



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA MECÂNICA
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA MECÂNICA
MESTRADO ACADÊMICO EM ENGENHARIA MECÂNICA

JAIRON ISAIAS LEAL

PREVISÃO PROBABILÍSTICA DE VENTO VIA MODELOS DINÂMICOS LINEARES
EM UM CONTEXTO DE PRODUÇÃO DE HIDROGÊNIO

FORTALEZA

2024

JAIRON ISAIAS LEAL

PREVISÃO PROBABILÍSTICA DE VENTO VIA MODELOS DINÂMICOS LINEARES EM
UM CONTEXTO DE PRODUÇÃO DE HIDROGÊNIO

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia Mecânica do Programa de Pós-Graduação em Engenharia Mecânica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Mecânica. Área de Concentração: Equipamentos, processos e sistemas para energias renováveis

Orientadora: Profa. Dra. Carla Freitas de Andrade

Coorientador: Prof. Dr. Anselmo Ramalho Pitombeira Neto

FORTALEZA

2024

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- L471p Leal, Jairon Isaias.
Previsão probabilística de vento via modelos dinâmicos lineares em um contexto de produção de hidrogênio / Jairon Isaias Leal. – 2024.
83 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia, Programa de Pós-Graduação em Engenharia Mecânica, Fortaleza, 2024.
Orientação: Profa. Dra. Carla Freitas de Andrade.
Coorientação: Prof. Dr. Anselmo Ramalho Pitombeira Neto.
1. Energia eólica. 2. Previsão probabilística. 3. Modelos dinâmicos lineares. 4. Produção de hidrogênio verde. I. Título.

CDD 620.1

JAIRON ISAIAS LEAL

PREVISÃO PROBABILÍSTICA DE VENTO VIA MODELOS DINÂMICOS LINEARES EM
UM CONTEXTO DE PRODUÇÃO DE HIDROGÊNIO

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia Mecânica do Programa de Pós-Graduação em Engenharia Mecânica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia Mecânica. Área de Concentração: Equipamentos, processos e sistemas para energias renováveis

Aprovada em: 26 de Junho de 2024

BANCA EXAMINADORA

Profa. Dra. Carla Freitas de Andrade (Orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Anselmo Ramalho Pitombeira
Neto (Coorientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Paulo Alexandre Costa Rocha
Universidade Federal do Ceará (UFC)

Prof. Dr. Francisco Olimpio Moura Carneiro
Universidade da Integração Internacional da
Lusofonia Afro-Brasileira (UNILAB)

Prof. Dr. Fabrício Gonzalez Nogueira
Universidade Federal do Ceará (UFC)

À minha família, por sempre ser minha fortaleza
em termos emocionais e financeiros.

AGRADECIMENTOS

A Deus.

Aos meus pais, irmãs e sobrinha, que nos momentos de minha ausência foram extremamente compreensivos e incentivadores. Faço um agradecimento especial à minha irmã Layane, que, mais do que palavras de incentivo, sempre me ajudou de forma incondicional. Também não posso deixar de mencionar minha sobrinha/afilhada Clara, a fofuruxinha, pela capacidade genuína de irradiar alegria com um simples sorriso.

Aos meus orientadores, Prof^a. Dr^a. Carla Freitas de Andrade e Prof. Dr. Anselmo Ramalho Pitombeira Neto, pelo envolvimento e disponibilidade oferecidos durante o desenvolvimento da pesquisa. O conhecimento compartilhado por pesquisadores com diferentes expertises foi primordial para o meu desenvolvimento acadêmico.

Aos membros da banca, Prof. Dr. Paulo Alexandre Costa Rocha, Prof. Dr. Francisco Olimpio Moura Carneiro e Prof. Dr. Francisco de Assis de Souza Filho por aceitarem participar na banca de defesa e pelas sugestões. Em especial, agradeço ao professor Paulo Rocha pelas valiosas, e sempre muito oportunas, considerações e sugestões à pesquisa desde a qualificação.

Ao Prof. Dr. Bruno de Athayde Prata pela condução exemplar em suas aulas e, especificamente, pelo primeiro incentivo para o uso do *Latex* e da plataforma *OverLeaf*. Além disso, agradeço-o pelas sugestões e contribuições durante sua participação em minha qualificação.

À Prof^a. Dr^a. Ruth Pastôra Saraiva Leão pela sua capacidade singular de transmitir conhecimento em suas maravilhosas aulas sobre energia eolioelétrica.

A todos os colegas de turma de mestrado, pelo companheirismo ao longo do curso e pelo compartilhamento de experiências tanto em momentos de dificuldade como de diversão.

Ao Valdi Matos, secretário do programa de Pós-Graduação em Engenharia Mecânica, pela excelência profissional e prestatividade.

À bibliotecária Wesleyne Nunes, da biblioteca da Pós-Graduação em Economia da Universidade Federal do Ceará, pela assistência quanto às dúvidas sobre como referências particulares.

À Hohana Sanders, funcionária da Biblioteca de Pós-Graduação em Engenharia, pela presteza no atendimento e por zelar para manter o ambiente de estudo agradável na biblioteca.

Ao grupo Pesquisa Operacional em Produção e Logística (OPL), por me proporcionar uma convivência com pessoas de grande capacidade intelectual e pela estrutura física que foi extremamente útil em momentos de dificuldade relativas ao esforço computacional.

Aos autores Michael Heberl e Ali Javaid pela gentileza na elucidação de dúvidas quanto às suas respectivas pesquisas.

À Empresa de Pesquisa Energética pela disponibilização dos dados utilizados nesta pesquisa.

E, por fim, à Fundação Cearense de Apoio ao Desenvolvimento (Funcap) pelo apoio financeiro via bolsa de estudos em parte do período do Mestrado.

“Ein Geist, der sich für eine neue Idee öffnet,
wird niemals zu seiner ursprünglichen Größe
zurückkehren.”

(Albert Einstein)

RESUMO

A consolidação do hidrogênio renovável como um dos pilares para a transição energética depende da superação de alguns desafios operacionais como a previsibilidade da fonte primária e os impactos causados pela operação intermitente. Motivada por essa premissa, a presente pesquisa tem como objetivo avaliar a aplicação de *Dynamic Linear Models* na previsão probabilística de velocidade de ventos em um contexto de produção de hidrogênio. Para concretização desse propósito, medições anemométricas de usinas eólicas situadas nos estados da Bahia (BA), do Ceará (CE) e do Rio Grande do Sul (RS) foram usadas como dados de entrada para um procedimento de previsão contínua. As 329 previsões diárias resultantes desse procedimento foram utilizadas para estimar a produção de hidrogênio a partir de três métodos de conversão diferenciados entre si pela calibragem da eficiência do processo eletrolítico. Os resultados mostram que o CE obteve os melhores resultados em termos de ajuste pontual e de probabilidade de cobertura, mas a BA teve intervalos de predição relativamente mais precisos. Os valores medianos para CE, RS e BA de *normalized Root Mean Squared Error* são 0,1501, 0,2855 e 0,3272. Os valores medianos para CE, BA e RS de *Prediction Interval Coverage Probability* são 97,92%, 79,17% e 70,83%. Já os valores medianos de *Prediction Interval Normalized Average Width* são 65,66%, 69,15% e 73,94% para BA, CE e RS. Dentre os métodos de conversão eletricidade-hidrogênio, o segundo método de conversão (C2), que é calibrado de forma teórica, resultou em menores diferenças entre os valores observados e previstos no horizonte mensal. Sob a perspectiva diária, a amostra da BA conteve a maior quantidade diária de períodos críticos decorrentes de flutuações superiores a $\pm 50\%$.

Palavras-chave: Energia eólica. Previsão probabilística. Modelos dinâmicos lineares. Produção de hidrogênio verde.

ABSTRACT

The consolidation of renewable hydrogen as one of the mainstays of the energy transition depends on overcoming some operational challenges, such as the predictability of the primary energy source and the impacts caused by intermittent operation. Building on this premise, the present research aims to evaluate the application of Dynamic Linear Models for the probabilistic forecasting of wind speed in the context of hydrogen production. To this end, wind measurements of wind power plants located in the states of Bahia (BA), Ceará (CE) and Rio Grande do Sul (RS) were used as input for a rolling forecasting procedure. The 329 daily forecasts resulting from this procedure were used to estimate hydrogen production from three conversion methods differentiated from each other by calibrating the efficiency of the electrolytic process. The results show that CE obtained the best results regarding point adjustment and probability of coverage but BA had relatively more accurate prediction intervals. The normalized Root Mean Squared Error (nRMSE) median values for CE, RS, and BA are 0.1501, 0.2855, and 0.3272. The Prediction Interval Coverage Probability (PICP) median values for CE, BA, and RS are 97.92%, 79.17%, and 70.83%. The Prediction Interval Normalized Average Width (PINAW) median values are 65.66%, 69.15%, and 73.94% for BA, CE, and RS. Among the electricity-hydrogen conversion methods, C2, which is theoretically calibrated, resulted in smaller differences between the observed and predicted values over the monthly horizon. From a daily perspective, BA's sample contained the highest daily amount of critical periods due to fluctuations greater than $\pm 50\%$.

Keywords: Wind energy. Probabilistic wind speed forecasting. Dynamic Linear Models. Green hydrogen production.

LISTA DE FIGURAS

Figura 1 – Componentes básicos de uma planta de eletrólise de água sob a perspectiva da planta	22
Figura 2 – Curva de potência padrão de aerogeradores	23
Figura 3 – Eficiência da produção de hidrogênio em função da potência de entrada . . .	25
Figura 4 – Medições do fluxo anual do rio Nilo no período de 1871 e 1970, em $10^8 m^3$.	33
Figura 5 – Valores filtrados do fluxo anual do rio Nilo para duas razões sinal-ruído diferentes	34
Figura 6 – Previsões um passo à frente do fluxo anual do rio Nilo, considerando a construção da barragem	35
Figura 7 – Medições anemométricas realizadas em um parque eólico situado na Bahia (2019-2021)	45
Figura 8 – Medições anemométricas realizadas em um parque eólico situado no Ceará (2019-2021)	46
Figura 9 – Medições anemométricas realizadas em um parque eólico situado no Rio Grande do Sul (2019-2021)	47
Figura 10 – Subconjuntos de treinamento (azul) e teste (laranja) para as séries temporais da Bahia, Ceará e Rio Grande do Sul	48
Figura 11 – Fluxograma da implementação dos DLMs	49
Figura 12 – Ilustração dos procedimentos de previsão contínua	51
Figura 13 – Diagrama de caixa dos valores de nRMSE com faixas de cores referentes aos estados de acurácia. As cores representam os estados excelente (amarelo creme), bom (amarelo esverdeado), regular (amarelo claro) e pobre (amarelo escuro)	59
Figura 14 – Amostra dos resultados da previsão para a Bahia. Os dados observados (cinza) são contrastados com os valores previstos (verde) e com os intervalos de predição de 95% (verde claro)	61
Figura 15 – Amostra dos valores previstos com intervalo de predição de 95% para o Ceará. Os dados observados (cinza) são contrastados com os valores previstos (verde) e com os intervalos de predição de 95% (verde claro)	61

Figura 16 – Amostra dos valores previstos com intervalo de predição de 95% para o Rio Grande do Sul. Os dados observados (cinza) são contrastados com os valores previstos (verde) e com os intervalos de predição de 95% (verde claro) . . .	62
Figura 17 – Comparação da estimativa da produção mensal de hidrogênio por método de conversão — Bahia	63
Figura 18 – Comparação da estimativa da produção mensal de hidrogênio por método de conversão — Ceará	64
Figura 19 – Comparação da estimativa da produção mensal de hidrogênio por método de conversão — Rio Grande do Sul	65
Figura 20 – Comparação da estimativa da produção diária de hidrogênio com dados observados no dia 06/06/2021. As barras destacadas representam flutuações de geração eólicoelétrica superiores a $\pm 50\%$ — Bahia	66
Figura 21 – Comparação da estimativa da produção diária de hidrogênio com dados observados no dia 29/11/2021. As barras destacadas representam flutuações de geração eólicoelétrica superiores a $\pm 50\%$ — Ceará	67
Figura 22 – Comparação da estimativa da produção diária de hidrogênio com dados observados no dia 06/12/2021. As barras destacadas representam flutuações de geração eólicoelétrica superiores a $\pm 50\%$ — Rio Grande do Sul	68
Figura 23 – Estimativa da produção de hidrogênio a partir dos dados observados (azul), previstos (preto) e dos valores de L_{sup} e L_{inf} (cinza) referentes ao dia 06/06/2021 — Bahia	69
Figura 24 – Estimativa da produção de hidrogênio a partir dos dados observados (azul), previstos (preto) e dos valores de L_{sup} e L_{inf} (cinza) referentes ao dia 29/11/2021 — Ceará	70
Figura 25 – Estimativa da produção de hidrogênio a partir dos dados observados (azul), previstos (preto) e dos valores de L_{sup} e L_{inf} (cinza) referentes ao dia 06/12/2021 — Rio Grande do Sul	72

LISTA DE TABELAS

Tabela 1 – Os 10 artigos mais recentes com aplicação de métodos bayesianos e FK para previsão de ventos	40
Tabela 2 – Informações gerais sobre cada série temporal após a imputação dos dados .	44
Tabela 3 – Diferentes classes de acurácia do modelo preditivo medida pelo nRMSE . .	52
Tabela 4 – Linearização trapezoidal da eficiência de um eletrolisador	54
Tabela 5 – Informações técnicas dos aerogeradores selecionados	56
Tabela 6 – Semelhanças e diferenças entre os métodos de modelagem do eletrolisador .	56
Tabela 7 – Comparação dos resultados da métrica nRMSE para as janelas de tempo semanal e diária	58
Tabela 8 – Comparação dos resultados das três séries temporais, segundo as métricas PICP e PINAW	60

LISTA DE ABREVIATURAS E SIGLAS

AC	Análise Caótica
ANN	<i>Artificial Neural Networks</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
BA	Bahia
C1	primeiro método de conversão
C2	segundo método de conversão
C3	terceiro método de conversão
CCP	Coefficiente de Correlação de Pearson
CE	Ceará
DLMs	<i>Dynamic Linear Models</i>
EKF	<i>Extended Kalman Filter</i>
EPE	Empresa de Pesquisa Energética
FK	Filtro de Kalman
MAE	<i>Mean Absolute Error</i>
MAPE	<i>Mean Absolute Percentage Error</i>
MCCM	Monte Carlo com Cadeias de Markov
MDPR	Modelo baseado na Densidade de Probabilidade de Rayleigh
MDPW	Modelo baseado na Densidade de Probabilidade de Weibull
MPPT	<i>Maximum Power Point Tracking</i>
MSE	<i>Mean Square Error</i>
nRMSE	<i>normalized Root Mean Squared Error</i>
PCE	<i>Power Curve Error</i>
PCI-H2	Poder Calorífico Inferior do hidrogênio
PG	Processo Gaussiano
PICP	<i>Prediction Interval Coverage Probability</i>
PINAW	<i>Prediction Interval Normalized Average Width</i>
Pot	Potência
RKF	<i>Recurrent Kalman Filter</i>
RMSE	<i>Root Mean Square Error</i>
RS	Rio Grande do Sul
SCEH2	Sistema de Conversão Eólica-Hidrogênio

SVM	<i>Support Vector Machines</i>
TMDLM	<i>Truncated bivariate Matrix Bayesian Dynamic Linear Model</i>
UEEs	Usinas Eolioelétricas
UKF	<i>Unscented Kalman Filter</i>
VV	velocidade do vento
WD	<i>Wavelet Decomposition</i>

SUMÁRIO

1	INTRODUÇÃO	17
1.1	Objetivos	19
1.2	Estrutura da Dissertação	20
2	REFERENCIAL TEÓRICO	21
2.1	Modelagem de sistemas de conversão eólica-hidrogênio	21
2.2	Influência da flutuação da fonte primária no desempenho de eletrolisadores	26
2.3	Previsão probabilística via DLMS	29
2.3.1	<i>Modelo de Nível Local</i>	36
2.3.2	<i>Modelo com Tendência Linear</i>	36
2.3.3	<i>Modelo com Sazonalidade representada por séries de Fourier</i>	37
2.3.4	<i>Estimativa de parâmetros desconhecidos</i>	38
2.4	Uso de métodos bayesianos e do filtro de Kalman na previsão de ventos .	39
3	METODOLOGIA	43
3.1	Análise descritiva dos dados e das áreas de estudo	43
3.2	Metodologia de implementação dos DLMS	47
3.2.1	<i>Métricas de desempenho</i>	52
3.3	Metodologia para estimativa da produção de hidrogênio	53
3.3.1	<i>Método de conversão utilizado por Javaid et al (2022)</i>	53
3.3.2	<i>Metodologia apresentada por Zhu et al (2022)</i>	54
3.3.3	<i>Metodologia apresentada por Hofrichter et al. (2023)</i>	55
3.3.4	<i>Síntese dos métodos utilizados para a modelagem dos eletrolisadores</i> . . .	56
4	RESULTADOS E DISCUSSÃO	58
4.1	Resultados da previsão do vento	58
4.2	Resultados da modelagem do SCEH2	63
5	CONCLUSÃO	73
	REFERÊNCIAS	75

1 INTRODUÇÃO

A temática da transição energética fomenta debates sobre as adequações necessárias para a substituição gradual das fontes fósseis pelas fontes tidas como alternativas e/ou renováveis. Nesse contexto, a integração de fontes renováveis com o hidrogênio ganha destaque como um alicerce desse processo de transformação energética (Gielen *et al.*, 2019; Kovač *et al.*, 2021).

Em particular, a relação potencialmente simbiótica entre os mercados de hidrogênio e eólico — em que o hidrogênio pode compor a estratégia de armazenamento de longo prazo e a energia eolielétrica pode ser o selo certificador do lastro renovável da produção de hidrogênio —, tem incentivado a cooperação entre os setores público e privado, além da academia, para a criação de uma estrutura de mercado capaz de tornar o hidrogênio uma *commodity* que movimentará o comércio internacional.

No Brasil, tal cenário representa uma oportunidade singular para o desenvolvimento industrial do país em face de seu grande potencial em recursos renováveis, em infraestrutura de mercado e em capital humano. Como iniciativas para materialização dessa expectativa, pode-se citar a elaboração do Programa Nacional do Hidrogênio, que delinea as políticas e diretrizes relacionadas ao mercado de hidrogênio para o Brasil (MME, 2023); a criação da Comissão Especial de Transição Energética e Produção de Hidrogênio Verde na Câmara dos Deputados, que acompanhará a implementação de medidas para transição energética no Brasil (Agência Câmara de Notícias, 2023); e a criação do Grupo de Trabalho no Ceará, que tem o objetivo de conduzir a implantação do centro de hidrogênio no estado (Ceará, 2023).

A concretização dessas iniciativas dependerá, no entanto, da superação de alguns desafios técnicos associados aos impactos causados pela integração de sistemas de produção de hidrogênio com fontes primárias flutuantes. Afinal, espera-se que plantas industriais de hidrogênio sejam capazes de atender a determinada demanda por hidrogênio da forma mais rentável possível, o que inclui a capacidade de otimizar a operação em termos de compatibilidade com o sistema energético (Grüger *et al.*, 2019).

Tal compatibilidade tem diferentes níveis de complexidade, pois depende da integração (ou não) com a rede elétrica. Segundo Ceylan e Devrim (2023), sistemas de produção de hidrogênio renovável podem apresentar duas configurações: sistemas conectados à rede ou sistemas desconectados da rede. Em situações em que a planta de hidrogênio renovável está conectada à rede elétrica, esta atua como suplemento para mitigar o impacto da intermitência da fonte primária, o que, conseqüentemente, agrega mais estabilidade ao sistema (Xu *et al.*, 2022;

Bahou, 2023).

Os sistemas desconectados da rede elétrica localizam-se em regiões em que a conexão com a rede elétrica é financeira e fisicamente inviável — geralmente isso ocorre em locais remotos ou isolados. Nesses casos, a intermitência incide de forma mais direta no processo de produção de hidrogênio. Os sistemas híbridos isolados e as plantas de eletrólise embarcadas em usinas eólicas em alto mar são exemplos de sistemas de produção de hidrogênio renovável desconectados da rede elétrica (Sahraie *et al.*, 2024; Ramakrishnan *et al.*, 2024).

Conforme destacado por Wang *et al.* (2024), a maior exposição às flutuações da fonte primária suscita uma otimização do gerenciamento da oferta e da demanda de energia. Diante dessa premissa, a previsibilidade da fonte primária se torna um fator fundamental para viabilidade desses sistemas, pois podem otimizar o planejamento operacional a fim de mitigar os riscos associados à incerteza da geração eólicoelétrica e, por conseguinte, evitar perdas econômicas e garantir a segurança de suprimento em um contexto de produção de hidrogênio.

Ao longo dos últimos anos, o campo de estudo sobre previsão de ventos — doravante tratada como sinônimo de previsão da geração eólicoelétrica e demais expressões correlatas — tem apresentado grande avanço científico devido à incorporação de métodos sofisticados que tornam as previsões determinísticas (ou pontuais) cada vez mais acuradas. É imprescindível considerar, no entanto, que todo processo preditivo envolve incerteza por si só. Em face disso, outra tipologia, denominada de previsão probabilística, é apontada como uma das principais tendências para literatura de previsão de ventos, pois essa abordagem possui a capacidade de quantificar as incertezas associadas aos valores previstos sob qualquer nível de confiança (Zhang *et al.*, 2019; Sun *et al.*, 2020; Xie *et al.*, 2023; Hu *et al.*, 2024).

Como uma alternativa, Li e Shi (2012) apontam que a adoção de métodos bayesianos tem grande potencial de tornar os processos de caracterização e previsão do vento mais precisos, confiáveis e adaptativos. Tal capacidade para lidar com as idiossincrasias do vento justifica-se pela premissa de que os métodos bayesianos são fundamentados no gerenciamento da incerteza a partir da representação e aferição desta por meio de leis de probabilidades, que, por sua vez, produzem inferências probabilísticas sobre o fenômeno estudado (West; Harrison, 1997).

Dentre os métodos bayesianos, os modelos de espaços de estados têm grande repercussão na literatura por flexibilizarem a modelagem de séries temporais e de processos espaço-temporais ao permitirem a previsão da evolução temporal de uma variável de resposta em um determinado domínio espacial (Bessac *et al.*, 2015). A modelagem bayesiana através

dos *Dynamic Linear Models* (DLMs) — que são casos particulares dos modelos de espaços de estados — configura-se como uma abordagem oportuna para a análise de séries temporais, pois, além de considerar os próprios dados históricos, também permite a incorporação de outros fatores como o conhecimento de fenômenos futuros e de outros eventos excepcionais.

Em comparação com outras técnicas de análise de séries temporais, como o modelo de *Autoregressive Integrated Moving Average* (ARIMA), os DLMs têm a vantagem de não exigirem a transformação preliminar dos dados para lidar com a não-estacionariedade. Ou seja, a modelagem dinâmica torna possível a análise de séries temporais que não se desenvolvem em torno de uma média constante (Souza, 2014). Além disso, quando comparado à modelagem do tipo caixa preta, como rede neural, os DLMs diminuem a subjetividade na interpretação da relação causal entre os parâmetros de entrada do modelo e o fenômeno analisado, o que é reflexo direto de sua parametrização dinâmica. Tais características explicitam a atratividade da implementação dessa técnica bayesiana como estratégia para previsão probabilística do recurso eólico em um contexto de produção de hidrogênio.

Em face de todo o contexto apresentado, o caráter inovativo da presente pesquisa consiste em avaliar a aplicação de uma modelagem preditiva do vento orientada à incerteza para o campo temático que trata sobre sistemas de conversão eólica-hidrogênio, além de reunir diferentes estudos que modelam esses sistemas. Ademais, a tempestividade da presente pesquisa consiste na contribuição para o avanço da literatura em um momento em que há um movimento global de transição energética.

1.1 Objetivos

A presente pesquisa tem como objetivo geral investigar a aplicação de modelos dinâmicos lineares bayesianos para a previsão probabilística de vento em um contexto de produção de hidrogênio.

Como objetivos específicos, podem ser destacados:

- Realizar um levantamento sobre os principais impactos da flutuação da fonte primária no funcionamento de eletrolisadores.
- Analisar a viabilidade da modelagem dinâmica bayesiana como recurso de otimização para sistemas de conversão eólica-hidrogênio.
- Validar os modelos preditivos a partir de estudos de caso com dados medidos em regiões com diferentes características climatológicas.

- Comparar diferentes métodos de calibragem da eficiência de eletrolisadores.

1.2 Estrutura da Dissertação

A estrutura deste trabalho está organizada em cinco capítulos. O primeiro deles é esta introdução, na qual é feita a contextualização sobre o tema em estudo, são apresentados os objetivos da pesquisa e a justificativa. O segundo capítulo apresenta um resumo dos métodos de conversão de energia eólicoelétrica em hidrogênio, uma revisão sobre a influência da flutuação da fonte primária no desempenho de eletrolisadores e a base teórica dos DLMS. O terceiro capítulo é destinado à descrição dos procedimentos metodológicos adotados tanto para a previsão bayesiana das séries temporais de vento quanto para a modelagem da conversão eólica-hidrogênio. No capítulo quatro são apresentados e discutidos os resultados desta pesquisa. No quinto capítulo, são apresentadas as principais conclusões e as sugestões de trabalhos futuros.

2 REFERENCIAL TEÓRICO

Neste Capítulo, as linhas temáticas que fundamentam a presente pesquisa são apresentadas em três seções. A Seção 2.1 apresenta alguns conceitos básicos e descreve métodos utilizados para a modelagem de sistemas de conversão eólica-hidrogênio. Na Seção 2.2, é realizado um levantamento sobre a influência da flutuação da fonte primária no desempenho de eletrolisadores. Na Seção 2.3, a modelagem dinâmica bayesiana é apresentada como uma alternativa promissora para a previsão de ventos em um contexto de produção de hidrogênio.

2.1 Modelagem de sistemas de conversão eólica-hidrogênio

Embora seja o elemento mais abundante no universo, o hidrogênio em sua forma gasosa — que é aquela demandada industrialmente como transportadora de energia e/ou matéria-prima de processos industriais — está disponível em quantidades muito baixas na natureza (Gupta *et al.*, 2022). Como consequência, processos termoquímicos, biológicos e bioeletroquímicos, além da hidrólise e da eletrólise da água, são usados como métodos de produção do hidrogênio (Chaudhary *et al.*, 2024).

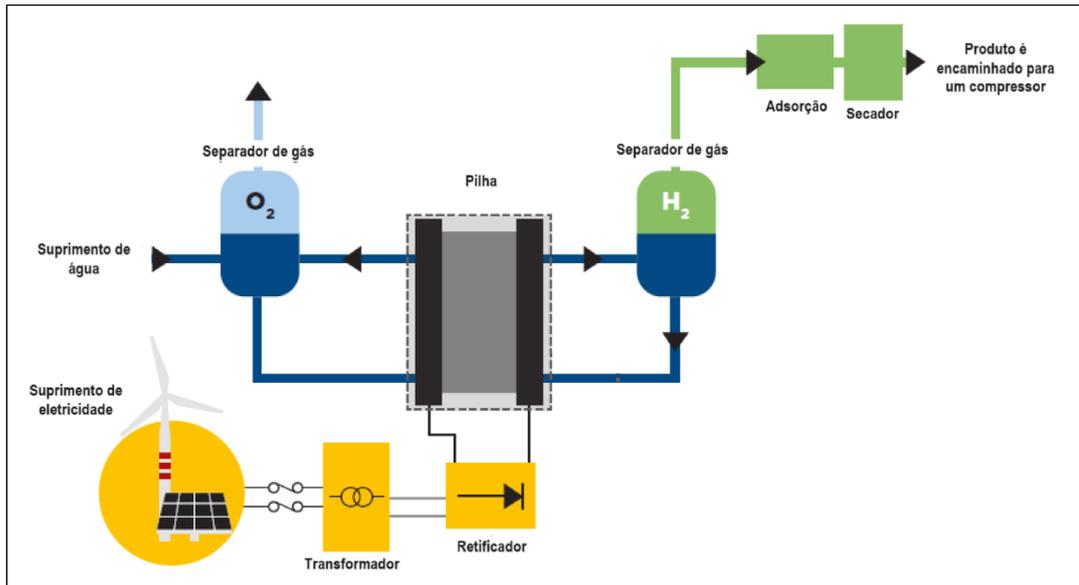
Com uma participação de 4% em termos de produção anual (El-Shafie, 2023), a eletrólise de água é um método que produz hidrogênio de alta pureza e é eletrointensivo, pois não ocorre de forma espontânea em temperaturas abaixo de 2.250 C° (Flamm *et al.*, 2021; Grigoriev; Fateev, 2023; Martinez Lopez *et al.*, 2023).

Os eletrolisadores são os equipamentos utilizados no processo eletrolítico e têm como elemento principal a célula, que é onde ocorre o processo eletroquímico propriamente dito. A célula é composta por dois eletrodos imersos em um eletrólito líquido ou adjacentes a uma membrana sólida, por duas camadas porosas de transporte e por placas bipolares. A pilha inclui múltiplas células conectadas em série, além de material isolante, vedação e moldura (IRENA, 2020). Além da pilha, uma planta de eletrólise de água inclui equipamentos para resfriamento, processamento do hidrogênio (para pureza e compressão), conversão da entrada de eletricidade (transformador e retificador), tratamento do abastecimento de água e saída de gás.

Resumidamente, o processo se inicia quando a água pura é injetada no sistema através de bombas e flui pelas camadas porosas de transporte até alcançar o eletrodo. No eletrodo, uma corrente elétrica é injetada para que ocorra a fragmentação da água em oxigênio e hidrogênio. Na prática, como a separação não é completa, há impurezas que são tratadas nos

respectivos separadores de gás, conforme mostrado na Figura 1.

Figura 1 – Componentes básicos de uma planta de eletrólise de água sob a perspectiva da planta



Fonte: Adaptado de IRENA (2020).

Após apresentadas as principais definições, será explanado a seguir um panorama da evolução científica na modelagem de Sistema de Conversão Eólica-Hidrogênio (SCEH2). Para a conversão vento-eletricidade, parte da literatura tem utilizado a Equação (2.1) — denominada curva de potência cúbica aproximada — para o cálculo da Potência Elétrica ($P(v)$), em Watts (W), gerada por um aerogerador (Carrillo *et al.*, 2013; Duca, 2020):

$$P(v) = \frac{1}{2} C_{p,eq} \rho A v^3, \quad (2.1)$$

sendo $C_{p,eq}$ uma constante equivalente ao coeficiente de potência, ρ representa a densidade do ar e é convencionalmente igual a $1,255 \text{ (kg/m}^3\text{)}$, A é a área de varredura do rotor em (m^2) e v é a velocidade do vento (m/s) .

A partir desse valor de $P(v)$, em um segundo estágio, a Equação (2.2) é utilizada para estimativa da produção de hidrogênio (Alavi *et al.*, 2016; Mostafaeipour *et al.*, 2019):

$$H_2 = \frac{\eta \times P(v)}{E_{cons}}, \quad (2.2)$$

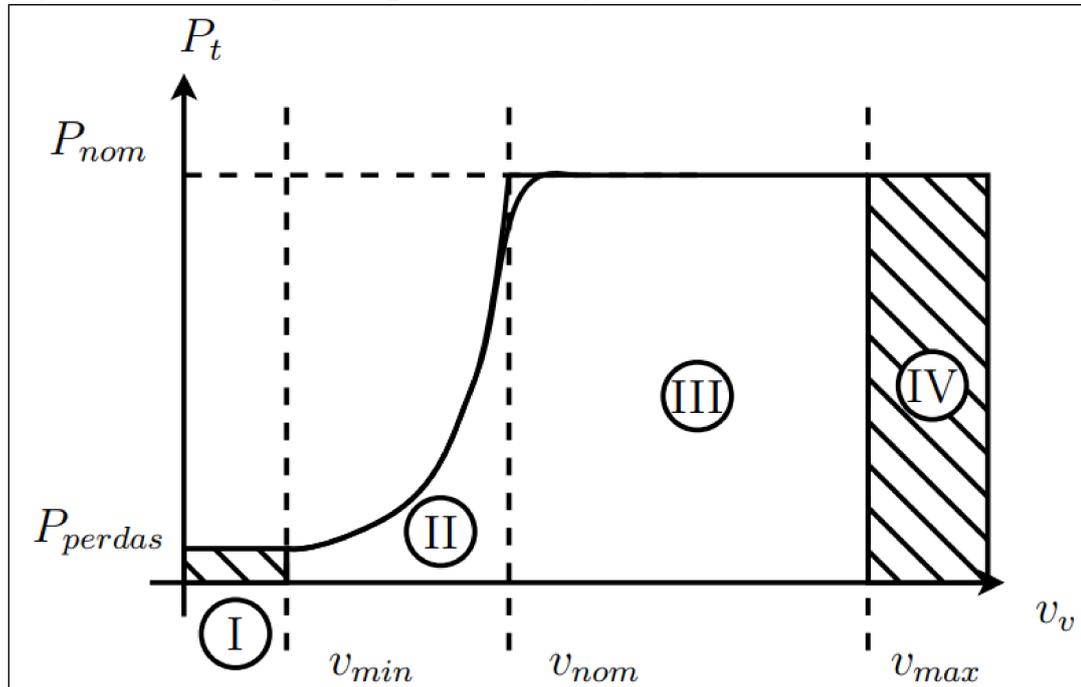
em que η é a eficiência do processo de eletrólise e E_{cons} o consumo de eletricidade do eletrolisador.

Um exemplo disso é o estudo realizado por Javaid *et al.* (2022) em que os autores avaliaram a viabilidade da produção de hidrogênio a partir da geração eólicoelétrica para localidades suburbanas. O cálculo da potência elétrica gerada pelos aerogeradores é inteiramente

baseado na Equação (2.1) e a estimativa da produção de hidrogênio é feita por um cálculo análogo ao da Equação (2.2), mas com a troca do parâmetro E_{cons} pelo poder calorífico mais baixo do hidrogênio. De toda forma, a eficiência do eletrolisador é considerada constante.

Essa abordagem apresenta, todavia, algumas limitações do ponto de vista técnico para a representação da operação real desses dois sistemas de conversão. No que diz respeito à conversão vento-potência elétrica, a Figura 2 ilustra um comportamento operacional típico de aerogeradores.

Figura 2 – Curva de potência padrão de aerogeradores



Fonte: Dobbin (2014).

As regiões com área tracejada (I e IV) na Figura 2 são consideradas críticas, uma vez que a conversão em energia ocorre para valores acima da velocidade mínima (v_{min}) e abaixo da velocidade máxima (v_{max}). Na região III, está a faixa da potência nominal, onde a potência é constante para valores entre a velocidade nominal (v_{nom}) e a velocidade máxima (v_{max}). Já região II é a faixa em a relação vento-potência é cúbica e, portanto, representa o intervalo de operação em que a Equação (2.1) pode ser coerentemente aplicada. A principal informação extraída da Figura 2 é a de que nem toda velocidade do vento é capaz de acionar o aerogerador

para produzir eletricidade. Matematicamente, essa relação é sintetizada pela Equação (2.3).

$$P(v) = \begin{cases} 0, & \text{se } v < v_{min} \text{ ou } v > v_{max} \\ q(v), & \text{se } v_{min} \leq v < v_{nom} \\ P_{nom}, & \text{se } v_{nom} \leq v \leq v_{max}, \end{cases} \quad (2.3)$$

onde $q(v)$ é uma função que descreve a curva na região II e P_{nom} é a potência nominal do aerogerador.

Com maior rigor técnico, pode-se ressaltar ainda que há um campo de conhecimento na literatura de sistemas de conversão vento-potência dedicado ao estudo de equações matemáticas para a representação da curva de potência. Em Carrillo *et al.* (2013), é realizada uma revisão que, além da Equação (2.1) — denominada de curva de potência cúbica —, inclui a curva de potência polinomial, a curva de potência exponencial e a curva de potência cúbica aproximada. A representação matemática dessas três últimas é dada, respectivamente, pelas Equações (2.4), (2.5) e (2.6).

$$q(v) = C_1 + C_2 v + C_3 v^2, \quad (2.4)$$

$$q(v) = \frac{1}{2} \rho A K_p (v^\beta - v_{min}^\beta), \quad (2.5)$$

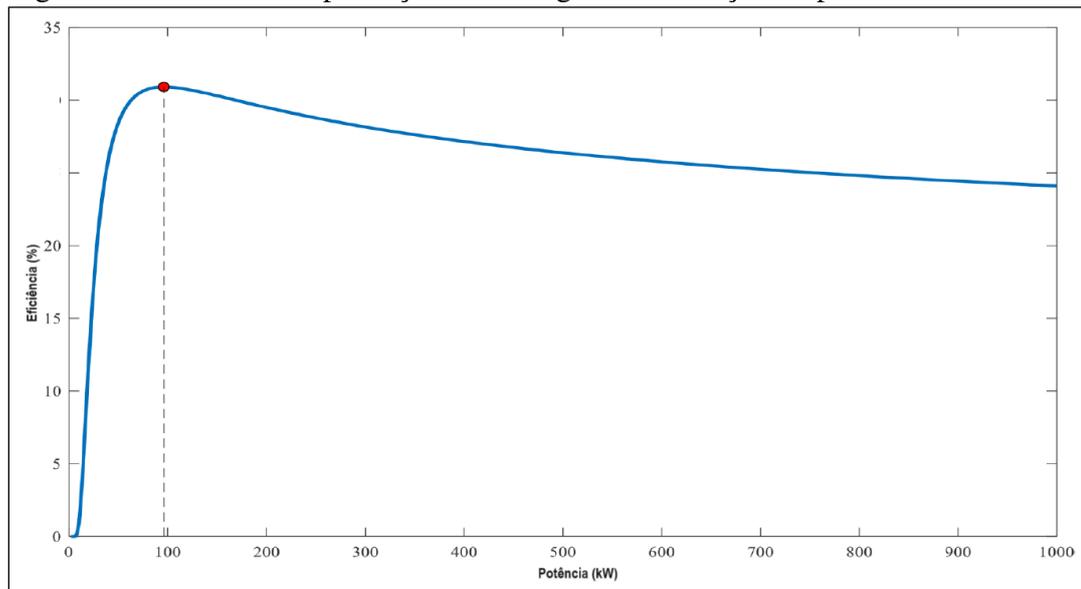
$$q(v) = \frac{1}{2} C_{p,max} \rho A v^3, \quad (2.6)$$

onde C_1 , C_2 e C_3 são coeficientes calculados a partir da v_{min} , da P_{nom} e da v_{nom} ; e $C_{p,max}$ é o coeficiente de potência máximo ou limite de Betz cujo valor é 0,593.

No que concerne à conversão eólica-hidrogênio, Martinez Lopez *et al.* (2023) destacam que a resposta das pilhas de um eletrolisador depende de parâmetros, como o fluxo de água, a temperatura e presença de bolhas, que mudam com o tempo. Por esse motivo, uma modelagem mais simples é tida como inadequada para representação de sistemas reais por ignorar o comportamento dinâmico em resposta à variação da potência de entrada.

Na Figura 3, o comportamento da eficiência do eletrolisador em resposta à potência de entrada fornecida por aerogeradores é ilustrado. Observa-se que, à medida que a potência de produção de hidrogênio aumenta de 0 para 100 kW, a eficiência do eletrolisador aumenta rapidamente, atingindo a eficiência máxima da produção de hidrogênio de cerca de 31% quando a potência atinge o valor de 100 kW. A partir desse patamar, a eficiência diminui gradualmente para 24,11% mesmo com aumento da potência fornecida. A partir da Figura 3, fica claro que o

Figura 3 – Eficiência da produção de hidrogênio em função da potência de entrada



Fonte: Adaptado de Hong *et al.* (2022).

processo de conversão de uma fonte renovável para hidrogênio caracteriza-se como um processo não linear, o que demonstra a limitação da Equação 2.2 em considerar a eficiência constante.

Nessa perspectiva, a literatura que trata de SCEH2 vem avançando na proposição de métodos de conversão de energia eólicoelétrica em hidrogênio que aproximem o cálculo teórico da dinâmica operacional desse tipo de sistema. Em Zhu *et al.* (2022), por exemplo, um modelo de eletrolisador em forma de circuito é construído com base na teoria eletroquímica. A partir desse modelo, a curva de eficiência do equipamento de produção de hidrogênio é derivada e linearizada.

Em Mazzeo *et al.* (2022), é apresentada uma metodologia para comparação energética entre sistemas de geração de eletricidade pelas fontes eólica e solar (sejam isolados ou híbridos) em um contexto de produção de hidrogênio. Tal comparação considera três cenários em que a rede elétrica interage com esses sistemas a fim de equalizar a relação entre a potência gerada pela fonte primária e a potência requerida pelo eletrolisador.

Dinh *et al.* (2021) desenvolvem um modelo para avaliação da viabilidade da produção de hidrogênio a partir de Usinas Eólicoelétricas (UEEs) instaladas em alto mar sem conexão com a rede elétrica. Ainda que considere uma eficiência constante para o eletrolisador, tal modelo dinamiza a estimativa da produção real de hidrogênio através de uma estrutura condicional que considera como parâmetros as potências máxima e mínima a serem fornecidas pelos aerogeradores, a eletricidade consumida por unidade de hidrogênio produzido e, inclusive, a eletricidade consumida para purificação de água, compressão do hidrogênio e por perdas ancilares.

A partir de algumas características operacionais de uma planta de hidrogênio em operação, Hofrichter *et al.* (2023) propõem um cálculo condicional da eficiência horária do eletrolisador em função do nível de potência utilizada para alimentar o mesmo. Especificamente, os autores consideram que na operação abaixo de 15% da potência nominal do eletrolisador o cálculo é baseado em uma função polinomial de quinta ordem e na operação acima de 15% essa função se torna linear. Na modelagem realizada por Ma *et al.* (2023), além da consideração da curva de potência para geração eólicoelétrica, os autores modelam a potência do eletrolisador por meio de uma função quadrática dependente da taxa do fluxo de massa de hidrogênio.

Em Hong *et al.* (2022), os autores idealizam uma estratégia sofisticada para melhorar o planejamento operacional de uma SCEH2. O parâmetro chave dessa estratégia é a otimização da eficiência do eletrolisador através de controle *fuzzy* segmentado. Zheng *et al.* (2023) propuseram uma estrutura de decisão baseada em um método robusto de otimização orientado por dados a fim de otimizar a operação diária de sistemas de conversão eólica-hidrogênio. Tal estrutura trata as incertezas inerentes às previsões de energia eólicoelétrica e de preço tendo a aversão ao risco como fator limitante.

2.2 Influência da flutuação da fonte primária no desempenho de eletrolisadores

Considerando a perspectiva de que a conversão da produção eólicoelétrica em hidrogênio não é linear, torna-se necessário debater também sobre os efeitos empíricos ocasionados pela flutuação da fonte primária sobre o desempenho de eletrolisadores.

Kojima *et al.* (2023), por exemplo, conduziram uma robusta revisão das características intrínsecas às fontes eólica e solar fotovoltaica e suas implicações na operação de sistemas de conversão de energia renovável em hidrogênio. Os autores realizam uma comparação entre tecnologias de eletrolisadores e destacam que os eletrolisadores com membrana de troca de prótons são, tecnologicamente, mais adequados para operarem em combinação com fontes de energia flutuantes do que os alcalinos. O estudo ainda destaca que a operação intermitente por si só pode levar à degradação dos componentes de um eletrolisador, pois gera forte influência sobre a temperatura da pilha/célula, pressão do gás, pureza do gás, além da própria quantidade de gás hidrogênio gerado. Em particular, a potência flutuante causa degradação do eletrodo devido às mudanças repentinas no potencial do mesmo.

O gás de cruzamento — produto da eletrólise incompleta — pode atingir altas concentrações em baixos valores de eletricidade e, em ocasiões extremas em que a potência de

entrada é inferior a determinado limite definido pelo fabricante, existe o risco da proporção de hidrogênio e oxigênio excederem o limite de explosão (Hong *et al.* 2022).

Em Martinez Lopez *et al.* (2023), é realizada uma revisão sobre a dinâmica operacional das três principais tecnologias de eletrolisadores de baixa temperatura – sabidamente os alcalinos, os de membrana trocadora de prótons e os de membrana trocadora de ânions – e a sua respectiva influência na integração destes com sistemas fotovoltaicos. Os autores destacam que a resposta elétrica dos eletrolisadores é muito rápida, o que os torna mais suscetíveis às limitações operacionais causadas por variáveis de resposta mais lenta – a exemplo da temperatura e da pressão – quando são alimentados eletricamente por plantas fotovoltaicas.

Haoran *et al.* (2023) investigam eventuais problemas de segurança e de eficiência para eletrolisadores alcalinos de grande porte conectados a sistemas de energia renovável. Na avaliação experimental, os autores identificam que esse tipo de eletrolisador, ao operar com potência de entrada flutuante, apresenta diferenças no estado de operação das células eletrolíticas e empiricamente nas áreas superficiais das próprias células. Esse fenômeno foi intitulado pelos autores como inconsistência e pode provocar a degradação e, portanto, encurtar a vida útil do eletrolisador. Além da inconsistência, a dinâmica da eficiência, a flexibilidade de regulação de potência e a pureza do gás são outros aspectos investigados na pesquisa.

Dobó e Palotás (2017) estudaram a relação entre flutuações de corrente e eficiência eletrolítica do AEL e verificaram que a aplicação de uma corrente contínua instável através do AEL resulta em uma perda de eficiência eletrolítica em comparação com condições de corrente contínua estável na mesma densidade de corrente.

Schnuelle *et al.* (2020) realizaram uma avaliação aprofundada da produção de hidrogênio sob entradas de energia flutuantes. O estudo mostra que indicadores-chave de desempenho, como eficiência de produção de hidrogênio, utilização de energia, rendimento de gás e custo líquido de produção, são significativamente afetados por padrões transitórios de mudança de energia e resposta flexível de eletrolisadores.

Enquanto os estudos citados acima têm como objeto de pesquisa apontar as implicações da flutuação da fonte primária na operação de eletrolisadores, outros estudos concentram esforços na perspectiva de propor meios práticos para atenuar esses problemas. Ganeshan *et al.* (2016), por exemplo, conduziram experimentos em laboratório e em campo para testar o funcionamento de um conversor redutor para manter constante a corrente contínua de um eletrolisador de 30 W conectado a um painel fotovoltaico de 68 W_p . Os resultados mostraram que a integração

desse conversor foi capaz de possibilitar uma taxa constante de produção de hidrogênio mesmo com um faixa de irradiância solar variando entre 440 e 975 W/m^2 .

Honsho *et al.* (2023) avaliaram a durabilidade de uma célula eletrolítica em face da flutuação de tensão causada pela geração eolioelétrica. Por meio de experimentos, os autores desenvolveram um protocolo de teste para padronizar a identificação da potencial flutuação dentro de um período de 24 horas. Os resultados obtidos sugerem que tanto as perdas reversíveis quanto as irreversíveis resultantes da degradação das camadas do catalisador anódico podem ser mitigadas pela incorporação de intervalos de descanso durante a operação.

Em De Battista *et al.* (2006), um novo sistema de controle é proposto para o cumprimento das especificações do eletrolisador quanto à taxa de variação da potência de entrada. Tal sistema é composto por um algoritmo de *Maximum Power Point Tracking* (MPPT) — que já é tradicionalmente utilizado em aerogeradores de velocidade variável — e por um circuito adicional para condicionamento da potência de referência. Os resultados das simulações demonstraram que o sistema proposto foi capaz de moldar adequadamente o perfil de potência, o que, por conseguinte, pode aprimorar a segurança no funcionamento do eletrolisador e melhorar as condições de produção de hidrogênio.

Outros estudos como o de Zheng *et al.* (2022), detalham o modo de operação dos eletrolisadores. Segundo os autores há três estágios: estágio de operação, estágio de espera e estágio de inatividade. Na mesma linha, Martinez Lopez *et al.* (2023) detalham que há dois mecanismos de acionamento de um eletrolisador. O primeiro deles é chamado de partida a frio, que ocorre quando o eletrolisador está desligado, frio e despressurizado. Nesse caso, o sistema precisa ser aquecido antes de começar a produzir hidrogênio. O segundo mecanismo é o arranque a quente, que ocorre em uma curta interrupção — análoga ao que Zheng *et al.* (2022) classifica como estágio de espera — em que a pressão e a temperatura podem ser mantidas e a produção pode continuar sem atrasos.

De modo geral, percebe-se que a literatura avança no sentido de compreender melhor as condições operacionais, sendo a capacidade de resposta rápida apontada como uma das características que requerem melhorias adicionais (Risbud *et al.*, 2023). É nesse contexto, por exemplo, que Qiu *et al.* (2022) apresentam um método fundamentado na inferência bayesiana para estimar parâmetros que contribuem para flexibilidade operacional de eletrolisadores alcalinos alimentados por fontes renováveis. Os resultados indicam que o erro na estimativa é reduzido em 71,1%, quando comparado com métodos *Extended Kalman Filter* (EKF) e *Unscented Kalman*

Filter (UKF). Como resultado, a eficiência computacional é aprimorada na ordem de 1000 vezes em relação a estimativa feita pela abordagem tradicional de Monte Carlo com Cadeias de Markov (MCCM).

Outro aspecto essencial deve ser considerado quando a produção de hidrogênio e fontes de energia flutuante são combinadas é uma estratégia de operação e de gestão de energia. Nesse sentido, destaca-se uma vertente da literatura sobre SCEH2 que potencializa a estratégia operacional desse tipo de sistema ao incorporar métodos de previsão de ventos em seu escopo. Alguns estudos pontuais tratam desse tema ao integrarem abordagens de previsão de ventos a estratégias de gerenciamento de geração (Cano *et al.*, 2015; Brka *et al.*, 2015; Brka *et al.*, 2016), ao usarem previsão de ventos baseada em heurística como uma estratégia operacional rentável (Grüger *et al.*, 2019), e ao incorporarem previsão de ventos por meio de técnicas de aprendizado de máquina para estimar a produção diária de hidrogênio (Javaid *et al.*, 2022).

A partir de uma análise convergente, infere-se que as duas linhas temáticas apresentadas acima apontam para um importante tópico de pesquisa que possui grandes valores acadêmicos e práticos para sistemas de produção de hidrogênio integrados às fontes renováveis. Tal tópico refere-se à previsão bayesiana da fonte primária.

2.3 Previsão probabilística via DLMs

Em um contexto de busca por técnicas avançadas de previsão de ventos, os DLMs têm se apresentado como uma alternativa bastante exequível e aplicável devido à incorporação da abordagem Bayesiana nesses modelos (García *et al.*, 2020; Duca *et al.*, 2021; Duca *et al.*, 2022). Com isso, a informação predita do conhecimento incerto atual passa a ser representado por distribuições de probabilidade (West; Harrison, 1997; Prado; West, 2010).

A base matemática da abordagem bayesiana origina-se do teorema de Bayes, que pressupõe que o processo de aprendizagem consiste no cálculo da probabilidade condicional de determinado evento de interesse, dado certa informação experimental (West; Harrison, 1997; Petris *et al.*, 2009; Gelman *et al.*, 2015). Tal teorema é simplificada expresso por:

$$p(\theta|Y) = \frac{p(\theta) \times L(Y|\theta)}{p(Y)}, \quad (2.7)$$

em que $p(\theta|Y)$ pode ser traduzido como a distribuição de probabilidade do evento θ acontecer dado um segundo evento Y . Já a componente $p(\theta)$ é a distribuição de probabilidade relacionada às observações que descrevem a ocorrência do evento θ , $L(Y|\theta)$ é a função de verossimilhança

que retrata a evidência Y quando o evento θ ocorreu e $p(Y)$ é uma constante de normalização.

A partir da Equação (2.7), fica evidente a pressuposição de que a incerteza sobre determinada variável pode ser representada, sob enfoque da inferência bayesiana, por duas funções de probabilidade tradicionalmente chamadas de *priori* ou *posteriori* — no caso, representadas por $p(\theta)$ e $p(\theta|Y)$, respectivamente.

Dessa forma, a estratégia de adotar os DLMS como ferramenta para previsão de ventos consiste em utilizar a estrutura probabilística que evolui ao longo do tempo a fim de garantir maior flexibilidade ao processo preditivo. Isso é especialmente relevante dada uma conjuntura em que a cadeia produtiva do hidrogênio depende de aspectos como a previsibilidade da fonte primária.

A fim de contemplar a formulação dos DLMS, considera-se uma série temporal representada pelo vetor Y_t de dimensão n ($Y_t \in \mathbb{R}^n$) e assume-se que os erros observacionais v_t e de evolução ω_t sejam mutuamente independentes no tempo, os DLMS são especificados pelas Equações (2.8) e (2.9), que são denominadas como Equação de Observação — relativa à série temporal observada — e de Evolução — que descreve como o vetor de estados latentes, θ_t ($\theta_t \in \mathbb{R}^p$), evolui ao longo do tempo —, respectivamente (West; Harrison, 1997; Gomes, 2018):

$$Y_t = F_t' \theta_t + v_t, \quad \text{onde: } v_t \sim \mathcal{N}(0, V_t), \quad (2.8)$$

$(n \times 1)$ $(n \times p)(p \times 1)$ $(n \times 1)$

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad \text{onde: } \omega_t \sim \mathcal{N}(0, W_t), \quad (2.9)$$

$(p \times 1)$ $(p \times p)(p \times 1)$ $(p \times 1)$

em que F_t' é a matriz de observação, θ_t é o vetor de estados latentes (ou não observados), G_t é a matriz de evolução dos estados e θ_{t-1} é o vetor de estados que representa as informações *a priori* (Gomes, 2018). É relevante salientar que os DLMS são comumente avaliados a partir da presunção de que o erro observacional (v_t) segue uma distribuição normal multivariada com média 0 e matriz de covariâncias V_t . Analogamente, também se assume que o erro de evolução (ω_t) segue uma distribuição normal com média 0, mas com uma matriz de evolução W_t . As dimensões matriciais n e p representam o número de observações e de parâmetros, respectivamente.

É oportuno observar que a Equação (2.8) guarda semelhança com a forma matemática dos modelos de regressão clássicos. Nestes modelos, a variável de resposta, Y , está relacionada a um conjunto de variáveis regressoras, F , a um conjunto de coeficientes de regressão, θ , — que pondera a influência que os regressores exercem sobre Y — e a um erro agregado, v . Essa

relação é representada pela Equação (2.10):

$$Y = \theta_0 + F_1 \theta_1 + \dots + F_p \theta_p + v. \quad (2.10)$$

Assumindo, agora, que Y_t denota um processo variante no tempo, a Equação (2.10) pode ser compactamente reescrita por:

$$Y_t = F_t' \theta + v_t, \quad (2.11)$$

em que

$$F_t = \begin{pmatrix} 1 \\ F_1 \\ \vdots \\ F_p \end{pmatrix} \quad \text{e} \quad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_p \end{pmatrix}.$$

No caso dos DLMS, no entanto, considera-se que todos os parâmetros mudam com o tempo e, dessa forma, θ passa a também ser indexado no tempo e sua evolução temporal passa a ser determinada pela Equação (2.9).

Após essa breve digressão por fins didáticos, pode-se retomar à descrição da especificação matemática dos DLMS para destacar que, em síntese, a formulação dos DLMS pressupõe a existência de um processo latente, ω_t — definido pela Equação (2.9) —, e um conjunto de observações, Y_t — definido pela Equação (2.8) —, em que ω_t quantifica as incertezas que geram a perda de informação com o passar do tempo e evolui a partir de alguma lógica. Em particular, W_t cria a dinâmica no vetor de estados e G_t controla a parte determinística da evolução do processo.

A definição dos DLMS também contempla o conjunto de informações disponíveis até o instante t e é representado por D_t e expresso como $D_t = (I_t, D_{t-1})$, sendo que I_t representa toda a informação adicional relevante obtida no tempo t (SOUZA, 2014). Nos casos em que a série temporal Y_t é a única fonte de informação, o sistema é tratado como fechado e o conjunto de informações disponíveis no tempo t passa a ser denotado por $D_t = (Y_t, D_{t-1})$. Assim, pode-se representar o conjunto das informações iniciais por D_0 , o que torna as Equações (2.8) e (2.9) implicitamente condicionais a D_{t-1} (Rodrigues, 2011).

Uma das características mais vantajosas dos DLMS é a disponibilidade das distribuições de suavização, filtragem e previsão, as quais estão disponíveis de forma fechada por meio do procedimento de estimação chamado Filtro de Kalman (FK) (King; Kowal, 2023). Para tanto, os parâmetros F_t' , G_t , W_t e V_t são assumidos conhecidos para todo tempo t .

Partindo dessa presunção e assumindo por convenção que o processo de previsão de uma nova observação é dado pela distribuição de probabilidades marginal de $(y_t | D_{t-1})$, a atualização dos parâmetros é obtida por meio da combinação da *priori* e a equação de verossimilhança, ambas no tempo t , conforme o teorema de Bayes. Em seguida, as distribuições condicionais do sistema $(\theta_s | D_t)$ são extraídas tomando como base as seguintes situações (Shumway; Stoffer, 2017):

- Quando $s < t$, o processo é chamado de suavização.
- Quando $s = t$, o processo é chamado de filtragem.
- Quando $s > t$, o processo é chamado de previsão.

O processo de suavização — conhecido também como análise retrospectiva — é utilizado com o propósito de analisar o comportamento de observações posteriores de determinado fenômeno (West; Harrison, 1997). Em uma aplicação desse recurso para EOL, por exemplo, seria possível comparar a eletricidade que foi efetivamente gerada com aquela que foi prevista em um momento anterior para delinear quais fenômenos impactaram o desempenho da usina avaliada. Contudo, por não ser objeto de estudo analisar retrospectivamente os dados, somente os processos de filtragem e previsão serão detalhados a seguir.

A fim de ilustrar visualmente uma aplicação de DLMs para uma série temporal, um exemplo didático apresentado em Petris *et al.* (2009) é utilizado. A Figura 4 mostra as medições do nível do rio Nilo em Ashwan entre 1871 a 1970, em $10^8 m^3$ que estão disponíveis no *software* estatístico R (R Core Team, 2022). É notório que, além de apresentar não estacionariedade, a série apresenta uma clara mudança de nível em torno do ano de 1900. Essa mudança ocorreu por conta do início da construção da barragem de Ashwan, que impactou drasticamente no fluxo do rio Nilo nessa região.

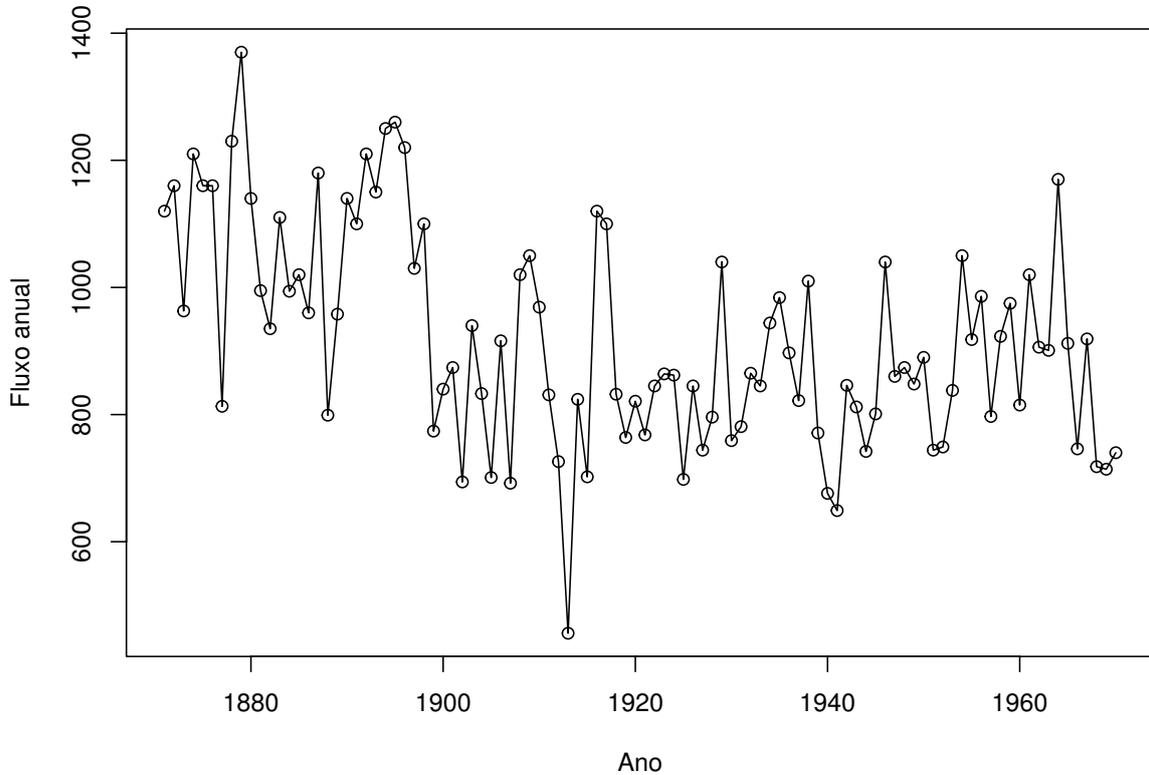
O processo de filtragem, amplamente difundido como FK, tem a finalidade de estimar iterativamente os parâmetros para a distribuição *a posteriori* do processo latente sempre que há uma nova observação. Considerando a especificação feita a partir das Equações (2.8) e (2.9) e assumindo uma informação atual denotada por

$$(\theta_{t-1} | y_{1:t-1}) \sim \mathcal{N}(m_{t-1}, C_{t-1}), \quad (2.12)$$

as seguintes relações passam a ser válidas (Petris *et al.*, 2009):

1. A distribuição preditiva um passo à frente de θ_t dado $y_{1:t-1}$ é gaussiana e possui os

Figura 4 – Medições do fluxo anual do rio Nilo no período de 1871 e 1970, em $10^8 m^3$



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

parâmetros:

$$a_t = E(\theta_t | y_{1:t-1}) = G_t m_{t-1} \quad e \quad (2.13)$$

$$R_t = Var(\theta_t | y_{1:t-1}) = G_t C_{t-1} G_t' + W_t.$$

2. A distribuição preditiva um passo à frente de Y_t dado $y_{1:t-1}$ é também gaussiana e possui os parâmetros:

$$f_t = E(Y_t | y_{1:t-1}) = F_t a_t \quad e \quad (2.14)$$

$$Q_t = Var(Y_t | y_{1:t-1}) = F_t R_t F_t' + V_t.$$

3. A distribuição preditiva um passo à frente de Y_t dado $y_{1:t-1}$ é também gaussiana e possui os parâmetros:

$$m_t = E(\theta_t | y_{1:t}) = a_t + R_t F_t' Q_t^{-1} \varepsilon_t \quad e \quad (2.15)$$

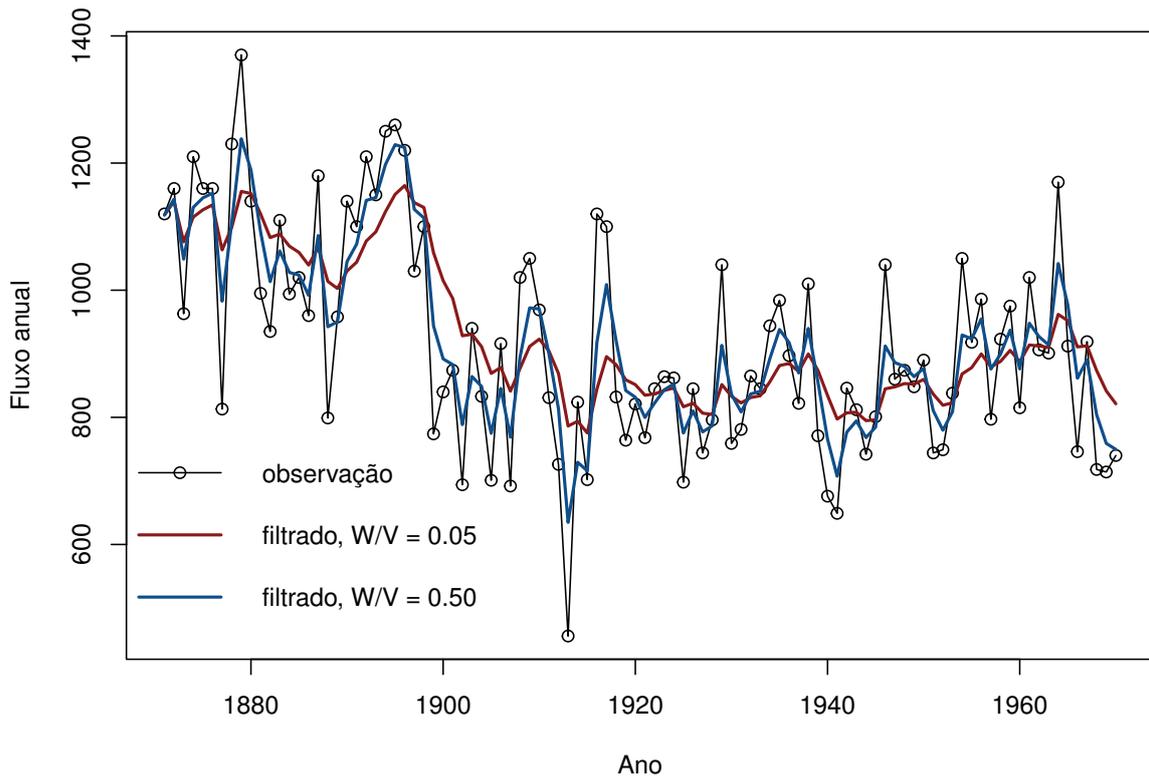
$$C_t = Var(\theta_t | y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t.$$

em que $\varepsilon_t = Y_t - f_t$.

Na Figura 5, o processo de filtragem é ilustrado por meio da aplicação para os dados do rio Nilo já mostrados na Figura 4. Com intuito de avaliar o melhor ajuste às curvas dos dados, Petris *et al.* (2009) utilizaram dois modelos de nível local com razões sinal-ruído distintas: W/V

= 0,05 e $W/V = 0,5$. Fica perceptível pela ilustração que modelo com a razão de $W/V = 0,5$ acompanha com mais precisão a série observada, o que justifica sua escolha para o processo seguinte de suavização.

Figura 5 – Valores filtrados do fluxo anual do rio Nilo para duas razões sinal-ruído diferentes



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

No processo de previsão via DLMs, as distribuições preditivas dos estados e das observações k passos à frente são sumarizadas pelas suas médias e variâncias por tratar-se de um processo gaussiano (Rodrigues, 2011; Souza, 2014). Partindo dessa premissa e considerando $a_t(0) = k_t$ e $R_t(0) = C_t$ para $k \geq 1$, as seguintes relações se tornam válidas (Rodrigues, 2011):

$$a_t(k) = G_{t+k}a_t(k-1), \quad (2.16)$$

$$R_t(k) = G_{t+k}R_t(k-1)G_{t+k}' + W_{t+k}, \quad (2.17)$$

$$f_t(k) = F_{t+k}a_t \text{ e} \quad (2.18)$$

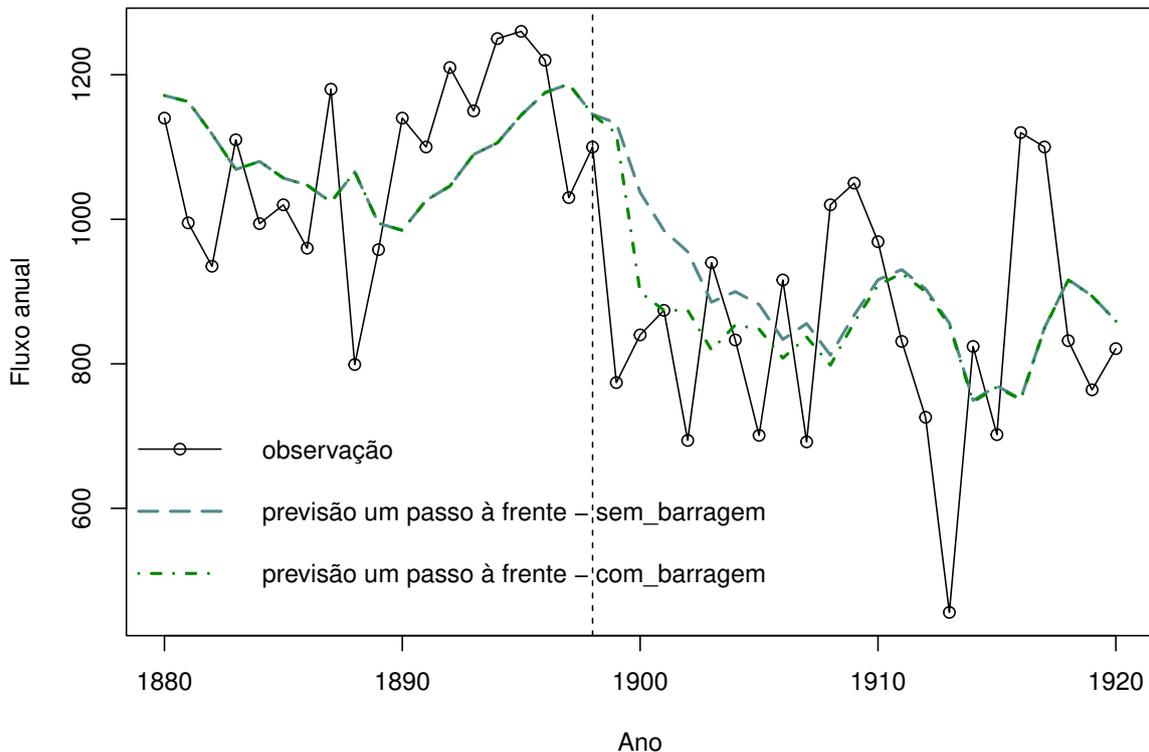
$$Q_t(k) = F_{t+k}R_t(k)F_{t+k}' + V_t, \quad (2.19)$$

em que $(\theta_{t+k} | D_t) \sim \mathcal{N}(a_t(k), R_t(k))$ e $(Y_{t+k} | D_t) \sim \mathcal{N}(f_t(k), Q_t(k))$. Rodrigues (2011) acentua que a Equação (2.18) é denominada como função de previsão e é ela que define a forma

da curva das previsões da série $Y_{t+1} \dots Y_{t+k}$.

A Figura 6 apresenta as previsões um passo à frente feitas por dois modelos para os dados do rio Nilo. A diferença básica entre eles é que apenas um tem a variância de evolução do sistema W_t readequada — com valor 12 vezes maior — para considerar o impacto da construção da barragem. Frente a essa possibilidade, Petris *et al.* (2009) concluíram que a importância em utilizar modelos preditivos mais flexíveis como os DLMs está na habilidade de incorporar as singularidades do fenômeno estudado.

Figura 6 – Previsões um passo à frente do fluxo anual do rio Nilo, considerando a construção da barragem



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

Em conformidade com o princípio da superposição de modelos, os DLMs tornam possível a modelagem de uma série temporal pela combinação de componentes básicas, como nível, tendência e sazonalidade — todas elas definidas por distribuições gaussianas (ou normais) — o que viabiliza o detalhamento estatístico da estrutura subjacente do processo que gerou os dados medidos (Mikkonen *et al.*, 2020). Apresenta-se a seguir a formulação dessas componentes com o objetivo de ilustrar como a estrutura descrita pelas Equações (2.8) e (2.9) pode caracterizar

uma grande variedade de modelos.

2.3.1 Modelo de Nível Local

O mais simples exemplo de um modelo dinâmico bayesiano aplicado para uma série temporal univariada Y_t é denominado modelo de nível local. Além desta, outras denominações encontradas na literatura são: modelo de passeio aleatório mais ruído, modelo de tendência estável e modelo polinomial de primeira ordem. Nesse caso, há apenas de um nível que varia de acordo com um passeio aleatório.

O nível permanece localmente constante, mas varia quando se considera longos períodos e, além disso, a variação das observações em torno dos níveis (medida por V) é bem maior que as variações temporais do nível ao longo do tempo (medidas por W). Para modelos de nível local que caracterizam séries univariadas, F_t' e G_t são escalares iguais a 1. Assim, esse tipo de modelo é definido por:

$$\begin{aligned} Y_t &= \mu_t + v_t, \quad \text{onde : } v_t \sim \mathcal{N}(0, V), \\ \mu_t &= \mu_{t-1} + \omega_t, \quad \text{onde : } \omega_t \sim \mathcal{N}(0, W), \end{aligned} \tag{2.20}$$

em que as sequências de erro v_t e ω_t são independentes, tanto internamente quanto entre elas. Seguindo com a notação adotada na Seção 2.3, a previsão k passos à frente para este modelo é representada pela seguinte distribuição preditiva:

$$(Y_{t+k} | D_t) \sim \mathcal{N}(m_t, Q_t(k)) \quad \text{para } k \geq 1, \tag{2.21}$$

em que $Q_t(k) = C_t + kW + V$ e a função de previsão, $f_t(k)$, nesse caso passa a ser rigorosamente igual a média calculada no tempo t , ou seja, $f_t(k) = m_t$.

2.3.2 Modelo com Tendência Linear

O modelo com tendência linear é um pouco mais sofisticado por incluir em sua estrutura dinâmica uma inclinação do nível. Com isso, o nível permanece localmente linear, mas a forma da reta pode variar com a passagem do tempo. O modelo com tendência linear é definido por:

$$\begin{aligned} Y_t &= \mu_t + v_t, \quad \text{onde : } v_t \sim \mathcal{N}(0, V), \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \omega_{t,1}, \quad \text{onde : } \omega_{t,1} \sim \mathcal{N}(0, \sigma_\mu^2), \\ \beta_t &= \beta_{t-1} + \omega_{t,2}, \quad \text{onde : } \omega_{t,2} \sim \mathcal{N}(0, \sigma_\beta^2), \end{aligned} \tag{2.22}$$

considerando que os erros v_t , $\omega_{t,1}$ e $\omega_{t,2}$ não sejam correlacionados. Este é um modelo dinâmico com

$$\theta_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix}, \quad G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad W = \begin{pmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{pmatrix} \text{ e } F' = \begin{pmatrix} 1 & 0 \end{pmatrix},$$

em que as variantes σ_μ^2 e σ_β^2 podem ser zero. A função de previsão k passos à frente para esse modelo é $f_t(k) = m_{1,t} + k \cdot m_{2,t}$.

2.3.3 Modelo com Sazonalidade representada por séries de Fourier

A sazonalidade denota a existência de um padrão cíclico que se repete a cada s períodos. Em séries temporais de velocidade do vento com medições feitas no Nordeste brasileiro, por exemplo, é comum encontrar dois tipos de comportamentos sazonais: um diário — em que as velocidades diminuem ao longo da manhã e aumentam até atingirem um valor máximo no início da noite — e outro anual — em que as velocidades são menores durante o primeiro semestre e maiores durante o segundo semestre, especialmente entre agosto e outubro (Gouveia, 2011). Por essa razão, há grande necessidade em usar técnicas de previsão que incorporem tais comportamentos.

A modelagem dinâmica bayesiana permite que séries temporais com a presença de componente sazonal sejam modeladas via representação de coeficientes de Fourier, que utiliza combinações lineares de harmônicos para tal finalidade (Pereira, 2017). Nesse sentido, utilizando como base nas deduções feitas em Souza (2014), considera-se uma função cíclica $g(t)$, que representa uma combinação linear de harmônicos (representados aqui por q) e está associada a um vetor de fatores sazonais $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_s)$, sendo s o período, que satisfaz a condição de $s > 1$. Assim, cada componente harmônica de $g(t)$ pode ser dada por:

$$S_q(t) = A_q \cos(\omega_q t + \gamma_q), \quad (2.23)$$

em que ω_q é a frequência e equivale a $(2\pi q)/s$, A_q é a amplitude e γ_q é a fase inicial. Considerando que o valor da função cosseno varia de -1 até 1 , pode-se desenvolver a Equação (2.23) da seguinte forma:

$$S_q(t) = A_q \cos \gamma_q \cos(\omega_q t) - A_q \sin \gamma_q \sin(\omega_q t) = a_q \cos(\omega_q t) + b_q \sin(\omega_q t), \quad (2.24)$$

sendo a_q e b_q denominados de coeficientes de Fourier. Após algumas transformações trigonomé-

tricas da Equação 2.24, a função $g(t)$ é finalmente dada por:

$$g(t) = a_0 + \sum_{q=1}^h [a_q \cos(\omega_q t) + b_q \sin(\omega_q t)], \quad (2.25)$$

considerando que o valor de h está condicionado ao fato de s ser um número par ou ímpar. Com isso, se s for ímpar, $h = (s - 1)/2$ e, se s for par, $h = s/2$.

Para que a evolução da componente harmônica ao longo do tempo possa ser avaliada, os valores de a_q e b_q devem ser conhecidos. Entretanto, na prática isso não ocorre e, assim, deve-se obter os valores de $S_q(t)$ e de seu conjugado, $S_q^*(t)$, para que a componente harmônica e seu conjugado no tempo $t + 1$ seja dada por:

$$\begin{aligned} S_q(t + 1) &= a_q \cos[\omega_q(t + 1)] + b_q \sin[\omega_q(t + 1)], \\ S_q^*(t + 1) &= -a_q \sin[\omega_q(t + 1)] + b_q \cos[\omega_q(t + 1)]. \end{aligned} \quad (2.26)$$

Finalmente, traduzindo essas relações para a forma matricial das Equações (2.8) e (2.9), obtém-se:

$$\begin{aligned} Y_t &= F_t' \begin{bmatrix} S_q(t + 1) \\ S_q^*(t + 1) \end{bmatrix} + v_t, \\ \underbrace{\begin{bmatrix} S_q(t + 1) \\ S_q^*(t + 1) \end{bmatrix}}_{\theta_t} &= \underbrace{\begin{bmatrix} \cos \omega_q & \sin \omega_q \\ -\sin \omega_q & \cos \omega_q \end{bmatrix}}_{G_t} \cdot \underbrace{\begin{bmatrix} S_q(t) \\ S_q^*(t) \end{bmatrix}}_{\theta_{t-1}}, \end{aligned} \quad (2.27)$$

em que $F_t = (1 \ 0)$. Por fim, a função de previsão k passos à frente desse modelo é dada por:

$$f_t(k) = \sum_{q=1}^h [m_{t,2q-1} \cos(\omega_q k) + m_{t,2q} \sin(\omega_q k)].$$

2.3.4 Estimativa de parâmetros desconhecidos

Até aqui, assumiu-se que a quádrupla $\langle F_t', G_t, W_t, V_t \rangle$ era conhecida a fim de facilitar o estudo do comportamento e das principais propriedades dos DLMS. Porém, em aplicações reais, dificilmente tais parâmetros são conhecidos, o que implica a não possibilidade de aplicação direta do FK.

Logo, adotando a notação de Petris *et al.* (2009), considera-se que a quádrupla seja dependente de um vetor de parâmetros desconhecidos denominado como ψ_t , que conserva a presunção de normalidade. A teoria dos DLMS permite que ψ_t seja estimado seja duas

perspectivas: a clássica — que pode utilizar uma função de máxima verossimilhança — e a bayesiana — que pode utilizar métodos de MCCM para prover uma simulação estocástica.

Por fins de economia computacional, somente a perspectiva clássica é usada na presente pesquisa para estimativa de ψ , que é realizada a partir da maximização da Equação 2.28 (Petris *et al.*, 2009):

$$l(\psi) = -\frac{1}{2} \sum_{t=1}^n \log |Q_t| - \frac{1}{2} \sum_{t=1}^n [(y_t - f_t)' Q_t^{-1} (y_t - f_t)], \quad (2.28)$$

Devido à impossibilidade de maximização analítica da Equação 2.28, métodos computacionais numéricos são utilizados para tal finalidade. Dentre eles, o algoritmo quasi-Newton Broyden-Fletcher-Goldfarb-Shanno (BFGS), que é baseado no método de Newton, é um dos mais usados para tal finalidade (ver Capítulo 7 de Durbin e Koopman (2012) e página 29 de Montenegro-Silva (2016) para maior detalhamento).

Por fim, após concluída a estimação dos parâmetros de interesse, o procedimento habitual de filtragem, suavização e previsão pode ser realizado.

2.4 Uso de métodos bayesianos e do filtro de Kalman na previsão de ventos

Apresentadas as características dos DLMS, faz-se necessária a realização de uma revisão para mapear a produção científica atual da utilização de métodos bayesianos e do FK na modelagem preditiva da velocidade do vento. Essa revisão foi realizada mediante consulta prévia na base de dados Scopus através da vinculação por operadores booleanos das seguintes palavras-chave: “*dynamic linear models*”, “*wind speed*”, “*state space models*”, “*kalman filter*”, “*dynamic bayesian*”, “*wind energy*” e “*bayesian forecasting*”.

Tais consultas foram realizadas em 22 de janeiro de 2024 e optou-se por não delimitar o horizonte temporal, mas o tipo de documento foi delimitado para artigos e artigos de revisão escritos na língua inglesa. A combinação das palavras-chave supracitada compôs a *string* de busca delimitada por título, resumo e palavras-chave e resultou em 293 documentos.

Após essa etapa de busca preliminar, o arquivo em formato *Bibtex* exportado do Scopus foi importado para o *software* RStudio para ser compilado, inclusive com a exclusão dos duplicados, a partir do pacote Bibliometrix da linguagem R em uma planilha eletrônica. Concluída essa etapa, restaram 280 documentos cujos títulos e resumos foram objeto de uma leitura exploratória que identificou uma gama de aplicações dos métodos bayesianos para além do escopo desta Dissertação.

De fato, esse resultado não foi surpreendente tendo em vista as robustas revisões realizadas por Li e Shi (2012) e Adedipe *et al.* (2020), que identificaram uma pluralidade de aplicações da inferência bayesiana em diversas áreas da indústria eólica como: avaliação de risco, diagnóstico/prognóstico de falhas, análise estrutural, planejamento de operação e manutenção e, naturalmente, previsão de ventos. Após uma restrição feita para estudos referentes à previsão de ventos, foi possível avaliar as contribuições dos métodos bayesianos e do FK nesse campo de estudo. Os 10 artigos mais recentes estão sumarizados na Tabela 1.

Tabela 1 – Os 10 artigos mais recentes com aplicação de métodos bayesianos e FK para previsão de ventos

Referência	Método(s)	Variável(is) de entrada	Métrica(s)
Dhakal <i>et al.</i> (2022)	WEB-RYM-ARIMA-FK-SVM-ANN	VV	[1], [2] e [3]
Duca <i>et al.</i> (2022)	DLGM-LGP-DTGM	VV e Pot	[1], [2] e [6]
Aly (2022)	ANFIS-RKF-WNN	Pot	[1] e [5]
Alanis <i>et al.</i> (2021)	RHONN-EKF	VV	[1], [2], [4] e [7]
Cai <i>et al.</i> (2021)	ARX-FK-GPR-NWP-UKF-APF	VV e dados de NWP	[1], [2] e [10]
Duca <i>et al.</i> (2021)	WGDM	VV	[1] e [8]
Lin <i>et al.</i> (2021)	CA-WD-FK	VV	[1], [2] e [3]
Aly (2020)	WNN-ANN-RKF-TS	VV e tempo	[1], [5] e [9]
García <i>et al.</i> (2020)	TMDLM	VV e DV	[3] e [4]
Zhao <i>et al.</i> (2020)	GP-UKF	VV	[1], [2], [3] e [11]

Fonte: elaborado pelo autor (2023).

Nota: [1] = *Mean Absolute Percentage Error* (MAPE), [2] = *Root Mean Square Error* (RMSE), [3] = *Mean Absolute Error* (MAE), [4] = *Mean Square Error* (MSE), [5] = *normalized Root Mean Squared Error* (nRMSE), [6] = *Interval Score*, [7] = Média, [8] = *Power Curve Error* (PCE), [9] = Coeficiente de Correlação de Pearson (CCP), [10] = Distribuição Residual e [11] = *Forecast Skill*.

Em Dhakal *et al.* (2022), é realizada uma análise comparativa do desempenho dos métodos Modelo baseado na Densidade de Probabilidade de Weibull (MDPW), Modelo baseado na Densidade de Probabilidade de Rayleigh (MDPR), ARIMA, FK, *Support Vector Machines* (SVM) e *Artificial Neural Networks* (ANN) para qual arranjo sequencial formaria o melhor modelo híbrido capaz de prever com precisão dados multivariados de velocidade do vento (VV) para regiões e horizontes temporais distintos. Os resultados obtidos sugeriram que o modelo de persistência foi o mais preciso para previsão de uma ou duas horas no futuro, um modelo baseado em Weibull performou melhor nos horizontes temporais de seis horas ou mais e um ARIMA ou modelo de aprendizado de máquina foi uma boa escolha para horizontes de previsão entre duas e seis horas. Na mesma linha, Aly testou e comparou diversas combinações entre métodos de previsão em Aly (2022) e em Aly (2020). Em ambas as pesquisas, os modelos híbridos que incorporaram o *Recurrent Kalman Filter* (RKF) obtiveram melhor desempenho.

Em Duca *et al.* (2022) foi proposta uma abordagem dinâmica não gaussiana para

modelar o comportamento estocástico de séries temporais de dados sintéticos de VV e de Potência (Pot). O cerne do estudo é a tentativa de eliminar a necessidade de transformações matemáticas para obtenção de normalidade, simetria e variância constante dos dados. Os resultados sugeriram que modelar o ano inteiro pode exigir abordagens híbridas para capturar características como assimetria, média não estacionária e variância em janelas temporais mais longas. Alanis *et al.* (2021) propõem uma modelagem preditiva que consiste em uma rede neural de alta ordem treinada online pelo algoritmo de EKF. A simplicidade de implementação, a baixa complexidade computacional e a operação em tempo real são diferenciais apontados em relação a outros métodos de previsão por inteligência artificial. Além da VV, a previsão de Pot gerada e preço da energia também é realizada.

Em Cai *et al.* (2021), dados resultantes de uma NWP são utilizados como entrada de um sistema da função de estado do filtro bayesiano para garantir a precisão da previsão de médio/longo prazo, porém os resultados mostraram que o método proposto produziu melhor desempenho de previsão em curto prazo. Duca *et al.* (2021) investigaram o desempenho de modelos de espaço de estados baseados nas distribuições de Weibull e Gamma para ajustar e prever a VV. Os resultados mostraram que tais modelos foram capazes de acomodar a não estacionariedade com boa eficiência computacional. As previsões foram obtidas por meio da geração de séries sintéticas com base em preceitos da inferência bayesiana.

Lin *et al.* (2021) utilizaram a *Wavelet Decomposition* (WD) para decompor os dados de VV em componentes de alta e baixa frequência. Também foi realizada uma Análise Caótica (AC) de cada componente. Além dessas duas técnicas de aprendizado de máquina, FK completou um modelo híbrido capaz de realizar previsões de curto prazo em escala diária e mensal. Em Zhao *et al.* (2020), o método Processo Gaussiano (PG) foi utilizado como função de transição não linear de um modelo de espaço de estado e a covariância obtida foi usada como o ruído do processo. Em seguida, a técnica do UKF foi usada para resolver o modelo de espaço de estado e atualizar a previsão inicial de curto prazo. García *et al.* (2020) construíram um *Truncated bivariate Matrix Bayesian Dynamic Linear Model* (TMDLM) para análise conjunta de componentes do vento e previsão de ventos de curto prazo. Os autores também inovam ao considerar observações de vento calmo como entrada. Os resultados mostram que a classe de DLMS proposta foi capaz de realizar previsões de curto prazo com maior acurácia para três locais nos meses de inverno e verão.

Em síntese, a maioria dos artigos reunidos na Tabela 1 recorrem a hibridização

para preencher lacunas dos modelos individuais. Curiosamente, boa parte dessas lacunas, a exemplo da não estacionariedade e da flexibilidade, já são potencialmente tratadas pelos DLMS. Notadamente, os dados de velocidade de vento predominam como dados de entrada. Sobretudo, não se identificou nenhuma aplicação na produção de hidrogênio.

3 METODOLOGIA

Apresenta-se a seguir o procedimento metodológico utilizado na presente pesquisa. Na Seção 3.1, são descritos os dados utilizados e das áreas que são objeto de estudo. Na Seção 3.2, o procedimento de implementação dos modelos é detalhado e também são apresentadas as métricas designadas para avaliação do desempenho. Por fim, a Seção 3.3 detalha os três métodos de conversão utilizados para a modelagem de SCEH2 a partir dos dados previstos.

3.1 Análise descritiva dos dados e das áreas de estudo

Para a execução da modelagem empregada nesta pesquisa, foram realizados previamente a coleta e o agrupamento de três séries temporais que reúnem registros históricos de velocidade do vento medidos em UEEs situadas nos estados da Bahia (BA), do Ceará (CE) e do Rio Grande do Sul (RS). A escolha dessa base de dados justifica-se pela sua confiabilidade, pois contém registros fornecidos pelas UEEs que operam no mercado regulado de energia (EPE, 2014).

Em decorrência de restrições relativas à confidencialidade das informações, tais registros foram obtidos junto à Empresa de Pesquisa Energética (EPE) por meio de procedimento administrativo em conformidade com a Lei nº 12.527/2011 (Lei de Acesso à Informação) (Brasil, 2011). Especificamente, os dados compreendem o período de três anos — nomeadamente, da 00:00h do dia 01/01/2019 até às 23:30h do dia 31/12/2021 — e, embora tenham sido medidos continuamente, estão integralizados a cada 30 minutos. Dessa forma, cada um dos três conjuntos de dados contém um total de 52.608 registros.

Após a coleta dos dados, realizou-se uma análise exploratória das três séries temporais e constatou-se a presença de dados anômalos valorados em “-99”. Estes valores foram excluídos e as células vazias foram preenchidas por meio de um procedimento de imputação denominado, por convenção, de sazonal. Na prática, essa imputação sazonal consistiu em preencher determinada célula vazia com a média dos respectivos valores correspondentes ao mesmo tempo e data dos outros anos. Assim, por exemplo, se a célula referente à 00:00h do dia 01 de janeiro de 2020 estiver vazia, ela deve ser preenchida pela média entre os valores referentes à 00:00h do dia 01 de janeiro de 2019 e à 00:00h do dia 01 de janeiro de 2021.

A Tabela 2 fornece informações estatísticas básicas, além da altura em que os dados foram medidos e da quantidade de dados imputados após o processo de imputação. Observa-se

na Tabela 2 que os conjuntos de dados do CE e do RS são medidos por anemômetros localizados a 120 metros do solo, enquanto o conjunto da BA refere-se às medições feitas na altura de 78 metros. Além de explicar o motivo pelo qual os valores de velocidade do vento da BA são menores em relação aos demais, essa informação da altura de medição é útil para a escolha dos aerogeradores utilizados na avaliação da produção de hidrogênio. Para o período considerado, a Tabela 2 também mostra que há presença de ventos calmos (0 m/s) nas três séries.

Outrossim, algumas informações presentes na Tabela 2 permitem notar que a série do CE tem algumas particularidades que a diferencia das demais séries. Primeiro, percebe-se que, para BA e RS, a média é maior que a mediana, o que indica que a assimetria das respectivas curvas de distribuição é positiva (ver Figuras 7c e 9c). O fato da diferença entre a média e a mediana ser menor para o CE também ajuda a explicar o maior grau de simetria dos dados do CE (ver Figura 8c) em relação aos outros. Outro destaque é que os dados do CE apresentaram um maior volume de anomalias e, portanto, é a série com maior quantidade de dados imputados (594), embora isso ainda seja pouco significativo frente à quantidade total (52.608).

Tabela 2 – Informações gerais sobre cada série temporal após a imputação dos dados

Estado	Altura da medição [m]	Mínimo [m/s]	Mediana [m/s]	Média [m/s]	Máximo [m/s]	Número de dados imputados	Percentual de dados imputados
BA	78	0,00	6,46	6,74	22,66	384	0,73%
CE	120	0,00	7,47	7,39	16,44	594	1,13%
RS	120	0,00	7,26	7,44	28,68	5	0,01%

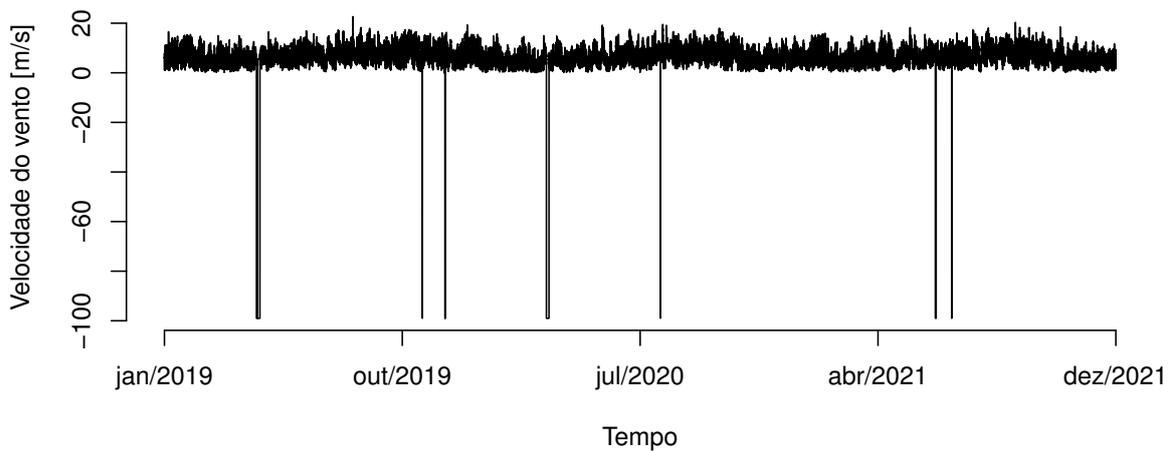
Fonte: elaborado pelo autor.

Quanto à delimitação para dados oriundos de três estados, a escolha justifica-se pela tentativa de estudar a previsão de ventos — assim como a modelagem do sistema de conversão eólica-hidrogênio — com dados reais de regiões com relevante potencial eólico, mas com diferentes características do regime de ventos. A seguir, serão apresentadas breves descrições dessas características.

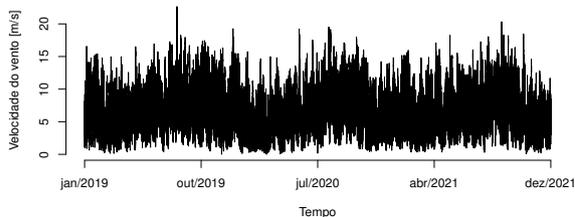
O estado da BA apresenta diferentes regimes de ventos resultantes da sobreposição de mecanismos atmosféricos globais — nomeadamente ventos alísios ao norte e ventos resultantes da interação do Anticiclone Subtropical do Atlântico com as ondas de massas polares ao sul — e regionais — nomeadamente brisas marinhas e terrestres e os jatos noturnos. Na região central do Estado — que concentra a maior quantidade das UEEs instaladas (ANEEL, 2017) —, o próprio deslocamento atmosférico e as formações montanhosas geram brisas montanha-vale com ciclos diários que tendem a aumentar de intensidade durante a noite (Santos *et al.*, 2013; Alessandra,

2017). A série temporal da BA é apresentada na Figura 7. A ideia central da Figura 7, assim como das Figuras 8 e 9, é ilustrar a diferenciação entre os dados originais (contendo valores anômalos) e os dados imputados.

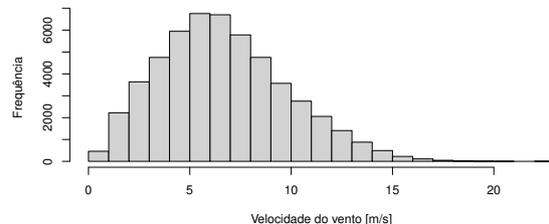
Figura 7 – Medições anemométricas realizadas em um parque eólico situado na Bahia (2019-2021)



(a) Dados originais



(b) Dados com imputação sazonal



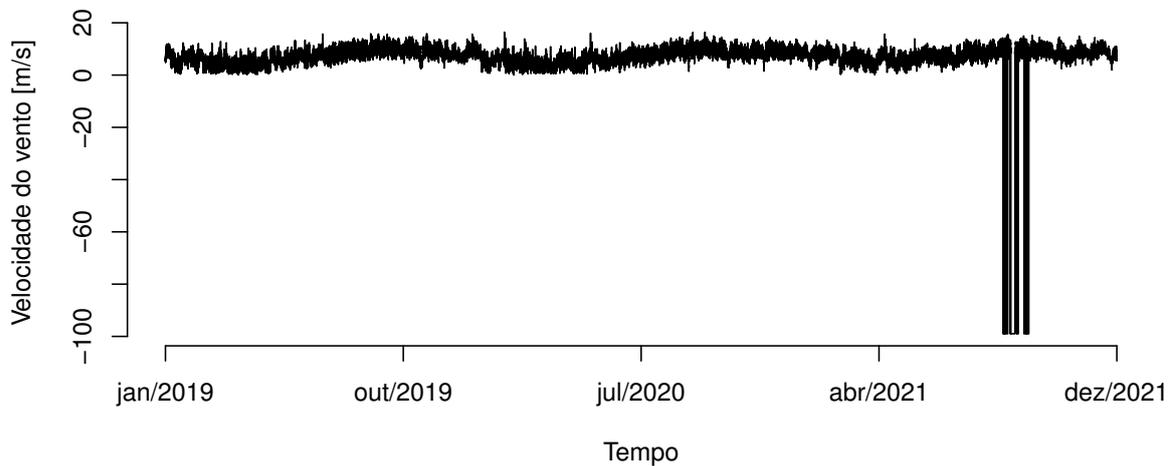
(c) Histograma

Fonte: Elaborado pelo autor (2024).

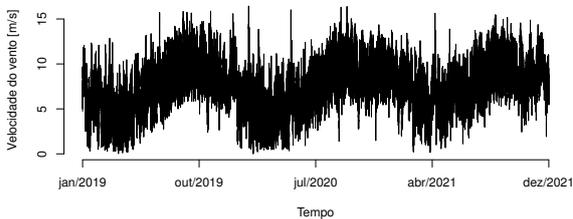
O CE está situado na zona litorânea Norte-Nordeste onde há a convergência do movimento das massas de ar oriundas dos hemisférios Norte e Sul que, por sua vez, originam os chamados ventos alísios caracterizados pela grande constância em intensidade e em direção. Ocorre, ainda, a intensificação desses ventos em decorrência das brisas marinhas nessa região gera anualmente ventos médios entre 6 e 9 m/s. De modo geral, o regime dos ventos no CE apresenta grande sazonalidade — como fica evidente pela Figura 8 —, em geral com a ocorrência de ventos mais intensos nos meses de final de inverno e primavera (agosto a novembro), e mais brandos nos meses de final de verão e outono (fevereiro a maio) (Alessandra, 2017; Mummey,

2017; ADECE, 2019).

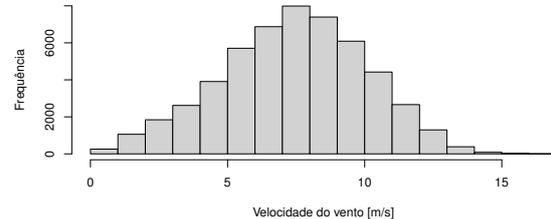
Figura 8 – Medições anemométricas realizadas em um parque eólico situado no Ceará (2019-2021)



(a) Dados originais



(b) Dados com imputação sazonal

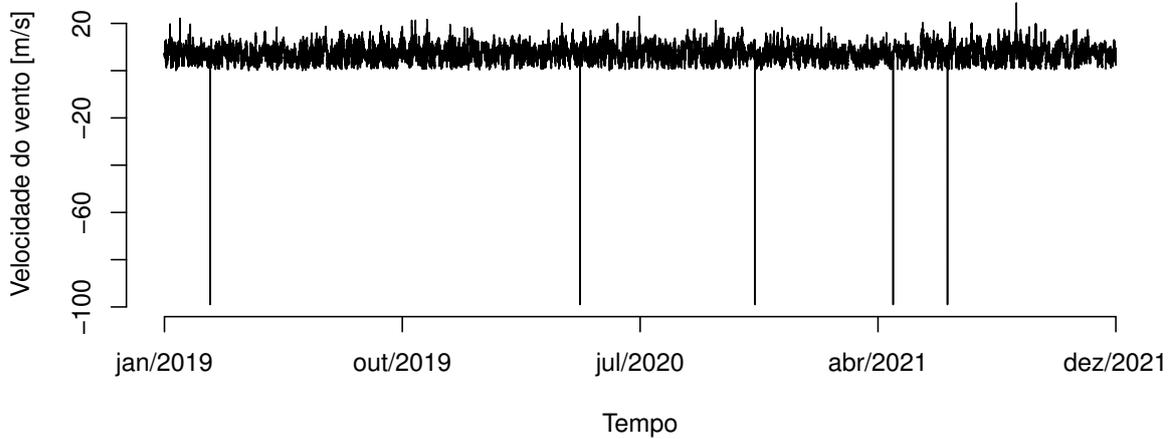


(c) Histograma

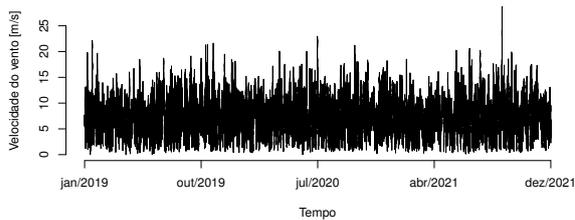
Fonte: Elaborado pelo autor (2024).

Já o Estado do RS possui uma climatologia bastante dinâmica por estar situado em uma zona de transição climatológica entre massas tropicais e massas de ar oriundas do polo Sul (Mummey, 2017). De modo geral, o litoral e o centro-oeste são as regiões do Estado que apresentam maior potencial eólico e, por consequência, as que concentram a maior parte dos empreendimentos eólicos (ANEEL, 2017). Na região litorânea, a maior intensidade dos ventos ocorre no fim de setembro até dezembro e no centro-oeste os ventos máximos ocorrem entre o fim de junho e o fim de setembro (Camargo Schubert; Eletrosul, 2014). Na Figura 9, é apresentada a série temporal do RS.

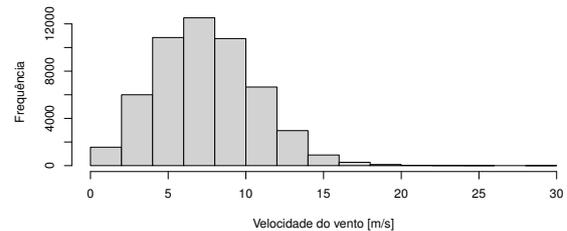
Figura 9 – Medições anemométricas realizadas em um parque eólico situado no Rio Grande do Sul (2019-2021)



(a) Dados originais



(b) Dados com imputação sazonal



(c) Histograma

Fonte: Elaborado pelo autor (2024).

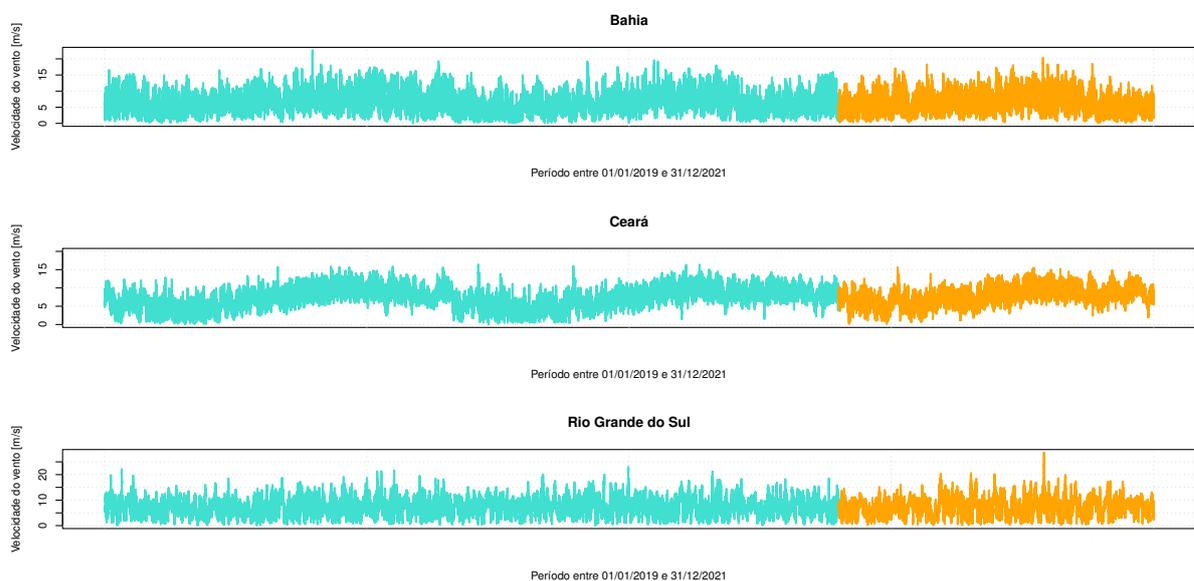
3.2 Metodologia de implementação dos DLMS

O ambiente experimental é baseado no sistema operacional Windows 11 em um computador portátil com processador Intel Core i5-10210U de 1,60GHz e 8 GB de RAM. Os algoritmos da modelagem são implementados na versão 4.1.3 do *software* estatístico livre R/RStudio (R Core Team, 2022) e na plataforma *Google Colab* com processamento em nuvem usando as bibliotecas *dlm* e *stats* da linguagem R. Assim, a metodologia da implementação da modelagem dinâmica bayesiana é, majoritariamente, baseada nas referências Petris *et al.* (2009) e Petris (2010).

Para a execução dos modelos, cada uma das três séries temporais foi dividida em dois subconjuntos: um para treinamento e para teste dos modelos de previsão. Os dados treinados consistem em um intervalo que compreende os valores entre a 1^a e a 36816^a observação —

da 00:00h do dia 01/01/2019 até as 23:30h do dia 05/02/2021 —, o que equivale à 70% da quantidade total de dados (52.608) com arredondamento para múltiplo de 48. Já o subconjunto de teste é constituído pelas observações posteriores ao subconjunto de treinamento, ou seja, está delimitado ao intervalo da 36.817^a até a 52.608^a observação — da 00:00h do dia 06/02/2021 até as 23:30h do dia 31/12/2021 —, o que equivale aos 30% restantes do total. Para fins ilustrativos, a Figura 10 apresenta os subconjuntos de treinamento e de teste para as três séries temporais estudadas.

Figura 10 – Subconjuntos de treinamento (azul) e teste (laranja) para as séries temporais da Bahia, Ceará e Rio Grande do Sul



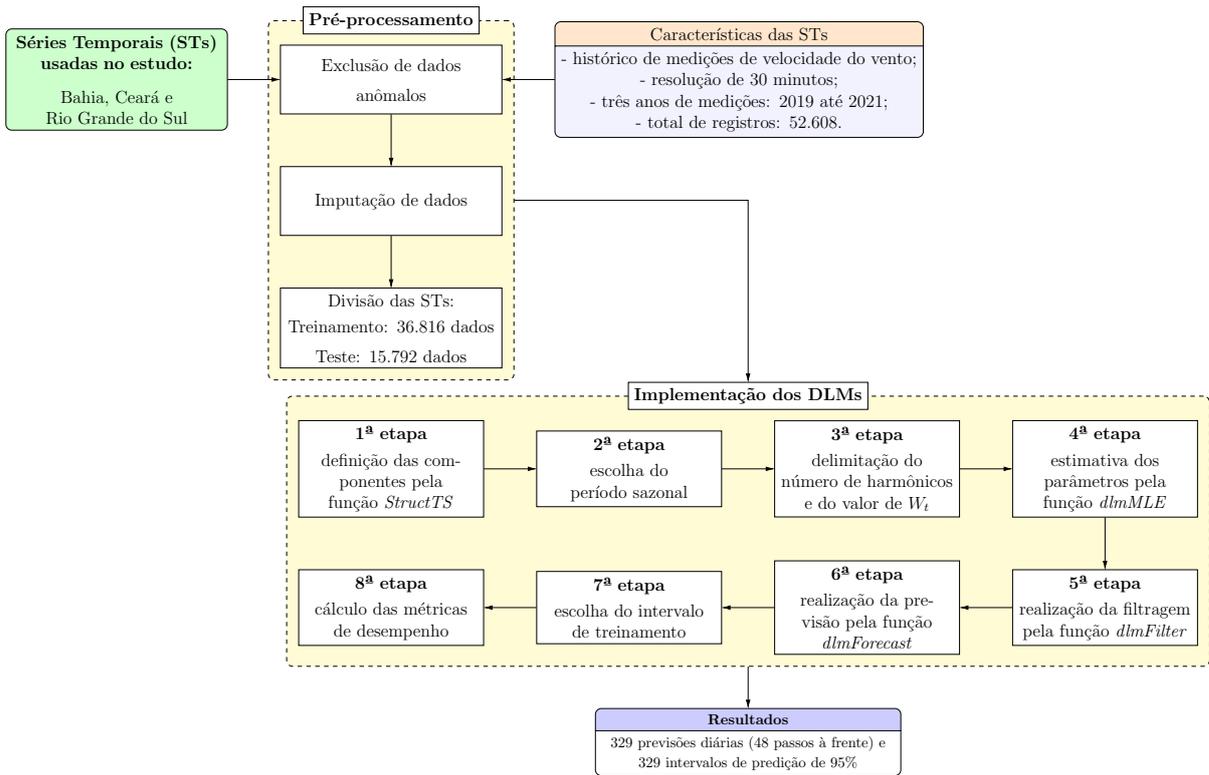
Fonte: Elaborado pelo autor (2024).

Todas as etapas de implementação dos DLMS, que serão descritas a seguir, estão sintetizadas na Figura 11.

A primeira etapa para implementação da modelagem bayesiana consistiu na definição dos blocos que integrariam os DLMS. Essa definição foi baseada na função *StructTS* da biblioteca *stats*, que, de acordo com Petris *et al.* (2009), permite a escolha das componentes de nível, de tendência e de sazonalidade a partir do tipo de argumento passado pelo usuário. Tal função utiliza o algoritmo *Limited-memory BFGS-boxed* (L-BFGS-B) para maximizar a função de verossimilhança e, por conseguinte, para estimar os parâmetros das componentes (Ripley, 2002).

Dado que, ao utilizar o argumento *type = "trend"* para os três conjuntos de dados, não se identificou inclinação (*slope*), somente as componentes de nível e de sazonalidade foram usadas para o ajuste dos modelos. Os resultados da função *StructTS* também permitiram

Figura 11 – Fluxograma da implementação dos DLMS



Fonte: Elaborado pelo autor (2024).

encontrar os valores das variâncias V_t e W_t através do argumento *type* = “level” para as respectivas componentes de nível.

A segunda etapa consistiu na escolha do melhor valor para o período (s) referente à componente de sazonalidade. Conforme descrito no início da Subseção 2.3.3, é possível que haja a coexistência de dois períodos sazonais em uma série temporal de vento. Considerando que as séries estudadas só possuem três anos de medições e, portanto, podem ser insuficientes para uma representação mais fidedigna de uma sazonalidade anual, somente a sazonalidade diária foi considerada. Como os dados são integralizados a cada 30 minutos, em um dia há um total de 48 dados e, portanto, a sazonalidade diária é representada por um período de 48.

Quanto ao número de harmônicos (q), a função *dlmModTrig()* delimita os valores de q até $s/2$. Ou seja, o valor máximo de q seria 24. Contudo, aumentar a quantidade de harmônicos de forma indiscriminada gera um sobreajuste (do inglês, “*overfitting*”) do modelo, pois os harmônicos tendem se ajustarem ao ruído da série e isso torna a representação da sazonalidade menos parcimoniosa (Barrio, 2022). Por essa razão, optou-se por delimitar o número de harmônicos até 10 e, dessa forma, aplicou-se uma variação de 1 até 10 harmônicos.

Além disso, adotou-se outra estratégia para melhorar o ajuste do modelo com base em um método análogo àquele utilizado por Petris *et al.* (2009) e que foi apresentado na Seção

2.3. Especificamente, um fator multiplicativo denominado de ζ , que variou de 10^{-5} até 10^5 em um passo de 10 em 10, foi aplicado ao valor da variância W_t do bloco de nível local. O plano de fundo dessa estratégia é a tentativa de otimizar o valor da variância da previsão Q_t , visto que, conforme às Equações 2.17 e 2.19, esse valor também depende de W_t .

Assim, os modelos foram construídos a partir de uma função que soma os comandos $dlmModPoly(order = 1, dV = \gamma, dW = \delta \times \zeta)$ e $dlmModTrig(s=48, q=\lambda)$. As variáveis γ , δ , ζ e λ são apenas ilustrativas em que γ e δ correspondem aos valores definidos na primeira etapa para V_t e W_t , respectivamente. Em síntese, a terceira etapa consiste em um processo iterativo em que, para cada um dos 11 valores de ζ , os dez valores possíveis de λ são avaliados para que o modelo mais acurado seja selecionado.

Na quarta etapa, a função $dlmMLE$ estimou os parâmetros dos modelos de acordo com a máxima verossimilhança. Essa função retorna sua declaração sobre a convergência das estimativas dos parâmetros, que, se for 0, indica que o algoritmo foi concluído com sucesso.

Na quinta etapa, o vetor de parâmetros, par , obtido da função $dlmMLE$ foi usado como argumento para construir o modelo. Em seguida, função $dlmFilter$ utilizou esse resultado como argumento para realização do processo de filtragem.

Na sexta etapa, a previsão 48 passos à frente foi, enfim, realizada pela função $dlmForecast$ a partir da lista resultante da filtragem. Importante destacar que a função $dlmForecast$ retorna uma lista contendo quatro componentes (duas matrizes e duas listas). Dentre as matrizes, uma delas é nomeada como f e refere-se aos valores previstos. Dentre as listas, uma delas é dedicada aos valores da variância Q .

Com isso, o cálculo dos limites superior (L_{sup}) e inferior (L_{inf}) do intervalo de predição é dado por:

$$L_{sup} = f + 2 \times \sqrt{Q}, \quad (3.1)$$

$$L_{inf} = f - 2 \times \sqrt{Q}, \quad (3.2)$$

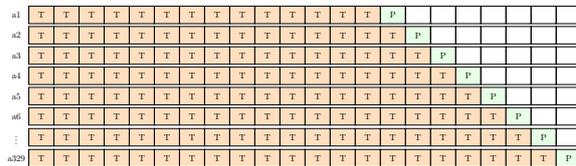
em que f é o dado previsto de velocidade do vento correspondente à componente f da função $dlmForecast$.

Na sétima etapa, o procedimento de previsão contínua foi adotado. Considerando que as previsões do dia seguinte são realizadas 48 passos à frente e que o subconjunto de teste contém 15.792 dados, foram realizadas 329 previsões diárias de forma sequencial. A ideia

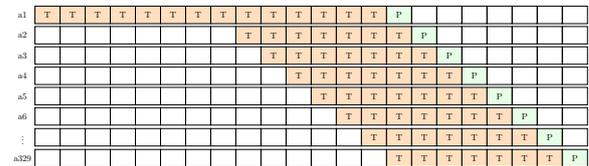
central para a adoção desse procedimento é emular a rotina de um sistema de gerenciamento operacional que contenha a previsão de dados em seu escopo.

A definição da janela de tempo ideal para o treinamento do modelo é um desafio relevante para previsões de ventos realizadas continuamente. Uma discussão mais aprofundada sobre a definição da janela de tempo dedicada ao treinamento para previsões contínuas pode ser encontrada em Liu e Wang (2021). Para a presente pesquisa, conjecturou-se inicialmente três procedimentos para definição da janela de tempo de treinamento. Tais procedimentos são ilustrados na Figura 12. Outros estudos abordam esse tema ao aplicarem o método de janela móvel (ou *windowing*) para previsão de fontes renováveis (Bezerra *et al.*, 2023) e ao evidenciarem a relação entre o erro de previsão e o tamanho dos dados (Liao *et al.*, 2023).

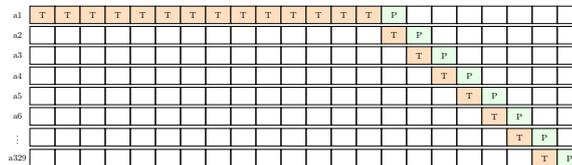
Figura 12 – Ilustração dos procedimentos de previsão contínua



(a) Janela de tempo expansiva.



(b) Janela de tempo semanalmente contínua.



(c) Janela de tempo diariamente contínua.

Fonte: Elaborado pelo autor (2024).

Nota: T = intervalo de treinamento. P = intervalo de previsão.

É relevante ressaltar que se tratam de esquemas meramente ilustrativos, visto que, na primeira amostra de teste (*a1*), os dados utilizados para treino são referentes a todo subconjunto de treinamento, ou seja, 36.816 dados, que correspondem a 767 dias. A partir da segunda amostra, o processo de treinamento é conduzido a partir da agregação de novos dados, conforme a quantidade de dados empregados para treinamento de cada procedimento.

Conforme ilustrado na Subfigura 12a, no procedimento em que a janela de tempo é expansiva, à medida em que novos dados são adquiridos, o treinamento é feito pela incorporação de todos os dados. Ao serem realizados testes com as 10 primeiras amostras, constatou-se a inviabilidade desse procedimento quanto ao custo computacional, pois cada treinamento demorou mais de uma hora para ser finalizada.

No caso dos procedimentos em que a janela de tempo utilizada para treinamento é semanal ou diária, o peso computacional não foi tão expressivo, ainda que seja importante destacar que a janela de tempo diariamente contínua tem a desvantagem de utilizar um horizonte de tempo muito reduzido. Uma comparação dos resultados para esses dois procedimentos é mostrada no Capítulo 4.

Na oitava e última etapa, as métricas de desempenho foram calculadas.

3.2.1 Métricas de desempenho

Dado que DLMS fornecem resultados de previsão pontual e probabilística, as métricas de avaliação utilizadas no presente estudo compreenderam métricas de previsão pontual e probabilística. A métrica usada para medir a acurácia do ajuste pontual dos modelos foi o nRMSE. Sua escolha justificou-se pela tentativa de evitar a dependência de escala e, assim, é possível comparar os resultados com outros modelos aplicados a séries temporais com dimensões distintas. Conceitualmente, o nRMSE indica a dispersão ou a variabilidade da precisão da previsão e é calculada por meio da Equação (3.3) (Bounoua *et al.*, 2021):

$$nRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_{obs} - f)^2}}{\overline{y_{obs}}}, \quad (3.3)$$

sendo n o tamanho da amostra (48, no caso), y_{obs} o dado observado de velocidade do vento e $\overline{y_{obs}}$ o valor médio dos dados observados. De modo geral, um bom modelo é aquele para o qual o valor de nRMSE é o mais baixo possível. Em Li *et al.* (2013), a acurácia do modelo medida pelo nRMSE é categorizada em quatro classes, conforme mostrado na Tabela 3.

Tabela 3 – Diferentes classes de acurácia do modelo preditivo medida pelo nRMSE

Classe	Valores de nRMSE (%)	Estado de acurácia
1	nRMSE < 10%	Excelente
2	10% < nRMSE < 20%	Bom
3	20% < nRMSE < 30%	Regular
4	nRMSE > 30%	Pobre

Fonte: Adaptado de Li *et al.* (2013).

A avaliação de previsões probabilísticas exclusivamente por métricas determinísticas pode levar a conclusões inválidas (van der Meer *et al.*, 2018). Dessa forma, os resultados dos DLMS também foram avaliados sob o viés probabilístico a partir de duas importantes propriedades: a confiabilidade e a precisão. Enquanto a confiabilidade avalia se os dados

observados podem ser efetivamente da distribuição de probabilidade prevista, a precisão avalia quão concentrada é essa distribuição (Mitrentsis; Lens, 2022).

A *Prediction Interval Coverage Probability* (PICP) foi usada como métrica da confiabilidade dos intervalos de predição e é expressa matematicamente pela Equação (3.4):

$$PICP = \left(\frac{1}{n} \sum_{i=1}^n c_i \right) \times 100\%, \quad \text{sendo } c_i = \begin{cases} 0, & \text{se } y_i^{obs} \notin (L_{sup}, L_{inf}) \\ 1, & \text{se } y_i^{obs} \in (L_{sup}, L_{inf}) \end{cases} \quad (3.4)$$

em que c_i é uma variável auxiliar. Deduz-se da Equação (3.4) que valores altos de PICP implicam que mais dados estão dentro dos limites intervalares de previsão, o que é evidentemente desejável (Khosravi *et al.*, 2013).

Contudo, a PICP por si só pode não ser um indicador suficiente para indicar se o desempenho do modelo preditivo é adequado, visto que intervalos muito largos se traduzem em altos valores de PICP, mas podem não conter informações significativas para o processo de tomada de decisão (He *et al.*, 2017). Nesse sentido, o *Prediction Interval Normalized Average Width* (PINAW) foi utilizado como métrica complementar para a avaliação da precisão dos intervalos de predição. O PINAW é calculado pela Equação (3.5) (Serrano-Guerreiro *et al.*, 2021):

$$PINAW = \left[\frac{1}{nR} \sum_{i=1}^n (L_{sup} - L_{inf}) \right] \times 100\%, \quad (3.5)$$

sendo que R denota a diferença entre os valores máximo e mínimo dos limites do intervalo de predição. Quanto menor for o valor de PINAW, mais adequado é o intervalo de predição.

3.3 Metodologia para estimativa da produção de hidrogênio

A etapa de modelagem do SCEH2 consistiu em utilizar os dados resultantes dos DLMs para estimar a produção de hidrogênio de um hipotético SCEH2. As estimativas fundamentaram-se em três abordagens que diferem entre si pelo modo com que o cálculo da eficiência do eletrolisador foi calibrado. Tanto o *Google Sheets* quanto o R/RStudio foram utilizados como plataformas de execução desta etapa. O detalhamento das três metodologias de calibragem é feito a seguir.

3.3.1 Método de conversão utilizado por Javaid *et al* (2022)

O primeiro método de conversão (C1) é utilizado em Javaid *et al.* (2022), no qual supõe-se um comportamento constante para a eficiência de um eletrolisador. Em C1, o cálculo

da potência elétrica gerada pelos aerogeradores é baseado na Equação (2.6) sem que a curva de potência seja considerada. Assim, a geração eólicoelétrica total, $E_{t,1}$, é calculada por meio da multiplicação da potência elétrica (dada pela Equação (2.1)) pelo número de aerogeradores, N_{AEG} , e por 0,5 — que é o componente temporal dado em horas, considerando a resolução de 30 minutos —, conforme a Equação (3.6):

$$E_{t,1}(v) = N_{AEG} \cdot P_1(v) \cdot 0.5. \quad (3.6)$$

Com o valor de $E_{t,1}$, a estimativa da produção de hidrogênio é feita por um cálculo análogo ao da Equação (2.2), em que o valor de Poder Calorífico Inferior do hidrogênio (PCI-H₂) é 33,3 kWh/kg e a eficiência do eletrolisador do C1 ($\eta_{ele,1}$) é fixada em 75%, conforme a Equação (3.7):

$$H_2^{C1} = \frac{\eta_{ele,1} \times E_{t,1}(v)}{33,3}, \quad (3.7)$$

3.3.2 Metodologia apresentada por Zhu et al (2022)

O C2 é baseado na metodologia apresentada em Zhu et al. (2022) para o comportamento variável da eficiência do eletrolisador. Os valores referentes a curva de linearização proposta pelos autores estão dispostos na Tabela 4. Os valores mostrados na coluna da faixa de potência são resultantes da relação entre a potência de entrada no eletrolisador e a potência nominal do eletrolisador.

Tabela 4 – Linearização trapezoidal da eficiência de um eletrolisador

Faixa de potência	Eficiência do eletrolisador ($\eta_{ele,2}$)
0-0,05	0%
0,05-0,2	82%
0,2-0,5	77%
0,5-1,0	73%
1,0-1,5	66%

Fonte: Adaptado de Zhu et al. (2022).

Com isso, o cálculo da geração eólicoelétrica, $E_{t,2}$, é baseado na Equação (2.3) para um $q(v)$ calculado pela Equação (2.6), o que resulta na Equação (3.8):

$$E_{t,2}(v) = N_{AEG} \cdot P_2(v) \cdot 0.5. \quad (3.8)$$

A partir disso, a estimativa da produção de hidrogênio é também calculada pela Equação (2.2), diferindo do C1 por utilizar uma eficiência do eletrolisador que varia segundo a

linearização apresentada na Tabela 4, conforme a Equação (3.9):

$$H_2^{C2} = \frac{\eta_{ele,2} \times E_{t,2}(v)}{33,3}. \quad (3.9)$$

3.3.3 Metodologia apresentada por Hofrichter et al. (2023)

O terceiro método de conversão (C3) é baseado na modelagem proposta por Hofrichter et al. (2023). De forma detalha, os autores consideram que, na operação com $P\%$ abaixo de 10%, a operação do eletrolisador é interrompida. Logo, $\eta_{ele,3}$ equivale a 0. Nas situações em que o valor de $P\%$ varia entre 10% e 15%, o valor de $\eta_{ele,3}$ é calculado através de uma função polinomial de quinta ordem. Para valores de $P\%$ acima de 15%, o valor de $\eta_{ele,3}$ é dado por uma função linear. Tais cenários são descritos pela Equação 3.10.

$$\eta_{ele,3} = \begin{cases} 0\%, & \text{se } P\% < 10\% \\ 100 + [0,00005 * (P\%)^5 - 0,0061 * (P\%)^4 + 0,2372 * (P\%)^3 - \\ 4,2014 * (P\%)^2 + 36,675 * P\% - 62,87], & \text{se } 10\% \leq P\% < 15\% \\ -0,149 * P\% + 74,977, & \text{se } 15\% \leq P\% \leq 100\% \\ 100\%, & \text{se } P\% > 100\% \end{cases} \quad (3.10)$$

Destaca-se que a Equação 3.10 possui três adaptações em relação a forma canônica disposta em Hofrichter et al. (2023):

- o fator de aumento anual da eficiência é desconsiderado, visto que o subconjunto de teste compreende um período inferior a um ano;
- o valor de 100 foi acrescentado para evitar incoerência nos valores resultantes da segunda condição; e
- em conformidade com Hofrichter et al. (2023), valores de potência superiores à potência nominal do eletrolisador foram desconsiderados. Ou seja, nos casos em que $P\% > 100\%$, somente a potência equivalente à potência nominal do eletrolisador é utilizada integralmente ($\eta_{ele,3} = 100\%$) para a estimativa da produção de hidrogênio.

Analogamente ao C2, no C3 a geração eolioelétrica, $E_{t,2}$, é calculada com base na Equação (2.3), resultando, assim, na Equação (3.11):

$$E_{t,3}(v) = N_{AEG} \cdot P_3(v) \cdot 0.5. \quad (3.11)$$

A partir disso, a estimativa da produção de hidrogênio, H_3^{C3} , é também calculada pela Equação (2.2), mas utilizando uma eficiência do eletrolisador que varia em conformidade com as condições vistas na Equação 3.10. Com isso, a H_3^{C3} é calculada pela Equação (3.12):

$$H_3^{C3} = \frac{\eta_{ele,3} \times E_{t,3}(V)}{33,3}. \quad (3.12)$$

3.3.4 Síntese dos métodos utilizados para a modelagem dos eletrolisadores

Considerando a informação apresentada na Tabela 2 sobre a altura de medição, os aerogeradores utilizados para o cálculo da geração eólicoelétrica foram selecionados de forma que a altura do *hub* fosse compatível com a altura de medição dos dados de vento. As principais informações técnicas dos aerogeradores estão dispostas na Tabela 5.

Tabela 5 – Informações técnicas dos aerogeradores selecionados

Informações	GE77/1500	AW125/3000
Fabricante	GE	Acciona
Potência nominal	1.500 kW	3.000 kW
Velocidade de corte inferior	3,5 m/s	3,5 m/s
Velocidade nominal	15 m/s	12 m/s
Velocidade de corte superior	25 m/s	25 m/s
Área de varredura do rotor	4.657 m ²	12.305 m ²
Altura do <i>hub</i>	80 m	120 m
Aplicação	BA	CE e RS

Fonte: Elaborado pelo autor (2024).

Os parâmetros utilizados nos três métodos de modelagem dos eletrolisadores estão sintetizados na Tabela 6.

Tabela 6 – Semelhanças e diferenças entre os métodos de modelagem do eletrolisador

Parâmetro	C1	C2	C3
$P_{nom,UEE}$	51 MW	51 MW	51 MW
N_{AEG}	BA: 34; CE e RS: 17	BA: 34; CE e RS: 17	BA: 34; CE e RS: 17
η_{AEG} (%)	BA: 40; CE e RS: 38,84	BA: 40; CE e RS: 38,84	BA: 40; CE e RS: 38,84
Curva de potência	Não	Sim	Sim
$P_{nom,ELE}$	40 MW	40 MW	40 MW
$PCI - H2$	33,3 kWh/kg	33,3 kWh/kg	33,3 kWh/kg
Modo de calibragem da η_{ELE}	Numérico	Teórico	Empírico

Fonte: Elaborado pelo autor (2024).

É perceptível que os valores de potência nominal das hipotéticas UEEs ($P_{nom,UEE}$), de N_{AEG} , de eficiência dos aerogeradores (η_{AEG}) e de potência nominal da hipotética planta

de eletrólise ($P_{nom,ELE}$) são iguais para todos, enquanto a curva de potência não é considerada apenas para o C1 e o tipo de calibragem da eficiência dos eletrolisadores difere para todos os métodos.

4 RESULTADOS E DISCUSSÃO

Este Capítulo está dividido em duas seções. A Seção 4.1 apresenta os resultados da previsão do vento e a Seção 4.2 mostra os resultados da modelagem do SCEH2.

4.1 Resultados da previsão do vento

Na Tabela 7, são comparados os resultados obtidos pela métrica $nRMSE$ para as 329 amostras de teste dos dados da BA ($nRMSE_{BA}$), do CE ($nRMSE_{CE}$) e do RS ($nRMSE_{RS}$) por meio dos procedimentos de janela de tempo diária e semanal.

Tabela 7 – Comparação dos resultados da métrica $nRMSE$ para as janelas de tempo semanal e diária

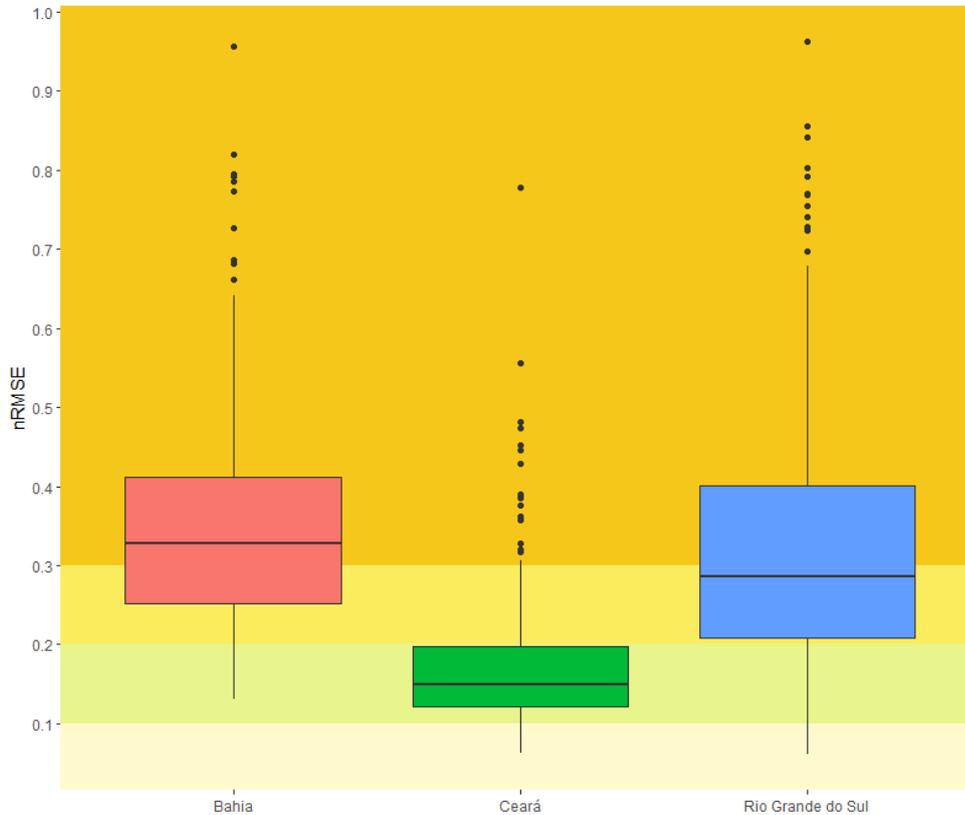
Parâmetro	Janela diária			Janela semanal		
	$nRMSE_{BA}$	$nRMSE_{CE}$	$nRMSE_{RS}$	$nRMSE_{BA}$	$nRMSE_{CE}$	$nRMSE_{RS}$
Mínimo	10,31%	7,56%	7,73%	13,04%	6,25%	6,08%
Mediano	35,77%	18,54%	33,57%	32,72%	15,01%	28,55%
Médio	36,24%	22,18%	40,95%	34,47%	17,31%	31,73%
Máximo	97,88%	96,37%	97,71%	95,69%	77,77%	96,27%
Desvio padrão	0,17	0,12	0,20	0,13	0,08	0,16

Fonte: Elaborado pelo autor (2024).

A principal informação extraída da comparação dos dados dispostos na Tabela 7 é que, de modo geral, os resultados foram mais acurados quando houve um aumento da janela de treinamento do horizonte temporal diário (contendo 48 dados) para o semanal (contendo 336 dados); com exceção ao valor mínimo encontrado para a BA. Conclui-se, portanto, que o aumento no volume de dados treinados melhorou a capacidade da maioria dos modelos para aprender o comportamento dos dados observados, o que é coerente com os resultados obtidos por Liao *et al.* (2023) para uma previsão probabilística baseada em algoritmos de aprendizado de máquina. Por essa razão, a janela semanal foi escolhida para a realização dos treinamentos a partir da segunda amostra de teste das séries temporais.

Na comparação entre as séries temporais na janela semanal, percebe-se que as previsões amostrais do CE apresentaram, em geral, melhores índices estatísticos, embora o menor valor de $nRMSE$ (6,08%) e, portanto, a previsão pontual mais acurada tenha sido verificada na previsão diária do dia 06/12/2021 para a série do RS. Outro destaque é a acurácia relativamente baixa dos modelos inerentes aos dados da BA.

Figura 13 – Diagrama de caixa dos valores de nRMSE com faixas de cores referentes aos estados de acurácia. As cores representam os estados excelente (amarelo creme), bom (amarelo esverdeado), regular (amarelo claro) e pobre (amarelo escuro)



Fonte: Elaborado pelo autor (2024).

A Figura 13 providencia uma perspectiva visual dos valores de nRMSE em referência às quatro classes de acurácia mostrados na Tabela 3. Do canto inferior ao superior, as quatro faixas em tonalidade de amarelo representam as classes de acurácia 1, 2, 3 e 4. Tomando como referência o terceiro percentil, representado pela linha horizontal que delimita os retângulos no lado superior, percebe-se que 75% dos valores de nRMSE das previsões amostrais do CE podem ser classificados, no mínimo, como boas; mesmo apresentando a maior quantidade de valores anômalos (15).

Quanto às previsões com os dados do RS, a linha horizontal em destaque dentro dos retângulos, que representa a mediana, permite concluir que 50% dos valores das previsões amostrais do RS podem ser classificados, pelo menos, como bons. Quanto às previsões com os dados da BA, nota-se grande proporção de previsões amostrais entre as classes 3 (regular) e 4 (pobre).

Na Tabela 8, encontram-se os resultados obtidos pelas métricas PICP e PINAW para as 329 previsões amostrais dos dados da BA, do CE e do RS. Além das medidas estatísticas

clássicas, também estão dispostas as respectivas quantidades de amostras de teste com PICP máximo ($N_{PICP,max}$), ou seja, o número de previsões em que seus intervalos de predição de 95% compreendem todos os dados observados aos quais seus dados previstos se referem.

Tabela 8 – Comparação dos resultados das três séries temporais, segundo as métricas PICP e PINAW

Parâmetro	PICP			PINAW		
	BA	CE	RS	BA	CE	RS
Mínimo	2,08%	27,08%	12,50%	38,14%	50,44%	44,41%
Mediano	79,17%	97,92%	70,83%	65,66%	69,15%	73,94%
Médio	76,10%	93,31%	71,80%	68,84%	72,84%	76,01%
Máximo	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Desvio padrão	20,72	9,69	24,01	0,17	0,13	0,14
$N_{PICP,max}$	57	133	84	-	-	-

Fonte: Elaborado pelo autor (2024).

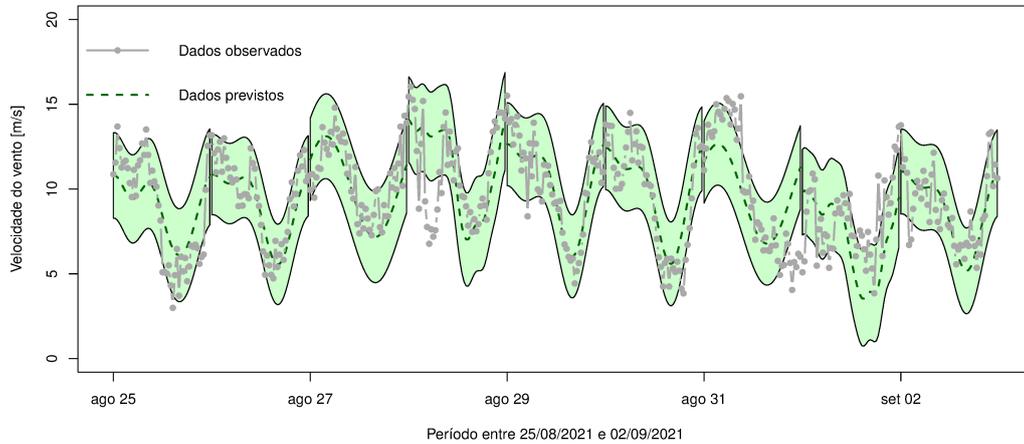
Os resultados da métrica PICP demonstram coerência com os resultados de nRMSE por indicarem que as previsões amostrais do CE tiveram maior quantidade de dados observados compreendidos pelos respectivos intervalos de predição. Em outras palavras, a probabilidade de cobertura dos intervalos de predição foi maior para os dados do CE. Em contraste, os resultados da métrica PINAW mostram que essa maior probabilidade é reflexo dos valores elevados de PINAW, que, por sua vez, indicam a ocorrência de intervalos de predição largos para o CE.

Adotando o valor médio como referência, verifica-se pela Tabela 8 que as previsões diárias do RS produzem intervalos de predição 9,43% e 4,17% mais largos do que as previsões da BA e do CE, respectivamente. É relevante destacar, sobretudo, que não há uma correlação entre o bom desempenho nas métricas nRMSE e PICP e o mau desempenho na métrica PINAW. Há previsões diárias com bons resultados de nRMSE e PICP que apresentam altos valores de PINAW, assim como há previsões diárias que possuem resultados relativamente bons ou intermediários em todas as métricas.

Para fins ilustrativos, as Figuras 14, 15 e 16 consistem em ilustrações de uma composição sequencial de nove previsões diárias das três séries temporais avaliadas. Optou-se por delimitar a apenas nove amostras para melhorar a percepção visual em que o critério de escolha foi baseado no ajuste pontual e, assim, a sequência escolhida para cada série foi aquela que apresentou a menor soma dos valores de nRMSE.

A Figura 14, em particular, apresenta o intervalo amostral que compreende o período entre 25/08/2021 e 02/09/2021 com dados da BA. A amostra correspondente ao dia 29/08/2021

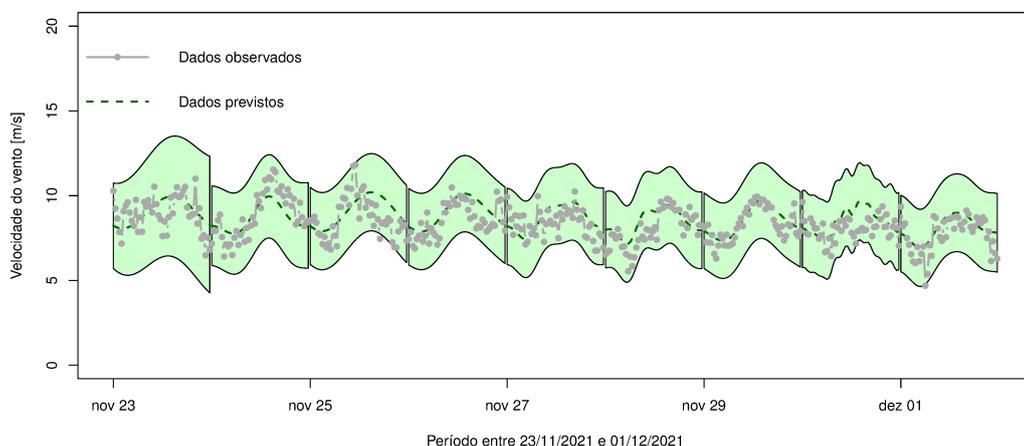
Figura 14 – Amostra dos resultados da previsão para a Bahia. Os dados observados (cinza) são contrastados com os valores previstos (verde) e com os intervalos de predição de 95% (verde claro)



Fonte: Elaborado pelo autor (2024).

obteve o melhor ajuste pontual (com nRMSE de 13,14%) entre todas 329 amostras da BA. Fica evidente pela Figura 14 que a flutuação da velocidade do vento é relevante no horizonte diário em relação aos dados dispostos nas Figuras 15 e 16, sendo esse um dos principais motivos que podem explicar a baixa acurácia dos modelos referentes aos dados da BA. Apesar de algumas previsões apresentarem valores altos para nRMSE — especialmente nos dias 31/08/2021 e 01/09/2021 com 25,86% e 32,11%, respectivamente —, é interessante notar que o intervalo de predição acompanha o perfil de flutuação dos dados.

Figura 15 – Amostra dos valores previstos com intervalo de predição de 95% para o Ceará. Os dados observados (cinza) são contrastados com os valores previstos (verde) e com os intervalos de predição de 95% (verde claro)

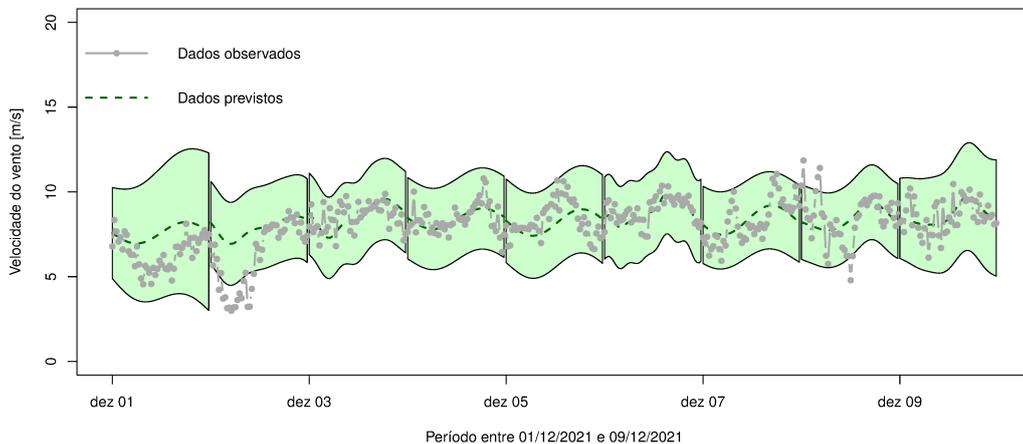


Fonte: Elaborado pelo autor (2024).

A Figura 15 apresenta as amostras referentes ao período de 25/11/2021 a 03/12/2021 do CE. É possível notar que as previsões apresentaram um bom ajuste pontual com base na categorização apresentada por Li *et al.* (2013), pois cinco das nove amostras podem ser categorizadas como boas e as demais podem ser categorizadas como excelentes, a exemplo da previsão do dia 29/11/2021 com o menor nRMSE dentre as amostras de teste do CE: 6,25%. Nesse período, os limites superiores dos intervalos de predição mantiveram-se no patamar de 5 a 13 *m/s*, enquanto que na Figura 14 tais limites excedem tal patamar, especialmente nos dias 28/11/2021 e 02/12/2021.

A Figura 16 apresenta o período de 01/12/2021 a 09/12/2021 para o RS. Embora a inspeção visual aponte distorções relevantes nos dias 01/12/2021, 02/12/2021 e 08/12/2021, há três previsões amostrais que podem ser categorizadas como excelentes. Tais previsões referem-se aos dias 04/12/2021 (com nRMSE de 8,93%), 06/12/2021 (com nRMSE de 6,08%) e 09/12/2021 (com nRMSE de 9,79%).

Figura 16 – Amostra dos valores previstos com intervalo de predição de 95% para o Rio Grande do Sul. Os dados observados (cinza) são contrastados com os valores previstos (verde) e com os intervalos de predição de 95% (verde claro)



Fonte: Elaborado pelo autor (2024).

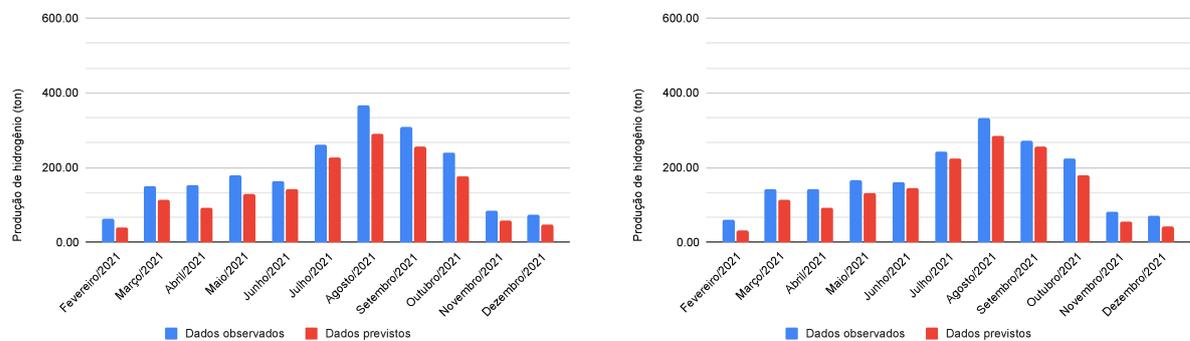
De modo geral, é perceptível que os dados do CE apresentaram melhores ajustes, o que deve estar relacionado à presença mais destacada da componente sazonal, conforme fica perceptível na Figura 8. O ajuste pouco acurado dos modelos, especialmente para os dados da BA, pode ser justificado basicamente por dois motivos: pelo alto grau de variabilidade dos dados estudados e pelo fato de os modelos compostos apenas com as componentes de nível e de

sazonalidade não terem sido capazes de capturar essa variabilidade sistemática.

4.2 Resultados da modelagem do SCEH2

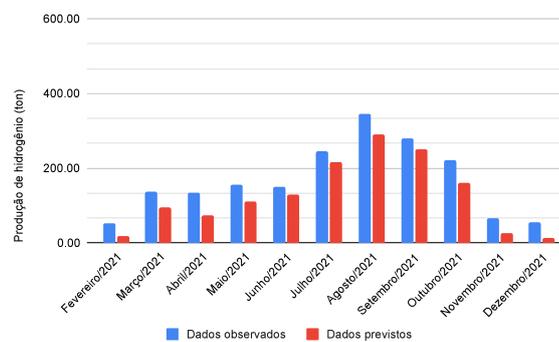
As Figuras 17, 18 e 19 apresentam a produção de hidrogênio projetada com dados observados e previstos para uma análise do perfil mensal. É interessante observar nos gráficos dispostos nas Subfiguras 17a, 17b e 17c que, para todos os métodos, os meses de agosto e setembro apresentaram maior produção de hidrogênio. Em contraste, nos meses de fevereiro, novembro e dezembro a produção atingiu níveis bem baixos de produção.

Figura 17 – Comparação da estimativa da produção mensal de hidrogênio por método de conversão — Bahia



(a) C1

(b) C2



(c) C3

Fonte: Elaborado pelo autor (2024).

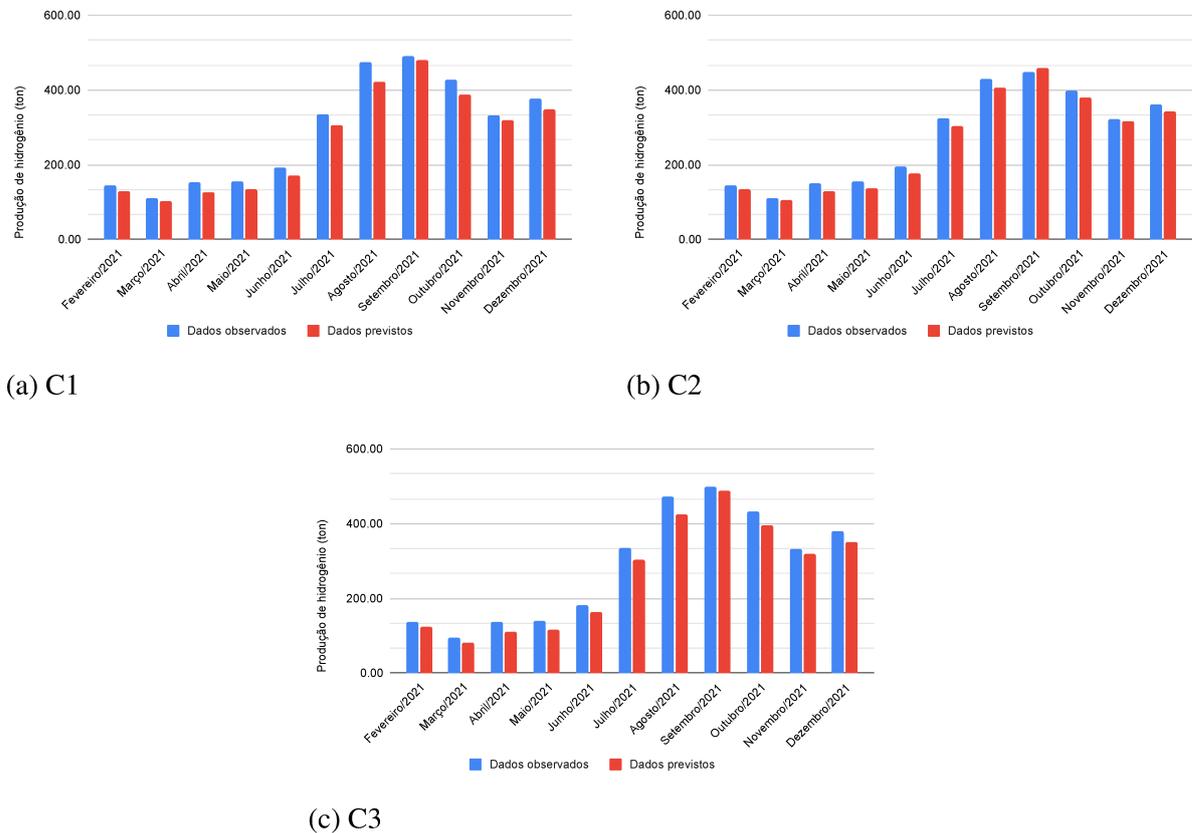
Em um contexto de produção de hidrogênio, isso pode ser objeto de uma análise mais detalhada para que a planta de eletrólise tenha suas campanhas de manutenção preventiva programadas para os períodos de baixa produção, como ocorre nas UEEs. Isso possibilita ganhos financeiros por se evitar perda de oportunidade.

Na maior parte dos casos, evidencia-se nas ilustrações da Figura 17 uma notável

diferença entre os resultados obtidos a partir dos valores observados e previstos, embora seja oportuno destacar que o cálculo realizado pelo C2 parece atenuar mais essa diferença.

No caso do CE, as Subfiguras 18a, 18b e 18c evidenciam o padrão sazonal esperado para produção de hidrogênio, conforme descrito na Seção 3.1. Com isso, a maior capacidade produtiva ocorre nos meses de agosto a novembro e os menores níveis de produção ocorrem entre fevereiro e junho. Quanto à comparação entre dados observados e previstos, percebe-se uma notável diferença nos meses de agosto e setembro para os métodos C1 e C3. Em contrapartida, observa-se uma menor diferença entre os valores observados e previstos para o C2, conforme mostra a Subfigura 18b.

Figura 18 – Comparação da estimativa da produção mensal de hidrogênio por método de conversão — Ceará

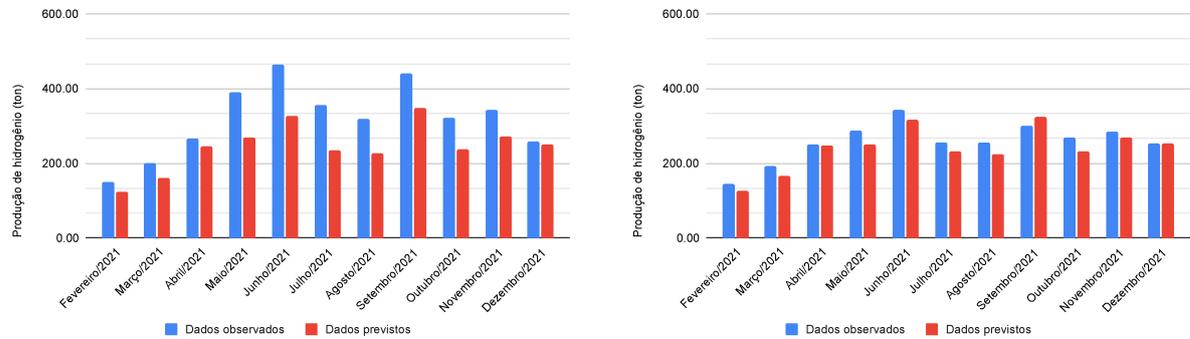


Fonte: Elaborado pelo autor (2024).

As Subfiguras 19a, 19b e 19c mostram o padrão sazonal esperado na região centro-oeste do estado do RS, conforme descrito na Seção 3.1, visto que a produção de hidrogênio é maior no mês de junho. Analogamente aos resultados apresentados pelas Figuras 17 e 18, fica evidente que o cálculo de produção de hidrogênio a partir do método C2 tornou a estimativa mais suave. Dessa forma, considerando o escopo avaliado, o C2 configura-se como o mais apropriado

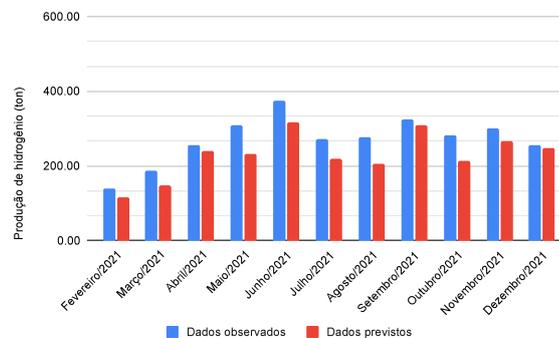
para avaliações do desempenho de um SCEH2 sob a perspectiva mensal.

Figura 19 – Comparação da estimativa da produção mensal de hidrogênio por método de conversão — Rio Grande do Sul



(a) C1

(b) C2



(c) C3

Fonte: Elaborado pelo autor (2024).

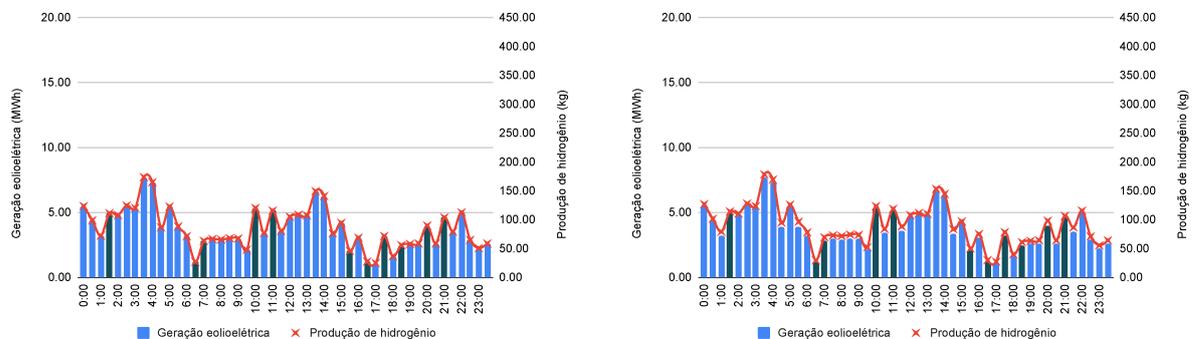
Os resultados apresentados nas Figuras 20, 21 e 22 referem-se aos valores estimados da produção diária por cada método de conversão. Por questão de conveniência, os dias selecionados foram aqueles em que as previsões pontuais apresentaram melhor desempenho. Também é relevante salientar que algumas dessas ilustrações possuem barras destacadas em azul escuro para evidenciar períodos em que a flutuação da geração eólicoelétrica excedeu diferenças na ordem de $\pm 50\%$ em relação ao período anterior.

No caso da BA, as Subfiguras 20a, 20b e 20c mostram os cálculos correspondentes aos dados observados no dia 06/06/2021. O padrão diário é caracterizado por muitas flutuações superiores a $\pm 50\%$ em relação ao período anterior. Esse padrão não muda de forma significativa entre os três métodos. Dentre as barras destacadas, duas delas representam períodos críticos pelo alto nível de flutuação. Entre 09:30h e 10:00h, há um incremento de 152% (de 2,12 para 5,33 *MWh*) e, entre 17:00h e 17:30h, há um novo aumento de 189% (de 1,10 para 3,17 *MWh*).

Na comparação entre os métodos para os dados da BA, nota-se que há instantes em

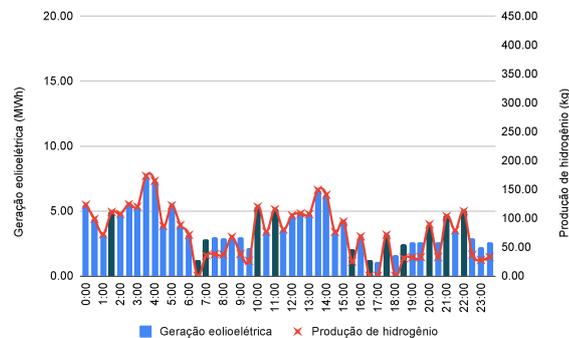
que a produção de hidrogênio calculada por C3 é diferente de C1 e C2. Em particular, no período de 06:30h até 09:30h, a produção mostrada na Subfigura 20c apresenta uma notável flutuação; o que contrasta com o perfil praticamente constante verificado nas Subfiguras 20a e 20b. Essa diferença ocorre porque, nesse período, a eficiência do eletrolisador calculada por C3 varia entre valores de 0%, 42%, 75% e 40%, enquanto a eficiência de C2 apresenta uma flutuação mais suave por variar entre 77% e 82%. Como discutido na Subseção 3.3.1, a eficiência em C1 é considerada constante para todo valor de potência de alimentação do eletrolisador.

Figura 20 – Comparação da estimativa da produção diária de hidrogênio com dados observados no dia 06/06/2021. As barras destacadas representam flutuações de geração eolielétrica superiores a $\pm 50\%$ — Bahia



(a) C1

(b) C2



(c) C3

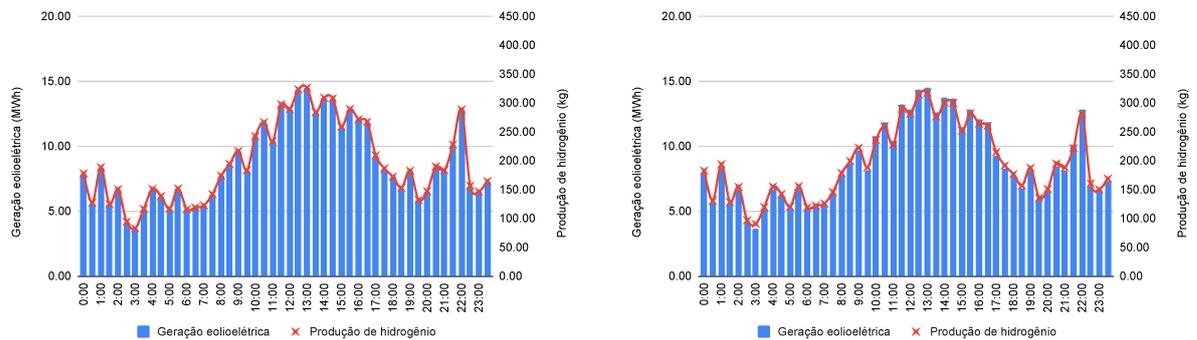
Fonte: Elaborado pelo autor (2024).

Outra diferença entre C3 e os demais métodos refere-se à ocorrência de eficiências com valor de 0%. Particularmente, verificou-se que a eficiência decresce do patamar de 74,95% (às 16:00h) para 0% (de 16:30h até 17:00h) e volta a atingir o valor de 74,95% (às 17:30h). Nesse tipo de situação em que há uma diminuição da produção seguida de um aumento minutos depois, é imprescindível que haja uma avaliação da compatibilidade do nível de flutuação com as especificações operacionais do eletrolisador no que concerne às partidas a quente e a frio. Um

importante parâmetro nesse contexto é a taxa em que a redução/aumento no fornecimento de energia elétrica ocorre (Wang *et al.*, 2018), uma vez que curtos períodos de flutuação na geração de eletricidade podem ocasionar os problemas discutidos na Seção 2.2.

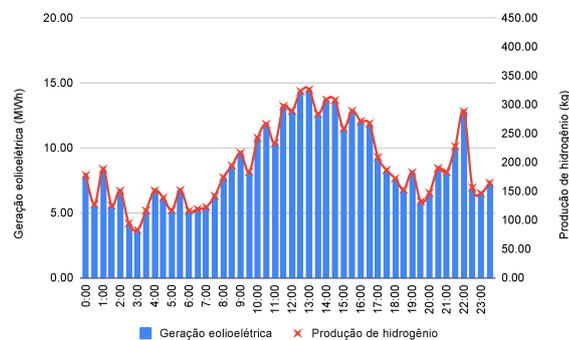
É oportuno reportar ainda que a geração eólicoelétrica total estimada para o dia 06/06/2021 foi de 182,26 *MWh* para os três métodos. Esse valor coincidente decorre do fato de que, nesse dia, as velocidades de vento estiveram dentro da região II da curva de potência (ver Figura 2). De fato, a faixa de velocidades, que é de 4,57 até 8,75 *m/s*, está entre as velocidades de corte inferior e nominal para o aerogerador GE77/1500, conforme mostrado na Tabela 5. Portanto, a conversão vento-potência elétrica é estabelecida integralmente pela relação cúbica entre essas duas grandezas. Além disso, a produção total estimada pelo método C1 é de 4,1 toneladas; pelo método C2, esse valor é de 4,3 toneladas; e, pelo método C3, esse valor é de 3,6 toneladas.

Figura 21 – Comparação da estimativa da produção diária de hidrogênio com dados observados no dia 29/11/2021. As barras destacadas representam flutuações de geração eólicoelétrica superiores a $\pm 50\%$ — Ceará



(a) C1

(b) C2



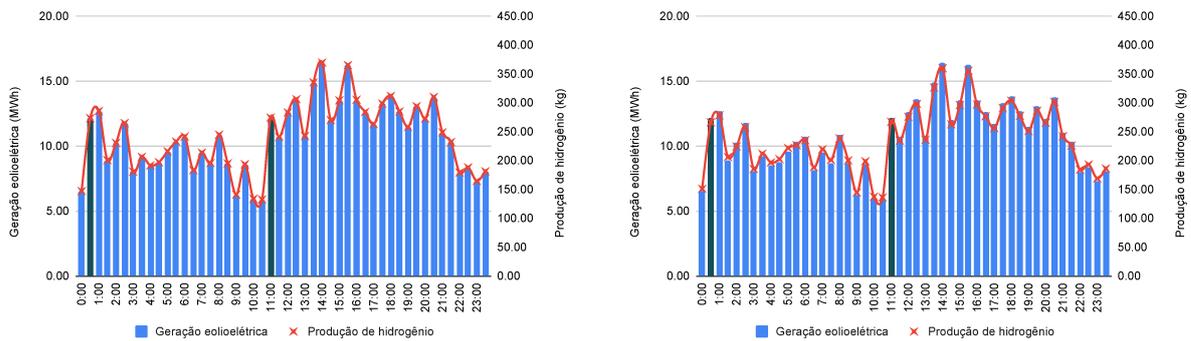
(c) C3

Fonte: Elaborado pelo autor (2024).

Para o CE, os dados referentes ao dia 29 de novembro de 2021 estão dispostos

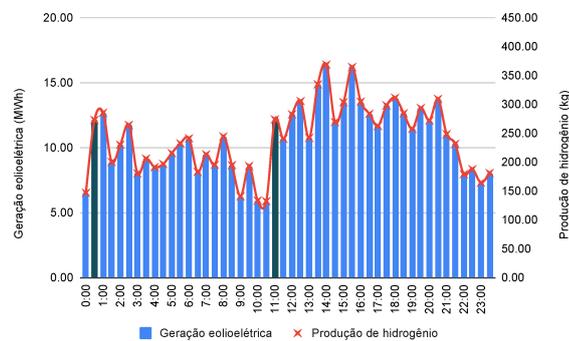
nas Subfiguras 21a, 21b e 21c. O perfil diário nesse dia é caracterizado por uma elevação da produção a partir das 08:00h, por um pico entre 12:30h e 13:00h, seguido de um decaimento até 18:30h. Há também dois picos de produção às 22:00h e às 23:30h. De modo geral, esse dia não apresentou variações maiores que $\pm 50\%$ e, dessa forma, não há barras em coloração destacada para o CE.

Figura 22 – Comparação da estimativa da produção diária de hidrogênio com dados observados no dia 06/12/2021. As barras destacadas representam flutuações de geração eólica superiores a $\pm 50\%$ — Rio Grande do Sul



(a) C1

(b) C2



(c) C3

Fonte: Elaborado pelo autor (2024).

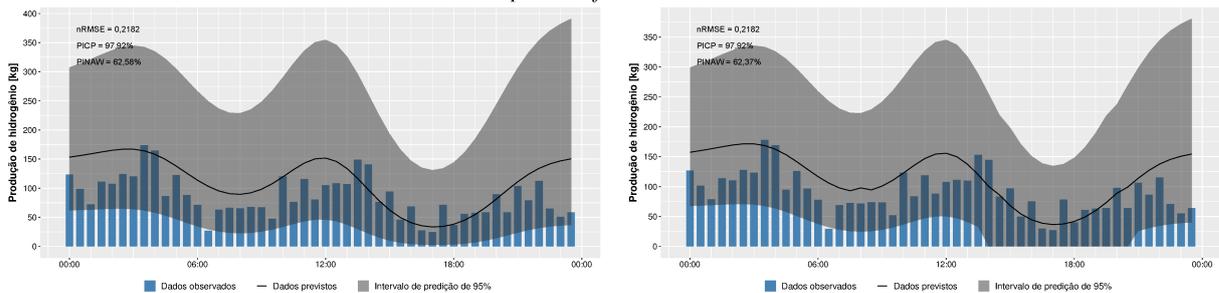
Comparando os três métodos para os dados do CE, percebe-se que também não há diferença da geração eólica total entre os métodos. A faixa de velocidades, que é de 6,29 até 9,94 m/s , está entre as velocidades de corte inferior e nominal para o aerogerador AW125/3000. Como resultado, os três métodos apresentaram o mesmo montante de 417,14 MWh em termos de eletricidade gerada. Essa faixa de velocidades também gerou uma sutil diferença entre os métodos quanto à produção de hidrogênio. Especificamente, os métodos C1, C2 e C3 para o CE apresentaram um montante de hidrogênio de 9,39, 9,41 e 9,38 toneladas, respectivamente.

Para o RS, os dados dispostos nas Subfiguras 22a, 22b e 22c são referentes ao dia 6 de dezembro de 2021. O perfil é relativamente comportado com valores médios de 10,7 *MWh* para a geração eólicoelétrica e de 240,95 *kg* para a produção de hidrogênio. Contudo, dois períodos destacam-se pelo alto nível de flutuação. Um deles ocorre no início do dia — entre 00:00h e 00:30h — com um incremento de geração na ordem de 86% e o outro no início da tarde — entre 10:30h e 11:00h.

Quanto à comparação entre os três métodos de conversão eletricidade-hidrogênio, a análise é similar àquela feita para os dados da BA e do CE, pois a faixa de velocidades é de 7,36 até 10,36 *m/s* e, portanto, está dentro da região II da curva de potência do aerogerador. Com isso, os três métodos apresentaram o mesmo montante de 513,51 *MWh* em termos de eletricidade gerada, além de valores similares de hidrogênio produzido: 11,57, 11,44 e 11,55 toneladas para C1, C2 e C3, respectivamente.

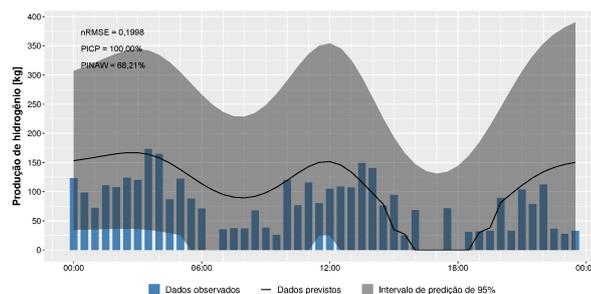
As Figuras 23, 24 e 25 apresentam a comparação das estimativas a partir dos dados observados, previstos e dos valores de L_{sup} e L_{inf} . Nas Subfiguras 23a e 23b, nota-se que as estimativas feitas pelos métodos C1 e C2 para os dados da BA possuem grande similaridade, ainda que os valores de L_{inf} para o período entre às 14:00h e às 20:30h sejam iguais a zero para o C2, conforme mostrado na Subfigura 23b.

Figura 23 – Estimativa da produção de hidrogênio a partir dos dados observados (azul), previstos (preto) e dos valores de L_{sup} e L_{inf} (cinza) referentes ao dia 06/06/2021 — Bahia



(a) C1

(b) C2



(c) C3

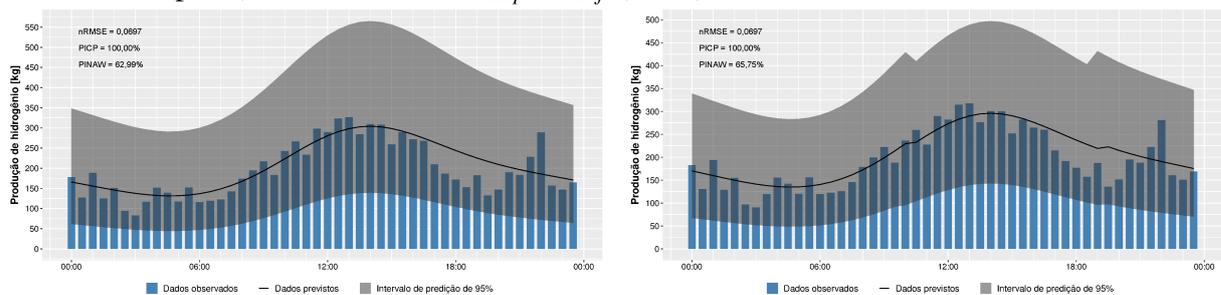
Fonte: Elaborado pelo autor (2024).

A similaridade visual entre C1 e C2 é confirmada pela igualdade entre os respectivos valores de nRMSE (0,2182) e de PICP (97,92%). As estimativas feitas a partir de C3 resultaram em uma acurácia de 19,98%, que é 8,4% melhor que C1 e C2. Esse melhor desempenho pontual é acompanhado por uma maior cobertura (PICP = 100%). A dimensão do intervalo de predição, porém, permanece em um patamar próximo ao dos métodos C1 e C2 com uma sutil diferença a menor.

De modo geral, avaliando o ajuste pontual da produção de hidrogênio com base nas classes de acurácia apresentadas na Tabelas 3, pode-se classificar as estimativas de C1 e C2 para o dia 06/06/2021 como regulares e a estimativa de C3 como boa. Cabe reportar que o erro pontual de 13,04% para previsão da variável explicativa passa a resultar em 21,82% (para C1 e C2) e 19,98% (para C3) após dois processos de conversão — nomeadamente vento-eletricidade e eletricidade-hidrogênio.

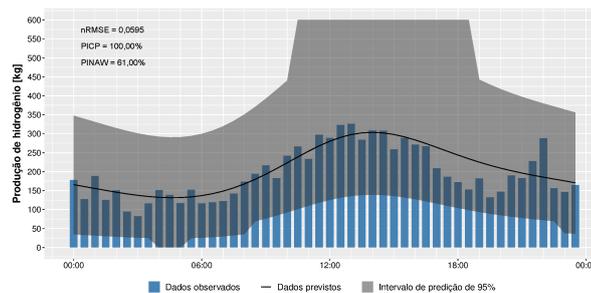
Para o CE, as Subfiguras 24a, 24b e 24c ilustram as estimativas de produção de hidrogênio para dia 29/11/2021 e, como esperado, os resultados foram mais acurados que aqueles obtidos para as amostras da BA e do RS. Especialmente entre 00:00h e 16:30h, a produção estimada com os dados previstos apresenta um bom ajuste em relação ao perfil da estimativa com os dados observados.

Figura 24 – Estimativa da produção de hidrogênio a partir dos dados observados (azul), previstos (preto) e dos valores de L_{sup} e L_{inf} (cinza) referentes ao dia 29/11/2021 — Ceará



(a) C1

(b) C2



(c) C3

Fonte: Elaborado pelo autor (2024).

Em referência às classes de acurácia, pode-se classificar as estimativas para o CE no dia 29/11/2021 como excelentes, pois os três métodos obtiveram valores de nRMSE compatíveis com essa classe — 6,97% para C1 e C2, e 5,95% para C3. Curiosamente, o resultado da métrica nRMSE para C3 foi melhor, inclusive, que o próprio valor obtido para previsão da variável explicativa (6,25%). O valor de PINAW para a estimativa de C3 (61,00%) é cerca de 6,5% menor que o PINAW para a previsão (65,26%).

Ao contrário de C1, que apresenta curvas suaves e alinhadas com o comportamento dos dados previstos, fica perceptível nas Subfiguras 24b e 24c que os limites dos intervalos de predição de C2 e de C3 possuem desníveis. Isso deriva do fato de que, para C2 e C3, a magnitude da eletricidade que alimenta o eletrolisador determina o patamar da eficiência do processo de eletrólise. Por conseguinte, é natural que a distância entre os valores previstos e de seus respectivos limites superior e inferior seja diferente.

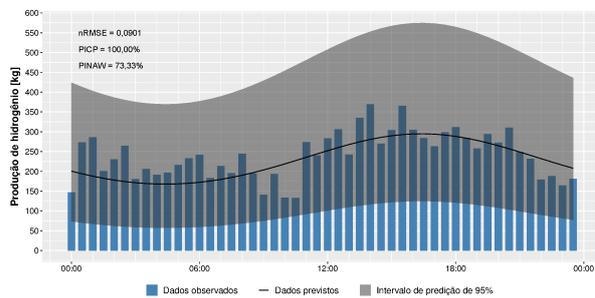
Na Subfigura 24b, por exemplo, o recorte entre 10:30h e 18:30h apresenta um desnível da curva que representa os valores de L_{sup} . Isso ocorre porque, nesse período, as potências estimadas a partir de velocidades, que variam entre 11,13 e 11,93 m/s , pertencem à faixa de potência que retorna uma eficiência do eletrolisador de 66% (ver Tabela 4), enquanto que, no resto do dia, a eficiência é de 74%.

Já na Subfigura 24c, percebe-se um ressalto da curva no referido intervalo de tempo. Isso acontece porque os valores de potência excedem a capacidade nominal do eletrolisador, o cálculo da conversão passa a ser regido pela quarta condição da Equação (3.10). Com isso, a quantidade de hidrogênio produzido nesse período a partir dos dados de L_{sup} é 600,60 kg ($40.000kW \times 0,5h \times 100\% \div 33,3kWh/kg$) a cada intervalo de 30 minutos.

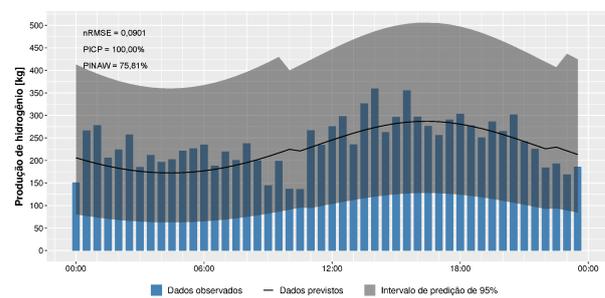
No que concerne ao RS, as estimativas de produção de hidrogênio para o dia 06/12/2021 podem ser classificadas como excelentes, dado que os valores de nRMSE foram 9,01% para C1 e C2, e 5,95% para C3. Diferente do resultado observado para o CE, os valores de nRMSE dos três métodos foram maiores do que aquele obtido para a previsão de velocidade de ventos (6,08%). Mas somente os valores da métrica PINAW para C1 (73,33%) e C3 (74,97%) foram ligeiramente menores que o PINAW para a previsão (74,44%).

Assim como observado para o CE, as estimativas a partir dos métodos C2 e de C3 para o RS resultaram em desníveis nas curvas dos limites dos intervalos de predição. Na Subfigura 25b, o período entre 10:00h e 22:30h apresenta um desnível nos valores de L_{sup} . Isso ocorre porque, nesse período, as velocidades variam entre 11,10 e 12,00 m/s e resultam em uma

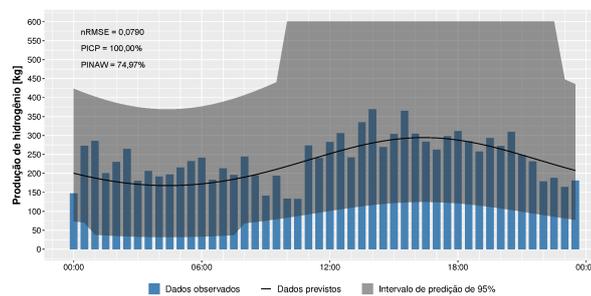
Figura 25 – Estimativa da produção de hidrogênio a partir dos dados observados (azul), previstos (preto) e dos valores de L_{sup} e L_{inf} (cinza) referentes ao dia 06/12/2021 — Rio Grande do Sul



(a) C1



(b) C2



(c) C3

Fonte: Elaborado pelo autor (2024).

eficiência do eletrolisador de 66%. No restante do dia, a eficiência é de 73%.

Na Subfigura 25c, notam-se dois períodos com desnível. Um deles, compreendido entre 01:30h e 07:30h, refere-se à curva do L_{inf} e apresenta velocidades que variam entre 5,57 e 5,87 m/s , o que resulta em uma eficiência média de 42%, que, por sua vez, é 44% menor que a eficiência média do restante do dia para os valores de L_{inf} (74,95%). O outro desnível refere-se à curva do L_{sup} e ocorre entre 10:00h e 22:30h. Nesse caso, a eficiência média é de 100%, que difere dos 74,83% médios verificados para o restante do dia.

5 CONCLUSÃO

Este trabalho apresentou os DLMs bayesianos como uma técnica de previsão probabilística de vento em um contexto de produção de hidrogênio. Além da previsão em si, o estudo também enfatizou a dinamicidade da modelagem de sistemas de conversão vento-eletricidade-hidrogênio.

Para isso, inicialmente, foi realizada uma investigação acerca dos principais métodos de conversão de energia eólicoelétrica em hidrogênio que apontou uma tendência da literatura que trata de SCEH2 em propor métodos de conversão de energia eólicoelétrica em hidrogênio que aproximem o cálculo teórico da dinâmica operacional desse tipo de sistema. Tal tendência mostra alinhamento com a parcela da comunidade científica que passa a realizar experimentos e/ou análises críticas sobre as influências da flutuação da fonte primária no desempenho de eletrolisadores.

A revisão feita neste estudo mostrou que a operação intermitente pode influenciar o desempenho dos eletrolisadores de diversas formas. Dentre elas, destaca-se que a potência flutuante pode causar a degradação do eletrodo em virtude das mudanças repentinas no potencial do eletrodo. Além disso, existe um maior risco operacional causado pelo gás de cruzamento, que é uma mistura indesejada dos gases hidrogênio e oxigênio.

Para validação da modelagem preditiva estudada na pesquisa, foram coletados dados de velocidade de vento medidos em usinas eólicoelétricas situadas nos estados da Bahia, do Ceará e do Rio Grande do Sul. Após o devido tratamento, tais dados foram modelados por meio de um procedimento de previsão contínua em que o período de uma semana foi escolhido como janela temporal para treinamento dos modelos preditivos. Tal procedimento gerou como resultado 329 previsões diárias para cada série temporal avaliada.

Entre as séries, aquela referente ao Ceará obteve os melhores resultados em termos de ajuste pontual e de probabilidade de cobertura. Especificamente, as previsões com os dados do Ceará tiveram valores medianos de 15% e de 97,92% nas métricas nRMSE e PICP, respectivamente. No caso do Rio Grande do Sul, os valores foram de 28,55% para o nRMSE mediano e 70,83% para a PICP mediana. Para a Bahia, os valores foram de 32,72% para o nRMSE mediano e 79,17% para a PICP mediana. Na avaliação probabilística pela métrica PINAW, verificou-se que as previsões da Bahia obtiveram intervalos de predição mais estreitos — com um PINAW mediano de 65,66% — do que os intervalos para o Ceará e Rio Grande do Sul — com valores medianos de PINAW de 69,15% e de 73,94%.

Dentre os três métodos de conversão utilizados para estimar a geração eólicoelétrica e a produção de hidrogênio correspondente, percebeu-se que o C2 resultou em menores diferenças entre os valores observados e previstos no horizonte mensal. A avaliação da estimativa da produção de hidrogênio sob a perspectiva diária fomentou a discussão acerca da taxa em que a redução e o aumento no fornecimento de energia elétrica ocorre. Por essa avaliação, a amostra da Bahia conteve a maior quantidade diária de períodos críticos decorrentes de variações superiores a $\pm 50\%$.

Por fim, a experiência obtida através deste trabalho proporciona a visualização de como a previsão probabilística de vento via DLMs fornece interpretações detalhadas dos resultados de previsão de ventos em um contexto de produção de hidrogênio. Assim, o conceito de tomada de decisão baseada em dados passa a ser uma importante estratégia para decisões gerenciais em plantas de eletrólise alimentadas por fontes intermitentes.

Nesse sentido, os seguintes tópicos de pesquisa são sugeridos para trabalhos futuros:

- inclusão da análise de ventos de rajada e de calmaria no processo preditivo a fim de melhor a captura da variabilidade sistemática do vento;
- investigação do desempenho de modelos dinâmicos generalizados a fim de analisar e prever séries temporais que não satisfaçam a presunção de normalidade;
- e
- previsão de produção de hidrogênio com dados reais de uma planta em operação.

REFERÊNCIAS

- ADEDIPE, T.; SHAFIEE, M.; ZIO, E. Bayesian network modelling for the wind energy industry: An overview. **Reliability Engineering System Safety**, s.l., v. 202, p. 107053, 2020.
- AGÊNCIA CÂMARA DE NOTÍCIAS. **Comissão Especial da Transição Energética é instalada**. 2023. <<https://www.camara.leg.br/noticias/968800-comissao-especial-da-transicao-energetica-e-instalada/>>. Acesso em: 12 set. 2023.
- AGÊNCIA DE DESENVOLVIMENTO DO ESTADO DO CEARÁ. **Atlas Eólico e Solar do Ceará**. 2019. <<http://atlas.adece.ce.gov.br>>. Acesso em: 03 mar. 2023.
- AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **Informações Geográficas do Setor Elétrico Brasileiro**. 2017. <<https://antigo.aneel.gov.br/web/guest/informacoes-geograficas>>. Acesso em: 07 mar. 2023.
- ALANIS, A. Y.; SANCHEZ, O. D.; ALVAREZ, J. G. Time series forecasting for wind energy systems based on high order neural networks. **Mathematics**, s.l., v. 9, 2021.
- ALAVI, O.; MOSTAFAEIPOUR, A.; QOLIPOUR, M. Analysis of hydrogen production from wind energy in the southeast of iran. **International Journal of Hydrogen Energy**, s.l., v. 41, n. 34, p. 15158–15171, 2016.
- ALY, H. H. An intelligent hybrid model of neuro wavelet, time series and recurrent kalman filter for wind speed forecasting. **Sustainable Energy Technologies and Assessments**, s.l., v. 41, p. 100802, 2020.
- ALY, H. H. A hybrid optimized model of adaptive neuro-fuzzy inference system, recurrent kalman filter and neuro-wavelet for wind power forecasting driven by dfig. **Energy**, s.l., v. 239, p. 122367, 2022.
- ASSOCIADOS, C. S. E.; S.A., E. C. E. **Atlas do potencial eólico brasileiro**. Porto Alegre: [s.n.], 2014. ISBN 0387947256.
- BAHOU, S. Techno-economic assessment of a hydrogen refuelling station powered by an on-grid photovoltaic solar system: A case study in morocco. **International Journal of Hydrogen Energy**, v. 48, p. 23363–23372, 2023.
- BARRIO, B. F. **Large scale Bayesian dynamic forecasting for count time series**. 2022. 166 f. Tese (Doctorado en Ingeniería Matemática, Estadística e Investigación Operativa) — Programa de doctorado en Ingeniería Matemática, Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas, Universidad Complutense de Madrid, Madrid, 2022.
- BARROS, A. M. de L. **Avaliação da precipitação como variável exógena na previsão de geração eólica de curto prazo de uma central eólica no Ceará**. 2017. 193 f. Dissertação (Doutorado em Engenharia Civil) — Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Pernambuco, Recife, 2017.
- BESSAC, J.; AILLIOT, P.; MONBET, V. Gaussian linear state-space model for wind fields in the north-east atlantic. **Environmetrics**, s.l., v. 26, n. 1, p. 29–38, 2015.

BEZERRA, F. D. V.; MARINHO, F. P.; ROCHA, P. A. C.; SANTOS, V. O.; THÉ, J. V. G.; GHARABAGHI, B. Machine learning dynamic ensemble methods for solar irradiance and wind speed predictions. **Atmosphere**, s.l., v. 14, 2023. Disponível em: <<https://www.mdpi.com/2073-4433/14/11/1635>>.

BOUNOUA, Z.; Ouazzani Chahidi, L.; MECHAQRANE, A. Estimation of daily global solar radiation using empirical and machine-learning methods: A case study of five moroccan locations. **Sustainable Materials and Technologies**, s.l., v. 28, p. e00261, 2021.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011. **Diário Oficial da República Federativa do Brasil**, Brasília, DF, 2011. Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2012/Lei/L12651.htm>.

BRKA, A.; AL-ABDELI, Y. M.; KOTHAPALLI, G. Predictive power management strategies for stand-alone hydrogen systems: Operational impact. **International Journal of Hydrogen Energy**, v. 41, n. 16, p. 6685–6698, 2016. ISSN 0360-3199. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360319915307862>>.

BRKA, A.; KOTHAPALLI, G.; AL-ABDELI, Y. M. Predictive power management strategies for stand-alone hydrogen systems: Lab-scale validation. **International Journal of Hydrogen Energy**, v. 40, n. 32, p. 9907–9916, 2015. ISSN 0360-3199. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360319915015669>>.

CAI, H.; JIA, X.; FENG, J.; YANG, Q.; LI, W.; LI, F.; LEE, J. A unified bayesian filtering framework for multi-horizon wind speed prediction with improved accuracy. **Renewable Energy**, s.l., v. 178, p. 709–719, 2021.

CANO, M. H.; KELOUWANI, S.; AGBOSSOU, K.; DUBÉ, Y. Power management system for off-grid hydrogen production based on uncertainty. **International Journal of Hydrogen Energy**, v. 40, n. 23, p. 7260–7272, 2015. ISSN 0360-3199. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360319915008277>>.

CARRILLO, C.; OBANDO MONTAÑO, A.; CIDRÁS, J.; DÍAZ-DORADO, E. Review of power curve modelling for wind turbines. **Renewable and Sustainable Energy Reviews**, s.l., v. 21, p. 572–581, 2013.

CEARÁ. Decreto nº 34.315, de 20 de outubro de 2021. **Diário Oficial do Estado do Ceará**, Fortaleza, CE, 2021. Disponível em: <<http://pesquisa.doe.seplag.ce.gov.br/doepesquisa/sead.do?page=ultimasEdicoes&cmd=11&action=Ultimas>>.

CEYLAN, C.; DEVRIM, Y. Green hydrogen based off-grid and on-grid hybrid energy systems. **International Journal of Hydrogen Energy**, v. 48, p. 39084–39096, 2023.

CHAUDHARY, K.; BHARDVAJ, K.; CHAUDHARY, A. A qualitative assessment of hydrogen generation techniques for fuel cell applications. **Fuel**, s.l., v. 358, p. 130090, 2024.

De Battista, H.; MANTZ, R. J.; GARELLI, F. Power conditioning for a wind–hydrogen energy system. **Journal of Power Sources**, s.l., v. 155, n. 2, p. 478–486, 2006.

DHAKAL, R.; SEDAI, A.; POL, S.; PARAMESWARAN, S.; NEJAT, A.; MOUSSA, H. A novel hybrid method for short-term wind speed prediction based on wind probability distribution function and machine learning models. **Applied Sciences**, s.l., v. 12, 2022.

- DINH, V. N.; LEAHY, P.; MCKEOGH, E.; MURPHY, J.; CUMMINS, V. Development of a viability assessment model for hydrogen production from dedicated offshore wind farms. **International Journal of Hydrogen Energy**, s.l., v. 46, n. 48, p. 24620–24631, 2021.
- DOBBIN, T. M. C. V. d. C. **Modelagem e controle de um aerogerador duplamente alimentado com turbina de velocidade variável para estudos de estabilidade eletromecânica**. 2014. 163 f. Dissertação (Mestrado em Engenharia Elétrica) — Programa de Pós-Graduação e pesquisa em Engenharia, Instituto Alberto Luiz Coimbra, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.
- DOBÓ, Z.; PALOTÁS, Á. B. Impact of the current fluctuation on the efficiency of alkaline water electrolysis. **International Journal of Hydrogen Energy**, s.l., v. 42, n. 9, p. 5649–5656, 2017.
- DUCA, V. E.; FONSECA, T. C.; CYRINO OLIVEIRA, F. L. A generalized dynamical model for wind speed forecasting. **Renewable and Sustainable Energy Reviews**, v. 136, p. 110421, 2021.
- DUCA, V. E.; FONSECA, T. C.; CYRINO OLIVEIRA, F. L. Joint modelling wind speed and power via bayesian dynamical models. **Energy**, v. 247, p. 123431, 2022.
- DUCA, V. E. L. de A. **Non-Gaussian dynamical modeling of wind power generation**. 2020. 82 f. Tese (Pós-doutorado em Estatística) — Instituto de Matemática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2020.
- DURBIN, J.; KOOPMAN, S. Maximum likelihood estimation of parameters. In: **Time Series Analysis by State Space Methods**. s.l.: Oxford University Press, 2012. ISBN 9780199641178.
- EL-SHAFIE, M. Hydrogen production by water electrolysis technologies: A review. **Results in Engineering**, s.l., v. 20, p. 101426, 2023.
- EMPRESA DE PESQUISA ENERGÉTICA. **Nota técnica: DEA 06/14**: Sistema de acompanhamento de medições anemométricas ama — manual do usuário. Rio de Janeiro: EPE, 2014.
- FLAMM, B.; PETER, C.; BÜCHI, F. N.; LYGEROS, J. Electrolyzer modeling and real-time control for optimized production of hydrogen gas. **Applied Energy**, s.l., v. 281, p. 116031, 2021.
- GANESHAN, I. S.; MANIKANDAN, V.; Ram Sundhar, V.; SAJIV, R.; SHANTHI, C.; KOTTAYIL, S. K.; RAMACHANDRAN, T. Regulated hydrogen production using solar powered electrolyser. **International Journal of Hydrogen Energy**, s.l., v. 41, n. 24, p. 10322–10326, 2016.
- GARCÍA, I.; HUO, S.; PRADO, R.; BRAVO, L. Dynamic bayesian temporal modeling and forecasting of short-term wind measurements. **Renewable Energy**, v. 161, p. 55–64, 2020.
- GELMAN, A.; CARLIN, J. B.; STERN, H. S.; DUNSON, D. B.; VEHTARI, A.; RUBIN, D. B. **Bayesian data analysis**. New York: Chapman and Hall/CRC, 2015. 675 p. ISBN 9780429113079.
- GIELEN, D.; BOSHELL, F.; SAYGIN, D.; BAZILIAN, M. D.; WAGNER, N.; GORINI, R. The role of renewable energy in the global energy transformation. **Energy Strategy Reviews**, s.l., v. 24, p. 38–50, 2019.

- GOMES, L. E. da S. **Calibração Espaço-Temporal de Previsões Numéricas do Modelo de Mesoescala Eta para a Velocidade do Vento em Minas Gerais**. Dissertação (Mestrado em Estatística) — Instituto de Matemática, Programa de Pós-Graduação em Estatística, Universidade Federal do Rio de Janeiro, 2018.
- GOUVEIA, H. T. V. **Previsão de ventos e geração eólica do sistema ne: analisando diversos sítios e buscando a melhor modelagem através da inteligência artificial**. 2011. 133 f. Dissertação (Mestrado em Engenharia Elétrica) — Centro de Tecnologia e Geociências, Programa de Pós-Graduação em Engenharia Elétrica, Universidade de Pernambuco, Recife, 2011.
- GRIGORIEV, S. A.; FATEEV, V. N. **Hydrogen Production by Water Electrolysis**. s.l.: John Wiley & Sons, Ltd, 2023. 233 p. ISBN 9781119283676.
- GRÜGER, F.; HOCH, O.; HARTMANN, J.; ROBINIUS, M.; STOLTEN, D. Optimized electrolyzer operation: Employing forecasts of wind energy availability, hydrogen demand, and electricity prices. **International Journal of Hydrogen Energy**, s.l., v. 44, n. 9, p. 4387–4397, 2019.
- GUPTA, A.; KUMAR, M.; KUMAR, V.; THAKUR, I. S. Chapter 10 - hydrogen production and carbon sequestration for biofuels and biomaterials. In: THAKUR, I. S.; PANDEY, A.; NGO, H. H.; SOCCOL, C. R.; LARROCHE, C. (Ed.). **Biomass, Biofuels, Biochemicals**. s.l.: Elsevier, 2022. p. 231–252. ISBN 978-0-12-823500-3.
- HAORAN, C.; XIA, Y.; WEI, W.; YONGZHI, Z.; BO, Z.; LEIQI, Z. Safety and efficiency problems of hydrogen production from alkaline water electrolyzers driven by renewable energy sources. **International Journal of Hydrogen Energy**, s.l., 2023.
- HE, Y.; LIU, R.; LI, H.; WANG, S.; LU, X. Short-term power load probability density forecasting method using kernel-based support vector quantile regression and copula theory. **Applied Energy**, s.l., v. 185, p. 254–266, 2017.
- HOFRICHTER, A.; RANK, D.; HEBERL, M.; STERNER, M. Determination of the optimal power ratio between electrolysis and renewable energy to investigate the effects on the hydrogen production costs. **International Journal of Hydrogen Energy**, s.l., v. 48, n. 5, p. 1651–1663, 2023.
- HONG, Z.; WEI, Z.; HAN, X. Optimization scheduling control strategy of wind-hydrogen system considering hydrogen production efficiency. **Journal of Energy Storage**, s.l., v. 47, p. 103609, 2022.
- HONSHO, Y.; NAGAYAMA, M.; MATSUDA, J.; ITO, K.; SASAKI, K.; HAYASHI, A. Durability of pem water electrolyzer against wind power voltage fluctuation. **Journal of Power Sources**, s.l., v. 564, p. 232826, 2023.
- HU, J.; HU, W.; CAO, D.; HUANG, Y.; CHEN, J.; LI, Y.; CHEN, Z.; BLAABJERG, F. Bayesian averaging-enabled transfer learning method for probabilistic wind power forecasting of newly built wind farms. **Applied Energy**, s.l., v. 355, p. 122185, 2024.
- INTERNATIONAL RENEWABLE ENERGY AGENCY. **Green hydrogen cost reduction: scaling up electrolyzers to meet the 1.5 climate goal**. [S.l.]: IRENA, 2020.

- JAVAID, A.; JAVAID, U.; SAJID, M.; RASHID, M.; UDDIN, E.; AYAZ, Y.; WAQAS, A. Forecasting hydrogen production from wind energy in a suburban environment using machine learning. **Energies**, s.l., v. 15, n. 23, 2022.
- KHOSRAVI, A.; NAHAVANDI, S.; CREIGHTON, D. Prediction intervals for short-term wind farm power generation forecasts. **IEEE Transactions on Sustainable Energy**, s.l., v. 4, p. 602–610, 2013.
- KING, B.; KOWAL, D. R. Warped dynamic linear models for time series of counts. **Bayesian Analysis**, s.l., v. 1, n. 1, 2023.
- KOJIMA, H.; NAGASAWA, K.; TODOROKI, N.; ITO, Y.; MATSUI, T.; NAKAJIMA, R. Influence of renewable energy power fluctuations on water electrolysis for green hydrogen production. **International Journal of Hydrogen Energy**, s.l., v. 48, n. 12, p. 4572–4593, 2023.
- KOVAČ, A.; PARANOS, M.; MARCIUŠ, D. Hydrogen in energy transition: A review. **International Journal of Hydrogen Energy**, s.l., v. 46, n. 16, p. 10016–10035, 2021.
- LI, G.; SHI, J. Applications of bayesian methods in wind energy conversion systems. **Renewable Energy**, s.l., v. 43, p. 1–8, 2012.
- LI, M.-F.; TANG, X.-P.; WU, W.; LIU, H.-B. General models for estimating daily global solar radiation for different solar radiation zones in mainland china. **Energy Conversion and Management**, s.l., v. 70, p. 139–148, 2013.
- LIAO, Q.; CAO, D.; CHEN, Z.; BLAABJERG, F.; HU, W. Probabilistic wind power forecasting for newly-built wind farms based on multi-task gaussian process method. **Renewable Energy**, s.l., v. 217, p. 119054, 2023.
- LIN, L.; XIA, D.; DAI, L.; ZHENG, Q.; QIN, Z. Chaotic analysis and prediction of wind speed based on wavelet decomposition. **Processes**, s.l., v. 9, 2021.
- LIU, L.; WANG, J. Super multi-step wind speed forecasting system with training set extension and horizontal–vertical integration neural network. **Applied Energy**, s.l., v. 292, p. 116908, 2021.
- MA, Z.; TIAN, T.; CUI, Q.; SHU, J.; ZHAO, J.; WANG, H. Rapid sizing of a hydrogen-battery storage for an offshore wind farm using convex programming. **International Journal of Hydrogen Energy**, s.l., v. 48, n. 58, p. 21946–21958, 2023.
- Martinez Lopez, V.; ZIAR, H.; HAVERKORT, J.; ZEMAN, M.; ISABELLA, O. Dynamic operation of water electrolyzers: A review for applications in photovoltaic systems integration. **Renewable and Sustainable Energy Reviews**, s.l., v. 182, p. 113407, 2023.
- MAZZEO, D.; HERDEM, M. S.; MATERA, N.; WEN, J. Z. Green hydrogen production: Analysis for different single or combined large-scale photovoltaic and wind renewable systems. **Renewable Energy**, s.l., v. 200, p. 360–378, 2022.
- MIKKONEN, S.; NÉMETH, Z.; VARGA, V.; WEIDINGER, T.; LEINONEN, V.; YLI-JUUTI, T.; SALMA, I. Decennial time trends and diurnal patterns of particle number concentrations in a central european city between 2008 and 2018. **Atmospheric Chemistry and Physics**, s.l., v. 20, p. 12247–12263, 2020.

MINISTÉRIO DE MINAS E ENERGIA. **Programa Nacional do Hidrogênio**. Disponível em: <<https://www.gov.br/mme/pt-br/programa-nacional-do-hidrogenio-1>>. Acesso em: 12 set. 2023.

MITRENTSIS, G.; LENS, H. An interpretable probabilistic model for short-term solar power forecasting using natural gradient boosting. **Applied Energy**, s.l., v. 309, p. 118473, 2022.

MONTENEGRO-SILVA, C. P. **Inferência Bayesiana em Modelos de Dinâmica de Populações Biológicas com Termo de Perturbação Assimétrico**. 2016. 114 f. Tese (Doutorado em Ciências) — Instituto de Matemática, Universidade de São Paulo, São Paulo, 2016.

MOSTAFAEIPOUR, A.; REZAEI, M.; MOFTAKHARZADEH, A.; QOLIPOUR, M.; SALIMI, M. Evaluation of hydrogen production by wind energy for agricultural and industrial sectors. **International Journal of Hydrogen Energy**, s.l., v. 44, n. 16, p. 7983–7995, 2019.

MUMMEY, J. F. C. **Uma contribuição metodológica para a otimização da operação e expansão do sistema hidrotérmico brasileiro mediante a representação estocástica da geração eólica**. 2017. 184 f. Tese (Doutorado em Ciências) — Programa de Pós-Graduação em Energia, Instituto de Energia e Ambiente, Universidade de São Paulo, São Paulo, 2017.

PEREIRA, J. da S. **Um modelo dinâmico para séries temporais contínuas com massa em zero**. 2017. 89 f. Dissertação (Mestrado em Matemática) — Programa de Pós-Graduação em Matemática, Universidade Federal do Amazonas, Manaus, 2017.

PETRIS, G. An r package for dynamic linear models. **Journal of Statistical Software**, s.l., v. 36, n. 12, p. 1–16, 2010.

PETRIS, G.; PETRONE, S.; CAMPAGNOLI, P. **Dynamic linear models with R**. New York: Springer Science & Business Media, 2009. ISBN 21975736.

PRADO, R.; WEST, M. **Time series: modeling, computation, and inference**. New York: Chapman and Hall/CRC, 2010. 368 p. ISBN 9780429184819.

QIU, X.; ZHANG, H.; QIU, Y.; ZHOU, B.; ZANG, T.; QI, R.; LIN, J.; WANG, J. Online dynamic parameter estimation of an alkaline electrolysis system based on bayesian inference. In: **2022 IEEE 5th International Electrical and Energy Conference (CIEEC)**. Nanjing: [s.n.], 2022. p. 3362–3367.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>.

RAMAKRISHNAN, S.; DELPISHEH, M.; CONVERY, C.; NIBLETT, D.; VINOTHKANNAN, M.; MAMLOUK, M. Offshore green hydrogen production from wind energy: Critical review and perspective. **Renewable and Sustainable Energy Reviews**, v. 195, p. 114320, 2024.

RIPLEY, B. D. Time series in r 1.5. 0. **The Newsletter of the R Project Volume**, s.l., v. 2, p. 2, 2002.

RISBUD, M.; KAUR, G.; DHAWALE, D. S.; ZHU, H.; HAQUE, N.; GIDDEY, S. Chapter 14 - electrolyzer technologies for hydrogen economy. In: SCIPIONI, A.; MANZARDO, A.; REN, J. (Ed.). **Hydrogen Economy (Second Edition)**. Second edition. s.l.: Academic Press, 2023. p. 459–485. ISBN 978-0-323-99514-6.

- RODRIGUES, G. S. **Modelos Dinâmicos Dirichlet**. 2011. 128 f. Dissertação (Mestrado em Estatística) — Departamento de Estatística, Instituto de Ciências Exatas, Universidade de Brasília, Brasília, 2011.
- SAHRAIE, E.; KAMWA, I.; MOEINI, A.; MOHSENI-BONAB, S. M. Component and system levels limitations in power-hydrogen systems: Analytical review. **Energy Strategy Reviews**, v. 54, p. 101476, 2024.
- SANTOS, A. Á. B.; CAMARGO, O.; BACK, A.; SILVA, F.; CATANI, F.; ESTANTE, F.; LIMA, G.; VIOLATO, G.; ARAÚJO, J.; ANDRADE, P. *et al.* Atlas eólico bahia. SECTI-SEINFRA-CIMATEC/SENAI, 2013.
- SCHNUELLE, C.; WASSERMANN, T.; FUHRLAENDER, D.; ZONDERVAN, E. Dynamic hydrogen production from pv wind direct electricity supply – modeling and techno-economic assessment. **International Journal of Hydrogen Energy**, s.l., v. 45, n. 55, p. 29938–29952, 2020.
- SERRANO-GUERRERO, X.; BRICEÑO-LEÓN, M.; CLAIRAND, J.-M.; ESCRIVÁ-ESCRIVÁ, G. A new interval prediction methodology for short-term electric load forecasting based on pattern recognition. **Applied Energy**, s.l., v. 297, p. 117173, 2021.
- SHUMWAY, R. H.; STOFFER, D. S. State space models. In: **Time Series Analysis and Its Applications: With R Examples**. 4th edition. ed. Cham: Springer International Publishing, 2017. p. 289–384. ISBN 978-3-319-52452-8.
- SOUZA, P. O. d. **Modelos dinâmicos para a distribuição Poisson Generalizada**. 2014. 154 f. Dissertação (Mestrado em Estatística) — Instituto de Ciências Exatas, Universidade de Brasília, Brasília, 2014.
- SUN, M.; FENG, C.; ZHANG, J. Multi-distribution ensemble probabilistic wind power forecasting. **Renewable Energy**, s.l., v. 148, p. 135–149, 2020.
- van der MEER, D.; WIDÉN, J.; MUNKHAMMAR, J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption. **Renewable and Sustainable Energy Reviews**, s.l., v. 81, p. 1484–1512, 2018.
- WANG, D.; MURATORI, M.; EICHMAN, J.; WEI, M.; SAXENA, S.; ZHANG, C. Quantifying the flexibility of hydrogen production systems to support large-scale renewable energy integration. **Journal of Power Sources**, s.l., v. 399, p. 383–391, 2018.
- WANG, J.; WEN, J.; WANG, J.; YANG, B.; JIANG, L. Coordinated scheduling of wind-solar-hydrogen-battery storage system for techno-economic-environmental optimization of hydrogen production. **Energy Conversion and Management**, v. 314, p. 118695, 2024.
- WEST, M.; HARRISON, J. **Bayesian Forecasting and Dynamic Models**. Berlin, Heidelberg: Springer-Verlag, 1997. ISBN 0387947256.
- XIE, Y.; LI, C.; LI, M.; LIU, F.; TAUKENOVA, M. An overview of deterministic and probabilistic forecasting methods of wind energy. **iScience**, s.l., v. 26, p. 105804, 2023.
- XU, X.; HU, W.; LIU, W.; DU, Y.; HUANG, Q.; CHEN, Z. Robust energy management for an on-grid hybrid hydrogen refueling and battery swapping station based on renewable energy. **Journal of Cleaner Production**, v. 331, p. 129954, 2022.

ZHANG, J.; YAN, J.; INFIELD, D.; LIU, Y.; LIEN, F. sang. Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and gaussian mixture model. **Applied Energy**, s.l., v. 241, p. 229–244, 2019.

ZHAO, X.; WEI, H.; LI, C.; ZHANG, K. A hybrid nonlinear forecasting strategy for short-term wind speed. **Energies**, s.l., v. 13, n. 7, 2020.

ZHENG, Y.; WANG, J.; YOU, S.; LI, X.; BINDNER, H. W.; MÜNSTER, M. Data-driven scheme for optimal day-ahead operation of a wind/hydrogen system under multiple uncertainties. **Applied Energy**, s.l., v. 329, p. 120201, 2023.

ZHENG, Y.; YOU, S.; BINDNER, H. W.; MÜNSTER, M. Data-driven scheme for optimal day-ahead operation of a wind/hydrogen system under multiple uncertainties. **Applied Energy**, s.l., v. 307, p. 118091, 2022.

ZHU, L.; YU, M.; WU, K.; DOU, Z.; ZHANG, C. Optimal operation strategy for power-to-hydrogen system considering dynamic efficiency. In: **2022 IEEE 6th Conference on Energy Internet and Energy System Integration (EI2)**. Chengdu: [s.n.], 2022. p. 836–842.