáveis de qualidade do produto, algumas entradas do sistema possuem um impacto maior. Segundo McAuley e McGregor (1991), as variáveis com maior influência no índice de fluidez envolvem

- a temperatura do reator;
- a relação entre a concentração de hidrogênio e a concentração de eteno no reator;
- a relação entre a concentração de buteno e a concentração de eteno;
- a concentração de impurezas frente à de eteno;
- a concentração de cocatalisador frente à de eteno.

Pode-se perceber que a maior parte das variáveis são relacionadas com a concentração de eteno, uma vez que é o monômero da reação. Assim, a concentração relativa das demais variáveis com o eteno influi diretamente na reação e, por consequência, nas propriedades poliméricas. Ainda, a relação entre as variáveis é não-linear, dado o alto grau de interação entre elas (MCAULEY; MCGREGOR, 1991).



### 3.1.2 Estudo de Caso

O foco do estudo é o primeiro reator da unidade industrial, com o objetivo de desenvolver um modelo empírico para o índice de fluidez, uma das variáveis de qualidade mais importantes do produto. Atualmente, para medição do IF do polímero formado, é realizada coleta manual na unidade industrial a cada quatro horas e posterior análise em laboratório. O tempo de análise, desde a coleta até disponibilização no historiador de dados, é de cerca de 30 minutos, sendo o dado preenchido retroativamente no historiador, representando o período de coleta, e não de disponibilização do dado.

Dessa forma, busca-se desenvolver um modelo que consiga relacionar as variáveis de entrada do processo e estimar o índice de fluidez no período entre amostras, de modo a permitir o ajuste mais frequente das condições de processo. Por meio dessa inferência de propriedade, é possível visualizar, com frequência maior, o valor do IF e verificar se está atendendo as especificações do produto desejado e, caso não esteja, realizar alterações no processo antecipadamente. Para tanto, é necessário avaliar as variáveis que influenciam no processo, coletá-las e desenvolver uma rede neuronal que possa inferir, de maneira satisfatória, esse indicador de qualidade. Como política operacional da unidade industrial estudada, considera-se satisfatório um erro dentro de uma faixa de 10% em relação ao valor real do índice de fluidez.

Escolheu-se o primeiro reator pelo fato da maior parte da especificação final do produto ser dependente da reação de polimerização no mesmo, além de ser mais conveniente e menos custoso, em termos de processo, realizar o ajuste das variáveis em relação ao segundo reator, cabendo a esse uma correção do IF. Somado à isso, o primeiro reator possui uma faixa mais ampla de valores de IF, de modo que podem ser modeladas diferentes faixas de processo. Mesmo que aplicada ao primeiro reator apenas, a estratégia desenvolvida pode ser utilizada no segundo reator, sendo necessário apenas realizar as devidas adequações em termos do estudo das variáveis e desenvolvimento da rede neuronal.

## 3.2 Modelagem do Sistema

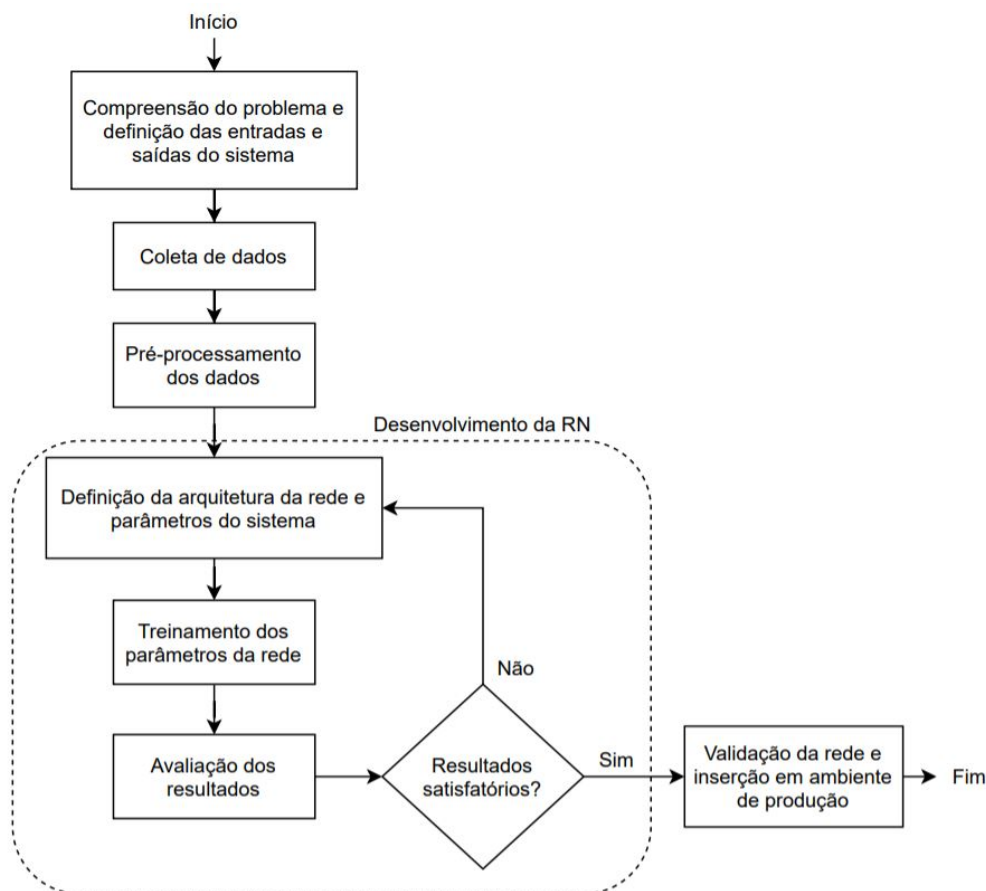
Com o intuito de desenvolver um modelo para predição de uma variável por meio de técnicas baseadas em dados, mais especificamente via redes neurais, é aconselhável seguir um conjunto de passos, que vão desde a compreensão do problema até a validação. A Figura 8 demonstra um fluxograma de desenvolvimento de rede neuronal, sendo aplicado para a elaboração do trabalho. Vale destacar que o fluxo é iterativo, de modo que as etapas de validação definem a continuação da otimização dos parâmetros.

### 3.2.1 Coleta e Pré-tratamento de Dados

Para coletar os dados da unidade industrial, foi realizada uma extração do historiador de dados da planta (PI Process Book<sup>®</sup>) do período de 2017 até 2021. No total, foram coletados cerca de 9000 valores de dados de 19 variáveis. As variáveis coletadas, bem como suas descrições são dadas pela Tabela 1. A coleta foi feita com base no conhecimento das variáveis que possuem maior influência no processo, sendo as vazões destacadas anteriormente na Figura 7, as concentrações de analisadores de linha da unidade, *grade* produzido, índice de fluidez conforme amostragem, além da temperatura, pressão e nível do reator.

Uma vez que o conjunto de dados coletados pode ter discrepâncias, tais como erro de coleta do programa ou falhas de leitura de instrumentos em campo, os dados com problema

**Figura 8:** Fluxograma de desenvolvimento de uma RN.



Fonte: Adaptado de Gabrys, Kadlec e Strandt (2009).

foram removidos. O tratamento realizado para a rede convencional é diferente daquele necessário para a rede recorrente, em função das características inerentes do algoritmo. Para a rede convencional, foram removidos dados fora do limite de engenharia definidos pela operação da unidade industrial, bem como erros de coleta, reduzindo o conjunto para cerca de 8700 dados. Já para a rede recorrente, dado que é necessário manter a dependência temporal, com o tempo sendo o eixo base, preferiu-se por não remover os dados discrepantes, mas sim alterá-los para que tivessem o valor zero como entrada na rede, mantendo o número de dados de entrada.

Uma vez que as variáveis coletadas possuem diferentes unidades de medida, a escala entre elas é diferente. Para tanto, recomenda-se a normalização para melhorar o desempenho e facilitar a convergência do algoritmo de otimização (ILYAS et al., 2018). Dessa forma, foi utilizado, para cada variável, o escalonamento dado por

$$X^{norm} = \frac{X - \bar{X}}{\sigma(X)}, \quad (7)$$

em que  $X^{norm}$  representa o valor normalizado,  $X$  o dado real,  $\bar{X}$  a média e  $\sigma(X)$  o desvio padrão amostral.

**Tabela 1:** Variáveis de processo do reator.

Variável	Descrição
$L$	% Nível em relação à capacidade máxima
$T$	Temperatura
$P$	Pressão
$P_{C_2}$	Pressão parcial de eteno
$[C_2]$	% Concentração de eteno
$[H_2]$	% Concentração de hidrogênio
$[C_4]$	% Concentração de buteno
$F_{p1}$	Vazão de gás de topo (válvula 1)
$F_{p2}$	Vazão de gás de topo (válvula 2)
$F_{p3}$	Vazão de gás de topo (válvula 3)
$F_{H_2}$	Vazão de hidrogênio
$F_{C_2}$	Vazão de eteno
$F_{hex}$	Vazão de hexano
$F_{lm}$	Vazão de licor-mãe
$F_{cat}$	Vazão de catalisador
$F_{ativ}$	Vazão de ativador
$[H_2]/[C_2]$	Relação de concentração de hidrogênio e eteno
$Grade$	$Grade$ em produção na unidade
IF	Índice de Fluidez (variável de predição)

Fonte: Do autor.

### 3.2.2 Defasagem Temporal dos Dados

Após a coleta dos dados, é necessário realizar o pré-processamento, conforme demonstrado anteriormente no fluxograma da Figura 8. Considerando o tempo de residência do reator e as características inerentes do processo, como o tempo para polimerização, é possível inferir que as variáveis coletadas em um dado instante não correspondem ao valor do índice de fluidez medido naquele mesmo instante.

No caso das concentrações, o reator possui uma calota, que concentra os gases não reagidos no processo, sendo analisados via analisador de linha (cromatógrafo). No entanto, a reação ocorre na fase líquida, de modo que existe uma dinâmica associada ao IF. Para as vazões, conforme são inseridos os reagentes no reator, ocorre o processo de solubilização e polimerização, impactando também a concentração desses materiais. Dessa forma, os dados foram defasados buscando relacionar esses aspectos dinâmicos, bem como o tempo de residência e a amostragem da variável. Uma vez que a estimação do tempo correto demanda conhecimento mais aprofundado, considerou-se uma defasagem aproximada de quatro horas para as variáveis, equivalendo ao tempo de residência do reator, buscando representar simplificadamente a dinâmica do processo.

Vale destacar que essa estratégia é necessária apenas para configurações de rede neuronal que não possuem realimentação de estado (estrutura não recorrente), uma vez que a defasagem busca corrigir temporalmente a sequência de entradas frente à saída. Em função da rede recorrente já possuir a capacidade de modelagem da dinâmica, ao fazer realimentações entre camadas, os dados não foram ajustados temporalmente pelo tempo de residência do reator.

### 3.2.3 Seleção de Variáveis

Um dos passos do pré-processamento de dados inclui o processo de seleção de variáveis. Essa estratégia, quantitativa e/ou qualitativa, visa reduzir o número de variáveis no modelo, de modo a realizar o mapeamento para a saída com um menor número de variáveis de entrada, privilegiando aquelas com uma relação mais representativa. A motivação do uso dessa estratégia, no caso do aprendizado de máquina, é obter uma melhor generalização da predição, diminuindo o erro associado. Isso ocorre pelo fato da rede neuronal modelar relações apenas para os dados representativos, ao invés de mapear todo o conjunto de variáveis para a saída (GABRYS; KADLEC; STRANDT, 2009).

Conforme exposto por Elisseff e Guyon (2003), existem diversas formas de realizar a seleção das variáveis, sendo a correlação uma das primeiras abordagens adotadas, tanto para remoção de variáveis menos representativas, como para a visualização da dependência entre os próprios dados de entrada. A correlação é uma abordagem quantitativa baseada na formulação do coeficiente de determinação, dado anteriormente pela Equação (6), levando-se em conta a covariância entre as variáveis. Para tanto, geralmente é utilizada uma matriz de correlação, em que são enumeradas as variáveis e, para cada célula da matriz, é computada a correlação entre duas variáveis. A Figura 9 demonstra um recorte da matriz de correlação do sistema, sendo retratadas apenas as interações das entradas com a saída, ocultando as relações das entradas entre si.

**Figura 9:** Correlação das variáveis de entrada do sistema com IF.



Fonte: Do autor.

Segundo Hosseinkhani, Ibrahim e Koupaie (2013), existem outras estratégias para seleção de variáveis que não levam em conta a relação somente entre duas variáveis, fornecendo um ordenamento mais representativo em termos das variáveis a serem escolhidas. Um dos métodos mais utilizados nesse quesito é o *stepwise backward elimination*, em que, a cada iteração do código, é feita uma ponderação da importância das variáveis para predição e a variável de entrada com pior atributo é eliminada, sendo sempre avaliado um conjunto de entradas, ao invés de uma individualmente. Isso garante que variáveis sejam escolhidas conforme a melhor combinação entre elas, não eliminando necessariamente aquela com menor influência individual. Aplicando o método para o conjunto de dados em questão, por meio da utilização de módulos disponíveis em bibliotecas da linguagem de programação utilizada (tais como adaptações do módulo *recursive feature elimination*), é obtido o ordenamento conforme demonstrado na Tabela 2.

Na Tabela 2, foram apresentadas, por simplicidade, apenas as oito variáveis mais importantes. Embora não exista um número padronizado para o número de variáveis a serem utilizadas em um sistema de regressão de aprendizado de máquina, é prática comum a escolha entre cinco e dez variáveis de entrada, a depender do problema abordado e da complexidade do mesmo. Dessa forma, optou-se pela utilização de oito variáveis, um valor intermediário recomendado e cerca de metade do número de variáveis originais.

Pode-se perceber que a variável de entrada com maior correlação com a saída é a concentração de hidrogênio em relação à de eteno. Conforme exposto na Seção 2.1, o hidrogênio funciona como um terminador de cadeia, de modo que quanto maior sua

**Tabela 2:** Ranking obtido para seleção de variáveis.

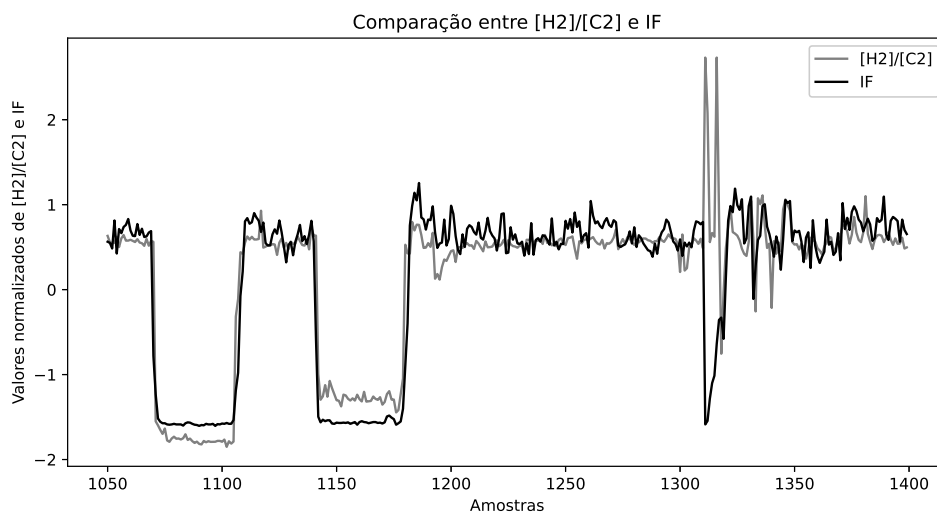
Posição	Variável
1	$[H_2]/[C_2]$
2	$[C_2]$
3	$[H_2]$
4	$F_{H_2}$
5	$F_{cat}$
6	$F_{C_2}$
7	$P_{C_2}$
8	$L$

Fonte: Do autor.

concentração em relação ao monômero da reação (eteno), menores serão as cadeias poliméricas, conferindo um índice de fluidez maior. Além disso, outras variáveis de entrada também impactam a relação  $[H_2]/[C_2]$ , tais como vazões de entrada, haja vista que tendem a aumentar ou diminuir a concentração dos gases solubilizados no sistema. Como exemplo disso, ao aumentar a vazão de catalisador, será aumentada a conversão de reagentes em polímero e, sendo o eteno o principal reagente, a concentração do mesmo tende a cair proporcionalmente mais que a de hidrogênio, aumentando a relação  $[H_2]/[C_2]$  e, portanto, o IF.

A Figura 10 retrata um período de dados de operação extraídos do historiador, apesar dos valores do IF e da relação de concentrações de hidrogênio e eteno. É possível verificar a correlação principalmente nas maiores variações, que ocorrem em transições de carga da unidade industrial, visto que cada produto exige uma determinada combinação de parâmetros de operação. Em função de confidencialidade industrial, os dados são apresentados normalizados conforme Equação (7).

**Figura 10:** Gráfico de IF e relação entre concentração de hidrogênio e eteno.



Fonte: Do autor.

### 3.2.4 Redes Neurais

Para implementação das redes neurais, foram utilizados os conjuntos de bibliotecas (conhecido como *frameworks*) *Keras* e *scikit-learn*, sendo os mais usuais para aplicações de *machine learning* na linguagem de programação Python. Por meio dessa estrutura, é possível desenvolver a arquitetura das redes de forma sequencial, agregando módulos referentes às camadas e adicionando tratamentos dos dados.

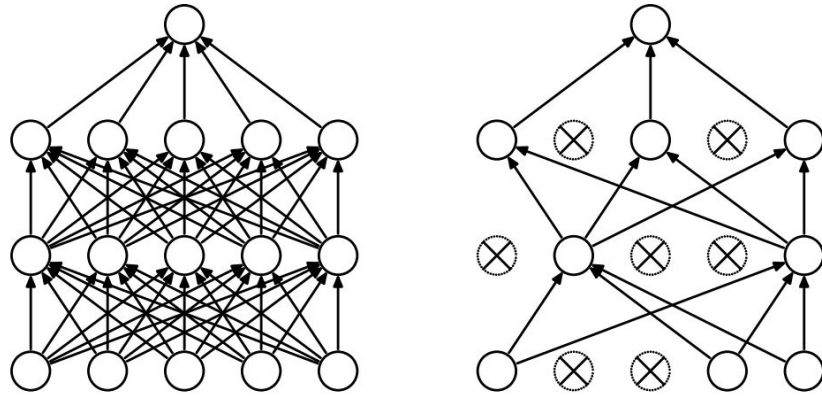
Buscando a convergência da resposta e menor erro, foram adotadas algumas estratégias específicas para o treinamento das redes convencionais e recorrentes. Antes de iniciar o treinamento de redes neurais, é boa prática normalizar os dados para que, além de ser diminuído o tempo de convergência, seja garantido que todas as variáveis tenham a mesma ponderação (peso) no cômputo dos gradientes, evitando problemas associados às escalas dos valores (ILYAS et al., 2018). No entanto, a média e desvio padrão utilizados na normalização de todo o conjunto de dados devem ser calculados apenas com os dados de treinamento da rede, evitando adicionar viés para a validação (CASARI; ZHENG, 2018). Deste modo, os dados foram separados, deixando 70% para o treino e 30% para validação e teste, utilizando a Equação (7) para normalização dos dados.

Para a etapa de treinamento, foi utilizada a inicialização aleatória dos parâmetros da rede neuronal, seguindo uma distribuição normal com desvio padrão obtido a partir da normalização. Essa estratégia foi empregada visto que, a depender dos parâmetros, a otimização pode convergir para um mínimo local, não resultando na melhor combinação final de parâmetros da rede (ILYAS et al., 2018). Dessa forma, inicializando os parâmetros aleatoriamente, visa-se mitigar esse problema, sendo realizado um conjunto de testes para a mesma estrutura para computar uma média dos resultados.

Uma vez que as redes neurais necessitam de iterações para treinamento (também conhecidas como épocas), o tempo computacional para otimização pode se estender sem que o resultado de validação seja, de fato, melhorado. A fim de contornar esse problema, foi implementada uma lógica na qual o treinamento é parado após 20 iterações sem redução do erro médio quadrático da inferência nos dados de validação. Essa técnica é conhecida como *early stopping*, e permite mitigar um dos problemas mais comuns no treinamento de redes neurais: o sobreajuste dos dados (do inglês, *overfitting*). Normalmente, após uma dada iteração, a rede começa a se ajustar de forma demasiada aos dados, tentando modelar pequenas variações (que podem ser inclusive erros de medição/ruído). Ao fazer isso, a rede diminui a capacidade de generalização, de modo que o erro diminui para o conjunto de dados de treino, mas aumenta para o conjunto de dados de validação (RASKUTTI; WAINWRIGHT; YU, 2014).

Segundo Hinton et al. (2014), outra forma de lidar com o sobreajuste é por meio do uso de uma técnica conhecida como *dropout*. Essa estratégia visa, a cada iteração, remover aleatoriamente (segundo uma dada probabilidade) alguns neurônios e conexões entre camadas do treinamento, ou seja, algumas variáveis de entrada e suas propagações de sinais pela rede são temporariamente desconsideradas na predição. Dessa forma, a rede busca realizar a predição apenas com aqueles sinais considerados, atualizando os pesos da rede como um todo e buscando mitigar a dependência que pode surgir em torno de alguma variável em específico, causando o sobreajuste. A Figura 11 ilustra o *dropout*, em que a rede da direita possui a mesma estrutura da rede da esquerda, mas com alguns neurônios e conexões inibidos na iteração. Ao passo que não existe um número definido para a probabilidade de inativação de neurônios por iteração, foi utilizado um valor de 0,2 entre camadas (definido empiricamente, sendo um valor adotado como prática comum), com cada neurônio tendo uma probabilidade de 80% de ser mantido por época.

**Figura 11:** Representação do *dropout* em uma rede neuronal convencional.



Fonte: Hinton et al. (2014).

#### 3.2.4.1 Estrutura da Rede Neuronal Convencional

Para a rede neuronal convencional, foi implementada a estrutura apresentada na Figura 4, com as camadas de entrada, intermediárias e de saída com conexões apenas no sentido de saída de rede, sem realimentação. Dessa forma, cada vetor de entrada (que representa um dado para cada variável) é processado pelo modelo, gerando apenas uma saída, a inferência do índice de fluidez. Assim, considerando a defasagem aplicada na rede, a saída é dada por

$$IF[k] = NN(x[k - \tau]), \quad (8)$$

em que  $k$  representa a amostra,  $NN$  a rede neuronal (modelo) e  $\tau$  o tempo de residência do reator (defasagem aplicada).

#### 3.2.4.2 Estrutura da Rede Neuronal Recorrente

O desenvolvimento da rede neuronal recorrente também seguiu os passos iniciais descritos anteriormente, realizando as adaptações necessárias para abranger a realimentação entre camadas da rede. Assim, foi implementada a estrutura conforme formulações dadas pelas Equações (2) e (3), em que os dados de entrada são adicionados no modelo e, para computar a saída predita, existe a consideração do estado de recorrência, que estabelece a dependência dos instantes anteriores no atual.

## 4 RESULTADOS

Nesta seção, são apresentados os resultados dos testes realizados com redes neuronais convencionais e recorrentes, analisando a raiz do erro médio quadrático entre predição e dado real, bem como o coeficiente de correlação, visando avaliar se atendem os requisitos de projeto. Vale destacar que todos os resultados apresentados são normalizados, em função de confidencialidade industrial.

### 4.1 Rede Convencional

Para a rede convencional, inicialmente foram feitos testes variando-se o número de neurônios e de camadas ocultas, para poder verificar o padrão de erro da resposta e direcionar análises posteriores. Uma vez que, após a seleção de variáveis, restaram oito entradas para regressão e que normalmente são empregados números de neurônios com base na potência de dois, escolheram-se valores para o número de neurônios por camada intermediária variando de 4 até 64. Os resultados da raiz do erro médio quadrático e do coeficiente de correlação para cada teste são apresentados na Tabela 3, considerando uma média de dez testes realizados para cada estrutura de rede.

**Tabela 3:** Testes com redes neuronais convencionais.

Camadas	Neurônios por camada	<i>RMSE</i>	$R^2$
1	4	0,377	0,872
1	8	0,358	0,884
1	16	0,350	0,890
1	32	0,358	0,884
1	64	0,359	0,884
2	4	0,362	0,881
2	8	0,350	0,889
2	16	0,338	0,896
2	32	0,351	0,888
2	64	0,346	0,892

Fonte: Do autor.

Pode-se perceber que a rede que apresentou menor erro e maior coeficiente de correlação foi a configuração com duas camadas intermediárias de 16 neurônios cada. No geral, as configurações com duas camadas intermediárias apresentaram resultado superior em relação àquelas com uma, em função da capacidade de predição adicional conferida pelo maior número de parâmetros. No entanto, a partir de certo ponto, aumentar o número de



parâmetros piora o resultado, dado que a rede pode se sobreajustar aos dados de treino e diminuir a capacidade de generalização, tal como observado nas configurações de 32 e 64 neurônios em relação à configuração com 16 neurônios. Dado que, para o problema em questão, o objetivo está na minimização do erro, foi escolhida a estrutura com menor RMSE, sem considerar sua complexidade em um primeiro momento.

#### 4.1.1 Pré-processamento dos dados

Foram realizados testes para verificar se os desenvolvimentos feitos na etapa de pré-processamento dos dados culminaram em melhora dos resultados obtidos com as redes. Para tanto, a Tabela 4 demonstra o RMSE e o coeficiente de determinação para diferentes combinações de características dos dados de entrada, em relação à defasagem temporal e seleção de variáveis. Nota-se que, ao utilizar todo o conjunto de variáveis, ao invés das selecionadas, o erro aumenta e o coeficiente de determinação diminui. No caso da defasagem, o efeito é o mesmo, possuindo, no entanto, um impacto maior no resultado. Desse modo, percebe-se que a seleção de variáveis e defasagem temporal dos dados, buscando representar a dinâmica do processo, melhoram as métricas, sendo aplicados para os demais testes realizados.

**Tabela 4:** Testes com pré-processamento de dados.

Pré-processamento	RMSE	$R^2$
Com seleção de variáveis e com defasagem	0,338	0,896
Sem seleção de variáveis e com defasagem	0,351	0,889
Com seleção de variáveis e sem defasagem	0,378	0,872
Sem seleção de variáveis e sem defasagem	0,398	0,858

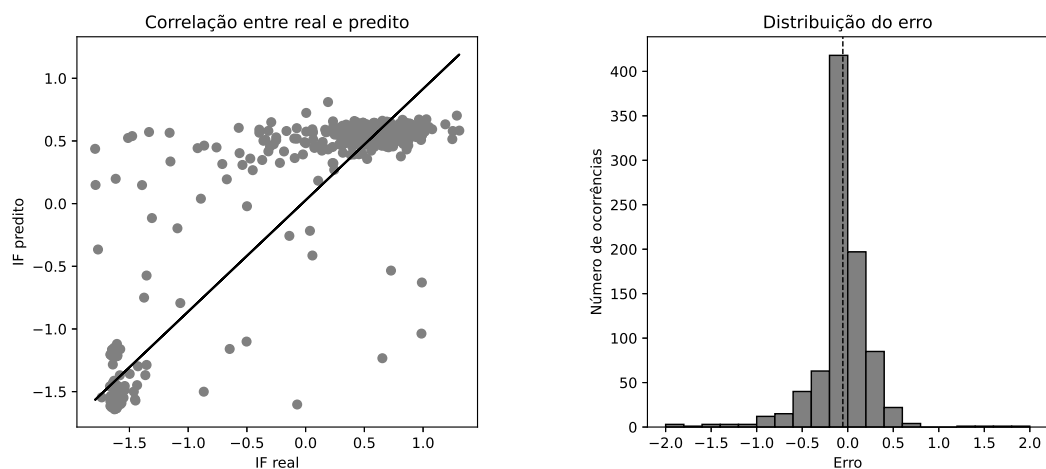
Fonte: Do autor.

#### 4.1.2 Inferência com a rede convencional

A partir da rede com duas camadas intermediárias de 16 neurônios cada, foram gerados os gráficos de correlação entre saída real e predita, bem como a distribuição de erro, demonstrados na Figura 12. É possível verificar, pelo gráfico de correlação, que existem dois grupos de dados, estando um próximo a um IF normalizado de -1,5 e o outro de 0,5. Esses grupos denotam índices de fluidez típicos para os *grades* mais comuns produzidos na unidade industrial. Ao avaliar o IF na faixa de 0,5, percebe-se que a rede apresentou uma dispersão de dados maior do que na faixa inferior. Pelo gráfico de distribuição do erro, observa-se que, mesmo com um erro médio próximo à 0 (representado pela linha tracejada), boa parte do desvio padrão está associado à variabilidade da predição para a faixa alta de IF, conforme observado no gráfico de correlação.

Para verificar o resultado da rede em termos da inferência em si, foi plotada a predição do índice de fluidez do conjunto de teste, comparando-o com o IF real, sendo demonstrado um trecho na Figura 13. Nota-se que a saída apresenta um comportamento semelhante à média dos dados, não representando de forma adequada as variações observadas nos dados reais. Além disso, a rede apresenta algumas alterações de valor mais acentuadas na predição que não foram verificados no conjunto real, como observado próximo à amostra 25 e entre as amostras 100 e 125. Após a amostra 125, é realizado uma transição de produto na unidade industrial, sendo representado de maneira adequada pela rede. Posteriormente, o IF permanece na faixa próxima a -1,5, com a predição apresentando um aparente problema

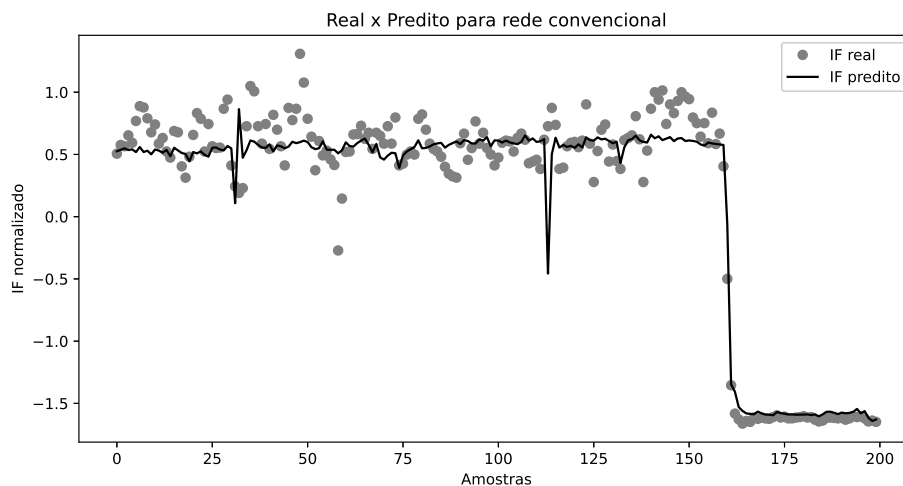
**Figura 12:** Gráfico de correlação e distribuição do erro - rede convencional.



Fonte: Do autor.

de viés. Esse problema também é verificado no gráfico de correlação da Figura 12, dada a dispersão de pontos nessa faixa.

**Figura 13:** Comparação entre saída real e predita - rede convencional.

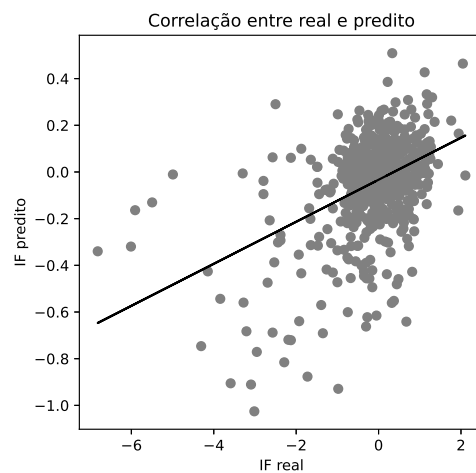


Fonte: Do autor.

Conforme observado pela Figura 12, no gráfico de correlação existem mais pontos na faixa de IF próxima a 0,5, denotando que nessa faixa encontram-se os principais *grades* produzidos na unidade industrial. Além disso, na Figura 13, é possível perceber que, enquanto a faixa de IF próxima a -1,5 possui uma variabilidade menor, a faixa de 0,5 possui uma dispersão maior, sem uma representação adequada por parte da rede neuronal. Então, foi realizado um teste utilizando a mesma estrutura de rede, com duas camadas ocultas de 16 neurônios cada, para realizar a predição exclusivamente para essa região, otimizando os parâmetros apenas para essa faixa de IF, buscando segregar comportamentos distintos entre faixas.

Para tanto, os dados passaram pelos mesmos tratamentos, de modo que, com a nova normalização, os pontos que antes estavam centralizados em 0,5 ficaram centralizados em zero. A Figura 14 demonstra o resultado obtido para a correlação entre valores reais e preditos, sendo observada uma distribuição maior de dados para o IF real entre -2 e 2, enquanto que para o predito concentraram-se entre -0,5 e 0,25. Por meio da distribuição dos pontos, pode-se perceber que o modelo não se ajusta adequadamente aos dados, dada a dispersão.

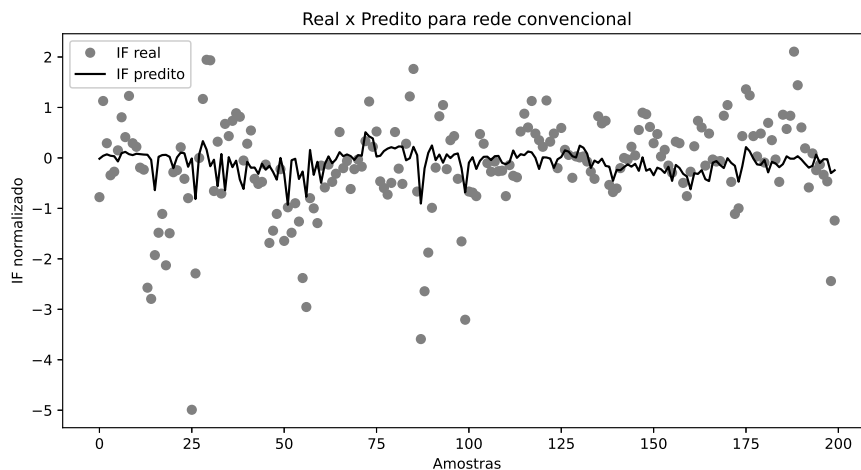
**Figura 14:** Correlação para região de IF - rede convencional.



Fonte: Do autor.

Ao ser comparada a evolução temporal do resultado predito frente ao real, é obtido o gráfico da Figura 15. Pode-se perceber que a rede não consegue representar de maneira adequada o comportamento do sistema, de modo que variações mais significativas dos dados reais não são verificadas no resultado predito. Isso ocorre pela falta de capacidade de predição da rede convencional obtida, não reproduzindo a variabilidade dos dados de entrada na saída, estando mais próximo à uma predição da média.

**Figura 15:** Comparação entre saída real e predita para região de IF - rede convencional.



Fonte: Do autor.

## 4.2 Rede Recorrente

Em função dos problemas de representação da rede convencional, foram realizados testes semelhantes com a rede recorrente. Para tanto, foram variados os números de neurônios e camadas intermediárias da rede, com número de neurônios entre 4 e 64 por camada oculta. O resultado obtido é demonstrado na Tabela 5, com a raiz do erro médio quadrático e o coeficiente de correlação para cada configuração testada, considerando uma média de dez testes realizados com cada estrutura de rede.

**Tabela 5:** Testes com redes neuronais recorrentes.

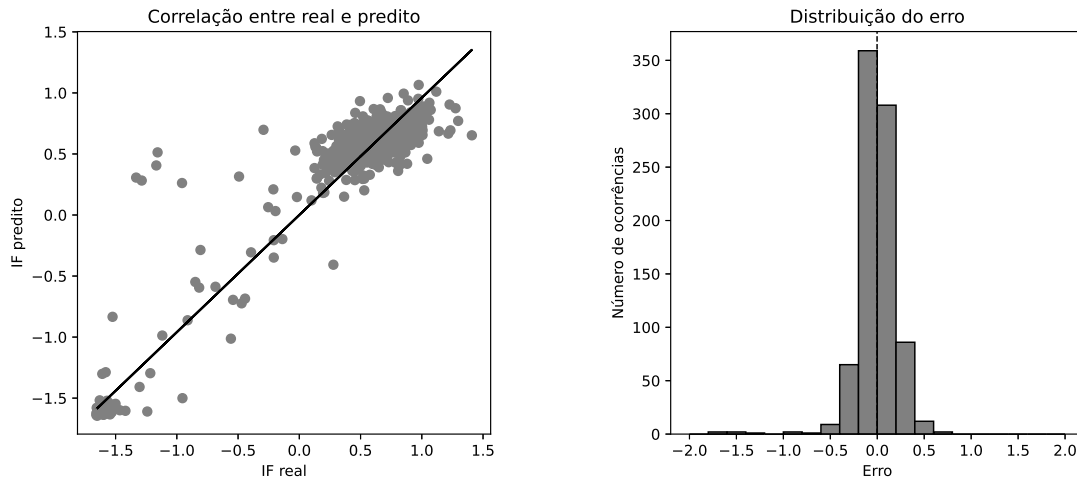
Camadas	Neurônios por camada	RMSE	$R^2$
1	4	0,249	0,936
1	8	0,234	0,943
1	16	0,232	0,944
1	32	0,228	0,946
1	64	0,228	0,946
2	4	0,244	0,939
2	8	0,233	0,944
2	16	0,226	0,946
2	32	0,223	0,949
2	64	0,224	0,948

Fonte: Do autor.

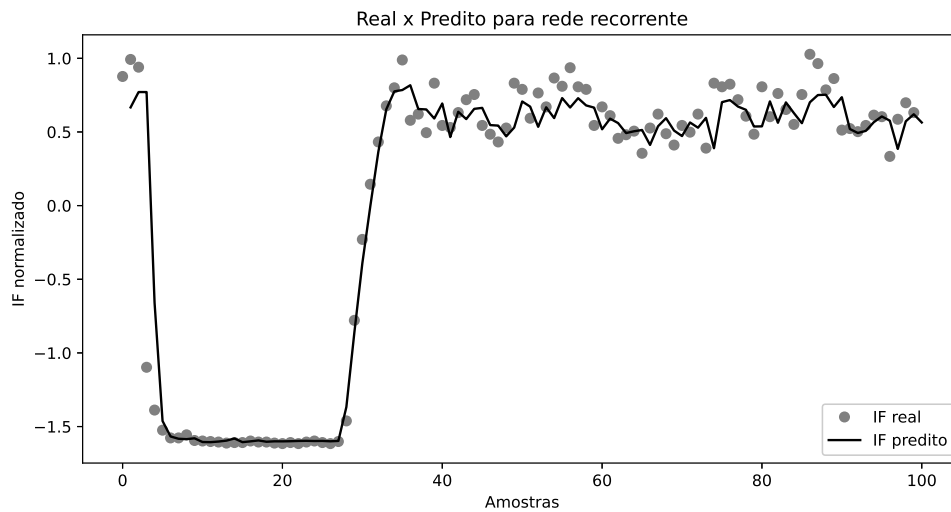
### 4.2.1 Inferência com a rede recorrente

A configuração com melhores resultados (menor erro e maior valor de coeficiente de determinação) foi a rede com duas camadas intermediárias de 32 neurônios cada. No entanto, mesmo apresentando melhores resultados, algumas configurações demonstraram resultados semelhantes, conforme valores observados para as configurações de duas camadas de 16 e 64 neurônios, de modo que poderiam ter sido escolhidas também. Assim, utilizando essa configuração, foram plotados os gráficos de correlação e distribuição do erro, conforme Figura 16.

Observa-se que o gráfico de correlação possui uma dispersão menor dos pontos em relação à rede convencional e que a distribuição do erro possui média de praticamente zero. O resultado melhor em relação à rede convencional, observado também pela raiz do erro médio quadrático, deve-se à maior capacidade de inferência da rede recorrente frente à convencional, dado o estado de recorrência. Desse modo, a rede consegue representar de forma mais adequada a dinâmica do sistema, ou seja, a evolução temporal dos dados de entrada e a relação com a saída. A Figura 17 ilustra temporalmente o comportamento geral da predição em relação aos dados reais. É possível perceber que os movimentos de transição são bem descritos pela rede, com a predição mudando de patamar juntamente com os dados reais do processo (próximo às amostras 10 e 30). Além disso, nota-se que movimentos mais significativos dentro de uma faixa de dados também são representados, tal como a redução do índice de fluidez próximo à amostra 60.

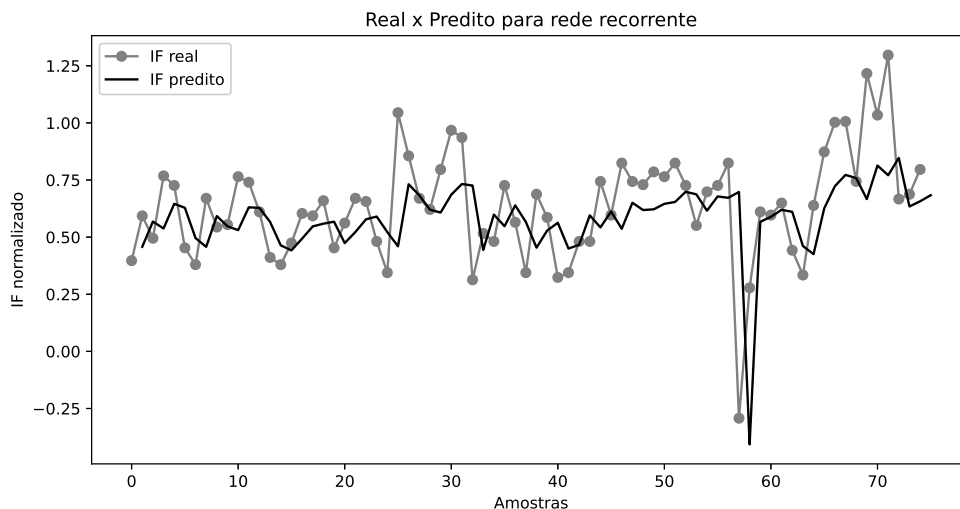
**Figura 16:** Gráfico de correlação e distribuição do erro - rede recorrente.

Fonte: Do autor.

**Figura 17:** Comparação entre saída real e predita - rede recorrente.

Fonte: Do autor.

No entanto, em função do estado de recorrência, em que as amostras anteriores são utilizadas para cômputo do estado atual, são observados problemas de inferência em alguns trechos. A Figura 18 retrata um momento em que a rede recorrente inicialmente fez uma predição média, sem representar adequadamente a variabilidade dos dados reais e, posteriormente, apresentou uma aparente defasagem temporal de uma amostra. Isso ocorre em função da ponderação mais expressiva do estado de recorrência obtida a partir da otimização dos parâmetros da rede recorrente. Dessa forma, é possível obter predições que não seguem o comportamento esperado, realizando movimentos atrasados em relação aos dados reais do processo. Desse modo, foram testadas outras estratégias para tentar mitigar esse problema.

**Figura 18:** Problema de representação da rede recorrente.

Fonte: Do autor.

### 4.3 Rede Combinada

Buscando desenvolver uma estratégia que mitigasse os problemas de representação apresentados da rede recorrente, foi utilizado o conceito de estrutura empilhada de camadas. Conforme exposto por Graves, Mohamed e Hinton (2013), ao adicionar camadas de rede recorrente de forma sequencial, tal como foi feito anteriormente com duas camadas, tende-se a aumentar a capacidade de predição da rede, ao passo em que é aumentado o número de parâmetros treináveis. É possível, também, empilhar diferentes tipos de redes, formando uma estrutura de rede combinada, normalmente com capacidade de predição melhor, uma vez que tende a mitigar problemas de estruturas individuais.

Para o desenvolvimento do modelo combinado, foram testadas diferentes estruturas utilizadas anteriormente nos testes das redes convencionais e recorrentes. Assim, por simplificação, foram selecionadas apenas as redes que apresentaram melhor resultado para cada tipo, tanto para uma camada como para duas. Para a rede convencional, as estruturas com melhores resultados foram as de 16 neurônios, para uma e duas camadas. Já para a rede recorrente, foram as redes com uma e duas camadas de 32 neurônios cada. Assim, conforme proposto por Graves, Mohamed e Hinton (2013), foram compostas quatro estruturas de modelos sequenciais combinados para testes, por meio dos modelos recorrente e convencional:

- a. 1 camada recorrente (32 neurônios) e 1 camada convencional (16 neurônios);
- b. 1 camada recorrente (32 neurônios) e 2 camadas convencionais (16 neurônios);
- c. 2 camadas recorrentes (32 neurônios) e 1 camada convencional (16 neurônios);
- d. 2 camadas recorrentes (32 neurônios) e 2 camadas convencionais (16 neurônios).

Com as estruturas definidas, foram feitos os mesmos testes com o conjunto de dados, avaliando RMSE e o coeficiente de determinação, conforme demonstrado na Tabela 6. Vale

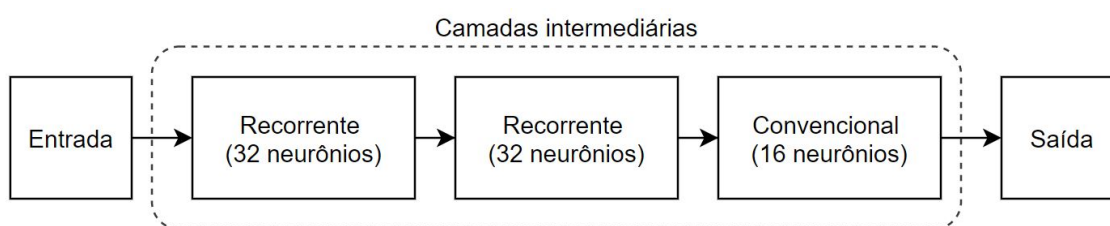
destacar que o treinamento da rede foi feito da mesma forma, por meio de inicialização aleatória dos parâmetros e épocas de treinamento para ajuste dos mesmos, tal como descrito pela Equação (4). Observa-se que o resultado obtido para as estruturas combinadas foi superior em relação às redes simplesmente convencionais ou recorrentes. A estrutura com menor erro dentre as avaliadas foi a com duas camadas recorrentes de 32 neurônios e uma camada convencional de 16 neurônios, conforme esquemático demonstrado na Figura 19, compondo o modelo combinado.

**Tabela 6:** Testes com estruturas combinadas.

Estrutura	RMSE	$R^2$
a	0,214	0,953
b	0,218	0,952
c	0,208	0,956
d	0,211	0,955

Fonte: Do autor.

**Figura 19:** Estrutura final da rede.



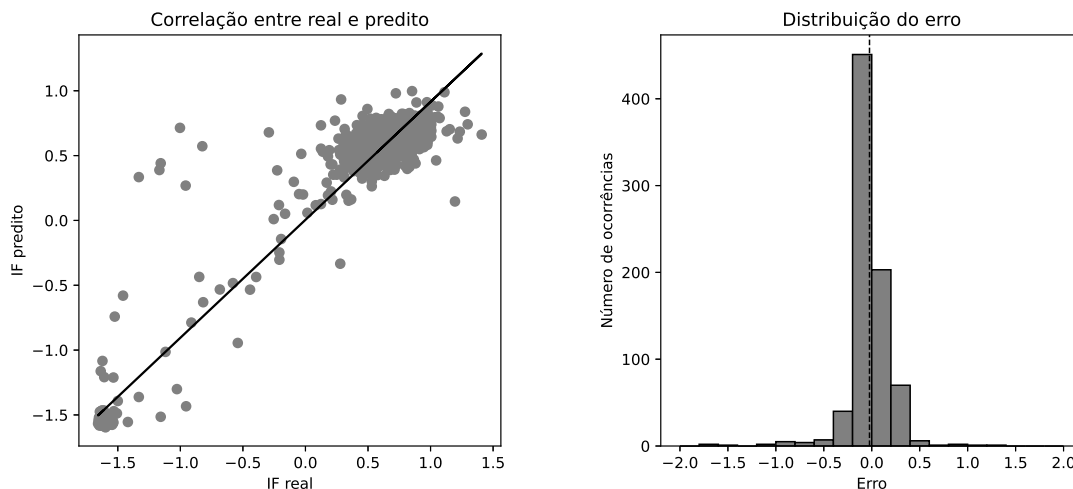
Fonte: Do autor.

#### 4.3.1 Inferência com a rede combinada

Os resultados de teste obtidos indicaram, então, RMSE de 0,208 e 0,956 para o coeficiente de determinação. Da mesma forma que antes, foram gerados os gráficos de correlação entre a saída real e predita, bem como a distribuição do erro, sendo demonstrados na Figura 20. Por meio do gráfico de correlação, é possível perceber que os dados estão mais concentrados no entorno dos pontos de operação típicos da planta, com alguns deles possuindo um desvio maior. Também é possível verificar que a distribuição do erro apresenta média de aproximadamente zero, com a maior parte do erro concentrado entre -0,25 e 0,25.

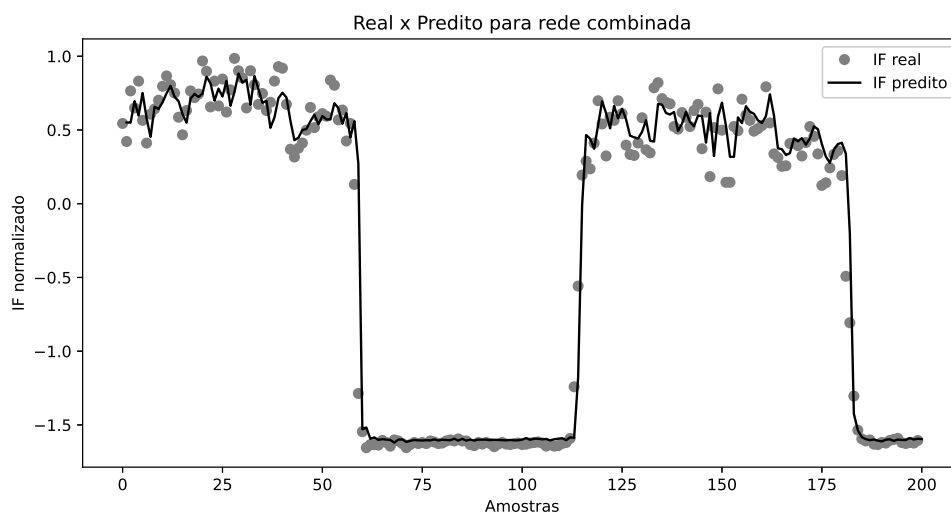
Para visualizar temporalmente a saída da rede e compará-la com os dados reais da unidade industrial, é demonstrado, pela Figura 21, um segmento do conjunto de testes. É possível perceber que boa parte da variabilidade dos dados reais é representada pela rede, principalmente nos períodos de variações mais expressivas do IF. No entanto, mesmo com o resultado melhor em termos de erro e coeficiente de determinação, a faixa de IF próxima a -1,5 apresenta um aparente problema de viés, além de não ter representado de forma satisfatória os dados reais de processo. Uma vez que os *grades* mais produzidos na unidade industrial estão contemplados no patamar superior de IF, o resultado observado para a faixa inferior não interferiu significativamente nos resultados da rede.

**Figura 20:** Gráfico de correlação e distribuição do erro - rede combinada.



Fonte: Do autor.

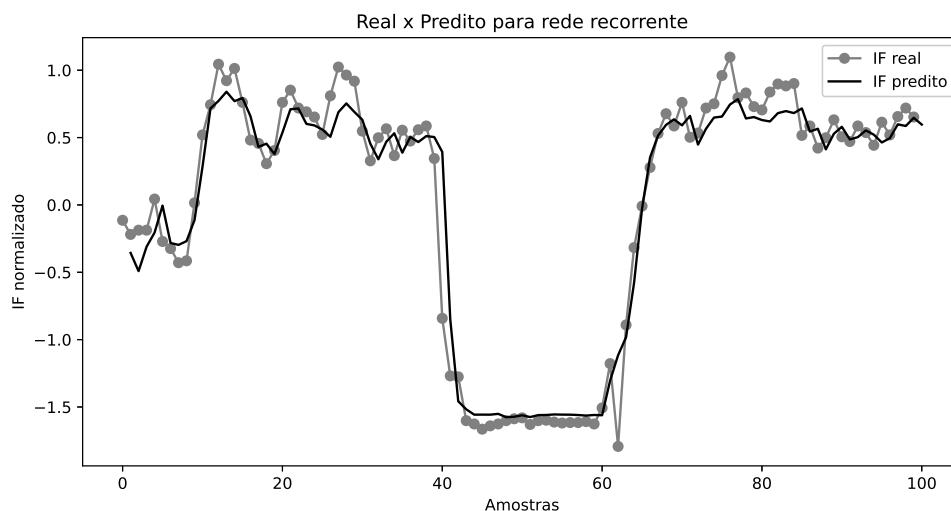
**Figura 21:** Comparação entre saída real e predita - rede combinada.



Fonte: Do autor.

Buscando detalhar esse comportamento, a Figura 22 demonstra 100 amostras do conjunto de testes. Pode-se perceber que os momentos de transição entre níveis de índice de fluidez são representados pela rede, ao passo que a faixa inferior (-1,5) carece de uma boa aderência aos dados reais. Um dos objetivos da arquitetura combinada era reduzir o problema da rede puramente recorrente, em que a ponderação para a saída predita ocasionava, em alguns trechos, defasagem em relação aos dados reais, além de não representar satisfatoriamente variações menos abruptas dos dados reais. Na arquitetura combinada proposta, nota-se que existem trechos em que ainda são observados problemas semelhantes. No entanto, mesmo não havendo uma forma quantitativa de avaliar a melhoria nesse caso, pode-se verificar um comportamento mais condizente da saída predita com os dados reais.



**Figura 22:** Detalhamento da saída real e predita - rede combinada.

Fonte: Do autor.

A fim de analisar se o modelo final desenvolvido cumpria os critérios mencionados anteriormente na Seção 3.1.2, em que é considerado satisfatório um erro médio de 10%, foi verificada a média do erro absoluto em relação aos dados reais, ou seja, qual a variação média da predição frente ao IF desejado. Assim, considerando o conjunto de testes, obteve-se um erro de aproximadamente 13% em relação aos dados reais, estando acima do valor desejado para o modelo. No entanto, uma vez que os produtos principais da unidade industrial são produzidos com IF mais elevado, para dados exclusivamente dessa faixa o erro médio foi de cerca de 8%. Dessa forma, pode-se dizer que o modelo obtido possui capacidade de predição satisfatória para boa parte do conjunto de dados, ao passo que necessita refinamento em alguns trechos.

Vale destacar também que o erro, além de estar associado à estrutura da rede e à capacidade de predição da mesma, também é impactado pelos dados de entrada e os tratamentos feitos nos mesmos, tais como defasagem temporal, escolha de variáveis, normalização. Então, o erro final obtido é, na verdade, um efeito combinado e propagado pela rede, sendo decorrente dos desenvolvimentos feitos, desde a etapa de coleta (estando sujeita a problemas de instrumentação), até a inferência dos modelos desenvolvidos. Acerca das variáveis, além da seleção das mesmas poder ocasionar perda de representação da dinâmica, existem ainda variáveis não medidas, ou de difícil medição, tais como grau de ativação do catalisador e presença de contaminantes na alimentação que não constam no modelo e que possuem impacto, direto ou indireto, no índice de fluidez do polímero.

## 5 CONCLUSÃO

No presente trabalho, foram apresentadas as etapas e estratégias envolvidas no desenvolvimento de um modelo empírico para o índice de fluidez de resinas de polietileno produzidas em reator *Slurry*. Já que o processo de polimerização é não-linear e envolve os efeitos de diversas variáveis de processo, foram abordadas diferentes metodologias para obter um modelo que atendesse aos objetivos propostos.

Por meio dos resultados obtidos, conclui-se que a rede convencional, sem realimentação, não consegue representar a característica dos dados reais de IF, dada a simplicidade do modelo frente ao problema. Por outro lado, a rede recorrente apresenta um resultado melhor, visto que o uso do estado passado consegue modelar, de forma aproximada, a dinâmica do sistema. No entanto, ainda que em menor grau, a rede puramente recorrente pode apresentar erros, uma vez que a otimização pode conduzir a um ponderamento elevado do estado de recorrência, conferindo uma característica de defasagem da predição em certos intervalos.

A rede combinada, em que foram utilizadas as estruturas da rede recorrente e convencional, apresentou melhora nas métricas de erro e coeficiente de determinação, visto que busca combinar dois modelos, mitigando problemas de representação das estruturas individuais. Contudo, mesmo com resultado satisfatório, observou-se que o modelo ainda não foi capaz de prever com exatidão valores na faixa inferior de IF, apresentando erro de viés. No caso de movimentos de transição de carga (observados pelas variações mais significativas do IF), o modelo conseguiu representar adequadamente esse período crítico para a unidade industrial.

Considerando as características da unidade industrial e do processo de polimerização em si, conseguiu-se obter um modelo capaz de inferir a característica geral do índice de fluidez, apresentando, no entanto, alguns problemas de representação de dinâmicas do sistema. Uma vez que, para o controle do processo, são realizadas atualmente alterações combinadas de variáveis, impactando de maneira diferente a saída, entendem-se algumas limitações do modelo. Vale destacar também que, mesmo que o modelo possa melhorar a capacidade de atuação sobre o processo desses períodos sem medições laboratoriais, o mesmo não substitui as análises, já que a complexidade e variação temporal do sistema implica atualização constante dos modelos, podendo-se, inclusive, utilizar os dados reais para realizar correções das predições.

Visando melhorar o desempenho do modelo, poderiam ser avaliadas situações específicas do funcionamento do reator e do processo em si, de modo a melhorar a ponderação proposta para as variáveis e representar de maneira mais significativa o sistema. Adicionalmente, como trabalhos futuros, poderiam ser testadas outras combinações de estruturas de redes neurais, além das estruturas convencionais e recorrentes, tais como as redes convolucionais. Uma alternativa também seria a composição com um modelo fenomenológico,

criando um modelo híbrido, de modo a tentar representar de maneira mais adequada as dinâmicas do sistema não modeladas plenamente pelas redes neuronais.

Em termos do resultado das redes neuronais, além da capacidade de predição das mesmas e da estrutura desenvolvida, uma das fontes de erro é decorrente das manipulações feitas com os dados de entrada. Ao passo que a utilização de uma técnica de tratamento de dados possa não impactar significativamente o resultado, ao combinar diferentes técnicas e aplicá-las ao conjunto de dados, o erro pode se tornar suficientemente expressivo frente ao erro de treinamento das redes. Para tanto, deve-se avaliar com mais detalhes o efeito dessas estratégias no resultado final e verificar se as considerações feitas foram adequadas.

## REFERÊNCIAS

- AHMED, F.; NAZIR, S.; YEO, Y. K. A recursive PLS-based soft sensor for prediction of the melt index during grade change operations ins HDPE plant. *Korean Journal of Chemical Engineering*, v. 26, n. 1, p. 14–10, 2009.
- ALHINDAWI, F.; ALLAF, R.; ALTARAZI, S. Machine Learning Models for Predicting and Classifying the Tensile Strenght of Polymeric Films Fabricated via Different Production Processes. *Journal of Materials*, v. 12, p. 1475–1488, 2019.
- AWAD, A. H. et al. Mechanical and Physical Properties of PP and HDPE. *Engineering Science*, v. 4, n. 2, p. 34–42, 2019.
- BRYCE, D. M. *Plastic Injection Molding: Manufacturing Process Fundamentals*. Michigan: Society of Manufacturing Engineers, 1996. p. 277. ISBN 0-08726-3472-8.
- CASARI, A.; ZHENG, A. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. Sebastopol: O’Reilly Media, 2018. p. 218. ISBN 0-14919-5324-2.
- CHIANG, L. H.; QIN, S. J. Advances and opportunities in machine learning for process data analytics. *Computers and Chemical Engineering*, v. 126, p. 465–473, 2019.
- CHICCO, D.; JURMAN, G.; WARRENS, M. J. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RSME in regression analysis evaluation. *PeerJ Computer Science*, v. 7, 2021.
- CHIKKALI, S.; PATEL, K.; SIVARAM, S. Ultrahigh molecular weight polyethylene: Catalysys, structure, properties, processing and applications. *Progress in Polymer Science*, v. 109, 2020.
- CHUN, T. H. et al. Industrial Revolution 4.0 Contibution of Chemical Engineer. *Turkish Journal of Computer and Mathematics Education*, v. 12, n. 11, p. 5037–5041, 2021.
- ELISSEEFF, A.; GUYON, I. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research*, v. 3, p. 1157–1182, 2003.
- GABRYS, B.; KADLEC, P.; STRANDT, S. Data-driven Soft Sensors in the process industry. *Computers and Chemical Engineering*, v. 33, p. 795–814, 2009.
- GRAVES, A.; MOHAMED, A.-R.; HINTON, G. Speech recognition with deep recurent neural networks. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, p. 6645–6649, 2013.
- HAYKIN, S.; LEUNG, H. The complex backpropagation algorithm. *IEEE Transactions on Signal Processing*, v. 39, n. 9, p. 2101–2104, 1991.

- HILAL-ALNAQBI, A.; MOURAD, A.; YOUSEF, B. F. Prediction of the Mechanical Properties of PE/PP Blends Using Artificial Neural Networks. *Procedia Engineering*, v. 10, p. 2713–2718, 2011.
- HINTON, G. et al. Dropout: A Simple Way to Prevent Neural Network from Overfitting. *Journal of Machine Learning Research*, v. 15, p. 1929–1958, 2014.
- HOSSEINKHANI, J.; IBRAHIM, S.; KOUPAIE, H. M. Outlier Detection in Stream Data by Machine Learning and Feature Selection Methods. *International Journal of Advanced Computer Science and Information Technology (IJACSIT)*, v. 2, n. 3, p. 17–24, 2013.
- ILYAS, A. et al. How Does Batch Normalization Help Optimization? *32nd Conference on Neural Information Processing Signals (NeurIPDS)*, Montreal, Canadá, 2018.
- JHA, R.; SAHOO, B. B.; SIGNH, A. Long Short-Term Memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys*, v. 67, p. 1471–1481, 2019.
- KANG, I.; KIM, Y.; PARK, S. On-line Property Prediction for Industrial Slurry High-Density Polyethylene Process with Various Grade Transition Modes. *Journal of Chemical Engineering of Japan*, v. 43, n. 7, p. 581–594, 2010.
- KIM, T. Y.; LEE, E. H.; YEO, Y. K. Prediction and quality control of the melt index during production of high-density polyethylene. *Korean Journal of Chemical Engineering*, v. 25, n. 4, p. 613–622, 2008.
- LEE, E. H. et al. Modeling and control of the melt index in HDPE process. *Korean Journal of Chemical Engineering*, v. 26, n. 1, p. 7–13, 2008.
- MAINKA, C. P. *Influência dos Parâmetros Moleculares nas Propriedades do Polietileno de Alta Densidade*. 2018. Tese (Mestrado em Engenharia) – Programa de Pós-Graduação em Engenharia de Minas, Metalúrgica e de Materiais, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- MCAULEY, K. B.; MCGREGOR, J. F. On-line inference of Polymer Properties in an industrial polyethylene reactor. *AIChE*, v. 37, n. 4, p. 825–835, 1991.
- MOHR, T. *Desenvolvimento de um Analisador Virtual para uma Planta de Polietileno de Alta Densidade*. 2004. Tese (Mestrado em Engenharia) – Programa de Pós-Graduação em Engenharia Química, Universidade Federal do Rio Grande do Sul, Porto Alegre.
- OSTERTAGOVA, E. Modelling using polynomial regression. *Procedia Engineering*, v. 48, p. 500–506, 2012.
- PEACOCK, A. J. *Handbook of Polyethylene: Structures, Properties and Applications*. New York: Marcel Dekker, Inc., 2000. p. 556. ISBN 0-08247-9546-6.
- RASKUTTI, G.; WAINWRIGHT, M. J.; YU, B. Early Stopping and Non-parametric Regression: An Optimal Data-dependent Stopping Rule. *Journal of Machine Learning Research*, v. 15, p. 335–366, 2014.
- SHERSTINSKY, A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena*, v. 404, 2020.