

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE
MINAS GERAIS - *CAMPUS* SABARÁ
BACHARELADO EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Gabriel Carmo Sampaio

**MODELAGEM PREDITIVA MULTI-SAÍDA DE UM REATOR
CONTÍNUO DE TANQUE AGITADO: estimacão de concentraão e
temperatura utilizando redes neurais LSTM**

Sabar
2026

GABRIEL CARMO SAMPAIO

**MODELAGEM PREDITIVA MULTI-SAÍDA DE UM REATOR
CONTÍNUO DE TANQUE AGITADO: estimação de concentração e
temperatura utilizando redes neurais LSTM**

Trabalho de Conclusão de Curso apresentado à banca examinadora do curso de Engenharia de Controle e Automação do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais *Campus* Sabará, como parte dos requisitos para obtenção do título de Bacharel em Engenharia de Controle e Automação.

Orientador: Profa. Dra. Mariela Maia Quadros

Sabará
2026

Sampaio, Gabriel Carmo

S192m

Modelagem preditiva multi-saída de um reator contínuo de tanque agitado [manuscrito] : estimação de concentração e temperatura utilizando redes neurais LSTM. / Gabriel Carmo Sampaio. - 2026.

64 f. : il.

Orientador(a): Dra. Mariela Maia Quadros.

Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Controle e Automação) – Instituto Federal de Minas Gerais, *Campus* Sabará.

1. Controle preditivo. – Monografia. 2. Redes neurais (Computação). – Monografia. 3. Simulação (Computadores). – Monografia. 4. Engenharia de sistemas. – Monografia. 5. Análise de séries temporais. – Monografia. I. Quadros, Mariela Maia. II. Instituto Federal de Minas Gerais, *Campus* Sabará. Bacharelado em Engenharia de Controle e Automação. III. Título.

CDU 681.5.017

César dos Santos Moreira / CRB6-2229
Biblioteca do IFMG *Campus* Sabará



MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS
Campus Sabará

Diretoria de Ensino, Pesquisa e Extensão
Coordenação do Curso de Engenharia de Controle e Automação
Rodovia MGC 262, Km 10 - Bairro Sobradinho - CEP 34590-390 - Sabará - MG - www.ifmg.edu.br

ATA DE DEFESA DE TRABALHO DE CONCLUSÃO DE CURSO

Aos quatorze dias do mês de julho do ano de dois mil e vinte e seis, às dezoito horas, sob a presidência da professora Mariella Maia Quadros, reuniu-se a banca examinadora composta pelos professores abaixo relacionados, para a defesa do Trabalho de Conclusão de Curso (TCC) do discente **Gabriel Carmo Sampaio**, matrícula nº 0059757, do curso de Engenharia de Controle e Automação, do IFMG *campus* Sabará.

O trabalho intitulado "MODELAGEM PREDITIVA MULTI-SAÍDA DE UM REATOR CONTÍNUO DE TANQUE AGITADO: estimação de concentração e temperatura utilizando redes neurais LSTM" foi apresentado e submetido à apreciação da banca. Após exposição, arguição e deliberação, a banca atribuiu a nota final de 75 pontos (de um total de 80), resultando em **aprovado**, condicionado ao cumprimento das orientações e prazos estabelecidos pelas normas acadêmicas institucionais.

Compuseram a Banca Examinadora:

- Membro 1 (Presidente): Profa. Mariella Maia Quadros (orientadora)
- Membro 2: Prof. Diego Oliveira Miranda
- Membro 3: Prof. Luiz Guilherme Hilel Drumond Silveira

O discente deverá apresentar a versão final do trabalho em formato PDF e depositá-la no repositório institucional até o dia vinte e sete de julho de dois mil e vinte e seis. O não cumprimento dessas exigências implicará na não contabilização das horas referentes aos componentes curriculares de TCC I e TCC II no sistema acadêmico.

A sessão foi encerrada às vinte horas. Para constar, eu, Mariella Maia Quadros, redigi a presente ata que após lida publicamente, foi aprovada e assinada pelos membros da banca examinadora.

Sabará, 14 de julho de 2026.



Documento assinado eletronicamente por **Mariella Maia Quadros, Professora EBTT**, em 14/07/2026, às 20:54, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Luiz Guilherme Hilel Drumond Silveira, Professor**, em 14/07/2026, às 21:16, conforme Decreto nº 10.543, de 13 de novembro de 2020.



Documento assinado eletronicamente por **Diego Oliveira Miranda, Professor EBTT**, em 15/07/2026, às 07:08, conforme Decreto nº 10.543, de 13 de novembro de 2020.



A autenticidade do documento pode ser conferida no site <https://sei.ifmg.edu.br/consultadocs> informando o código verificador **2809111** e o código CRC **C7E59A31**.

23714.001466/2025-48	2600082v1
----------------------	-----------

Dedico este trabalho a Deus, fonte de força e sabedoria, que me guiou em cada etapa desta caminhada. À minha família, pelo amor, apoio e incentivo constantes. E aos meus amigos, que estiveram ao meu lado, tornando esta trajetória mais leve e significativa.

“Seu foco determina sua realidade.”

Qui-Gon Jinn, *Star Wars: A Ameaça Fantasma*

RESUMO

Este trabalho apresenta o desenvolvimento e a avaliação de um modelo preditivo multi-saída baseado em rede neural de memória de curto e longo prazo (LSTM, do inglês *Long Short-Term Memory*) para representar a dinâmica de um reator contínuo de tanque agitado (CSTR, do inglês *Continuous Stirred-Tank Reactor*). O estudo utiliza o conjunto de dados CSTR da base DaISy, composto por 7500 amostras com intervalo de amostragem de 0,1 minuto, equivalente a 6 segundos. A proposta tem como objetivo estimar simultaneamente a concentração e a temperatura do processo a partir do histórico recente de operação, sendo analisada como um componente preditivo orientado a dados. A implementação foi desenvolvida em Python, utilizando como entradas históricas a vazão de resfriamento, a concentração e a temperatura, organizadas em janelas temporais de 50 amostras. Os dados foram normalizados por média e desvio padrão calculados apenas no trecho de treinamento, e o modelo foi treinado com função de perda baseada no erro quadrático médio, otimizador Adam e mecanismos de parada antecipada, redução da taxa de aprendizado e salvamento do melhor modelo. A avaliação considerou testes de um passo à frente, teste externo, avaliação recursiva e estimação em múltiplos passos, sempre comparando a rede LSTM com um modelo de persistência. Os resultados indicaram que a rede LSTM apresentou melhor capacidade de estimar as duas variáveis de saída, acompanhando de forma mais próxima a dinâmica do processo do que a estratégia de persistência, principalmente em horizontes superiores a um passo. Conclui-se que a abordagem proposta é adequada como componente preditivo para o CSTR analisado, respeitando os limites do conjunto de dados utilizado. Entretanto, o modelo desenvolvido não constitui, por si só, um gêmeo digital completo, pois não inclui integração em tempo real com uma planta física, monitoramento operacional contínuo ou atuação automática sobre o processo.

Palavras-chave: reator contínuo de tanque agitado; LSTM; gêmeo digital; identificação de sistemas; séries temporais.

ABSTRACT

This work presents the development and evaluation of a multi-output predictive model based on a Long Short-Term Memory (LSTM) neural network to represent the dynamics of a Continuous Stirred-Tank Reactor (CSTR). The study uses the CSTR dataset from the DaISy database, composed of 7500 samples with a sampling interval of 0.1 minute, equivalent to 6 seconds. The proposal aims to simultaneously estimate the process concentration and temperature from the recent operating history and is analyzed as a data-driven predictive component. The implementation was developed in Python, using cooling flow rate, concentration and temperature as historical inputs, organized into temporal windows of 50 samples. The data were standardized using mean and standard deviation computed only from the training segment, and the model was trained with a mean squared error loss function, Adam optimizer, early stopping, learning rate reduction and best model checkpointing. The evaluation considered one-step-ahead tests, external testing, recursive evaluation and multi-step forecasting, always comparing the LSTM network with a persistence model. The results indicated that the LSTM network was better able to estimate both output variables, following the process dynamics more closely than the persistence strategy, especially for horizons longer than one step. It is concluded that the proposed approach is suitable as a predictive component for the analyzed CSTR, within the limits of the dataset used. However, the developed model does not constitute a complete digital twin by itself, since it does not include real-time integration with a physical plant, continuous operational monitoring or automatic actuation on the process.

Keywords: continuous stirred-tank reactor; LSTM; digital twin; system identification; time series.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação conceitual do CSTR e das variáveis utilizadas no modelo. . .	17
Figura 2 – Representação esquemática de um neurônio artificial.	20
Figura 3 – Representação conceitual de uma rede neural artificial em camadas.	21
Figura 4 – Representação conceitual de uma célula LSTM.	24
Figura 5 – Relação entre temperatura e concentração no conjunto de dados e resíduos do ajuste linear.	41
Figura 6 – Relação entre temperatura e concentração no conjunto de dados, com destaque para o trecho usado no treinamento.	42
Figura 7 – Comportamento temporal de q_c , C_A e T no conjunto de dados CSTR.	43
Figura 8 – Evolução da função de perda durante o treinamento da rede LSTM.	44
Figura 9 – Comparação entre valores reais, LSTM e persistência no teste interno.	45
Figura 10 – Visão ampliada da comparação entre valores reais, LSTM e persistência no teste interno.	46
Figura 11 – Erro de estimação da LSTM no teste interno.	46
Figura 12 – Comparação entre valores reais, LSTM e persistência no teste externo de um passo à frente.	47
Figura 13 – Visão ampliada da comparação entre valores reais, LSTM e persistência no teste externo de um passo à frente.	48
Figura 14 – Estimação externa recursiva da LSTM para concentração e temperatura. . .	49
Figura 15 – Erro da estimação externa recursiva da LSTM.	49
Figura 16 – Estimação de múltiplos passos da concentração para horizonte de 10 passos. .	51
Figura 17 – Estimação de múltiplos passos da temperatura para horizonte de 50 passos. .	51
Figura 18 – Evolução da função de perda da LSTM no experimento complementar com horizonte de 5 passos.	53
Figura 19 – Comparação entre valores reais, LSTM e persistência no experimento complementar com horizonte de 5 passos.	53
Figura 20 – Visão ampliada da comparação entre valores reais, LSTM e persistência no experimento complementar com horizonte de 5 passos.	54
Figura 21 – Erro de estimação da LSTM no experimento complementar com horizonte de 5 passos.	54

LISTA DE TABELAS

Tabela 1	– Principais variáveis do processo CSTR.	18
Tabela 2	– Variáveis utilizadas no modelo preditivo.	33
Tabela 3	– Parâmetros de normalização utilizados no experimento.	34
Tabela 4	– Arquitetura da rede LSTM utilizada.	36
Tabela 5	– Resultados no teste interno de um passo à frente.	44
Tabela 6	– Resultados no teste externo de um passo à frente.	47
Tabela 7	– Resultados da avaliação externa recursiva da LSTM.	48
Tabela 8	– Resultados da avaliação recursiva de múltiplos passos.	50
Tabela 9	– Resultados da LSTM e da persistência no experimento complementar com horizonte de 5 passos.	52

LISTA DE ABREVIATURAS E SIGLAS

CSTR	<i>Continuous Stirred-Tank Reactor</i> – Reator Contínuo de Tanque Agitado
DaISy	<i>Database for the Identification of Systems</i> – Base de Dados para Identificação de Sistemas
DT	<i>Digital Twin</i> – Gêmeo Digital
LSTM	<i>Long Short-Term Memory</i> – Memória de Curto e Longo Prazo
MAE	<i>Mean Absolute Error</i> – Erro Absoluto Médio
ML	<i>Machine Learning</i> – Aprendizado de Máquina
MLP	<i>Multilayer Perceptron</i> – Perceptron Multicamadas
MSE	<i>Mean Squared Error</i> – Erro Quadrático Médio
RMSE	<i>Root Mean Squared Error</i> – Raiz do Erro Quadrático Médio
RNN	<i>Recurrent Neural Networks</i> – Redes Neurais Recorrentes

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Objetivos	13
1.1.1	<i>Objetivo geral</i>	13
1.1.2	<i>Objetivos específicos</i>	13
1.2	Justificativa	13
2	FUNDAMENTAÇÃO TEÓRICA	16
2.1	Reator Contínuo de Tanque Agitado	16
2.2	Identificação de Sistemas e Modelagem Orientada a Dados	19
2.3	Aprendizado de Máquina e Redes Neurais Artificiais	19
2.4	Redes de Memória de Curto e Longo Prazo	22
2.5	Gêmeos Digitais Preditivos	25
2.6	Normalização e Organização Temporal de Séries	25
2.7	Validação Temporal, Métricas e Persistência	26
3	TRABALHOS RELACIONADOS	28
4	METODOLOGIA	32
4.1	Conjunto de Dados e Variáveis Utilizadas	32
4.2	Recortes Temporais e Divisão dos Dados	33
4.3	Normalização dos Dados	34
4.4	Construção das Janelas Temporais	34
4.5	Arquitetura da Rede LSTM	36
4.6	Treinamento	37
4.7	Modelo Simples de Comparação	38
4.8	Avaliação Interna	38
4.9	Avaliação Externa de Um Passo à Frente	38
4.10	Experimento Complementar com Horizonte de 5 Passos	39
4.11	Avaliação Recursiva e Multi-step	39
5	RESULTADOS	41
5.1	Representatividade do Treinamento e Relação entre Variáveis	41
5.2	Caracterização do Conjunto de Dados	43

5.3	Evolução do Treinamento	44
5.4	Teste Interno de Um Passo à Frente	44
5.5	Teste Externo de Um Passo à Frente	47
5.6	Avaliação Externa Recursiva	48
5.7	Avaliação Recursiva Multi-step	50
5.8	Análise Complementar com Horizonte de Estimação de 5 Passos	51
5.9	Discussão dos Resultados	54
6	CONCLUSÃO	56
	REFERÊNCIAS	58
	APÊNDICE A – TRECHOS PRINCIPAIS DA IMPLEMENTAÇÃO COMPUTACIONAL	60
A.1	Configuração do experimento	60
A.2	Normalização	61
A.3	Construção das janelas temporais	61
A.4	Arquitetura e treinamento da LSTM	62
A.5	Modelo simples de comparação	63

1 INTRODUÇÃO

A modelagem de sistemas dinâmicos é uma atividade importante na Engenharia de Controle e Automação, pois permite compreender, estimar e analisar o comportamento de processos ao longo do tempo. Em ambientes industriais, essa tarefa pode ser desafiadora devido à presença de não linearidades, ruídos de medição, atrasos, perturbações e interação entre diferentes grandezas do processo.

Tradicionalmente, muitos processos são modelados a partir de princípios físicos, como balanços de massa, balanços de energia e equações diferenciais. Essa abordagem é essencial para interpretar os fenômenos envolvidos e continua sendo uma base importante na engenharia (OGATA, 2010; SEBORG *et al.*, 2010). Entretanto, em alguns casos, a obtenção de um modelo analítico suficientemente preciso pode exigir informações detalhadas da planta, parâmetros difíceis de medir e hipóteses simplificadoras que nem sempre representam bem o comportamento real do sistema (LJUNG, 1999; AGUIRRE, 2007).

Nesse contexto, modelos orientados a dados surgem como uma alternativa complementar para a representação de processos dinâmicos. Em vez de depender apenas de equações obtidas a partir da descrição física completa do sistema, esses modelos utilizam dados históricos de operação para aprender relações entre entradas e saídas. Essa estratégia tem se tornado cada vez mais relevante com o avanço de técnicas de inteligência artificial e com a maior disponibilidade de dados em sistemas industriais.

O processo utilizado como estudo de caso neste trabalho é um CSTR. Esse tipo de reator é bastante utilizado em estudos de controle, identificação de sistemas e modelagem de processos por reunir características comuns em plantas industriais, como comportamento não linear, dependência temporal e acoplamento entre variáveis. Os dados utilizados pertencem ao conjunto CSTR da base DaISy, uma base pública voltada à identificação de sistemas dinâmicos. Esse conjunto está associado ao processo apresentado por Morningred *et al.* (1992), no qual a vazão de resfriamento influencia a concentração do reagente e a temperatura do reator.

De forma geral, o interesse deste trabalho está em construir uma representação computacional capaz de estimar o comportamento futuro de grandezas relevantes do reator a partir de seu histórico recente de operação. Para isso, utiliza-se uma rede LSTM, arquitetura adequada para problemas envolvendo séries temporais e dependências entre amostras ao longo do tempo (HOCHREITER; SCHMIDHUBER, 1997).

A proposta se relaciona ao conceito de gêmeo digital, entendido de forma geral como uma representação virtual associada a um sistema físico, capaz de acompanhar, simular ou estimar seu comportamento (GRIEVES, 2014; TAO *et al.*, 2019). Entretanto, o trabalho não implementa um gêmeo digital completo conectado a uma planta em tempo real. O foco está no desenvolvimento de um modelo preditivo orientado a dados que pode compor uma etapa inicial ou o núcleo de estimação de um gêmeo digital.

Dessa forma, este trabalho busca integrar conceitos de modelagem de sistemas dinâmicos, identificação de sistemas, redes neurais e gêmeos digitais em uma aplicação voltada à Engenharia de Controle e Automação. A pesquisa não tem como objetivo substituir a modelagem física do processo, controlar a planta ou entregar uma estrutura completa de gêmeo digital, mas avaliar a capacidade de um modelo baseado em dados de representar a dinâmica do reator em condições definidas pelo conjunto de dados utilizado.

1.1 Objetivos

1.1.1 Objetivo geral

O objetivo geral deste trabalho consiste em desenvolver e avaliar um modelo preditivo multi-saída baseado em rede neural LSTM para estimar a concentração e a temperatura de um CSTR a partir do histórico de operação disponível no conjunto de dados DaISy.

1.1.2 Objetivos específicos

Para alcançar esse objetivo geral, foram definidos os seguintes objetivos específicos:

- caracterizar o conjunto de dados CSTR da base DaISy, identificando suas variáveis, intervalo de amostragem e comportamento temporal;
- estruturar janelas temporais com histórico de vazão de resfriamento, concentração e temperatura para treinamento supervisionado da rede;
- implementar um modelo LSTM multi-saída para estimar simultaneamente concentração e temperatura;
- avaliar o desempenho do modelo por meio de MAE, MSE, RMSE e R^2 , comparando-o com a persistência em testes de um passo, testes externos e avaliações recursivas;
- analisar os limites da abordagem considerando suavidade da série, acúmulo de erro em simulação recursiva e ausência de integração com uma planta real.

1.2 Justificativa

A modelagem de processos industriais é fundamental para análise, simulação, monitoramento e apoio à tomada de decisão. Em sistemas não lineares, como reatores químicos, essa tarefa se torna ainda mais relevante, pois pequenas alterações nas condições de operação podem afetar significativamente o comportamento do processo. A capacidade de estimar a evolução de variáveis como concentração e temperatura contribui para uma operação mais estável, permitindo identificar tendências de comportamento antes que elas resultem em condições indesejadas.

Do ponto de vista industrial, modelos preditivos podem auxiliar na redução de paradas não planejadas, no acompanhamento da qualidade do produto, no uso mais eficiente de energia e insumos e no suporte a estratégias de controle e supervisão. Em processos químicos, esses aspectos têm impacto direto na produtividade e na segurança operacional, pois permitem que engenheiros e operadores compreendam melhor a dinâmica do sistema e tomem decisões com maior embasamento. Assim, o desenvolvimento de modelos capazes de representar adequadamente processos dinâmicos é relevante para a modernização de plantas industriais e para a adoção de ferramentas associadas à transformação digital.

Do ponto de vista social, melhorias em processos industriais também podem refletir em benefícios indiretos para a sociedade. Operações mais bem monitoradas e estimadas tendem a reduzir desperdícios, consumo desnecessário de recursos, geração de produtos fora de especificação e riscos associados a condições operacionais inadequadas. Embora este trabalho utilize uma base pública de identificação de sistemas, e não medições coletadas diretamente de uma planta em operação, a investigação de modelos preditivos para processos dinâmicos contribui para uma linha de pesquisa relacionada a sistemas industriais mais seguros, eficientes e sustentáveis.

Embora modelos fenomenológicos sejam indispensáveis para compreender a natureza física do sistema, modelos orientados a dados podem complementar essa abordagem quando há disponibilidade de medições históricas representativas. Essa combinação é especialmente interessante em aplicações nas quais se busca estimar a evolução do processo sem depender exclusivamente de uma descrição analítica completa. Nesse contexto, redes LSTM são relevantes por sua capacidade de tratar séries temporais e dependências entre amostras, características comuns em processos industriais dinâmicos.

A escolha do CSTR se justifica por sua importância na literatura de controle e automação e por sua capacidade de representar desafios encontrados em processos industriais reais, como não linearidade, acoplamento entre variáveis e dependência do histórico de operação. Além disso, trabalhos recentes têm explorado o uso de inteligência artificial nesse tipo de sistema, inclusive em aplicações relacionadas à classificação de falhas com redes MLP (MORAIS, 2024). Em paralelo, redes LSTM têm sido aplicadas à identificação e previsão em sistemas dinâmicos e processos térmicos (LEITE, 2023; RODRIGUES, 2022; CORREIA, 2025), o que reforça a atualidade do tema e delimita a diferença em relação à proposta deste trabalho, voltada à regressão multi-saída com LSTM aplicada ao CSTR.

A contribuição deste trabalho está no desenvolvimento e na avaliação de um modelo LSTM multi-saída para estimar simultaneamente variáveis relevantes do CSTR, considerando diferentes formas de teste e comparação com um modelo de referência. Com isso, o estudo busca fornecer uma análise aplicada sobre o uso de redes neurais recorrentes como componente preditivo orientado a dados, discutindo tanto seu potencial quanto seus limites para aplicações associadas a gêmeos digitais. Assim, o trabalho se justifica pela relevância técnica do processo estudado, pela relação com demandas industriais por estimação e monitoramento e pela contribuição para

a aplicação de inteligência artificial em sistemas dinâmicos de engenharia.

2 FUNDAMENTAÇÃO TEÓRICA

Sistemas dinâmicos são aqueles cujo comportamento se desenvolve ao longo do tempo em resposta a entradas, condições iniciais, perturbações e características internas do próprio processo. Esse conceito é essencial na Engenharia de Controle e Automação, pois muitos processos industriais envolvem variáveis que não respondem instantaneamente às ações aplicadas. Temperatura, nível, vazão, pressão e concentração são exemplos de grandezas que variam ao longo do tempo e, por isso, necessitam de modelos capazes de representar sua evolução dinâmica (OGATA, 2010; SEBORG *et al.*, 2010).

A modelagem desses sistemas permite compreender a relação entre entradas e saídas, estimar respostas futuras e analisar características como estabilidade, atraso, comportamento transitório e regime permanente. Na engenharia de controle, essa modelagem é tradicionalmente realizada por meio de equações diferenciais, funções de transferência e representações em espaço de estados (OGATA, 2010).

Em processos reais, entretanto, a obtenção de um modelo matemático exato pode ser difícil. Sistemas industriais frequentemente apresentam não linearidades, ruídos de medição, atrasos, incertezas paramétricas e interações entre múltiplas variáveis. Por esse motivo, modelos orientados a dados podem complementar a modelagem fenomenológica, principalmente quando há disponibilidade de séries históricas representativas do processo.

2.1 Reator Contínuo de Tanque Agitado

O CSTR é um sistema no qual reagentes entram continuamente em um tanque, enquanto produtos são removidos de forma também contínua. A agitação interna permite considerar, em uma aproximação ideal, que as propriedades no interior do reator sejam uniformes.

O processo de referência é baseado no sistema apresentado por Morningred *et al.* (1992), no qual ocorre uma reação irreversível em um reator de volume constante resfriado por uma corrente de fluido refrigerante. Em uma formulação de controle e identificação, a vazão de resfriamento, representada por q_c , pode ser tratada como variável manipulada, enquanto a concentração do reagente A, representada por C_A , e a temperatura do reator, representada por T , são variáveis de saída medidas. A Tabela 1 resume essas grandezas do ponto de vista do processo físico.

A Figura 1 apresenta uma representação conceitual do processo físico. O diagrama destaca a vazão de resfriamento q_c , tratada como variável manipulada, e as variáveis de saída C_A e T , correspondentes à concentração do reagente A e à temperatura do reator.

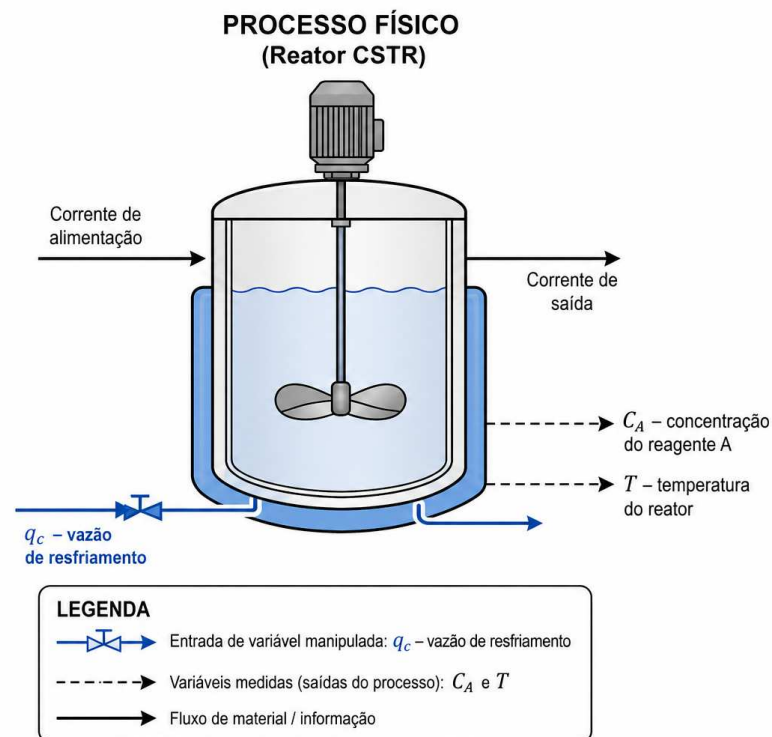


Figura 1 – Representação conceitual do CSTR e das variáveis utilizadas no modelo.

Fonte: elaboração própria.

Na Figura 1, observa-se um reator de tanque continuamente agitado com uma corrente de alimentação, uma corrente de saída e um sistema de agitação mecânica. A corrente de alimentação introduz reagentes no tanque, enquanto a corrente de saída remove continuamente o conteúdo processado. A presença do agitador busca manter o fluido interno homogêneo, permitindo assumir, em uma aproximação ideal, que concentração e temperatura sejam uniformes no interior do reator. Essa hipótese de mistura perfeita é uma das características fundamentais de um CSTR.

O reator é representado com uma camisa de resfriamento, por onde circula o fluido refrigerante. A vazão desse fluido, indicada por q_c , é a variável manipulada considerada no trabalho. Alterações em q_c modificam a taxa de remoção de calor do sistema e, consequentemente, influenciam a temperatura do reator. Como a temperatura afeta a cinética da reação, mudanças na remoção de calor também podem se refletir na evolução da concentração do reagente A. Dessa forma, q_c atua como uma entrada importante para explicar a dinâmica conjunta de C_A e T .

As setas tracejadas indicam as variáveis medidas utilizadas como saídas do processo: C_A , que representa a concentração do reagente A, e T , que representa a temperatura do reator. Em problemas de identificação orientada a dados, saídas medidas também podem ser reutilizadas como regressoras temporais, pois seus valores passados carregam informação sobre o estado recente do processo.

Essa representação física ajuda a justificar o uso de modelos temporais. Em um tanque

agitado, uma alteração na entrada não produz necessariamente uma mudança instantânea na saída medida. A resposta depende do volume do tanque, das vazões, da eficiência de mistura e das relações químicas e térmicas envolvidas. Assim, valores passados das entradas e das saídas carregam informação relevante sobre o estado atual e sobre a evolução futura do processo.

Tabela 1 – Principais variáveis do processo CSTR.

Símbolo	Descrição	Papel no processo
q_c	Vazão de resfriamento	Variável manipulada
C_A	Concentração do reagente A	Saída medida
T	Temperatura do reator	Saída medida

Em um CSTR exotérmico, a reação libera calor durante sua ocorrência. Assim, a temperatura do reator depende tanto da geração de calor provocada pela reação quanto da remoção de calor realizada pelo sistema de resfriamento. Como a taxa de reação é influenciada pela temperatura, há acoplamento entre o balanço de massa e o balanço de energia. De forma geral, a dinâmica pode ser associada às seguintes relações:

$$\frac{dC_A}{dt} = \frac{F}{V}(C_{A0} - C_A) - r_A, \quad (2.1)$$

$$\frac{dT}{dt} = \frac{F}{V}(T_0 - T) + \frac{-\Delta H}{\rho C_p} r_A - \frac{UA}{\rho C_p V}(T - T_c), \quad (2.2)$$

Em que F representa a vazão de alimentação, V o volume do reator, C_{A0} a concentração de entrada, T_0 a temperatura de entrada, r_A a taxa de reação, ΔH a entalpia de reação, ρ a densidade, C_p o calor específico, U o coeficiente global de transferência de calor, A a área de troca térmica e T_c a temperatura associada ao sistema de resfriamento. Essas equações são utilizadas apenas para contextualizar os fenômenos físicos envolvidos na dinâmica do CSTR. O modelo desenvolvido neste trabalho não resolve explicitamente os balanços de massa e energia; em vez disso, aprende relações temporais diretamente a partir das séries disponíveis no conjunto de dados.

A não linearidade do CSTR está fortemente associada à cinética da reação química. Em muitos casos, a taxa de reação pode ser representada por uma relação do tipo Arrhenius:

$$r_A = k_0 e^{-\frac{E}{RT}} C_A, \quad (2.3)$$

Em que k_0 é o fator pré-exponencial, E é a energia de ativação e R é a constante universal dos gases. Como a taxa depende exponencialmente da temperatura, pequenas variações em T podem provocar alterações relevantes na velocidade de consumo do reagente. Esse comportamento reforça a necessidade de modelos capazes de representar relações não lineares e dependências temporais.

2.2 Identificação de Sistemas e Modelagem Orientada a Dados

A identificação de sistemas consiste na construção de modelos matemáticos a partir de dados de entrada e saída. Diferentemente da modelagem puramente fenomenológica, baseada diretamente nas leis físicas que descrevem o processo, a identificação utiliza medições para estimar uma representação capaz de reproduzir o comportamento observado (LJUNG, 1999; AGUIRRE, 2007).

Segundo a abordagem clássica de identificação, o processo envolve etapas como escolha das variáveis, definição da estrutura do modelo, estimação dos parâmetros e validação com dados não utilizados no ajuste. Essa validação é fundamental, pois um modelo pode se ajustar bem ao conjunto de treinamento e, ainda assim, apresentar baixa capacidade de generalização para outros trechos da série (LJUNG, 1999; AGUIRRE, 2007).

Em uma abordagem orientada a dados, o objetivo não é substituir completamente a modelagem física do processo, mas construir uma representação computacional capaz de estimar saídas futuras a partir de medições históricas. Esse tipo de modelo é particularmente útil quando a formulação fenomenológica completa não está disponível para uso direto, quando há incertezas paramétricas relevantes ou quando se deseja uma aproximação empírica para tarefas de estimação e monitoramento.

2.3 Aprendizado de Máquina e Redes Neurais Artificiais

O aprendizado de máquina é uma área da inteligência artificial dedicada ao desenvolvimento de algoritmos capazes de identificar padrões em dados. Em problemas supervisionados, o modelo é treinado a partir de pares de entrada e saída, ajustando seus parâmetros internos para reduzir a diferença entre a resposta estimada e a resposta real (GOODFELLOW; BENGIO; COURVILLE, 2016).

As redes neurais artificiais são modelos computacionais compostos por unidades de processamento interligadas, chamadas neurônios artificiais. Cada neurônio realiza uma combinação ponderada das entradas e aplica uma função de ativação, permitindo que a rede represente relações não lineares entre variáveis (HAYKIN, 2009).

Uma forma básica de representar um neurônio artificial é o perceptron. Nesse modelo, as entradas $x_1, x_2, \dots, x_n$ são multiplicadas por pesos $w_1, w_2, \dots, w_n$, somadas a um viés b e então passadas por uma função de ativação ϕ . A saída do neurônio pode ser escrita como:

$$a = \phi \left(\sum_{j=1}^n w_j x_j + b \right). \quad (2.4)$$

Na Equação 2.4, x_j representa a j -ésima entrada do neurônio, w_j é o peso associado a essa entrada, b é o viés, ϕ é a função de ativação e a é a saída produzida pelo neurônio. O viés

permite deslocar a resposta do neurônio, enquanto os pesos determinam a contribuição de cada entrada. Durante o treinamento, pesos e vieses são ajustados para reduzir o erro do modelo. A Figura 2 apresenta essa estrutura de forma esquemática.

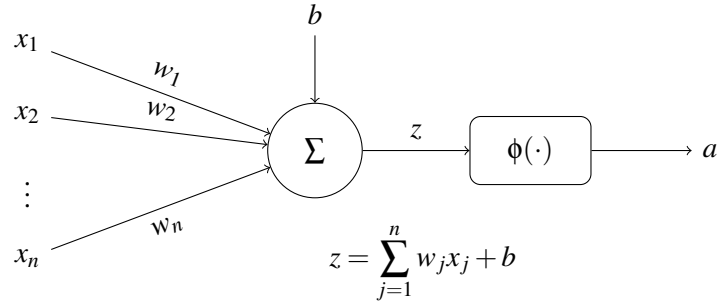


Figura 2 – Representação esquemática de um neurônio artificial.

Fonte: elaboração própria.

Embora o perceptron seja uma representação simples, ele é útil para compreender a lógica básica das redes neurais: combinar entradas ponderadas, aplicar uma transformação não linear e produzir uma saída. Redes mais complexas são formadas pela conexão de muitos neurônios organizados em camadas. A Figura 3 apresenta uma representação esquemática de uma rede neural em camadas.

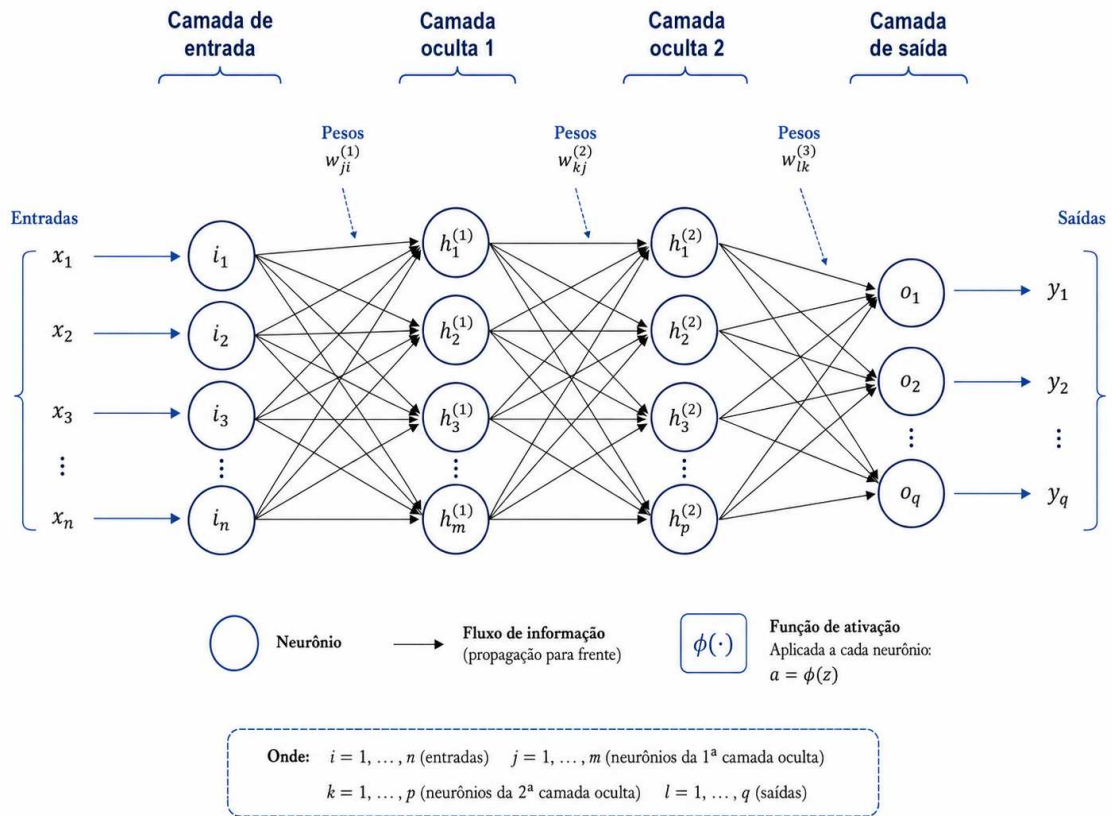


Figura 3 – Representação conceitual de uma rede neural artificial em camadas.

Fonte: elaboração própria.

Na Figura 3, a camada de entrada recebe as variáveis do problema, representadas por x_1, x_2, x_3 , até x_n . Esses valores são repassados aos neurônios das camadas ocultas por meio de conexões ponderadas. Os pesos, indicados genericamente por w , determinam a intensidade com que a saída de um neurônio influencia o neurônio seguinte. Durante o treinamento, esses pesos são ajustados para reduzir o erro entre a saída estimada e a saída real.

As camadas ocultas são responsáveis por construir representações intermediárias dos dados. Cada neurônio combina as entradas recebidas, aplica uma função de ativação e envia o resultado para a próxima camada. Essa composição sucessiva permite que a rede represente relações não lineares, pois a resposta final não depende apenas de uma combinação linear simples das entradas. Na camada de saída, os neurônios produzem os valores estimados pelo modelo. Em problemas de regressão multi-saída, como o tratado neste trabalho, a rede pode produzir simultaneamente mais de uma variável estimada, como concentração e temperatura.

Em contextos industriais, redes neurais podem ser empregadas na estimação de variáveis de processo, identificação de falhas, modelagem de sistemas não lineares e suporte a estratégias de controle (HAYKIN, 2009; GOODFELLOW; BENGIO; COURVILLE, 2016). Sua principal vantagem está na capacidade de aprender relações complexas diretamente a partir dos dados. Por outro lado, seu desempenho depende da qualidade dos dados, da seleção das variáveis de entrada,

da normalização, da arquitetura e do procedimento de treinamento. Além disso, o uso de redes neurais não elimina a necessidade de conhecer as variáveis medidas e as condições de operação do processo; quando a modelagem fenomenológica não é viável ou não é suficientemente prática para a tarefa, esses modelos podem oferecer uma alternativa empírica, com custo computacional e necessidade de validação próprios.

2.4 Redes de Memória de Curto e Longo Prazo

As redes neurais recorrentes foram desenvolvidas para processar dados sequenciais. Diferentemente das redes neurais convencionais, elas possuem conexões que permitem utilizar informações de instantes anteriores no cálculo da saída atual. Essa característica torna esse tipo de rede adequado para problemas nos quais a ordem temporal dos dados é relevante (GOODFELLOW; BENGIO; COURVILLE, 2016).

Redes recorrentes tradicionais podem apresentar dificuldade para aprender dependências temporais longas devido ao desaparecimento ou à explosão do gradiente. Para superar essa limitação, Hochreiter e Schmidhuber (1997) propuseram a arquitetura LSTM. A célula LSTM utiliza mecanismos internos chamados portas, que controlam quais informações devem ser preservadas, atualizadas ou descartadas ao longo da sequência.

A escolha de uma rede LSTM, em vez de um modelo puramente estático ou de uma referência simples, está relacionada à natureza de processos dinâmicos amostrados no tempo. No CSTR, as saídas futuras não dependem apenas do valor instantâneo das entradas, mas também da evolução anterior do processo. Além disso, concentração e temperatura apresentam acoplamento e comportamento não linear. Modelos mais simples podem ser úteis como referência, especialmente quando as variáveis mudam lentamente, mas tendem a ter maior dificuldade para representar dependências temporais e relações não lineares acumuladas ao longo de uma janela de operação.

De forma simplificada, uma célula LSTM pode ser descrita pelas Equações 2.5 a 2.10. Essas equações representam o cálculo interno realizado pela célula a cada instante da sequência de entrada. No contexto deste trabalho, elas ajudam a explicar como a rede processa uma janela temporal contendo valores passados de vazão de resfriamento, concentração e temperatura. O objetivo dessas operações é atualizar uma memória interna da rede e produzir uma representação temporal que será usada, ao final da janela, para estimar os valores futuros de C_A e T .

Os parâmetros das equações não são definidos manualmente. As matrizes de pesos e os vetores de vieses são ajustados durante o treinamento, por meio da minimização da função de perda. Assim, a rede aprende, a partir dos dados, quais informações da sequência devem ser preservadas, atualizadas ou descartadas.

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad (2.5)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad (2.6)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C), \quad (2.7)$$

$$C_t = f_t C_{t-1} + i_t \tilde{C}_t, \quad (2.8)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (2.9)$$

$$h_t = o_t \tanh(C_t), \quad (2.10)$$

Em que x_t representa a entrada da célula no instante t da sequência. Neste trabalho, esse instante corresponde a uma posição dentro da janela temporal usada como entrada da LSTM, e não necessariamente a uma nova execução independente do modelo. O termo h_t representa o estado oculto produzido no instante t , enquanto h_{t-1} representa o estado oculto do instante anterior da mesma sequência. Esse estado carrega informação resumida sobre as amostras já processadas pela célula.

O termo C_t representa o estado da célula, também chamado de memória interna da LSTM. De modo análogo, C_{t-1} representa a memória interna vinda do instante anterior. A porta de esquecimento f_t controla quanto da memória anterior C_{t-1} será preservado. A porta de entrada i_t controla quanto da nova informação candidata será incorporado à memória. A informação candidata é representada por $\tilde{C}_t$, calculada a partir da entrada atual e do estado oculto anterior. A porta de saída o_t controla quanto da memória atualizada será exposto no estado oculto h_t .

As matrizes W_f , W_i , W_C e W_o são matrizes de pesos associadas, respectivamente, à porta de esquecimento, à porta de entrada, ao cálculo do estado candidato e à porta de saída. Os vetores b_f , b_i , b_C e b_o são os vieses correspondentes a esses mesmos cálculos. Portanto, os subíndices indicam a qual parte da célula LSTM cada conjunto de parâmetros pertence. Esses pesos e vieses são os parâmetros aprendidos durante o treinamento da rede.

A notação $[h_{t-1}, x_t]$ indica a concatenação entre o estado oculto anterior e a entrada atual. Dessa forma, cada porta decide seu valor com base tanto na nova amostra da sequência quanto na informação acumulada até o instante anterior. A função σ representa a função sigmoide, que produz valores entre 0 e 1, permitindo interpretar as portas como mecanismos de controle parcial da passagem de informação. A função $\tanh$ limita os valores ao intervalo entre -1 e 1, sendo usada no cálculo da informação candidata e na formação do estado oculto. A Figura 4 apresenta o fluxo interno de informação em uma célula LSTM.

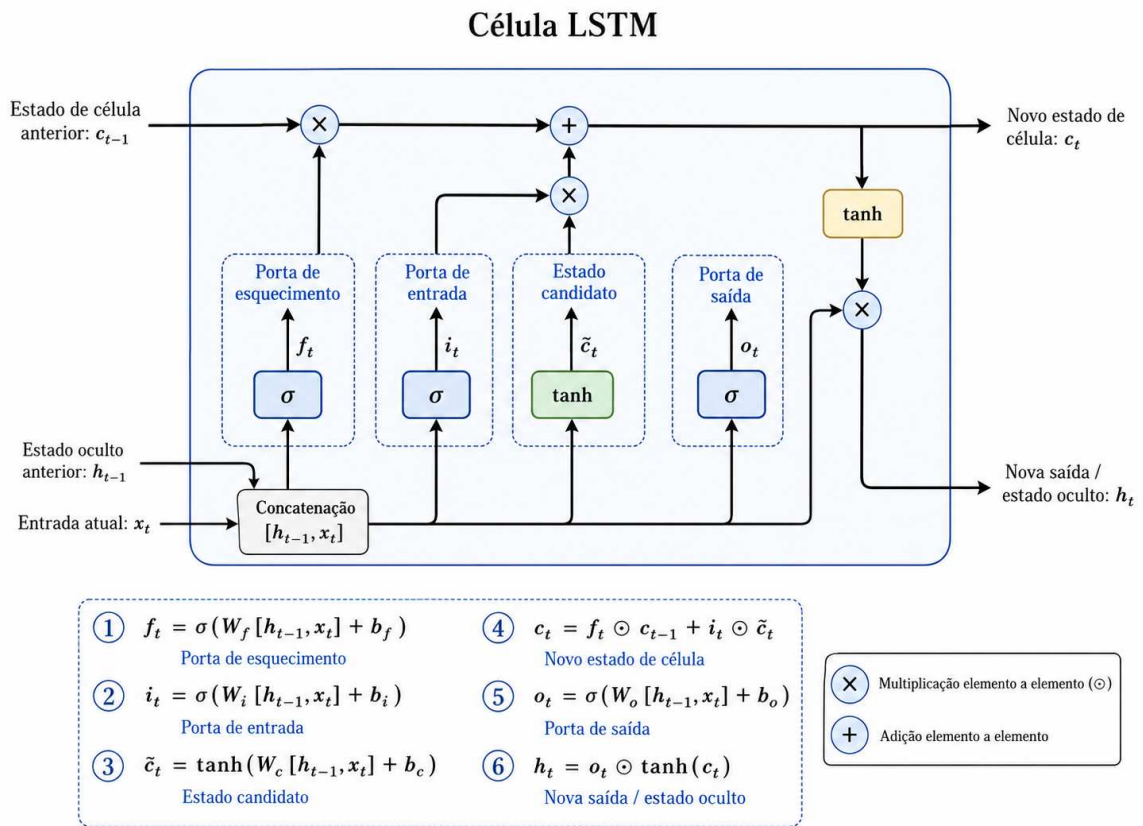


Figura 4 – Representação conceitual de uma célula LSTM.

Fonte: elaboração própria.

Na Figura 4, a célula recebe três informações principais: a entrada atual x_t , o estado oculto anterior h_{t-1} e o estado de célula anterior c_{t-1} . A entrada atual corresponde à amostra processada naquele instante da sequência, enquanto h_{t-1} resume a informação já propagada pela célula até o instante anterior. O estado c_{t-1} representa a memória interna acumulada pela LSTM.

A porta de esquecimento produz f_t e controla quanto do estado de célula anterior deve ser mantido. Quando os valores dessa porta se aproximam de 1, a informação anterior tende a ser preservada; quando se aproximam de 0, ela tende a ser descartada. A porta de entrada produz i_t e controla quanto da nova informação candidata será incorporado ao estado da célula. Essa nova informação candidata, representada por $\tilde{c}_t$, é calculada com a função $\tanh$ a partir da entrada atual e do estado oculto anterior.

O novo estado de célula c_t resulta da combinação entre a memória anterior filtrada pela porta de esquecimento e a informação candidata filtrada pela porta de entrada. Em seguida, a porta de saída produz o_t , que define quanto da memória atualizada será exposto como novo estado oculto h_t . Dessa forma, a LSTM consegue controlar separadamente o que deve ser esquecido, o que deve ser incorporado e o que deve ser disponibilizado como saída para os próximos instantes da sequência.

No caso do CSTR, a concentração e a temperatura futuras dependem do histórico da vazão de resfriamento, da própria concentração e da própria temperatura. Dessa forma, uma rede LSTM pode utilizar janelas temporais para aprender padrões dinâmicos que modelos estáticos teriam dificuldade em representar.

2.5 Gêmeos Digitais Preditivos

O conceito de gêmeo digital refere-se à criação de uma representação virtual de um sistema físico, capaz de acompanhar, simular, analisar ou estimar seu comportamento (GRIEVES, 2014; TAO *et al.*, 2019). Em aplicações industriais, essa abordagem pode contribuir para monitoramento, manutenção preditiva, diagnóstico de falhas, otimização operacional e suporte à tomada de decisão (FULLER *et al.*, 2020).

A construção de um gêmeo digital pode utilizar modelos físicos, modelos empíricos, modelos orientados a dados ou combinações híbridas. Em um exemplo funcional, uma planta instrumentada envia medições de sensores para uma representação computacional; o modelo virtual estima estados não medidos, simula cenários de operação ou antecipa desvios, e essas informações podem apoiar monitoramento, diagnóstico ou decisão operacional. Para que essa estrutura seja caracterizada como gêmeo digital em sentido mais completo, é necessário haver vínculo explícito entre sistema físico, dados atualizados e modelo computacional.

Neste trabalho, não se implementa um gêmeo digital completo, pois não há conexão em tempo real com uma planta física nem atuação automática sobre o processo. O modelo desenvolvido é tratado como um componente preditivo orientado a dados, cuja função principal é estimar o comportamento futuro de C_A e T com base no histórico das variáveis do processo. Assim, o conceito de gêmeo digital é usado como motivação de aplicação, e não como contribuição principal ou como sistema entregue.

2.6 Normalização e Organização Temporal de Séries

A normalização dos dados consiste em transformar os valores numéricos das variáveis utilizadas pelo modelo para uma escala comum. Em problemas com variáveis físicas de unidades distintas, como vazão, concentração e temperatura, esse procedimento evita que variáveis com maior escala numérica dominem o treinamento da rede neural e contribui para tornar o processo de otimização mais estável. Uma forma comum de normalização é a padronização pela média e pelo desvio padrão:

$$z = \frac{x - \mu}{\Phi}. \quad (2.11)$$

Em que x representa o valor original de uma variável do conjunto de dados, μ representa a média dessa variável, Φ representa seu desvio padrão e z representa o valor normalizado

correspondente. Em séries temporais, os parâmetros de normalização devem ser calculados apenas com dados disponíveis no conjunto de treinamento, evitando que informações futuras influenciem a preparação dos dados.

Além da normalização, problemas de estimação temporal exigem que a ordem das amostras seja preservada. Para isso, é comum organizar os dados em janelas temporais, nas quais um conjunto de amostras passadas é usado para estimar uma ou mais amostras futuras. Essa estrutura permite transformar uma série temporal em um problema de aprendizado supervisionado, mantendo a dependência entre o histórico recente do processo e a saída a ser estimada. A definição do tamanho da janela, do horizonte de estimação e da forma do alvo de treinamento faz parte da metodologia experimental adotada em cada aplicação.

2.7 Validação Temporal, Métricas e Persistência

Em problemas de séries temporais, a divisão dos dados deve preservar a ordem cronológica. A divisão aleatória pode causar vazamento de informação, permitindo que padrões futuros influenciem o treinamento. Por isso, neste trabalho são utilizados segmentos cronológicos para treinamento, validação, teste interno e teste externo.

Além da estimação de um passo à frente, também é importante avaliar o comportamento recursivo. Nesse tipo de teste, estimativas produzidas pelo modelo são reaproveitadas como parte da próxima janela temporal. Essa avaliação é mais rigorosa porque pequenos erros podem se acumular ao longo da simulação.

As métricas utilizadas incluem MSE, RMSE, MAE e o coeficiente de determinação (R^2):

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2. \quad (2.12)$$

Na Equação 2.12 e nas demais métricas apresentadas a seguir, N representa o número total de amostras avaliadas, y_i representa o valor real da variável de saída na i -ésima amostra de avaliação e $\hat{y}_i$ representa o valor estimado pelo modelo para essa mesma amostra. O índice i não indica uma nova variável do processo; ele é apenas um índice de contagem usado no cálculo das métricas. Em termos da notação temporal adotada nas equações de estimação, cada amostra i corresponde a um determinado instante de avaliação da série, associado ao par formado pelo valor real futuro e pela estimação produzida a partir da janela temporal correspondente. Quando as métricas são calculadas separadamente para concentração e temperatura, y_i e $\hat{y}_i$ se referem à variável de saída considerada em cada caso. No coeficiente de determinação, $\bar{y}$ representa a média dos valores reais avaliados.

$$RMSE = \sqrt{MSE}, \quad (2.13)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (2.14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad (2.15)$$

Como referência simples, utiliza-se o modelo de persistência, no qual o próximo valor estimado é igual ao último valor observado:

$$\hat{y}(k+1) = y(k). \quad (2.16)$$

Embora simples, a persistência pode apresentar bom desempenho em sistemas lentos ou com pequenas variações entre amostras consecutivas. Assim, a comparação com esse modelo ajuda a verificar se a LSTM realmente acrescenta capacidade preditiva ao problema analisado.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta trabalhos relacionados ao tema desta pesquisa, com foco em reatores contínuos de tanque agitado, identificação de sistemas, redes neurais aplicadas a processos dinâmicos, estimação de séries temporais e gêmeos digitais. Como a fundamentação teórica já apresentou os conceitos de CSTR, redes neurais, LSTM, validação temporal e métricas de regressão, esta seção concentra-se em discutir contribuições próximas ao problema investigado.

O CSTR é frequentemente utilizado como estudo de caso em controle de processos, identificação de sistemas e modelagem de sistemas não lineares. Sua importância decorre do acoplamento entre fenômenos de transporte, reação química, geração de calor e remoção de energia. Por apresentar uma dinâmica relativamente compacta, mas rica em não linearidades, o CSTR permite avaliar métodos de modelagem e controle em um ambiente representativo de processos industriais.

Entre os trabalhos clássicos relacionados ao processo utilizado neste estudo, destaca-se Morningred *et al.* (1992), que apresentou um controlador preditivo não linear adaptativo aplicado a um CSTR exotérmico. O sistema descrito pelos autores envolve uma reação irreversível em um reator de volume constante resfriado por uma corrente de fluido refrigerante. A manipulação da vazão de resfriamento para influenciar o comportamento interno do reator evidencia a forte relação entre entrada manipulada, temperatura e concentração no processo.

Além dos estudos clássicos de controle, pesquisas recentes também têm explorado o uso de inteligência artificial em CSTR. Morais (2024) desenvolveu um sistema de detecção e diagnóstico de falhas em um reator de tanque com agitação contínua utilizando uma rede neural do tipo perceptron multicamadas (MLP, do inglês *Multilayer Perceptron*). O trabalho é especialmente próximo do contexto desta pesquisa por combinar CSTR, inteligência artificial e Engenharia de Controle e Automação em um ambiente acadêmico semelhante. A diferença central é que Morais (2024) trata um problema de classificação de falhas, enquanto este trabalho trata um problema de regressão temporal multi-saída para estimação de concentração e temperatura.

Essa aproximação entre modelos computacionais e processos químicos também é relevante do ponto de vista operacional. Em sistemas industriais, a ocorrência de falhas em sensores, atuadores ou no próprio processo pode comprometer desempenho, confiabilidade e segurança. Nesse sentido, o uso de inteligência artificial em CSTR não se limita à estimação de variáveis, mas também se estende a diagnóstico, classificação de condições operacionais e suporte a estratégias de controle tolerante a falhas, como observado no trabalho de Morais (2024). A presença desse trabalho na revisão, portanto, não tem a função de comparar diretamente métricas de regressão, mas de mostrar uma aplicação correlata de redes neurais ao mesmo tipo de processo.

A identificação de sistemas fornece a base metodológica para construir modelos a partir de dados experimentais ou históricos. Ljung (1999) apresenta a identificação como um processo sistemático que envolve escolha da estrutura do modelo, estimação de parâmetros e validação com

dados não utilizados no ajuste. No Brasil, Aguirre (2007) também é uma referência importante para a área, discutindo técnicas lineares e não lineares aplicadas a sistemas reais.

Essas contribuições são importantes porque mostram que a qualidade de um modelo identificado não depende apenas do algoritmo utilizado. A escolha das variáveis, a excitação do sistema, o tratamento dos dados e a forma de validação também influenciam diretamente a capacidade de generalização do modelo. Em sistemas dinâmicos, essa preocupação é ainda mais evidente, pois a preservação da ordem temporal dos dados é essencial para avaliar se o modelo consegue representar a evolução futura do processo a partir de informações passadas.

No contexto de processos industriais não lineares, modelos orientados a dados podem complementar abordagens fenomenológicas. Quando a modelagem analítica completa não é viável para a finalidade de estimação, ou quando o comportamento da planta varia com o ponto de operação, redes neurais e outros métodos de aprendizado de máquina podem ser utilizados para aprender relações entrada-saída diretamente a partir das medições disponíveis (GOODFELLOW; BENGIO; COURVILLE, 2016; HAYKIN, 2009). Essa escolha não dispensa o conhecimento das variáveis do processo nem elimina o custo computacional do treinamento; ela desloca parte da dificuldade da identificação de parâmetros físicos para a seleção de dados, arquitetura e validação do modelo.

Entre as arquiteturas utilizadas para dados sequenciais, as redes neurais recorrentes se destacam por considerar a ordem temporal das amostras. Dentro desse grupo, as redes LSTM são frequentemente empregadas por sua capacidade de lidar com dependências temporais mais longas (HOCHREITER; SCHMIDHUBER, 1997). Essa característica é especialmente relevante em processos dinâmicos, nos quais o estado futuro depende não apenas das entradas atuais, mas também do histórico recente do sistema.

Leite (2023) aplicou uma rede neural recorrente LSTM à identificação de um sistema térmico composto por dois aquecedores e dois sensores. O estudo utilizou a capacidade das células LSTM de capturar dependências em dados sequenciais para representar o comportamento térmico, avaliando a metodologia em um sistema físico com entradas de aquecimento e respostas de temperatura. Essa contribuição é relevante porque demonstra a aplicação de LSTM em um processo dinâmico real no qual o comportamento futuro depende do histórico das variáveis, aproximando-se da lógica empregada neste trabalho para o CSTR.

Outros trabalhos também indicam a aplicabilidade de redes neurais em processos industriais. Rodrigues (2022) aplicou redes LSTM à previsão de padrões em um forno de reaquecimento de tarugos, processo térmico no qual a temperatura precisa ser acompanhada com precisão para reduzir desvios operacionais. Correia (2025) investigou redes recorrentes profundas para modelagem e identificação de sistemas dinâmicos, considerando aplicações em pedais de distorção e robôs móveis e avaliando o ajuste por métricas de erro e coeficiente de determinação. Em conjunto, esses estudos reforçam que redes LSTM são adequadas para problemas em que o histórico recente das variáveis contém informação relevante para a estimação, ainda que os

domínios de aplicação sejam distintos do CSTR.

A estimação de séries temporais também aparece em diferentes áreas de aplicação, como processos industriais, sistemas térmicos, clima, energia e manutenção. Além de redes neurais, modelos estatísticos clássicos continuam sendo utilizados como referência em muitos problemas, especialmente quando as séries apresentam estrutura linear ou sazonalidade bem definida. Essa comparação é importante porque evita assumir, de forma automática, que uma arquitetura mais complexa será sempre superior.

Nesse sentido, Nunes (2025) comparou modelos autorregressivos integrados de médias móveis sazonais (SARIMA, do inglês *Seasonal Autoregressive Integrated Moving Average*), modelos de vetores autorregressivos (VAR, do inglês *Vector Autoregressive*) e redes LSTM na estimação de variáveis climáticas, como precipitação, velocidade do vento, temperatura e umidade. Embora o domínio de aplicação seja diferente do CSTR, o trabalho contribui para esta revisão por discutir a comparação entre modelos estatísticos e redes LSTM em séries temporais, utilizando métricas como raiz do erro quadrático médio (RMSE, do inglês *Root Mean Squared Error*), erro quadrático médio (MSE, do inglês *Mean Squared Error*) e erro absoluto médio (MAE, do inglês *Mean Absolute Error*). Os resultados apresentados pela autora indicam que o desempenho depende da natureza da série analisada e da variável prevista, reforçando a importância de avaliar cada técnica no contexto específico do problema e de apresentar métricas separadas por saída.

Em séries temporais industriais, modelos simples, como persistência, podem apresentar desempenho competitivo quando as variáveis variam lentamente. Por isso, a comparação com uma referência simples é uma etapa importante para interpretar corretamente o ganho obtido por modelos neurais. Essa preocupação dialoga diretamente com trabalhos de identificação e estimação, nos quais a avaliação deve considerar não apenas o erro médio, mas também a forma como o modelo se comporta em trechos não utilizados no treinamento.

O conceito de gêmeo digital também se relaciona com essa discussão, pois envolve a criação de uma representação virtual associada a um sistema físico. Grieves (2014) apresenta o gêmeo digital como uma representação virtual do sistema, enquanto Tao *et al.* (2019) discutem sua aplicação na manufatura inteligente, destacando a integração entre dados, modelos digitais e processos físicos. De forma complementar, Fuller *et al.* (2020) apresenta uma revisão sobre tecnologias habilitadoras, desafios e questões em aberto relacionadas aos gêmeos digitais, incluindo integração de dados, escolha adequada de modelos computacionais e capacidade de monitorar ou estimar o comportamento do sistema físico.

Em aplicações orientadas a dados, um gêmeo digital pode utilizar modelos treinados com séries históricas do processo como parte de sua estrutura computacional. Nesse caso, sua qualidade depende da representatividade dos dados, da arquitetura do modelo, do procedimento de validação e da capacidade de generalização para trechos não utilizados durante o treinamento. Essa perspectiva é coerente com a proposta deste trabalho apenas no nível do componente preditivo: utiliza-se uma rede LSTM para estimar a evolução da concentração e da temperatura

em um CSTR, sem implementar uma estrutura completa de gêmeo digital conectada a uma planta real.

De modo geral, a literatura analisada mostra que o CSTR é um processo relevante para estudos de controle, identificação e inteligência artificial aplicada a sistemas dinâmicos. O trabalho de Morningred *et al.* (1992) fornece a base clássica do processo utilizado, enquanto Moraes (2024) aproxima o tema de aplicações recentes de inteligência artificial em CSTR, especialmente no contexto de detecção de falhas.

Na área de identificação de sistemas, Ljung (1999) e Aguirre (2007) sustentam a importância de modelar sistemas dinâmicos a partir de dados, considerando seleção de variáveis, estrutura do modelo e validação. Já os trabalhos de Leite (2023), Rodrigues (2022) e Correia (2025) demonstram o uso de redes neurais, em especial LSTM, na identificação e estimação de sistemas dinâmicos e processos industriais.

Por fim, estudos sobre estimação de séries temporais e gêmeos digitais reforçam a necessidade de avaliar modelos com métricas quantitativas, comparação com referências simples e validação em dados não utilizados no treinamento. Esses elementos fundamentam o desenvolvimento de um modelo preditivo multi-saída para concentração e temperatura de um CSTR.

4 METODOLOGIA

A pesquisa possui caráter aplicado e abordagem quantitativa. É aplicada porque busca desenvolver um modelo computacional capaz de representar o comportamento dinâmico de um processo industrial. É quantitativa porque a análise é realizada por meio de dados numéricos, métricas de erro e gráficos comparativos entre valores reais e valores estimados.

O procedimento metodológico foi dividido nas seguintes etapas: seleção do conjunto de dados, definição das variáveis de entrada e saída, normalização dos dados, construção de janelas temporais, divisão cronológica da série, treinamento da rede LSTM, comparação com modelos simples e avaliação em diferentes condições de teste. Essa organização foi adotada para separar claramente as decisões relacionadas aos dados, à construção do modelo e à validação dos resultados, reduzindo o risco de misturar informações de treinamento e teste. A configuração principal do experimento é apresentada no Apêndice A.1.

4.1 Conjunto de Dados e Variáveis Utilizadas

O conjunto de dados utilizado corresponde ao conjunto CSTR da base DaISy, *Database for the Identification of Systems* (MOOR, 2008). A DaISy reúne bases públicas para identificação de sistemas dinâmicos, e o conjunto CSTR pertence à seção de sistemas da indústria de processos. A base foi atualizada pela última vez em 2008 e consultada para este trabalho em 2026. A escolha desse conjunto se justifica por se tratar de uma série pública, documentada e voltada a problemas de identificação, o que permite avaliar o modelo em um processo compatível com o objetivo do trabalho. O processo está associado ao sistema estudado por Morningred *et al.* (1992), no qual ocorre uma reação irreversível do tipo $A \rightarrow B$ em um reator de volume constante resfriado por uma corrente de fluido refrigerante.

A série possui 7500 amostras, com intervalo de amostragem de 0,1 minuto, equivalente a 6 segundos, totalizando aproximadamente 750 minutos de operação registrada. Esse intervalo não foi uma escolha deste trabalho, mas uma característica do conjunto de dados. Como consequência, a estimação de um passo à frente corresponde a prever apenas 6 segundos adiante, o que tende a favorecer modelos simples em trechos de variação lenta. Por isso, além do teste de um passo, foram incluídas avaliações recursivas e de múltiplos passos. Foram consideradas três grandezas do processo: vazão de resfriamento, concentração do reagente e temperatura do reator. No restante do trabalho, a vazão de resfriamento é representada por q_c , a concentração por C_A e a temperatura por T .

As variáveis usadas como entrada do modelo foram q_c , C_A e T . As saídas estimadas foram C_A e T . Essa aparente repetição ocorre porque concentração e temperatura são utilizadas em instantes diferentes: seus valores passados compõem a janela de entrada, enquanto seus valores futuros são os alvos de estimação. Essa escolha considera que a vazão de resfriamento influencia diretamente a remoção de calor, enquanto concentração e temperatura apresentam

comportamento acoplado na dinâmica do CSTR. Portanto, usar apenas a variável manipulada não seria suficiente para representar o estado recente do processo, e usar também as saídas anteriores permite que o modelo aprenda a evolução temporal das grandezas que serão estimadas. A Tabela 2 resume o papel de cada variável no modelo.

Tabela 2 – Variáveis utilizadas no modelo preditivo.

Variável	Notação	Papel no modelo
Vazão de resfriamento	q_c	Entrada histórica
Concentração	C_A	Entrada histórica e saída estimada
Temperatura	T	Entrada histórica e saída estimada

4.2 Recortes Temporais e Divisão dos Dados

O experimento foi estruturado com dois recortes temporais principais. O primeiro recorte contém as 5000 primeiras amostras da série e foi utilizado para treinamento, validação e teste interno. O segundo recorte contém as amostras posteriores ao índice 5000 e foi reservado para teste externo. Essa separação foi adotada para criar duas etapas de avaliação: uma avaliação interna, ainda dentro da região usada para desenvolvimento do modelo, e uma avaliação externa, em uma região temporal posterior e completamente separada. Em séries temporais, essa escolha é importante porque amostras vizinhas tendem a ser correlacionadas; portanto, avaliar apenas pontos misturados ao longo da série poderia produzir uma impressão excessivamente otimista do desempenho.

No primeiro recorte, as janelas foram construídas por deslizamento de uma amostra. Essa estratégia aumenta a quantidade de pares supervisionados, mas também cria janelas fortemente sobrepostas. Assim, a última janela de treinamento e a primeira janela de validação, por exemplo, podem compartilhar grande parte do histórico. Por esse motivo, o teste interno deve ser interpretado como uma avaliação inicial do ajuste no recorte de desenvolvimento, enquanto o teste externo, separado no índice 5000 antes da preparação das janelas externas, é a avaliação mais relevante para discutir generalização temporal.

Após a construção das janelas temporais no primeiro recorte, os dados foram divididos de forma cronológica em 70% para treinamento, 15% para validação e 15% para teste interno. A proporção de 70% para treinamento foi escolhida para fornecer uma quantidade maior de amostras ao ajuste dos pesos da rede, o que é necessário em modelos neurais com número elevado de parâmetros. As parcelas de 15% para validação e 15% para teste interno foram mantidas com tamanhos iguais para separar duas funções distintas: a validação acompanha o treinamento e influencia decisões como parada antecipada e redução da taxa de aprendizado, enquanto o teste interno é usado apenas depois do treinamento para uma avaliação inicial em dados não vistos pelo otimizador.

Como se trata de uma série temporal, a divisão de validação e teste foi mantida sem

embaralhamento. Essa decisão preserva a ordem dos acontecimentos e evita que amostras futuras sejam misturadas com amostras anteriores na separação dos conjuntos. O embaralhamento foi aplicado apenas aos lotes do conjunto de treinamento durante o processo de otimização, já depois da separação cronológica dos conjuntos. Nesse caso, o objetivo é melhorar o processo numérico de treinamento, sem alterar quais amostras pertencem a treino, validação ou teste.

4.3 Normalização dos Dados

A normalização foi realizada por média e desvio padrão, conforme a Equação (2.11). Essa transformação foi adotada porque as variáveis possuem escalas físicas diferentes, especialmente temperatura, concentração e vazão de resfriamento. Sem normalização, variáveis de maior magnitude poderiam dominar o cálculo da perda e dificultar o ajuste dos pesos da rede. Os parâmetros de normalização foram calculados somente com base na parcela de treinamento do primeiro recorte. Essa escolha evita que informações da validação, do teste interno ou do teste externo influenciem a escala usada no treinamento, preservando a independência das avaliações.

Os parâmetros obtidos para as variáveis de entrada são apresentados na Tabela 3. Para as saídas C_A e T , foram usados os mesmos valores correspondentes a essas variáveis. Depois da estimação, os resultados foram convertidos novamente para a escala original. Essa conversão é necessária porque as métricas em escala normalizada são úteis para treinamento, mas a interpretação dos resultados do processo deve ocorrer nas unidades físicas das variáveis. O trecho de implementação da normalização é apresentado no Apêndice A.2.

Tabela 3 – Parâmetros de normalização utilizados no experimento.

Variável	Média	Desvio padrão
q_c	100,214249	4,066911
C_A	0,090216	0,014177
T	440,987701	3,232321

4.4 Construção das Janelas Temporais

Cada amostra de entrada fornecida ao modelo foi formada por uma janela temporal com 50 pontos consecutivos das três variáveis de entrada. Como o intervalo de amostragem é de 6 segundos, cada janela representa 300 segundos, ou seja, 5 minutos de histórico do processo. O uso de uma janela temporal é necessário porque o CSTR possui dinâmica própria: o valor futuro de concentração e temperatura não depende apenas do instante atual, mas também da evolução recente do processo. A janela de 50 amostras foi adotada por cobrir um intervalo maior que os horizontes avaliados posteriormente, incluindo o horizonte de 50 passos, e por permitir que a LSTM observe a resposta recente de C_A e T às mudanças em q_c . Portanto, a escolha não foi baseada apenas em custo computacional; ela também buscou manter uma memória temporal compatível com a escala de variação observada no conjunto de dados.

Em cada instante k , o vetor de entrada do processo foi definido por:

$$\mathbf{x}(k) = \begin{bmatrix} q_c(k) \\ C_A(k) \\ T(k) \end{bmatrix}. \quad (4.1)$$

A saída de interesse foi definida por:

$$\mathbf{y}(k) = \begin{bmatrix} C_A(k) \\ T(k) \end{bmatrix}. \quad (4.2)$$

O horizonte de estimação adotado foi de um passo à frente. Essa escolha é coerente com a proposta de modelagem preditiva baseada em dados, pois permite avaliar se o modelo consegue acompanhar a dinâmica local do processo a partir do histórico recente. Além disso, a estimação de um passo à frente reduz a propagação de erro durante o treinamento supervisionado e serve como base para avaliações recursivas posteriores. Assim, a partir do histórico recente de q_c , C_A e T , o modelo estima os valores de C_A e T no instante seguinte. O valor real do instante seguinte está disponível durante a preparação dos dados porque a série histórica completa é conhecida. Esse valor é usado como rótulo supervisionado e como referência para cálculo do erro naquela janela; ele não é fornecido como entrada da mesma estimação. Como as janelas são deslizantes, um ponto usado como rótulo em uma janela pode aparecer como entrada histórica em uma janela posterior, respeitando a ordem temporal da série.

Com uma janela de $n = 50$ amostras e horizonte de um passo, a estimação pode ser representada por:

$$\hat{\mathbf{y}}(k+1) = f(\mathbf{x}(k-n+1), \dots, \mathbf{x}(k)). \quad (4.3)$$

O alvo de treinamento foi definido no modo de variação, denominado `delta` na implementação. Essa escolha foi adotada porque, em sistemas amostrados com intervalo curto, os valores consecutivos de saída tendem a ser próximos. Se o modelo aprendesse apenas o valor absoluto futuro, poderia se aproximar de uma solução parecida com a persistência. Ao aprender a variação, a rede é direcionada a representar a mudança local do processo em relação ao último valor conhecido. Nesse caso, a rede não aprende diretamente o valor absoluto futuro, mas a diferença entre o valor futuro e o último valor conhecido da janela. Para uma saída genérica y , o alvo de treinamento pode ser escrito como:

$$\Delta y(k+1) = y(k+1) - y(k). \quad (4.4)$$

Nessa equação, $y(k)$ representa o último valor conhecido da saída na janela de entrada, $y(k+1)$ representa o valor real no instante futuro usado como rótulo supervisionado, e $\Delta y(k+1)$

representa a variação real entre esses dois instantes. Durante o treinamento, a rede ajusta seus parâmetros para estimar essa variação. A saída produzida pelo modelo é indicada por $\Delta\hat{y}(k+1)$, que representa a variação estimada pela rede para o instante $k+1$.

Após a estimação, a variação estimada $\Delta\hat{y}(k+1)$ é somada ao último valor conhecido da janela para reconstruir a saída estimada:

$$\hat{y}(k+1) = y(k) + \Delta\hat{y}(k+1). \quad (4.5)$$

O procedimento de construção das janelas e do alvo supervisionado é apresentado no Apêndice A.3.

4.5 Arquitetura da Rede LSTM

A rede LSTM foi adotada porque o problema envolve uma série temporal com dependência dinâmica entre variáveis. Em um CSTR, concentração e temperatura não dependem apenas dos valores instantâneos das entradas, mas também da trajetória recente do processo. Além disso, a relação entre vazão de resfriamento, concentração e temperatura é acoplada e não linear. Por esse motivo, a LSTM foi escolhida como modelo principal, pois é capaz de processar sequências e aprender relações temporais a partir de uma janela de histórico. A opção por esse tipo de rede também se justifica pela necessidade de compará-la com modelos mais simples: se a LSTM apresentar desempenho superior, isso indica que a memória temporal e a representação não linear acrescentam informação relevante ao problema.

A entrada da rede possui formato $(50, 3)$, correspondente a 50 instantes de tempo e três variáveis por instante. A saída possui dois neurônios com ativação linear, correspondentes às variações estimadas de C_A e T . A formulação multi-saída foi adotada porque concentração e temperatura pertencem ao mesmo processo e evoluem de forma acoplada; portanto, estimar as duas variáveis no mesmo modelo permite que a rede compartilhe a representação temporal aprendida. A arquitetura utilizada é resumida na Tabela 4, e o trecho correspondente da implementação é apresentado no Apêndice A.4.

Tabela 4 – Arquitetura da rede LSTM utilizada.

Camada	Configuração
Entrada	Janela temporal com formato $(50, 3)$
LSTM	128 unidades
Dropout	Taxa de 0,02
Densa intermediária	64 neurônios, ativação ReLU
Saída	2 neurônios, ativação linear

A camada LSTM resume a informação temporal da janela em uma representação interna. O uso de 128 unidades foi adotado para fornecer capacidade suficiente de representação sem

tornar a rede excessivamente profunda. O dropout foi utilizado com taxa baixa, buscando reduzir sobreajuste sem retirar capacidade de aprendizado, o que é importante porque taxas muito altas poderiam prejudicar a aprendizagem das pequenas variações entre amostras consecutivas. A camada densa intermediária permite representar relações não lineares antes da camada final de regressão. A ativação linear na saída foi utilizada porque o problema é de regressão, e não de classificação.

4.6 Treinamento

O modelo foi treinado com o otimizador Adam, taxa de aprendizado igual a 5×10^{-4} e função de perda MSE. O Adam foi utilizado por combinar adaptação individual da taxa de aprendizado dos parâmetros com boa estabilidade em problemas de redes neurais. A taxa de aprendizado foi mantida em valor moderado para reduzir oscilações no ajuste, já que o alvo em modo `delta` envolve variações pequenas entre amostras consecutivas. A função de perda MSE foi adotada por penalizar erros maiores com mais intensidade, característica desejável em estimação de variáveis de processo, nas quais desvios elevados são mais críticos. Durante o treinamento, também foram acompanhadas as métricas MAE e RMSE. A MAE facilita a interpretação do erro médio absoluto, enquanto a RMSE mantém relação direta com a função quadrática de perda.

O tamanho de lote utilizado foi 16 e o número máximo de épocas foi 500. O lote pequeno permite atualizações frequentes dos pesos e reduz a necessidade de memória, enquanto o limite elevado de épocas funciona como teto de treinamento. Na prática, o treinamento não depende apenas desse limite, pois os mecanismos de parada antecipada interrompem o processo quando a validação deixa de melhorar.

Foram utilizados três mecanismos de controle do treinamento. O primeiro foi a parada antecipada, com paciência de 60 épocas e restauração dos pesos associados ao menor erro de validação. Essa estratégia evita continuar o treinamento quando não há melhora consistente e reduz o risco de sobreajuste. O segundo foi a redução automática da taxa de aprendizado quando o erro de validação deixava de melhorar por 15 épocas. A taxa foi reduzida, e não aumentada, porque a ausência de melhora indica que passos menores podem refinar os pesos próximos de uma região de mínimo, enquanto passos maiores poderiam aumentar oscilações na função de perda. O terceiro foi o salvamento do melhor modelo observado ao longo do treinamento, garantindo que a avaliação posterior utilize a melhor configuração encontrada durante o processo. Os detalhes desses procedimentos são mostrados no Apêndice A.4.

A semente aleatória foi fixada em 42 para tornar a execução mais reprodutível. Isso não significa ausência de aleatoriedade no treinamento, mas sim controle da inicialização e de operações pseudoaleatórias usadas pelo ambiente. Outra semente poderia levar a pequenas diferenças nos pesos finais e, consequentemente, nas métricas. Por isso, os resultados devem ser

interpretados como referentes à configuração experimental especificada.

4.7 Modelo Simples de Comparação

Para justificar o uso da LSTM, foi considerado um modelo simples como referência: o modelo de persistência, que assume que o valor futuro será igual ao último valor observado na janela. Esse modelo é importante em séries temporais com variações suaves, pois fornece uma referência mínima difícil de superar quando a amostragem é curta. Assim, a persistência permite verificar se a rede realmente aprendeu a dinâmica do processo ou se apenas reproduziu o comportamento quase constante entre amostras consecutivas. O trecho essencial desse baseline é apresentado no Apêndice A.5.

4.8 Avaliação Interna

A avaliação interna foi realizada no segmento final das janelas construídas a partir das 5000 primeiras amostras. Esse teste foi mantido porque permite verificar o desempenho do modelo em uma parcela do mesmo recorte temporal usado para desenvolvimento, mas que não participou do ajuste dos pesos. Nesse teste, cada janela de entrada contém apenas valores reais da própria série, e o modelo estima o próximo ponto de concentração e temperatura.

As estimativas normalizadas foram reconstruídas conforme o modo de variação, convertidas para a escala original e comparadas com os valores reais. As mesmas janelas foram usadas para avaliar o modelo de persistência, garantindo que a comparação fosse realizada sob as mesmas condições. As métricas calculadas foram MAE, MSE, RMSE e coeficiente de determinação R^2 . O uso conjunto dessas métricas permite observar o erro absoluto, a penalização quadrática de desvios maiores, o erro na unidade da variável e a proporção da variabilidade explicada pelo modelo. A matriz de confusão não foi utilizada porque o problema tratado é de regressão, com estimação de valores contínuos de concentração e temperatura, e não de classificação em categorias discretas.

4.9 Avaliação Externa de Um Passo à Frente

O teste externo de um passo à frente foi realizado no trecho da série iniciado após as 5000 primeiras amostras. Esse trecho não participou do treinamento, da validação nem do teste interno. A separação externa foi incluída para avaliar a capacidade de generalização temporal do modelo em uma região posterior da série, reduzindo a dependência da avaliação em relação ao recorte usado no desenvolvimento e evitando a sobreposição direta com as janelas do treinamento. Em cada estimação, a janela de entrada foi formada por valores reais do trecho externo, e o modelo estimou o ponto seguinte.

Essa avaliação de um passo à frente foi mantida porque isola a capacidade de estimação local do modelo: como a entrada usa valores reais da janela, o erro medido corresponde principalmente ao erro de estimação do próximo ponto, sem acúmulo de erros anteriores. A comparação foi feita entre valores reais, estimativas da LSTM e modelo de persistência.

4.10 Experimento Complementar com Horizonte de 5 Passos

Além do experimento principal com horizonte de um passo à frente, foi realizado um experimento complementar com horizonte de 5 passos. Essa avaliação não substitui o experimento principal, mas foi incluída para observar o comportamento dos modelos em uma tarefa mais exigente. Como o tempo de amostragem do conjunto de dados é $T_s = 6$ s, a janela temporal de 80 amostras representa 480 s de histórico do processo. Com `prediction_horizon = 5`, os modelos estimam concentração e temperatura 30 s à frente.

O experimento complementar manteve o conjunto de dados CSTR, as entradas q_c , C_A e T , as saídas C_A e T e o modo de alvo `delta`. A rede LSTM foi configurada com 128 unidades, camada densa com 32 neurônios, taxa de dropout de 0,20, tamanho de lote 8, taxa de aprendizado 5×10^{-5} , parada antecipada com paciência de 30 épocas e redução da taxa de aprendizado com paciência de 15 épocas e fator 0,5. A semente aleatória utilizada foi 42.

Nessa análise, o baseline de persistência continuou sendo definido como a estimativa do valor futuro a partir do último valor conhecido da janela. Assim, o experimento complementar compara a LSTM com a mesma referência simples utilizada nas demais avaliações do trabalho. Como esse experimento modifica também hiperparâmetros e tamanho de janela, seus resultados devem ser interpretados como uma análise complementar de configuração alternativa, e não como prova isolada de que apenas o aumento do horizonte para 5 passos causou a diferença observada.

4.11 Avaliação Recursiva e Multi-step

Além da avaliação de um passo à frente, foi realizada uma avaliação recursiva. Esse teste foi incluído porque representa uma condição mais exigente: em vez de usar sempre valores reais das saídas na janela, o modelo passa a usar suas próprias estimativas em passos seguintes. Nesse procedimento, as estimativas geradas para C_A e T são reinseridas nas janelas seguintes. Assim, o modelo passa a depender progressivamente de suas próprias estimativas para as variáveis de saída.

Também foi realizada avaliação recursiva de múltiplos passos para horizontes fixos de 10, 20 e 50 passos. Esses horizontes foram escolhidos para observar o comportamento do erro em curto e médio prazo dentro da escala temporal do conjunto de dados. Como cada amostra representa 6 segundos, os horizontes correspondem a aproximadamente 1 minuto, 2 minutos e 5 minutos de estimação. Para cada horizonte, o modelo executa estimativas sucessivas até

alcançar o instante desejado. Esse teste é mais exigente que a estimação de um passo à frente, pois evidencia o acúmulo de erro ao longo do horizonte de estimação.

5 RESULTADOS

Este capítulo apresenta os resultados obtidos com o modelo LSTM multi-saída aplicado ao conjunto de dados do CSTR. Os valores numéricos foram extraídos dos arquivos JSON e CSV gerados pela implementação Python, incluindo as métricas do teste interno, do teste externo e da avaliação de múltiplos passos. Os gráficos são usados como apoio para interpretar o comportamento temporal das estimativas, especialmente nos trechos em que a sobreposição das curvas dificulta a distinção visual entre valor real, LSTM e persistência.

5.1 Representatividade do Treinamento e Relação entre Variáveis

Antes da análise direta das previsões, foi avaliada a relação entre concentração e temperatura no conjunto de dados. Essa verificação é relevante porque o modelo foi treinado de forma cronológica, usando apenas a parte inicial da série para ajuste dos pesos. Assim, além de preservar a ordem temporal, é necessário verificar se o trecho de treinamento cobre de forma adequada a faixa operacional observada no restante do conjunto.

A Figura 5 apresenta a relação entre temperatura e concentração em todo o conjunto de dados, juntamente com um ajuste linear simples e os resíduos desse ajuste. Observa-se um comportamento aproximadamente linear no núcleo da faixa operacional, com $R^2 = 0,985$ para o ajuste linear simples. Entretanto, os resíduos exibem um padrão sistemático em forma de U, mais pronunciado nas extremidades da faixa de concentração observada, especialmente para C_A próximo de 0,06 mol/L e de 0,15 mol/L. Esse padrão é consistente com a natureza exponencial da cinética de Arrhenius, apresentada na Equação 2.3, cuja curvatura se torna mais evidente longe do ponto central de operação.

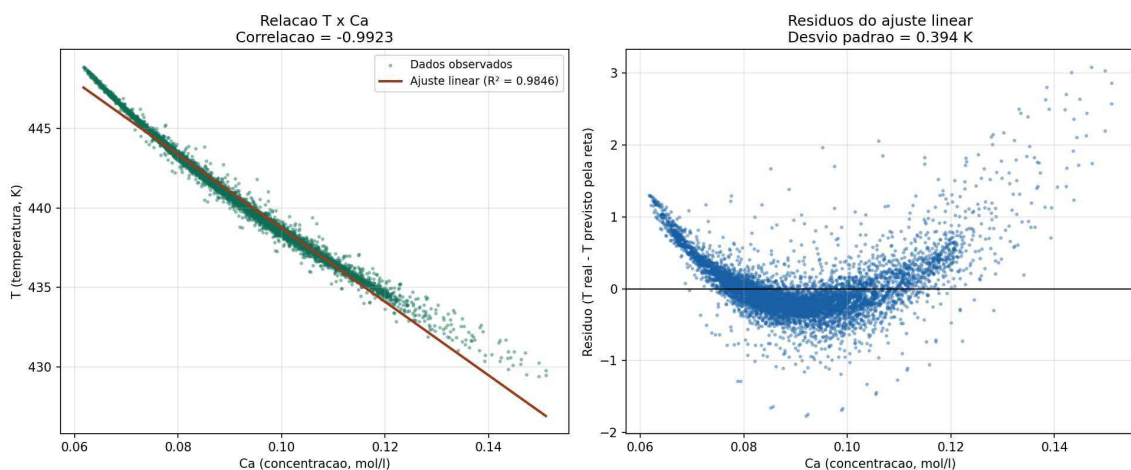


Figura 5 – Relação entre temperatura e concentração no conjunto de dados e resíduos do ajuste linear.

Essa análise sugere que a não linearidade do processo, embora discreta no núcleo dos

dados, concentra-se principalmente nas regiões de fronteira da faixa operacional coberta pelo conjunto. Essas regiões são importantes porque podem exigir maior capacidade de generalização do modelo, especialmente quando aparecem com menor frequência ou quando não estão igualmente distribuídas entre treinamento, validação e teste.

Para verificar a representatividade do recorte usado no treinamento, a Figura 6 destaca as 3500 amostras do trecho inicial usadas como base do treinamento e do ajuste dos normalizadores, em comparação com as demais amostras, pertencentes à validação, ao teste interno e ao teste externo.

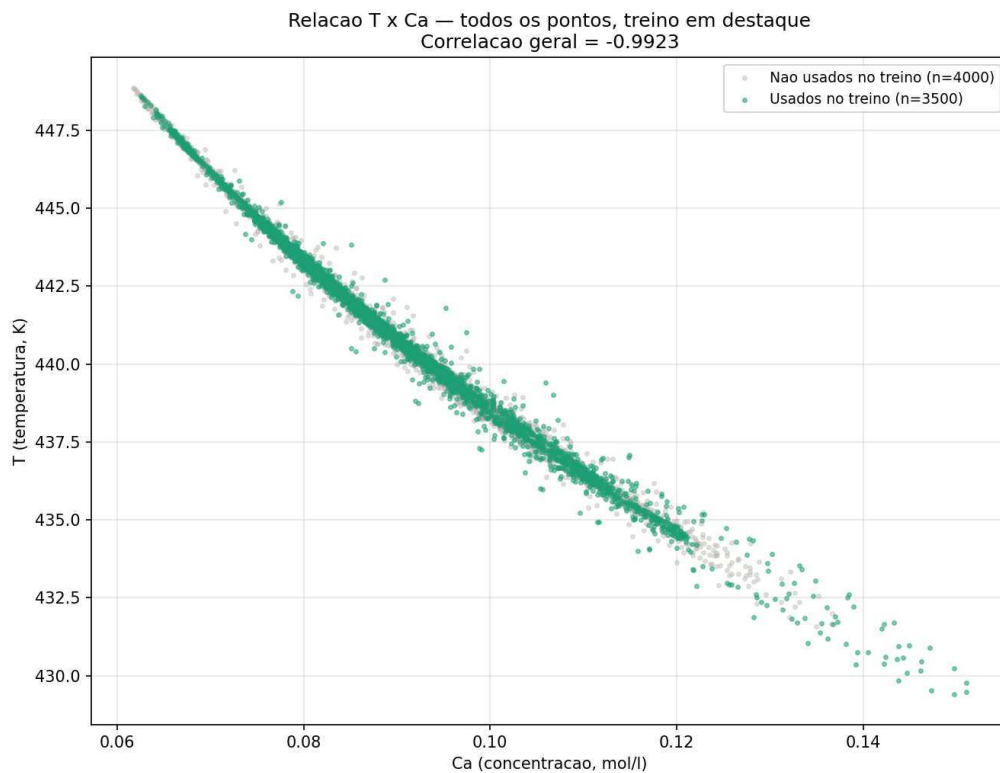


Figura 6 – Relação entre temperatura e concentração no conjunto de dados, com destaque para o trecho usado no treinamento.

Observa-se que as amostras do trecho de treinamento cobrem a maior parte da faixa de valores de concentração e temperatura observada no conjunto completo. A correlação entre as duas variáveis no recorte de treinamento foi de -0,9919, próxima da correlação geral do conjunto completo, igual a -0,9923. Isso indica que, apesar de a divisão cronológica ter alocado ao treinamento apenas os primeiros 350 dos 750 minutos totais de operação, a faixa de valores de concentração e temperatura vista pelo modelo durante o ajuste é representativa da faixa observada no restante da série.

Ainda assim, verificou-se que a amostra com menor concentração e maior temperatura de todo o conjunto de dados, $C_A = 0,0617$ mol/L e $T = 448,87$ K, não está contida no intervalo de treinamento, ocorrendo no início do trecho reservado ao teste externo. Essa observação não

compromete a validade da divisão cronológica adotada, mas ressalta que a representatividade da faixa operacional do treinamento em relação ao restante da série não é automaticamente garantida por esse tipo de divisão. Por isso, essa verificação complementar é importante para interpretar os resultados apresentados nas seções seguintes.

5.2 Caracterização do Conjunto de Dados

A Figura 7 apresenta as três variáveis utilizadas no experimento ao longo das 7500 amostras do conjunto CSTR da base DaISy. A linha vertical tracejada indica o início do trecho externo, a partir da amostra 5000, reservado para avaliação fora do recorte de desenvolvimento.

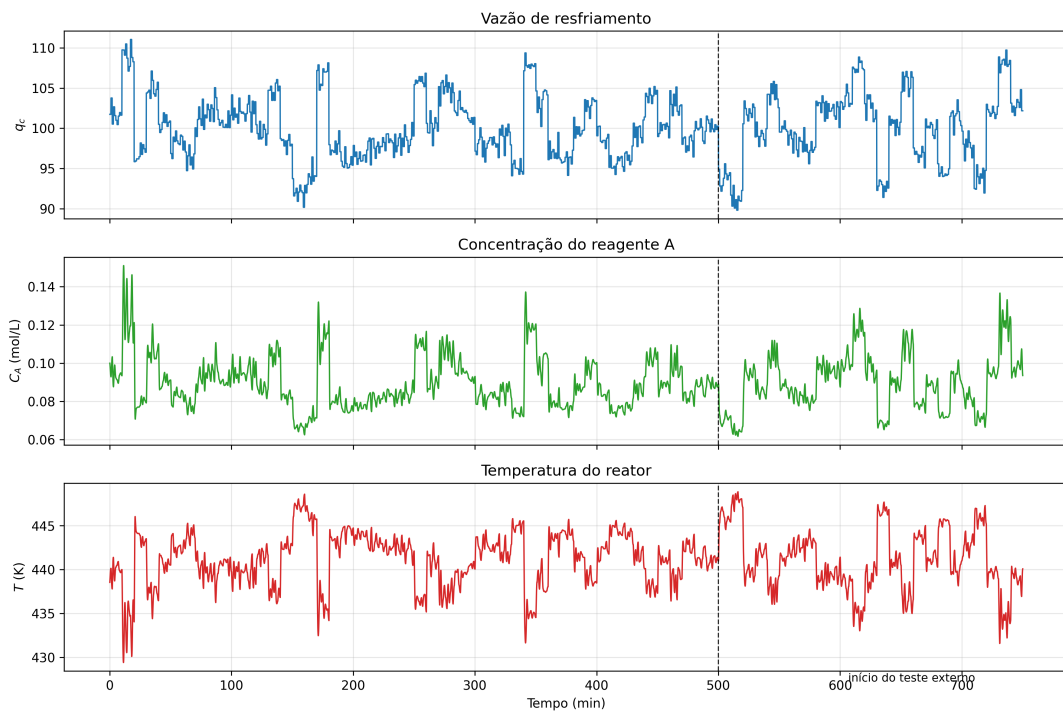


Figura 7 – Comportamento temporal de q_c , C_A e T no conjunto de dados CSTR.

Observa-se que a vazão de resfriamento apresenta mudanças por patamares e trechos de variação, enquanto concentração e temperatura respondem de forma dinâmica. As cristas e vales observados em C_A e T estão associados às alterações em q_c e ao acoplamento físico do reator: mudanças na remoção de calor afetam a temperatura e, pela cinética da reação, também influenciam a concentração. Esse comportamento justifica o uso de uma janela temporal com valores passados das três variáveis, pois a resposta do reator não é instantânea. Também explica por que a persistência pode apresentar bom desempenho em um passo à frente: com amostragem de 6 segundos, pontos consecutivos tendem a ser próximos, principalmente fora dos trechos de transição.

5.3 Evolução do Treinamento

A evolução da função de perda durante o treinamento é apresentada na Figura 8. O gráfico compara a perda no conjunto de treinamento e no conjunto de validação ao longo das épocas. Essa análise é importante para verificar se o modelo conseguiu reduzir o erro sem apresentar divergência evidente entre treinamento e validação.

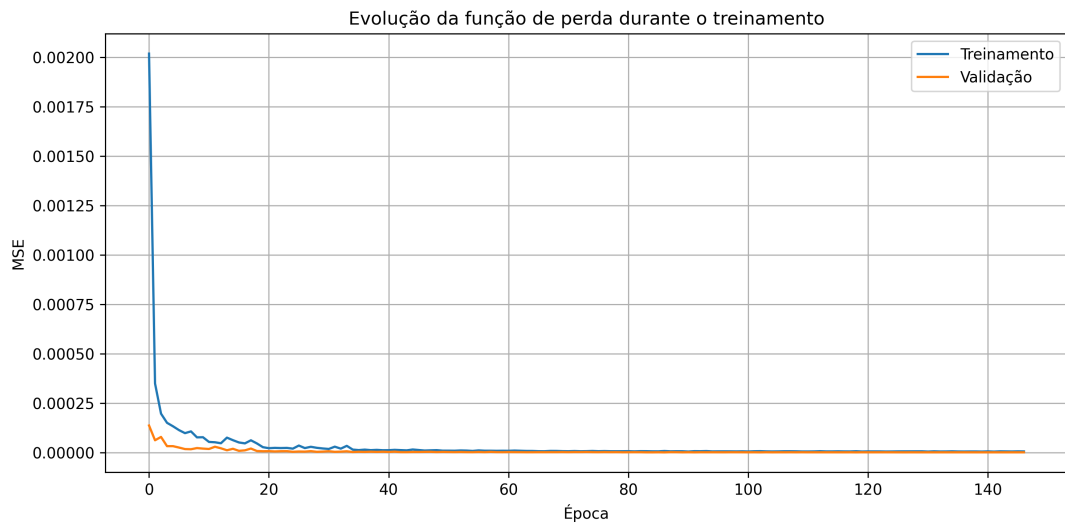


Figura 8 – Evolução da função de perda durante o treinamento da rede LSTM.

Como o treinamento utilizou *EarlyStopping* com restauração dos melhores pesos, o modelo final salvo foi mantido com os parâmetros associados ao melhor desempenho de validação alcançado durante o processo. A proximidade entre as curvas de treinamento e validação indica que não houve divergência evidente entre os conjuntos durante o ajuste. Ainda assim, essa leitura deve considerar que as janelas internas são deslizantes e, portanto, fortemente correlacionadas entre si.

5.4 Teste Interno de Um Passo à Frente

O teste interno foi realizado no segmento final das janelas geradas a partir das 5000 primeiras amostras do conjunto de dados. A Tabela 5 apresenta a comparação entre o modelo de persistência e a rede LSTM para as duas variáveis de saída.

Tabela 5 – Resultados no teste interno de um passo à frente.

Saída	Modelo	MAE	MSE	RMSE	R^2
C_A	Persistência	$5,3185 \times 10^{-4}$	$6,5297 \times 10^{-7}$	$8,0806 \times 10^{-4}$	0,989470
C_A	LSTM	$1,3947 \times 10^{-5}$	$5,2730 \times 10^{-10}$	$2,2963 \times 10^{-5}$	0,999991
T	Persistência	0,126644	0,038817	0,197021	0,988597
T	LSTM	0,003037	$2,2529 \times 10^{-5}$	0,004746	0,999993

No teste interno, a LSTM apresentou erro menor que o modelo de persistência nas duas saídas. Para C_A , a redução percentual do RMSE foi de aproximadamente 97,16%. Para T , a redução foi de aproximadamente 97,59%. Esses resultados indicam que o modelo aprendeu uma representação mais informativa que a simples repetição do último valor observado. Entretanto, a persistência também obteve valores absolutos baixos, com RMSE de $8,0806 \times 10^{-4}$ para C_A e 0,197021 para T . Isso confirma que o problema de um passo à frente é favorecido pela amostragem curta e pela suavidade local da série.

A Figura 9 apresenta a comparação gráfica entre valores reais, valores estimados pela LSTM e valores estimados pelo modelo de persistência. Como as curvas ficam muito próximas em parte do intervalo, a Figura 10 apresenta uma ampliação dos primeiros pontos do teste interno. A Figura 11 mostra o erro da LSTM ao longo do teste interno.

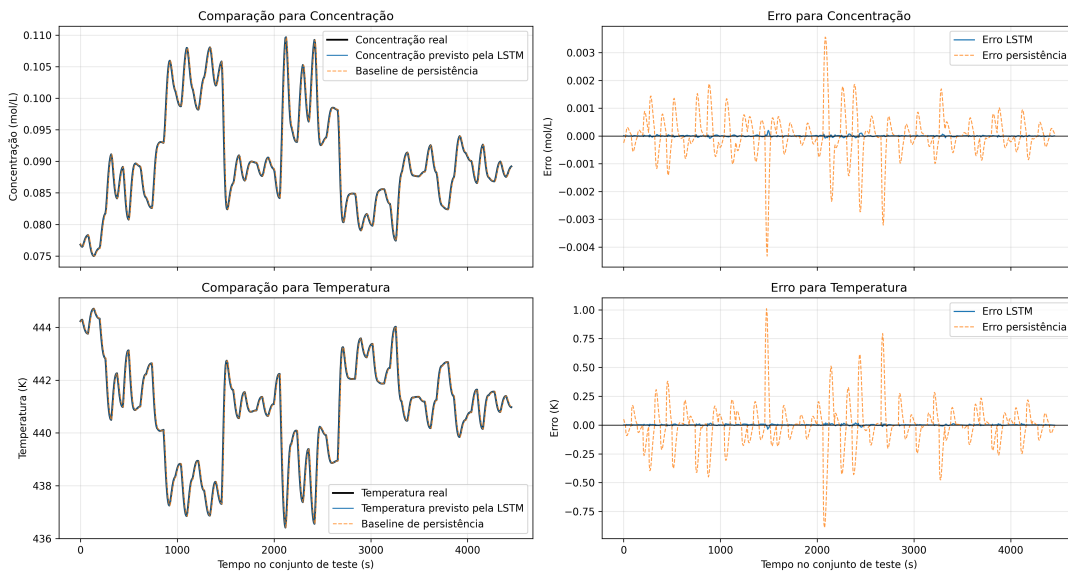


Figura 9 – Comparação entre valores reais, LSTM e persistência no teste interno.

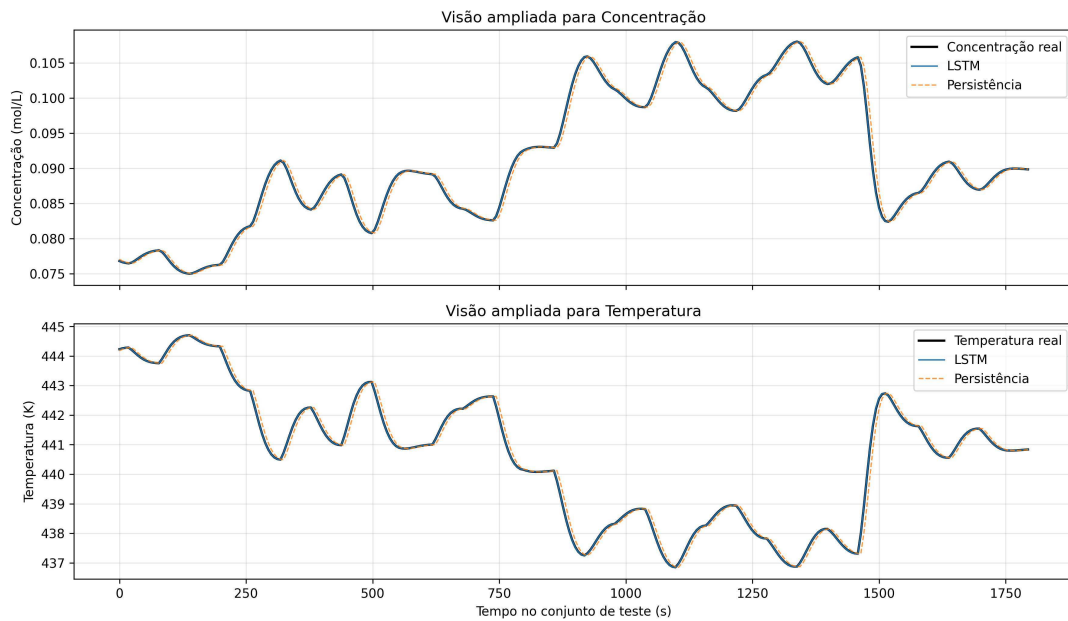


Figura 10 – Visão ampliada da comparação entre valores reais, LSTM e persistência no teste interno.

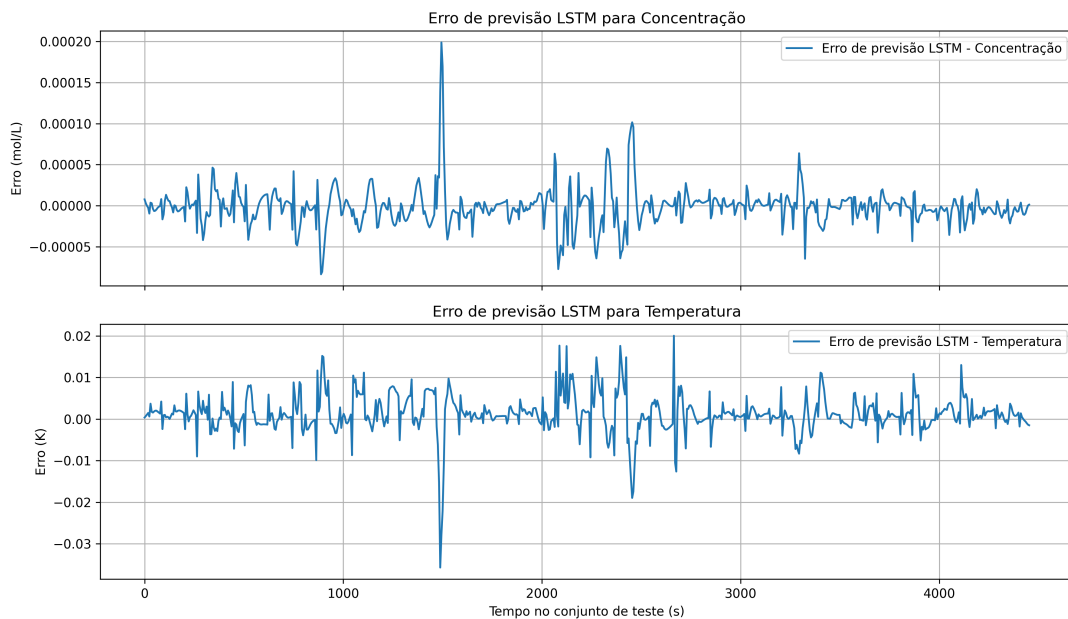


Figura 11 – Erro de estimação da LSTM no teste interno.

Na Figura 9, a curva da LSTM aparece quase sobreposta ao valor real porque o erro de um passo à frente é pequeno na escala das variáveis. Por esse motivo, a Figura 10 e a Figura 11 são mais adequadas para visualizar diferenças entre os modelos. O erro tende a aumentar nos trechos de maior inclinação das séries, isto é, próximos a cristas, vales e mudanças de regime. Nesses pontos, a persistência fica naturalmente atrasada, pois repete o último valor observado, enquanto a LSTM tenta estimar a variação local em relação ao último ponto da janela.

5.5 Teste Externo de Um Passo à Frente

O teste externo utilizou o trecho do conjunto de dados a partir do índice 5000. Esse trecho não participou do treinamento, da validação nem do teste interno. A Tabela 6 apresenta os resultados obtidos.

Tabela 6 – Resultados no teste externo de um passo à frente.

Saída	Modelo	MAE	MSE	RMSE	R^2
C_A	Persistência	$6,3582 \times 10^{-4}$	$1,0155 \times 10^{-6}$	$1,0077 \times 10^{-3}$	0,995303
C_A	LSTM	$1,8772 \times 10^{-5}$	$1,1241 \times 10^{-9}$	$3,3528 \times 10^{-5}$	0,999995
T	Persistência	0,149945	0,058186	0,241218	0,995237
T	LSTM	0,004058	$4,7844 \times 10^{-5}$	0,006917	0,999996

No teste externo de um passo à frente, a LSTM também superou o modelo de persistência. A redução percentual do RMSE foi de aproximadamente 96,67% para C_A e 97,13% para T . Esse resultado é relevante porque o trecho externo corresponde a uma região posterior da série temporal, não utilizada na etapa de desenvolvimento do modelo. Ao mesmo tempo, a persistência permaneceu competitiva em termos absolutos, com R^2 superior a 0,995 nas duas variáveis. Assim, o ganho da LSTM deve ser interpretado como melhoria sobre uma referência já forte para previsão local de 6 segundos, e não como correção de um baseline fraco.

A Figura 12 apresenta a comparação gráfica do teste externo de um passo à frente. A Figura 13 mostra uma ampliação do início do trecho externo para facilitar a análise visual da diferença entre as curvas.

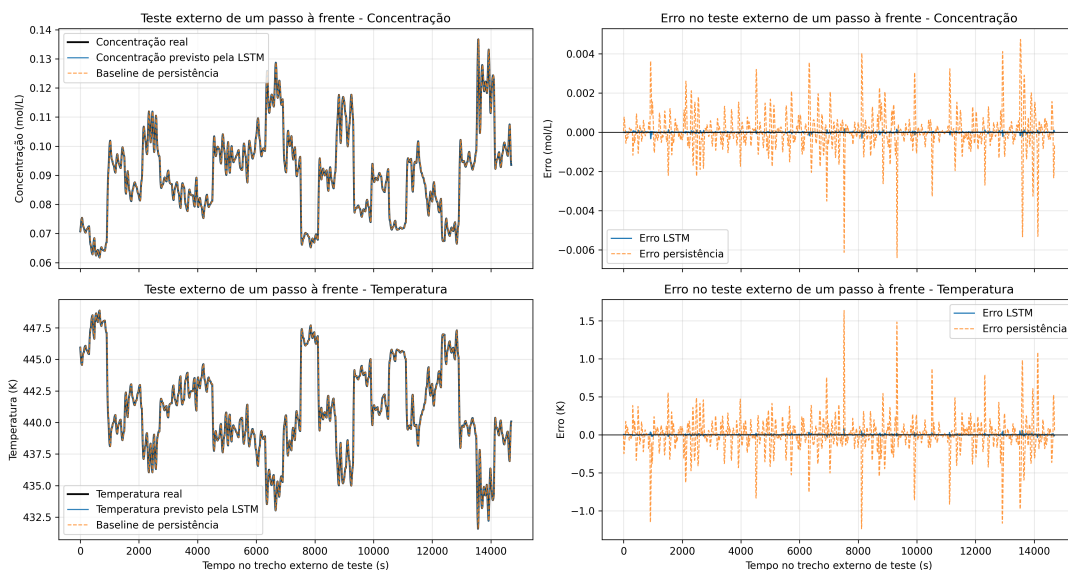


Figura 12 – Comparação entre valores reais, LSTM e persistência no teste externo de um passo à frente.

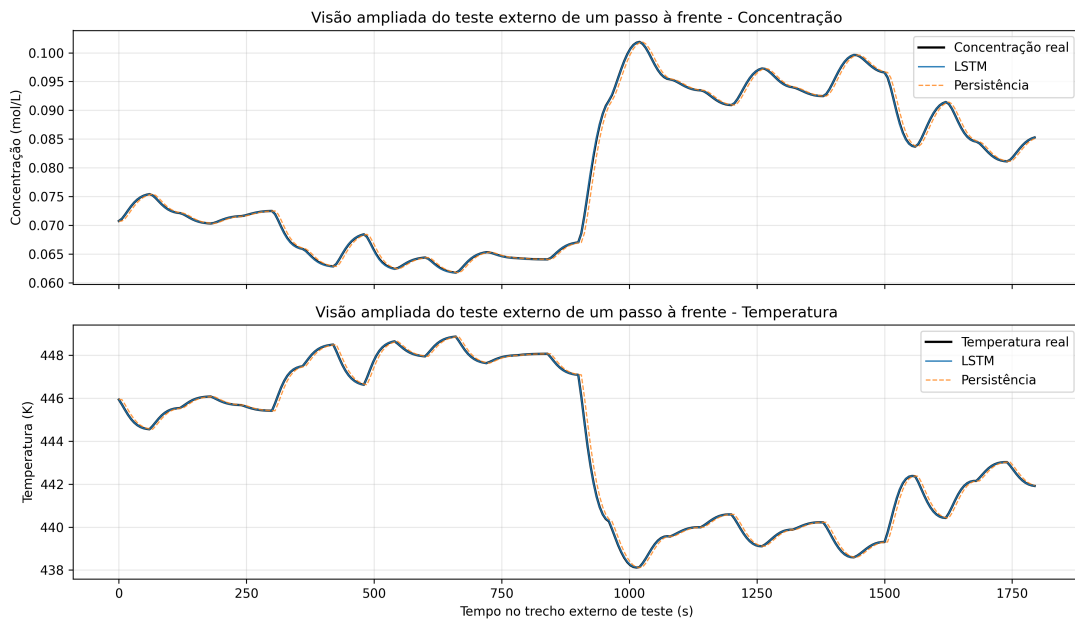


Figura 13 – Visão ampliada da comparação entre valores reais, LSTM e persistência no teste externo de um passo à frente.

As Figuras 12 e 13 mostram que a diferença entre LSTM e valor real também fica pequena no trecho externo. A distinção aparece com mais clareza nos instantes em que a série muda de direção: nesses pontos, a persistência apresenta erro de atraso, enquanto a LSTM reduz esse atraso ao estimar a variação do próximo passo. Essa observação é importante porque, em uma série com amostragem curta, apenas a sobreposição visual das curvas pode sugerir desempenho equivalente entre os modelos, embora as métricas mostrem diferença consistente.

5.6 Avaliação Externa Recursiva

A avaliação recursiva externa é mais rigorosa que a estimação de um passo à frente, pois as estimativas anteriores de C_A e T são reinseridas nas janelas seguintes. Dessa forma, o modelo passa a depender progressivamente de suas próprias estimativas para as variáveis de saída. A Tabela 7 apresenta as métricas obtidas pela LSTM nesse teste.

Tabela 7 – Resultados da avaliação externa recursiva da LSTM.

Saída	MAE	MSE	RMSE	R^2
C_A	$1,0825 \times 10^{-4}$	$2,8751 \times 10^{-8}$	$1,6956 \times 10^{-4}$	0,999867
T	0,029307	0,001689	0,041094	0,999862

Observa-se aumento do erro em relação ao teste externo de um passo à frente. Esse aumento ocorre porque, na avaliação recursiva, os valores estimados de C_A e T substituem medições reais nas janelas seguintes. Assim, um pequeno erro no instante atual altera a condição de entrada do instante posterior, podendo deslocar a trajetória simulada ao longo do tempo.

Mesmo assim, os valores de R^2 permaneceram elevados para as duas saídas, indicando que o modelo manteve boa capacidade de acompanhamento da tendência geral do processo.

As Figuras 14 e 15 apresentam, respectivamente, as estimativas recursivas e os erros associados.

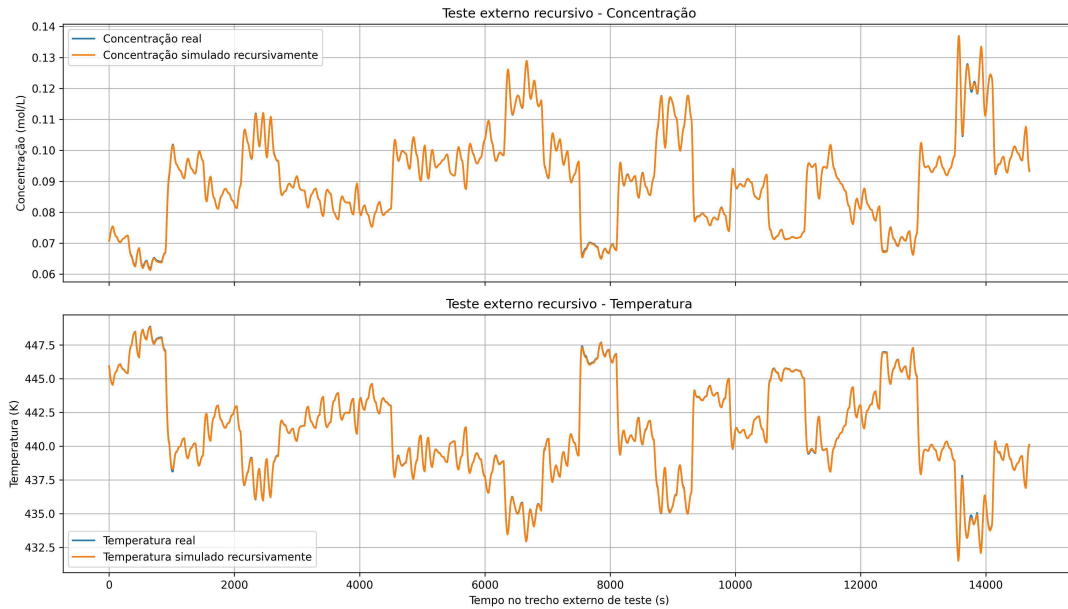


Figura 14 – Estimação externa recursiva da LSTM para concentração e temperatura.

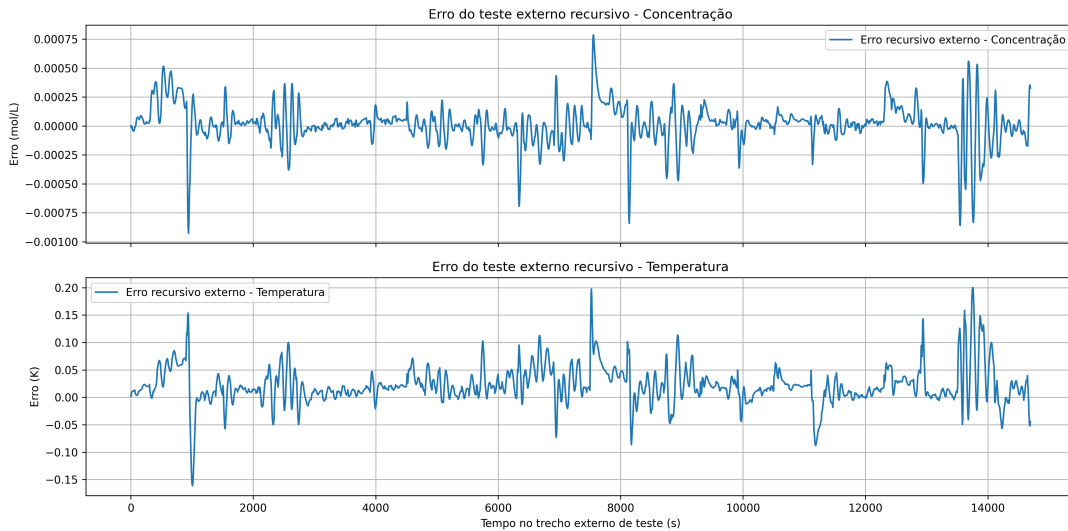


Figura 15 – Erro da estimação externa recursiva da LSTM.

Na avaliação recursiva, a análise visual deve ser mais cuidadosa que no caso de um passo à frente. A rede não apenas estima pontos isolados: ela passa a reconstruir uma trajetória usando parte de suas próprias saídas anteriores. Por isso, cristas e vales são pontos críticos, pois pequenos deslocamentos temporais podem produzir erro mesmo quando a forma geral da curva é

preservada. Esse comportamento é semelhante ao observado na persistência: uma curva atrasada pode apresentar erro métrico elevado sem necessariamente indicar perda completa da tendência física do processo.

5.7 Avaliação Recursiva Multi-step

Além da avaliação recursiva ao longo de todo o trecho externo, foram analisados horizontes fixos de 10, 20 e 50 passos. A Tabela 8 apresenta os resultados da LSTM e do modelo de persistência nesses horizontes.

Tabela 8 – Resultados da avaliação recursiva de múltiplos passos.

Horizonte	Saída	Modelo	MAE	RMSE	R^2
10	C_A	LSTM	$8,2695 \times 10^{-5}$	$1,4302 \times 10^{-4}$	0,999905
10	C_A	Persistência	0,005066	0,007520	0,738014
10	T	LSTM	0,019759	0,030711	0,999923
10	T	Persistência	1,200302	1,747806	0,749549
20	C_A	LSTM	$9,4848 \times 10^{-5}$	$1,5780 \times 10^{-4}$	0,999885
20	C_A	Persistência	0,006408	0,009613	0,571465
20	T	LSTM	0,024752	0,036616	0,999890
20	T	Persistência	1,522251	2,273735	0,575505
50	C_A	LSTM	$1,0678 \times 10^{-4}$	$1,6889 \times 10^{-4}$	0,999867
50	C_A	Persistência	0,010249	0,013899	0,096401
50	T	LSTM	0,028815	0,040761	0,999862
50	T	Persistência	2,443070	3,309512	0,090618

Os resultados de múltiplos passos mostram que o erro da LSTM cresce com o aumento do horizonte, mas permanece substancialmente menor que o erro da persistência. Para o horizonte de 50 passos, por exemplo, o RMSE da temperatura foi 0,040761 para a LSTM e 3,309512 para a persistência. Para a concentração, o RMSE foi $1,6889 \times 10^{-4}$ para a LSTM e 0,013899 para a persistência.

Essa diferença ajuda a interpretar a utilidade da LSTM. Em um passo à frente, a persistência já é uma referência forte porque o intervalo de amostragem é de apenas 6 segundos. À medida que o horizonte aumenta para 10, 20 e 50 passos, a persistência deixa de acompanhar a trajetória e passa a representar apenas um valor atrasado. A LSTM, por outro lado, utiliza o histórico de q_c , C_A e T para estimar a direção da variação. Assim, o ganho da rede torna-se mais evidente nos horizontes correspondentes a aproximadamente 1, 2 e 5 minutos.

As Figuras 16 e 17 ilustram dois casos da avaliação de múltiplos passos: concentração no horizonte 10 e temperatura no horizonte 50. Esses gráficos ajudam a visualizar a diferença entre a estratégia recursiva da LSTM e o modelo de persistência.

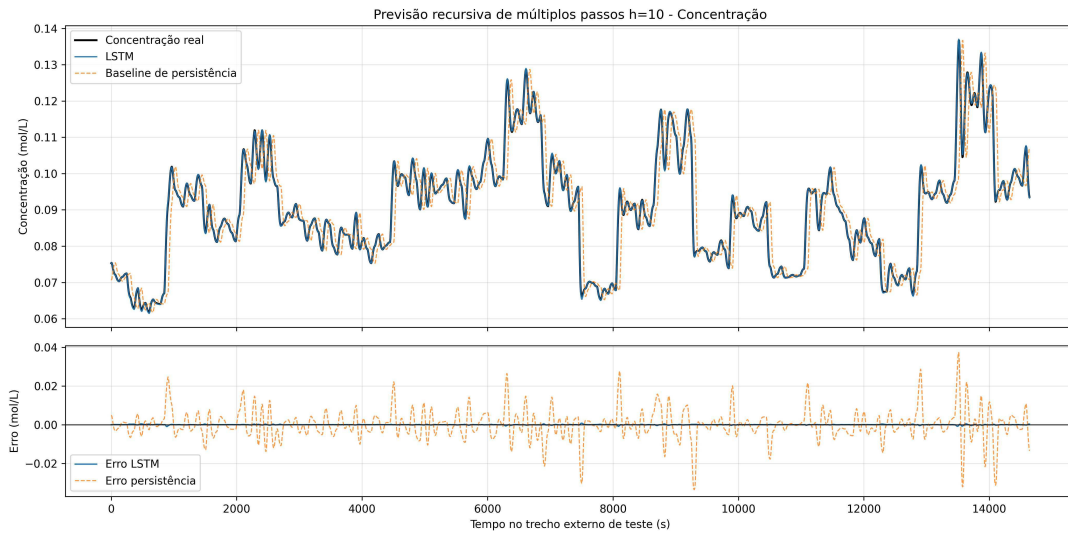


Figura 16 – Estimação de múltiplos passos da concentração para horizonte de 10 passos.

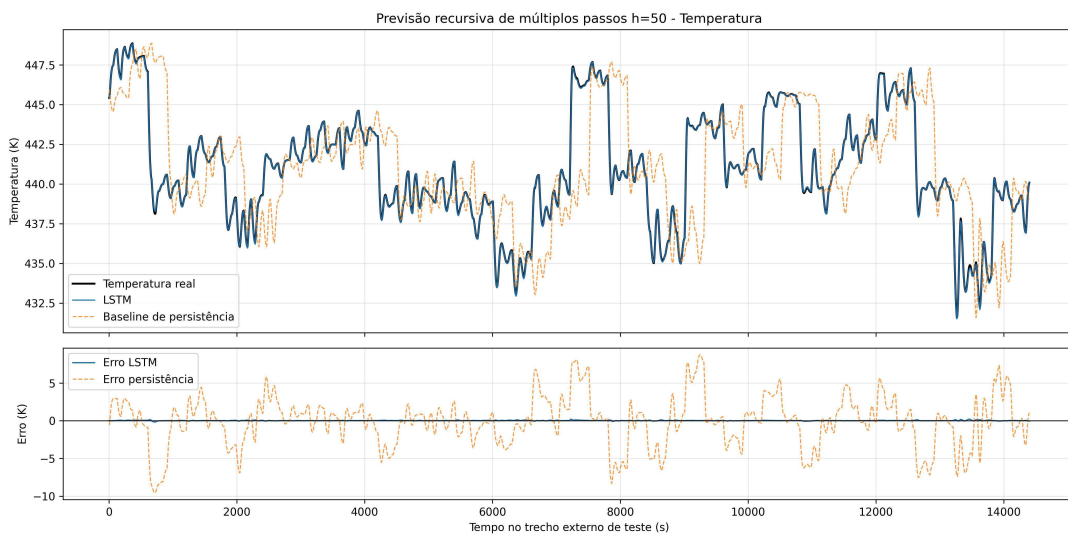


Figura 17 – Estimação de múltiplos passos da temperatura para horizonte de 50 passos.

As figuras de múltiplos passos evidenciam que o erro se concentra principalmente nas regiões de mudança mais intensa, e não em regime quase estacionário. Esse comportamento é coerente com a natureza do processo: quando a série está em patamar ou varia lentamente, repetir o último valor já aproxima bem o ponto futuro; quando há transição, o modelo precisa capturar a direção e a velocidade da resposta dinâmica.

5.8 Análise Complementar com Horizonte de Estimação de 5 Passos

Embora o experimento principal de um passo à frente apresente erros bastante reduzidos, esse cenário é naturalmente favorecido pela característica lenta do processo e pelo uso do último

valor medido na janela. Por esse motivo, a avaliação complementar com horizonte de 5 passos foi incluída para observar o comportamento dos modelos em uma tarefa de estimação mais desafiadora.

Nesse experimento complementar, o modelo utilizou uma janela de 80 amostras, equivalente a 480 s de histórico do processo, para estimar concentração e temperatura 5 passos à frente. Como o intervalo de amostragem é de 6 s, esse horizonte corresponde a uma estimação 30 s à frente. A Tabela 9 apresenta os resultados obtidos para persistência e LSTM.

Tabela 9 – Resultados da LSTM e da persistência no experimento complementar com horizonte de 5 passos.

Saída	Modelo	MAE	MSE	RMSE	R^2
C_A	Persistência	0,002515	$1,3638 \times 10^{-5}$	0,003693	0,777068
C_A	LSTM	$7,1700 \times 10^{-4}$	$1,7061 \times 10^{-6}$	0,001306	0,972112
T	Persistência	0,605139	0,778458	0,882302	0,767436
T	LSTM	0,235035	0,176144	0,419696	0,947377

Os resultados mostram que a LSTM superou o baseline de persistência nas duas variáveis. Para C_A , o RMSE passou de 0,003693 na persistência para 0,001306 na LSTM, com redução percentual de aproximadamente 64,63%. Para T , o RMSE passou de 0,882302 para 0,419696, com redução percentual de aproximadamente 52,43%. A diferença indica que, nesse cenário complementar, o uso do histórico temporal permitiu estimar melhor a evolução do processo do que a simples repetição do último valor conhecido.

Em comparação com os resultados de um passo à frente, os erros do horizonte de 5 passos foram maiores, o que é coerente com o aumento da distância entre a janela de entrada e o ponto estimado. Ainda assim, os valores de R^2 permaneceram elevados para concentração e temperatura, indicando que a LSTM manteve capacidade de acompanhar a tendência das duas variáveis no cenário complementar. Como essa análise também alterou tamanho de janela, dropout, tamanho de lote, camada densa e taxa de aprendizado, ela não permite atribuir a diferença exclusivamente ao horizonte de 5 passos. O resultado deve ser entendido como evidência adicional de uma configuração alternativa, não como comparação controlada de um único hiperparâmetro.

A Figura 18 apresenta a evolução da perda durante o treinamento da LSTM no experimento complementar. As Figuras 19, 20 e 21 mostram, respectivamente, a comparação entre valores reais, LSTM e persistência, uma visão ampliada da comparação e o erro de estimação da LSTM.

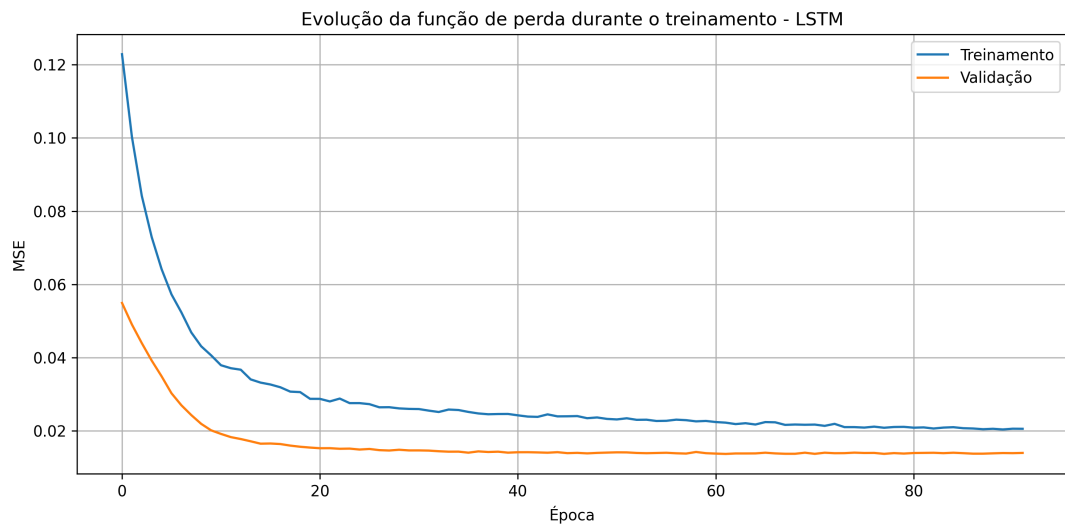


Figura 18 – Evolução da função de perda da LSTM no experimento complementar com horizonte de 5 passos.

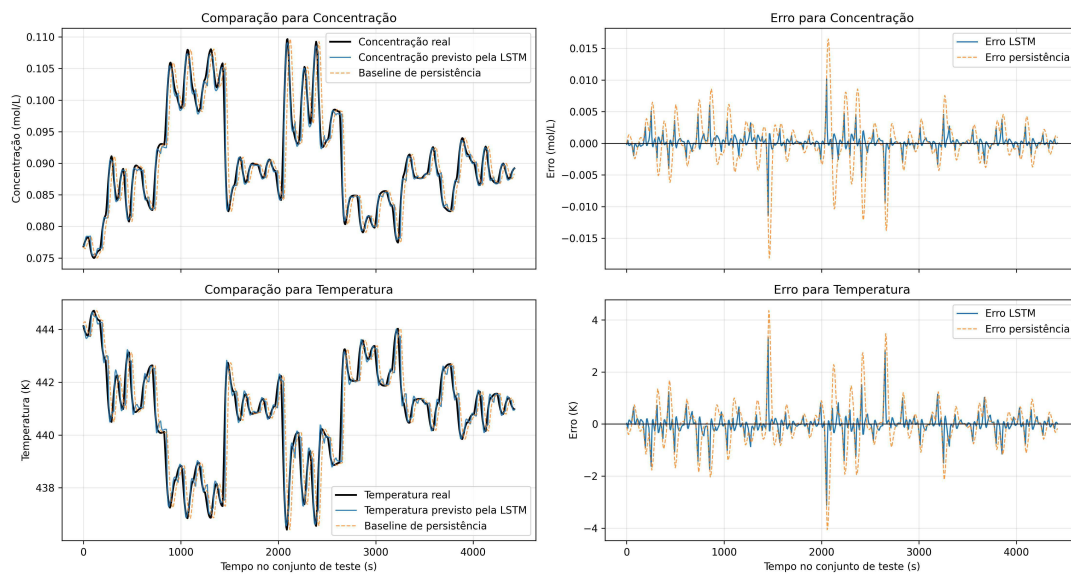


Figura 19 – Comparação entre valores reais, LSTM e persistência no experimento complementar com horizonte de 5 passos.

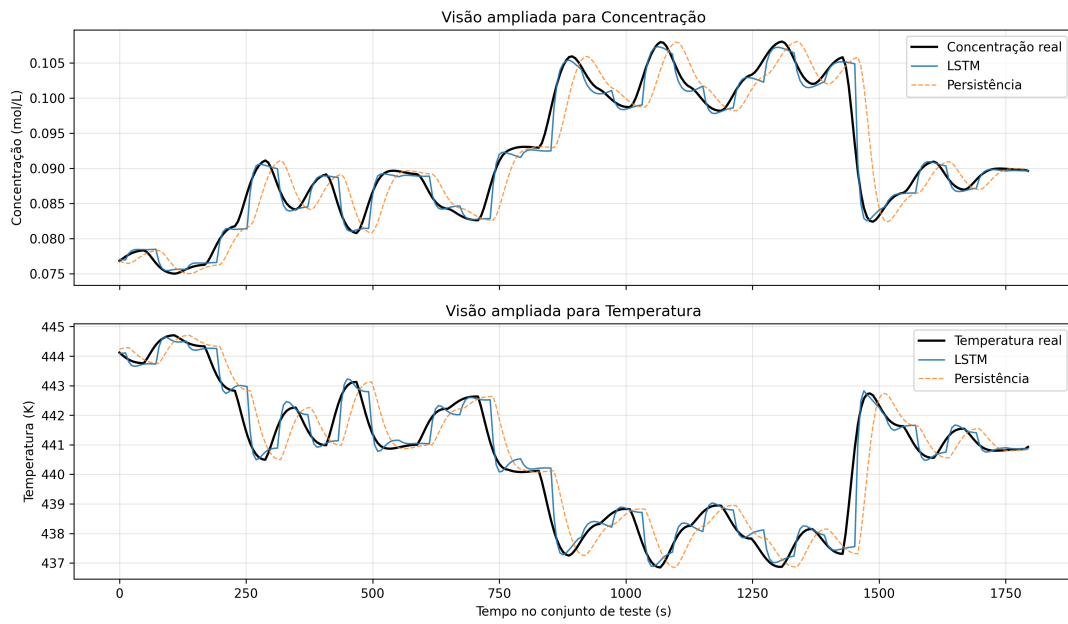


Figura 20 – Visão ampliada da comparação entre valores reais, LSTM e persistência no experimento complementar com horizonte de 5 passos.

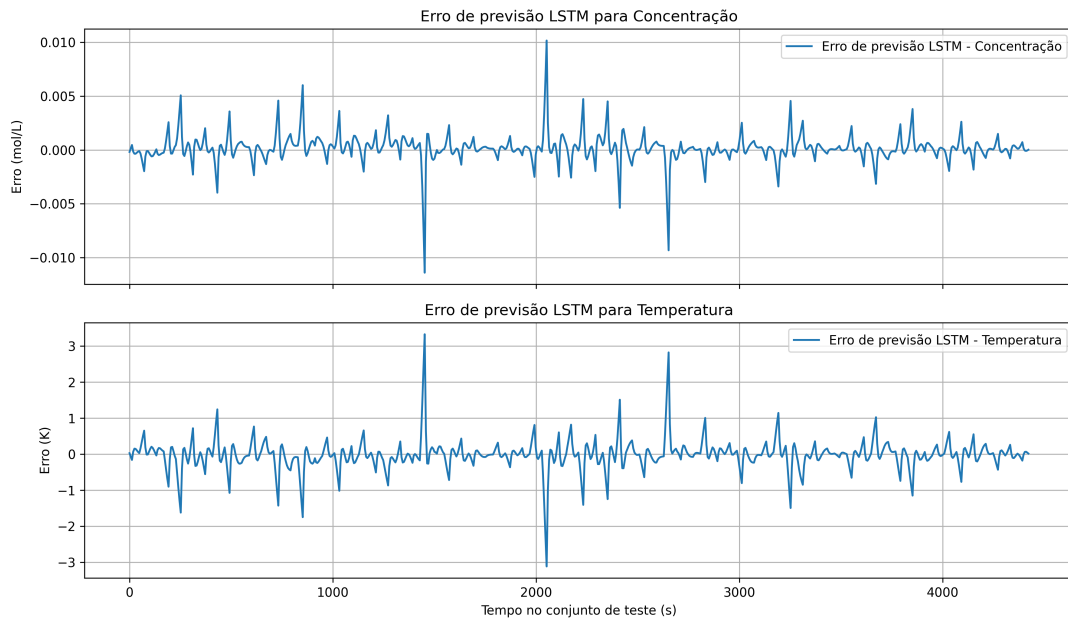


Figura 21 – Erro de estimação da LSTM no experimento complementar com horizonte de 5 passos.

5.9 Discussão dos Resultados

Os resultados indicam que a rede LSTM conseguiu representar a dinâmica conjunta de concentração e temperatura do CSTR com erros baixos nos testes realizados. A comparação com o modelo de persistência é especialmente importante porque o CSTR amostrado a cada 6

segundos apresenta variações relativamente suaves entre pontos consecutivos. Nesse contexto, a persistência é um baseline forte para um passo à frente, e sua boa performance não deve ser tratada como falha experimental. A contribuição da LSTM aparece com mais clareza ao reduzir o atraso em mudanças de direção e ao manter erro baixo em avaliações recursivas e de múltiplos passos.

A diferença entre o teste de um passo à frente e o teste recursivo também é relevante. No teste de um passo à frente, a janela de entrada contém apenas valores reais. No teste recursivo, parte das entradas passa a depender das estimativas anteriores do próprio modelo. O aumento de erro, portanto, não deve ser explicado apenas como algo “esperado”, mas como consequência da realimentação das estimativas e da sensibilidade dos trechos de transição. Ainda assim, os resultados indicam que o modelo manteve estabilidade suficiente para acompanhar a evolução das duas variáveis ao longo do trecho externo.

Do ponto de vista de uma possível aplicação em gêmeos digitais, os resultados mostram apenas que o modelo desenvolvido pode atuar como um componente preditivo orientado a dados, capaz de estimar simultaneamente C_A e T a partir do histórico recente de operação. Essa conclusão deve ser interpretada dentro dos limites do conjunto de dados utilizado: o modelo foi avaliado sobre a base CSTR disponível e não sobre uma planta física real em operação, não havendo integração em tempo real nem controle automático do processo.

6 CONCLUSÃO

Este trabalho teve como objetivo desenvolver e avaliar um modelo preditivo multi-saída baseado em rede neural LSTM para representar a dinâmica de concentração e temperatura de um CSTR. A proposta foi construída a partir do conjunto de dados CSTR da base DaISy e analisada como um componente preditivo orientado a dados, sem caracterizar a implementação como um gêmeo digital completo.

O desenvolvimento partiu da caracterização do processo, da definição das variáveis de entrada e saída e da preparação do conjunto de dados. A implementação utilizou como entradas a vazão de resfriamento q_c , a concentração C_A e a temperatura T , e como saídas estimadas C_A e T . A normalização foi realizada por média e desvio padrão, calculados apenas no trecho de treinamento, e as amostras foram organizadas em janelas temporais de 50 pontos, correspondentes a 5 minutos de histórico do processo.

A arquitetura implementada foi composta por uma camada LSTM com 128 unidades, uma camada de dropout, uma camada densa intermediária com 64 neurônios e uma camada de saída linear com dois neurônios. O treinamento utilizou função de perda MSE, otimizador Adam, batch size 16, até 500 épocas e callbacks de parada antecipada, redução da taxa de aprendizado e salvamento do melhor modelo.

Os resultados mostraram que a LSTM apresentou desempenho superior ao modelo de persistência nas avaliações realizadas. No teste interno de um passo à frente, o RMSE foi reduzido de $8,0806 \times 10^{-4}$ para $2,2963 \times 10^{-5}$ em C_A e de 0,197021 para 0,004746 em T . No teste externo de um passo à frente, o RMSE foi reduzido de $1,0077 \times 10^{-3}$ para $3,3528 \times 10^{-5}$ em C_A e de 0,241218 para 0,006917 em T . Ainda assim, a persistência apresentou erros absolutos baixos em um passo à frente, o que é coerente com a amostragem de 6 segundos e com a suavidade local da série.

Na avaliação recursiva externa, observou-se aumento do erro em relação à estimação de um passo à frente. Esse comportamento ocorre porque as estimativas anteriores de C_A e T passam a compor janelas futuras, de modo que pequenos desvios podem deslocar a trajetória simulada, principalmente em cristas, vales e trechos de maior variação. Ainda assim, o modelo manteve valores elevados de R^2 e erros compatíveis com uma boa capacidade de acompanhamento da dinâmica geral do processo. A avaliação de múltiplos passos para horizontes de 10, 20 e 50 passos também indicou desempenho superior da LSTM em relação à persistência, mesmo com crescimento gradual do erro à medida que o horizonte aumentou.

De forma complementar, a análise com horizonte de 5 passos indicou que uma configuração alternativa da LSTM também conseguiu estimar a dinâmica do processo em um cenário mais desafiador do que a estimação de um passo à frente. Embora os erros tenham aumentado em relação ao experimento principal, a rede manteve desempenho superior à persistência e valores elevados de R^2 para concentração e temperatura. Como esse experimento alterou outros

hiperparâmetros além do horizonte, ele deve ser interpretado como avaliação complementar, e não como análise isolada do efeito do horizonte de 5 passos.

Dessa forma, conclui-se que o modelo LSTM desenvolvido foi capaz de aprender relações temporais relevantes entre vazão de resfriamento, concentração e temperatura, representando de forma satisfatória a dinâmica multi-saída do CSTR no conjunto de dados analisado. A abordagem se mostrou adequada como componente preditivo orientado a dados, especialmente por combinar estimação simultânea de variáveis acopladas, avaliação em trecho externo e análise recursiva.

Apesar dos resultados positivos, algumas limitações devem ser destacadas. O modelo foi avaliado sobre uma base de dados específica, e não em uma planta física operando em tempo real. Além disso, a seleção dos hiperparâmetros foi definida na implementação do projeto e não foi acompanhada, neste trabalho, por uma busca sistemática ampla de arquiteturas alternativas. O teste interno também deve ser interpretado com cautela, pois janelas deslizantes vizinhas compartilham grande parte do histórico; por isso, o teste externo teve papel central na análise de generalização temporal.

Como trabalhos futuros, recomenda-se avaliar o modelo em outros trechos operacionais ou em dados obtidos de uma planta experimental, investigar arquiteturas alternativas como GRU e LSTM empilhada, realizar uma análise sistemática de hiperparâmetros e estudar a integração do modelo preditivo com estratégias de controle, diagnóstico de falhas ou monitoramento em tempo real. Também seria interessante comparar a LSTM com modelos clássicos de identificação de sistemas e com técnicas orientadas a dados mais simples, a fim de quantificar com maior clareza o custo-benefício da abordagem neural.

REFERÊNCIAS

- AGUIRRE, L. A. **Introdução à Identificação de Sistemas: Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais**. 3. ed. Belo Horizonte: Editora UFMG, 2007. Citado 4 vezes nas páginas 12, 19, 29 e 31.
- CORREIA, M. G. **Modelagem e Identificação de Sistemas Dinâmicos com Redes Recorrentes Profundas aplicados em Pedais de Distorção e Robôs Móveis**. Dissertação (Dissertação (Mestrado em Engenharia Elétrica e de Computação)) — Universidade Federal de Goiás, Goiânia, 2025. Citado 3 vezes nas páginas 14, 29 e 31.
- FULLER, A.; FAN, Z.; DAY, C.; BARLOW, C. Digital twin: Enabling technologies, challenges and open research. **IEEE Access**, v. 8, p. 108952–108971, 2020. Citado 2 vezes nas páginas 25 e 30.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge: MIT Press, 2016. Citado 4 vezes nas páginas 19, 21, 22 e 29.
- GRIEVES, M. **Digital Twin: Manufacturing Excellence through Virtual Factory Replication**. [S.l.], 2014. Citado 3 vezes nas páginas 12, 25 e 30.
- HAYKIN, S. **Neural Networks and Learning Machines**. 3. ed. New York: Prentice Hall, 2009. Citado 3 vezes nas páginas 19, 21 e 29.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. **Neural Computation**, v. 9, n. 8, p. 1735–1780, 1997. Citado 3 vezes nas páginas 12, 22 e 29.
- LEITE, G. N. M. **Rede Neural Recorrente LSTM para Identificação de um Sistema Térmico**. Monografia (Trabalho de Conclusão de Curso (Bacharelado em Engenharia Elétrica)) — Universidade Tecnológica Federal do Paraná, Apucarana, 2023. Citado 3 vezes nas páginas 14, 29 e 31.
- LJUNG, L. **System Identification: Theory for the User**. 2. ed. Upper Saddle River: Prentice Hall, 1999. Citado 4 vezes nas páginas 12, 19, 28 e 31.
- MOOR, B. L. R. D. **DaISy: Continuous Stirred Tank Reactor Dataset**. 2008. Department of Electrical Engineering, ESAT/STADIUS, KU Leuven, Belgium. Disponível em: <<http://homes.esat.kuleuven.be/~smc/daisy/>>. Atualizado pela última vez em 2008. Acesso em: 1 jun. 2026. Dataset utilizado: Continuous Stirred Tank Reactor, seção Process Industry Systems. Citado na página 32.
- MORAIS, V. R. **Deteção de falhas em um sistema reator de tanque com agitação contínua utilizando inteligência artificial**. Monografia (Trabalho de Conclusão de Curso (Bacharelado em Engenharia de Controle e Automação)) — Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais, Campus Sabará, Sabará, 2024. Citado 3 vezes nas páginas 14, 28 e 31.
- MORNINGRED, J. D.; PADEN, B. E.; SEBORG, D. E.; MELLICHAMP, D. A. An adaptive nonlinear predictive controller. **Chemical Engineering Science**, v. 47, n. 4, p. 755–762, 1992. Citado 5 vezes nas páginas 12, 16, 28, 31 e 32.
- NUNES, I. C. **Modelagem e Previsão de Variáveis Climáticas Usando Modelos SARIMA, VAR e LSTM**. Monografia (Trabalho de Conclusão de Curso (Bacharelado em Estatística)) — Universidade Federal de Uberlândia, Uberlândia, 2025. Citado na página 30.

OGATA, K. **Modern Control Engineering**. 5. ed. Upper Saddle River: Prentice Hall, 2010. Citado 2 vezes nas páginas 12 e 16.

RODRIGUES, N. T. J. **Redes Neurais Artificiais LSTM aplicadas na previsão de padrões de um forno de reaquecimento de tarugos**. Dissertação (Dissertação (Mestrado em Engenharia Química)) — Universidade Federal de Campina Grande, Campina Grande, 2022. Citado 3 vezes nas páginas 14, 29 e 31.

SEBORG, D. E.; EDGAR, T. F.; MELLICHAMP, D. A.; DOYLE, F. J. **Process Dynamics and Control**. 3. ed. Hoboken: Wiley, 2010. Citado 2 vezes nas páginas 12 e 16.

TAO, F.; ZHANG, M.; LIU, Y.; NEE, A. Y. C. **Digital Twin Driven Smart Manufacturing**. London: Academic Press, 2019. Citado 3 vezes nas páginas 12, 25 e 30.

APÊNDICE A – TRECHOS PRINCIPAIS DA IMPLEMENTAÇÃO COMPUTACIONAL

Este apêndice apresenta os trechos principais da implementação utilizados para conferir os parâmetros descritos na metodologia. Os códigos foram reduzidos aos pontos essenciais para facilitar a leitura e evitar repetição de rotinas auxiliares.

A.1 Configuração do experimento

O trecho a seguir resume as variáveis de entrada e saída, o tamanho da janela temporal, o horizonte de estimação, a divisão dos dados e os principais hiperparâmetros de treinamento.

```
1 dataset_name = "cstr"
2 dataset_path = "data/cstr.dat"
3 sampling_interval_seconds = 6.0
4
5 feature_columns = (
6     "coolant_flow",
7     "concentration",
8     "temperature",
9 )
10
11 target_columns = (
12     "concentration",
13     "temperature",
14 )
15
16 dataset_start_index = 0
17 dataset_end_index = 5000
18 external_test_start_index = 5000
19 external_test_end_index = None
20
21 window_size = 50
22 prediction_horizon = 1
23 target_mode = "delta"
24
25 train_ratio = 0.70
26 validation_ratio = 0.15
27 test_ratio = 0.15
28
29 lstm_units = 128
30 dense_units = 64
31 dropout_rate = 0.02
32 batch_size = 16
33 epochs = 500
34 learning_rate = 5e-4
35 early_stopping_patience = 60
```

```
36 reduce_lr_patience = 15
37 reduce_lr_factor = 0.5
```

Listing A.1 – Configuração principal do experimento com o CSTR.

A.2 Normalização

A normalização foi feita por média e desvio padrão. Os parâmetros são calculados no trecho de treinamento e reutilizados nos demais conjuntos.

```
1 def fit_standard_scaler(data):
2     mean = np.mean(data, axis=0)
3     std = np.std(data, axis=0)
4     std = np.where(std == 0.0, 1.0, std)
5     return StandardScalerParameters(mean=mean, std=std)
6
7
8 def transform_standard_scaler(data, scaler):
9     return (data - scaler.mean) / scaler.std
10
11
12 def inverse_transform_outputs(normalized_data, scaler):
13     return (normalized_data * scaler.std) + scaler.mean
```

Listing A.2 – Normalização padrão aplicada às variáveis.

A.3 Construção das janelas temporais

O alvo utilizado no treinamento corresponde à variação entre o valor futuro e o último valor conhecido da janela. Por isso, no modo delta, o modelo aprende a diferença entre dois instantes consecutivos.

```
1 for start_index in range(last_index):
2     end_index = start_index + window_size
3     target_index = end_index + prediction_horizon - 1
4
5     current_output = target_matrix[end_index - 1, :]
6     future_output = target_matrix[target_index, :]
7
8     input_windows.append(feature_matrix[start_index:end_index, :])
9
10    if target_mode in ["absolute", "direct"]:
11        target_values.append(future_output)
12    elif target_mode == "delta":
13        target_values.append(future_output - current_output)
```

Listing A.3 – Criação das janelas temporais e do alvo em modo delta.

A.4 Arquitetura e treinamento da LSTM

A rede recebe uma janela temporal com 50 amostras e três variáveis por amostra. A saída possui dois neurônios, correspondentes às variações estimadas de concentração e temperatura.

```
1 input_layer = tf.keras.layers.Input(
2     shape=(window_size, number_of_features),
3     name="time_window_input",
4 )
5
6 x = tf.keras.layers.LSTM(
7     units=lstm_units,
8     return_sequences=False,
9     name="lstm_layer",
10 )(input_layer)
11
12 x = tf.keras.layers.Dropout(rate=dropout_rate)(x)
13 x = tf.keras.layers.Dense(units=dense_units, activation="relu")(x)
14
15 output_layer = tf.keras.layers.Dense(
16     units=number_of_outputs,
17     activation="linear",
18     dtype="float32",
19 )(x)
20
21 model = tf.keras.Model(inputs=input_layer, outputs=output_layer)
22 model.compile(
23     optimizer=tf.keras.optimizers.Adam(learning_rate=learning_rate),
24     loss="mse",
25     metrics=[
26         tf.keras.metrics.MeanAbsoluteError(name="mae"),
27         tf.keras.metrics.RootMeanSquaredError(name="rmse"),
28     ],
29 )
```

Listing A.4 – Arquitetura LSTM utilizada no modelo preditivo.

```
1 early_stopping = tf.keras.callbacks.EarlyStopping(
2     monitor="val_loss",
3     patience=60,
4     restore_best_weights=True,
5 )
6
7 reduce_learning_rate = tf.keras.callbacks.ReduceLROnPlateau(
8     monitor="val_loss",
9     factor=0.5,
10    patience=15,
11    min_lr=1e-6,
```

```
12 )
13
14 model_checkpoint = tf.keras.callbacks.ModelCheckpoint(
15     filepath="best_lstm_model.keras",
16     monitor="val_loss",
17     save_best_only=True,
18 )
```

Listing A.5 – Callbacks utilizados durante o treinamento.

A.5 Modelo simples de comparação

Além da LSTM, foi utilizado um modelo simples de persistência para comparação. Esse modelo assume que o próximo valor será igual ao último valor observado na janela temporal, servindo como referência mínima para avaliar se a rede neural acrescenta capacidade preditiva ao problema analisado.

```
1 last_normalized_output = X_test[:, -1, output_feature_indices]
2 baseline_prediction = inverse_transform_outputs(
3     normalized_data=last_normalized_output,
4     scaler=target_scaler,
5 )
```

Listing A.6 – Baseline de persistência.