Universidade Federal de Juiz de Fora

Faculdade de Engenharia

Programa de Pós-Graduação em Engenharia Elétrica

Paulo Sérgio de Castro Nascimento

Desenvolvimento de uma Metodologia Integrada para Estimação e Previsão de Geração Eólica em Larga Escala para o Planejamento e a Operação de Sistemas Elétricos

Juiz de Fora

2026

Paulo Sérgio de Castro Nascimento

Desenvolvimento de uma Metodologia Integrada para Estimação e Previsão de Geração Eólica em Larga Escala para o Planejamento e a Operação de Sistemas Elétricos

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora, na área de concentração em Sistemas de Energia Elétrica, como requisito parcial para obtenção do título de Doutor em Engenharia Elétrica.

Orientador: André Luís Marques Marcato, D.Sc.

Juiz de Fora

2026

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF com os dados fornecidos pelo(a) autor(a)

Nascimento, Paulo Sérgio de Castro.
Desenvolvimento de uma Metodologia Integrada para Estimação e Previsão de Geração Eólica em Larga Escala para o Planejamento e a Operação de Sistemas Elétricos / Paulo Sérgio de Castro Nascimento. – 2026.
297 f. : il.

Orientador: André Luís Marques Marcato, D.Sc.
Tese (Doutorado) – Universidade Federal de Juiz de Fora, Faculdade de Engenharia. Programa de Pós-Graduação em Engenharia Elétrica, 2026.

1. Modelos de Previsão e Estimação. 2. Tratamento de Dados. 3. Previsão Eólica em Tempo Real. 4. Previsão Eólica de Curtíssimo e Curto Prazo . I. Marcato, André Luís Marques, orient. II. Título.

Paulo Sérgio de Castro Nascimento

Desenvolvimento de uma Metodologia Integrada para Estimação e Previsão de Geração Eólica em Larga Escala para o Planejamento e a Operação de Sistemas Elétricos

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Sistemas de Energia Elétrica

Aprovada em 24 de setembro de 2026.

BANCA EXAMINADORA

Prof. Dr. André Luis Marques Marcato - Orientador
Universidade Federal de Juiz de Fora

Prof. Dr. Fernando Ramos Martins
Universidade Federal de São Paulo

Prof. Dr. André Rodrigues Gonçalves
Instituto Nacional de Pesquisas Espaciais

Prof. Dr. João Alberto Passos Filho
Universidade Federal de Juiz de Fora

Prof. Dr. Bruno Henriques Dias
Universidade Federal de Juiz de Fora

Juiz de Fora, 31/08/2026.

Documento assinado eletronicamente por **BRUNO HENRIQUES DIAS**, **Professor(a)**, em 24/09/2026, às 12:39, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Joao Alberto Passos Filho**, **Professor(a)**, em 24/09/2026, às 12:39, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Andre Luis Marques Marcato**, **Professor(a)**, em 24/09/2026, às 12:39, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Andre Rodrigues Goncalves**, **Usuário Externo**, em 28/09/2026, às 10:53, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fernando Ramos Martins**, **Usuário Externo**, em 20/10/2026, às 17:37, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **3183859** e o código CRC **BC710323**.

Dedico este trabalho a todos meus familiares e amigos.

AGRADECIMENTOS

Ao Professor André Luís Marques Marcato pelas valiosas ideias e recomendações. Também por todos os ensinamentos passados, que contribuíram muito para elaboração deste trabalho.

Ao Professor José Luiz Rezende Pereira (In Memoriam) pelo incentivo para início nessa empreitada. Além do empenho durante todo o decorrer do curso de graduação e pós-graduação, trazendo enormes contribuições para minha formação acadêmica.

Aos Professores da pós-graduação pelos ensinamentos e dedicação nas disciplinas lecionadas. Especialmente ao Prof. André Marcato pela confiança e incentivos. Ao Operador Nacional do Sistema Elétrico (ONS), pela disponibilidade de utilização dos recursos e informações necessários a este trabalho.

Aos amigos da Universidade Federal de Juiz de Fora, pelo companheirismo ao longo destes anos de formação acadêmica, em especial aos amigos do programa de pós-graduação pelos conselhos técnicos e computacionais relacionados ao trabalho. Aos amigos do ONS pela disposição e contribuição de suas experiências na elaboração deste projeto.

A minha esposa, Ana Carolyne, por me incentivar a participar do processo seletivo de doutorado, assim como pela cumplicidade e suporte contínuo ao longo desta pesquisa.

A todos amigos e familiares da Cachoeira, Descoberto, São João Nepomuceno, Juiz de Fora, Rio de Janeiro e região, pela ajuda e incentivos ao longo destes anos dedicados ao curso de graduação e pós-graduação.

Este trabalho foi apoiado pela Universidade Federal de Juiz de Fora; pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), sob os processos nº 307843/2023-8 e 404859/2025-9; pela Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), sob o processo TEC APQ 02800/23; e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

"*Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution.*"

Albert Einstein

RESUMO

A busca por um futuro mais sustentável é cada vez mais necessária, e a participação das fontes renováveis na matriz energética dos países cresce continuamente. É inegável a importância dessa transformação energética; entretanto, o crescimento acentuado de fontes propensas à alta variabilidade e intermitência torna a operação dos sistemas elétricos mais complexa. Essa variabilidade gera custos operativos e complexidade, principalmente para os modelos de previsão, tanto de oferta quanto de demanda. Além disso, diferentemente das fontes convencionais, controladas pelos operadores, essas fontes são de certa forma controladas pela natureza, gerando incertezas na segurança energética e impactos nos sistemas elétricos. No Brasil, além da alta variabilidade e da dependência de fatores meteorológicos, essas fontes são não despacháveis, isto é, em condição normal de operação sua geração é diretamente associada ao recurso natural que chega à usina. Sobrepõe-se a esse quadro o crescimento expressivo do corte de geração (*curtailment*) observado nos últimos anos, sobretudo na região Nordeste, que faz a geração verificada deixar de refletir o recurso eólico disponível e perturba ainda mais as séries históricas utilizadas pelos modelos. Sendo assim, torna-se imprescindível a realização de previsões acuradas de sua geração para compor o portfólio otimizado de produção energética do SIN (Sistema Interligado Nacional). Tais previsões norteiam as decisões dos operadores em tempo real e são insumos importantes para o planejamento de curtíssimo prazo (Programação Diária da Operação Eletroenergética), além de úteis para o planejamento de curto e médio prazo. A maior acurácia na previsão dessas fontes possibilita melhor uso dos recursos, reduzindo a necessidade de reserva de potência operativa ou de geração termoelétrica adicional àquela definida por mérito econômico pelos modelos de otimização oficialmente em uso. Diante desse cenário, o objetivo principal deste trabalho é o desenvolvimento de modelos de previsão de geração eólica para o caso brasileiro, abrangendo desde o tempo real até um mês à frente, nos horizontes necessários à operação e ao planejamento do ONS (Operador Nacional do Sistema Elétrico) e incorporando informações dos principais modelos meteorológicos mundiais. Como etapas de suporte a esse objetivo, são propostos um tratamento e uma recomposição dos dados históricos de vento e geração, que integram medições de diferentes instituições do setor, e um modelo de estimação da geração potencial, capaz de reconstruir a geração na ausência de corte e de quantificar as perdas por *curtailment*. As metodologias foram aplicadas em larga escala, considerando mais de mil parques eólicos de grande porte conectados à rede de transmissão brasileira, um dos maiores sistemas do mundo em geração eólica centralizada, contribuindo para aumentar a confiabilidade da operação e do planejamento eletroenergético do SIN.

Palavras-chave: Modelos de Previsão e Estimação; Tratamento de Dados; Previsão Eólica em Tempo Real; Previsão Eólica de Curtíssimo e Curto Prazo.

ABSTRACT

The search for a more sustainable future is increasingly necessary, and the participation of renewable sources in the energy matrix of countries keeps growing. The importance of this energy transformation is undeniable; however, the sharp growth of sources prone to high variability and intermittency makes the operation of electrical systems more complex. This variability generates operational costs and complexity, mainly for forecast models, both supply and demand. Furthermore, unlike conventional sources, controlled by operators, these energy sources are somewhat controlled by nature, generating uncertainty in energy security and impacts on electrical systems. In Brazil, in addition to the high variability and the dependence on meteorological factors, these sources are non-dispatchable, that is, under normal operating conditions their generation is directly associated with the natural resource that arrives at the plant. Superimposed on this scenario is the significant growth of generation curtailment observed in recent years, especially in the Northeast region, which causes the verified generation to no longer reflect the available wind resource and further disturbs the historical time series used by the models. Therefore, it is essential to produce accurate forecasts of their generation to compose the optimized energy production portfolio of the SIN (National Interconnected System). Such forecasts guide operators' decisions in real time and are important inputs for very short-term planning (Daily Scheduling of Electroenergetic Operations), as well as useful for short- and medium-term planning. Greater accuracy in forecasting these sources enables better use of resources, reducing the need for operating power reserves or thermoelectric generation beyond that defined by economic merit by the optimization models officially in use. Given this scenario, the main objective of this work is the development of wind power generation forecasting models for the Brazilian case, covering horizons from real time up to one month ahead, as required for the operation and planning carried out by the ONS (National Electric System Operator) and incorporating information from the main numerical weather prediction models worldwide. As supporting stages for this objective, a treatment and reconstruction of the historical wind and generation data is proposed, integrating measurements from different institutions of the sector, together with a potential generation estimation model, able to reconstruct the generation that would occur in the absence of curtailment and to quantify the associated losses. The methodologies were applied at large scale, considering more than one thousand utility-scale wind farms connected to the Brazilian transmission grid, one of the largest centralized wind power systems in the world, contributing to increase the reliability of the operation and the electroenergetic planning of the SIN.

Keywords: Forecasting and Estimation Models; Data Processing; Real-Time Wind Forecast; Very Short- and Short-Term Wind Forecast.

LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma simplificado do modelo proposto, com os quatro blocos encadeados. 35
Figura 2 – Evolução da capacidade instalada no SIN. 37
Figura 3 – Distribuição geográfica das usinas eólicas no território brasileiro. 38
Figura 4 – Panorama das famílias de modelos de previsão de geração eólica por horizonte de previsão. 46
Figura 5 – Classificação dos modelos de previsão de geração eólica segundo os três eixos. 46
Figura 6 – Processo de limpeza e composição dos dados históricos, incluindo as estimações cruzadas entre geração e vento. 63
Figura 7 – Dados reunidos para o desenvolvimento dos modelos de previsão eólica. . . 64
Figura 8 – Identificação de não conformidade entre a geração horária do ONS e CCEE. 68
Figura 9 – Identificação de não conformidade com alta correlação entre curvas de geração do ONS e da CCEE. 68
Figura 10 – Processo de montagem do histórico de geração verificada, com preenchimento sequencial das falhas. 72
Figura 11 – Processo de montagem do melhor histórico de vento verificado, com preenchimento sequencial das falhas. 75
Figura 12 – Dispersão V $\times$ P para identificação de dados espúrios de uma usina, combinando as fontes de geração (ONS e CCEE) e de vento (ONS e EPE). 79
Figura 13 – Identificação e seleção dos dados verificados de geração, antes e após a limpeza, ao longo de cinco anos; dias com problema em cor mais escura. . . 80
Figura 14 – Identificação e seleção dos dados verificados de vento, antes e após a limpeza, ao longo de cinco anos; dias com problema em cor mais escura. 81
Figura 15 – Histórico consistido de geração e vento de uma usina, após limpeza, composição e preenchimento. 82
Figura 16 – Identificação de dados espúrios e resultado do melhor histórico, em dois dias com problemas nos dados de vento (valores congelados e dados espúrios). . 82
Figura 17 – Associação do dado a ser filtrado ao padrão similar, para uma curva diária de vento com valores congelados. 84
Figura 18 – Regressão linear através do dado a ser filtrado e o padrão ajustado, utilizada na identificação de dados espúrios. 85
Figura 19 – Identificação de dados espúrios pelos erros de ajuste e seus limites, e resultado após a remoção dos pontos espúrios. 86
Figura 20 – Previsão do Filtro de Kalman com regiões de falha (com e sem inovação). . 88
Figura 21 – Classificação de um *cluster* com poucas falhas: antes e após o pré-tratamento, em séries diárias de vento. 89
Figura 22 – Classificação de um *cluster* com muitas falhas: antes e após o pré-tratamento, em séries diárias de vento. 89

Figura 23 – Correção dos dados de geração verificada para dois dias distintos. 90
Figura 24 – Correção dos dados de vento verificado para dois dias distintos. 91
Figura 25 – *Cluster* com poucas falhas: antes do pré-tratamento e após o preenchimento final, em séries diárias de vento. 92
Figura 26 – *Cluster* com muitas falhas: antes do pré-tratamento e após o preenchimento final, em séries diárias de vento. 92
Figura 27 – Estrutura do modelo de estimação da geração. 96
Figura 28 – Dispersão Vento $\times$ Potência antes e após o ajuste pela potência disponível para uma usina eólica do SIN. 97
Figura 29 – Evolução temporal da geração verificada e prevista sem a correção pela potência disponível. 99
Figura 30 – Evolução temporal da geração verificada e prevista com a correção pela potência disponível. 100
Figura 31 – Curva de potência de uma turbina eólica: comportamento teórico e real. . . 100
Figura 32 – Representação esquemática da relação Vento $\times$ Potência teórica. 101
Figura 33 – Curva genérica Vento $\times$ Potência com limites operacionais. 102
Figura 34 – Comparação curvas V $\times$ P estimadas pela sigmoide e pela regressão logística. 103
Figura 35 – Aplicação das curvas limites, com a dispersão V $\times$ P antes e após a eliminação dos dados incompatíveis. 105
Figura 36 – Dispersão Vento $\times$ Potência com os limites inferiores e superiores, evidenciando outras "dispersões" associadas a restrições operativas. 106
Figura 37 – Comparação entre curvas V $\times$ P estimadas sem e com a inserção de pontos fictícios. 107
Figura 38 – Curvas V $\times$ P de parques eólicos com saturação e decaimento na geração. . 109
Figura 39 – Comparação entre regressão logística e redes neurais para o ajuste com a média, nos conjuntos de treino, validação e teste. 110
Figura 40 – Comparação das dispersões V $\times$ P utilizando o vento de 1, 3 e 14 naceles (cabines dos aerogeradores). 111
Figura 41 – Comparação curvas V $\times$ P por vento médio de 3 e 14 pontos de medição. . . 112
Figura 42 – Comparação dos erros para diferentes formas de cálculo do vento equivalente, nos conjuntos de treinamento e teste. 113
Figura 43 – Estimação da curva de potência para 0h e 6h, para uma usina do interior da Bahia. 114
Figura 44 – Estimação da curva de potência para 12h e 18h, para uma usina do interior da Bahia. 115
Figura 45 – Estrutura dos modelos desenvolvidos neste trabalho. 119
Figura 46 – Modelos meteorológicos utilizados na previsão de geração eólica, comparando horizonte de previsão e resoluções temporal e espacial. 122

Figura 47 – Dispersão do erro (verificado − previsto) ao longo do horizonte de previsão — ponto Brumado II 230. 124
Figura 48 – Dispersão do erro (verificado − previsto) ao longo do horizonte de previsão — ponto Açu II 230. 125
Figura 49 – Dispersão do erro (verificado − previsto) por ponto de análise para os três modelos meteorológicos — primeiras 24 horas de previsão. 126
Figura 50 – Dispersão do erro (verificado − previsto) agregada por estado para os três modelos meteorológicos — primeiras 24 horas de previsão. 127
Figura 51 – Dessazonalização de séries de vento verificado e previsto, para o modelo GEFS. 130
Figura 52 – Regressão ajustada para os intervalos das 00:00 e 09:30 do dia, entre vento previsto e verificado. 132
Figura 53 – Regressão ajustada para os intervalos das 19:00 e 02:30 do dia, com ajuste pobre, entre vento previsto e verificado. 133
Figura 54 – Impacto da correção de vento nas curvas de vento previsto — modelos ECMWF e GFS, para um ponto de conexão do Nordeste. 134
Figura 55 – Características do *boxplot*. 135
Figura 56 – Erros brutos do vento previsto e corrigido, pelo GFS, para uma usina. 135
Figura 57 – Evolução da capacidade instalada de um determinado ponto estabilizada, com geração total e média das usinas do ponto. 137
Figura 58 – Evolução da capacidade instalada de um determinado ponto em variação, com geração total e média das usinas do ponto. 138
Figura 59 – Regressão ajustada entre geração média e geração total, usada para reconverter a previsão ao montante total do ponto. 139
Figura 60 – Aplicação do modelo de combinação — subsistema Nordeste (séries, RMSE histórico e pesos por horizonte). 147
Figura 61 – Estrutura do modelo de curtíssimo prazo. 150
Figura 62 – Geração verificada e previsão para o grupo e para as usinas individuais. . . . 151
Figura 63 – Separação das previsões por usinas utilizando fator proporcional à potência instalada. 151
Figura 64 – Separação das previsões por usinas com violação da potência instalada. . . . 153
Figura 65 – Separação das previsões por usinas após a correção pela geração verificada. 155
Figura 66 – Separação das previsões por usinas após a correção pela geração prevista. . 156
Figura 67 – Separação das previsões por usinas após a correção simultânea pela geração verificada e pela previsão. 157
Figura 68 – Geração eólica média mensal do Nordeste dos últimos anos, estimada pela resolução e verificada. 159
Figura 69 – Estrutura do modelo de curto prazo. 160
Figura 70 – Distribuição do erro da previsão diária — GEFS bruto (sem correção). . . . 161

Figura 71 – Perfis da geração e velocidade do vento verificadas e previstas, e curva V × P do conjunto de usinas analisado. 163
Figura 72 – Distribuição do erro da previsão diária — GEFS após a correção do viés. . . 164
Figura 73 – Ajuste da regressão linear para inclusão da geração das usinas do tipo III — Subsistema Nordeste. 165
Figura 74 – Fluxograma do modelo de previsão de geração eólica em tempo real. 167
Figura 75 – Fluxograma do modelo proposto, em sua versão detalhada. 175
Figura 76 – Dados de vento brutos, filtrados e completados, juntamente com a geração eólica. 178
Figura 77 – Percentual de falhas nas séries de geração de energia por usina em cada etapa do processo de tratamento de dados. 178
Figura 78 – Percentual de falhas nas séries de velocidade de vento por usina em cada etapa do processo de tratamento de dados. 179
Figura 79 – Dispersão entre velocidade do vento e potência gerada, com os dados reais de operação e os registros sob restrição operacional. 181
Figura 80 – Dispersão entre geração verificada e geração estimada com boa identificação de restrições. 182
Figura 81 – Dispersão entre geração verificada e geração estimada com identificação insatisfatória de restrições. 183
Figura 82 – Correção de viés da geração estimada para o subsistema Sul: séries temporais e histogramas do erro médio antes e depois da correção. 185
Figura 83 – Correção de viés da geração estimada para o subsistema Nordeste: séries temporais e histogramas do erro médio antes e depois da correção. 186
Figura 84 – Geração estimada e verificada com indicação de cortes para o subsistema Sul. 187
Figura 85 – Geração estimada e verificada com indicação de cortes para o subsistema Nordeste. 188
Figura 86 – Evolução mensal da geração estimada e verificada com indicação de cortes para o subsistema Sul. 190
Figura 87 – Evolução mensal da geração estimada e verificada com indicação de cortes para o subsistema Nordeste. 191
Figura 88 – Evolução mensal da geração estimada e verificada com indicação de cortes para o SIN. 192
Figura 89 – Dispersão entre geração verificada e geração estimada em resolução semi-horária para os subsistemas Sul e Nordeste. 193
Figura 90 – Vento médio por ponto na Região Sul — verificação e previsões (primeiras 24 horas). 195
Figura 91 – Vento médio por ponto na Região Nordeste — verificação e previsões (primeiras 24 horas). 196

Figura 92 – Vento médio por ponto na Região Sul — verificação e média corrigida (primeiras 24 horas). 197
Figura 93 – Vento médio por ponto na Região Nordeste — verificação e média corrigida (primeiras 24 horas). 197
Figura 94 – Viés médio do vento do GEFS por estado (primeiros 8 dias) — previsões brutas. 198
Figura 95 – Viés médio do vento do GEFS por estado (primeiros 8 dias) — previsões corrigidas. 198
Figura 96 – Geração prevista, verificada tratada e verificada real com indicação de cortes para o subsistema Sul. 201
Figura 97 – Geração prevista, verificada tratada e verificada real com indicação de cortes para o subsistema Nordeste. 201
Figura 98 – Evolução mensal da geração prevista, verificada tratada e verificada real com indicação de cortes para o subsistema Sul. 204
Figura 99 – Evolução mensal da geração prevista, verificada tratada e verificada real com indicação de cortes para o subsistema Nordeste. 204
Figura 100 – Evolução mensal da geração prevista, verificada tratada e verificada real com indicação de cortes para o SIN. 205
Figura 101 – Dispersão entre geração verificada tratada e geração prevista em resolução semi-horária para os subsistemas Sul e Nordeste. 206
Figura 102 – Comparação dos erros horários percentuais, normalizados pela potência instalada, das previsões de geração eólica obtidas pelos modelos GFS, GEFS e ECMWF ao longo das primeiras 24 horas de previsão. 208
Figura 103 – Distribuição dos erros horários, em MWm, das previsões de geração eólica obtidas pelos modelos GFS, GEFS e ECMWF, agrupados por faixas do horizonte de previsão. 209
Figura 104 – Comparação dos erros médios diários percentuais, normalizados pela potência instalada, das previsões de geração eólica obtidas pelos modelos GFS, GEFS, ECMWF e pela combinação *Comb*, ao longo dos primeiros 15 dias do horizonte de previsão. 210
Figura 105 – Densidade dos erros horários, em MWm, para os modelos GFS, GEFS, ECMWF e Combinado, considerando o conjunto agregado das 24 primeiras horas do horizonte de previsão. 211
Figura 106 – Erro de previsão do modelo, em termos percentuais e normalizado pela potência instalada, para os subsistemas Sul e Nordeste ao longo do horizonte de previsão de 35 dias. 212
Figura 107 – Distribuição dos erros percentuais, normalizados pela potência instalada, por passo semanal e patamar de carga para o subsistema Sul, comparando o modelo proposto e o modelo de resolução. 216

Figura 108 – Distribuição dos erros percentuais, normalizados pela potência instalada, por passo semanal e patamar de carga para o subsistema Nordeste, comparando o modelo proposto e o modelo de resolução. 216
Figura 109 – Comparação dos erros médios percentuais (nMAPE, normalizado pela potência instalada) por passo semanal e patamar de carga para o subsistema Sul, entre o modelo proposto e o modelo de resolução. 217
Figura 110 – Comparação dos erros médios percentuais (nMAPE, normalizado pela potência instalada) por passo semanal e patamar de carga para o subsistema Nordeste, entre o modelo proposto e o modelo de resolução. 218
Figura 111 – Comparação dos erros médios percentuais (nMAPE, normalizado pela potência instalada) por passo semanal e patamar de carga para o SIN, entre o modelo proposto e o modelo de resolução. 219
Figura 112 – Distribuição dos erros de previsão eólica de alguns países. 220
Figura 113 – Distribuição dos erros horários normalizados e Normal ajustada para os modelos GFS, GEFS, ECMWF e Combinado, considerando o horizonte D+1 (25 a 48 horas à frente). 221
Figura 114 – Comportamento conjunto de múltiplos modelos no subsistema Sul para o dia 08/01/2025. 222
Figura 115 – Comportamento conjunto de múltiplos modelos no subsistema Nordeste para o dia 08/01/2025. 222
Figura 116 – MAE normalizado pela geração máxima do período em função do horizonte de previsão no subsistema Sul. 224
Figura 117 – MAE normalizado pela geração máxima do período em função do horizonte de previsão no subsistema Nordeste. 224
Figura 118 – Distribuição dos pesos dos modelos na combinação por inverso do erro (1/MAE) e frequência de exclusão de cada modelo, para os agrupamentos de usinas do SIN e o horizonte de até 9 horas. 225
Figura 119 – Peso médio de cada modelo na combinação por inverso do erro (1/MAE) em função do horizonte de previsão, para os agrupamentos de usinas do SIN. . . 226
Figura 120 – Distribuição do erro normalizado por modelo nos horizontes de 30 minutos, 3 horas e 24 horas no subsistema Sul. 228
Figura 121 – Distribuição do erro normalizado por modelo nos horizontes de 30 minutos, 3 horas e 24 horas no subsistema Nordeste. 228
Figura 122 – Exemplo 1 de aplicação do MAPE e nMAPE, com dois dias hipotéticos de mesmo erro absoluto. 266
Figura 123 – Exemplo 2 de aplicação do MAPE e nMAPE, com MAPE maior para um erro absoluto menor. 267
Figura 124 – Exemplo de erro para um período do Sul, com erros percentuais elevados pelo MAPE apesar de erros baixos em MW. 268

Figura 125 – Comportamento da geração de energia eólica na região Nordeste, decomposta em tendência, sazonalidades e série residual. 270
Figura 126 – Comportamento da geração de energia eólica na região Sul, decomposta em tendência, sazonalidades e série residual. 271
Figura 127 – Métricas de erro normalizadas pela amplitude da série (nRMSE e nMAE) por usina eólica, ordenadas pelo nRMSE. 272
Figura 128 – Distribuição do erro médio normalizado (nME) por usina eólica, normalizado pela amplitude da série. 272
Figura 129 – Desempenho do modelo para cada grupo de parques eólicos, normalizado pela amplitude da série. 273
Figura 130 – Distribuição do erro médio normalizado (nME) por agrupamento de usinas, normalizado pela amplitude da série. 274
Figura 131 – Distribuição do erro normalizado de todos os pontos dos agrupamentos de usinas. 275
Figura 132 – Gráfico de floresta dos erros médios dos grupos de parques eólicos, com seus intervalos de confiança. 276
Figura 133 – Dispersão do erro de vento ao longo do horizonte para o ponto Brumado II 230, antes e após a correção. 278
Figura 134 – Dispersão do erro de vento ao longo do horizonte para o ponto Açu II 230, antes e após a correção. 279
Figura 135 – Dispersão do erro de vento do ECMWF por ponto de análise, antes e após a correção. 279
Figura 136 – Dispersão do erro de vento do GFS por ponto de análise, antes e após a correção. 280
Figura 137 – Dispersão do erro de vento do GEFS por ponto de análise, antes e após a correção. 280
Figura 138 – Dispersão do erro de vento do ECMWF agregada para o Rio Grande do Norte, antes e após a correção. 282
Figura 139 – Dispersão do erro de vento do GFS agregada para o Rio Grande do Norte, antes e após a correção. 282
Figura 140 – Dispersão do erro de vento do GEFS agregada para o Rio Grande do Norte, antes e após a correção. 283
Figura 141 – Dispersão do erro de vento do ECMWF agregada para a Bahia, antes e após a correção. 284
Figura 142 – Dispersão do erro de vento do GFS agregada para a Bahia, antes e após a correção. 284
Figura 143 – Dispersão do erro de vento do GEFS agregada para a Bahia, antes e após a correção. 285

Figura 144 – Dispersão do erro de vento do ECMWF agregada por estado, antes e após a correção. 285
Figura 145 – Dispersão do erro de vento do GFS agregada por estado, antes e após a correção. 286
Figura 146 – Dispersão do erro de vento do GEFS agregada por estado, antes e após a correção. 286
Figura 147 – Métricas de erro normalizadas pela amplitude da série (nRMSE e nMAE) da previsão de geração com o ECMWF para o primeiro dia de previsão, por agrupamento de usinas. 288
Figura 148 – Comparação do nRMSE, normalizado pela amplitude da série, entre o modelo de estimação e a previsão de geração com o ECMWF para o primeiro dia de previsão. 288
Figura 149 – Comparação do nRMSE, normalizado pela amplitude da série, da previsão de geração entre os modelos GFS, ECMWF e GEFS para o primeiro dia de previsão. 289
Figura 150 – Distribuição do erro normalizado de previsão semi-horário para os modelos GFS, ECMWF e GEFS. 290
Figura 151 – Previsões de geração baseadas nos modelos NWP individuais (ECMWF, GFS, GEFS) e sua combinação (NWP Combinado) no ponto Acaraú II para o dia 08/01/2025. 291
Figura 152 – Leque de previsões da Persistência Inteligente convergindo ao NWP Combinado no ponto Acaraú II para o dia 08/01/2025. 291
Figura 153 – Leque de previsões da Rede Neural (MLP) tendendo ao ECMWF nos horizontes finais no ponto Acaraú II para o dia 08/01/2025. 292
Figura 154 – Previsão do SARIMA em comparação ao NWP Combinado e ao verificado no ponto Acaraú II para o dia 08/01/2025. 293
Figura 155 – Previsão do SARIMA em comparação ao NWP Combinado e ao verificado no ponto Atlântica para o dia 08/01/2025. 293
Figura 156 – Comportamento conjunto de múltiplos modelos no ponto Acaraú II para o dia 08/01/2025. 294
Figura 157 – Comportamento conjunto de múltiplos modelos no ponto Atlântica para o dia 08/01/2025. 294
Figura 158 – MAE normalizado pela geração máxima do período em função do horizonte de previsão no ponto Acaraú II. 295
Figura 159 – MAE normalizado pela geração máxima do período em função do horizonte de previsão no ponto Atlântica. 295
Figura 160 – Distribuição do erro normalizado por modelo nos horizontes de 30 minutos, 3 horas e 24 horas no ponto Acaraú II. 296
Figura 161 – Distribuição do erro normalizado por modelo nos horizontes de 30 minutos, 3 horas e 24 horas no ponto Atlântica. 297

LISTA DE TABELAS

Tabela 1 – Síntese da literatura quanto à abrangência e à maturidade das abordagens, e posicionamento da tese. 58
Tabela 2 – Síntese da literatura quanto ao escopo do problema abordado, e posicionamento da tese. 59
Tabela 3 – Proveniência, discretização e disponibilidade dos dados utilizados. 65
Tabela 4 – Papel de cada base de dados observados nas etapas de limpeza e composição das séries históricas. 94
Tabela 5 – Parâmetros do modelo de combinação por inverso do erro quadrático médio. 146
Tabela 6 – Síntese dos modelos do *ensemble* de tempo real e das combinações. 174
Tabela 7 – Métricas de erro normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, por subsistema antes da correção de viés. 189
Tabela 8 – Métricas de erro normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, por subsistema após a correção de viés. 189
Tabela 9 – Métricas de erro normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio do modelo de previsão por subsistema. 202
Tabela 10 – Métricas de erro do modelo combinado para o subsistema Sul, por horizonte de previsão (D0 a D21), normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, exceto o $\text{nMAPE}_{\text{Pinst}}$, normalizado pela potência instalada. 213
Tabela 11 – Métricas de erro do modelo combinado para o subsistema Nordeste, por horizonte de previsão (D0 a D21), normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, exceto o $\text{nMAPE}_{\text{Pinst}}$, normalizado pela potência instalada. 214
Tabela 12 – Métricas de erro do modelo combinado para o agregado do SIN, por horizonte de previsão (D0 a D21), normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, exceto o $\text{nMAPE}_{\text{Pinst}}$, normalizado pela potência instalada. 215
Tabela 13 – Síntese das contribuições e evidências associadas ao atendimento dos objetivos da tese. 231
Tabela 14 – Janela de treinamento e principais hiperparâmetros dos modelos calibráveis. 253
Tabela 15 – Regime dominante da persistência inteligente por faixa de horizonte. 256
Tabela 16 – *Features* de base dos modelos de aprendizado de máquina (instante t). . . . 259

LISTA DE ABREVIATURAS E SIGLAS

ACR	Ambiente de Contratação Regulada
AEMO	*Australian Energy Market Operator*
AMA	Sistema de Acompanhamento de Medições Anemométricas
ANA	Agência Nacional de Águas
ANEEL	Agência Nacional de Energia Elétrica
ANN	*Artificial Neural Network*
ARIMA	*Autoregressive Integrated Moving Average*
BAM	*Brazilian Atmospheric Model*
CCEE	Câmara de Comercialização de Energia Elétrica
CEPEL	Centro de Pesquisas de Energia Elétrica
CPTEC	Centro de Previsão de Tempo e Estudos Climáticos
DECOMP	Modelo de Determinação da Coordenação da Operação a Curto Prazo
DESSEM	Modelo de Despacho Hidrotérmico de Curto/Curtíssimo Prazo
ECMWF	*European Centre for Medium-Range Weather Forecasts*
ELM	*Extreme Learning Machine*
EMS	*Energy Management System*
EPE	Empresa de Pesquisa Energética
ERCOT	*Electric Reliability Council of Texas*
Eta	Modelo Atmosférico de Mesoescala
ETS	*Error-Trend-Seasonal* (suavização exponencial)
GA	*Genetic Algorithm*
GAN	*Generative Adversarial Network*
GEFS	*Global Ensemble Forecast System*
GFS	*Global Forecast System*
GMM	*Gaussian Mixture Model* (Modelo de Mistura de Gaussianas)

GRNN	*General Regression Neural Network*
HRES	*High Resolution*
IFS	*Integrated Forecasting System*
IQR	*Interquartile Range* (intervalo interquartil)
ISO	*Independent System Operator*
LOESS	*Locally Estimated Scatterplot Smoothing*
LSTM	*Long Short-Term Memory*
MAE	*Mean Absolute Error*
MAPE	*Mean Absolute Percentage Error*
ME	*Mean Error*
MHG	Melhor Histórico de Geração
MHV	Melhor Histórico de Vento
MLP	*Multi Layer Perceptron*
MMGD	Micro e Minigeração Distribuída
MOC	*Model Output Calibration*
MOS	*Model Output Statistics*
MPCP	Modelo de Planejamento de Curto Prazo
MPCTP	Modelo de Planejamento de Curtíssimo Prazo
MPMP	Modelo de Planejamento de Médio Prazo
MSE	*Mean Squared Error*
NCEP	*National Centers for Environmental Prediction*
NEWAVE	Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo
nMAE	*normalized Mean Absolute Error* (erro absoluto médio normalizado)
nMAPE	*normalized Mean Absolute Percentage Error* (erro percentual absoluto médio normalizado)
nME	*normalized Mean Error* (erro médio normalizado)

nRMSE	*normalized Root Mean Square Error* (raiz do erro quadrático médio normalizada)
NOAA	*National Oceanic and Atmospheric Administration*
NSE	*Nash-Sutcliffe Efficiency*
NWP	*Numerical Weather Prediction* (Previsão Numérica do Tempo)
ONS	Operador Nacional do Sistema Elétrico
PCH	Pequena Central Hidrelétrica
PI	*Plant Information*
PMO	Programa Mensal da Operação
RBFN	*Radial Basis Function Network*
REE	*Red Eléctrica de España*
REGER	Rede de Gerenciamento de Energia
REN	Resolução Normativa
RF	*Random Forest*
RMSE	*Root Mean Square Error*
RNA	Rede Neural Artificial (equivalente a ANN)
SARIMA	*Seasonal Autoregressive Integrated Moving Average*
SCADA	*Supervisory Control and Data Acquisition*
SES	*Simple Exponential Smoothing*
SIN	Sistema Interligado Nacional
SMF	Sistema de Medição para Faturamento
SSC	Sistema de Supervisão e Controle
SVM	*Support Vector Machine*
VRE	*Variable Renewable Energy* (energia renovável variável)
WRF	*Weather Research and Forecasting*
XGBoost	*Extreme Gradient Boosting*

LISTA DE SÍMBOLOS

Convenção: a natureza da grandeza (verificada, estimada ou prevista) e os índices (usina, grupo, intervalo, fonte de dados etc.) são indicados no subscrito, separados por vírgula, na ordem *natureza, índices* — por exemplo, $G_{v,gr,j}$ denota a geração verificada do grupo gr no intervalo j.

G_i	Geração, em que o índice i pode assumir v (verificada), e (estimada) ou p (prevista), resultando em G_v, G_e e G_p, respectivamente
$G_{v,t}$	Geração verificada no instante t
$G_{e,t}$	Geração estimada no instante t
$G_{p,t}$	Geração prevista no instante t (também empregada para a geração prevista total do ponto)
$G_{v,gr,j}$	Geração verificada total do grupo gr no intervalo j
$G_{v,i,j}$	Geração verificada da usina i no intervalo j
$G_{p,gr,k}$	Geração prevista do grupo gr no intervalo k
$G_{v,\mathrm{CCEE}}$	Geração verificada pela CCEE
G_{aj}	Geração ajustada pela disponibilidade de potência
NG_v	Geração normalizada
$G_{\min}$	Geração verificada mínima
$G_{\max}$	Geração verificada máxima
P_{inst}	Potência instalada
$P_{inst,i}$	Potência instalada da usina i
$P_{\min}$	Potência mínima (assíntota inferior da curva V $\times$ P)
$P_{\max}$	Potência máxima (assíntota superior da curva V $\times$ P; também a potência disponível máxima no ajuste por disponibilidade)
P_{disp}	Potência disponível informada ao ONS
V_i	Vento, em que o índice i pode assumir v (verificado), e (estimado) ou p (previsto), resultando em V_v, V_e e V_p, respectivamente
V_{pa}	Vento previsto ajustado

$V_{v,\text{ONS}}$	Vento verificado do ONS (sistema PI)
$V_{v,\text{EPE}}$	Vento verificado da EPE
V_{iss}	Vento sem sazonalidade, em que o índice i pode assumir v (verificado), p (previsto) ou a (ajustado), resultando em V_{vss}, V_{pss} e V_{ass}, respectivamente
V_{mid}	Velocidade de vento no ponto de inflexão da curva V $\times$ P
NV_v	Vento normalizado
V $\times$ P	Curva de Vento $\times$ Potência
α_a	Coeficiente angular do ajuste das previsões de vento
β_a	Coeficiente linear do ajuste das previsões de vento
α_k	Coeficiente angular da regressão por horizonte k
α_{mt}	Coeficiente angular da regressão entre geração média e total
β_{mt}	Coeficiente linear da regressão entre geração média e total
α_{gr}	Coeficiente angular da regressão entre grupos de usinas
β_{gr}	Coeficiente linear da regressão entre grupos de usinas
$\varphi_{v,\Omega,k}$	Fator de redução, a partir da geração verificada, do conjunto de usinas Ω no intervalo k
$\varphi_{p,\Omega,k}$	Fator de redução, a partir da geração prevista, do conjunto de usinas Ω no intervalo k
$f_{i,j}$	Fator de separação da usina i no intervalo j
θ_t	Vetor de parâmetros (estado) do modelo de Kalman no instante t
Y_t	Vetor de observações no instante t
F_t, G_t	Matrizes de observação e de evolução, respectivamente, do modelo de Kalman
V_t, W_t	Matrizes de covariância do modelo de Kalman
μ_t	Componente de nível do modelo de Kalman
β_t	Componente de tendência do modelo de Kalman
σ_X	Desvio-padrão da variável X
$\overline{X}$	Valor médio da variável X

SUMÁRIO

1 – INTRODUÇÃO **27**

1.1 Contextualização 27
1.2 Motivação e justificativa 28
1.3 Objetivos do trabalho 30
1.4 Publicações decorrentes 30
1.5 Principais contribuições do trabalho 32
1.5.1 Contribuições gerais 32
1.5.2 Contribuições específicas 32
1.6 Estrutura do trabalho 34

2 – REVISÃO BIBLIOGRÁFICA **36**

2.1 A geração eólica no Brasil e no mundo 36
2.2 Aspectos regulatórios da previsão de energia eólica 39
2.2.1 Sistemas de previsão centralizados vs. descentralizados 39
2.2.2 Experiência internacional 40
2.3 Modelos para tratamento de dados 41
2.4 Modelos para ajuste de previsões numéricas de tempo 42
2.5 Modelos para previsão de geração eólica 45
2.5.1 Classificação dos modelos 45
2.5.2 Combinação de previsões 49
2.5.3 Posicionamento do trabalho 49
2.5.4 Desenvolvimentos por ISOs e instituições de pesquisa 49
2.6 *Curtailment* na geração eólica 51
2.6.1 Práticas de corte de geração em escala global 51
2.6.2 Dinâmica do corte de geração no sistema elétrico brasileiro 52
2.6.3 Abordagens para estimação de geração eólica 53
2.7 Representação eólica na cadeia de planejamento da operação do SIN 54
2.7.1 Planejamento e programação da operação 54
2.7.2 Representação das usinas eólicas no horizonte de curto prazo 56
2.8 Desafios na comparabilidade de modelos 56
2.9 Considerações finais do capítulo 57

3 – MODELOS PARA TRATAMENTO DE DADOS **62**

3.1 Dados de entrada 63
3.2 Limpeza de dados históricos 66
3.2.1 Limpeza dos dados de geração verificada 66

3.2.2 Limpeza dos dados de vento verificado 70
3.3 Composição e preenchimento dos históricos de geração 71
3.3.1 Composição dos históricos de geração 71
3.3.2 Preenchimento dos históricos de geração 72
3.4 Composição e preenchimento dos históricos de vento 74
3.4.1 Composição dos históricos de vento 75
3.4.2 Preenchimento dos históricos de vento 76
3.5 Aplicação dos processos de tratamento 78
3.6 Modelo estatístico para identificação de dados espúrios e preenchimento de lacunas 81
3.6.1 Identificação dos dados espúrios 83
3.6.2 Modelo de correção de dados faltantes 86
3.6.3 Aplicação do processo de classificação 88
3.6.4 Processo de preenchimento de dados faltantes 90
3.7 Considerações finais do capítulo 93

4 – MODELO PARA ESTIMAÇÃO DE GERAÇÃO EÓLICA 95
4.1 Tratamentos da geração para períodos de *curtailment* 95
4.2 Ajuste da geração com as informações de potência disponível 96
4.3 Metodologias para estimação das curvas Vento × Potência 98
4.4 Processo de refinamento da dispersão Vento × Potência das usinas 104
4.5 Estimação das curvas Vento × Potência 105
4.6 Metodologias alternativas para elaboração das curvas V × P 109
4.7 Processos de modelagem das curvas Vento × Potência 113
4.8 Considerações finais do capítulo 116

5 – MODELO PARA PREVISÃO DE GERAÇÃO EÓLICA 118
5.1 Considerações iniciais 118
5.2 Estrutura do modelo proposto 119
5.3 Modelagem das previsões numéricas de vento 120
5.3.1 Modelos Numéricos de Tempo avaliados 120
5.3.2 Avaliação dos insumos meteorológicos utilizados pelo modelo 121
5.3.3 Metodologia utilizada para ajuste das previsões de vento 128
5.4 Criação dinâmica das curvas Vento × Potência 136
5.5 Modelo final de conversão de vento em potência 141
5.6 Modelo de combinação de previsões 142
5.6.1 Motivação e contexto 142
5.6.2 Formulação geral 142
5.6.3 Cálculo dos pesos pelo inverso do erro quadrático médio 142
5.6.4 Correção dos pesos na presença de previsões ausentes 144
5.6.5 Diagrama do fluxo metodológico 145

5.6.6 Propriedades do método . . . 145
5.6.7 Parâmetros do modelo . . . 145
5.6.8 Aplicação do modelo . . . 146
5.7 Interpolação horária das previsões de geração . . . 148
5.8 Particularidades da metodologia de previsão de geração de curtíssimo prazo . . . 149
5.8.1 Metodologia de separação das previsões por usinas . . . 150
5.9 Particularidades da metodologia de previsão de geração de curto prazo . . . 158
5.9.1 Estrutura básica do modelo . . . 158
5.9.2 Metodologia proposta para o bloco 1 . . . 160
5.9.3 Estimativa da geração eólica para o bloco 2 . . . 163
5.9.4 Estimativa da geração eólica para o bloco 3 . . . 165
5.10 Modelo de previsão de geração em tempo real . . . 166
5.10.1 Características principais . . . 167
5.10.2 Protocolo de treinamento e simulação . . . 168
5.10.3 Modelos do *ensemble* . . . 169
5.10.4 Estratégias de combinação . . . 172
5.10.5 Síntese dos modelos . . . 173
5.11 Considerações finais do capítulo . . . 173

6 – RESULTADOS **177**
6.1 Resultados do processo de tratamento dos dados observados . . . 177
6.2 Resultados da aplicação do modelo de estimação por subsistema . . . 180
6.2.1 Premissas e organização dos resultados . . . 180
6.2.2 Resultados por subsistema . . . 184
6.3 Resultados da aplicação da metodologia de ajuste nas previsões de vento . . . 193
6.3.1 Análise da magnitude do vento previsto . . . 194
6.3.2 Panorama nacional do viés médio . . . 196
6.4 Resultados do modelo de previsão de geração por subsistema . . . 199
6.4.1 Previsões para o primeiro dia à frente . . . 199
6.4.2 Previsões horárias para o horizonte de 35 dias . . . 207
6.4.3 Previsões semanais para o horizonte até 5 semanas à frente . . . 215
6.4.4 Discussões complementares . . . 219
6.5 Resultados dos modelos de previsão em tempo real por subsistema . . . 221
6.5.1 Visão conjunta dos modelos . . . 222
6.5.2 Erro médio absoluto por horizonte . . . 223
6.5.3 Comportamento dos pesos da combinação . . . 225
6.5.4 Distribuição do erro por horizonte . . . 227
6.5.5 Síntese dos resultados por subsistema . . . 227
6.6 Considerações finais do capítulo . . . 229

7 – CONSIDERAÇÕES FINAIS **231**
7.1 Trabalhos futuros 237

REFERÊNCIAS BIBLIOGRÁFICAS **239**

APÊNDICE A – DETALHAMENTO DOS MODELOS DE TEMPO REAL **252**
A.1 Persistência pura 253
A.2 Persistência inteligente 253
A.2.1 Etapa 1, correção de viés de longo prazo 254
A.2.2 Etapa 2, ancoragem no valor atual 254
A.2.3 Etapa 3, componente autorregressiva 255
A.2.4 Etapa 4, correção fina ancorada no verificado atual 256
A.3 Persistência com derivada 256
A.4 Regressão linear 257
A.5 ARIMA sobre o resíduo do NWP 257
A.6 Suavização exponencial com reversão à média 258
A.7 *Random Forest* e *XGBoost* 259
A.8 Rede neural (MLP) 260
A.9 SARIMA 261
A.10 Combinação por inverso do erro (1/MAE) 261
A.11 Combinação por filtro de Kalman 262

APÊNDICE B – MÉTRICAS DE VALIDAÇÃO E DESEMPENHO **264**

APÊNDICE C – ANÁLISES DESAGREGADAS DOS RESULTADOS **269**
C.1 Comportamento da geração eólica no Brasil 269
C.2 Resultados da estimação por usinas e agrupamentos 270
C.2.1 Resultados por usinas individuais 270
C.2.2 Resultados por agrupamentos de usinas 273
C.3 Resultados do ajuste das previsões de vento por ponto e por estado 277
C.3.1 Análise por ponto de análise 277
C.3.2 Análise por estado 281
C.4 Resultados da previsão de geração por agrupamentos 287
C.5 Resultados dos modelos de tempo real por agrupamentos 290
C.5.1 A combinação das previsões de geração baseadas nos modelos NWP . . 290
C.5.2 Comportamento dos modelos ao longo do horizonte 291
C.5.3 O modelo de controle sem NWP (SARIMA) 292
C.5.4 Visão conjunta dos modelos 293
C.5.5 Erro médio absoluto por horizonte 294
C.5.6 Distribuição do erro por horizonte 296
C.5.7 Síntese dos resultados por agrupamentos 297

1 INTRODUÇÃO

A demanda mundial por eletricidade cresce de forma contínua e, aliada à busca por um desenvolvimento mais sustentável, consolidou as fontes renováveis variáveis (VREs, do inglês *Variable Renewable Energy*) como peça central da transição energética. Entre elas, a energia eólica destaca-se como uma das mais promissoras [1, 2]. Na última década, sua capacidade instalada global atingiu 1 TW, o que corresponde a aproximadamente 25% da capacidade total de geração renovável instalada mundialmente [3], e as projeções indicam que, até 2030, a geração eólica superará a das usinas hidrelétricas em escala global, ficando atrás apenas da solar fotovoltaica [4]. Esse avanço é impulsionado pela contínua redução dos custos de instalação, pelos ganhos de eficiência dos aerogeradores e pelas metas de descarbonização assumidas por diversos países. A expansão acelerada, contudo, não se traduz apenas em maior capacidade instalada: à medida que a participação eólica cresce, ampliam-se também os desafios técnicos associados à sua natureza variável e pouco despachável, o que exige dos operadores de sistema ferramentas cada vez mais robustas de previsão e planejamento.

No Brasil, a fonte eólica ocupa posição de destaque nessa transformação. A matriz elétrica nacional é composta por aproximadamente 90% de fontes renováveis, sendo cerca de 44% da capacidade total correspondente às grandes usinas e às pequenas centrais hidrelétricas [5]; o restante provém de outras fontes renováveis, como a solar e a eólica. A energia eólica, em particular, detém a segunda maior participação nessa matriz, respondendo por aproximadamente 14% da capacidade instalada total [5]. Com mais de mil parques eólicos de grande porte em operação, que totalizam cerca de 36 GW, o Brasil figura entre os cinco maiores países do mundo em capacidade instalada de geração eólica, atrás apenas de China, Estados Unidos, Alemanha e Índia [6], com expansão concentrada especialmente nas regiões Nordeste e Sul.

É nesse cenário de rápida expansão, e diante dos desafios que a alta penetração eólica impõe à operação e ao planejamento dos sistemas elétricos, que se insere a presente pesquisa. Para enfrentar esses desafios, de modo a permitir o aproveitamento seguro e eficiente da fonte, são necessárias metodologias capazes de tratar os dados de geração, estimar a produção efetiva das usinas e prever a geração futura. É em torno da superação desses desafios que este trabalho se desenvolve. Embora parte deles também se aplique à geração solar fotovoltaica, o escopo desta tese restringe-se ao tratamento de dados, à estimação e à previsão da geração eólica.

1.1 Contextualização

No contexto brasileiro, os desafios associados à integração de fontes renováveis variáveis apresentam características particulares em razão da estrutura institucional e operacional do setor elétrico, cuja operação é conduzida de forma centralizada pelo ONS (Operador Nacional do Sistema Elétrico). O mercado atacadista de eletricidade de curto prazo adota uma estrutura baseada em *pool*, mais especificamente do tipo *tight pool* [7]. Nesse contexto, o ONS é

responsável pelo despacho centralizado das unidades de geração de grande porte conectadas ao Sistema Interligado Nacional (SIN), rede nacional de transmissão de energia elétrica [7, 8].

Essas decisões de despacho são tomadas no âmbito do planejamento da operação, estruturado em horizontes semi-horários (um dia à frente), diários (até sete dias à frente), semanais (até um mês à frente) e mensais (até cinco anos à frente), com o objetivo de minimizar os custos operacionais presentes e futuros [7, 9]. Em última análise, essa tarefa envolve atender à demanda prevista de carga, otimizar os fluxos de transmissão no SIN e gerenciar estrategicamente os níveis dos reservatórios hidrelétricos [9]. Esses horizontes são operacionalizados por uma cadeia de modelos oficiais desenvolvidos pelo CEPEL (Centro de Pesquisas de Energia Elétrica): o Modelo de Planejamento de Curtíssimo Prazo (MPCTP), representado pelo DESSEM (Modelo de Despacho Hidrotérmico de Curto/Curtíssimo Prazo) [10]; o Modelo de Planejamento de Curto Prazo (MPCP), representado pelo DECOMP (Modelo de Determinação da Coordenação da Operação a Curto Prazo) [11]; e o Modelo de Planejamento de Médio Prazo (MPMP), representado pelo NEWAVE (Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo) [12].

Nesse ambiente de operação centralizada, a integração da fonte eólica impõe desafios de diferentes naturezas e magnitudes [13]. O principal decorre da elevada variabilidade da geração e da dificuldade de prever sua condição futura, dada sua dependência de variáveis climatológicas. Essas características exigem o acionamento de outras fontes para absorver essa intermitência, efeito que se intensifica com o aumento da capacidade eólica conectada ao sistema. Entre os impactos mais imediatos destacam-se a crescente necessidade de reservas de potência e energia [13–15] e as rampas de geração, que constituem um dos principais desafios para os operadores, especialmente na operação em tempo real [16]. Soma-se a isso o crescimento recente do corte de geração (*curtailment*) no Brasil, impulsionado pela expansão acelerada da geração renovável e por sua concentração em regiões com restrições de transmissão. A partir de 2023, esse fenômeno passou a ocorrer de forma expressiva, sobretudo no Nordeste [17]. Nessas situações, a usina gera menos do que poderia devido a restrições operativas, e não à condição de vento, fazendo com que a geração observada deixe de refletir o recurso eólico disponível.

1.2 Motivação e justificativa

Diante do cenário e dos desafios descritos, a previsão de energia eólica torna-se um insumo crítico para a operação e o planejamento do sistema [13–15, 18]. A previsão de energia eólica, com resolução temporal horária e alcance de até um mês à frente conforme considerado neste trabalho, é particularmente relevante para estudos de planejamento operacional e programação da operação conduzidos por Operadores Independentes do Sistema (ISOs, do inglês *Independent System Operators*) [15]. Essas previsões diferenciam-se conforme o horizonte temporal: o de tempo real, que na literatura abrange desde minutos até algumas horas [19], mas que, neste trabalho, se estende até 24 horas à frente e admite diferentes classes de modelos; o de

curtíssimo prazo apoia-se sobretudo em modelos meteorológicos; e o de curto prazo combina modelos meteorológicos e estatísticos. Essas previsões são fundamentais para estabelecer critérios que garantam a operação segura da rede de transmissão e para coordenar o despacho e o possível corte de diversas tecnologias de geração de maneira eficiente [15, 18, 20].

A qualidade das previsões desempenha um papel decisivo na eficácia de tais estudos e, consequentemente, no desempenho geral do sistema elétrico. A experiência internacional mostra que previsões subestimadas podem levar ao despacho de unidades de geração convencional mais caras, fora da ordem de mérito econômico, aumentando desnecessariamente os custos operacionais e afetando negativamente os consumidores [15]. Por outro lado, previsões superestimadas podem resultar em déficits de energia durante certos períodos, exigindo o desligamento forçado de dezenas de milhares de consumidores [18].

Diversos modelos voltados à previsão de energia eólica de curto prazo foram desenvolvidos e relatados na literatura especializada [14, 21–24]. Apesar dos sucessos relatados, algumas questões surgem da perspectiva dos ISOs. Muitos desses modelos foram desenvolvidos ou validados com base em um número limitado de parques eólicos [14, 21, 24], e frequentemente requerem dados de entrada que nem sempre estão disponíveis para os ISOs, como ângulos de passo das pás da turbina, orientação da nacele e extensas séries temporais históricas de dados operacionais das usinas [21, 23, 24]. Simultaneamente, as informações publicamente disponíveis sobre os modelos oficialmente adotados pelos ISOs são escassas e pouco claras na literatura, com detalhes e desempenho dispersos em documentos técnicos de difícil acesso ou desatualizados [15, 20, 25–28]. Essa opacidade pode decorrer de decisões estratégicas de segurança operacional ou de restrições contratuais, uma vez que muitos operadores dependem de previsões fornecidas por terceiros [14, 15, 26, 27], mas parece conflitar com a crescente demanda, sobretudo nos mercados atacadistas mais avançados, por maior transparência nas decisões operacionais do sistema [7, 20].

No contexto do planejamento eletroenergético realizado no Brasil pelo ONS, a previsão de geração eólica de curto prazo seguiu, até recentemente, metodologia estabelecida por Resoluções Normativas da Agência Nacional de Energia Elétrica (ANEEL) [29], baseada na média histórica da geração observada e, portanto, limitada à reprodução das variações sazonais das séries. Diante desse cenário, tornou-se fundamental para o ONS dispor de modelos capazes de produzir previsões eólicas mais acuradas. Esse contexto impulsionou a elaboração do presente trabalho, cujos resultados já contribuíram para o desenvolvimento de um novo modelo atualmente adotado pelo ONS, com aprovação da ANEEL e dos demais agentes do setor. As metodologias oriundas desta pesquisa já alimentam o MPCTP, no qual essa necessidade é mais crítica, além de subsidiarem o MPCP, o MPMP e a operação em tempo real. Adicionalmente, os modelos desenvolvidos permitem quantificar a geração não realizada por *curtailment*, informação necessária para a apuração de *constrained-off* e para a transparência regulatória entre os agentes e o operador.

1.3 Objetivos do trabalho

Com o objetivo de contribuir para uma representação mais precisa da geração eólica no planejamento e na operação de sistemas elétricos de grande porte com elevada participação dessa fonte, este trabalho propõe o desenvolvimento de novas metodologias e modelos matemáticos e computacionais para a estimação e a previsão da geração eólica. Busca-se, assim, mitigar os impactos decorrentes da variabilidade inerente à geração eólica, fornecendo subsídios tanto para operadores de sistemas quanto para a comunidade acadêmica dedicada ao tema.

Para alcançar esse objetivo, o trabalho contempla: (i) o desenvolvimento de metodologias para tratamento, recomposição e qualificação das séries históricas de vento e geração; (ii) a proposição de modelos para estimação da geração eólica e quantificação dos efeitos de eventos de *curtailment*; (iii) o desenvolvimento de técnicas de pós-processamento de previsões meteorológicas; (iv) a construção de modelos de previsão da geração eólica para diferentes horizontes temporais; (v) o desenvolvimento de métodos de combinação de previsões; (vi) a análise comparativa de abordagens clássicas e de aprendizado de máquina aplicadas à previsão; e (vii) a caracterização do comportamento da geração eólica no Brasil e de seus fatores de variabilidade.

1.4 Publicações decorrentes

Em decorrência das investigações realizadas ao longo deste trabalho, foram produzidos os seguintes artigos, publicados em periódicos e congressos nacionais e internacionais:

Ref. [30] NASCIMENTO, P. S. C.; MARCATO, A. L. M. A Statistical Approach for Missing Data Imputation and Outlier Detection in Wind Power Time Series for Large-Scale Power Systems. In: 61st International Universities Power Engineering Conference (UPEC), 2026.

Ref. [31] NASCIMENTO, P. S. C.; COSSICH, W.; ARAUJO, L.; SANTOS, I.; ALMEIDA, K.; MARCATO, A. L. M. An Integrated Methodology for Assessing Wind Power Curtailment Using Anemometric Measurements and Operational Data in the Brazilian Context. Atmosphere, v. 17, n. 4, p. 333, 2026. DOI: https://doi.org/10.3390/atmos17040333.

Ref. [32] NASCIMENTO, P. S. C.; DEOTTI, L. M. P.; MARCATO, A. L. M.; FARIAS, W. C. M. Wind Power Forecasting for Independent System Operators: Modeling Approaches, Practical Applications, and Operational Challenges in the Brazilian Context. IEEE Access, 2026. DOI: https://doi.org/10.1109/ACCESS.2026.3665895.

Ref. [33] NASCIMENTO, P. S. C.; ARAUJO, L. F.; ALVES, R. J. M.; FARIAS, W. C. M.; JUNIOR, C. A. A.; MARCATO, A. L. M. Avanços Metodológicos na Previsão Eólica para o Planejamento de Curto Prazo do ONS. In: Brazil Windpower, 2024, São Paulo. Brazil Windpower, 2024.

Ref. [34] NASCIMENTO, P. S. C.; ARAUJO, L. F.; MUNDSTOCK, T. F. G.; MARCATO, A. L. M.; NASCIMENTO, G. S. V. Modelo de Correção das Previsões de Geração Eólica em Tempo Real. In: Brazil Windpower, 2022, São Paulo. Brazil Windpower, 2022.

Ref. [35] NASCIMENTO, P. S. C.; ALVES, R. J. M.; MUNDSTOCK, T. F. G.; ALMEIDA JR, A. J. I.; MARCATO, A. L. M. Impacto da Disponibilidade de Potência na Previsão e Estimação da Geração Eólica. In: Brazil Windpower, 2022, São Paulo. Brazil Windpower, 2022.

Ref. [36] ARAUJO, L. F.; NASCIMENTO, P. S. C.; GONCALVES, G. A.; MARCATO, A. L. M.; MADEIRA, F.; COSSICH, W. Modelo de Estimação de Geração Eólica: Uma Proposta no Contexto de Recursos Energéticos Distribuídos. In: Brazil Windpower, 2022, São Paulo. Brazil Windpower, 2022.

Ref. [37] NASCIMENTO, P. S. C.; ARAUJO, L. F.; MARCATO, A. L. M.; ARAUJO JUNIOR, C. A.; BARROS, A. M. L. Modelo de Previsão Eólica Semanal para Estudos Prospectivos de Curto Prazo do ONS. In: LIV Simpósio Brasileiro de Pesquisa Operacional, 2022, Juiz de Fora. Anais do Simpósio Brasileiro de Pesquisa Operacional, 2022.

Ref. [38] NASCIMENTO, P. S. C.; KHENAYFIS, L. S.; ALVES, R. J. M.; MARCATO, A. L. M.; FONSECA, C. M. S.; SILVA, P. S. O. Desenvolvimento de um modelo de clusterização de usinas eólicas para atendimento dos modelos de previsão do ONS. In: IX Simpósio Brasileiro de Sistemas Elétricos, 2022, Santa Maria. Anais do Simpósio Brasileiro de Sistemas Elétricos, 2022.

Ref. [39] NASCIMENTO, P. S. C.; MENDES, E. L.; MARCATO, A. L. M.; ARAUJO, L. F.; KHENAYFIS, L. S. Statistical model for treatment of missing and outliers in time series for wind power forecast. In: XIV Latin-American Congress on Electricity Generation and Transmission, 2022, Rio de Janeiro. Anais do Congresso Latino-Americano de Geração e Transmissão de Energia Elétrica, 2022.

Ref. [40] NASCIMENTO, P. S. C.; ARAUJO, L. F.; MADEIRA, F.; KHENAYFIS, L. S.; AZAMBUJA, R. R. Desenvolvimento de um Modelo de Previsão Eólica Semanal para Planejamento de Curto Prazo. In: Brazil Windpower, 2021, São Paulo. Brazil Windpower, 2021.

Ref. [41] ARAUJO, L. F.; NASCIMENTO, P. S. C.; MADEIRA, F.; BARROS, A. M. L.; ALVES, R. J. M. Modelo de estimação de geração eólica a partir de dados de vento observado e curvas vento x potência. In: Brazil Windpower, 2021, São Paulo. Brazil Windpower, 2021.

Ref. [42] NASCIMENTO, P. S. C.; FRANCA, A. C. G.; KHENAYFIS, L. S.; DUARTE, V. S.; CAMARGO JUNIOR, H. Desenvolvimento e Análise de Modelos Clássicos e de Machine Learning na Previsão de Geração de Energia Eólica. In: Brazil Windpower, 2020, São Paulo. Brazil Windpower, 2020.

1.5 Principais contribuições do trabalho

Esta tese apresenta contribuições de natureza geral, assim como outras de natureza mais específica. As subseções a seguir descrevem algumas dessas contribuições.

1.5.1 Contribuições gerais

Dentre as contribuições de natureza mais gerais do trabalho, merece destaque:

I. A literatura raramente documenta de forma detalhada o comportamento operacional da geração eólica em nível de sistema apoiada em dados reais de um grande número de usinas, o que limita a compreensão das peculiaridades dessa fonte em larga escala, especialmente no sistema brasileiro. Para preencher essa lacuna, o trabalho caracteriza a operação eólica brasileira a partir de exemplos e situações reais de parques individuais e agrupados, abordando a variabilidade regional da geração, as diferenças entre os regimes de vento das regiões Sul e Nordeste e de suas sub-regiões, e os impactos das restrições operativas. O resultado é um panorama construído sobre um sistema com mais de mil usinas eólicas de grande porte, um dos maiores do mundo em geração centralizada, que confere abrangência incomum a essa caracterização. (Capítulo 6 e Apêndice C; Publicações [31, 32].)

II. As diferenças de desempenho e o viés dos modelos numéricos de previsão do tempo no contexto brasileiro são pouco documentados na literatura, sobretudo quanto à forma como esse viés se manifesta ao longo dos diferentes horizontes e passos de tempo. Para suprir essa necessidade, o trabalho analisa e compara os principais modelos meteorológicos internacionais aplicados às localidades dos parques nacionais. Como resultado, caracterizam-se as limitações e as diferenças de desempenho desses modelos, oferecendo subsídios objetivos para a seleção das fontes de previsão e para a correção de viés empregada nas etapas seguintes. (Capítulos 5 e 6; Publicação [32].)

1.5.2 Contribuições específicas

Os resultados obtidos com a pesquisa possibilitam a avaliação de algumas contribuições mais específicas, em comparação com as produções encontradas na literatura e/ou realizadas por outras instituições. É importante ressaltar que neste momento será oferecida apenas uma breve descrição das contribuições, sendo que os detalhes serão mais aprofundados no trabalho por meio dos textos, análises e resultados.

I. **Processamento e tratamento dos dados para séries temporais de velocidade do vento e geração de energia eólica**. Não foram encontrados na literatura trabalhos que reúnam, em uma única metodologia, todos os requisitos do tratamento de dados eólicos em um sistema de grande escala: detecção de dados anômalos, preenchimento de falhas e composição de múltiplas fontes de medição para mais de mil parques. Para contornar essa lacuna,

desenvolveu-se uma metodologia que integra e trata dados de vento e geração provenientes de diversas fontes (ONS, CCEE, EPE, ANEEL), com particularidades específicas de falhas, construindo séries temporais consistentes para cada parque com o apoio de informações de usinas vizinhas por correlação espacial. Como resultado, obtêm-se séries históricas contínuas e confiáveis em escala sistêmica, que servem de base para os modelos de estimação e de previsão. (Capítulo 3; Publicações [30, 39].)

II. **Estimação da geração potencial de usinas eólicas a partir de modelos de conversão de vento em potência**. As técnicas usuais de conversão de vento em potência são majoritariamente regressivas e sensíveis a dados contaminados por restrições operacionais, limites de transmissão e alocação energética, o que compromete sua aplicação direta em grande escala quando esses dados não são previamente tratados. Para resolver isso, desenvolveu-se um modelo que constrói de forma automática e dinâmica as curvas de conversão a partir de dados medidos, com filtros que detectam e removem automaticamente as situações inadequadas ao ajuste antes da regressão, e que estima a geração potencial, isto é, aquela que ocorreria na ausência de corte. Como resultado, é possível reconstruir séries históricas e quantificar as perdas por *curtailment* em escala sistêmica, o que viabiliza dois processos do ONS: a previsão eólica para a Programação Diária da Operação e o planejamento de curto prazo, e a apuração de eventos de *constrained-off*. (Capítulo 4; Publicações [31, 35, 36, 41].)

III. **Ajustes nas previsões de vento para diversos modelos meteorológicos e intervalos de tempo**. A forma como o viés das previsões de vento se manifesta de maneira distinta a cada passo de tempo do horizonte, especialmente no contexto brasileiro, é pouco explorada na literatura, embora seja determinante para a qualidade da conversão em potência. Para tratar esse ponto, empregaram-se técnicas de formulação simples, porém aplicadas de modo criterioso, que modelam e corrigem o viés de forma específica por modelo meteorológico e por horizonte de previsão. Como resultado, além de melhorar as previsões, a metodologia permitiu investigar e caracterizar as diferenças de desempenho entre os modelos meteorológicos ao longo dos horizontes. (Capítulo 5; Publicações [32, 42].)

IV. **Avanços na previsão e representação eólica nos horizontes do planejamento eletroenergético de curtíssimo e curto prazo**. Embora existam muitos modelos preditivos na literatura, poucos tratam da previsão eólica em nível sistêmico com o detalhamento metodológico e a dimensão exigidos por um dos maiores sistemas do mundo em capacidade instalada, e a maioria depende de estruturas temporais das séries pouco adequadas ao caráter estocástico da fonte nos horizontes de interesse. Para atender a essa necessidade, desenvolveu-se um modelo único de previsão, com alcance horário de até um mês à frente, baseado em informações meteorológicas previstas em vez da autocorrelação das séries históricas de geração, que representa conjuntamente múltiplos parques por agrupamento espacial para mitigar a alta variabilidade da fonte, e do qual derivam os

produtos de curtíssimo e de curto prazo com o detalhamento próprio de cada horizonte. Como resultado, aprimora-se a representação eólica nas ferramentas de planejamento em uso no setor elétrico brasileiro, contribuindo para uma evolução em relação à metodologia anteriormente adotada pelo setor. (Capítulo 5; Publicações [32, 33, 36–38, 40, 42].)

V. **Avanços metodológicos para realização de previsão para operação em tempo real**. Existe uma variedade de modelos de previsão em tempo real na literatura, mas a maioria adota uma abordagem isolada e nem sempre aproveita a informação mais recente observada, ao passo que os modelos meteorológicos apresentam viés nos primeiros passos de tempo. Este trabalho realiza uma aplicação de diversas técnicas de previsão de tempo real, com adaptações e formulações inéditas, que vão desde variações da persistência (pura, inteligente e com derivada) até modelos estatísticos e de aprendizado de máquina. Além disso, outra contribuição reside em uma combinação dinâmica que integra tanto as previsões dos modelos meteorológicos quanto as dos modelos de tempo real, ajustando-se continuamente ao desempenho recente de cada um. Como resultado, a previsão corrige o viés meteorológico nos primeiros horizontes e converge de forma suave às previsões meteorológicas ao final do horizonte de tempo real, mostrando-se mais adaptativa do que as abordagens usuais da literatura. (Capítulo 5; Publicação [34].)

Além das contribuições discutidas anteriormente nesta seção, algumas outras contribuições vinculadas aos avanços e às publicações resultantes deste trabalho serão detalhadas mais adiante neste documento.

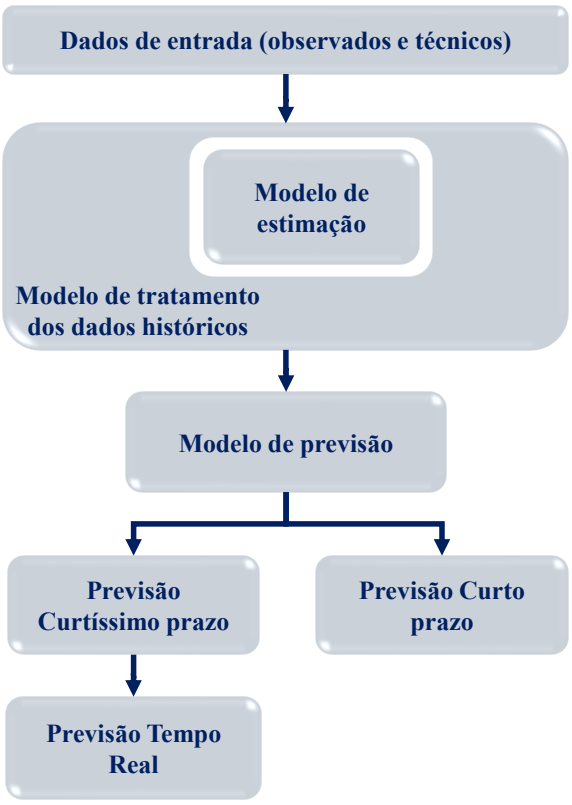
1.6 Estrutura do trabalho

O modelo de previsão de geração eólica proposto é estruturado em quatro grandes blocos funcionais encadeados: os **dados de entrada**, o **modelo de tratamento dos dados históricos**, o **modelo de estimação** e o **modelo de previsão**. A Figura 1 apresenta uma visão simplificada dessa estrutura, cuja versão detalhada é retomada ao final do Capítulo 5.

Com o objetivo de descrever de forma abrangente a metodologia proposta, seus fundamentos teóricos e as etapas de desenvolvimento do modelo, esta tese está organizada em sete capítulos. O presente capítulo apresenta o contexto da pesquisa, sua motivação, os objetivos e as principais contribuições do trabalho, além de introduzir os conceitos essenciais para a compreensão dos capítulos subsequentes.

O Capítulo 2 apresenta um levantamento bibliográfico sobre o tema da pesquisa. Inicialmente, são apresentadas algumas referências de métodos de tratamento de dados de fontes renováveis. Em seguida, é feito um detalhamento das informações meteorológicas e das possíveis técnicas de correção dessas previsões. Por fim, apresenta-se uma revisão crítica acerca dos modelos de previsão de geração, especialmente de geração eólica.

Figura 1 – Fluxograma simplificado do modelo proposto, com os quatro blocos encadeados.



Fonte: Elaboração própria.

O Capítulo 3, associado ao bloco de **tratamento dos dados históricos**, destaca alguns conceitos fundamentais que estão por trás dos principais métodos de tratamento de dados. Além disso, explana toda a metodologia desenvolvida para a seleção de dados e a composição das séries temporais históricas necessárias aos modelos de estimação e de previsão.

O Capítulo 4 detalha o **modelo de estimação** da geração das usinas eólicas, construído a partir de dados medidos, como registros anemométricos, de geração e de potência disponível. São detalhadas as técnicas de ajuste da geração e de construção automática e dinâmica dos modelos de conversão de vento em potência, aplicáveis à reconstrução de séries históricas e à estimação da geração potencial.

O Capítulo 5 apresenta o **modelo de previsão** de geração em seus três horizontes. São descritos o modelo de previsão em tempo real, baseado em um *ensemble* de modelos combinados dinamicamente, e o modelo que abrange desde o curtíssimo prazo até o curto prazo, com o detalhamento específico exigido por cada horizonte. Ao final deste capítulo, é retomada a versão completa do fluxograma do modelo proposto.

O Capítulo 6, por fim, descreve os principais resultados obtidos, na mesma ótica de separação entre a previsão em tempo real e a previsão de curtíssimo e curto prazo.

E ao final, o Capítulo 7 sintetiza as principais constatações e contribuições do trabalho.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta um levantamento bibliográfico acerca do tema de pesquisa do presente trabalho. Inicialmente, são apresentados conceitos introdutórios sobre a fonte eólica no Brasil e no mundo, seguidos de uma discussão sobre os aspectos regulatórios da previsão de energia eólica e os modelos de previsão centralizados e descentralizados adotados pelos operadores. Em seguida, abordam-se as principais características e técnicas dos modelos para tratamento de dados e para o ajuste das previsões dos modelos meteorológicos. Na sequência, são detalhadas as principais abordagens metodológicas para a previsão de geração eólica, bem como uma discussão sobre o fenômeno do corte de geração (*curtailment*). Por fim, discute-se a representação da geração eólica na cadeia de planejamento da operação do SIN e os desafios relacionados à comparabilidade entre modelos de previsão.

2.1 A geração eólica no Brasil e no mundo

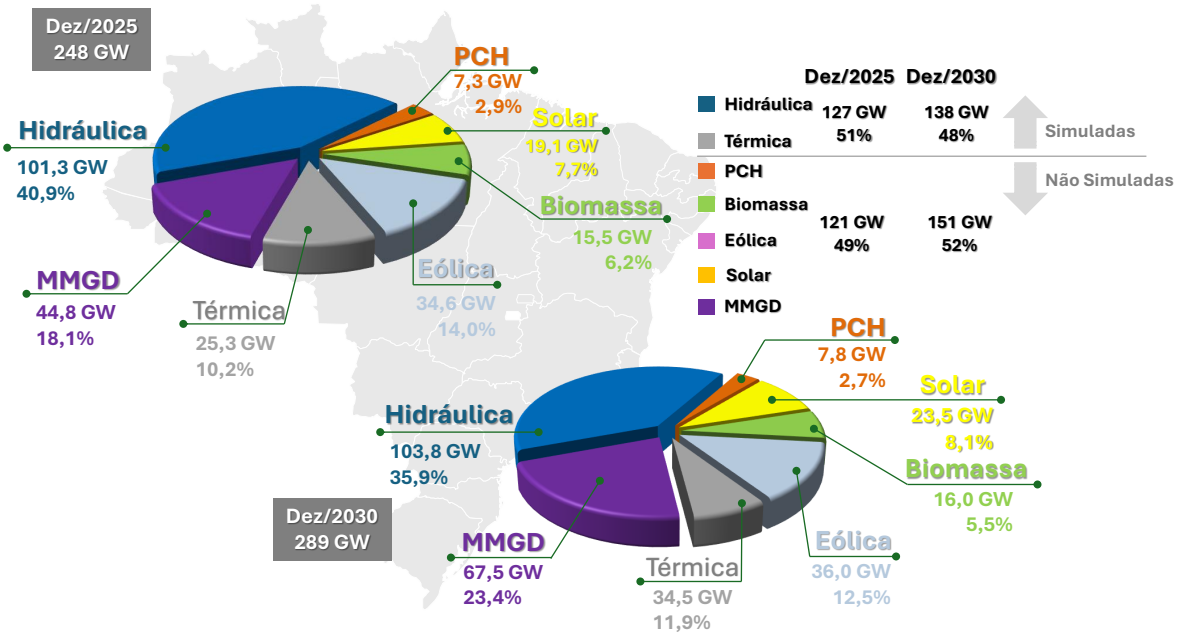
O Brasil é um país privilegiado em se tratando dos recursos energéticos. A matriz eletroenergética do país é majoritariamente renovável, dado que é capaz de gerar a energia necessária para atender a demanda do país a partir de fontes renováveis como a solar, a eólica e a hidráulica, além de fontes menores como a biomassa, ver Figura 2. Poucos países do mundo possuem uma matriz eletroenergética com tamanha proporção de fontes renováveis; países como Noruega, Dinamarca e Islândia também apresentam números expressivos de geração através de energia limpa.

Em 2024, o Brasil já ocupava o 5º lugar no ranking mundial de capacidade instalada de energia eólica *onshore*, com perspectiva de crescimento [6]. A capacidade instalada passará de 34,6 GW em 2025 para 36,0 GW em 2030, ver Figura 2.

Anteriormente, a matriz eletroenergética do país era majoritariamente hidrotérmica, com os modelos de planejamento bem adaptados e consolidados. Essas duas fontes (térmica e hidráulica) são as únicas que possuem seus despachos simulados para efeito da otimização energética, as demais fontes (eólica, solar etc.) são representadas por despachos fixos nos modelos de planejamento, sendo nomeadas como usinas não simuladas. Vale esclarecer que o termo “não simuladas” se refere apenas à consideração das fontes nos modelos do planejamento energético; no entanto, seus despachos são estimados por outros modelos, alguns deles alvos deste trabalho.

O grau de penetração e crescimento das fontes não simuladas, especialmente nos últimos anos, tem demandado ao ONS investimento nos seus modelos e processos, de forma a abarcar o novo paradigma de operação e planejamento do SIN. Aliado ao crescimento, há inúmeros desafios para a operação, sendo a alta variabilidade e a dificuldade de previsão os maiores deles. Impactos na estabilidade do sistema são frequentes, podendo até levar o sistema elétrico ao colapso. O caso mais recente foi o ocorrido em agosto de 2023, em que vários estados do país

Figura 2 – Evolução da capacidade instalada no SIN.



Fonte: Obtido de [43].

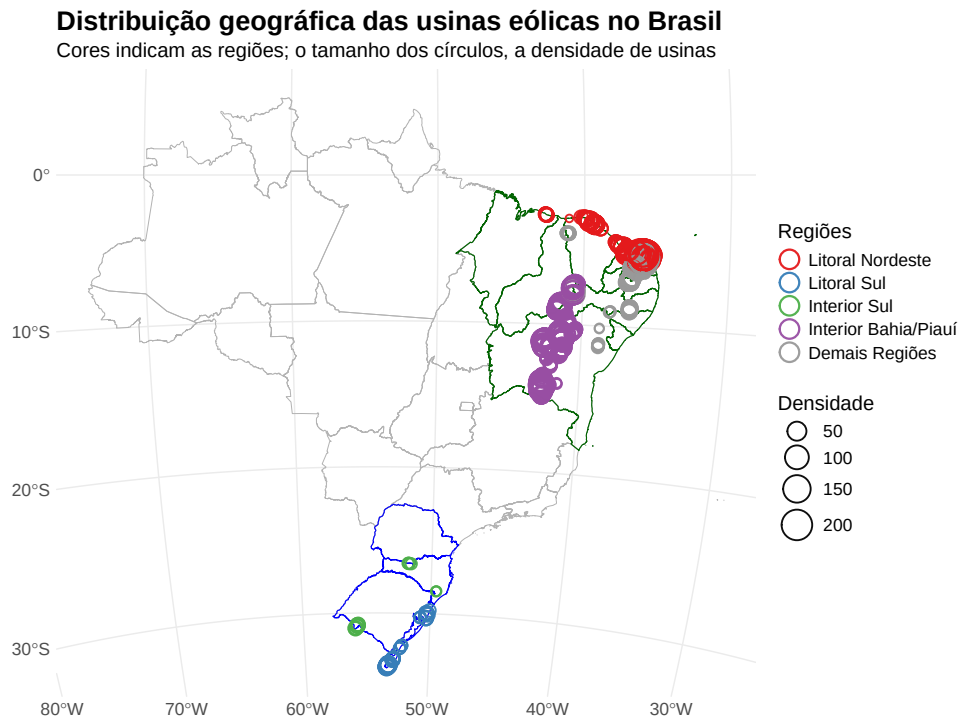
foram prejudicados devido a uma perturbação no subsistema da região Nordeste. Nessa situação, ocorreu uma falha de desempenho não esperada de parques eólicos e fotovoltaicos, fazendo com que o sistema de proteção atuasse, realizando a abertura do terminal de Quixadá da LT 500 kV Quixadá–Fortaleza II [44].

Tal ocorrência reforça que ainda são necessários ajustes nos processos e modelos da operação, a fim de garantir uma operação ainda mais segura e confiável para o fornecimento de energia a partir de fontes renováveis.

Além disso, a forma como a geração eólica se distribui no território brasileiro traz consigo desafios adicionais à operação e ao planejamento, bem como aos modelos e processos do ONS. A Figura 3 ilustra a distribuição geográfica dessas usinas em todo o território nacional. Os círculos representam parques eólicos, com seus tamanhos proporcionais à densidade de pontos adjacentes, permitindo a identificação de áreas com maior concentração. A densidade foi calculada com base nas distâncias geográficas entre parques eólicos, considerando usinas separadas por menos de 0,5 grau como pontos vizinhos.

A frota eólica brasileira é composta exclusivamente por parques eólicos *onshore*. Sua distribuição espacial é fortemente concentrada nas regiões Nordeste e Sul, conforme ilustrado na Figura 3, destacadas pelas áreas delineadas em verde e azul, respectivamente. Aproximadamente 94% dos parques eólicos estão localizados no Nordeste, concentração que aumenta a complexidade da operação do sistema elétrico nessa região, particularmente durante períodos de alta disponibilidade de recurso eólico. Em contraste, cerca de 6% dos parques eólicos estão localizados no Sul.

Figura 3 – Distribuição geográfica das usinas eólicas no território brasileiro.



Fonte: Elaboração própria.

Além das diferenças na concentração espacial, existem também diferenças estruturais relevantes entre os regimes de vento das regiões Nordeste e Sul, bem como de suas respectivas sub-regiões. Essas particularidades resultam em padrões distintos de geração eólica observada. Conforme mostrado na Figura 3, é possível identificar comportamentos mais característicos em áreas costeiras, representadas como Litoral Nordeste e Litoral Sul, que incluem parques eólicos localizados a até 40 km da linha da costa, e em áreas interioranas, classificadas como Interior Sul, Interior Bahia/Piauí e Outras Regiões.

Essas diferenças regionais e sub-regionais influenciam diretamente os modelos de previsão e de estimação da geração potencial. Nesses casos, uma caracterização precisa do comportamento típico e da variabilidade intrínseca do recurso eólico é essencial para evitar vieses na quantificação das perdas de energia.

Embora o Brasil apresente um alto potencial técnico para geração eólica *offshore* [45], a participação dessa fonte na matriz elétrica brasileira permanece incipiente, e atualmente não existem previsões consolidadas para uma penetração significativa no curto prazo. Ainda assim, estudos indicam que a energia eólica *offshore* deverá ser progressivamente incorporada aos cenários energéticos de longo prazo.

2.2 Aspectos regulatórios da previsão de energia eólica

A previsão de energia eólica desempenha um papel crítico na eficiência operacional do sistema elétrico e é, portanto, um elemento-chave na governança setorial [13, 15]. Nesse contexto, uma questão fundamental diz respeito à alocação de papéis e responsabilidades entre os agentes envolvidos. A literatura identifica dois paradigmas principais para abordar isso: sistemas de previsão centralizados e descentralizados [15, 25, 46, 47].

Os aspectos regulatórios adotados por cada país influenciam diretamente a forma como os operadores de sistema e de rede (ISOs) gerenciam as atividades de previsão da geração proveniente de fontes renováveis variáveis. As diretrizes regulatórias estabelecem, por exemplo, a distribuição de responsabilidades entre operadores, agentes geradores e provedores de serviços de previsão, além de definirem requisitos relacionados à qualidade, à periodicidade e ao compartilhamento das informações previstas.

Os modelos desenvolvidos neste trabalho possuem como público-alvo principal os operadores de sistema, com foco específico no ONS. Dessa forma, a compreensão do contexto regulatório associado à previsão de geração renovável torna-se relevante, uma vez que diversos aspectos da estruturação metodológica e do desenvolvimento dos modelos foram orientados por requisitos operacionais e regulatórios observados nesse ambiente institucional.

2.2.1 Sistemas de previsão centralizados vs. descentralizados

Em um *sistema centralizado*, uma entidade autônoma e imparcial dentro do setor elétrico, tipicamente o ISO ou a autoridade reguladora, assume a responsabilidade pela previsão de geração de todos os parques eólicos elegíveis [15, 18, 46, 47]. Essa entidade define os dados operacionais e informações que devem ser obrigatoriamente fornecidos pelos agentes correspondentes [15, 46, 47].

Em contraste, em um *sistema descentralizado*, a responsabilidade pela previsão recai sobre os agentes que operam os respectivos parques eólicos elegíveis [15, 18, 46, 47]. Uma vez que as previsões devem, em última instância, ser entregues ao ISO, cabe ao operador definir os requisitos técnicos e procedimentais a serem atendidos, como níveis mínimos de precisão, resolução temporal, horizonte de previsão, prazos de submissão e frequência de atualização [15, 46, 47].

Cada sistema apresenta seu próprio conjunto de vantagens e desafios. O sistema descentralizado, por exemplo, tem o potencial de produzir previsões mais precisas devido ao conhecimento detalhado dos agentes sobre as condições meteorológicas locais, ao acesso direto às especificações técnicas dos componentes dos parques eólicos e ao conhecimento imediato do estado operacional [15, 46–48]. No entanto, a ausência ou fragilidade de um arcabouço legal e administrativo que obrigue a entrega de previsões de alta qualidade por tais agentes permanece uma questão crítica para a eficácia desse sistema [15, 18, 46].

Por outro lado, os sistemas de previsão centralizados tendem a produzir resultados mais consistentes e alinhados com os interesses do sistema como um todo [15, 25, 46, 47]. O uso de metodologias padronizadas, a agregação de dados operacionais de usinas geograficamente próximas e a mitigação da incerteza por meio da diversidade espacial são fatores que contribuem para essa consistência [15,25,46,47]. No entanto, a falta de avanços metodológicos contínuos e o monitoramento insuficiente da qualidade dos dados fornecidos pelos agentes podem comprometer a eficácia geral do sistema [25, 46].

2.2.2 Experiência internacional

A decisão entre um arcabouço centralizado ou descentralizado de previsão de energia eólica é tipicamente definida pelas características específicas do setor elétrico de cada país ou região. Aspectos estruturais, como a organização do mercado atacadista de eletricidade de curto prazo e o esquema de despacho econômico correspondente, podem influenciar essa decisão [7]. No entanto, tais fatores estruturais são condicionados por considerações técnicas, como a segurança da rede e a confiabilidade operacional [15, 18].

Em mercados atacadistas de eletricidade baseados em *pool*, as decisões de despacho são tipicamente conduzidas por uma entidade independente, frequentemente o próprio ISO [7, 8, 49]. Isso naturalmente favorece a implementação de sistemas centralizados de previsão de energia eólica. Exemplos incluem o Alberta Electric System Operator (AESO) no Canadá e o Electric Reliability Council of Texas (ERCOT) nos Estados Unidos [46]. Notavelmente, alguns ISOs adotaram um esquema híbrido de previsão [15], no qual o operador permanece responsável pela previsão das usinas elegíveis, mas permite que os participantes do mercado submetam suas próprias previsões [15, 25, 46, 47]. No caso do Australian Energy Market Operator (AEMO), a participação é opcional [25,46], enquanto no Chile, a participação é obrigatória sob o Coordinador Eléctrico Nacional (CEN) [15]. Independentemente do modo de participação, os sistemas híbridos permitem o acesso às previsões mais precisas disponíveis.

Por outro lado, em mercados de autodespacho, as empresas de geração são responsáveis pelas decisões de despacho com base em contratos bilaterais registrados em ambientes designados [7, 49]. Embora esse modelo sugira compatibilidade com a previsão descentralizada, isso não é comumente observado na prática [15, 46, 50]. É essencial reconhecer que as previsões de energia eólica atendem a diversos propósitos operacionais, desde o planejamento da operação semanal até avaliações de confiabilidade e análises de redespacho [20, 51]. Portanto, mesmo em mercados de autodespacho, os ISOs frequentemente mantêm suas próprias previsões de energia eólica para apoiar avaliações técnicas internas e validar as previsões submetidas pelos participantes do mercado. Este é o caso da Red Eléctrica de España (REE) na Espanha e do National Energy System Operator (NESO) no Reino Unido [46, 50].

Dado esse contexto, estima-se que aproximadamente 80% dos ISOs em todo o mundo mantêm suas próprias previsões de energia eólica de curto prazo (assim como previsões de

energia solar), seja por meio de modelos centralizados ou híbridos [15]. Essas previsões podem ser terceirizadas para provedores especializados ou desenvolvidas internamente usando ferramentas e expertise próprias [15, 26, 28, 48, 50]. A decisão depende de fatores como a disponibilidade no mercado de provedores de serviços de previsão, as capacidades técnicas internas, a relevância operacional da geração renovável variável e os recursos financeiros e de infraestrutura disponíveis [47, 48]. Por exemplo, o ERCOT atualmente terceiriza a previsão para dois provedores externos, enquanto o AEMO e a REE desenvolvem e mantêm seus próprios modelos internos de previsão [26, 28, 50].

2.3 Modelos para tratamento de dados

Para atender os objetivos ou propósitos deste trabalho, será necessário o uso de um grande volume de dados. Sendo assim, foram reunidas informações de diversos centros e instituições do setor elétrico e de meteorologia, em nível nacional e internacional.

O conjunto de informações cadastrais das usinas são oriundas da ANEEL (Agência Nacional de Energia Elétrica) e do ONS. Os dados históricos observados de geração, velocidade do vento e outros, são obtidos do Sistema de Supervisão e Controle (SSC) do ONS. Em complemento, será utilizado dados históricos de geração observada do SMF (Sistema de Medição para Faturamento) da CCEE (Câmara de Comercialização de Energia Elétrica). Completando, dados anemométricos observados para as usinas eólicas também são provenientes do sistema AMA (Sistema de Acompanhamento de Medições Anemométricas), gerenciado pela EPE (Empresa de Pesquisa Energética). Vale salientar que os dados da EPE se restringem às usinas eólicas do Ambiente de Contratação Regulada (ACR), porém o volume de usinas programadas pelo ONS se estende um pouco mais.

Atualmente, o SIN é composto por mais de mil usinas eólicas conectadas na rede básica, com projeção de crescimento. Dessa forma, é imprescindível desenvolver técnicas/algoritmos inteligentes de análise e processamento de grandes massas de dados, técnicas de mineração de dados (*data mining*), tratamento de dados faltantes (*missing data*) e dados espúrios (*outliers*) [52]. É importante destacar que, na fase de tratamento de dados, as usinas são tratadas individualmente. No entanto, na etapa de previsão, essas usinas podem ser agrupadas, como será explicado em detalhes mais adiante neste documento.

As séries de geração e de vento adquiridas em tempo real são utilizadas no planejamento de curtíssimo, curto e médio prazo da operação do Sistema Interligado Nacional (SIN). No planejamento de curtíssimo prazo, os montantes diários de geração das usinas, em patamares semi-horários, são estimados com base nas previsões de geração eólica para o dia seguinte da operação, sendo os desvios de previsão um dos indicadores da qualidade do planejamento. Nesse processo, são utilizados modelos de previsão de séries temporais, e um de seus principais requisitos é que as séries estejam livres de dados faltantes, de dados espúrios e de efeitos causados por eventos especiais.

A presença de dados faltantes e dados espúrios pode comprometer a análise de séries temporais, assim como mudanças estruturais que ocorrem ao longo do tempo. Apesar de sua relevância, os impactos desses eventos frequentemente são negligenciados, em parte pela dificuldade de representar adequadamente a dinâmica temporal dos dados e de aplicar métodos robustos para sua identificação e correção. No modelo de tratamento de dados proposto nesta tese, descrito no Capítulo 3, são empregadas técnicas estatísticas e computacionais para identificação e correção de inconsistências em séries temporais, incluindo uma abordagem baseada no filtro de Kalman para sua recomposição. A seguir, são apresentados trabalhos da literatura que fundamentam essa escolha metodológica.

Diversas abordagens têm sido propostas na literatura para lidar com esse problema. Métodos estatísticos são amplamente empregados na detecção e correção de dados espúrios, assim como técnicas de inteligência artificial, incluindo redes neurais artificiais e outros métodos de *machine learning*, que vêm sendo aplicados com sucesso em séries do setor elétrico, como em [53].

Os modelos baseados em filtro de Kalman constituem uma das principais bases teóricas para o tratamento, a estimação e a previsão de séries temporais dinâmicas [54]. Nesse contexto, [55] propõe um modelo para monitoração da aquisição de dados baseado em modelos estruturais em espaço de estados e algoritmos iterativos, capaz de corrigir grandes lacunas em séries históricas armazenadas, bem como identificar e corrigir, em tempo real, falhas em medições de carga. De forma semelhante, [56] desenvolve uma metodologia para tratamento de dados faltantes e dados espúrios em séries temporais de carga minuto a minuto, combinando modelos lineares dinâmicos, redes neurais e filtro de Kalman.

A utilização do filtro de Kalman também é recorrente em aplicações relacionadas à energia eólica. Em [57], essa técnica é utilizada para previsão uma hora à frente da velocidade do vento em um parque eólico. Adicionalmente, filtros de Kalman são amplamente empregados em aplicações envolvendo amostragem e processamento de dados em intervalos curtos de tempo. Em [58], é apresentada uma abordagem baseada no filtro de Kalman para detecção *online* de falhas em turbinas eólicas. Esses trabalhos evidenciam a versatilidade dessa técnica em problemas de tratamento, estimação e acompanhamento dinâmico de séries temporais, motivando sua adoção na metodologia desenvolvida nesta tese.

2.4 Modelos para ajuste de previsões numéricas de tempo

Em geral, os modelos de previsão de geração eólica de curtíssimo prazo, comumente, utilizam informações previstas por modelos de previsão numérica de tempo NWP (*Numerical Weather Prediction*) [59]. Os modelos numéricos de tempo e clima, também chamados modelos meteorológicos, descrevem a física da atmosfera através de um conjunto de equações, como a equação de Navier–Stokes [60], que ao serem solucionadas permitem descrever o comportamento passado, presente e/ou futuro das grandezas meteorológicas.

Os resultados derivados desta pesquisa, apresentados tanto nos artigos publicados quanto nesta própria tese, são baseados em análises, testes e experimentos realizados com previsões provenientes de diferentes NWPs, destacando-se: (i) o modelo atmosférico de mesoescala Eta, (ii) o WRF (*Weather Research and Forecasting*) e (iii) o BAM (*Brazilian Atmospheric Model*), todos do CPTEC (Centro de Previsão de Tempo e Estudos Climáticos); (iv) o GFS (*Global Forecast System*) e o GEFS (*Global Ensemble Forecast System*), da NOAA (*National Oceanic and Atmospheric Administration*); e (v) o IFS (*Integrated Forecasting System*) do ECMWF (*European Centre for Medium-Range Weather Forecasts*).

Embora as informações provenientes dos NWPs apresentem os melhores resultados de previsão para as grandezas meteorológicas, comumente, é necessária a aplicação de modelos estatísticos de pós-processamento dessas informações. Nesse processo, a saída dos modelos NWP é tratada com o uso de observações, com intuito de encontrar alguma relação estatística e, dessa forma, corrigir os desvios dessas informações, [61].

Existem muitas outras soluções para pós-processamento das previsões numéricas, de forma a trazer uma maior assertividade aos modelos. Tais soluções envolvem métodos estocásticos, técnicas baseadas em análises de séries temporais e reanálises de dados meteorológicos. Cada uma dessas técnicas possui algumas vantagens e desvantagens e não são adequadas para todos os casos. Mas com o avanço dos estudos nessa área, novas técnicas mais avançadas foram desenvolvidas como os modelos numéricos de alta resolução, técnicas de aprendizado de máquina e a integração de múltiplas fontes (satélites, anemômetros etc.).

Uma necessidade é o ajuste do viés de vento, que consiste em desvios sistemáticos entre os valores de vento previsto e vento verificado; esses desvios podem ser causados por diversos fatores. Esse comportamento pode estar relacionado a erros nos parâmetros de entrada dos modelos, erros nos dados atmosféricos coletados, limites da resolução espacial, dificuldade em representar os fenômenos locais, condições iniciais da previsão, variações na atmosfera e um viés do próprio modelo meteorológico etc. [62].

Entre as técnicas de pós-processamento, especialmente para ajustar os vieses dos modelos, destaca-se o método conhecido como *Delta Change*. Este é um dos métodos mais simples e menos robustos para ajustar vieses em previsões meteorológicas. Consiste em calcular as diferenças entre os dados observados e previstos [63]. Esse método funciona assumindo que a variabilidade dos dados ao longo da série temporal possui a mesma magnitude.

Outra técnica bastante difundida na literatura é o mapeamento quantil-quantil, é utilizada por agências que tratam dados climáticos em seus modelos oficiais, por exemplo, a ANA (Agência Nacional de Águas) e a EPE (Empresa de Pesquisa Energética). Os quantis são valores que dividem um conjunto de dados em intervalos regulares, ou a probabilidade de uma variável aleatória assumir um valor superior ou inferior ao quantil desejado. A função quantil também pode ser descrita como a função inversa de uma função de distribuição acumulada.

O mapeamento quantil e suas variações são técnicas estatísticas que podem ser utilizadas na remoção de viés de variáveis climáticas, assumindo que a probabilidade da distribuição dos dados presentes é a mesma dos dados futuros. Esse mapeamento busca avaliar as distribuições da variável prevista e observada, comparando o grau de ajuste entre essas variáveis [64]. A técnica de mapeamento da distribuição normal é umas das variações, tal técnica assume que a variável aleatória possui uma distribuição normal, também conhecida como distribuição Gaussiana. Utilizando a função inversa da distribuição normal, é possível calcular os quantis das variáveis aleatórias (vento observado e vento previsto). O que permite verificar a diferença entre as distribuições das variáveis utilizadas. Esse método busca igualar as distribuições das variáveis observadas e previstas a partir de um valor ΔQ, dado pela diferença entre os quantis calculados [65]. Ou seja, os quantis das variáveis de vento observado e previsto são comparados e um ΔQ é estimado, com esse valor é possível atribuir os ajustes necessários aos valores previstos para que eles se alinhem e o viés seja removido.

Outra variação, o mapeamento quantil robusto empírico tem o mesmo princípio do modelo apresentando anteriormente, no entanto, utiliza algumas técnicas mais robustas para fazer o mapeamento quantil. Essa técnica consegue tratar os dados espúrios por utilizar técnicas e estimadores mais robustos, assim, consegue evitar distorções que podem aparecer nas distribuições [66]. Já o mapeamento quantil delta é uma abordagem mais refinada e consegue ajustar as distribuições de probabilidade, além de conseguir corrigir o viés sistemático, permite uma melhor adaptação dos dados previstos aos dados observados [67].

O MOS (*Model Output Statistics*) é uma técnica de pós-processamento estatística que pode ser utilizada para correção de dados meteorológicos de modelos numéricos. O objetivo desse modelo é reduzir a variação/viés entre os dados do previstos pelo modelo numérico e os dados observados [68, 69]. O modelo utiliza dados históricos de previsão e de observações para identificar conexões estatísticas através de análises de correlações, regressões simples, regressões multivariadas e outras técnicas que explicam a relação entre as variáveis.

Uma outra técnica é o escalonamento linear, que consiste em ajustar os dados de previsão de vento aos dados observados com o objetivo de diminuir ao máximo a variação entre esses dados. A técnica é baseada em identificar um fator de escala que seja capaz de minimizar as diferenças entre os dados previstos e observados [70]. O ajuste dos dados previstos é feito identificando um fator de escala que será aplicado às previsões originais. O fator de escala pode ser identificando utilizando técnicas estatísticas, como a regressão linear, regressão polinomial, regressão multivariada, entre outras.

Algumas técnicas mais avançadas, como as redes neurais de regressão generalizada GRNN (*General Regression Neural Network*) [71], também podem ser utilizadas. Essa rede pode ser projetada para realizar regressões para modelar relações não lineares e efetuar previsões. Essa rede possui quatro camadas em sua estrutura: camada de entrada (camada base radial), camada padrão, camada somatória e camada de saída (camada linear especial). A GRNN é

treinada a partir dos ajustes dos pesos na camada de saída da rede, a rede busca identificar os padrões dos dados de entradas, para cada padrão identificado um peso é associado e a saída da rede é a média ponderada dos resultados. A GRNN é conhecida por sua capacidade de adaptação a diferentes tipos de dados e sua habilidade em modelar relações complexas de maneira eficiente. Além disso, ela tem a vantagem de não exigir um longo período de treinamento, já que a fase de treinamento envolve apenas o armazenamento dos padrões.

Dado o exposto anteriormente, foi necessário o desenvolvimento de algoritmos dinâmicos e inteligentes para tratamentos das previsões dos modelos NWPs. Tais desenvolvimentos serão tratados mais adiante neste documento.

2.5 Modelos para previsão de geração eólica

Na literatura científica, diversos modelos de previsão de energia eólica de curto prazo estão disponíveis. Os modelos de previsão que deverão ser desenvolvidos podem ser classificados como métodos físicos, estatísticos, de aprendizado de máquina e/ou híbridos [72]. Uma breve visão geral desses modelos, bem como sua classificação sistemática, é apresentada a seguir.

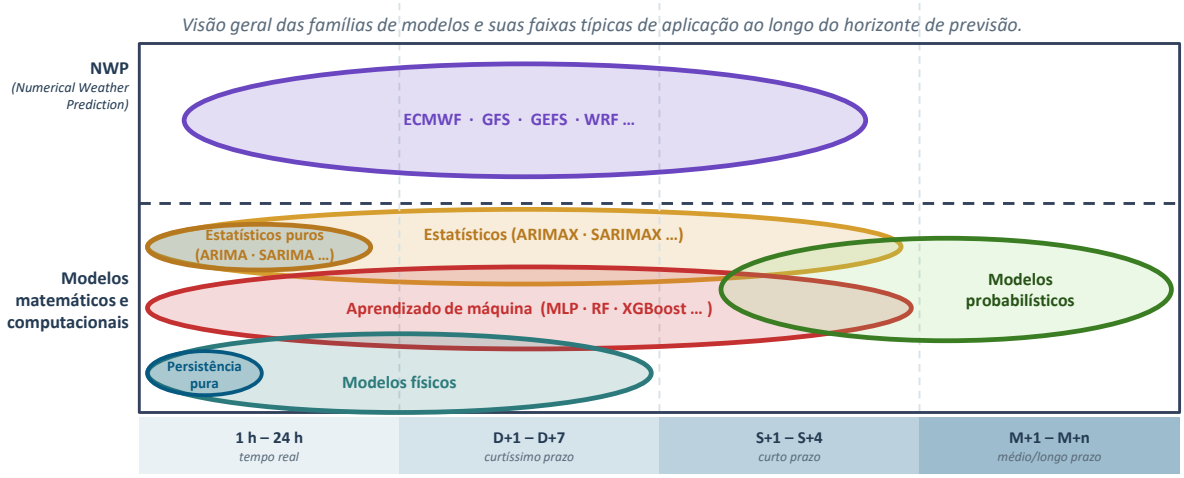
De forma geral, as famílias de modelos diferem tanto quanto à abordagem metodológica adotada (física, estatística, de aprendizado de máquina ou híbrida) quanto à faixa de horizonte em que são tipicamente aplicadas. A Figura 4 apresenta um panorama dessa relação, organizando as principais famílias de modelos ao longo do horizonte de previsão, desde o tempo real (instante atual a vinte e quatro horas à frente) até os horizontes de médio e longo prazo. Observa-se que os modelos baseados na própria série histórica, como a persistência e os modelos estatísticos puros, predominam nos horizontes mais curtos, enquanto as previsões dos modelos de NWP e as abordagens de aprendizado de máquina ganham relevância à medida que o horizonte se estende. Para os horizontes mais distantes, as abordagens probabilísticas tornam-se predominantes, em razão do aumento da incerteza. Os modelos híbridos, por combinarem componentes de mais de uma família, não constituem um bloco próprio na figura.

2.5.1 Classificação dos modelos

Os modelos de previsão de geração eólica podem ser sistematicamente classificados ao longo de três eixos independentes e complementares: os *dados de entrada* utilizados, a *abordagem metodológica* empregada e as *características da saída* produzida [73, 74]. Esses eixos são ortogonais entre si, de modo que um mesmo modelo é caracterizado por uma posição em cada um deles. Assim, por exemplo, um modelo de uma dada abordagem metodológica pode utilizar diferentes conjuntos de dados de entrada e produzir tanto saídas determinísticas quanto probabilísticas. A Figura 5 sintetiza essa classificação, resumindo as categorias de cada eixo que serão detalhadas a seguir.

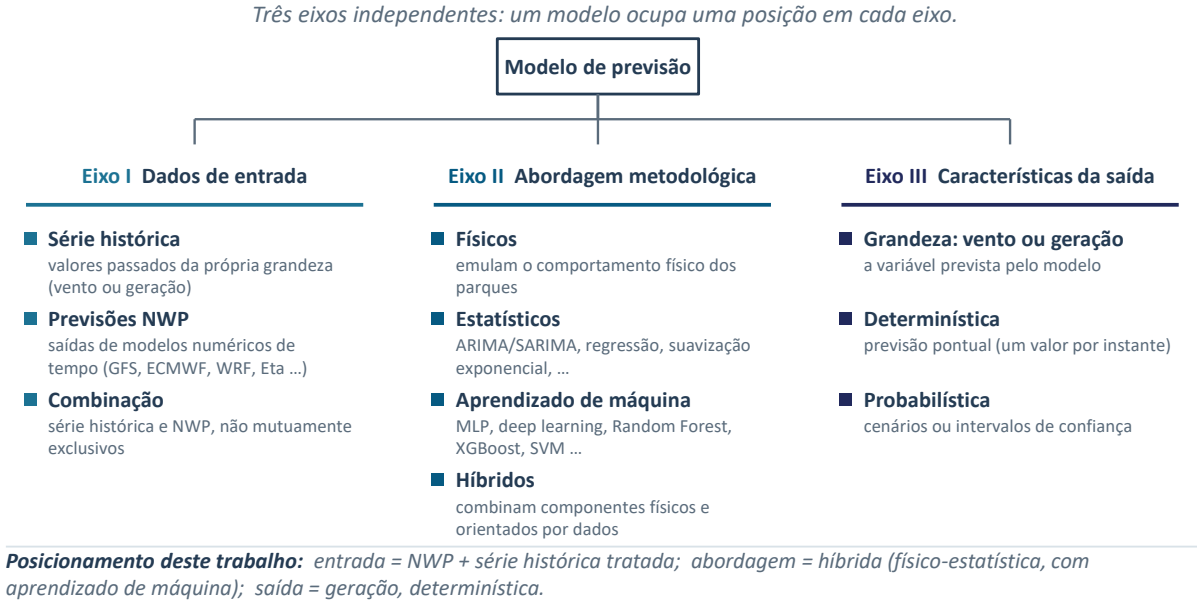
A seguir, cada eixo é descrito separadamente.

Figura 4 – Panorama das famílias de modelos de previsão de geração eólica por horizonte de previsão.



Fonte: Elaboração própria.

Figura 5 – Classificação dos modelos de previsão de geração eólica segundo os três eixos.



Fonte: Elaboração própria.

2.5.1.1 Dados de entrada

O primeiro eixo diz respeito aos dados de entrada utilizados pelo modelo [73, 74]. De modo geral, distinguem-se os modelos que utilizam apenas a série histórica de registros operacionais da própria variável a ser prevista (velocidade do vento e/ou geração) e os modelos que incorporam saídas de modelos de NWP. Os modelos baseados exclusivamente em dados históricos destacam-se pela relativa simplicidade e pela independência de previsões meteorológicas externas [73, 75]; em contrapartida, são inerentemente limitados em antecipar mudanças abruptas nas condições atmosféricas, o que restringe sua precisão em horizontes mais longos ou em contextos de alta variabilidade [42, 76]. Já os modelos que incorporam dados de NWP (como saídas do IFS (*Integrated Forecasting System* [77]), GFS (*Global Forecast System* [78]) ou WRF (*Weather Research and Forecasting Model* [79])) apresentam capacidade aprimorada de capturar mudanças meteorológicas significativas com antecedência [80], beneficiando-se da riqueza descritiva das simulações atmosféricas, embora permaneçam sensíveis à precisão e à granularidade espacial das previsões disponíveis [74, 81]. Vale destacar que esses dois conjuntos de entrada não são mutuamente exclusivos: parte dos modelos combina a série histórica com previsões de NWP.

2.5.1.2 Abordagem metodológica

O segundo eixo refere-se à abordagem metodológica adotada [73, 74]. Nesse aspecto, distinguem-se quatro grandes classes: modelos físicos, modelos estatísticos, modelos de aprendizado de máquina e modelos híbridos.

Os modelos físicos buscam representar explicitamente os processos atmosféricos e as características geográficas locais, emulando o comportamento físico dos parques a partir de suas características técnicas e de variáveis como vento, umidade e temperatura, permitindo a reprodução detalhada dos padrões de fluxo de vento. Apesar de sua alta demanda computacional e de sua dependência da qualidade dos dados de entrada, tais modelos são amplamente aplicados em contextos que requerem alta resolução espacial e temporal [81–83]. Nesta classificação, adota-se como critério a predominância da representação física na estrutura do modelo, ainda que alguns de seus parâmetros, como as parametrizações dos modelos de NWP ou as curvas de potência, sejam calibrados a partir de dados.

Os modelos estatísticos exploram relações entre variáveis a partir de dados, sem representar explicitamente a física do fenômeno. Incluem os métodos clássicos da classe *Box–Jenkins*, como ARIMA e SARIMA [84, 85], e os modelos de regressão. São reconhecidos por sua interpretabilidade e por seu bom desempenho em séries temporais, ainda que limitados na representação de relações fortemente não lineares.

Os modelos de aprendizado de máquina, por sua vez, oferecem maior flexibilidade e capacidade de capturar relações não lineares e de lidar com grandes conjuntos de dados. Entre

eles, destacam-se as redes neurais artificiais (ANN, *Artificial Neural Network*) [86], inclusive com *deep learning* [87]; as máquinas de vetores de suporte (SVM, *Support Vector Machine*) [88]; as máquinas de aprendizado extremo (ELM, *Extreme Learning Machine*) [89]; os sistemas de inferência *fuzzy*; os algoritmos genéticos (GA, *Genetic Algorithm*); e as florestas aleatórias [90], entre outros. Apesar de seu poder preditivo, frequentemente enfrentam desafios relacionados à interpretabilidade e ao risco de sobreajuste.

Cabe observar que, embora as classes estatística e de aprendizado de máquina sejam aqui distinguidas por clareza didática, ambas são, em essência, abordagens orientadas por dados (*data-driven*) e compartilham fundamentos estatísticos. Cabe ressaltar, ainda, que a fronteira entre modelos físicos e estatísticos nem sempre é rígida: modelos classificados como físicos, como os de NWP, dependem de assimilação de dados e de parametrizações calibradas empiricamente, enquanto modelos estatísticos frequentemente utilizam variáveis físicas como preditores. A classificação adotada deve, portanto, ser entendida como uma distinção de predominância, e não de exclusão.

Por fim, os modelos híbridos, cujo uso é cada vez mais comum, combinam componentes físicos e orientados por dados, visando explorar sinergias entre diferentes fontes de informação e metodologias [80, 91]. Esses modelos têm demonstrado desempenho superior em precisão e robustez em diversos cenários operacionais, embora sua calibração e complexidade operacional permaneçam um desafio técnico relevante [72, 92].

2.5.1.3 Características da saída

O terceiro eixo diz respeito às características da saída produzida pelo modelo [73, 74], podendo ser caracterizado por dois aspectos complementares.

O primeiro aspecto refere-se à grandeza prevista: alguns modelos preveem a velocidade do vento (adequados para aplicações que requerem essa variável como entrada, como na conversão posterior em geração), enquanto outros estimam diretamente a geração de energia elétrica.

O segundo aspecto refere-se ao tipo de resultado. As abordagens determinísticas fornecem uma previsão pontual, isto é, um único valor esperado para cada instante futuro, de fácil interpretação e aplicação direta. Já as abordagens probabilísticas quantificam as incertezas inerentes à previsão eólica, fornecendo intervalos de confiança ou cenários [93, 94]. Estas últimas oferecem suporte aprimorado à gestão de riscos, alinhando-se mais efetivamente aos requisitos de operação e confiabilidade do sistema elétrico [23, 95]. É importante ressaltar que o tipo de resultado independe da abordagem metodológica: uma mesma classe de modelo (física, estatística, de aprendizado de máquina ou híbrida) pode ser formulada para gerar tanto saídas determinísticas quanto probabilísticas.

Por fim, independentemente da posição nos três eixos, as previsões de geração eólica

diferem quanto ao horizonte e à granularidade temporal, conforme a aplicação pretendida [81,92]. No contexto brasileiro, as previsões de tempo real abrangem horizontes de até 24 horas à frente; as de curtíssimo prazo, horizontes de até sete dias à frente, críticos para a programação da operação; e as de curto prazo cobrem horizontes de até um mês. A granularidade temporal pode variar de minutos a várias horas, sendo resoluções horárias ou sub-horárias preferidas para a operação em tempo real e o despacho, enquanto granularidades agregadas, diárias, semanais ou mensais, são adotadas no planejamento de curto prazo.

2.5.2 Combinação de previsões

Uma vez que, em um mesmo sistema, podem coexistir diversos modelos de previsão atuando em paralelo (sejam diferentes modelos meteorológicos, sejam diferentes modelos matemáticos), e considerando que as aplicações operacionais geralmente requerem uma única previsão, torna-se necessário o desenvolvimento de algoritmos dinâmicos de combinação de previsões [80,91]. Tais algoritmos buscam consolidar, de forma ponderada, as previsões provenientes das múltiplas fontes, explorando suas complementaridades e mitigando os erros individuais. Essa estratégia é adotada no presente trabalho, conforme detalhado no Capítulo 5.

2.5.3 Posicionamento do trabalho

Posicionando o presente trabalho nos três eixos descritos, os modelos de previsão propostos utilizam como entrada tanto as previsões de NWP quanto as séries históricas tratadas (eixo de entrada); adotam uma abordagem híbrida, combinando fundamentos físicos e estatísticos (eixo metodológico); e produzem, como saída, previsões determinísticas de geração de energia (eixo de saída), em conformidade com os requisitos dos modelos de planejamento e operação oficialmente em uso no SIN, que demandam previsões pontuais. A opção por uma abordagem determinística e fundamentada em NWP justifica-se, portanto, por dois fatores: de um lado, a cadeia de modelos de otimização do SIN (MPMP, MPCP e MPCTP, atualmente o NEWAVE, o DECOMP e o DESSEM) requer previsões pontuais como insumo; de outro, avaliações comparativas conduzidas no âmbito desta pesquisa indicaram que os métodos clássicos de séries temporais, baseados apenas no histórico, apresentam desempenho limitado diante da baixa autocorrelação da geração eólica para horizontes do dia seguinte a vários dias à frente, enquanto as técnicas de aprendizado de máquina avaliadas não proporcionaram ganhos expressivos sobre a abordagem híbrida adotada [42,76]. Cabe ressaltar que técnicas de aprendizado de máquina foram avaliadas em publicações decorrentes desta pesquisa e são efetivamente empregadas no modelo de previsão em tempo real, descrito no Capítulo 5.

2.5.4 Desenvolvimentos por ISOs e instituições de pesquisa

Nas últimas décadas, diversos ISOs, bem como instituições de pesquisa internacionalmente reconhecidas, financiaram e desenvolveram modelos especializados para a previsão de

energia eólica. A seleção de dados de entrada, metodologias aplicadas e saídas esperadas varia de acordo com os requisitos operacionais e as responsabilidades de cada ISO.

Do ponto de vista metodológico, entre os modelos de base física destacam-se o SIPREÓLICO, desenvolvido pela Universidad Carlos III de Madrid (Espanha) e oficialmente adotado pelo operador espanhol REE [96]. Outros modelos representativos dessa categoria são o Prediktor, desenvolvido pelo Laboratório Nacional Dinamarquês (Dinamarca) [97]; o Previento, da Universidade de Oldenburg (Alemanha) [82, 98]; e o HIRPOM, desenvolvido pelo University College Cork (Irlanda) [99].

No contexto dos modelos estatísticos, entre os principais modelos encontram-se o WPPT, desenvolvido pela Universidade de Copenhague (Dinamarca) [100]; bem como os modelos britânicos RAL, do Appleton Rutherford Laboratory (Reino Unido), e GH Forecaster, desenvolvido pela antiga Garrad Hassan (Reino Unido) [101].

Em relação aos modelos híbridos, destaca-se o Australian Wind Energy Forecasting System (AWEFS), desenvolvido pelo Australian Energy Market Operator (AEMO) [28]. Esse sistema integra dados meteorológicos de múltiplos provedores juntamente com medições em tempo real de parques eólicos, fornecendo previsões em horizontes que vão do ultracurto prazo (atualizado a cada 5 minutos) até sete dias à frente [28]. Outras referências relevantes nessa categoria incluem o Zephyr, desenvolvido pelo Risø National Laboratory (Dinamarca) [102], e o WPMS, do ISET (atualmente Fraunhofer IEE, Alemanha) [103]. De uma perspectiva colaborativa, o consórcio europeu ANEMOS, que reúne 23 instituições de pesquisa, é notável por sua plataforma integrando metodologias tanto físicas quanto estatísticas [104]. No domínio corporativo, o modelo eWind, desenvolvido pela empresa norte-americana TrueWind USA, também é uma referência relevante [105].

No âmbito acadêmico, com destaque para o contexto nacional, diversas aplicações de técnicas estatísticas e de aprendizado de máquina à previsão de geração eólica também merecem registro, ilustrando o potencial dessas abordagens em diferentes horizontes. Em [106], foi aplicada a rede *Long Short Term Memory* (LSTM) à previsão de geração eólica, com a característica de atribuir pesos distintos a acontecimentos recentes e distantes, obtendo boa precisão e estabilidade nos resultados. Em [107], foram comparadas redes *Radial Basis Function Network* (RBFN) e *Multi-Layer Perceptron* (MLP) para prever a geração de um parque eólico a partir de suas gerações passadas, com menor erro médio absoluto para a RBFN. No trabalho desenvolvido em [108], a técnica de redes neurais artificiais foi aplicada à previsão de velocidade do vento e de geração na Usina Eólio-Elétrica de Palmas (PR), com resultados satisfatórios para o horizonte de curtíssimo prazo, assim como em [109], que aplicou a modelagem *Box & Jenkins* [85]. Já em [110], foram comparados cinco modelos, a saber: SARIMA (*Seasonal Autoregressive Integrated Moving Average*), SARIMA + *Wavelet*, redes neurais, SARIMA + redes neurais e um modelo híbrido proposto (SARIMA + duas redes neurais), sendo o modelo híbrido o de melhor desempenho. Esses trabalhos reforçam a tendência, observada na literatura,

de que abordagens orientadas por dados e, em especial, modelos híbridos tendem a apresentar bom desempenho preditivo [111].

2.6 *Curtailment* na geração eólica

O *curtailment* (corte de geração) eólica apresenta impacto relevante nos modelos e nos resultados desenvolvidos neste trabalho. Assim, algumas de suas principais características são apresentadas a seguir, com o objetivo de ilustrar os impactos desse fenômeno na previsão de geração eólica.

2.6.1 Práticas de corte de geração em escala global

2.6.1.1 Conceito e causas do corte de geração

A participação crescente de fontes de energia renovável variável, particularmente eólica e solar, no mix global de eletricidade tornou o corte de geração um desafio cada vez mais relevante para a operação de sistemas elétricos em muitos países [112].

O corte de geração é definido como a redução da geração de eletricidade de uma usina para níveis abaixo de seu potencial disponível [113]. Nesses casos, mesmo quando recursos energéticos como vento ou irradiância solar estão disponíveis, a geração é limitada por restrições operacionais ou estruturais do sistema elétrico, incluindo congestionamento de transmissão, acesso restrito à rede ou excesso de geração durante períodos de baixa demanda. Requisitos de segurança operacional, como controle de tensão, limites de interconexão e regulação de frequência, também podem levar ao corte de geração, particularmente em sistemas pequenos ou isolados [114].

De acordo com [113], o corte frequentemente ocorre devido a congestionamento de transmissão ou restrições de rede local associadas a limites térmicos ou de estabilidade, levando os operadores do sistema a reduzir a geração para prevenir sobrecargas e instabilidades. Também pode surgir de desequilíbrios entre geração e demanda na ausência de flexibilidade ou armazenamento suficiente, uma condição comum em sistemas com alta penetração solar ou eólica, bem como da necessidade de garantir inércia adequada do sistema e resposta apropriada a perturbações. Além dos fatores técnicos, o corte econômico também é observado, estando relacionado à lógica de despacho e considerações de ordem de mérito, incluindo preços de eletricidade baixos ou negativos e comportamento estratégico de oferta em mercados de energia [115].

2.6.1.2 Evolução do corte de geração em sistemas com alta penetração renovável

Em nível global, o corte de geração eólica intensificou-se em regiões como Europa, China, Canadá e Estados Unidos, acompanhando a expansão das fontes de energia renovável variável [112]. Essa tendência reflete tanto o crescimento dessas fontes quanto as limitações dos sistemas elétricos em lidar com alta variabilidade. Nesse contexto, o estudo de [112] analisa

o corte de geração por meio de uma abordagem comparativa internacional que relaciona as taxas de corte à participação da geração eólica e solar nos mixes nacionais de eletricidade. Com base em pelo menos dez anos de dados históricos, o estudo classifica os sistemas de acordo com a magnitude do corte e sua evolução temporal, mostrando que alta penetração renovável não implica necessariamente altos níveis de corte e destacando o papel da infraestrutura e da flexibilidade. Por exemplo, enquanto o corte aumentou em países como Reino Unido, Alemanha e Irlanda, diversas regiões da China experimentaram reduções apesar de maiores participações renováveis.

O corte de geração tornou-se um indicador estratégico para os operadores de sistemas de transmissão, pois reflete restrições estruturais e operacionais. Ele sinaliza a necessidade de expansão da transmissão, maior flexibilidade e serviços ancilares, além de apoiar decisões regulatórias e de mercado [112]. Do ponto de vista técnico, a capacidade limitada de transmissão tende a aumentar o corte. Em países como Alemanha e Reino Unido, maiores participações renováveis foram acompanhadas por custos aumentados e maior dependência de redespacho, destacando sua relevância para a gestão de reservas e serviços ancilares [116].

Do ponto de vista econômico, o corte de geração reduz a energia comercializável e a receita esperada de projetos de Energia Renovável Variável (VRE, do inglês *Variable Renewable Energy*), podendo afetar o desempenho financeiro e aumentar o risco do investidor. Evidências indicam que, à medida que a participação dessas fontes no sistema elétrico cresce, tanto a perda econômica por evento quanto o custo marginal do corte aumentam [117]. Para mitigar esses efeitos, mecanismos de compensação são adotados, como pagamentos por energia não despachada e instrumentos de proteção de receita, que dependem de métodos robustos para estimar a geração na ausência de corte [118].

2.6.2 Dinâmica do corte de geração no sistema elétrico brasileiro

No contexto brasileiro, caracterizado pela rápida expansão da geração eólica e sua concentração em regiões com restrições estruturais de transmissão, a estimação da geração não realizada desempenha um papel estratégico. Além de uma etapa descritiva, essa estimação apoia diretamente análises operacionais e regulatórias e contribui para maior transparência entre os agentes de mercado e o operador do sistema. Nesse sentido, modelos robustos são essenciais para aprimorar os processos de planejamento, operação e previsão.

Os resultados obtidos neste estudo devem ser interpretados à luz dos principais impulsionadores do corte de geração, que estão geralmente associados a restrições de transmissão e limitações no equilíbrio entre carga e geração [113, 119]. No Brasil, a classificação regulatória em indisponibilidade externa, confiabilidade e corte baseado em energia reflete a natureza diferente desses eventos, variando de restrições físicas externas às usinas até requisitos de segurança do sistema e condições de excesso de oferta de energia [17, 119]. Essa diversidade de causas impõe desafios adicionais para a modelagem, pois diferentes mecanismos operacionais podem produzir

efeitos distintos nos dados observados.

O aumento significativo do corte entre 2023 e 2025, particularmente em eventos relacionados à confiabilidade e à indisponibilidade externa, destaca a defasagem entre a expansão da geração renovável e a capacidade de transmissão disponível, especialmente na região Nordeste [17]. Além disso, a intensificação do corte baseado em energia durante as horas diurnas, impulsionado pela alta geração solar e pelo crescimento da geração distribuída, desloca os ajustes de equilíbrio para as usinas centralizadas, com maior impacto durante a estação de ventos [17, 120].

Nesse contexto, a estimação da geração de referência (também denominada geração contrafactual, geração estimada ou geração potencial) tem implicações diretas na aplicação de mecanismos como o *constrained-off*. A precisão dessa estimativa é essencial para garantir uma compensação financeira adequada e evitar distorções operacionais e econômicas, especialmente na presença de limitações estruturais e de incertezas associadas aos dados.

2.6.3 Abordagens para estimação de geração eólica

A estimação de geração eólica é essencial para operadores de sistema na análise de corte de geração, pois permite a separação dos efeitos causados por restrições operacionais daqueles relacionados à variabilidade natural do recurso eólico. Ao fornecer uma referência para a geração esperada na ausência de corte, essas estimativas apoiam o monitoramento operacional, análises pós-operativas e a avaliação de impactos energéticos e econômicos, além de contribuir para a tomada de decisão em tempo real, a programação da operação, o planejamento e a implementação de mecanismos de compensação. Diversas abordagens têm sido utilizadas para estimar a geração em condições de corte [121], incluindo métodos orientados por dados baseados em variáveis meteorológicas, bem como técnicas empíricas, estatísticas e de aprendizado de máquina [122]. Esta seção revisa essas abordagens, com foco no tratamento de dados, preenchimento de lacunas e estimação de geração potencial.

A geração eólica apresenta forte dependência espacial e temporal, permitindo o uso de informações de parques eólicos próximos para complementar séries temporais incompletas, reduzir ruído e melhorar a robustez das estimativas [123]. Técnicas comuns incluem o uso de parques eólicos vizinhos, interpolação espacial, agrupamento de usinas e combinações dessas estratégias para imputação de dados faltantes. Dados de parques eólicos ou estações meteorológicas próximas podem ser utilizados para calibração de modelos ou como variáveis auxiliares. Uma abordagem relevante é o uso de defasagens espaciais, que representam a dependência inter-regional por meio de médias ponderadas [124]. Na ausência de medições locais, a geração pode ser estimada combinando métodos vento-potência com técnicas de interpolação espacial como Vizinho Natural e Ponderação pelo Inverso da Distância [125, 126].

O agrupamento de parques eólicos com padrões semelhantes reduz a heterogeneidade dos dados e permite o desenvolvimento de modelos específicos por *cluster* baseados em perfis

de potência, séries de vento ou correlações entre parques [127]. Essas abordagens são particularmente úteis em cenários envolvendo lacunas de dados ou parques eólicos recém-comissionados, proporcionando benefícios operacionais relevantes [128].

Os métodos de estimação podem ser classificados em abordagens físicas, estatísticas e de aprendizado de máquina [21]. Os métodos físicos baseiam-se em características técnicas e na representação física do processo, com pouca ou nenhuma dependência de dados históricos, ainda que seus parâmetros possam ser calibrados com dados. Os métodos estatísticos exploram relações entre variáveis e são adequados para séries temporais incompletas e contextos em que a interpretabilidade é importante. Os métodos de aprendizado de máquina, também fundamentados em princípios estatísticos, oferecem maior flexibilidade e são capazes de capturar relações tanto lineares quanto não lineares de forma mais eficiente.

Os métodos baseados em curvas vento-potência modelam a relação entre velocidade do vento e potência gerada e podem ser paramétricos ou não paramétricos [129, 130]. Métodos de regressão não linear, como Modelos Aditivos Generalizados (GAM) e Splines de Regressão Adaptativa Multivariada (MARS) [130,131], são amplamente aplicados. Técnicas de aprendizado de máquina, incluindo Regressão por Vetores de Suporte (SVR), árvores de decisão e métodos de *Gradient Boosting*, particularmente XGBoost [132–134], destacam-se por sua robustez e capacidade de modelar relações complexas.

2.7 Representação eólica na cadeia de planejamento da operação do SIN

Os conceitos que regem a representação da geração eólica na cadeia de planejamento da operação estão definidos no Submódulo 2.4 (Premissas, Critérios e Metodologia para Estudos Energéticos) dos Procedimentos de Rede do ONS [135]. Nesta seção, apresenta-se a forma como as VREs são atualmente representadas no SIN.

2.7.1 Planejamento e programação da operação

No Brasil, o planejamento e a programação da operação do sistema elétrico são conduzidos por meio de três modelos principais de otimização, associados a diferentes horizontes: o Modelo de Planejamento de Médio Prazo (MPMP), atualmente o NEWAVE (Modelo de Planejamento da Operação de Sistemas Hidrotérmicos de Longo e Médio Prazo); o Modelo de Planejamento de Curto Prazo (MPCP), atualmente o DECOMP (Modelo de Coordenação da Operação de Curto Prazo); e o Modelo de Planejamento de Curtíssimo Prazo (MPCTP), atualmente o DESSEM (Modelo de Despacho Hidrotérmico de Curtíssimo Prazo em Tempo Real).

O MPMP [12] é utilizado no planejamento energético de médio e longo prazo, considerando um horizonte de até cinco anos pelo ONS e de até 30 anos pela EPE, ambos com resolução mensal. Seu objetivo é determinar a política de operação que minimize o custo esperado de

operação ao longo de todo o horizonte de planejamento.

O MPCP [11], por sua vez, é responsável pelo planejamento operacional de curto prazo (até dois meses à frente), embora possa ser utilizado em estudos que se estendem por até dois anos. Diferentemente do MPMP, o MPCP adota um arcabouço de modelagem mais detalhado, considerando restrições operacionais específicas. A discretização temporal é semanal para o primeiro mês e mensal a partir de então. Além disso, o modelo captura a variabilidade intradiária da demanda de eletricidade dividindo o dia em três patamares de carga: carga leve (períodos de baixo consumo), carga média (demanda intermediária) e carga pesada (períodos de pico).

O MPCTP [10] é utilizado no planejamento da operação de curtíssimo prazo, considerando horizontes de 24 a 168 horas (até sete dias), com resolução temporal semi-horária no primeiro dia e diária para os demais. Entre os modelos da cadeia de planejamento, é o que apresenta o maior nível de detalhamento, contemplando a representação individual das unidades geradoras e da rede elétrica completa.

Esses três modelos são oficialmente executados de forma sequencial e integrada para garantir a consistência entre decisões estratégicas e operacionais dentro do SIN. O ONS é responsável pela execução desses modelos e pela coleta e preparação de todos os dados de entrada necessários para garantir a execução adequada [135].

Para o horizonte de curto/médio prazo, atualmente, a representação da fonte eólica no Planejamento Mensal da Operação é feita com base na Resolução Normativa ANEEL nº. 843/2019, atual REN 1032/2022 [136]. De acordo com a resolução, para as usinas existentes, é calculada a média mensal do histórico dos últimos cinco anos de geração líquida disponibilizada ao SIN por cada usina. Essa média é agregada por subsistema, por mês e por patamar de carga. Cada mês do horizonte de planejamento tem o valor correspondente à média calculada do respectivo mês do histórico. No primeiro mês, o valor mensal é repetido para todas as revisões semanais do MPCP [136]. Para as usinas da expansão, é calculado um fator de capacidade por mês, por fonte e por subsistema que consiste no somatório da razão entre a média da geração líquida e a média da potência dos últimos cinco anos de cada usina, dividido pelo número total de usinas. Esse fator é usado para estimar a geração das usinas futuras e é considerado para cada mês do horizonte de planejamento.

No planejamento energético de longo prazo, a representação da geração das VREs, como a eólica, por meio de cenários é essencial devido à alta variabilidade e incerteza associadas a essas fontes [137]. Em uma matriz elétrica altamente dependente de recursos hídricos, eólicos e solares, a consideração das complementaridades entre fontes [138], em diferentes escalas espaciais e temporais, é crucial para reduzir riscos e explorar sinergias, e, portanto, deve ser incorporada à construção de cenários futuros. Por outro lado, em horizontes de mais curto prazo, previsões determinísticas são predominantemente utilizadas. No entanto, os três modelos de otimização atualmente aplicados no planejamento e programação da operação do SIN, incluindo estudos de longo prazo, dependem exclusivamente de previsões determinísticas de energia

eólica [10–12]. Consequentemente, o ONS deve contar com previsões pontuais de energia eólica em diferentes horizontes temporais para viabilizar a execução precisa dos processos de planejamento e operação do sistema [135]. Em consonância com esse requisito operacional, o modelo de previsão proposto neste trabalho adota uma abordagem determinística.

2.7.2 Representação das usinas eólicas no horizonte de curto prazo

A representação das usinas eólicas no SIN para fins de planejamento e programação da operação é regida pelos Procedimentos de Rede do ONS, que estabelecem as premissas, critérios e metodologias aplicáveis aos estudos energéticos [135].

De acordo com esse arcabouço regulatório, a previsão de geração eólica de curto prazo deve contemplar três categorias distintas de usinas [135]:

- **Usinas supervisionadas (referidas como tipos I e II):** São parques eólicos diretamente monitorados pelo ONS, que podem ou não estar sujeitos ao despacho centralizado, conforme definido no Submódulo 26.2 dos Procedimentos de Rede. Para essas unidades, estão disponíveis dados históricos de geração e medições de velocidade do vento, permitindo uma calibração mais precisa do modelo.
- **Usinas indiretamente supervisionadas (referidas como tipo III):** Este grupo compreende parques eólicos que, embora não sejam diretamente supervisionados pelo ONS, podem influenciar a operação do sistema em função de sua capacidade instalada. Para essas usinas, estão disponíveis apenas os dados históricos de geração fornecidos pela CCEE, não havendo medições diretas de vento, o que impõe restrições metodológicas aos modelos de previsão.
- **Usinas futuras ainda não em operação:** Esta categoria refere-se a projetos que devem entrar em operação dentro do horizonte de planejamento. Como séries históricas de medições não estão disponíveis, apenas informações técnicas como localização geográfica e capacidade instalada podem ser utilizadas, limitando significativamente a precisão da previsão.

Também é relevante notar que os parques eólicos de pequeno porte, incluindo aqueles categorizados sob micro e minigeração distribuída, são considerados nos estudos de planejamento e programação do SIN. No entanto, devido ao escopo deste trabalho, sua representação não será detalhada aqui.

2.8 Desafios na comparabilidade de modelos

Conforme introduzido na Seção 1.2, as informações disponíveis sobre modelos operacionais de previsão de energia eólica, tanto os utilizados internacionalmente quanto os apresentados

na literatura científica, são, na maioria dos casos, limitadas e insuficientemente detalhadas. Essa falta de transparência e padronização dificulta o desenvolvimento de análises comparativas mais robustas. A mesma limitação se aplica ao relato de resultados operacionais, que frequentemente carecem de informações consistentes e reprodutíveis.

Vale destacar alguns estudos que tentam superar essas limitações. Por exemplo, [24] apresenta uma análise estatística abrangente dos erros de previsão de energia eólica, considerando fatores como capacidade da usina, horizontes de previsão e categorias metodológicas. De forma semelhante, o estudo de [139] está entre os poucos que realizam uma comparação em nível de sistema do desempenho de previsão entre diferentes países, oferecendo uma perspectiva mais ampla sobre a eficácia dos modelos.

Adicionalmente, o trabalho de [140] compara erros de previsão entre Polônia e Portugal, destacando diferenças significativas e discutindo suas implicações para os preços do mercado de eletricidade. Em [141], a eficácia de diversas técnicas de inteligência artificial é avaliada para a previsão de energia eólica em Portugal, com redes neurais artificiais e sistemas de inferência fuzzy adaptativos apresentando desempenho superior.

2.9 Considerações finais do capítulo

Este capítulo apresentou uma revisão dos principais temas que fundamentam o trabalho, abrangendo desde a caracterização da geração eólica no Brasil e no mundo e os aspectos regulatórios da previsão, até os modelos de tratamento de dados, de ajuste de previsões numéricas de tempo, de previsão de geração e a representação eólica na cadeia de planejamento da operação do SIN. A presente seção sintetiza os principais modelos, métodos e estratégias revisados que efetivamente embasam as metodologias desenvolvidas ao longo desta tese.

Como forma de sistematizar essa síntese e posicionar o presente trabalho em relação à literatura, as Tabelas 1 e 2 comparam as principais referências correlatas segundo, respectivamente, a abrangência e a maturidade das abordagens (escala de análise, horizonte, uso de dados observados e de modelos NWP, foco em operadores de sistema, aplicação operacional e apresentação de resultados sistêmicos) e o escopo do problema abordado (tratamento de dados, estimação sob *curtailment*, correção de viés das previsões numéricas, previsão de geração e combinação de previsões). Foram consideradas apenas referências que apresentam modelos, metodologias ou resultados, excluindo-se documentos de caráter estritamente contextual ou regulatório.

A leitura conjunta das duas tabelas evidencia que a literatura, embora extensa e madura em vários eixos isolados, tende a concentrar-se em subconjuntos específicos do problema. Trabalhos voltados à previsão de geração são numerosos e frequentemente sofisticados, mas raramente incorporam o tratamento de dados em larga escala, a estimação da geração sob *curtailment* ou a validação em nível de sistema. Estudos de *curtailment*, por sua vez, caracterizam

o fenômeno em escala sistêmica, mas não o integram a um encadeamento de previsão. Da mesma forma, são escassos os trabalhos que reportam resultados efetivamente sistêmicos, apoiados em dados observados de um grande número de usinas reais. É justamente nessa lacuna que este trabalho se insere: em vez de aprimorar um único eixo, a tese propõe um arcabouço completo e encadeado, que vai do tratamento dos dados à estimação da geração, passando pela correção das previsões numéricas, pela previsão propriamente dita e pela combinação dinâmica de modelos, aplicado a mais de mil usinas eólicas e validado em nível de sistema, no contexto operacional de um operador nacional. As contribuições que sustentam cada um desses eixos, indicadas nas linhas finais das tabelas, foram formalizadas em publicações decorrentes deste trabalho.

De modo a permitir a associação direta entre essas contribuições e os eixos comparados nas tabelas, elas são resumidas nos seguintes tópicos: (i) tratamento de dados de vento e geração em larga escala, com detecção de dados espúrios e preenchimento de dados faltantes; (ii) integração e composição, em uma base única, de dados verificados de múltiplas fontes de medição; (iii) integração e composição, em uma base única, de dados previstos de diferentes modelos numéricos de previsão do tempo; (iv) estimação da geração por meio de modelos dinâmicos de conversão de vento em potência, construídos automaticamente a partir de dados medidos; (v) estimação da geração potencial (contrafactual) e quantificação das perdas associadas a eventos de *curtailment*; (vi) correção de viés das previsões numéricas de vento, específica por modelo meteorológico e por passo de tempo do horizonte; (vii) modelo único de previsão, do curtíssimo ao curto prazo, apoiado nas previsões dos modelos numéricos de tempo, sem depender exclusivamente da estrutura temporal das séries históricas de geração, do qual derivam os produtos de cada horizonte; (viii) representação conjunta de múltiplos parques por meio de agrupamento espacial, mitigando os efeitos da alta variabilidade da fonte; (ix) previsão em nível sistêmico, validada com dados reais em um dos maiores parques eólicos do mundo; (x) aplicação operacional nas ferramentas de planejamento do setor elétrico brasileiro; e (xi) *ensemble* de modelos combinados dinamicamente para a previsão em tempo real.

Tabela 1 – Síntese da literatura quanto à abrangência e à maturidade das abordagens, e posicionamento da tese.

Referência	**Escala**	**Horizonte**	**Dados observados**	**Modelos NWP**	**Foco em operador (ISO)**	**Aplicação operacional**	**Resultados sistêmicos**
[73, 81]	revisão	diversos	–	○	○	○	–
[21, 92]	revisão	diversos	–	○	–	–	–
[74]	revisão	diversos	○	○	○	○	–
[24]	múltiplos	diversos	●	●	○	○	–
[98]	múltiplos	curtíssimo	●	○	●	●	–
[96]	sistema	curtíssimo	●	○	●	●	○
[104]	múltiplos	curtíssimo	●	●	●	●	–

continua na próxima página

(continuação da Tabela 1)

Referência	**Escala**	**Horizonte**	**Dados observados**	**Modelos NWP**	**Foco em operador (ISO)**	**Aplicação operacional**	**Resultados sistêmicos**
[102]	múltiplos	curtíssimo	●	●	●	●	–
[95]	múltiplos	curtíssimo	●	○	○	○	–
[80]	múltiplos	curtíssimo	●	●	–	–	–
[87]	múltiplos	curtíssimo	●	○	–	–	–
[72, 86]	único	curtíssimo	●	○	–	–	–
[58]	único	tempo real	●	○	–	–	–
[142, 143]	–	–	–	–	–	–	–
[62, 144]	–	curtíssimo	●	●	○	○	–
[53]	múltiplos	–	●	–	○	○	–
[55, 56]	múltiplos	curtíssimo	●	○	○	○	–
[76]	múltiplos	curtíssimo	●	○	●	○	–
[145]	único	–	●	–	–	–	–
[112, 113]	sistema	–	●	–	●	●	○
[139]	sistema	curtíssimo	●	●	●	●	○
[26]	sistema	–	●	–	●	●	○
[146]	sistema	–	●	–	●	●	○
Esta tese	**sistema (>1000)**	**tempo real–curto**	●	●	●	●	●
Contribuição desta tese	viii [30–33, 38]	vii [32, 37, 40]	i [30, 39]	iii [32]	x [32–36, 38]	x, xi [30, 32, 34]	ix [30–33, 37]

Legenda: ● aborda; ○ aborda parcialmente; – não aborda.

Os algarismos romanos remetem às contribuições listadas anteriormente.

Fonte: Elaboração própria.

Tabela 2 – Síntese da literatura quanto ao escopo do problema abordado, e posicionamento da tese.

Referência	**Integração de múltiplas fontes**	**Tratamento de dados**	**Estimação sob *curtailment***	**Correção de viés de NWP**	**Previsão de geração**	**Combinação de previsões**
[73, 81]	–	○	–	○	●	○
[21, 92]	–	○	–	○	●	●
[74]	–	○	–	○	●	○
[24]	○	○	–	○	●	–
[98]	○	●	–	○	●	○
[96]	○	●	–	○	●	○
[104]	○	○	–	○	●	●
[102]	○	○	–	○	●	○
[95]	–	○	–	○	●	●

continua na próxima página

(continuação da Tabela 2)

Referência	Integração de múltiplas fontes	Tratamento de dados	Estimação sob *curtailment*	Correção de viés de NWP	Previsão de geração	Combinação de previsões
[80]	–	○	–	–	●	○
[87]	–	●	–	○	●	○
[72, 86]	–	○	–	○	●	●
[58]	–	○	–	–	○	●
[142, 143]	–	–	–	–	–	●
[62, 144]	–	–	–	●	–	–
[53]	○	●	–	–	–	–
[55, 56]	○	●	–	○	●	○
[76]	–	●	–	–	●	–
[145]	–	○	–	–	–	–
[112, 113]	○	–	○	–	–	–
[139]	○	–	–	–	●	–
[26]	–	–	○	–	–	–
[146]	○	○	○	–	–	–
[17, 119]	–	–	●	–	–	–
Esta tese	●	●	●	●	●	●
Contribuição desta tese	ii [30, 32, 36]	i [30, 39]	v [31, 35, 41]	vi [32, 42]	iv, vii [32, 33, 42]	xi [34, 42]

Legenda: ● aborda; ○ aborda parcialmente; – não aborda.

Os algarismos romanos remetem às contribuições listadas anteriormente.

Fonte: Elaboração própria.

Estabelecido esse panorama comparativo, os parágrafos seguintes retomam, eixo a eixo, os principais métodos e estratégias revisados neste capítulo, destacando de que forma cada um deles fundamenta as metodologias desenvolvidas nos capítulos subsequentes. No que se refere ao tratamento de dados, a revisão evidenciou a importância de métodos estatísticos e de aprendizado de máquina para a detecção e correção de dados espúrios e o preenchimento de dados faltantes, com destaque para abordagens baseadas em modelos lineares dinâmicos e no filtro de Kalman em formulação de espaço de estados. Esses conceitos sustentam diretamente o modelo de tratamento adotado na tese, no qual o filtro de Kalman, sob uma abordagem bayesiana, é empregado em conjunto com a composição hierárquica de múltiplas fontes de medição e o preenchimento de lacunas baseado em correlação espacial entre parques vizinhos.

Quanto ao ajuste das previsões numéricas de tempo, foram revisadas técnicas de pós-processamento que vão desde abordagens simples, como o *Delta Change* e o escalonamento linear, até o mapeamento quantil-quantil e suas variações, o *Model Output Statistics* (MOS) e abordagens baseadas em redes neurais. Esse panorama embasa a estratégia adotada na tese para a dessazonalização e a correção de viés das previsões de vento dos modelos NWP, com

destaque para o uso de regressões e janelas móveis adaptativas. A revisão dos próprios modelos meteorológicos (Eta, WRF e BAM (CPTEC), GFS/GEFS (NOAA/NCEP) e IFS (ECMWF)) também orientou a seleção das fontes de previsão efetivamente utilizadas.

Em relação aos modelos de previsão de geração, a revisão classificou as abordagens segundo os dados de entrada, a metodologia (física, estatística, de aprendizado de máquina ou híbrida) e as características da saída (determinística ou probabilística, de vento ou de potência). Esse arcabouço sustenta as escolhas metodológicas da tese, que adota uma abordagem híbrida, combinando informações físicas dos modelos meteorológicos com técnicas estatísticas e de aprendizado de máquina. Nesse contexto, integram o conjunto de métodos investigados e empregados ao longo do trabalho técnicas clássicas de séries temporais, como modelos da classe *Box–Jenkins* (ARIMA e ARIMAX) e a suavização exponencial; modelos de regressão linear; métodos de aprendizado de máquina, como redes neurais artificiais (MLP), *Random Forest* e *XGBoost*, tanto em formulações por horizonte (*step*) quanto de múltiplas saídas (*multi-output*); e estratégias de persistência em suas diferentes formas (pura, inteligente e com derivada). Adicionalmente, dada a coexistência de múltiplos modelos meteorológicos e matemáticos, a revisão reforçou a necessidade de algoritmos dinâmicos de combinação de previsões, estratégia adotada na tese por meio de esquemas de ponderação, como a combinação pelo inverso do erro e a combinação baseada no filtro de Kalman.

Por fim, a revisão dos aspectos regulatórios, da dinâmica de *curtailment* no sistema brasileiro e da representação eólica na cadeia de planejamento da operação (MPMP, MPCP e MPCTP, atualmente o NEWAVE, o DECOMP e o DESSEM) contextualiza as aplicações desenvolvidas, evidenciando os requisitos operacionais e os diferentes horizontes temporais que as metodologias propostas nos capítulos seguintes buscam atender.

3 MODELOS PARA TRATAMENTO DE DADOS

Neste capítulo, será apresentado o detalhamento da metodologia implementada para seleção e tratamento dos dados históricos observados, além do modelo de construção das séries históricas necessárias aos processos de estimação e previsão.

Os dados utilizados neste trabalho referem-se a dados medidos e previstos. Os dados medidos serão descritos como dados verificados ou dados observados. São três os tipos de dados necessários para os modelos de previsão e/ou estimação que serão apresentados: (i) dados de geração verificada; (ii) dados de vento verificado; e (iii) dados de vento previsto. Cada um desses conjuntos possui características específicas, exigindo procedimentos de tratamento distintos. Além desses, informações complementares, como a disponibilidade de potência dos parques eólicos e os dados cadastrais das usinas, também desempenham papel importante no desenvolvimento e na aplicação dos modelos propostos.

Os dados observados do ONS são recebidos dos parques eólicos através do SSC (Sistema de Supervisão e Controle) em tempo real (a cada 4 segundos). Esses dados são adquiridos pelo PI System, que integra o SSC utilizado pela Rede de Gerenciamento de Energia (REGER), a qual, por sua vez, integra uma rede de sistemas SCADA/EMS (*Supervisory Control and Data Acquisition System / Energy Management System*).

A construção de modelos de estimação e previsão de geração eólica com boa acurácia depende diretamente da qualidade e da consistência dos dados históricos utilizados em sua calibração. Nesse sentido, foi desenvolvido um modelo de tratamento dos dados históricos, cujo fluxo metodológico é ilustrado na Figura 6 e descrito de forma resumida a seguir.

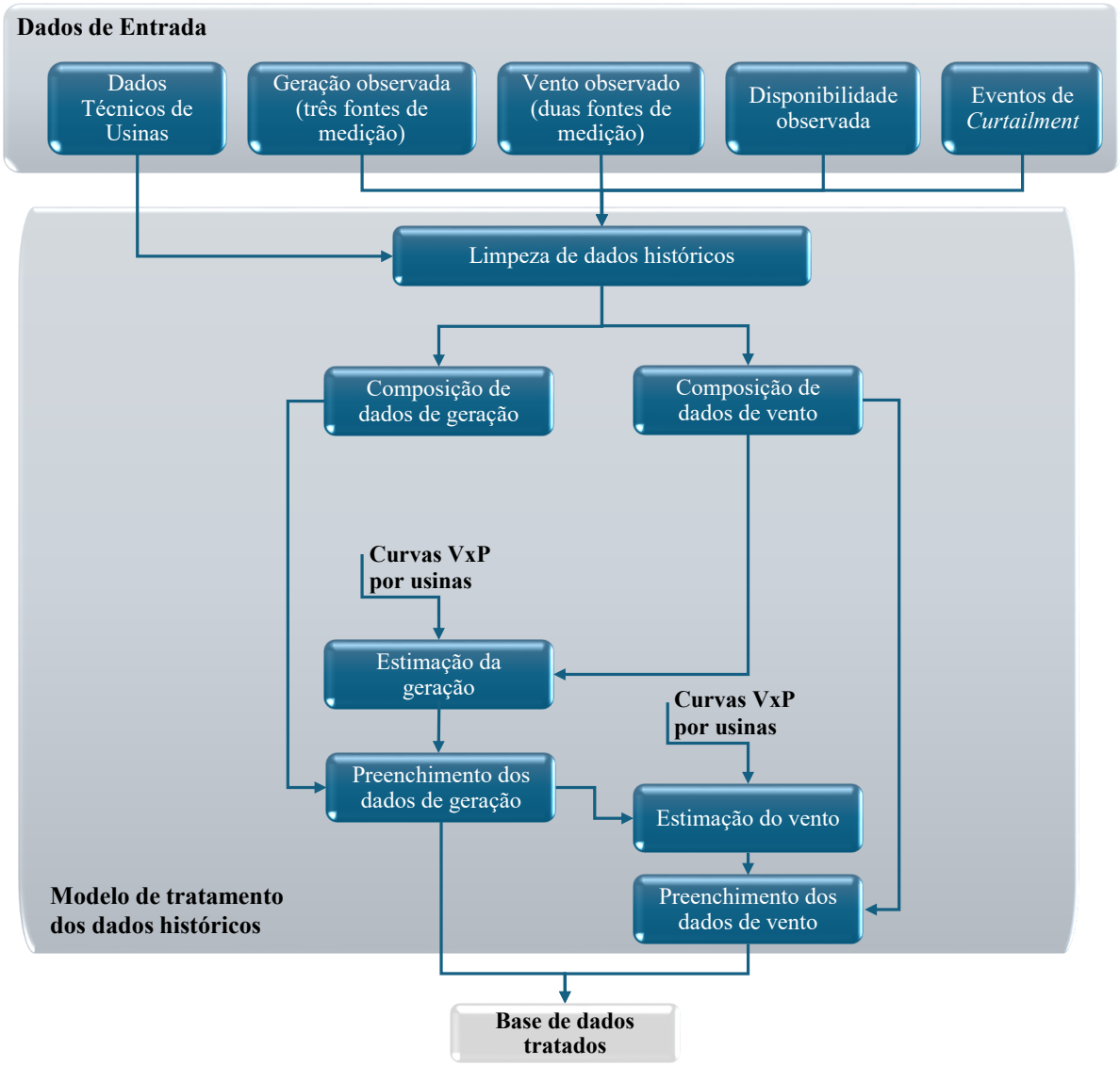
O processo recebe como dados de entrada cinco conjuntos de informações: (i) dados técnicos das usinas, (ii) geração observada, proveniente de duas fontes distintas de medição, (iii) vento observado, obtido a partir de duas fontes de medição, (iv) disponibilidade observada das unidades geradoras, e (v) registros de eventos de *curtailment*.

Na primeira etapa, todos esses dados passam por um procedimento de limpeza, no qual inconsistências, valores espúrios e registros incompatíveis são identificados e tratados. Em seguida, o fluxo se divide em dois ramos paralelos: a composição dos dados de geração e a composição dos dados de vento, nos quais as múltiplas fontes de medição são consolidadas em séries únicas e coerentes para cada usina.

A partir das séries compostas e das curvas de potência em função da velocidade do vento (curvas V $\times$ P) de cada usina, são realizadas estimações cruzadas: a geração é estimada a partir dos dados de vento e, de forma análoga, a velocidade do vento é estimada a partir dos dados de geração, utilizando as mesmas curvas. Essas estimações permitem o preenchimento de lacunas nas séries históricas de geração e de vento. Adicionalmente, mecanismos complementares são empregados para garantir a completude das séries, eliminando a ocorrência de dados faltantes.

O resultado final desse processo é uma base de dados tratados, consistente e completa, que serve como insumo fundamental para as etapas subsequentes de calibração e validação dos modelos de previsão de geração eólica. As seções a seguir detalham cada uma das etapas descritas.

Figura 6 – Processo de limpeza e composição dos dados históricos, incluindo as estimações cruzadas entre geração e vento.



Fonte: Elaboração própria.

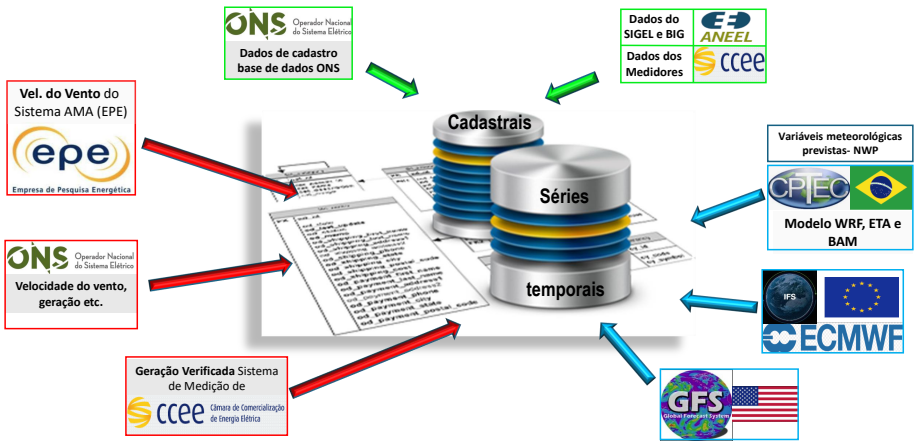
3.1 Dados de entrada

Conforme comentado previamente neste documento, as informações observadas e cadastrais das usinas provêm de diversas instituições do setor, reunindo-se todos os insumos necessários ao desenvolvimento dos modelos apresentados neste trabalho.

Os dados históricos de velocidade do vento e geração verificada foram obtidos do SSC do ONS, alguns disponíveis em [147]. Para completar a cobertura dos dados históricos de geração verificada foram utilizados dados em base horária e de 5 minutos provenientes do Sistema de Medição para Faturamento da CCEE [148], sendo o dado em base horária o mais consistido. Em relação aos dados históricos verificados de vento foram utilizados os dados do AMA, gerenciado pela EPE. As medições de vento do ONS, realizadas em naceles ou em torres, e as da EPE, obrigatoriamente realizadas em torres anemométricas, referem-se às alturas de instalação dos sensores em cada usina, que em geral se situam em torno de 100 metros, mas podem divergir desse valor. As informações cadastrais das usinas, por sua vez, foram obtidas do SIGA (Sistema de Informações de Geração), disponibilizado pela ANEEL [149]. A Figura 7 ilustra as informações que são utilizadas de cada instituição.

É importante distinguir o papel de cada base de dados no processo de tratamento. Os dados de geração e de vento do ONS constituem a principal base utilizada neste trabalho. Por serem transmitidos pelas usinas em tempo real, estão mais suscetíveis a falhas de comunicação, perdas de informação e inconsistências operacionais. Por outro lado, os dados de geração da CCEE não são enviados em tempo real, o que reduz significativamente a ocorrência desse tipo de problema. Situação semelhante ocorre com os dados de vento da EPE, que são medidos, armazenados e encaminhados periodicamente em lotes, com intervalos de vários dias, em vez de serem transmitidos continuamente em tempo real. Essa sistemática reduz impactos associados a falhas de comunicação e indisponibilidades temporárias de coleta. Dessa forma, as bases da CCEE e da EPE não são utilizadas como substitutas das informações do ONS, mas sim como fontes complementares, com características e propósitos distintos. Sua utilização permite realizar validações cruzadas, identificar inconsistências nos dados do ONS e recuperar informações ausentes quando necessário.

Figura 7 – Dados reunidos para o desenvolvimento dos modelos de previsão eólica.



Fonte: Elaboração própria.

Como pode ser visto na Figura 7, além dos dados observados e cadastrais, são reunidos previsões de vento de alguns modelos numéricos de tempo, provenientes de instituições nacionais e internacionais, destacando o CPTEC (Centro de Previsão do Tempo e Estudos Climáticos) [150], o ECMWF (*European Centre for Medium-Range Weather Forecasts*) [77] e o NCEP (*National Centers for Environmental Prediction*) [78].

Parte dos dados utilizados neste estudo é confidencial e pertence a empresas do setor elétrico brasileiro; ainda assim, boa parte deles está publicamente disponível. As informações cadastrais das usinas (potência instalada, coordenadas, data de entrada em operação, entre outras) são de acesso aberto por meio do SIGA [149], assim como os dados de geração verificada, disponibilizados nos portais de dados abertos do ONS [147] e da CCEE [148]. Já os dados anemométricos do AMA são de propriedade da EPE e não são de acesso aberto.

Quanto aos modelos numéricos de previsão do tempo, o GFS e o GEFS são públicos tanto em suas versões operacionais quanto em seus arquivos históricos [78]. O ECMWF, por outro lado, não é de acesso público em sua operação, exigindo aquisição, embora diversos de seus históricos estejam disponíveis [77]. Os modelos Eta, WRF e BAM, por sua vez, são desenvolvidos pelo CPTEC e disponibilizados publicamente por meio de seu Sistema de Transferência de Dados [150].

A Tabela 3 sintetiza, para cada conjunto de dados, a instituição de origem, as discretizações espacial e temporal, o horizonte (no caso das previsões) e a forma de disponibilização.

Tabela 3 – Proveniência, discretização e disponibilidade dos dados utilizados.

Dado	**Instituição (fonte)**	**Discr. espacial**	**Discr. temporal**	**Horizonte**	**Disponibilidade**
Informações técnicas e cadastrais	ANEEL / ONS	–	–	–	Público [149]
Geração verificada	ONS (SSC)	Usinas	4 s	–	Parcialmente público [147]
Geração verificada	CCEE (SMF)	Usinas	Horária e 5 min	–	Público [148]
Vento verificado	ONS (SSC)	Usinas	4 s	–	Parcialmente público [147]
Vento verificado	EPE (AMA)	Usinas	10 min	–	Privado
Disponibilidade e *curtailment*	ONS	Usinas	Minuto	–	Parcialmente público [147]
Previsão de vento (IFS-HRES)	ECMWF	$\sim$10 km	1 h; 3 h; 6 h	10 dias	Restrito (aquisição paga) [77]
Previsão de vento (IFS-ENS)	ECMWF	$\sim$20 km	3 h	8 dias	Restrito (aquisição paga) [77]
Previsão de vento (GFS)	NCEP / NOAA	$\sim$25 km	1 h; 3 h	15 dias	Público [78]
Previsão de vento (GEFS)	NCEP / NOAA	$\sim$50 km	3 h; 6 h	35 dias	Público [78]
Previsão de vento (Eta/WRF)	CPTEC	$\sim$5 km	1 h	3 dias	Público [150]
Previsão de vento (BAM)	CPTEC	$\sim$18 km	1 h	10 dias	Público [150]

Fonte: Elaboração própria.

Embora esta tese cite e contextualize um conjunto amplo de modelos numéricos de previsão do tempo, nem todos foram incluídos na avaliação final dos resultados. Os modelos Eta, WRF e BAM foram avaliados e utilizados em diferentes etapas da pesquisa, além de constarem

em trabalhos e publicações decorrentes, o que justifica sua presença na revisão bibliográfica e na Tabela 3.

O modelo Eta, em sua configuração de 5 km, foi descontinuado como produto operacional do CPTEC e substituído pelo WRF. Em razão disso, não há histórico compatível com todo o período analisado nesta tese, o que inviabilizou sua utilização nas avaliações finais. Ainda assim, o modelo foi empregado em estudos anteriores e em artigos publicados ao longo da pesquisa.

Já os modelos WRF e BAM, embora apresentem resultados satisfatórios para outras variáveis meteorológicas, mostraram desempenho inferior na previsão da velocidade do vento quando comparados aos modelos adotados na solução final (ECMWF, GFS e GEFS). Sua inclusão nos modelos de combinação acabou introduzindo ruído adicional, motivo pelo qual foram desconsiderados na etapa final de desenvolvimento.

Dessa forma, os modelos Eta, WRF e BAM são apresentados e contextualizados ao longo da tese, mas a avaliação final concentra-se exclusivamente nos modelos ECMWF, GFS e GEFS.

3.2 Limpeza de dados históricos

Conjuntos de dados reais frequentemente apresentam dados faltantes (*missing data*) e/ou dados espúrios, também chamados de dados aberrantes (*outliers*), o que pode comprometer a qualidade dos resultados. Nos dados obtidos da base histórica de tempo real do ONS, através do *SCADA*, a situação não é diferente. Sendo assim, os dados que serão utilizados podem apresentar características indesejadas aos modelos de previsão e que precisam ser tratadas. Esta seção visa apresentar as etapas de identificação e remoção de dados considerados inválidos.

As falhas observadas têm origens diversas. Entre as principais causas, destacam-se os problemas de comunicação e a perda de enlace na transmissão em tempo real, as falhas de equipamentos e de sensores, as medições congeladas, em que o valor permanece constante por travamento do sensor ou do canal de aquisição, e a baixa qualidade histórica da infraestrutura de medição em parte das usinas.

3.2.1 Limpeza dos dados de geração verificada

Os dados obtidos da base histórica de tempo real do ONS, através do *PI System*, possuem indicadores (*flags*) de qualidade das medidas coletadas. Dessa forma, já é possível identificar uma série de problemas de origem das medições nos parques eólicos; esta etapa pode ser considerada como um pré-filtro.

Especificamente em relação aos dados de geração, é possível contar com informações estimadas por modelos de estimadores de estado. Esses estimadores de estado são modelos matemáticos que, através de cálculos numéricos e estatísticos, processam e manipulam dados redundantes medidos provenientes da rede, para assim estimar algumas grandezas elétricas.

As falhas nos dados de geração são, de forma geral, identificadas na origem pelos *flags* de qualidade e substituídas pelos valores de geração provenientes do estimador de estado. Porém, como ainda persistem problemas que não são capturados nesta etapa, é necessária a identificação e retirada dos dados espúrios da série histórica de geração verificada.

O processo de identificação dos dados considerados espúrios pode ser resumido nas seguintes etapas:

I. **Identificação por avaliação cruzada entre os dados do ONS e os dados da CCEE**

Esta etapa busca identificar discordâncias entre os dados de geração do ONS e os da CCEE, no patamar de uma hora, tratando as duas bases como fontes independentes de uma mesma grandeza. A CCEE não é tomada como referência absoluta para validar o ONS: a comparação serve para sinalizar os pontos em que as duas fontes divergem e que, por isso, devem ser descartados do histórico do ONS. Para isso, a geração do ONS é integralizada em uma hora e ajusta-se uma regressão linear entre ela e a geração da CCEE. A partir dos parâmetros da regressão, utiliza-se o ângulo calculado para criar retas limites, que definem a diferença máxima aceita entre os dados do ONS e os dados horários da CCEE.

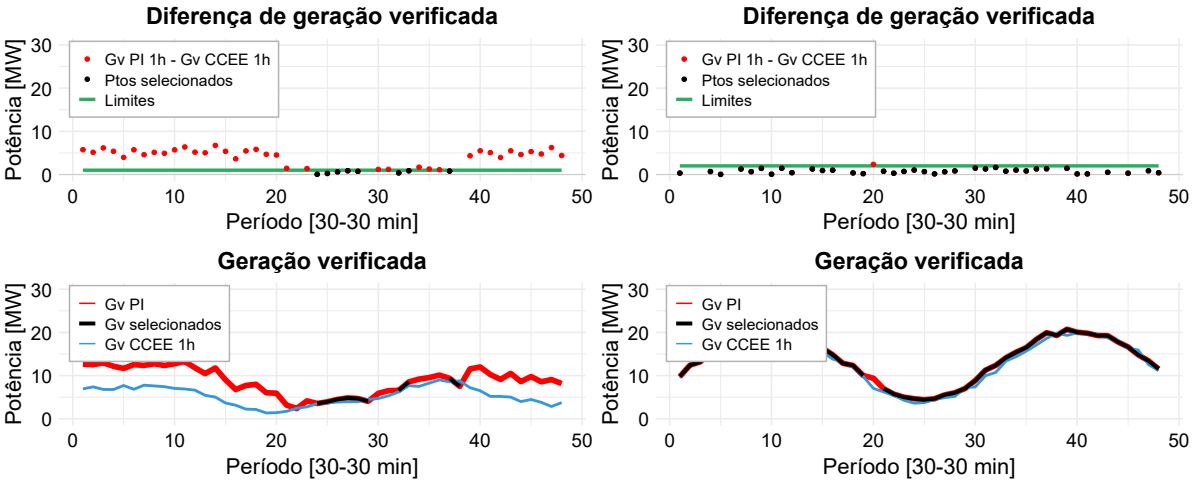
A Figura 8 ilustra dois exemplos de curvas referentes a dias distintos de geração (gráficos inferiores). Nos gráficos superiores, a reta em verde, obtida a partir da regressão mencionada, representa o limite máximo aceitável para as diferenças entre os dados horários de geração do PI/ONS (curva vermelha nos gráficos inferiores) e os dados horários da CCEE (curva azul nos gráficos inferiores). Os pontos em vermelho representam essas diferenças após a integralização da geração em intervalos horários.

Os pontos em preto nos gráficos superiores e as curvas em preto nos gráficos inferiores correspondem aos dados do ONS considerados válidos após a aplicação do critério de consistência. Na Figura 8, bem como em outras figuras deste capítulo que apresentam curvas diárias, o eixo horizontal representa os 48 períodos de 30 minutos que compõem um dia, correspondentes ao intervalo entre 00:00 e 23:30.

Observa-se que, no exemplo à esquerda, existem desconformidades significativas entre as bases de dados. O ONS apresenta valores sistematicamente superiores aos da CCEE no início e no final do dia, resultando em diferenças que ultrapassam amplamente o limite estabelecido. Como consequência, a maior parte dos pontos é descartada, permanecendo válida apenas a faixa central da curva, na qual as duas bases coincidem. Já no exemplo à direita, as diferenças entre os dados do ONS e da CCEE permanecem dentro dos limites aceitáveis na quase totalidade do dia, fazendo com que praticamente todos os pontos do ONS sejam classificados como válidos.

Uma opção mais simples para estimar o fator que limita o montante do desvio entre os dados seria utilizar a correlação diretamente como alternativa ao uso do ângulo da reta

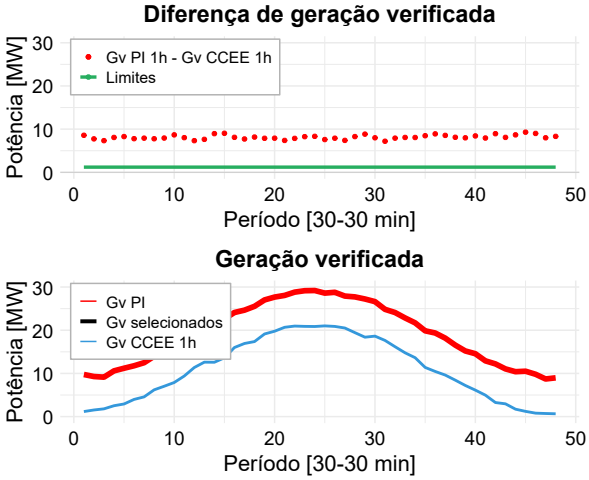
Figura 8 – Identificação de não conformidade entre a geração horária do ONS e CCEE.



Fonte: Elaboração própria.

ajustada. Entretanto, certas curvas apresentam ótima correlação, mas são bem distintas entre si, conforme observado na Figura 9, por vezes apresentando apenas um viés constante entre elas. Na figura, o coeficiente de correlação entre as duas curvas é de aproximadamente 0,98 (ou seja, elas variam de forma quase idêntica ao longo do dia), porém apresentam um viés sistemático considerável: a curva do ONS permanece consistentemente acima da curva da CCEE. Como consequência, a diferença ponto a ponto ultrapassa o limite estabelecido em todos os instantes do dia, e nenhum ponto do ONS é considerado válido, conforme se observa no gráfico superior pela ausência de pontos pretos sobre os pontos vermelhos, e no gráfico inferior pela ausência da curva preta.

Figura 9 – Identificação de não conformidade com alta correlação entre curvas de geração do ONS e da CCEE.



Fonte: Elaboração própria.

II. **Identificação de falhas no Estimador de Estados**

Algumas vezes o estimador apresenta resultados incoerentes para a geração verificada. Isso ocorre, geralmente, quando as informações redundantes utilizadas não são de qualidade, ou seja, o volume de falhas de medição passa a ser extenso demais, comprometendo outras medidas e a redundância do processo. Nestes casos, o estimador pode até gerar valores negativos para a geração. Além disso, na vizinhança desses valores, a curva de geração fica distorcida. Como é uma tarefa complicada identificar valores incoerentes dos estimadores, sempre que ocorrem valores negativos, esses são eliminados, assim como os dois valores vizinhos no tempo.

III. **Identificação de dados fora de limites físicos**

As séries históricas de geração verificada muitas vezes apresentam valores espúrios, com dados negativos e até acima da capacidade instalada da usina eólica. Dessa forma, são eliminados os valores negativos de geração, bem como os valores de geração superiores à capacidade instalada acrescida de 10%, de acordo com a inequação:

$$G_v < 0\,\text{MW} \text{ ou } G_v > 1{,}1 \cdot P_{inst}\,\text{MW} \tag{3.1}$$

em que G_v é a geração verificada e P_{inst} a potência instalada. A tolerância de 10% deve-se às situações reais em que a usina eólica pode gerar acima da potência declarada, seja por operação momentânea acima da capacidade nominal dos aerogeradores em condições ambientais favoráveis, seja pela incerteza inerente à medição. Esse valor representa uma folga usualmente adotada na prática do setor, suficiente para acomodar tais efeitos sem descartar dados fisicamente plausíveis.

IV. **Identificação de dados congelados**

Dados congelados representam um grande volume nas falhas encontradas nesses dados. Podem existir duas situações de dados congelados: (i) quando os dados verificados de geração apresentam-se congelados ao longo do tempo, podendo ficar congelados por algumas ou muitas horas; (ii) quando existe a variação do dado ao longo das horas de um dia, porém com repetição ao longo dos dias. Nesses casos, os dados são identificados e eliminados da série histórica.

V. **Identificação de dias com dados não representativos**

A última etapa de eliminação de dados ocorre pois os filtros anteriores acabam deixando passar algumas informações com aparente comprometimento. Desta forma, quando são eliminados muitos dados de um mesmo dia, durante as etapas anteriores, há grande chance de os dados serem de má qualidade. Assim, definiu-se um número mínimo de dados válidos para que os dados do dia sejam considerados válidos. Para a discretização de trinta minutos, têm-se 48 observações em um dia; é necessário que existam pelo menos cinco valores válidos, ou seja, se o dia não tiver pelo menos o correspondente a duas horas e meia de dados válidos, todo o dia é descartado.

3.2.2 Limpeza dos dados de vento verificado

Assim como os dados de geração, os dados de vento verificados obtidos do ONS possuem indicadores de qualidade das medições. No entanto, por não se tratarem de medidas elétricas, não existem estimadores capazes de reproduzir os valores nos momentos de erros na coleta dos dados. Como resultado, o volume de dados comprometidos nas séries históricas de vento é significativamente maior do que nas séries de geração, o que torna a aplicação do processo de limpeza indispensável. O processo de seleção dos dados válidos de vento verificado pode ser resumido em algumas etapas:

I. **Identificação por avaliação cruzada entre os dados de vento verificado do ONS, com o vento verificado da EPE e com a geração da CCEE**
Esta etapa busca identificar inconsistências nos dados de vento do ONS por meio da comparação com duas fontes independentes: o vento medido da EPE, que representa a mesma grandeza na mesma usina, e a geração verificada da CCEE. Assim como na geração, essas bases não são tomadas como referência absoluta, servindo apenas para sinalizar os dias em que o vento do ONS é incoerente com as demais informações. Para isso, calcula-se a correlação do vento do ONS com cada uma delas, e os dados do ONS do respectivo dia são descartados quando as inequações a seguir não são atendidas:

$$\mathrm{Cor}\big(V_{v,\mathrm{ONS}}, V_{v,\mathrm{EPE}}\big) > 0{,}6 \;\text{ e }\; \mathrm{Cor}\big(V_{v,\mathrm{ONS}}, G_{v,\mathrm{CCEE}}\big) > 0{,}3 \tag{3.2}$$

em que $V_{v,\mathrm{ONS}}$ representa o vento verificado do ONS, $V_{v,\mathrm{EPE}}$ o vento verificado da EPE e $G_{v,\mathrm{CCEE}}$ a geração verificada da CCEE em patamares de uma hora. Os valores 0,6 e 0,3 foram estimados empiricamente após exaustivos testes para um número extenso de usinas eólicas. A diferença entre os dois limiares é esperada: a primeira condição compara duas medições da mesma grandeza (vento) na mesma usina, das quais se admite correlação relativamente alta, justificando o limiar de 0,6; já a segunda relaciona grandezas distintas (vento medido e potência gerada), cuja dependência é intrinsecamente não linear, em razão da forma da curva de conversão de vento em potência, o que atenua a correlação linear e torna adequado um limiar bem menor, de 0,3.

II. **Identificação de dados fora de limites físicos**
Algumas vezes são observados valores espúrios nas séries históricas de vento verificado, tais como valores negativos e valores muito altos. Assim sendo, esses dados são eliminados sempre que os valores de vento são negativos ou maiores que trinta metros por segundo. Embora ventos acima de trinta metros por segundo possam ocorrer, isso não é frequente; além disso, os parques eólicos não geram para esse patamar de velocidade do vento. A maioria dos casos com registros acima de trinta metros por segundo deve-se a erros de medição e não a eventos extremos. Dessa forma, esses dados são eliminados quando violam a inequação:

$$V_v < 0\,\mathrm{m/s} \text{ ou } V_v > 30\,\mathrm{m/s} \tag{3.3}$$

em que V_v representa o vento verificado.

III. **Identificação de dados congelados**
O dado de vento medido pode estar congelado ao longo das horas de um dia, ou pode variar ao longo das horas, mas com repetição ao longo dos dias. Portanto, esses dados são identificados e eliminados da série histórica, sendo classificados como falha. Esse problema ocorre com muita frequência nas séries históricas de vento verificado, sendo o seu tratamento um ponto de atenção.

IV. **Identificação de dias com dados não representativos**
Quando há muitos dados eliminados ao longo de um dia, há uma grande probabilidade de o dado diário não ser de boa qualidade, o que foi verificado em muitos casos. Portanto, para a discretização de trinta minutos, com 48 observações em um dia, é necessário que existam pelo menos cinco valores válidos, ou seja, se o dia não tiver pelo menos o correspondente a duas horas e meia de dados válidos, todo o dia é descartado. Trata-se de um critério deliberadamente permissivo, que corresponde a cerca de 10% das observações diárias e visa descartar apenas os dias praticamente sem dados válidos, preservando o máximo de informação para as etapas posteriores de composição e preenchimento das séries.

3.3 Composição e preenchimento dos históricos de geração

Após a etapa de identificação e remoção dos dados espúrios, a próxima etapa é selecionar os dados que irão compor as séries temporais e tratar o que for necessário antes da entrada no modelo de previsão, sejam elas as séries de vento ou geração verificada.

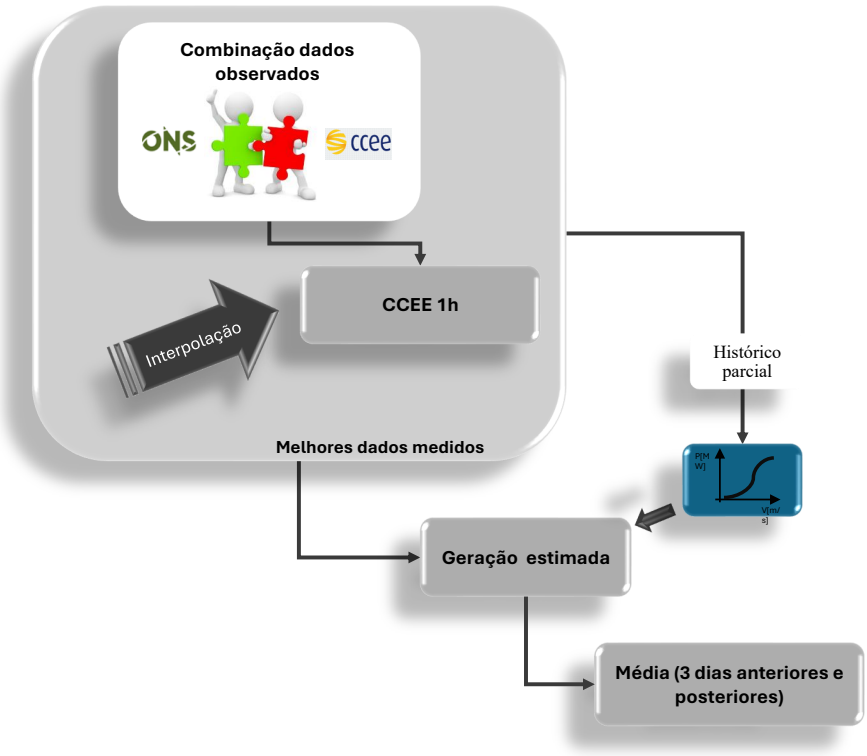
3.3.1 Composição dos históricos de geração

Como mencionado anteriormente, os dados de geração verificada possuem mais de uma origem; dessa forma, tem-se redundância dessas informações, relembrando, são de origem no histórico de tempo real (*PI System*) e do SMF da CCEE. Logo, é possível montar uma série histórica de geração verificada em intervalos de 30 minutos, unificada, com os melhores dados possíveis das duas fontes. Ainda assim, nem sempre é possível obter os dados medidos para todos os intervalos de tempo a partir destas fontes. Para tanto, foram criados alguns processos de reconstrução das informações de geração verificada. A Figura 10 destaca a ordem das opções consideradas no processo para composição do histórico de geração.

O processo de montagem do histórico de geração, apresentado na Figura 10, considera cinco opções para preenchimentos de falhas. Dessa forma, em cada intervalo de discretização, o dado é preenchido sequencialmente de acordo com a ordenação a seguir:

I. **Dado do ONS**
Uma vez que o dado do ONS, de forma geral, apresenta boa qualidade, os dados do

Figura 10 – Processo de montagem do histórico de geração verificada, com preenchimento sequencial das falhas.



Fonte: Elaboração própria.

histórico são preenchidos, como inicialização, em todas as posições em que esse dado do ONS seja válido.

II. **Dado da CCEE (discretização 5 min)**
Se, após a inicialização com os dados do ONS, o histórico ainda apresentar falhas, essas são preenchidas com o dado da CCEE (discretização 5 min), quando este for válido.

III. **Dado da CCEE (discretização 1 hora)**
Embora o dado com discretização horária não atenda à discretização semi-horária necessária aos modelos, ele pode ser interpolado gerando séries semi-horárias. Tais séries apresentam uma alternativa com boa qualidade; logo, são utilizadas para suprir as lacunas que as opções anteriores deixaram.

3.3.2 Preenchimento dos históricos de geração

I. **Geração estimada**
Se, após os passos anteriores, o histórico de geração ainda apresentar falhas, uma quarta opção é realizada. Essa é uma opção mais complexa e envolve modelos de estimação dos dados, logo não representa dados medidos efetivamente. De posse do histórico parcial das três primeiras opções, calibra-se, para essa usina, a curva Vento $\times$ Potência (V $\times$ P).

Selecionados os dados, é possível estimar as curvas V $\times$ P. Uma abordagem para essa estimação é o uso de regressão logística, uma vez que esse modelo se ajusta adequadamente à dispersão observada entre velocidade do vento e potência. A equação da regressão logística é dada por:

$$G_e = P_{\min} + \frac{P_{\max} - P_{\min}}{\left(1 + 10^{b\,(V_{mid} - V_v)}\right)^s} \tag{3.4}$$

em que G_e representa a geração *estimada*; $P_{\min}$ e $P_{\max}$ são as assíntotas inferior e superior (potência mínima e máxima); b e V_{mid} são, respectivamente, a inclinação e a coordenada-x do ponto de inflexão; s é um coeficiente; e V_v é o vento verificado.

Para isso, é necessário um conjunto mínimo de pontos (vento e geração). Como condição para a obtenção desses pontos, exige-se que o histórico contenha pelo menos trinta dias com medições válidas de vento e geração para cada intervalo semi-horário. No caso de geração nula, o par vento–geração é sempre considerado inválido, pois, em geral, está associado a condições operacionais atípicas ou a restrições de operação. No Capítulo 4, serão apresentados em detalhes os mecanismos de construção dessas curvas.

Todavia, existem mais de uma fonte para as informações de vento, sendo que uma fonte apenas pode não ter dados de vento suficientes para a usina. Dessa forma, existem algumas etapas para selecionar os dados de vento necessários; a ordem de escolha do histórico mínimo de vento segue a seguinte ordem:

a) **Vento verificado do ONS**
   A primeira fonte utilizada para compor o histórico de vento é o conjunto de dados de vento verificado proveniente do ONS. Ressalta-se que apenas os registros considerados válidos são incorporados à composição desse histórico.

b) **Vento verificado da EPE**
   Caso os dados do ONS não sejam suficientes para constituição do histórico completo, os dados de vento verificado da EPE são utilizados como segunda opção. Ressalta-se que os dados da EPE somente estão disponíveis para usinas que passaram por leilão e, portanto, não cobrem todo o conjunto de usinas.

c) **Vizinhança**
   Se, após a aplicação dos passos (a) e (b), ainda houver lacunas no histórico de vento, são utilizados dados de vento verificado de usinas vizinhas, considerando-se as n usinas mais próximas em um raio de até 50 km. Esse raio foi adotado por ser compatível com a escala em que o regime de vento tende a preservar correlação espacial relevante entre parques de uma mesma região, sob a influência de sistemas meteorológicos comuns. Trata-se de um limite superior de busca: como as usinas candidatas são posteriormente ordenadas e filtradas por correlação, eventuais vizinhas pouco correlacionadas dentro desse raio acabam descartadas nas etapas seguintes. Para cada uma dessas usinas, são aplicados os mesmos procedimentos descritos

nos itens (a) e (b). Em seguida, as usinas vizinhas são ordenadas de acordo com a correlação entre suas séries de vento e a série da usina cujo histórico está sendo preenchido. Posteriormente, os históricos das usinas vizinhas são ajustados por meio de regressão linear, com o objetivo de remover possíveis vieses em relação à geração da usina em análise. Por fim, o procedimento de preenchimento é aplicado sequencialmente a cada usina, de $i = 1$ a n, seguindo essa ordenação e avançando para a próxima usina sempre que a condição necessária não for atendida.

Um detalhamento mais completo das etapas de seleção dos dados de vento encontra-se na Seção 3.4.1 e no primeiro item da Seção 3.4.2.

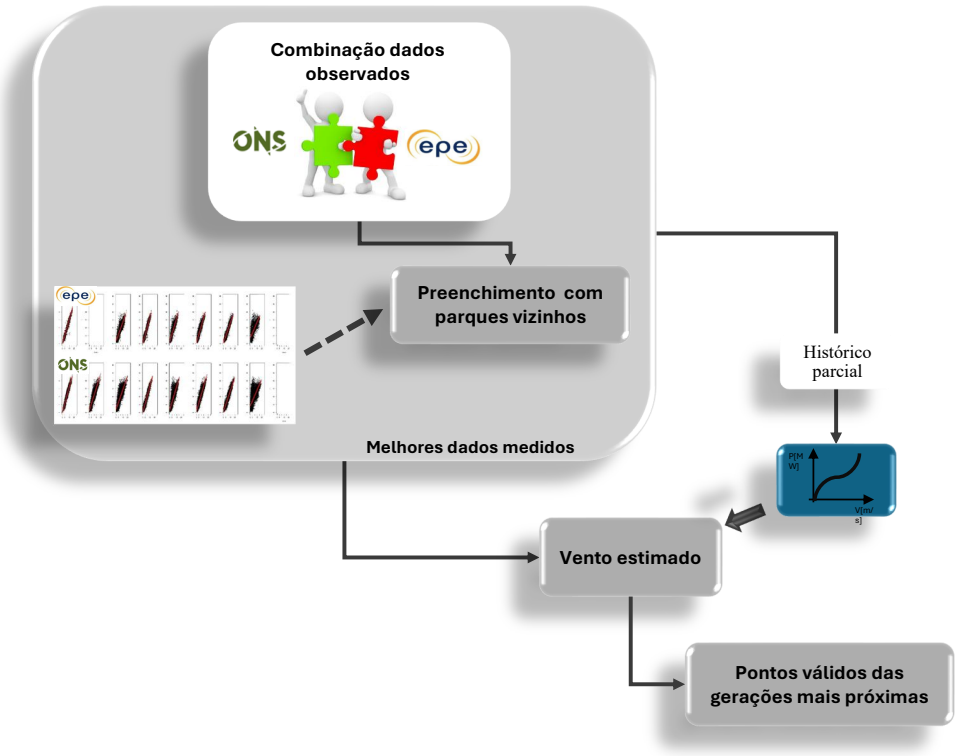
II. **Modelo estatístico de preenchimento**
Caso as etapas anteriores não sejam suficientes para completar todos os dados faltantes do histórico, duas abordagens complementares podem ser empregadas para o preenchimento das lacunas remanescentes. A primeira, mais simples, baseia-se na média aritmética dos valores correspondentes ao mesmo intervalo horário dos três dias anteriores e dos três dias posteriores à falha. A segunda utiliza um método estatístico de previsão voltado especificamente para o preenchimento de dados, cujos detalhes são apresentados na Seção 3.6. A abordagem por média aritmética é computacionalmente mais rápida, porém tende a produzir resultados menos aderentes ao comportamento real da série. O método estatístico, por sua vez, apresenta a vantagem de um preenchimento mais adequado quando as falhas se restringem a algumas horas do dia, uma vez que incorpora informações do próprio padrão diário a ser corrigido. Contudo, para falhas superiores a um dia, essa vantagem se reduz consideravelmente. Ambas as opções constituem um último recurso para tornar a série temporal completa. Espera-se, portanto, que sejam utilizadas com frequência significativamente menor que as demais etapas, de forma a não comprometer a qualidade do histórico reconstituído.

3.4 Composição e preenchimento dos históricos de vento

A montagem do melhor histórico de vento apresenta características similares à montagem do melhor histórico de geração. Para tanto, são utilizadas as informações advindas do ONS e da EPE, fontes com vento verificado disponível. Como ocorre na geração, há necessidade de alguns processos de reconstrução das informações de vento verificado, uma vez que nem sempre é possível obter os dados medidos para todos os intervalos de tempo com essas fontes. A Figura 11 destaca os processos para composição do histórico de vento verificado.

O processo de montagem do histórico de vento ilustrado na Figura 11 consiste em cinco possibilidades para preenchimentos de falhas nos históricos de vento. Destaca-se que os dados provenientes do ONS já passaram pelo processo de identificação de dados espúrios. Sendo assim, cada intervalo de tempo tem seu dado faltante preenchido sequencialmente.

Figura 11 – Processo de montagem do melhor histórico de vento verificado, com preenchimento sequencial das falhas.



Fonte: Elaboração própria.

3.4.1 Composição dos históricos de vento

Após a etapa de limpeza e consistência dos dados históricos, as séries temporais de velocidade do vento de diferentes fontes de medição (ONS e EPE) são integradas utilizando um modelo hierárquico de combinação de fontes. O objetivo desta etapa é maximizar o uso das informações disponíveis, priorizando dados observados locais e garantindo que qualquer suplementação seja estatisticamente consistente, fisicamente coerente e totalmente rastreável. O fluxo metodológico para a combinação de fontes de medição é estruturado da seguinte forma:

(i) Verificação de disponibilidade de dados válidos: avaliação da presença de observações de vento limpas e válidas dos bancos de dados do ONS e da EPE ao longo do período de interesse. Nem todos os parques eólicos possuem ambas as fontes disponíveis simultaneamente.

(ii) Ajuste de modelo de regressão linear entre fontes: quando ambas as fontes fornecem dados válidos, um modelo linear é ajustado entre as séries temporais correspondentes de velocidade do vento do ONS e da EPE, com a consistência estatística avaliada por meio do coeficiente de determinação (R^2, Equação (3.9)).

(iii) Filtro de consistência geométrica: aplicação de um critério adicional de qualidade baseado na distância geométrica dos pontos à linha de regressão ajustada, removendo observações

além de um limiar de distância predefinido para eliminar pares de medições inconsistentes.

(iv) Teste de desempenho do modelo: validação do ajuste linear com base em um critério mínimo de desempenho estatístico, com o modelo considerado elegível para suplementação apenas quando R^2 excede o limiar estabelecido ($R^2 > 0,7$). Adota-se aqui um limiar mais exigente por se tratar de duas fontes de medição de uma mesma usina (ONS e EPE), das quais se espera forte concordância; um ajuste com R^2 abaixo desse valor indicaria inconsistência entre medições que, a princípio, deveriam ser equivalentes.

(v) Suplementação hierárquica de dados: quando o modelo é validado, lacunas na série do ONS são preenchidas usando estimativas derivadas da série ajustada da EPE, preservando integralmente os valores originalmente observados e substituindo apenas as posições com dados faltantes.

Como resultado desse processo, obtêm-se séries temporais de vento integradas e consistentes, nas quais a fonte de informação utilizada em cada passo de tempo é explicitamente identificada. Essa abordagem garante transparência, rastreabilidade e robustez na combinação de fontes de medição, fornecendo uma base sólida para a etapa subsequente de preenchimento de lacunas baseado em correlação espacial.

3.4.2 Preenchimento dos históricos de vento

I. **Preenchimento com parques vizinhos**

Após a integração das fontes de medição, as séries temporais de velocidade do vento ainda podem conter lacunas temporais resultantes de falhas persistentes de medição ou da ausência simultânea de dados válidos nas diferentes bases de dados. Para abordar esses casos, um modelo de preenchimento de lacunas baseado em correlação espacial é empregado. Essa abordagem utiliza informações de parques eólicos geograficamente próximos e estatisticamente correlacionados, garantindo coerência física, consistência estatística e preservação dos dados observados. O fluxo metodológico adotado para o preenchimento de lacunas baseado em correlação espacial é estruturado da seguinte forma:

(i) Identificação de parques eólicos vizinhos: para cada parque eólico de referência, parques eólicos geograficamente próximos dentro de uma distância máxima de 50 km são selecionados, garantindo relevância espacial.

(ii) Ajuste linear das séries vizinhas: as séries temporais de vento de cada parque eólico vizinho são individualmente ajustadas à série do parque eólico de referência usando modelos de regressão linear da forma $y \sim x$, com estimação de coeficientes e consistência estatística avaliada por meio do coeficiente de determinação (R^2).

(iii) Classificação e seleção de parques eólicos: os parques eólicos vizinhos são classificados de acordo com seus valores de R^2, e apenas aqueles que atendem a um critério

mínimo de correlação estatística ($R^2 \geq 0,6$) são considerados elegíveis para preenchimento de lacunas, garantindo que apenas séries representativas contribuam para a suplementação dos dados. O limiar é ligeiramente menos restritivo do que o adotado na combinação de fontes de uma mesma usina, uma vez que se comparam parques distintos, geograficamente separados, cuja correlação é naturalmente atenuada por diferenças locais de regime de vento e relevo; um valor demasiadamente exigente eliminaria vizinhos ainda úteis ao preenchimento.

(iv) Preenchimento hierárquico de lacunas: valores faltantes na série do parque eólico de referência são preenchidos hierarquicamente, priorizando as séries vizinhas mais fortemente correlacionadas, substituindo apenas posições com dados faltantes e preservando integralmente os valores originalmente observados.

Essa abordagem de preenchimento de lacunas baseada em correlação espacial permite a exploração sistemática da redundância espacial entre parques eólicos próximos, resultando em séries temporais de vento finalizadas que combinam robustez estatística, coerência física e rastreabilidade completa dos dados utilizados.

II. **Vento estimado**

Caso o histórico de vento, após os passos anteriores, ainda apresente falhas, uma terceira opção é realizada. Para isso, a regressão logística obtida na opção de *Geração estimada* para montagem dos históricos de geração, Equação (3.4), é utilizada. Após a calibração dos parâmetros, essa equação é *invertida* e, assim, conhecida a geração para um determinado instante de tempo, é possível estimar o vento correspondente. A forma inversa pode ser escrita como:

$$V_e \;=\; V_{mid} \;-\; \frac{\ln\!\left(\left(\frac{G_v - P_{\min}}{P_{\max} - P_{\min}}\right)^{-1/s} - 1\right)}{b \ln 10} \tag{3.5}$$

em que V_e é o vento *estimado* e G_v a geração verificada.

III. **Pontos válidos próximos**

De forma análoga ao tratamento aplicado aos dados de geração, os dados de vento também dispõem de uma última opção de preenchimento, composta por duas abordagens distintas. A primeira, diferentemente da estimativa por média aritmética empregada na composição dos históricos de geração, calcula os dados faltantes a partir de *pontos próximos válidos* da dispersão V $\times$ P. Essa escolha se justifica pelo fato de que o uso de uma média baseada exclusivamente nos dados de vento, sem considerar os dados de geração correspondentes, poderia introduzir um espalhamento irreal na dispersão V $\times$ P, comprometendo os modelos subsequentes de estimação e previsão da geração. A segunda abordagem utiliza o mesmo método estatístico de previsão descrito anteriormente, voltado especificamente para o preenchimento de dados, cujos detalhes são apresentados na Seção 3.6. Ambas as opções constituem um último recurso para tornar a série temporal completa. Espera-se, portanto,

que sejam utilizadas com frequência significativamente menor que as demais etapas, de forma a não comprometer a qualidade do histórico reconstituído.

Há algumas usinas, no entanto, para as quais as opções anteriores não são suficientes para o preenchimento dos dados de vento. Nesses casos, não existem dados alternativos nem usinas próximas que possam ser utilizadas como referência, e as abordagens de preenchimento descritas anteriormente não são aplicáveis, uma vez que dependem de algum histórico de vento válido da própria usina, o qual, nessas situações, é inexistente. Excepcionalmente, para esses casos, as curvas V $\times$ P são construídas a partir do histórico de geração da usina combinado com os dados de previsão de vento do modelo meteorológico de melhor desempenho.

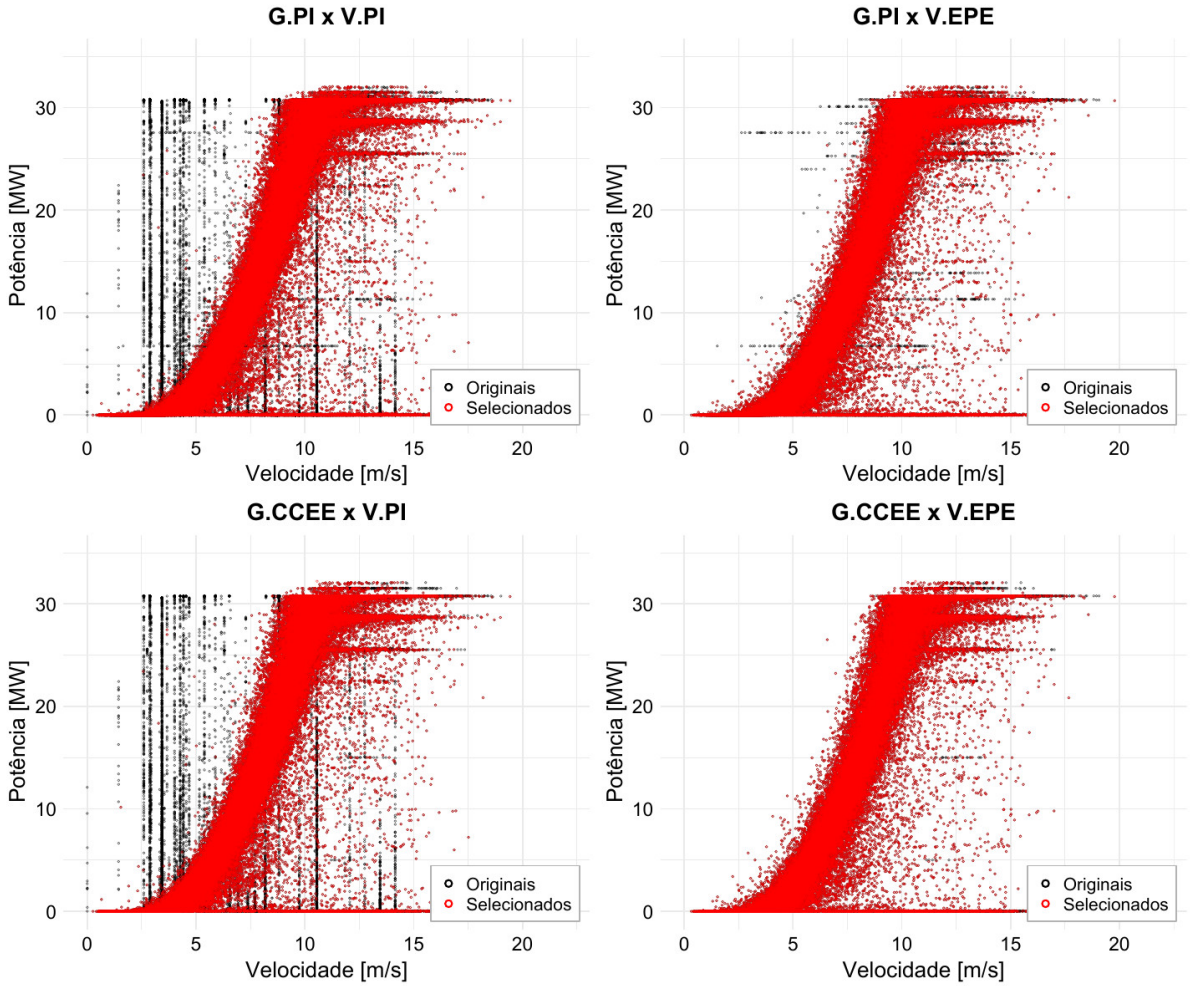
3.5 Aplicação dos processos de tratamento

Esta seção destina-se a apresentar alguns exemplos do processo de limpeza e montagem dos históricos de dados observados. A Figura 12 apresenta um exemplo da dispersão V $\times$ P construída com os dados verificados disponíveis para uma usina eólica, fornecendo uma visão geral dos dados aproveitados e descartados durante a etapa de limpeza. Na figura, os quatro gráficos representam a dispersão de V $\times$ P da usina, combinando as duas fontes de geração (PI/ONS e CCEE) e as duas fontes de vento (PI/ONS e EPE). Em cada painel, os pontos em preto representam os dados originais e os pontos em vermelho representam os dados que foram considerados válidos após a aplicação dos filtros de limpeza.

Os gráficos da Figura 12 demonstram o impacto do processo de limpeza dos dados, sendo possível observar que grande parte dos problemas do exemplo está relacionada ao vento verificado. No gráfico superior à esquerda (G.PI $\times$ V.PI), nota-se a presença de valores de geração que variam enquanto o vento se mantém constante, formando linhas verticais características de falhas no sensor de vento. Ao confrontar este gráfico com o gráfico superior à direita (G.PI $\times$ V.EPE), verifica-se que, quando se utiliza o vento da EPE, o volume de dados selecionados como válidos é consideravelmente maior, indicando que os dados de vento do PI apresentam mais inconsistências. Os dados de geração também apresentam problemas, em frequência bem menor, conforme se observa no mesmo gráfico superior à direita: pontos com geração congelada enquanto o vento varia, formando linhas horizontais características de falhas no sensor de geração.

Para oferecer uma compreensão mais abrangente dos tipos de problemas, as Figuras 13 e 14 apresentam as curvas diárias dos dados verificados ao longo de cinco anos, comparando as séries originais (gráficos à esquerda) e as séries após a etapa de limpeza (gráficos à direita). Nos gráficos da esquerda, as curvas em cor mais clara representam o conjunto completo de dias, enquanto as curvas destacadas em cor mais escura correspondem aos dias que apresentaram algum tipo de problema.

Figura 12 – Dispersão V × P para identificação de dados espúrios de uma usina, combinando as fontes de geração (ONS e CCEE) e de vento (ONS e EPE).



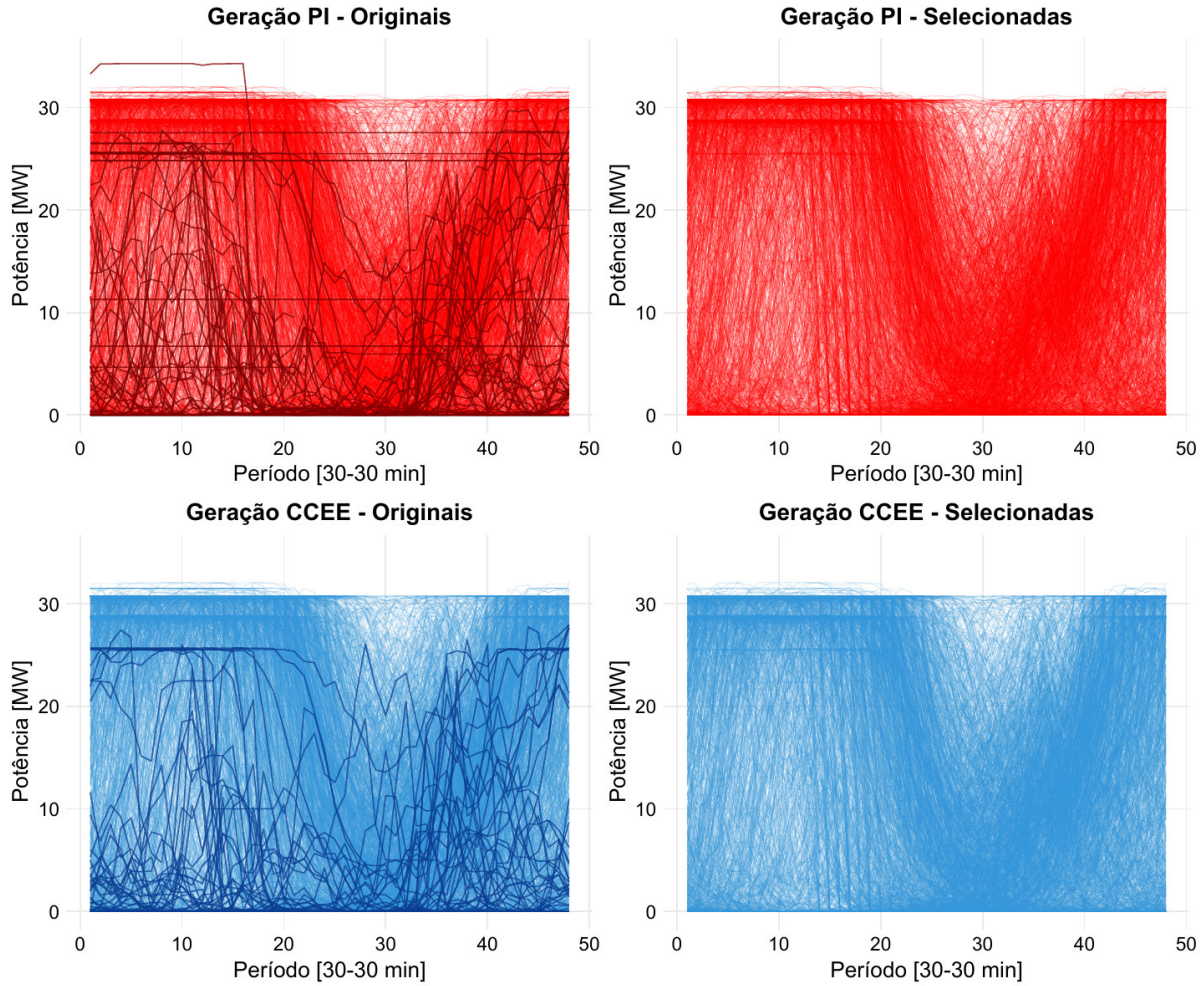
Fonte: Elaboração própria.

A Figura 13 apresenta as curvas diárias de geração para as duas fontes (PI e CCEE). Observa-se que ambas as fontes de geração apresentam alguns dias problemáticos, destacados em vermelho escuro (PI) e azul escuro (CCEE), porém em volume relativamente baixo, a maior parte das curvas se concentra no comportamento típico esperado.

Já a Figura 14, referente ao vento verificado, evidencia que o volume de problemas nos dados é significativamente maior, sobretudo para a fonte PI. Os dias destacados em vermelho escuro (V.PI) representam uma parcela considerável do total, com curvas distorcidas em relação ao padrão típico do vento diário da usina. A fonte EPE também apresenta dias problemáticos, em menor quantidade, identificáveis pelas curvas em verde escuro. Esse resultado é coerente com o observado na Figura 12, em que os painéis com vento do PI eliminam mais pontos da dispersão V × P do que os painéis com vento da EPE.

Após o processo de limpeza, as lacunas nos dados de vento precisam ser preenchidas, o que significa que os dados de vento devem ser reconstituídos ou selecionados de outra fonte,

Figura 13 – Identificação e seleção dos dados verificados de geração, antes e após a limpeza, ao longo de cinco anos; dias com problema em cor mais escura.

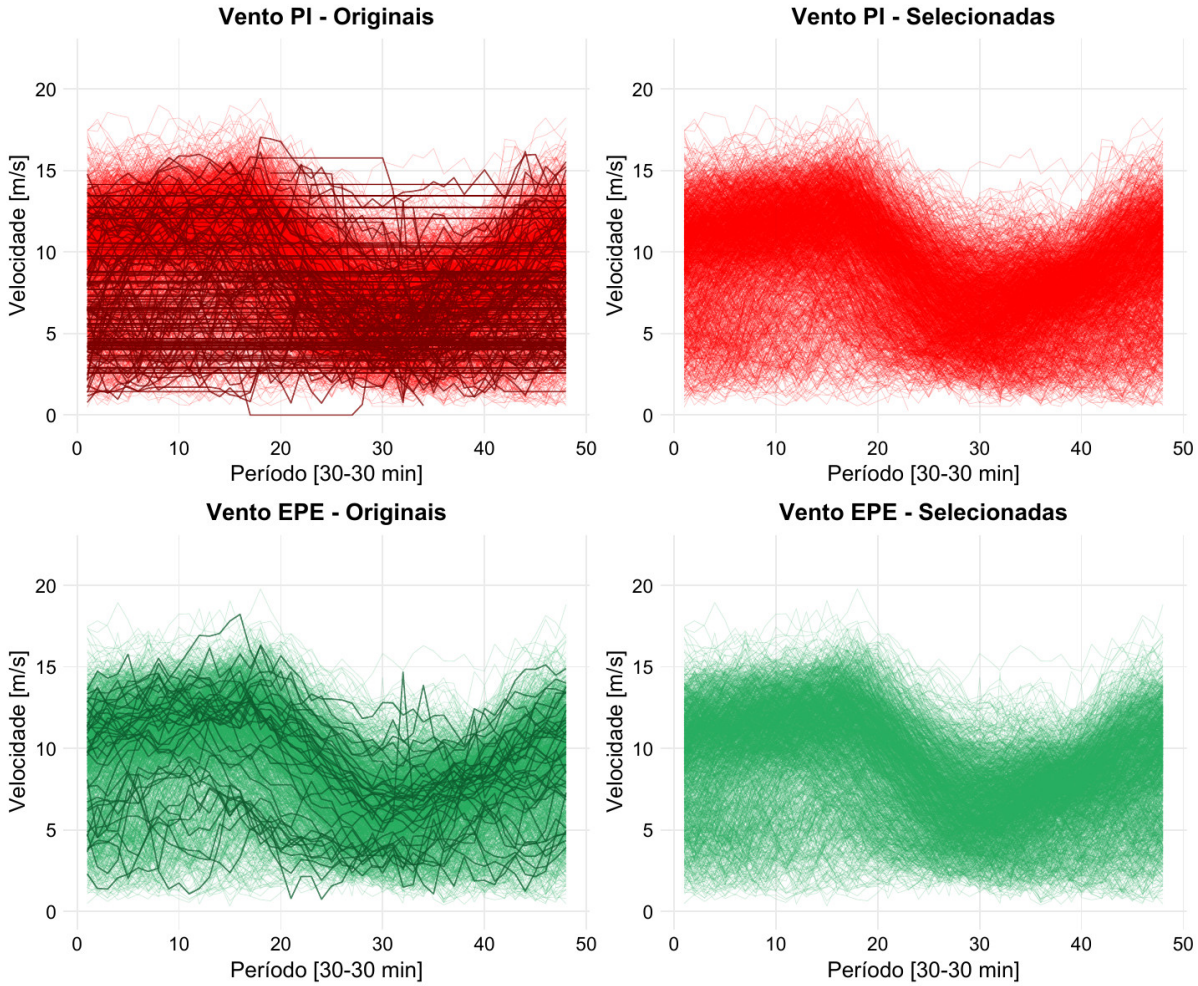


Fonte: Elaboração própria.

conforme detalhado na seção de montagem dos históricos de vento verificado. A Figura 15 apresenta o resultado final do processo, exibindo o histórico consistido, já preenchido e tratado, tanto para a geração (gráfico à esquerda) quanto para o vento (gráfico à direita) da usina. Comparando os gráficos da Figura 15 com os painéis correspondentes das Figuras 13 e 14, percebe-se que o histórico consistido elimina os dias problemáticos previamente destacados e apresenta um conjunto coerente de curvas diárias, com comportamento típico em relação ao perfil esperado da usina.

A Figura 16 apresenta as curvas de geração e vento para dois dias, ilustrando o comportamento de dois tipos de problemas nos dados de vento verificado: valores congelados (Dia 1) e dados espúrios (Dia 2). Nos gráficos superiores, as curvas representam, em vermelho, o vento do PI; em rosa, o vento do PI após o processo de eliminação de dados espúrios, nos dois dias essa curva é inexistente, uma vez que todo o histórico de vento do PI foi eliminado; em verde, o vento da EPE; e, em preto, o melhor vento selecionado (MHV).

Figura 14 – Identificação e seleção dos dados verificados de vento, antes e após a limpeza, ao longo de cinco anos; dias com problema em cor mais escura.



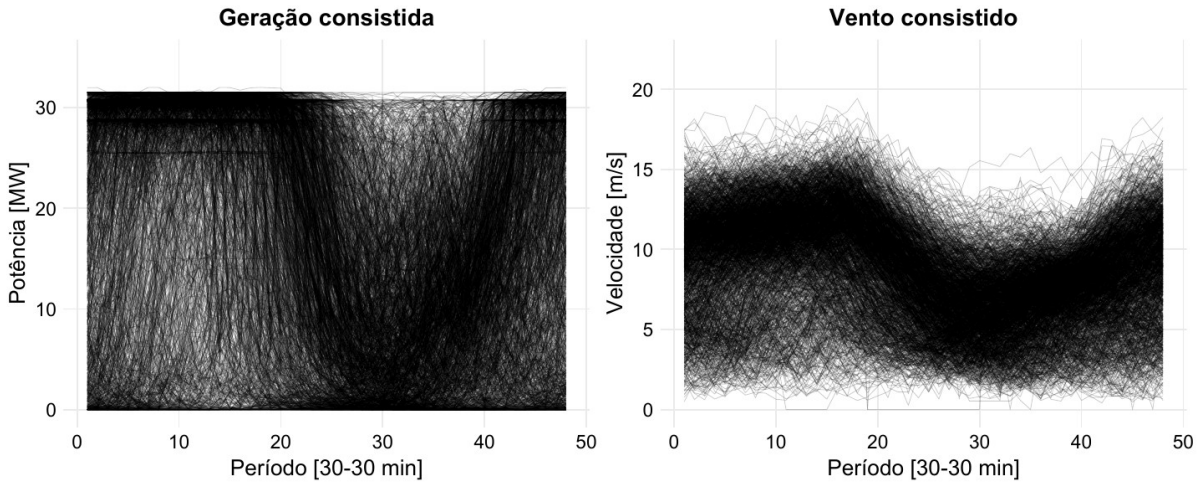
Fonte: Elaboração própria.

Os gráficos inferiores da Figura 16 apresentam as curvas de geração e são incluídos na figura apenas para dar uma ideia do comportamento do vento nesses dias, sendo: em vermelho, a geração do PI; em rosa, a geração do PI após o processo de eliminação de dados espúrios, nos dois dias essa curva está exatamente alinhada com a curva em vermelho, dado que os dados de geração se mostraram coerentes; em verde, a geração da CCEE de cinco minutos; em azul, a geração da CCEE no patamar de uma hora; e, em preto, o melhor histórico de geração (geração consistida).

3.6 Modelo estatístico para identificação de dados espúrios e preenchimento de lacunas

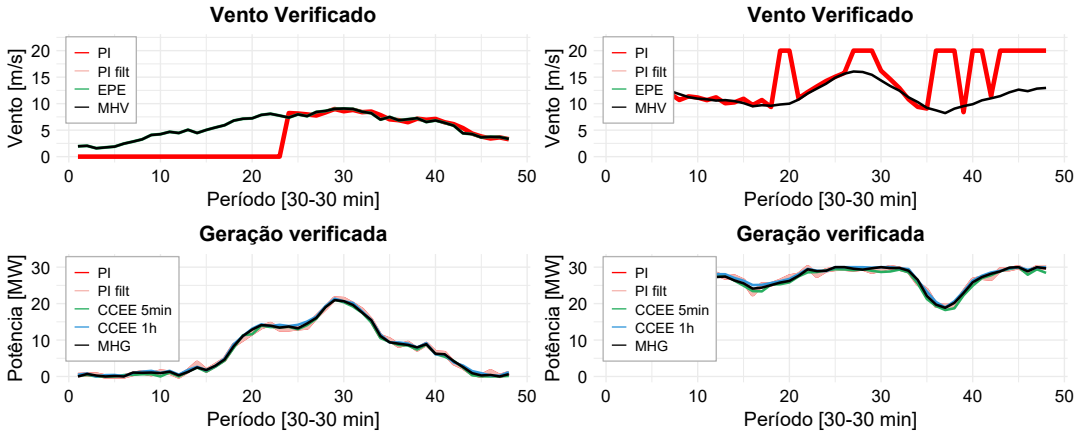
As etapas de limpeza dos históricos observados nem sempre são capazes de capturar todas as falhas existentes nos dados. Além disso, as etapas de montagem dos históricos podem ser insuficientes para gerar séries com a qualidade adequada aos modelos. Algumas delas dependem de informações de usinas próximas, possibilidade nem sempre factível, dado que diversas usinas

Figura 15 – Histórico consistido de geração e vento de uma usina, após limpeza, composição e preenchimento.



Fonte: Elaboração própria.

Figura 16 – Identificação de dados espúrios e resultado do melhor histórico, em dois dias com problemas nos dados de vento (valores congelados e dados espúrios).



Fonte: Elaboração própria.

podem estar localizadas a grandes distâncias umas das outras. Outras se baseiam em técnicas simplificadas, como médias aritméticas, no caso da montagem da geração, ou na identificação de pontos válidos próximos, no caso da montagem de vento.

Diante dessas limitações, foi desenvolvido um módulo alternativo de caráter mais genérico, aplicável a situações em que os tratamentos anteriormente descritos não se mostram adequados, conferindo maior robustez ao processo de preenchimento. Esse módulo adota uma abordagem de cunho estatístico e de aprendizado de máquina *não supervisionado*, com formulação em *espaço de estados* e perspectiva *bayesiana*. Sua implementação consiste na aplicação direta do filtro de Kalman [151], cujas previsões são derivadas de distribuições de probabilidade obtidas a partir de padrões (*clusters*) específicos dos dados, identificados por meio de uma metodologia de *clustering*. Os fundamentos teóricos utilizados baseiam-se principalmente no

trabalho de [54]. A seguir, apresenta-se o detalhamento da metodologia desenvolvida.

3.6.1 Identificação dos dados espúrios

Os dados que apresentam maior número de problemas são os de vento verificado; assim, algumas das etapas serão apresentadas para essa grandeza, sendo idênticas para a geração verificada.

3.6.1.1 Classificação das curvas dos dados e identificação dos padrões característicos

O processo de identificação de dados espúrios e de preenchimento de valores faltantes baseia-se na extração de padrões característicos presentes nos dados. Assim, o objetivo consiste em identificar padrões, realizar classificações e estimar variâncias. Segundo a literatura especializada, diversas técnicas podem ser empregadas nessa tarefa, como modelos hierárquicos, métodos nebulosos e redes neurais de Kohonen. Neste trabalho, optou-se pela utilização do método *K-means* [152], uma técnica de agrupamento simples que atende adequadamente aos requisitos da aplicação. Trata-se de um método de aprendizado não supervisionado com elevada capacidade de processamento e boa adaptação a problemas envolvendo dados de alta dimensionalidade.

O parâmetro k representa o número de *clusters* (ou grupos) em que os dados serão particionados e constitui um hiperparâmetro do processo. O algoritmo compreende quatro etapas principais [55]: (i) determinar as posições iniciais dos centroides dos k *clusters*; (ii) alocar cada elemento ao *cluster* cujo centroide esteja mais próximo; (iii) recalcular os centroides a partir da média dos elementos alocados; e (iv) repetir as etapas (ii) e (iii) até que o critério de convergência seja atingido. A convergência ocorre quando a alocação dos elementos aos *clusters* deixa de apresentar alterações significativas entre iterações consecutivas, sendo esse limite de significância previamente definido como um erro suficientemente pequeno e aceitável [55].

3.6.1.2 Eliminação de padrões muito atípicos

Uma vez que os padrões possuem muito melhor qualidade, em virtude da possibilidade de uso dos dados pré-tratados, conforme descrito nas seções anteriores de composição e preenchimento dos históricos, esta etapa é necessária apenas para eliminar os padrões extremamente atípicos. A eliminação ocorre quando o número de séries alocadas NSA num *cluster* é menor que 10% do número médio de séries dado pela relação do total de séries TS pelo número de padrões criados NP. Conforme:

$$\frac{NSA}{TS/NP} \times 100\% \leq 10\% \tag{3.6}$$

Assim, um ano de dados agrupado em 10 *clusters* exige que cada *cluster* tenha mais do que 3,6 séries alocadas, ou seja, ao menos quatro séries; caso contrário, o padrão é considerado muito atípico e eliminado. Cabe destacar que o critério é definido de forma relativa ao tamanho médio dos agrupamentos, e não por um valor absoluto, o que o torna robusto ao número de padrões

adotado. Seu único objetivo é descartar agrupamentos degenerados, formados por pouquíssimas séries, sem afetar os padrões representativos.

3.6.1.3 Associar curva a ser filtrada ao *cluster* correspondente

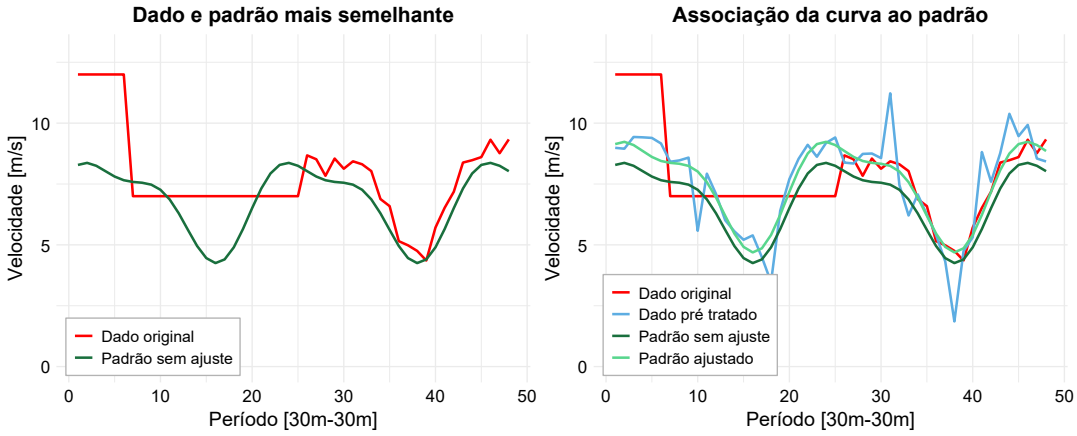
O próximo passo é associar o dado que será filtrado ao *cluster* correspondente, isso é realizado através do cálculo da distância Euclidiana entre o valor dos dados da série e o padrão ajustado de um nível. O padrão que apresentar menor distância é então escolhido, de acordo com a equação:

$$\text{dist}_k = \sqrt{\sum_{j=1}^{48}\left(Y_j - Y_{PO_k} \cdot PO_{jk}\right)}, \qquad \min\left\{\text{dist}_1, \text{dist}_2, \ldots, \text{dist}_{NP}\right\} \tag{3.7}$$

em que Y é o vetor de dados de um dia a ser filtrado; Y_j é o dado a ser filtrado no intervalo j; PO_k é o padrão de um dia do *cluster* k e PO_{jk} é o padrão do *cluster* k no intervalo j.

A Figura 17 (parte da esquerda) ilustra uma curva de vento diária discretizada em quarenta e oito intervalos (semi-horas do dia), sendo que há problemas de valores congelados na quase totalidade do dia. Na parte direita da mesma Figura 17, são adicionadas duas novas curvas: uma correspondente ao padrão ajustado e outra aos dados após a aplicação da consistência, conforme os filtros descritos anteriormente. Ressalta-se que a curva pré-tratada pode divergir da curva original mesmo em períodos considerados consistentes, uma vez que, para ilustrar a capacidade geral da metodologia proposta, foi inserido ruído na série de forma a simular a presença de dados espúrios.

Figura 17 – Associação do dado a ser filtrado ao padrão similar, para uma curva diária de vento com valores congelados.



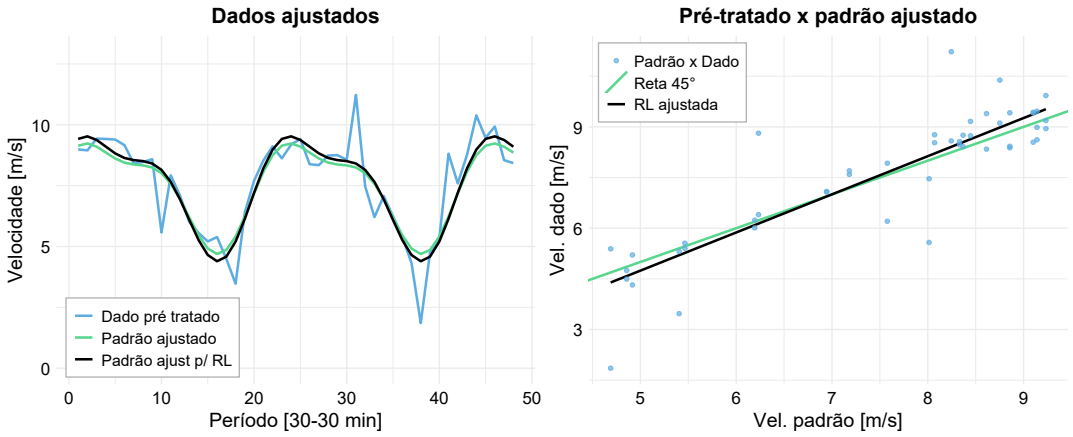
Fonte: Elaboração própria.

Observa-se que o dado consistido, mesmo alterado, apresenta boa coerência em relação ao padrão ajustado. Conclui-se, portanto, que a consistência de dados é fundamental para a qualidade das análises subsequentes.

3.6.1.4 Identificação dos dados espúrios e transformação em dados faltantes

A metodologia proposta para identificação dos dados espúrios é baseada no gráfico de dispersão entre a curva a ser corrigida e o padrão ajustado associado. A primeira etapa consiste em ajustar uma regressão linear entre o dado a ser filtrado e o padrão ajustado. Isto é realizado considerando que se o dado a ser filtrado for exatamente igual ao padrão ajustado, realizando um gráfico de dispersão entre eles tem-se uma relação de correlação unitária, ou seja, o ajuste de uma regressão linear seria uma reta de quarenta e cinco graus. A Figura 18 à esquerda ilustra a curva do dado a ser filtrado, do padrão ajustado e a curva obtida pela regressão. A figura à direita mostra o gráfico de dispersão entre os dados a serem filtrados (eixo y) e o padrão (eixo x), juntamente com a linha ajustada na dispersão.

Figura 18 – Regressão linear através do dado a ser filtrado e o padrão ajustado, utilizada na identificação de dados espúrios.



Fonte: Elaboração própria.

Na segunda etapa, calcula-se um limite inferior e superior, que será utilizado para identificar os dados espúrios, de forma a transformá-los em dados faltantes para posterior preenchimento desses. O processo é realizado primeiramente calculando o erro associado a cada intervalo de tempo entre o dado a ser filtrado e o padrão ajustado. O limite inferior e o superior, respectivamente, são calculados pelo desvio-padrão do erro multiplicado pela quantidade de desvios-padrão definida com base na equação:

$$\text{Lim} \;=\; \pm\sigma\,(Y-\hat{Y})\times N_{\sigma} \tag{3.8}$$

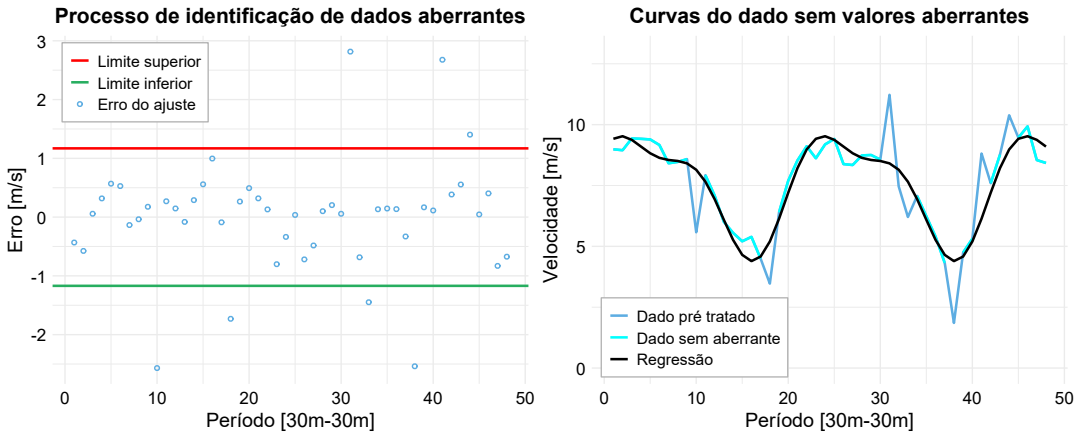
em que N_{σ} é estimado com base no coeficiente de determinação, Equação (3.9), da regressão linear mostrada na Figura 18. Quanto melhor a qualidade do ajuste menor o valor de N_{σ}, ou seja, mais restritivos serão os limites.

$$R^2 \;=\; 1 \;-\; \frac{\sum_{j=1}^{n}(Y_j-\hat{Y}_j)^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}, \qquad n=48. \tag{3.9}$$

Em que Y_j é o dado que está sendo filtrado para o intervalo j; $\hat{Y}_j$ é o dado estimado pela regressão; Y_i é o dado filtrado no intervalo i e $\bar{Y}$ é a média desse dado; n representa a discretização temporal dos dados. Neste trabalho, os dados possuem discretização semi-horária, portanto, $n = 48$.

Para o mesmo exemplo ilustrado nas Figura 17 e 18, é traçado o gráfico com os erros de ajuste e os limites inferior e superior, parte esquerda da Figura 19. Assim sendo, todo ponto que está situado fora do intervalo é considerado como dado espúrio, sendo retirado da série que será filtrada. A metade direita da Figura 19 mostra o dado pré-tratado e como ele aparece após a remoção dos pontos considerados dados espúrios.

Figura 19 – Identificação de dados espúrios pelos erros de ajuste e seus limites, e resultado após a remoção dos pontos espúrios.



Fonte: Elaboração própria.

3.6.2 Modelo de correção de dados faltantes

A correção dos dados que foram considerados como dados espúrios é realizada substituindo esses dados, que já foram transformados em dados faltantes, por valores mais plausíveis. Esses valores previstos são obtidos pelo modelo proposto baseado no filtro de Kalman. Existem algumas variações do filtro de Kalman, sendo que neste trabalho é utilizada uma formulação em espaço de estados de nível e tendência local. Neste modelo, o valor observado da série é presumido como a soma de duas componentes, que são o nível da série e uma componente aleatória com distribuição Gaussiana (Normal). A forma geral e específica desse modelo é apresentada a seguir:

$$y_t = \mu_t + v_t, \qquad v_t \sim \mathcal{N}(0, V_t) \tag{3.10}$$

$$\mu_t = \mu_{t-1} + \beta_t + \omega_t^{(1)}, \qquad \omega_t^{(1)} \sim \mathcal{N}(0, W_t^{(1)}) \tag{3.11}$$

$$\beta_t = \beta_{t-1} + \omega_t^{(2)}, \qquad \omega_t^{(2)} \sim \mathcal{N}(0, W_t^{(2)}) \tag{3.12}$$

Reescrevendo na forma matricial de espaço de estados:

$$Y_t = F'\theta_t + v_t, \qquad v_t \sim \mathcal{N}(0, V_t) \tag{3.13}$$

$$\theta_t = G\theta_{t-1} + \omega_t, \qquad \omega_t \sim \mathcal{N}(0, W_t) \tag{3.14}$$

em que Y_t é o vetor de observações; θ_t é o vetor de parâmetros; F_t e G_t são matrizes conhecidas; V_t e W_t são matrizes de covariâncias conhecidas. O modelo geralmente é definido pela quádrupla F, G, V, W, e o vetor de parâmetros θ, que são definidos do seguinte modo:[1]

$$F_t = (1\ 0), \qquad G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \theta_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} \tag{3.15}$$

Neste estudo, será empregada uma abordagem *Bayesiana* para ajustar os modelos lineares dinâmicos. Além disso, será utilizado um filtro de Kalman para atualizar os parâmetros θ. O processo se inicia a partir da distribuição a priori dos parâmetros no instante t. Em seguida, de posse da distribuição a priori, pode-se efetuar a previsão do valor futuro da série Y_t. Por fim, ao receber o novo valor dado da série, também chamado de inovação, efetua-se a atualização da distribuição a posteriori dos parâmetros do modelo para o instante t, a partir do qual se repete o processo para os instantes seguintes.

3.6.2.1 Priori em t:

$$\theta_t \mid D_{t-1} \sim \mathcal{N}(a_t, R_t), \qquad a_t = G_t m_{t-1}, \qquad R_t = G_t C_{t-1} G_t' + W_t \tag{3.16}$$

3.6.2.2 Previsão 1 passo à frente:

$$Y_t \mid D_{t-1} \sim \mathcal{N}(f_t, Q_t), \qquad f_t = F_t' a_t, \qquad Q_t = F_t' R_t F_t + V_t \tag{3.17}$$

3.6.2.3 Posteriori em t:

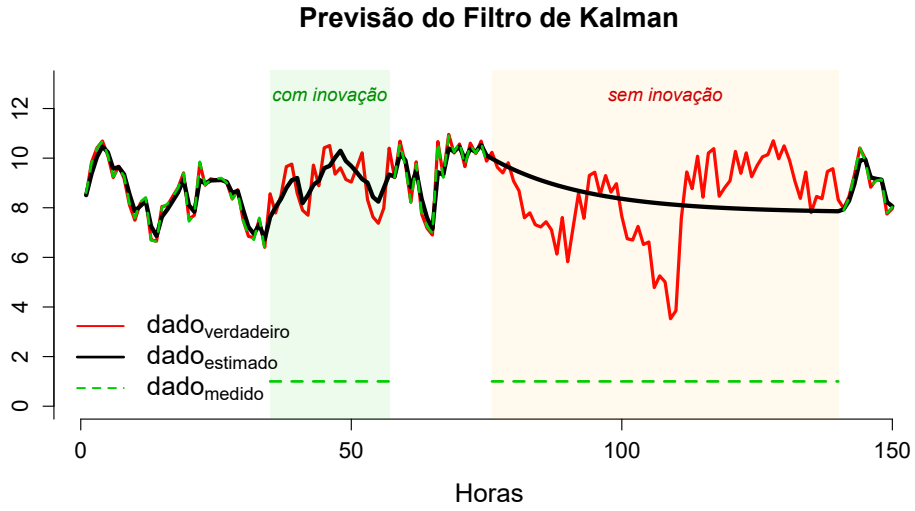
$$A_t = R_t F_t Q_t^{-1}, \qquad e_t = Y_t - f_t, \qquad m_t = a_t + A_t e_t, \qquad C_t = R_t - A_t A_t' Q_t \tag{3.18}$$

Resumidamente, o modelo aplica um mecanismo de identificação dos dados espúrios e a sua substituição para dados faltantes. Esses dados são preenchidos por estimativas do filtro de Kalman modificado.

Quando ocorrem falhas prolongadas, caracterizadas pela ausência contínua de inovações, o filtro tende a convergir para uma trajetória aproximadamente linear, conforme discutido em [56] e ilustrado na região sombreada em amarelo da Figura 20. Nesse contexto, ao realizar a atualização a posteriori no instante t, isto é, quando a inovação é necessária, propõe-se substituir o dado observado nesse instante pelo valor correspondente ao padrão de referência, conforme exemplificado na região sombreada em verde da Figura 20.

[1] Nesta seção, os símbolos F, G, V e W seguem a notação clássica de modelos em espaço de estados [54], designando as matrizes do filtro de Kalman. Esse uso é local e não se confunde com a notação adotada no restante do trabalho, em que G e V denotam, respectivamente, geração e velocidade de vento.

Figura 20 – Previsão do Filtro de Kalman com regiões de falha (com e sem inovação).



Fonte: Elaboração própria.

Diferentemente de [56], que utiliza informações de curvas de dias similares anteriores, a proposta deste trabalho de utilizar informações no mesmo instante t do padrão de referência se torna mais robusta e mais adequada ao tipo de dado (geração ou vento), já que o padrão teoricamente apresenta características similares ao dado que está sendo tratado, pois pertencem a um mesmo *cluster* de dados. Desta forma, a inovação pode ser obtida tomando-se a probabilidade do valor observado no dado no instante t, em uma distribuição gaussiana de média e variância do *cluster* escolhido para esse mesmo instante. A inovação no instante t seria o valor que acarreta esse montante de probabilidade na distribuição gaussiana com média e variância do *cluster* para o instante t. Assim, em vez de uma reta projetada pelo filtro de Kalman, o modelo seguirá um trajeto similar ao do padrão de referência ficando, portanto, mais próximo do dado real.

É bom reforçar que o principal objetivo da correção dos dados eólicos é fornecer insumos com qualidade para os modelos de previsão. Sendo que alguns dados estimados ou até mesmo reais, podem possuir rampas indesejadas. Neste contexto, os dados corrigidos passam por um filtro de suavização, um filtro *loess* (*locally estimated scatterplot smoothing*). Esse método, proposto em [153], baseia-se no ajuste sucessivo de n modelos de regressão, baseado em dados vizinhos, ajustados por mínimos quadrados ponderados.

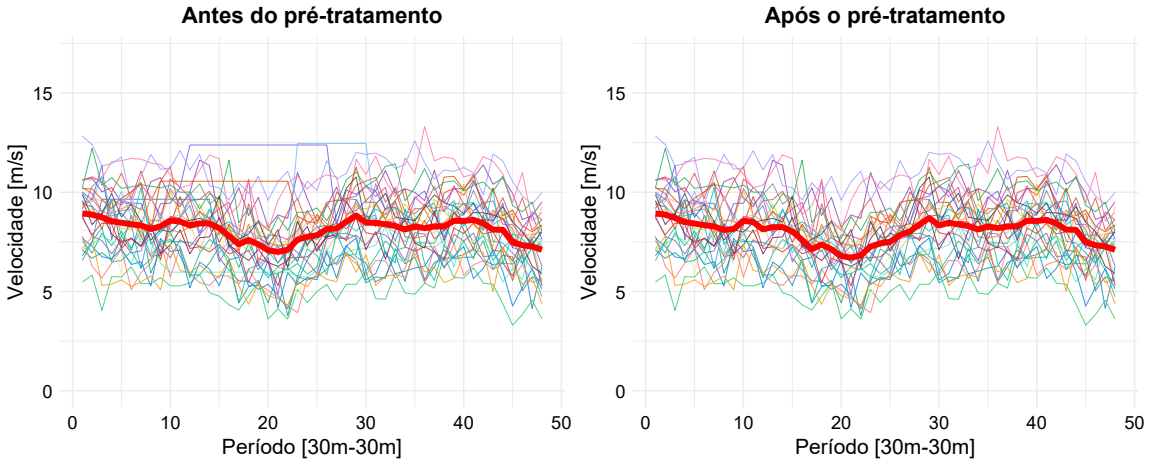
3.6.3 Aplicação do processo de classificação

Para ilustrar a capacidade da metodologia de identificação de dados espúrios, foram utilizados os dados originais do ONS, preservando-se todas as inconsistências presentes na série.

As Figuras 21 e 22 apresentam algumas séries diárias de vento pertencentes a um determinado *cluster*, ilustrando a classificação antes e após a etapa de pré-tratamento (identificação dos dados espúrios), na qual os valores identificados como espúrios são marcados como dados

faltantes. Na Figura 21, em que as séries em sua maioria não apresentam problemas, os centroides do *cluster* não diferem significativamente entre o painel da esquerda (antes do pré-tratamento) e o painel da direita (após o pré-tratamento), uma vez que poucos pontos foram removidos.

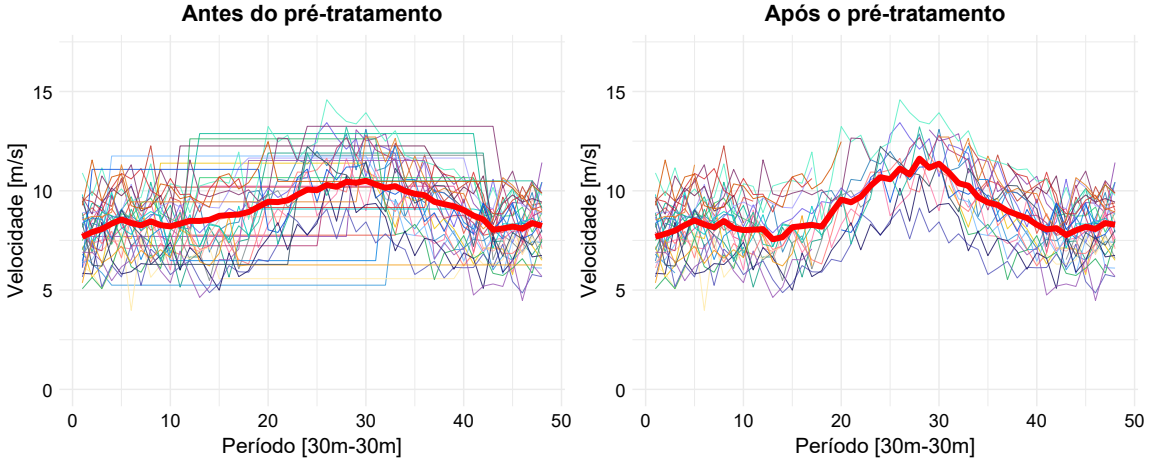
Figura 21 – Classificação de um *cluster* com poucas falhas: antes e após o pré-tratamento, em séries diárias de vento.



Fonte: Elaboração própria.

Já a Figura 22 ilustra um *cluster* em que há um grande número de séries com valores comprometidos, sendo que muitas dessas séries permanecem com valores congelados por longos períodos do dia. Como consequência, o centroide desse *cluster*, representado pela curva vermelha em destaque, fica visivelmente distorcido no painel da esquerda, deixando de representar adequadamente o comportamento típico das séries do agrupamento. No painel da direita, após a remoção dos pontos espúrios, o centroide passa a refletir de forma mais fidedigna o perfil característico do *cluster*.

Figura 22 – Classificação de um *cluster* com muitas falhas: antes e após o pré-tratamento, em séries diárias de vento.



Fonte: Elaboração própria.

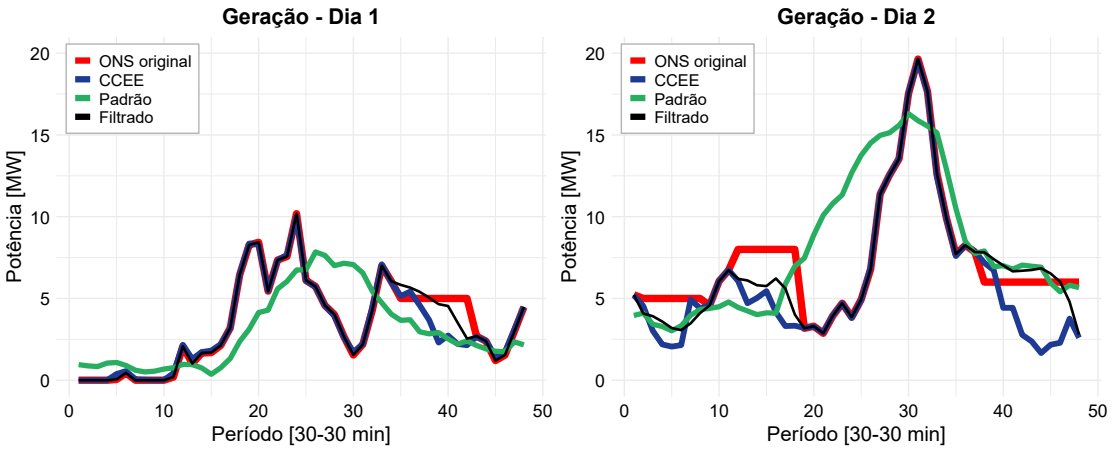
3.6.4 Processo de preenchimento de dados faltantes

As Figuras 23 e 24 ilustram o resultado do modelo de preenchimento de dados faltantes para uma usina real, considerando dois dias distintos da operação. Os dados observados pelo ONS são apresentados em vermelho, enquanto os dados das bases de comparação, CCEE para geração e EPE para vento, são apresentados em azul. A curva em verde representa o padrão dos dados que mais se identifica com o dado a ser corrigido, fornecendo informações para o filtro de Kalman. Após a correção pelo modelo proposto, tem-se a curva em preto. Cabe destacar que ONS, CCEE e EPE são fontes independentes de medição, podendo apresentar pequenas divergências entre si mesmo em períodos sem falhas, em decorrência de diferenças de calibração e de processos de medição.

Uma ressalva importante deve ser feita quanto ao papel dos dados da CCEE e da EPE nas figuras a seguir. Essas bases não são utilizadas em nenhuma etapa do processo de filtragem: a correção dos dados do ONS é realizada exclusivamente a partir do histórico do próprio ONS e do padrão associado (curva verde). As curvas da CCEE (para geração) e da EPE (para vento) são apresentadas apenas como referências externas, servindo como uma aproximação daquilo que seriam os dados reais caso não houvesse falhas de medição, permitindo avaliar o desempenho do filtro por comparação.

A Figura 23 apresenta a aplicação do filtro nos dados de geração do ONS. Como pode ser observado, os dados apresentam falhas de medição consideráveis, com janelas em que a potência permanece travada em um valor constante. No painel à esquerda (Dia 1), em que há apenas uma janela de falhas, a aplicação do filtro é bastante robusta: a curva estimada (preto) acompanha de perto a curva de referência da CCEE (azul) ao longo de praticamente todo o dia, incluindo o interior da janela travada, evidenciando a eficácia da correção quando o volume de informação disponível é suficiente.

Figura 23 – Correção dos dados de geração verificada para dois dias distintos.



Fonte: Elaboração própria.

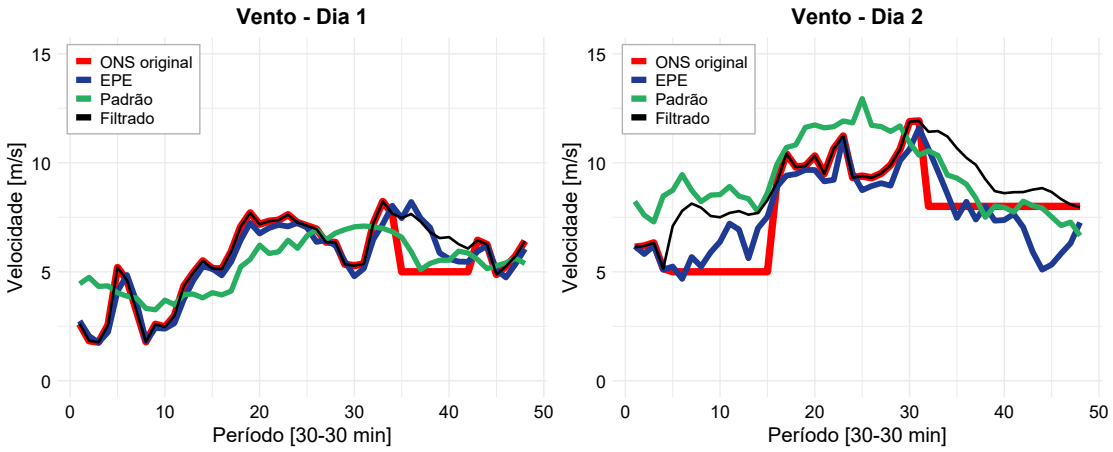
Já no painel à direita (Dia 2), o comportamento é mais desafiador. A ocorrência de

múltiplas janelas de falhas ao longo do dia reduz consideravelmente o volume de dados úteis para o filtro, o que afeta diretamente a qualidade da estimação. Observa-se que a curva estimada (preto) ainda fornece informações relevantes nos períodos intermediários, mas apresenta descolamento visível em relação à curva de referência da CCEE (azul), sobretudo no final do dia, onde a extensão prolongada da falha faz com que a estimativa passe a ser dominada pelo padrão histórico (curva verde), afastando-se do comportamento observado nos dados da CCEE. Esse tipo de degradação é esperado e evidencia o limite prático de qualquer método de preenchimento: quanto maior a proporção de dados faltantes em um dia, mais a estimativa depende de informações externas ao próprio dia e, portanto, maior a incerteza associada.

A Figura 24 ilustra situação análoga para a correção do vento observado, considerando os mesmos dois dias da Figura 23. Os dados de comparação, oriundos da base da EPE, são também representados pela curva em azul, com o mesmo papel de referência externa descrito anteriormente. Vale observar que as falhas de medição do vento não necessariamente coincidem temporalmente com as falhas de geração no mesmo dia, uma vez que os sensores de cada variável operam de forma independente.

No Dia 1 (painel à esquerda), com poucas falhas de medição, o filtro recupera bem a curva de velocidade do vento: a estimativa (preto) acompanha a referência da EPE (azul) ao longo de todo o horizonte, mesmo no interior da janela travada. Já no Dia 2 (painel à direita), com falhas mais extensas concentradas no início e no final do dia, observa-se que a estimativa fica sujeita a uma maior influência do padrão histórico (curva verde), causando descolamentos visíveis entre a curva estimada (preto) e a curva de referência da EPE (azul) justamente nesses períodos. Nos horários intermediários, quando há observações confiáveis do ONS disponíveis, a estimativa retorna próxima à referência, evidenciando que o filtro consegue se recuperar assim que dispõe de informação válida.

Figura 24 – Correção dos dados de vento verificado para dois dias distintos.

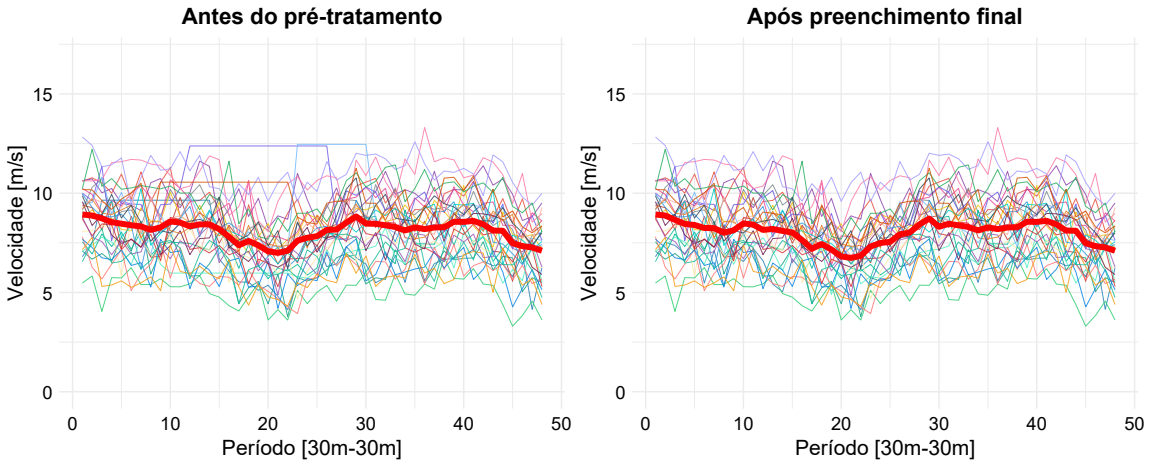


Fonte: Elaboração própria.

Em complemento às Figuras 23 e 24, as Figuras 25 e 26 apresentam o resultado da

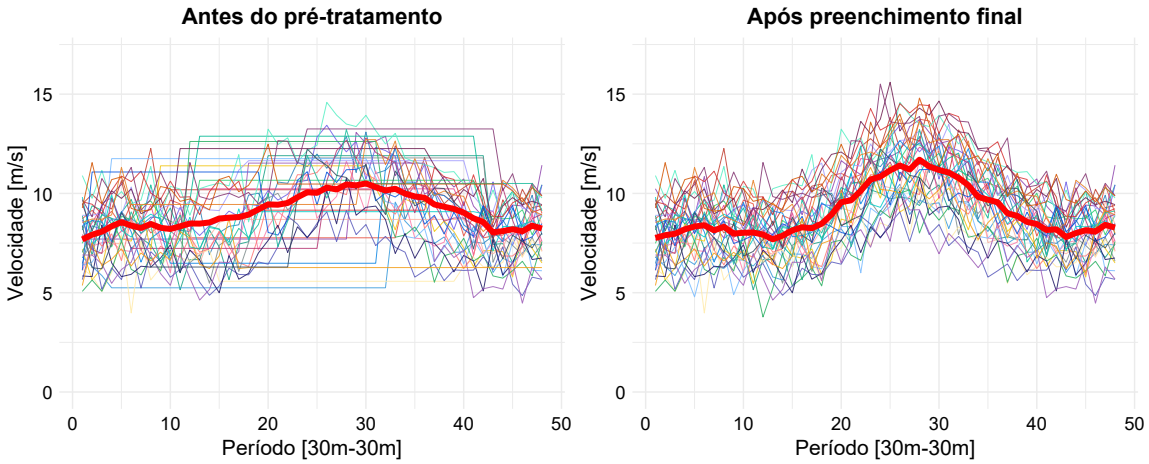
etapa de preenchimento final, na qual os pontos previamente marcados como dados faltantes são reconstruídos a partir do padrão ajustado associado ao *cluster*. É possível notar que a correção elimina as falhas e gera séries mais aderentes ao comportamento típico do agrupamento. Esses resultados são valiosos para alimentar os modelos de previsão, uma vez que mitigam o efeito de dados comprometidos sobre o aprendizado.

Figura 25 – *Cluster* com poucas falhas: antes do pré-tratamento e após o preenchimento final, em séries diárias de vento.



Fonte: Elaboração própria.

Figura 26 – *Cluster* com muitas falhas: antes do pré-tratamento e após o preenchimento final, em séries diárias de vento.



Fonte: Elaboração própria.

O filtro demonstra elevada capacidade de identificar corretamente a maioria dos dados espúrios. Todavia, como qualquer método estatístico, está sujeito à ocorrência de falsos positivos, nos quais dados válidos são eventualmente identificados e substituídos como se fossem espúrios. Apesar dessa limitação, tais ocorrências são esporádicas e representam uma parcela reduzida do conjunto de dados, não comprometendo a robustez e a confiabilidade do procedimento de tratamento adotado. Desta forma, conclui-se que o modelo proposto contribui de forma

substancial para o bom desempenho do modelo completo de tratamento dos dados e dos modelos de previsão. Em se tratando de empregabilidade, esse modelo pode ser facilmente aplicado nos processos, devido à baixa complexidade do algoritmo e da velocidade de processamento.

A principal limitação desta metodologia é que ela trata as variáveis de forma independente, o que pode resultar em pares distorcidos de potência versus vento. Portanto, é importante destacar que os resultados do modelo ainda podem ser aprimorados ao extrapolá-lo para casos multivariados, nos quais informações de geração e outras grandezas, como velocidade e direção do vento, temperatura, irradiância etc., são modeladas em conjunto. Apesar disso, os resultados promissores dessa abordagem alternativa para o tratamento e composição das séries históricas motivaram a publicação de um trabalho, conforme apresentado em [39].

3.7 Considerações finais do capítulo

Este capítulo apresentou, em detalhe, a metodologia desenvolvida para o tratamento e a recomposição das séries históricas de vento e geração, etapa que constitui a base sobre a qual se apoiam os modelos de estimação e de previsão discutidos nos capítulos seguintes. Foram descritos os procedimentos de limpeza e controle de qualidade, a composição hierárquica de múltiplas fontes de medição (ONS, CCEE e EPE), o preenchimento de lacunas baseado em correlação espacial entre parques vizinhos e o modelo estatístico de imputação fundamentado no filtro de Kalman, em uma formulação em espaço de estados de nível e tendência local.

A Tabela 4 sintetiza o papel de cada base de dados observados ao longo do tratamento, por grandeza e por etapa. Os dados do ONS, de geração e de vento, são a base principal: inicializam as séries históricas e passam pelos filtros de limpeza apresentados na Seção 3.2. Os dados da CCEE (geração) e da EPE (vento) atuam como fontes complementares e independentes: nas avaliações cruzadas da limpeza, permitem identificar inconsistências nos dados do ONS e, na composição e no preenchimento das séries (Seções 3.3 e 3.4), permitem recuperar informações ausentes. Em nenhuma dessas etapas as bases da CCEE e da EPE substituem as do ONS ou são tratadas como referência absoluta.

Os resultados obtidos com a aplicação dessas etapas de tratamento são apresentados na Seção 6.1, no capítulo de resultados.

Parte dos conceitos e resultados aqui apresentados foi sintetizada em produções científicas decorrentes desta pesquisa. Em [39], foi proposto o modelo estatístico para tratamento de dados faltantes e dados espúrios em séries temporais voltadas à previsão de geração eólica, que serviu de base para a abordagem de imputação adotada nos próximos capítulos. Mais recentemente, em [30], a metodologia foi consolidada e avaliada em larga escala, considerando mais de mil parques eólicos do SIN, demonstrando sua capacidade de reduzir substancialmente o percentual de dados faltantes e inconsistentes nas séries de vento e geração. Em relação a esses trabalhos, a abordagem apresentada nesta tese é mais abrangente, detalhando integralmente cada etapa

Tabela 4 – Papel de cada base de dados observados nas etapas de limpeza e composição das séries históricas.

Grandeza e base	**Característica do envio**	**Função no tratamento**	**Etapas em que é utilizada**
Geração – ONS (SSC)	Tempo real (4 s); mais suscetível a falhas de comunicação, perdas de informação e inconsistências operacionais	Base principal; inicializa a série histórica	Limpeza: *flags* de qualidade, estimador de estado, limites físicos, dados congelados e dias não representativos. Composição: primeira prioridade
Geração – CCEE (SMF)	Envio não ocorre em tempo real (discretização horária e de 5 min); menor ocorrência de falhas	Complementar: validação cruzada, identificação de inconsistências e recuperação de lacunas; também serve de referência de correlação para o vento do ONS	Limpeza: avaliação cruzada com a geração do ONS. Composição: preenchimento das lacunas (5 min e, em seguida, horária)
Vento – ONS (SSC)	Tempo real (4 s); mais suscetível a falhas de comunicação, perdas de informação e inconsistências operacionais	Base principal; base da série histórica de vento	Limpeza: *flags* de qualidade, limites físicos, dados congelados e dias não representativos. Composição: série a ser suplementada
Vento – EPE (AMA)	Medido e armazenado nos parques, e enviado em lotes, com intervalos de vários dias (discretização de 10 min)	Complementar: validação cruzada, identificação de inconsistências e recuperação de lacunas	Limpeza: avaliação cruzada com o vento do ONS e com a geração da CCEE. Composição: regressão linear ONS $\times$ EPE e suplementação das lacunas

Fonte: Elaboração própria.

do fluxo metodológico, as fontes de dados envolvidas e as estratégias específicas adotadas para cada grandeza, além de tratá-la como insumo direto para os modelos de estimação e previsão desenvolvidos na sequência.

4 MODELO PARA ESTIMAÇÃO DE GERAÇÃO EÓLICA

Concluída a etapa de tratamento e recomposição das séries históricas, descrita no Capítulo 3, dispõe-se de séries consistentes de vento e geração que servem de base para a modelagem da relação entre essas duas grandezas. Este capítulo apresenta o modelo de estimação da geração eólica desenvolvido neste trabalho, fundamentado na construção das curvas de Vento $\times$ Potência (V $\times$ P). Tais curvas são essenciais não apenas para a estimação da geração potencial, inclusive em situações de corte de geração, mas também constituem um insumo direto para os modelos de previsão apresentados no Capítulo 5.

A Figura 27 apresenta a estrutura do modelo de estimação, correspondente ao bloco de estimação do fluxograma apresentado na Figura 1. O processo utiliza como entradas as séries de dados observados previamente tratadas na etapa anterior e os registros de eventos de *curtailment*. Inicialmente, os eventos de *curtailment* são empregados para identificar e remover os períodos em que a geração esteve sujeita a restrições operativas. Em seguida, as séries de geração são corrigidas em função da disponibilidade de potência da usina, de modo a compensar intervalos em que a capacidade instalada esteve parcialmente indisponível.

Sobre as séries ajustadas, aplica-se um filtro de dispersão V $\times$ P individualizado por usina, com o objetivo de eliminar pontos anômalos ainda presentes nos dados. A partir do conjunto de observações selecionadas, são então construídas dinamicamente as curvas V $\times$ P específicas de cada usina. Por fim, utilizando as curvas obtidas e os registros de vento observado, o modelo calcula a geração estimada por usina.

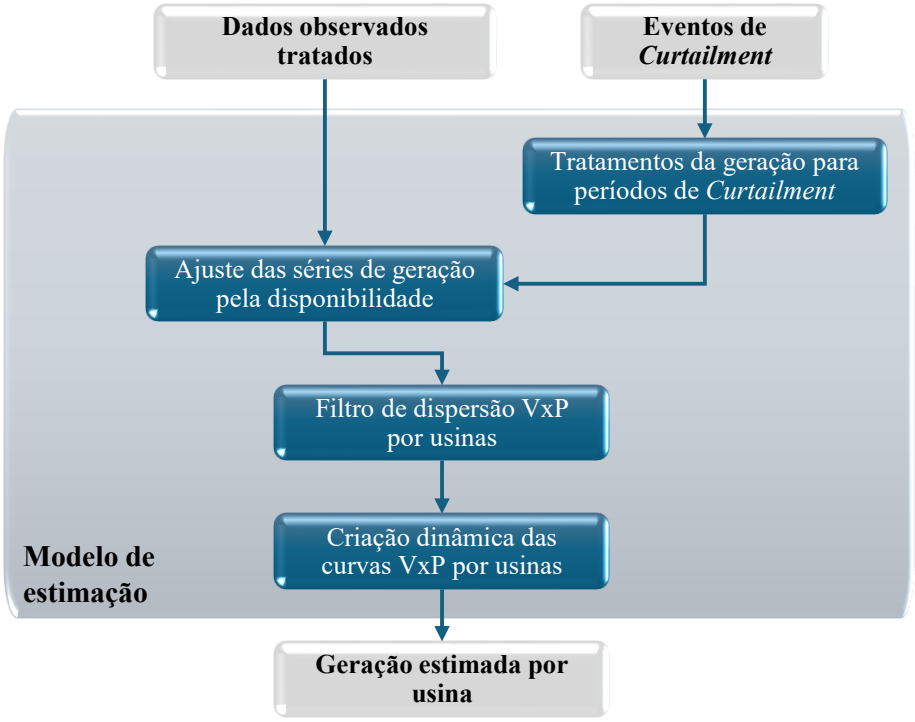
As seções a seguir detalham cada uma dessas etapas. Como a metodologia resumida depende da obtenção das curvas V $\times$ P, que permitem estimar a geração a partir de um valor de vento conhecido, descreve-se inicialmente como essas curvas foram levantadas neste trabalho.

4.1 Tratamentos da geração para períodos de *curtailment*

O procedimento atualmente adotado para o tratamento da geração nos períodos com eventos de *curtailment* é relativamente simples, consistindo apenas na remoção desses intervalos da série histórica. Entretanto, outras abordagens podem ser consideradas com o objetivo de reduzir a quantidade de dados descartados, sem comprometer a qualidade do processo de estimação, especialmente diante do crescimento significativo no volume de eventos registrados.

Uma estratégia possível consiste em avaliar, para cada período marcado com evento de *curtailment*, se a restrição esteve efetivamente ativa do ponto de vista da geração da usina. Em muitos casos, embora exista a sinalização de um evento de *curtailment*, o recurso energético disponível, como o vento, pode estar em níveis reduzidos, fazendo com que a geração da usina não alcance o limite de potência estabelecido pelo ONS. Nessas situações, a remoção do dado pode ser desnecessária, uma vez que a geração observada não foi efetivamente afetada pela

Figura 27 – Estrutura do modelo de estimação da geração.



Fonte: Elaboração própria.

restrição operativa.

4.2 Ajuste da geração com as informações de potência disponível

Os dados de potência disponível, depois de tratados, são utilizados para ajustar valores históricos de geração que foram observados em períodos nos quais a usina passava por alguma restrição de geração, seja por manutenção ou por falha de algum aerogerador. Esse ajuste, realizado a partir da Equação (4.1), reduz a dispersão V $\times$ P da usina e melhora o desempenho do modelo de estimação de geração.

$$G_{aj} = G_v \cdot \frac{P_{\max}}{P_{disp}} \tag{4.1}$$

Em que G_{aj} representa o dado de geração ajustado; G_v, a geração verificada; $P_{\max}$, o valor que representa a potência disponível máxima atingida pela usina; e P_{disp}, o dado de potência disponível recebido pelo ONS. A potência máxima $P_{\max}$ é definida como o menor valor entre um quantil elevado da potência disponível histórica e a potência instalada multiplicada por 1,1.

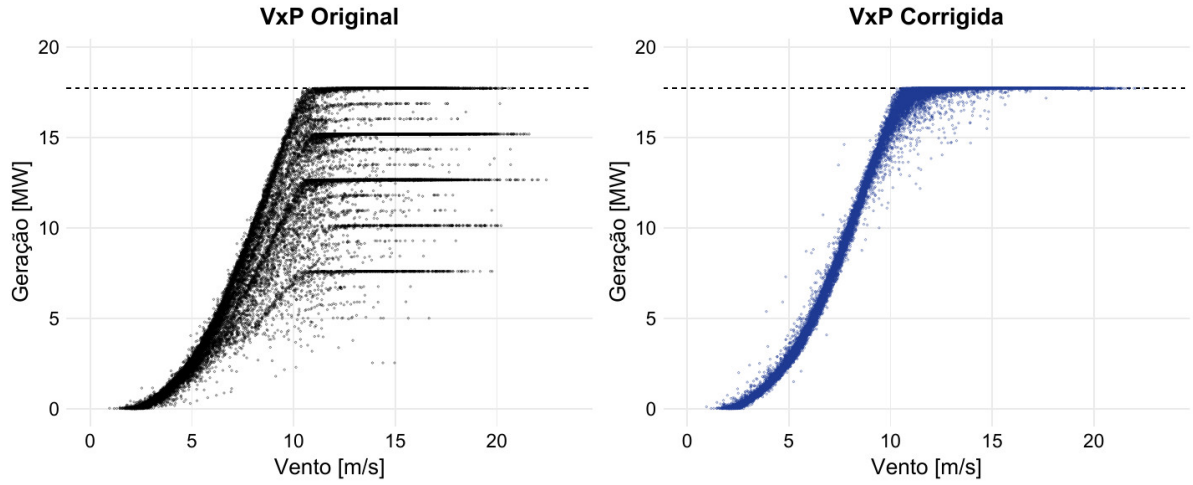
O algoritmo implementado utiliza uma matriz para armazenar os fatores pelos quais a geração observada será multiplicada, isto é, a relação entre a potência máxima e a potência disponível ($P_{\max}/P_{disp}$). Em um primeiro momento, a matriz é inicializada como uma matriz

unitária. O valor unitário será mantido sempre que a potência disponível for igual a P_{max}. Neste caso, a geração verificada não será alterada.

Por outro lado, quando a potência disponível for inferior a P_{max}, o fator de ajuste assumirá um valor maior que 1. Como consequência, os valores de geração serão deslocados para valores mais próximos da dispersão V $\times$ P que seria obtida caso o parque estivesse operando à plena potência.

A Figura 28 mostra o resultado do ajuste da geração para uma usina eólica do SIN. É visível que, no histórico selecionado, a usina operou em regiões de potência consideravelmente abaixo da sua capacidade instalada, o que se observa pelos patamares horizontais marcantes no painel da dispersão original. Após o ajuste pela potência disponível, a dispersão V $\times$ P apresenta uma variabilidade significativamente menor, com os pontos se concentrando em torno da curva característica esperada da turbina, refletindo o potencial real de geração da usina em condição plena.

Figura 28 – Dispersão Vento $\times$ Potência antes e após o ajuste pela potência disponível para uma usina eólica do SIN.



Fonte: Elaboração própria.

Alguns cuidados devem ser levados em consideração para que o ajuste proposto não cause distorções na dispersão Vento $\times$ Potência. O primeiro consiste em verificar se o valor ajustado de geração viola o limite estabelecido para a potência disponível máxima (P_{max}). Em caso afirmativo, a geração deverá ser substituída por P_{max}, com exceção dos casos em que a geração verificada original já violava esse limite. Nessa situação, deve-se reestabelecer o valor observado no histórico.

Um outro ponto de atenção é a situação em que a restrição de geração é significativa, ou seja, quando se observam valores consideravelmente baixos de potência disponível. Nesses casos, opta-se por eliminar os pontos correspondentes da dispersão V $\times$ P, o que é feito zerando-se o fator correspondente da matriz de ajuste. Observe que, ao se eliminar esses pontos, obtém-se

uma redução ainda maior da dispersão. Valores muito baixos de potência disponível podem indicar uma operação com um número muito reduzido de máquinas e, portanto, o ajuste dado pela Equação (4.1) pode não ser suficiente para corrigir os respectivos pontos de geração. Dessa forma, a eliminação desses dados acarreta maiores ganhos ao final do processo.

Para ilustrar o efeito da correção da geração verificada sobre o desempenho do modelo de previsão, as Figuras 29 e 30 apresentam a evolução temporal da geração verificada (em vermelho) sobreposta à geração prevista pelo modelo, com a previsão segmentada entre o período de treino (em amarelo) e o período de teste (em preto). Em cada figura, o painel superior exibe a série completa, abrangendo os períodos de treino e teste, enquanto o painel inferior apresenta um detalhamento (zoom) sobre o conjunto de teste, permitindo melhor inspeção visual das diferenças entre a previsão e a geração verificada nesse período.

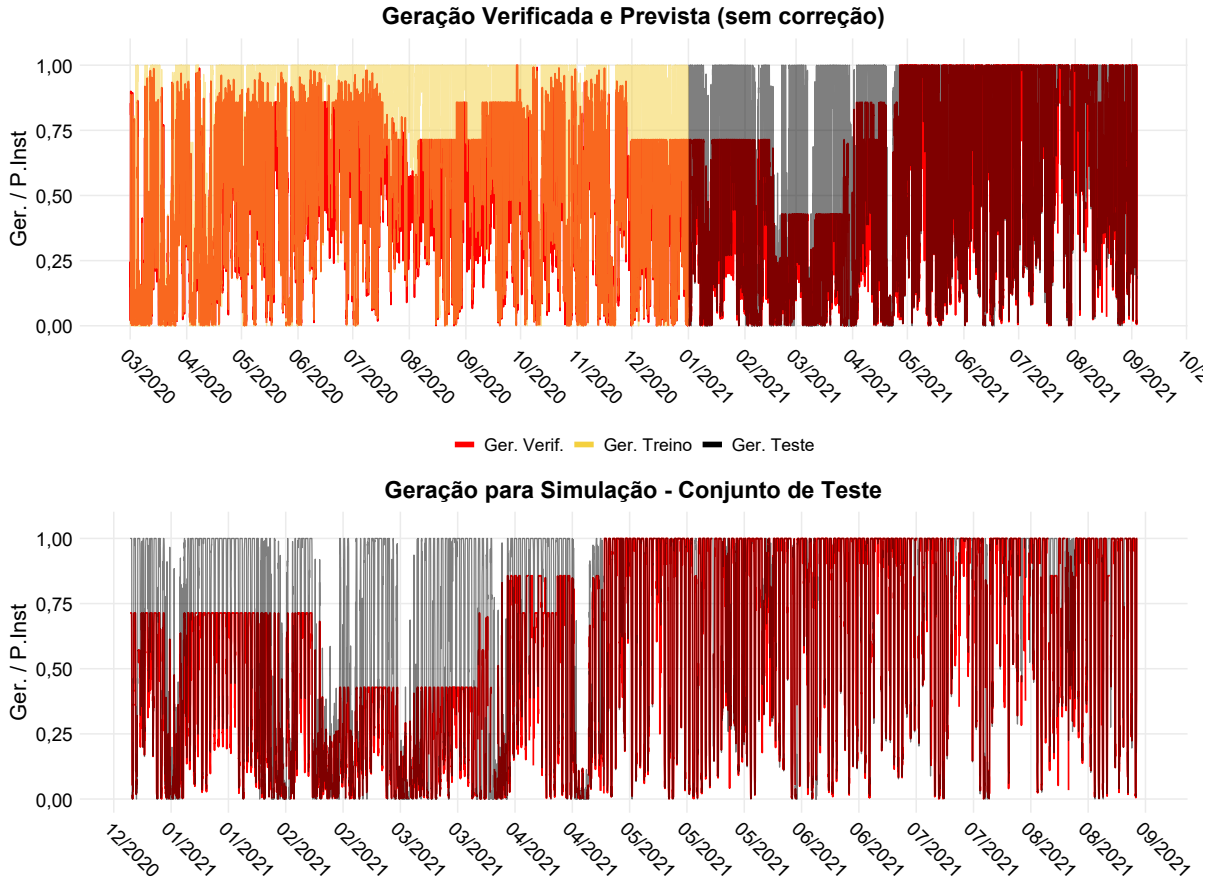
A Figura 29 apresenta o resultado obtido quando o modelo é treinado utilizando a geração verificada original, sem aplicação da correção pela potência disponível. Nesse caso, o modelo busca aprender o potencial máximo de geração da usina a partir das relações entre vento e potência observadas. Entretanto, como existem diversos instantes em que a usina opera com máquinas paradas ou com restrições operacionais (ou seja, com potência disponível inferior à capacidade instalada), a curva verificada (vermelha) apresenta valores reduzidos em vários momentos, mesmo quando as condições de vento permitiriam geração elevada. Como consequência, a previsão (curvas amarela e preta) frequentemente supera a geração verificada nesses momentos, conforme se observa nitidamente no painel inferior. Esse descolamento aparece como um erro de previsão, quando, na verdade, reflete apenas a operação da usina abaixo de seu potencial pleno em razão de fatores operacionais.

Já a Figura 30 apresenta o resultado quando se aplica a correção pela potência disponível. A previsão é inicialmente gerada considerando o potencial pleno da usina (isto é, supondo que todas as máquinas estejam disponíveis) e, em seguida, é ajustada pelo valor da potência disponível em cada instante de tempo, trazendo a estimativa para a realidade operativa da usina em cada momento. Dessa forma, quando a usina está operando com restrições, a previsão também acompanha esses limites; quando opera em condição plena, a previsão reflete o potencial completo. O resultado é uma aderência consideravelmente maior entre a curva prevista (preta) e a curva verificada (vermelha) no período de teste, conforme se observa no painel inferior, o que reduz substancialmente o viés de superestimação verificado na Figura 29. Essa comparação evidencia a importância de incorporar a informação de potência disponível na construção da previsão, especialmente em usinas com histórico significativo de restrições operacionais.

4.3 Metodologias para estimação das curvas Vento $\times$ Potência

De modo geral, a curva V $\times$ P é sempre disponibilizada pelo fabricante para cada aerogerador da usina eólica. Desconsiderando as limitações físicas dos equipamentos, a potência disponível no vento tem relação cúbica com a velocidade; isso significa que um desvio na veloci-

Figura 29 – Evolução temporal da geração verificada e prevista sem a correção pela potência disponível.



Fonte: Elaboração própria.

dade poderia ser potencializado no caso do uso da curva teórica diretamente. A Equação (4.2) relaciona a potência teórica disponível no vento:

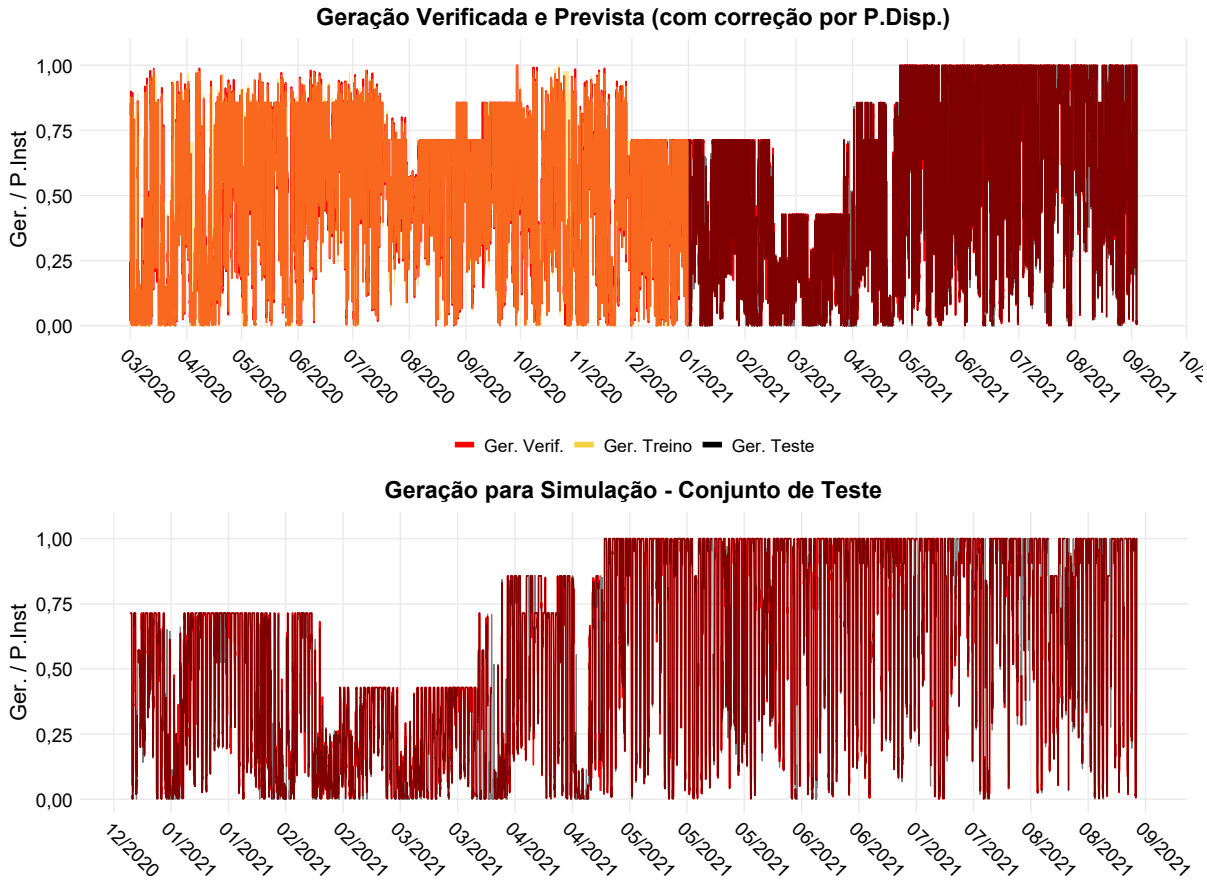
$$P(v) = \frac{1}{2}\rho A v^3 \tag{4.2}$$

em que $P(v)$ é a potência dependente do vento v; A é a área varrida pelo rotor (m^2); ρ é a densidade do ar (kg/m^3); e v é velocidade do vento (m/s).

De acordo com essa equação, a geração aumentaria infinitamente com o vento. Entretanto, a relação real, considerando as limitações físicas de engenharia do gerador ao transformar vento em potência, tem-se a curva V $\times$ P fornecida pelo fabricante. A Figura 31 mostra, para o aerogerador ENERCON E-70 E4, a comparação da curva utilizando a Equação (4.2) com os parâmetros desse aerogerador (curva em vermelho) e a curva V $\times$ P fornecida pelo fabricante (em verde).

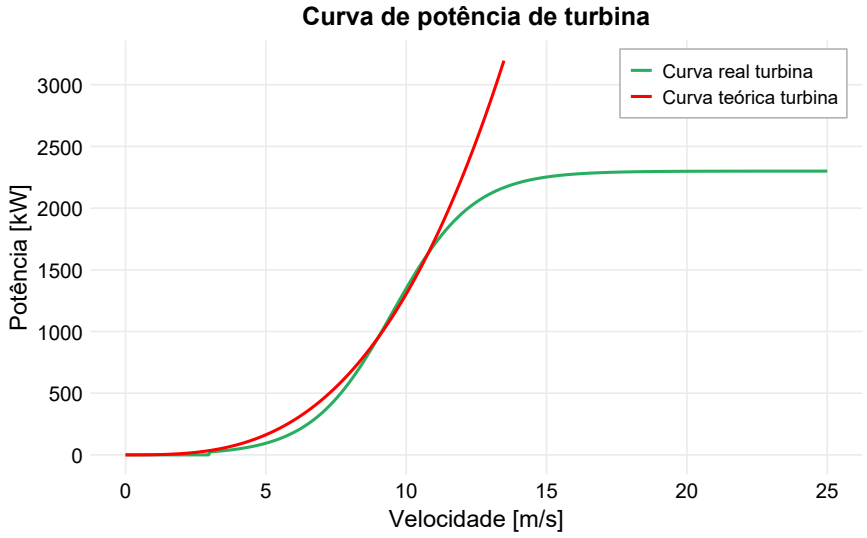
Embora o uso da curva fornecida pelo fabricante possa parecer retratar bem a relação entre vento e potência, pode haver discrepâncias entre a curva dada pelo fabricante e os dados verificados, devido a diversos fatores, tais como as condições de operação em campo serem

Figura 30 – Evolução temporal da geração verificada e prevista com a correção pela potência disponível.



Fonte: Elaboração própria.

Figura 31 – Curva de potência de uma turbina eólica: comportamento teórico e real.



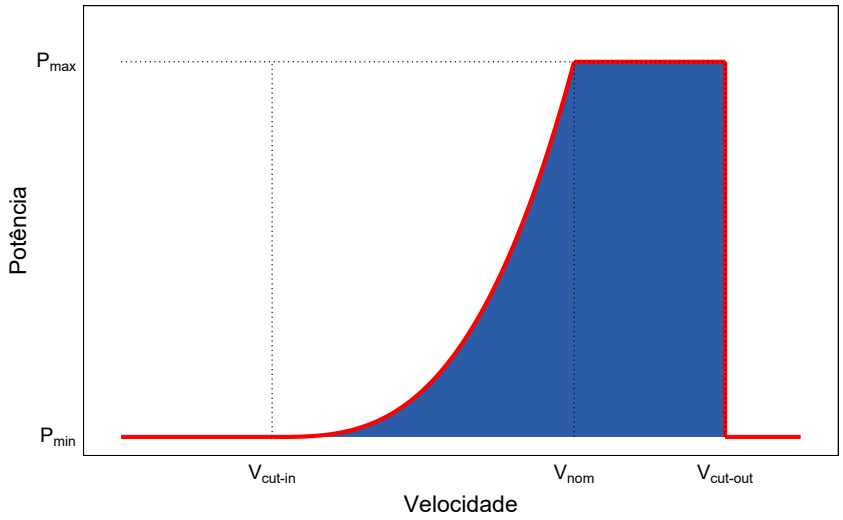
Fonte: Elaboração própria.

diferentes das condições sob as quais tais curvas foram levantadas. Sendo assim, é preferível estimar essa curva a partir de dados verificados de geração e vento dos parques.

Atualmente não há disponibilidade de dados observados por aerogerador para todas as usinas do SIN e o vento previsto que será utilizado para compor a previsão de geração é o mesmo para a usina como um todo. Sendo assim, não há necessidade de levantar uma curva para cada aerogerador, bastando estimar uma curva para o parque como todo.

Em geral, a curva de potência física pode ser representada por dois parâmetros principais de velocidade denominados *cut-in* e *cut-out*. A velocidade *cut-in* é a velocidade mínima de vento que permite o início de geração de energia elétrica pelo aerogerador. A velocidade *cut-out* é a velocidade máxima de vento que a máquina ainda poderá operar sem ser desligada. A Figura 32 ilustra o comportamento teórico de um aerogerador.

Figura 32 – Representação esquemática da relação Vento $\times$ Potência teórica.



Fonte: Elaboração própria.

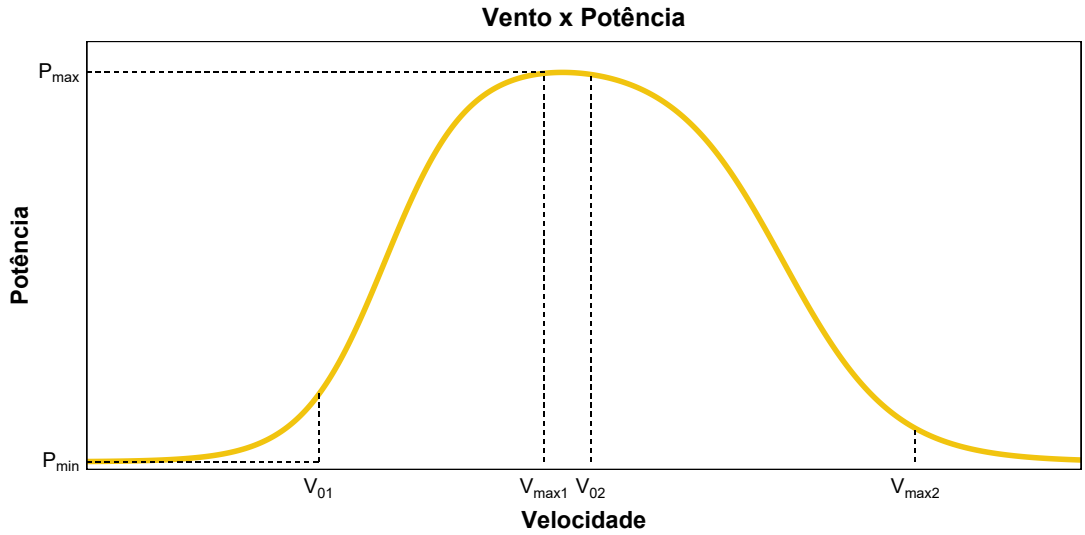
Há vários métodos para a estimação da curva V $\times$ P, como ajustes polinomiais, redes neurais e equações paramétricas, a exemplo das regressões logísticas e das funções sigmoides. Cada abordagem apresenta vantagens e limitações. Os ajustes polinomiais de diferentes graus poderiam ser utilizados, sendo o de terceiro grau o mais intuitivo, dada a relação cúbica entre o vento e a geração; contudo, a região de saturação não seria adequadamente capturada. O uso de redes neurais é outra possibilidade, e uma rede relativamente pequena já leva a bons resultados. Existem, ainda, equações paramétricas que se adaptam muito bem ao perfil da curva V $\times$ P, como é o caso da regressão logística apresentada na Equação (3.4).

Para a maioria das usinas eólicas existentes no Brasil, a regressão logística se adapta muito bem, pois a tecnologia dos aerogeradores satura a geração para um determinado valor de velocidade de vento até atingir um ponto de corte e partir dali a máquina passa a não gerar. Entretanto, algumas usinas possuem tecnologias mais avançadas que permitem às máquinas gerar energia para velocidades mais altas, mesmo que com redução do montante de geração.

Sendo assim, as curvas dessas usinas não seriam adequadamente modeladas, necessitando de outra classe de funções.

Uma classe das funções que se adapta muito bem nas duas situações acima descritas é a das funções sigmoides. A Equação (4.3) ilustra o arranjo dos parâmetros que formam essa função, sendo que os parâmetros $P_{\min}$, $P_{\max}$, V_{01}, $V_{\max 1}$, V_{02} e $V_{\max 2}$ representam as características próprias de cada parque eólico, conforme ilustrado na Figura 33.

Figura 33 – Curva genérica Vento × Potência com limites operacionais.



Fonte: Elaboração própria.

$$G_e = P_{\min} + \frac{P_{\max} - P_{\min}}{1 + e^{-\frac{8}{t_{s1}}(V_v - t_{c1})}} - \frac{P_{\max}}{1 + e^{-\frac{8}{t_{s2}}(V_v - t_{c2})}} \tag{4.3}$$

$$t_{s1} = V_{\max 1} - V_{01} \qquad \text{e} \qquad t_{s2} = V_{\max 2} - V_{02} \tag{4.4}$$

$$t_{c1} = \frac{V_{\max 1} + V_{01}}{2} \qquad \text{e} \qquad t_{c2} = \frac{V_{\max 2} + V_{02}}{2} \tag{4.5}$$

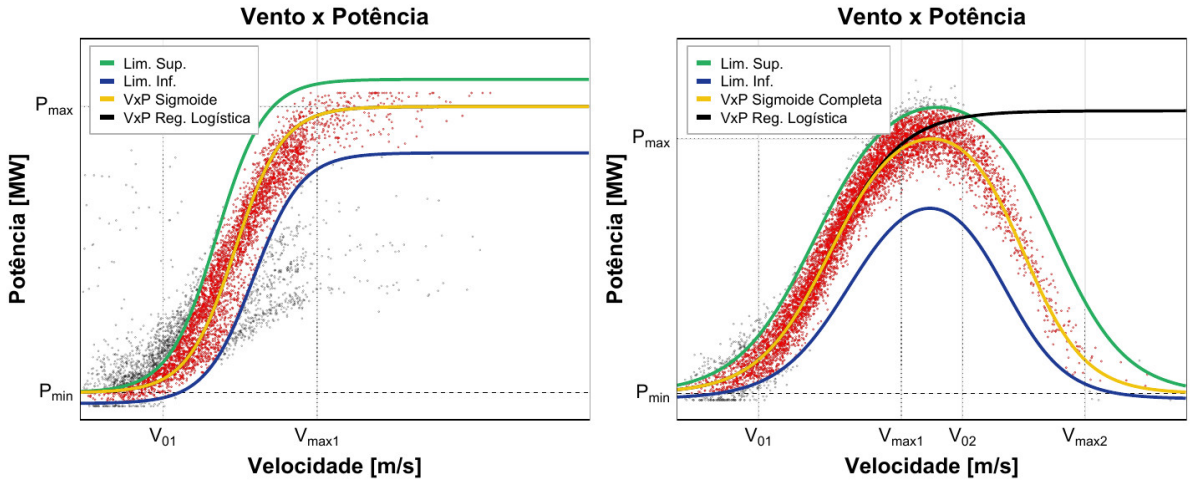
Em que G_e é a geração estimada; $P_{\min}$ representa a assíntota inferior à região de potência mínima da curva V × P; $P_{\max}$ a assíntota superior à região de potência máxima; V_{01} a coordenada x do ponto de inflexão para a região de potência e velocidade de vento baixas; $V_{\max 1}$ a coordenada x do ponto de inflexão para a região de potência alta com derivada positiva, isto é, com velocidade do vento e potência crescentes; V_{02} a coordenada x do ponto de inflexão para a região de potência alta com derivada negativa, isto é, com velocidade do vento crescente e potência decrescente; $V_{\max 2}$ a coordenada x do ponto de inflexão para a região de potência e velocidade de vento altas; e V_v o vento verificado.

Conforme supracitado, cada usina eólica possui características distintas. As curvas V × P de usinas que possuem uma região de saturação bem definida para velocidades elevadas

de vento podem ser modeladas através de uma função com a primeira e a segunda parcela da Equação (4.3). Em caso de usinas em que a geração eólica a partir de dada velocidade do vento começa a decair, é necessário utilizar também a terceira parcela da Equação (4.3) para modelar as respectivas curvas V $\times$ P.

Na sequência, são apresentadas imagens que comparam os resultados obtidos através das duas metodologias comentadas até aqui. Observa-se que o ajuste das curvas das usinas com saturação bem definida obtido através de ambas as metodologias é bastante semelhante, conforme mostrado nas curvas preta e amarela do painel à esquerda da Figura 34. Na referida figura, os rótulos dos eixos foram substituídos pelos parâmetros da Equação (4.3) (V_{01}, $V_{\max 1}$, V_{02}, $V_{\max 2}$, $P_{\min}$ e $P_{\max}$) para evidenciar as regiões características da curva. Já no caso das usinas com decaimento, conforme mostra o painel à direita, apenas a metodologia de estimação dos parâmetros da sigmoide completa é capaz de representar o comportamento de redução de geração em situações em que ventos mais elevados são registrados.

Figura 34 – Comparação curvas V $\times$ P estimadas pela sigmoide e pela regressão logística.



Fonte: Elaboração própria.

A curva em preto representa o ajuste obtido pelo uso da função logística simples, Equação (3.4), enquanto a curva em amarelo representa o ajuste utilizando a sigmoide completa, Equação (4.3). Antes do cálculo da curva V $\times$ P, é necessário aplicar um filtro sobre a dispersão dos dados verificados, sendo apresentado na Figura 34 um exemplo desse filtro pelas curvas limites (em verde e azul). A metodologia desse filtro é baseada em [145] e será apresentada na próxima seção.

Vale salientar que a curva em amarelo, obtida pela função sigmoide completa, respeita o formato conferido pelas curvas limites. Já a curva em preto, obtida pela regressão logística, por construção, cria uma região de saturação que não se encaixa adequadamente à dispersão real dos dados no caso das usinas com decaimento, conforme se observa no painel à direita: a curva preta acompanha a sigmoide completa durante a subida e até parte da saturação, mas mantém-se em

platô horizontal enquanto a sigmoide completa decai com o aumento da velocidade do vento. Entretanto, as duas abordagens serão utilizadas para apresentação de testes neste trabalho, dado que, no caso do painel à esquerda da Figura 34, ambas se adequam muito bem.

4.4 Processo de refinamento da dispersão Vento × Potência das usinas

Para uma boa estimação da curva V × P, inicialmente é necessário um tratamento específico da dispersão de cada usina. O conhecimento do comportamento padrão da dispersão V × P fornece informações úteis para a identificação e correção de dados espúrios (*outliers*), conforme modelos descritos em [52] e [146]. A técnica empregada foi baseada no modelo apresentado em [145], sendo desenvolvida uma série de artifícios matemáticos e computacionais para eliminar automaticamente os pontos da dispersão que não representam valores plausíveis, tentando manter o maior número possível de pontos corretos.

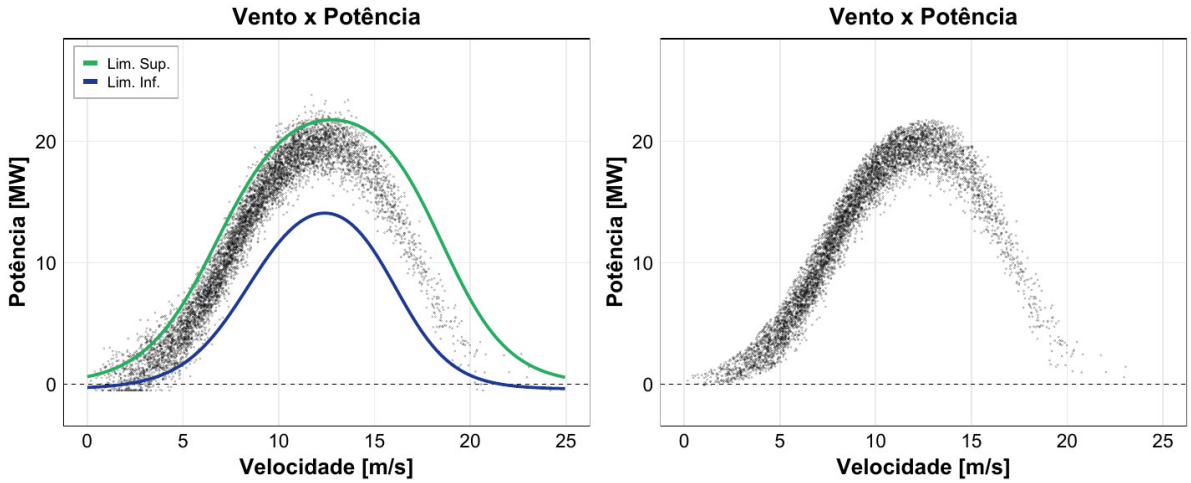
Sendo assim, cria-se uma função superior e uma inferior, de forma a considerar apenas os dados que se encontram dentro dos limites. A criação desses limites é feita com funções sigmoides, Equação (4.3). De posse dos parâmetros $P_{\min}$, $P_{\max}$, V_{01}, $V_{\max 1}$, V_{02} e $V_{\max 2}$ é possível, portanto, a criação dos limites inferiores e superiores para filtragem dos dados. Entretanto, esses parâmetros devem ser estimados a partir da própria dispersão dos dados verificados, uma vez que os parâmetros teóricos nem sempre são compatíveis com os estimados pela regressão na dispersão real. Ressalta-se que os parâmetros da sigmoide são calculados de forma automática pelo modelo que estima as curvas V × P. A ideia geral consiste em selecionar, dentro do conjunto de pontos verificados, aqueles pertencentes aos intervalos prováveis de cada parâmetro da equação. Assim, para cada coeficiente a ser especificado, um conjunto de pontos é selecionado e um determinado quantil desse vetor é calculado e atribuído como o valor do parâmetro desejado.

Por exemplo, na estimação do parâmetro V_{01}, seleciona-se previamente uma faixa de pontos da dispersão considerados representativos do início da rampa de subida; sobre essa faixa, calcula-se um quantil de velocidade que corresponde ao valor de V_{01} adotado. Procedimento análogo é realizado para a estimação dos demais parâmetros das curvas.

Nesse contexto, um dos principais benefícios da metodologia desenvolvida é a capacidade de estimar automaticamente os parâmetros das curvas limite. Isso é crucial para situações práticas, como as operações do ONS, que envolvem centenas de usinas. Tomando como exemplo o cenário brasileiro, com mais de mil usinas eólicas, seria impraticável criar manualmente as curvas limite para todas elas, especialmente considerando o grande número de usinas e a possibilidade de alterações no comportamento da dispersão V × P ao longo do tempo.

A Figura 35 apresenta um exemplo da criação das curvas limites e de como se torna a dispersão após a eliminação de dados incompatíveis. O painel à esquerda mostra a dispersão original sobreposta às curvas limite estimadas, e o painel à direita apresenta apenas os pontos retidos após a aplicação do filtro.

Figura 35 – Aplicação das curvas limites, com a dispersão V × P antes e após a eliminação dos dados incompatíveis.



Fonte: Elaboração própria.

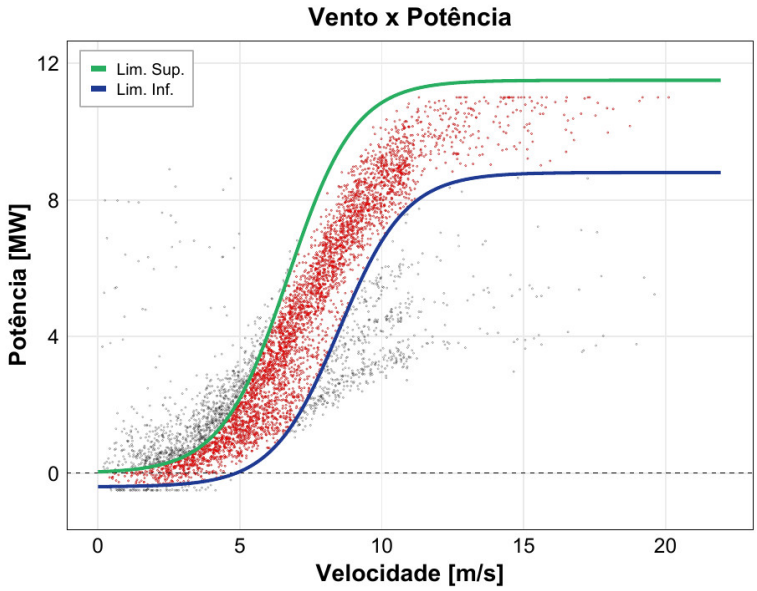
Ressalta-se que os dados eliminados nem sempre representam problemas de medição. Muitos pontos podem corresponder a medições precisas e serem cortados por não estarem dentro dos limites calculados. Esses dados podem ser originários de restrições operativas diversas, tais como: máquinas em manutenção, equipamentos em falha, restrições de geração e/ou de transmissão. Esses fatores fazem com que a usina não gere a respectiva energia mesmo com a disponibilidade de recurso (vento). Dessa forma, embora os dados que estão sendo eliminados possam ser reais, não é interessante mantê-los na calibração dos modelos, pois o que importa são os pontos em que o parque opera em condições usuais.

Em situações normais, a dispersão da curva de potência verificada não é muito acentuada e, neste caso, as limitações impostas apenas melhoram a qualidade do ajuste. Entretanto, em várias situações podem ser encontradas restrições operativas, comportando-se como se houvesse mais de uma curva de potência dentro da mesma figura. Tal comportamento é observado na Figura 36, em que se identifica a presença de uma "segunda dispersão" à direita, com geração saturada em níveis bem inferiores ao P_{max} esperado da usina. Muito provavelmente essa restrição se relaciona a limitações de transmissão, que restringem o escoamento da potência ativa gerada e obrigam a usina a não aproveitar plenamente ventos mais fortes. Os limites superior e inferior calculados automaticamente eliminam esses pontos, mantendo na calibração apenas a dispersão correspondente ao regime de operação plena.

4.5 Estimação das curvas Vento × Potência

Após o tratamento da dispersão V × P, a equação que representa cada usina pode ser selecionada. Nesse sentido, o modelo desenvolvido estima os parâmetros da função sigmoide descrita na Equação (4.3). Assim, a mesma função utilizada no processo de filtragem dos dados

Figura 36 – Dispersão Vento × Potência com os limites inferiores e superiores, evidenciando outras "dispersões" associadas a restrições operativas.



Fonte: Elaboração própria.

verificados é adotada para modelar a curva média V × P de parques eólicos.

A estimação dos parâmetros que permite melhor ajuste da curva à dispersão V × P pode ser realizada através de uma função capaz de selecionar os parâmetros de um modelo não-linear que resultem nos menores desvios quadráticos entre os dados calculados e observados. A função adotada para obter a curva de produtividade utiliza o algoritmo Gauss–Newton no processo de estimação dos parâmetros ótimos.

Em relação às informações que devem ser fornecidas como dados de entrada para a seleção dos parâmetros da curva V × P, pode-se listar a fórmula da sigmoide (incluindo parâmetros e variáveis) e o conjunto de dados verificados de velocidade do vento e geração filtrados pelas curvas limites estabelecidas anteriormente.

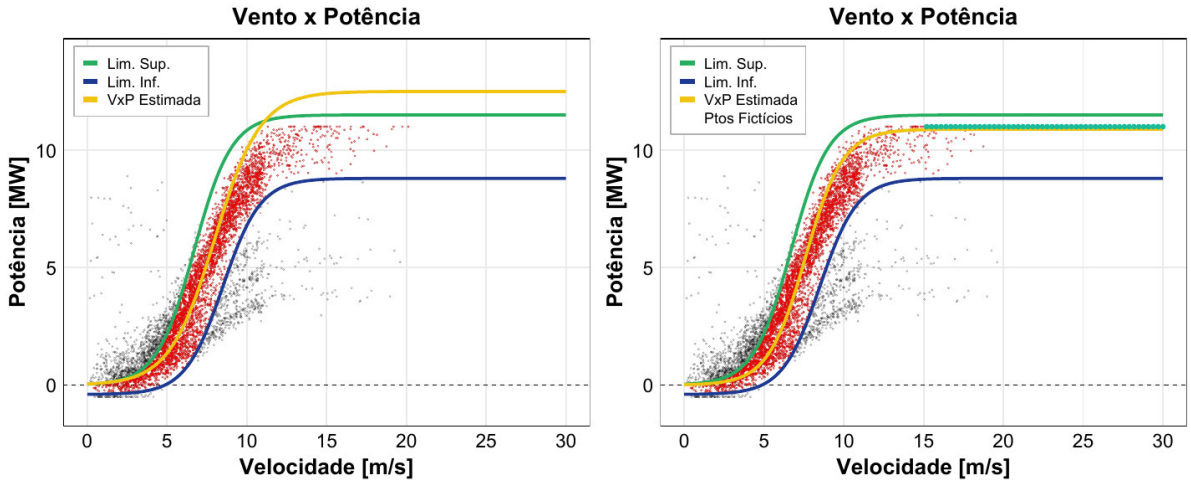
Por fim, o algoritmo Gauss–Newton utiliza uma estimativa inicial dos parâmetros da função sigmoide em seu processo de estimação. É fundamental que os parâmetros iniciais selecionados estejam dentro de uma faixa de parâmetros em potencial, o que evita estimações sem sentido e/ou a não convergência do algoritmo. Como neste momento os limites superior e inferior da curva V × P já são conhecidos, é possível utilizar os parâmetros médios das referidas curvas como dados iniciais para os parâmetros a serem estimados.

Além de ser influenciado pela estimativa inicial fornecida para os parâmetros da função, a convergência do algoritmo Gauss–Newton depende também da quantidade e da qualidade dos dados fornecidos. Um artifício utilizado para evitar a ocorrência de erros e estimações imprecisas é a inclusão de pontos fictícios ao conjunto de dados utilizado no processo de estimação. O artifício em questão é fundamental especialmente nas usinas que possuem um histórico de

medições reduzido e/ou cuja região de saturação não é bem definida.

A Figura 37 ilustra uma usina eólica cujo histórico é escasso na região de saturação. É possível observar no painel à esquerda que o número de pontos V $\times$ P verificados nessa região é reduzido, o que compromete a qualidade da estimação da curva: a curva V $\times$ P estimada (em amarelo) chega a ultrapassar o limite superior de geração da usina, indicando uma extrapolação inadequada para velocidades de vento elevadas.

Figura 37 – Comparação entre curvas V $\times$ P estimadas sem e com a inserção de pontos fictícios.



Fonte: Elaboração própria.

Para construir os pontos fictícios de velocidade do vento e geração na região de saturação, é necessário, em um primeiro momento, selecionar um vetor com os dez valores mais elevados de velocidade do vento verificados, acrescido de um valor adicional correspondente ao maior vento registrado multiplicado por um fator de 1,4. Como condição inicial, verifica-se se os valores de geração associados aos ventos selecionados estão contidos entre as curvas limites superior e inferior estabelecidas.

Quando esse critério é atendido, avalia-se se a mediana do vetor de ventos máximos é inferior ao vento mais rápido registrado no histórico. Em caso positivo, o vetor de ventos fictícios é iniciado no valor da mediana e composto por valores crescentes até a velocidade de 30 m/s. Caso contrário, o vetor fictício parte do próprio valor de vento máximo registrado. O vetor de gerações fictícias, por sua vez, é formado por valores constantes correspondentes ao percentil 80% do vetor de gerações máximas previamente selecionado, no intuito de representar de forma adequada o nível típico da região de saturação.

Existe ainda a situação em que as gerações máximas selecionadas não se encontram contidas entre as curvas limites, em decorrência da escassez de dados na região de saturação. Nesse caso, os vetores de vento e geração máxima passam a respeitar uma premissa menos restritiva: as gerações máximas correspondentes aos ventos mais elevados devem apenas ser inferiores à geração máxima estabelecida pela curva limite superior da usina. A partir desse ponto,

o mesmo método de seleção dos ventos fictícios é adotado, mas as gerações correspondentes passam a ser estimadas aplicando-se os ventos fictícios nas equações das curvas limites e calculando-se uma média ponderada dos valores resultantes, sendo atribuído peso 0,7 à curva superior e 0,3 à curva inferior.

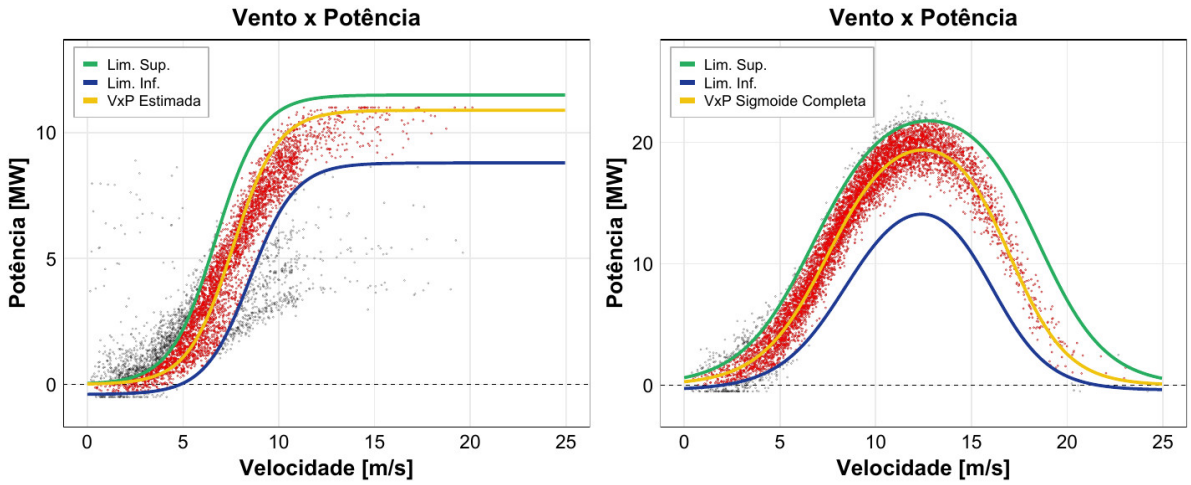
Uma situação prática que pode afetar de forma significativa a estimação dos pontos fictícios é a alta frequência de ocorrência de cortes em condições de vento elevado. Em cenários nos quais esses cortes ocorrem de maneira recorrente, impedindo que a usina opere próxima à sua capacidade nominal nessas faixas de vento ao longo do ano, adota-se uma premissa alternativa: a capacidade instalada máxima disponível da usina eólica no ano anterior à atualização da curva é utilizada como referência para a geração dos pontos fictícios. Essa abordagem, contudo, aplica-se exclusivamente a usinas que não dispõem de dados observados na região de saturação em decorrência dos cortes. Quando existem observações nessa região, as premissas anteriores permanecem mais adequadas, por proporcionarem uma estimação mais representativa do comportamento real da usina.

No painel à direita da Figura 37, os pontos fictícios são plotados em ciano, formando uma reta horizontal no nível de saturação. A inclusão desses pontos no processo de estimação garante um ajuste da curva V $\times$ P bem mais aderente à região de saturação, conforme demonstrado pela curva em amarelo, que agora se mantém alinhada com os pontos fictícios e contida dentro dos limites superior e inferior estabelecidos.

A metodologia de estimação das curvas de potência de usinas eólicas descrita acima possui a vantagem de ser capaz de modelar tanto usinas com a região de saturação bem definida quanto usinas que possuem a característica de decaimento, já explicada anteriormente, conforme observado na Figura 38. O painel à esquerda apresenta o resultado para uma usina com saturação bem definida (a mesma do painel à direita da Figura 37, agora com a curva já ajustada após a inclusão dos pontos fictícios), enquanto o painel à direita ilustra uma usina com decaimento na geração para ventos mais elevados.

Os avanços metodológicos para a estimação das curvas V $\times$ P descritos até o momento possibilitaram a elaboração de outro estudo, detalhado em [41]. Este trabalho pode ser especialmente relevante nos processos de pós-operacional da geração de energia eólica. Conforme estabelecido pela Resolução Normativa ANEEL nº 1.073 [154], de 12 de setembro de 2023, que altera a Resolução Normativa ANEEL nº 1.030 [155], de 26 de julho de 2022, é necessário calcular a referência de geração de energia resultante de eventos de restrição de operação por *constrained-off* a partir da curva de produtividade, que relaciona a potência de saída da usina e o vento verificado. Nesse sentido, o trabalho desenvolvido [41] pode ser diretamente aplicado nessa tarefa.

Figura 38 – Curvas V × P de parques eólicos com saturação e decaimento na geração.



Fonte: Elaboração própria.

4.6 Metodologias alternativas para elaboração das curvas V × P

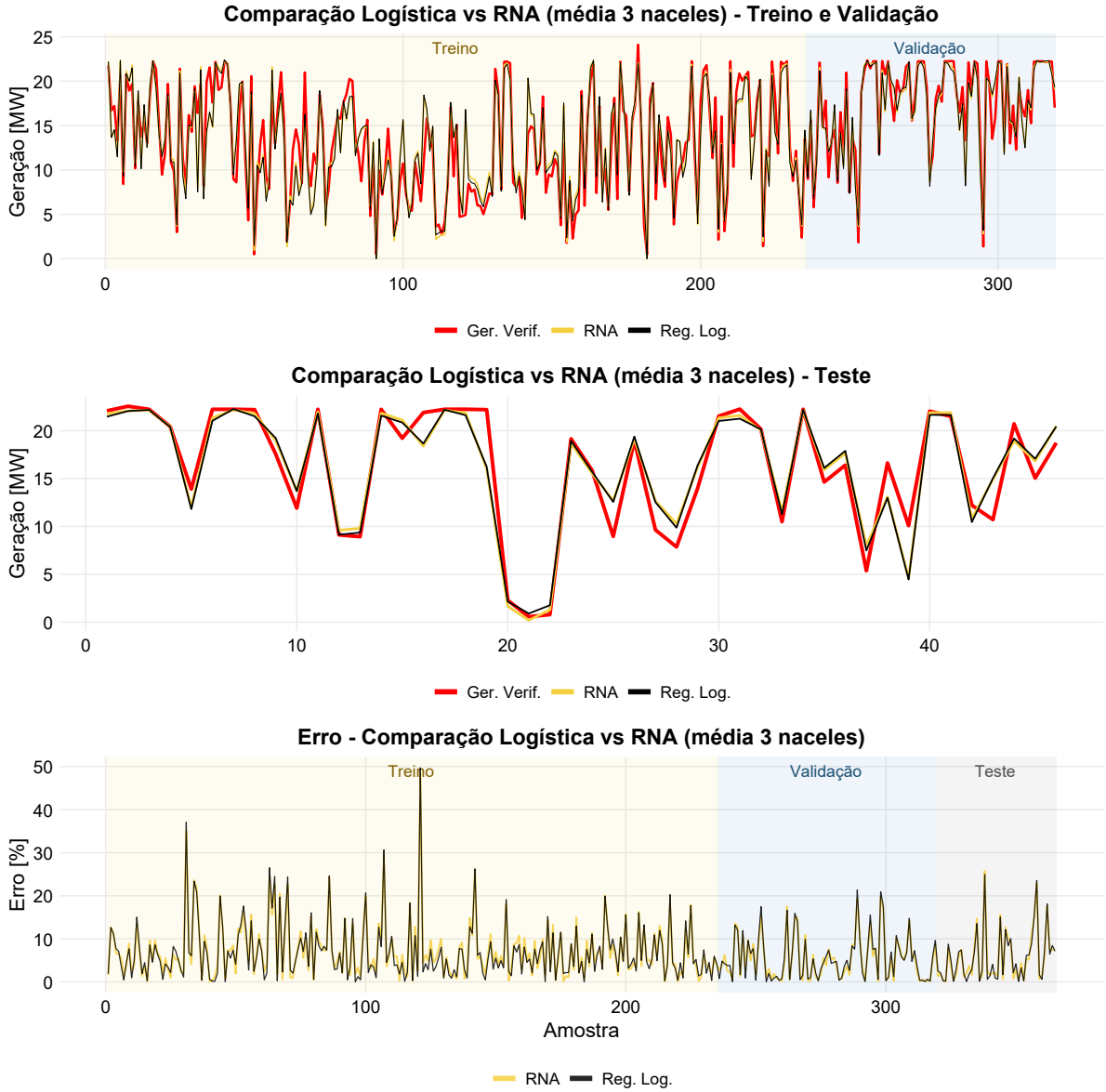
Nesta seção, é apresentada uma metodologia alternativa para obtenção das curvas V × P através da técnica de redes neurais artificiais, com o intuito de validar o uso das regressões paramétricas apresentadas anteriormente. Adicionalmente, é apresentada uma análise da redução da dispersão da curva V × P com o aumento no número de pontos de medição da velocidade do vento.

A Figura 39 apresenta uma comparação entre a geração verificada e a geração estimada pelas duas metodologias, regressão logística e redes neurais, ao longo dos conjuntos de dados utilizados no estudo. A figura é organizada em três painéis: o painel superior exibe os conjuntos de treino (em fundo amarelo claro) e validação (em fundo azul claro); o painel central apresenta o conjunto de teste; e o painel inferior mostra o erro absoluto percentual em relação à potência instalada da usina, calculado como $|G_v - G_p|/P_{inst} \cdot 100\%$, abrangendo treino, validação e teste de forma contínua, com os mesmos sombreados de fundo do painel superior.

De acordo com a Figura 39, observa-se uma boa previsão de geração, com valores estimados próximos aos valores verificados em todos os conjuntos. Entretanto, ao comparar o erro das duas metodologias, não se verifica um ganho significativo da rede neural em relação à regressão logística, indicando que a utilização de redes neurais não traz benefícios expressivos frente à metodologia vigente baseada em regressão.

Foram utilizadas redes Perceptron com múltiplas camadas, ou MLP (*Multi Layer Perceptron*). Nos modelos foram empregadas: função de treinamento *traingdx* (*Gradient Descent With Momentum And Adaptive Learning Rate Backpropagation*); função de transferência *tansig* (*Tan-Sigmoid*); número máximo de 500 épocas; função de desempenho MSE (*Mean Squared Normalized Error*); 1 camada escondida com 12 neurônios.

Figura 39 – Comparação entre regressão logística e redes neurais para o ajuste com a média, nos conjuntos de treino, validação e teste.

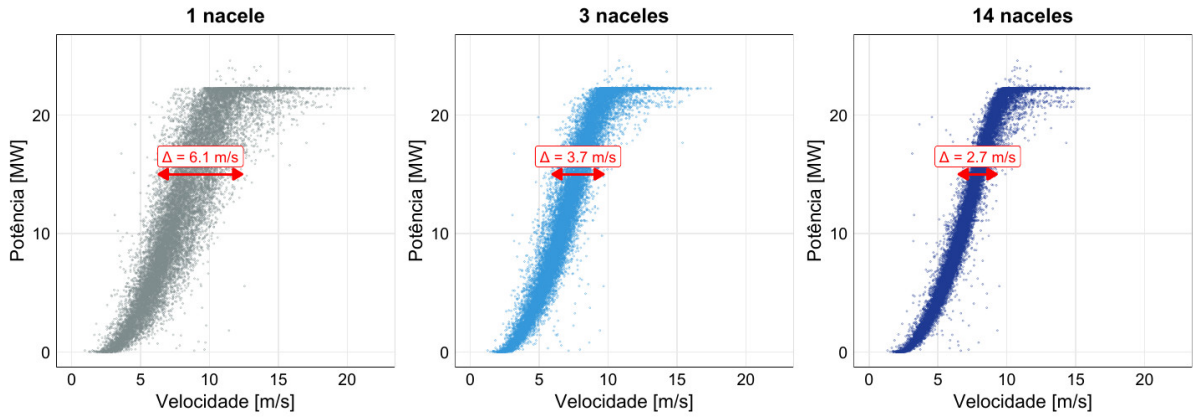


Fonte: Elaboração própria.

Como a utilização de modelos mais sofisticados não apresentou ganhos significativos em relação ao modelo de regressão apresentado, foi realizado um outro estudo com aumento de informações para a estimação das curvas V $\times$ P. Para ilustrar os benefícios da utilização de mais pontos de medições no processo, foi escolhida uma usina com 14 aerogeradores. A Figura 40 apresenta uma análise comparativa da dispersão da curva V $\times$ P utilizando o vento de uma única nacele, a média da velocidade do vento de três naceles distribuídas pelo parque e a média de todas as 14 naceles. Em cada painel, uma seta horizontal vermelha, posicionada no nível de 15 MW, indica o intervalo de velocidades de vento associadas a essa faixa de geração, com o valor Δ correspondente expresso em m/s. Observa-se que o Δ se reduz progressivamente à medida que

se aumenta o número de naceles utilizadas no cálculo do vento equivalente, evidenciando a redução da dispersão da curva V × P. Esse comportamento mostra que o aumento do número de pontos de medição traz maior representatividade da velocidade média do parque, contribuindo para um melhor ajuste da curva V × P, tanto pela regressão logística quanto pela sigmoide, e, consequentemente, para uma estimação mais precisa da geração.

Figura 40 – Comparação das dispersões V × P utilizando o vento de 1, 3 e 14 naceles (cabines dos aerogeradores).



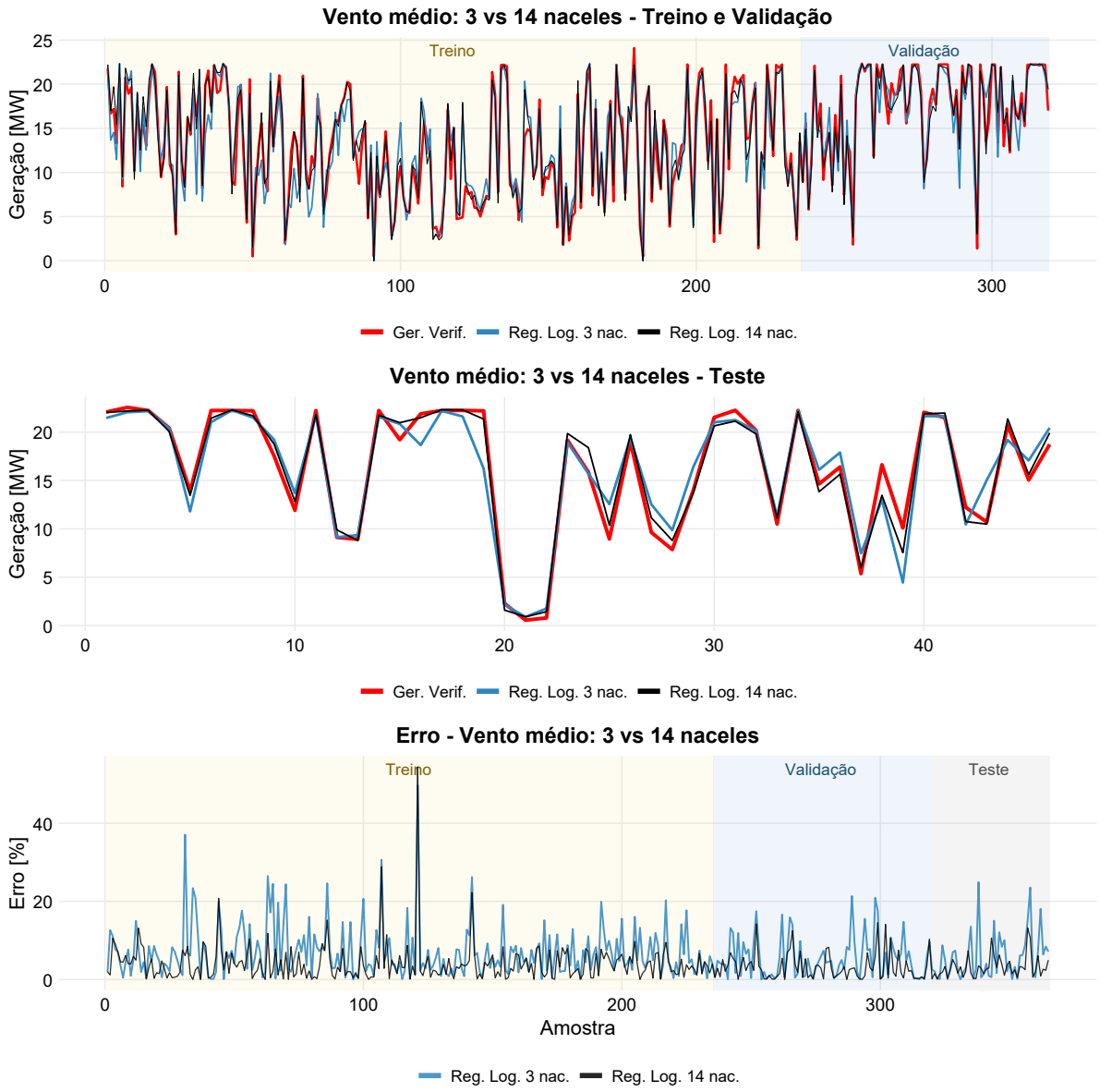
Fonte: Elaboração própria.

A Figura 41 apresenta a evolução temporal da geração verificada e da geração estimada pela regressão logística utilizando o vento médio calculado a partir de 3 e de 14 pontos de medição de nacele. A estrutura da figura é análoga à da Figura 39: o painel superior mostra os conjuntos de treino e validação, o painel central destaca o conjunto de teste e o painel inferior apresenta o erro absoluto percentual ao longo de todos os conjuntos. Observa-se uma boa aderência entre as curvas de geração estimada e observada em ambas as opções. Em relação ao erro, percebe-se valores menores quando se utiliza a média de 14 naceles, embora a utilização de apenas 3 medições já proporcione um ganho relevante no processo de estimação.

Para corroborar as afirmações anteriores, foram construídas curvas V × P considerando todos os dados de vento disponíveis para essa usina. Na Figura 42 são apresentados os erros percentuais para os conjuntos de treinamento e de teste considerando as medições de cada uma das 14 naceles individualmente, bem como uma análise comparativa com a média simples, a média ponderada e o melhor histórico de velocidade do vento (MHV) utilizado na modelagem atual, detalhado no capítulo de tratamento de dados. No cálculo da média ponderada, as velocidades do vento medidas em cada nacele são ponderadas pela geração verificada em cada ponto de medição. Observa-se uma redução significativa dos erros quando se utiliza a média simples da velocidade do vento de todas as naceles, evidenciando que o aumento do número de medições proporciona maior representatividade da velocidade média do parque eólico e um ganho expressivo na modelagem e na estimativa da geração prevista.

As análises anteriores são importantes para evidenciar que o uso da ferramenta adequada

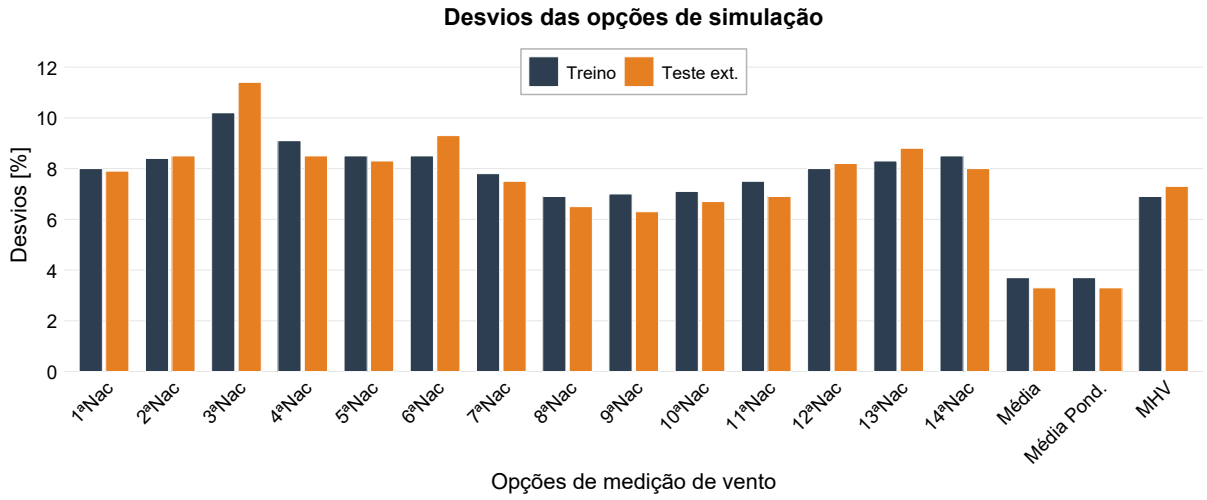
Figura 41 – Comparação curvas V × P por vento médio de 3 e 14 pontos de medição.



Fonte: Elaboração própria.

(seja ela de natureza mais simples, como a regressão logística ou a sigmoide, seja em base metodológica mais complexa, como as RNAs) conduz a resultados similares. Por outro lado, entende-se que a quantidade de observações, quando dispersas ao longo do parque eólico, como a medição em cada nacele, aumenta a acurácia da conversão de vento em potência. No entanto, a aquisição de informações de cada aerogerador torna-se uma tarefa difícil na prática, pois atualmente existem mais de mil parques eólicos conectados ao SIN, com dezenas de aerogeradores em cada um. Dessa forma, o envio de informações de cada aerogerador ao ONS aumentaria exageradamente o volume de dados do SSC do ONS, sendo a única saída viável o envio de uma única medição de vento por usina. Quando essa medição é precedida do cálculo da média das naceles, conforme apresentado nesta seção, traz ganhos expressivos ao processo de

Figura 42 – Comparação dos erros para diferentes formas de cálculo do vento equivalente, nos conjuntos de treinamento e teste.



Fonte: Elaboração própria.

previsão.

4.7 Processos de modelagem das curvas Vento × Potência

As curvas V × P podem ser construídas considerando diferentes níveis de agregação espacial e temporal. A agregação espacial pode ser realizada por usina ou por agrupamentos de usinas, enquanto a agregação temporal pode considerar, por exemplo, diferentes horas do dia ou períodos do ano.

A escolha do grau de agregação depende da aplicação de interesse. A segmentação por intervalos do dia, embora não seja indispensável, permite representar individualmente o comportamento de cada horário e capturar eventuais particularidades intradiárias, proporcionando pequenos ganhos de desempenho quando existem diferenças relevantes ao longo do dia. Nesse contexto, pode-se estimar uma curva específica para cada intervalo temporal adotado. Assim, utilizando-se uma discretização de 30 minutos, obtêm-se 48 curvas distintas para cada dia. Neste trabalho, essa segmentação é adotada tanto no modelo de estimação quanto no de previsão.

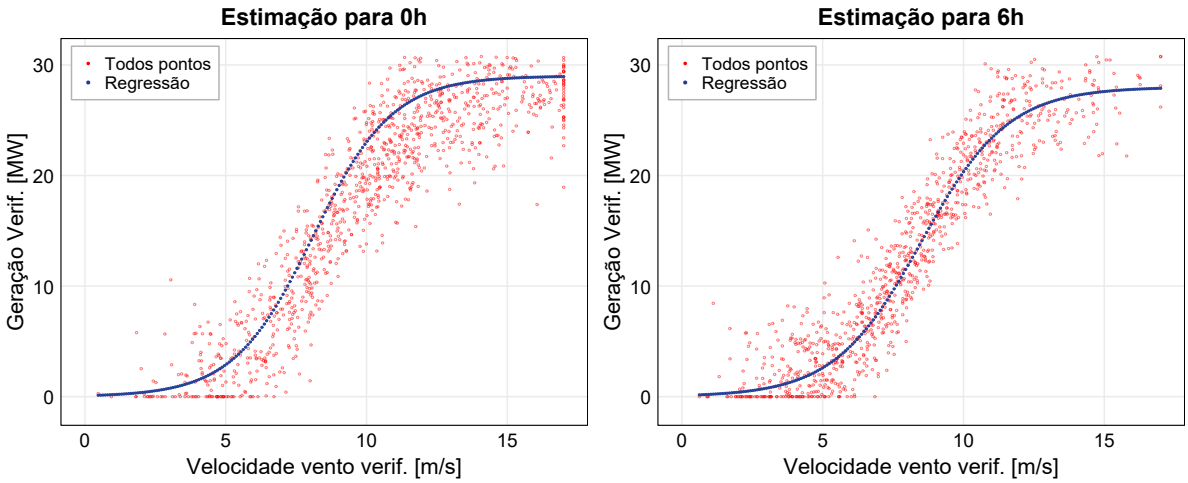
No processo de estimação das curvas, o principal objetivo é fornecer subsídios para a correção dos dados das usinas. Por essa razão, é desejável preservar a representação individual de cada usina, mantendo sua espacialidade. Já no contexto da previsão, o foco passa a ser a estimativa da geração em níveis mais agregados, de maior interesse para o ONS. Nessa situação, torna-se possível trabalhar com agrupamentos de usinas, como será apresentado no Capítulo 5. Essa abordagem traz benefícios adicionais, entre os quais se destaca a redução dos efeitos da variabilidade natural das séries, resultando em curvas mais estáveis e representativas.

As Figuras 43 e 44 apresentam, para uma mesma usina eólica localizada no interior da Bahia, as curvas V × P estimadas em quatro horários representativos do dia: 0h, 6h, 12h

e 18h. Em cada painel, a curva V $\times$ P ajustada é representada em azul, e os pontos em vermelho correspondem aos pares geração–vento verificado naquele horário específico. Ao agregar mentalmente os pontos dos quatro horários, recompõe-se a dispersão V $\times$ P completa da usina ao longo do dia; cada horário, no entanto, "ilumina" apenas uma fração dessa dispersão, em decorrência do perfil característico de vento daquele intervalo.

Observa-se que, às 0h, a usina opera com ventos elevados típicos do período noturno em parques eólicos da região Nordeste, com a geração se concentrando próxima ao limite superior da capacidade da usina; às 6h, o regime de vento se reduz parcialmente, mas a geração ainda permanece em níveis moderado-altos; às 12h, a velocidade do vento atinge seu mínimo diário, com a geração se concentrando em níveis bem baixos, comportamento característico das regiões do interior do semiárido nordestino, onde a forte radiação solar diurna gera uma camada limite atmosférica turbulenta e enfraquece o escoamento médio em altura; já às 18h, o regime se intensifica novamente, retornando aos níveis observados pela manhã.

Figura 43 – Estimação da curva de potência para 0h e 6h, para uma usina do interior da Bahia.

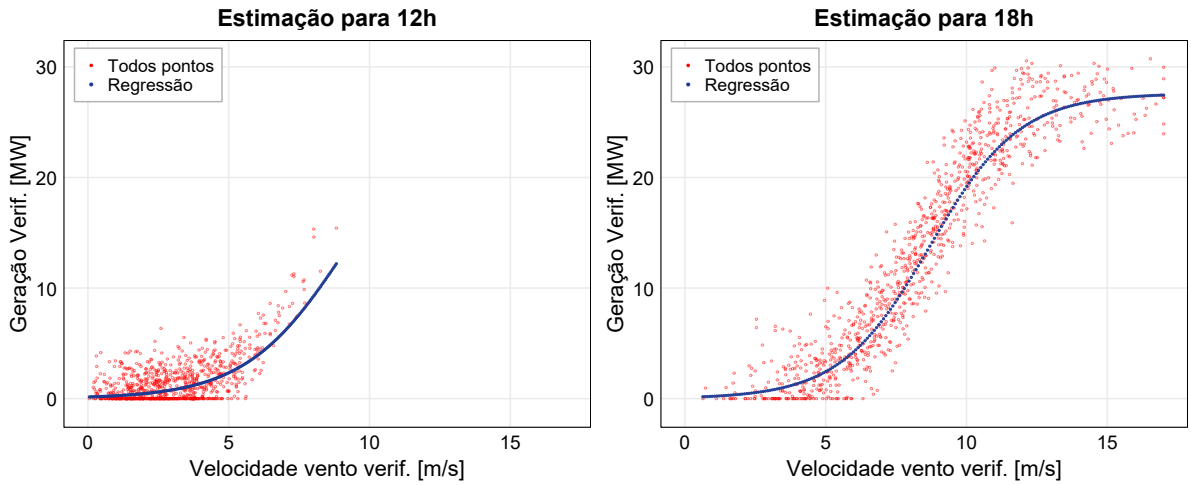


Fonte: Elaboração própria.

Outro aspecto relevante é que, mesmo se tratando da mesma usina, as curvas estimadas para cada horário passam em posições ligeiramente diferentes, mesmo na região central da dispersão. Esse comportamento reflete a influência de variáveis não modeladas explicitamente pela regressão (tais como a direção predominante do vento, a temperatura ambiente, a umidade ou as condições de turbulência), que apresentam padrões distintos ao longo do dia e impactam de forma desigual a relação entre vento e potência em cada horário. Como a curva é reajustada para cada intervalo, esses efeitos são naturalmente incorporados ao modelo, sem a necessidade de modelagem explícita.

Uma vantagem significativa dessa discretização é observada em parques onde a direção do vento muda em determinados horários e os parques eólicos não conseguem aproveitar o recurso com a mesma eficiência, em virtude do que é comumente chamado de efeito esteira

Figura 44 – Estimação da curva de potência para 12h e 18h, para uma usina do interior da Bahia.



Fonte: Elaboração própria.

(*wake effect*). Esse efeito ocorre quando o vento passa por um aerogerador e segue em direção a outro localizado atrás dele, resultando em redução da energia que chega ao segundo aerogerador, além de aumento de turbulência no escoamento entre o primeiro e o segundo.

A variação na relação V $\times$ P em intervalos de meia hora pode estar relacionada a diversos fatores, entre eles a direção do vento. Em regiões litorâneas, é sabido que durante o dia os ventos geralmente sopram do mar em direção à terra, já que o ar se desloca do ponto de pressão mais alta para o ponto de pressão mais baixa. Devido ao calor específico elevado da água, o continente sofre variações de temperatura muito maiores que o mar; assim, durante o dia, o continente está em temperatura mais alta do que o oceano, apresentando menor pressão. À noite, a situação se inverte: a água demora mais tempo para perder calor que a terra, sendo a pressão sobre o oceano mais baixa e fazendo com que o vento se direcione do continente para o mar.

No caso de usinas localizadas no interior, como a do exemplo, esse mecanismo não atua diretamente, mas o ciclo diurno do vento permanece marcante. Durante o dia, o aquecimento intenso da superfície intensifica a convecção e a mistura turbulenta na camada limite atmosférica, o que pode reduzir a velocidade média do vento na altura do *hub* dos aerogeradores. À noite, por outro lado, o resfriamento radiativo da superfície favorece uma atmosfera mais estável, com escoamento menos turbulento e, em muitos casos, ventos mais intensos na altura do rotor. Esse comportamento explica, em boa medida, o padrão observado nas Figuras 43 e 44, em que se verifica a redução acentuada da geração ao meio-dia e a recuperação no fim da tarde e durante a noite.

A forma mais natural, portanto, para se utilizar a curva de potência seria ajustá-la com todo o histórico de vento e geração verificada e, assim, realizar a previsão de posse do vento previsto. Todavia, a utilização do histórico completo para estimação dos parâmetros da regressão logística, além de não ser necessária para um bom ajuste, pode capturar características

indesejadas. Todo equipamento possui uma vida útil e, geralmente, vai perdendo rendimento com o passar dos anos, o que pode alterar as características da curva V $\times$ P. Além disso, outros fatores contribuem para mudanças na curva de potência, como equipamentos em falhas, substituição desses por outros com características levemente diferentes, e ainda ocorrências de curto prazo, como quedas de linhas e torres dos aerogeradores, que podem fazer com que as gerações tomem valores diferentes do esperado. Alguns desses efeitos, se conhecidos pelo usuário do previsor eólico, podem ser mitigados; entretanto, se não conhecidos, podem causar grandes desvios na previsão.

Com o objetivo de mitigar esses efeitos negativos, optou-se por estimar os parâmetros das curvas V $\times$ P utilizando janelas históricas distintas conforme a finalidade do modelo. Para os modelos de previsão, adotou-se uma janela de seis meses de dados, enquanto, para os modelos de estimação, considerou-se uma janela de um ano, em atendimento às regras estabelecidas na Resolução Normativa ANEEL nº 1.073 [154] para apuração de eventos de *constrained-off*.

A utilização de janelas históricas mais recentes reduz os desvios associados a fenômenos ocorridos em períodos mais distantes, que podem não estar presentes ou não ser representativos das condições observadas no momento da previsão. Dessa forma, evita-se que eventos passados, capazes de distorcer a relação V $\times$ P em determinado intervalo, comprometam o desempenho do modelo em sua aplicação atual.

Embora, à primeira vista, o número de pontos utilizados para representar o equacionamento possa parecer insuficiente, esses dados tendem a refletir melhor o comportamento recente da usina. Assim, caso tenha ocorrido algum fenômeno atípico em um período distante no tempo, seu efeito sobre a curva ajustada é reduzido, preservando a aderência da estimativa às condições operativas mais atuais.

Foram realizados testes com o objetivo de verificar se a adoção dessas janelas históricas comprometeria a qualidade das curvas ajustadas. Os resultados indicaram que não houve aumento dos desvios; ao contrário, foram observados ganhos de desempenho, conforme esperado.

4.8 Considerações finais do capítulo

Este capítulo apresentou o modelo de estimação da geração eólica a partir da construção de curvas de Vento $\times$ Potência (V $\times$ P), abordando desde o tratamento da dispersão V $\times$ P e a definição automática de curvas de fronteira até o ajuste dinâmico das curvas por meio de funções logísticas estimadas pelo algoritmo de Gauss–Newton. Discutiu-se, ainda, a inclusão de pontos fictícios para estabilizar o ajuste na região de saturação, bem como a reestimação periódica das curvas, de modo a acompanhar a evolução temporal do comportamento operacional dos parques. A estimação foi conduzida de forma individualizada por usina, sendo os resultados também avaliados de forma agregada por conjuntos de parques.

Os conceitos e desenvolvimentos apresentados neste capítulo também compuseram

um trabalho publicado em [31], no qual a metodologia de estimação da geração potencial foi aplicada em larga escala ao sistema brasileiro, com ênfase na quantificação das perdas de energia associadas a eventos de corte de geração, ou *curtailment*. Nesse trabalho, demonstrou-se que a estimação aplicada a conjuntos de parques apresenta elevada robustez e acurácia, e que a metodologia alcança altos níveis de aderência na reconstrução da geração em períodos sem restrição operativa.

O modelo de estimação descrito neste capítulo teve origem no trabalho apresentado em [41], no qual foi proposta a estimação da geração a partir de dados de vento observado e de curvas Vento $\times$ Potência, estabelecendo as bases da abordagem posteriormente aprofundada nesta tese e em [31]. Adicionalmente, em [35], foi investigada a incorporação da disponibilidade de potência como elemento de correção das curvas V $\times$ P, permitindo representar de forma mais adequada os períodos em que a usina opera com capacidade reduzida.

Em comparação com esses trabalhos anteriores, de caráter mais sucinto, a abordagem desenvolvida nesta tese apresenta um detalhamento mais amplo do modelo de estimação, incluindo sua formulação matemática, hipóteses, limitações e critérios de aplicação. Além disso, aprofunda-se a integração da etapa de estimação com o tratamento dos dados observados e com os modelos de previsão, consolidando a metodologia em um arcabouço mais completo para reconstrução e previsão da geração eólica.

5 MODELO PARA PREVISÃO DE GERAÇÃO EÓLICA

Este capítulo destina-se a apresentar as metodologias propostas para a realização das previsões, que se estendem desde a operação em tempo real até o planejamento de curtíssimo e curto prazo, e as seções a seguir detalharão as características principais dos modelos para cada horizonte temporal.

5.1 Considerações iniciais

Conforme já explanado neste documento, as previsões se diferenciam conforme o horizonte de tempo em que são feitas. Em se tratando especificamente dos seus usos na cadeia de processos da operação do SIN a situação não é diferente, a discretização do horizonte de tempo também se faz necessária. A operação em tempo real, por exemplo, demanda modelos capazes de prever a geração eólica para os próximos minutos até algumas horas à frente, isso é importante para as tomadas de decisão dos operadores na sala de controle. Para o horizonte de curtíssimo prazo, que se estende do dia seguinte até 7 dias à frente, essas previsões são insumos para os estudos elétricos, programação diária da operação eletroenergética e para o Modelo de Planejamento de Curtíssimo Prazo (MPCTP), atualmente o DESSEM (Modelo de Despacho Hidrotérmico de Curto/Curtíssimo Prazo).

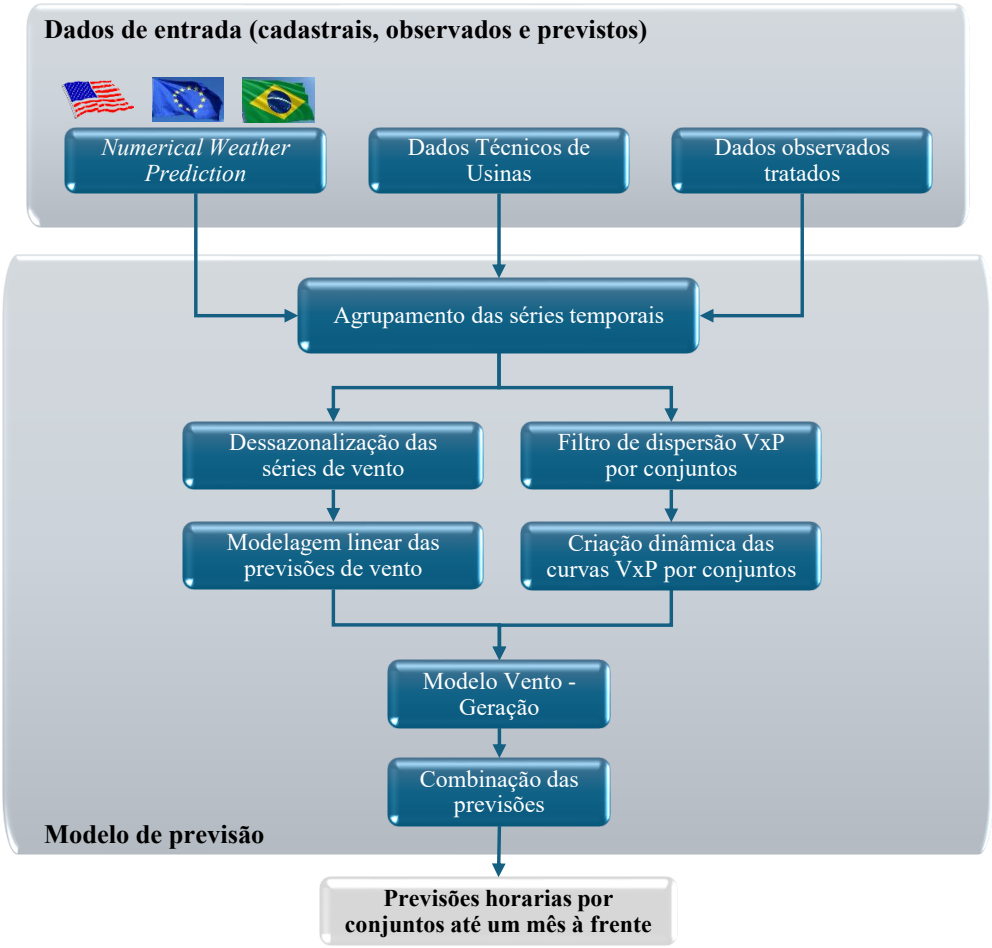
Já no horizonte de curto prazo, até um mês à frente, é de suma importância previsões semanais para o planejamento de curto prazo, conduzido com o Modelo de Planejamento de Curto Prazo (MPCP), atualmente o DECOMP (Modelo de Determinação da Coordenação da Operação a Curto Prazo). Para os horizontes mais extensos, como médio e longo prazo, que se estendem até 5 anos à frente, cujo planejamento é realizado com o Modelo de Planejamento de Médio Prazo (MPMP), atualmente o NEWAVE (Modelo de Planejamento da Operação de Sistemas Hidrotérmicos Interligados de Longo e Médio Prazo), a geração eólica é representada ainda deterministicamente, porém já se vislumbra a aplicação de cenários para essa fonte, assim como já feito para a fonte hídrica.

Para embasar e direcionar a metodologia proposta, foram implementados e testados diversos modelos para um grupo de usinas eólicas do SIN, incluindo métodos clássicos como *Box–Jenkins* [84], modelos de regressão e abordagens baseadas em inteligência artificial, como *ANN (Artificial Neural Network)* [86]. Esse processo de investigação, detalhado em [42], envolveu análises comparativas entre métodos de *machine learning*, séries temporais da classe *Box–Jenkins* e uma versão preliminar do modelo de previsão apresentado neste trabalho, tendo evidenciado que, para o horizonte considerado, os modelos numéricos de tempo são os que possuem maior poder preditivo, o que orientou a abordagem metodológica a privilegiar esses insumos como os principais.

5.2 Estrutura do modelo proposto

A Figura 45 resume a estrutura do modelo de previsão de geração eólica desenvolvido neste trabalho.

Figura 45 – Estrutura dos modelos desenvolvidos neste trabalho.



Fonte: Elaboração própria.

O modelo recebe três categorias de dados de entrada: (i) previsões de velocidade do vento provenientes de modelos de Previsão Numérica do Tempo (*Numerical Weather Prediction* – NWP), oriundos de diferentes centros meteorológicos internacionais e nacionais; (ii) dados técnicos cadastrais das usinas eólicas; e (iii) dados observados tratados, resultantes do processo de tratamento de dados históricos descrito na seção anterior.

A Figura 45 organiza as etapas de processamento que conduzem à previsão final de geração. Na primeira etapa, as séries temporais de vento previsto e observado, juntamente com os dados técnicos das usinas, são submetidas a um processo de agrupamento, no qual as usinas são organizadas em conjuntos, também denominados pontos ou agrupamentos. Atualmente, são considerados 76 agrupamentos, que representam, em conjunto, mais de mil usinas. A forma de

definição desses pontos, conjuntos ou agrupamentos será detalhada adiante neste capítulo. A partir desse agrupamento inicial, o fluxo de processamento se divide em dois ramos paralelos e complementares, responsáveis por fornecer as informações necessárias à construção da previsão final de geração.

No primeiro ramo, as séries de vento são dessazonalizadas para remoção de padrões sazonais sistemáticos, e em seguida é realizada uma modelagem linear das previsões de vento dos modelos NWP em relação ao vento observado, com o objetivo de corrigir vieses e ajustar as previsões meteorológicas às condições locais de cada conjunto de usinas.

No segundo ramo, é aplicado um filtro de dispersão V $\times$ P por conjuntos, seguido da criação dinâmica das curvas de potência em função da velocidade do vento (curvas V $\times$ P) para cada conjunto. Essas curvas representam a relação não linear entre a velocidade do vento corrigida e a geração elétrica esperada.

Os resultados de ambos os ramos convergem no modelo de conversão de vento em potência, que converte as previsões de vento ajustadas em estimativas de geração elétrica por meio das curvas V $\times$ P calibradas. Na etapa final, é realizada a combinação das previsões provenientes dos diferentes modelos NWP, gerando como produto final previsões horárias de geração eólica por conjuntos de usinas, com horizonte de até um mês à frente. As seções a seguir detalham cada uma das etapas descritas.

5.3 Modelagem das previsões numéricas de vento

As previsões de vento advindas dos NWPs ou comumente chamados modelos meteorológicos, em geral, apresentam desvios (erros) em relação ao vento que de fato ocorre. Observa-se, no entanto, comportamento de desvio típico, ou seja, a previsão de vento apresenta em quase todas as usinas eólicas um viés em relação ao vento verificado, necessitando de tratamentos específicos antes de serem utilizadas nos modelos de previsão.

5.3.1 Modelos Numéricos de Tempo avaliados

A realização de previsões com horizonte de até um mês requer a utilização de modelos de previsão numérica capazes de fornecer informações meteorológicas nesse intervalo temporal. Neste trabalho, foram avaliados os seguintes modelos:

- IFS (*Integrated Forecasting System*) $-$ HRES (*High Resolution*) do ECMWF (*European Centre for Medium-Range Weather Forecasts*), que possui horizonte de 10 dias e resolução espacial de $\sim$ 10 km, com resolução temporal de 1 hora nos $\sim$ 4 primeiros dias, de 3 horas nos $\sim$ 2 dias seguintes e de 6 horas nos demais;
- IFS $-$ ENS (*Ensemble*) do ECMWF, que possui horizonte de 8 dias e resoluções espacial e temporal de $\sim$ 20 km e 3 horas, respectivamente;

- GFS (*Global Forecast System*) do NCEP (*National Centers for Environmental Prediction*), com resolução espacial de 25 km e temporal de 1 hora nos primeiros 5 dias de previsão e de 3 horas no restante do horizonte (até 15 dias à frente);
- GEFS (*Global Ensemble Forecast System*) do NCEP, com resolução espacial de $\sim$ 50 km e resolução temporal de 3 horas nos primeiros 10 dias de previsão e de 6 horas no restante do horizonte, estendendo-se por 35 dias de previsão (até 34 dias à frente);
- WRF do CPTEC (*Centro de Previsão do Tempo e Estudos Climáticos*), que possui horizonte de 3 dias e resoluções espacial e temporal de 5 km e 1 hora, respectivamente;
- BAM do CPTEC, que possui horizonte de 10 dias e resoluções espacial e temporal de 18 km e 1 hora, respectivamente.

Considerando o horizonte de previsão de interesse, nenhum dos modelos meteorológicos avaliados é capaz de atendê-lo integralmente de forma isolada. Dessa forma, as previsões serão geradas a partir da combinação de diferentes insumos, explorando suas características complementares para garantir a cobertura de todo o horizonte analisado.

A Figura 46 sintetiza de forma comparativa os modelos NWPs disponíveis considerados. Nela, destacam-se o horizonte de previsão (comprimento da barra), a resolução temporal (segmentos coloridos da barra) e a resolução espacial (círculo à direita da barra, com raio proporcional).

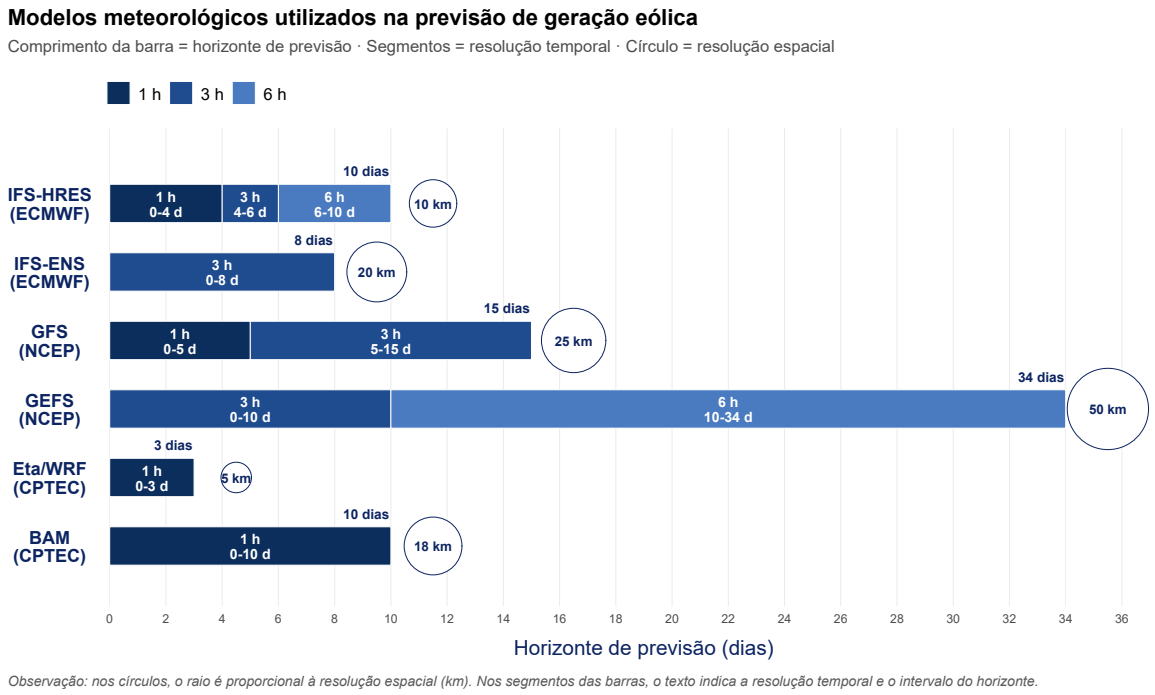
Existe também um modelo do ECMWF estendido, indo até 45 dias à frente, porém ainda não é possível utilizá-lo, porque não se dispõe de históricos longos o suficiente para avaliar o desempenho dessas previsões até o momento.

5.3.2 Avaliação dos insumos meteorológicos utilizados pelo modelo

Nesta seção, tem-se uma caracterização preliminar dos dados meteorológicos a serem utilizados pelo modelo proposto, com foco na identificação do viés associado às previsões de vento. Essa caracterização tem por objetivo motivar e justificar a metodologia de correção apresentada na sequência.

Os modelos numéricos de tempo geram previsões por pontos de grade, ou seja, para os modelos meteorológicos globais, o globo é representado por quadrículas. O espaçamento entre as quadrículas é a resolução espacial do modelo. Sendo assim, é necessário associar cada usina à sua respectiva quadrícula. Essa associação é feita com base na distância entre a usina e os centroides das quadrículas, sendo utilizado o centroide de menor distância. Duas métricas foram consideradas para o cálculo dessa distância: a distância euclidiana e a fórmula de Haversine.

Figura 46 – Modelos meteorológicos utilizados na previsão de geração eólica, comparando horizonte de previsão e resoluções temporal e espacial.



Fonte: Elaboração própria.

A distância euclidiana, calculada diretamente sobre as coordenadas geográficas, é dada por:

$$\text{dist}_{euc} = \sqrt{(\text{lat}_{quad} - \text{lat}_{usi})^2 + (\text{long}_{quad} - \text{long}_{usi})^2} \tag{5.1}$$

em que dist_{euc} representa a distância euclidiana; lat_{quad} e lat_{usi} são a latitude da quadrícula e da usina, respectivamente; e long_{quad} e long_{usi} são a longitude da quadrícula e da usina, respectivamente.

A fórmula de Haversine, por sua vez, considera a curvatura da Terra e fornece a distância sobre a superfície esférica entre dois pontos, sendo expressa por:

$$\text{dist}_{hav} = 2R \arcsin\left(\sqrt{\sin^2\left(\frac{\Delta\text{lat}}{2}\right) + \cos(\text{lat}_{usi})\cos\left(\text{lat}_{quad}\right)\sin^2\left(\frac{\Delta\text{long}}{2}\right)}\right) \tag{5.2}$$

em que dist_{hav} é a distância geodésica em quilômetros; $R = 6371$ km é o raio médio da Terra; $\Delta\text{lat} = \text{lat}_{quad} - \text{lat}_{usi}$ e $\Delta\text{long} = \text{long}_{quad} - \text{long}_{usi}$ são as diferenças de latitude e longitude em radianos, respectivamente.

Existem usinas litorâneas cujos centroides dos modelos de vento previsto encontram-se no mar. Nesses casos, são utilizados os centroides vizinhos em terra mais próximos, por representarem efeitos mais coerentes com a realidade física dos ventos no parque. Após a

associação das usinas às respectivas quadrículas, as previsões são agrupadas em pontos de análise por meio da média simples das usinas que os compõem.

Além da associação horizontal da usina à quadrícula do modelo, destaca-se que a variável meteorológica utilizada neste trabalho é a velocidade do vento prevista a 100 metros de altura para todos os modelos considerados. As medições de vento empregadas como referência correspondem às alturas efetivas de instalação dos sensores em cada usina, os quais podem estar localizados em naceles ou torres e, no caso dos dados do ONS, não seguem um padrão único de instalação. Em geral, contudo, esses sensores encontram-se a 100 metros de altura ou em níveis próximos. Neste trabalho, não é realizada uma extrapolação vertical explícita da velocidade do vento, por meio, por exemplo, da lei de potência ou do perfil logarítmico. Em vez disso, a diferença entre as alturas de previsão e medição é tratada de forma implícita pelo modelo de correção das previsões de vento. Como as alturas de medição tendem a ser próximas de 100 metros, testes preliminares, realizados para um grupo de usinas e para alguns horizontes de previsão, indicaram que uma correção vertical explícita e não linear não produz resultados significativamente distintos daqueles obtidos pela correção linear adotada, que trata conjuntamente todos os fatores responsáveis pelos desvios observados. Como limitação, essa abordagem depende do histórico de medições disponível para cada usina e não permite dissociar o efeito das diferenças de altura dos demais fatores que contribuem para o viés das previsões.

5.3.2.1 Caracterização do erro de previsão

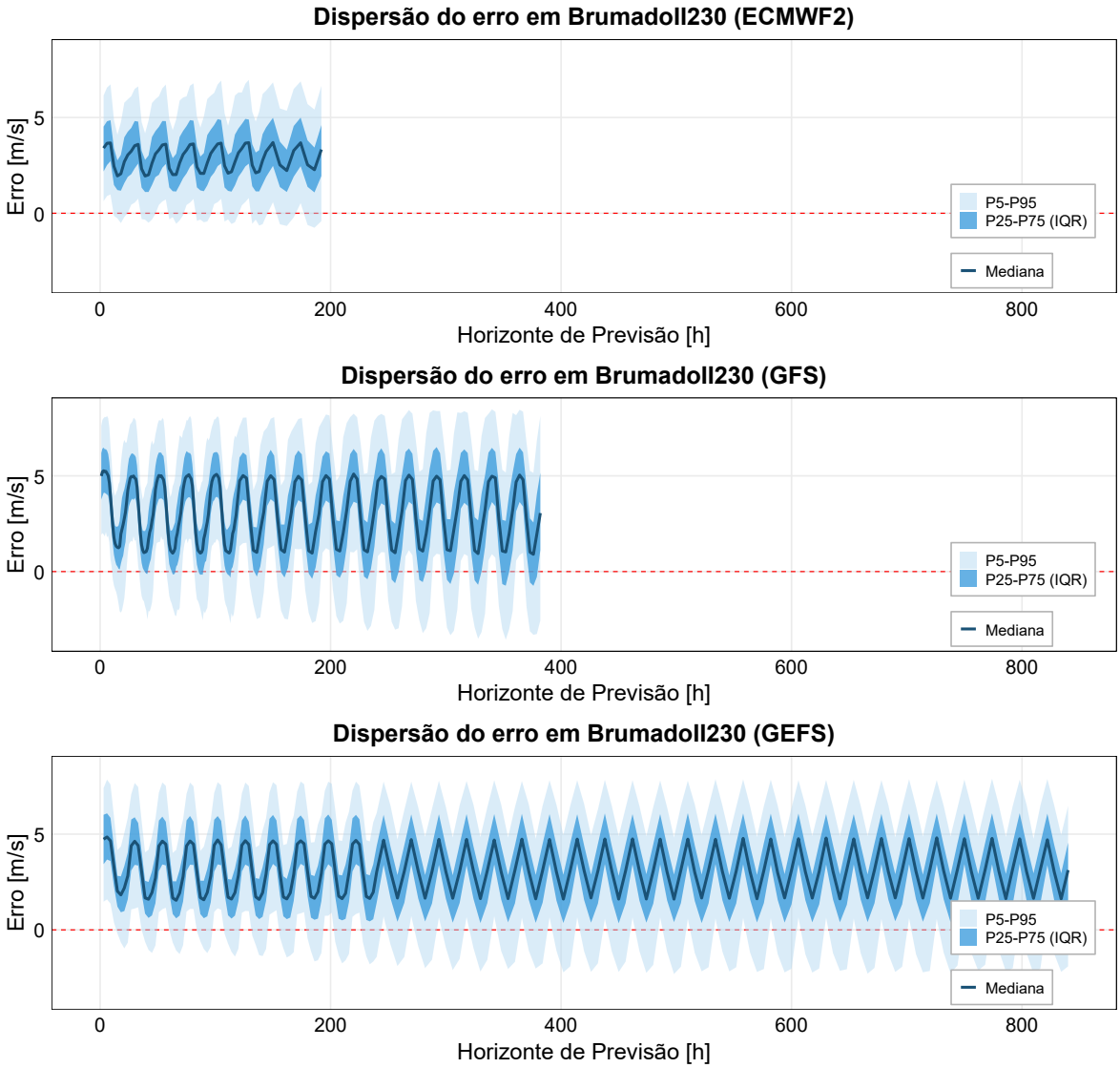
Para investigar o comportamento do erro de previsão de vento e identificar a presença de viés sistemático, foram realizadas análises em três escalas espaciais complementares: ponto individual (com detalhamento ao longo do horizonte de previsão), conjunto consolidado de todos os pontos de análise e agregado por estado.

5.3.2.1.1 Análise pontual ao longo do horizonte.

As Figuras 47 e 48 apresentam a dispersão do erro ao longo do horizonte de previsão para dois pontos selecionados como representativos de comportamentos distintos: Brumado II 230 (interior da Bahia) e Açu II 230 (litoral do Rio Grande do Norte). Em cada figura, os três painéis correspondem aos modelos ECMWF, GFS e GEFS, e mostram a mediana do erro (linha azul-escura), o intervalo interquartil P25–P75 (banda azul intermediária) e o intervalo P5–P95 (banda azul clara), calculados por horizonte de previsão. Cabe esclarecer que, nas figuras e nas análises seguintes, "ECMWF" designa o produto *ensemble* do IFS (IFS-ENS), identificado como ECMWF2 nas figuras.

A comparação entre os dois pontos evidencia que o viés da previsão pode ser tanto positivo quanto negativo, dependendo da localização e do modelo. No ponto Brumado II 230, observa-se uma mediana do erro persistentemente positiva em todos os modelos e em praticamente todos os horizontes, indicando subestimação sistemática do vento previsto em relação ao

Figura 47 – Dispersão do erro (verificado − previsto) ao longo do horizonte de previsão — ponto Brumado II 230.

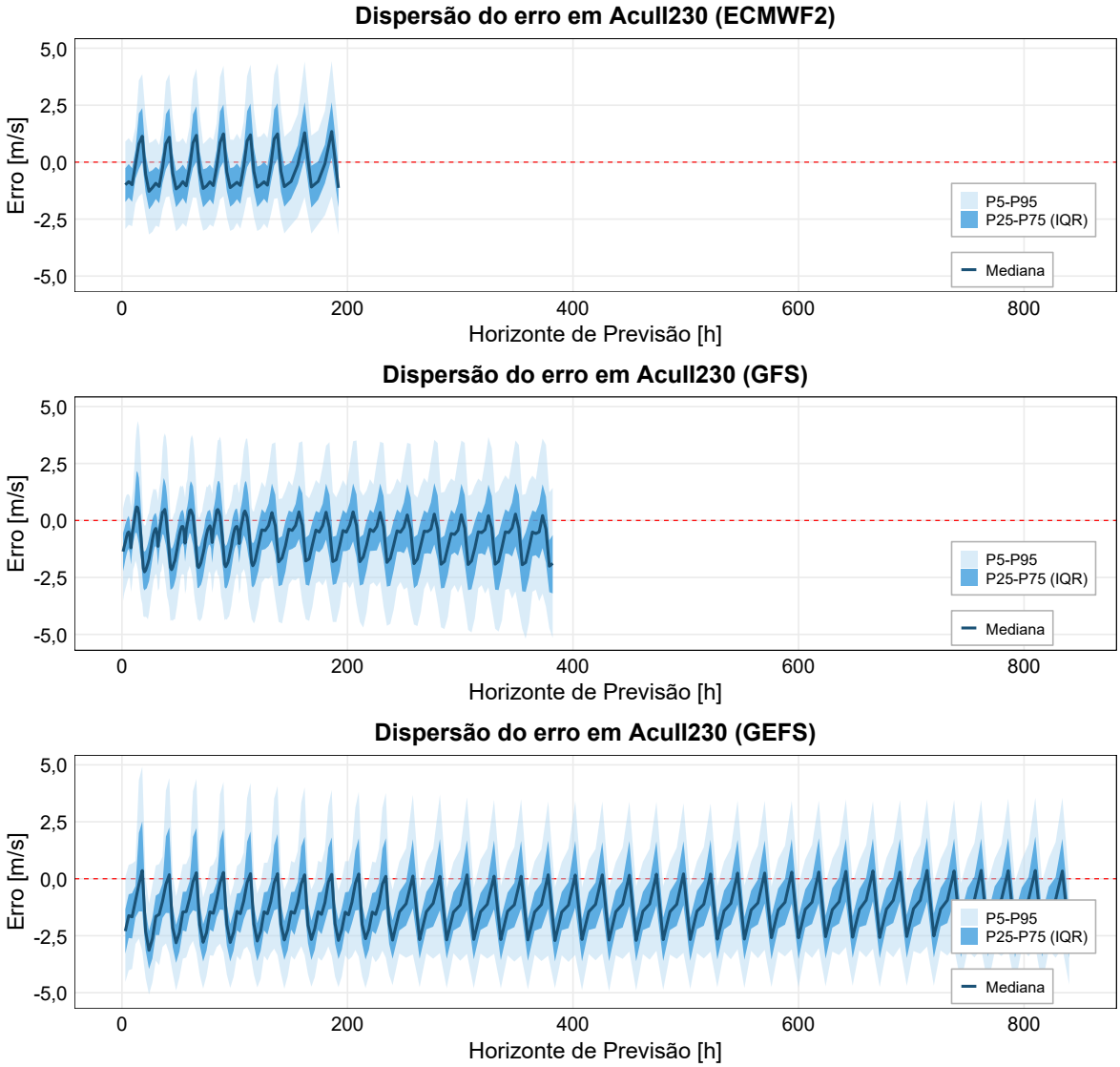


Fonte: Elaboração própria.

verificado. Já no ponto Açu II 230, o comportamento é distinto: a mediana é predominantemente negativa nos três modelos, com oscilação diária, mas com magnitudes do viés inferiores ao exemplo de Brumado II 230, indicando que, nesse ponto, as previsões dos modelos são mais aderentes ao vento verificado e apresentam leve superestimação.

Outro aspecto relevante é a presença de oscilações regulares ao longo do horizonte de previsão, com período aproximado de 24 horas. Esse padrão revela que o viés não é constante ao longo do dia: há horários em que a previsão tende a superestimar e outros em que tende a subestimar o vento. Esse comportamento está associado ao ciclo diurno do vento e indica que os modelos meteorológicos representam de forma imperfeita a variabilidade horária do escoamento atmosférico, como os efeitos de aquecimento e resfriamento da camada limite ao longo do dia, entre outros processos discutidos adiante. Tal padrão se mantém perceptível mesmo em

Figura 48 – Dispersão do erro (verificado − previsto) ao longo do horizonte de previsão — ponto Açu II 230.



Fonte: Elaboração própria.

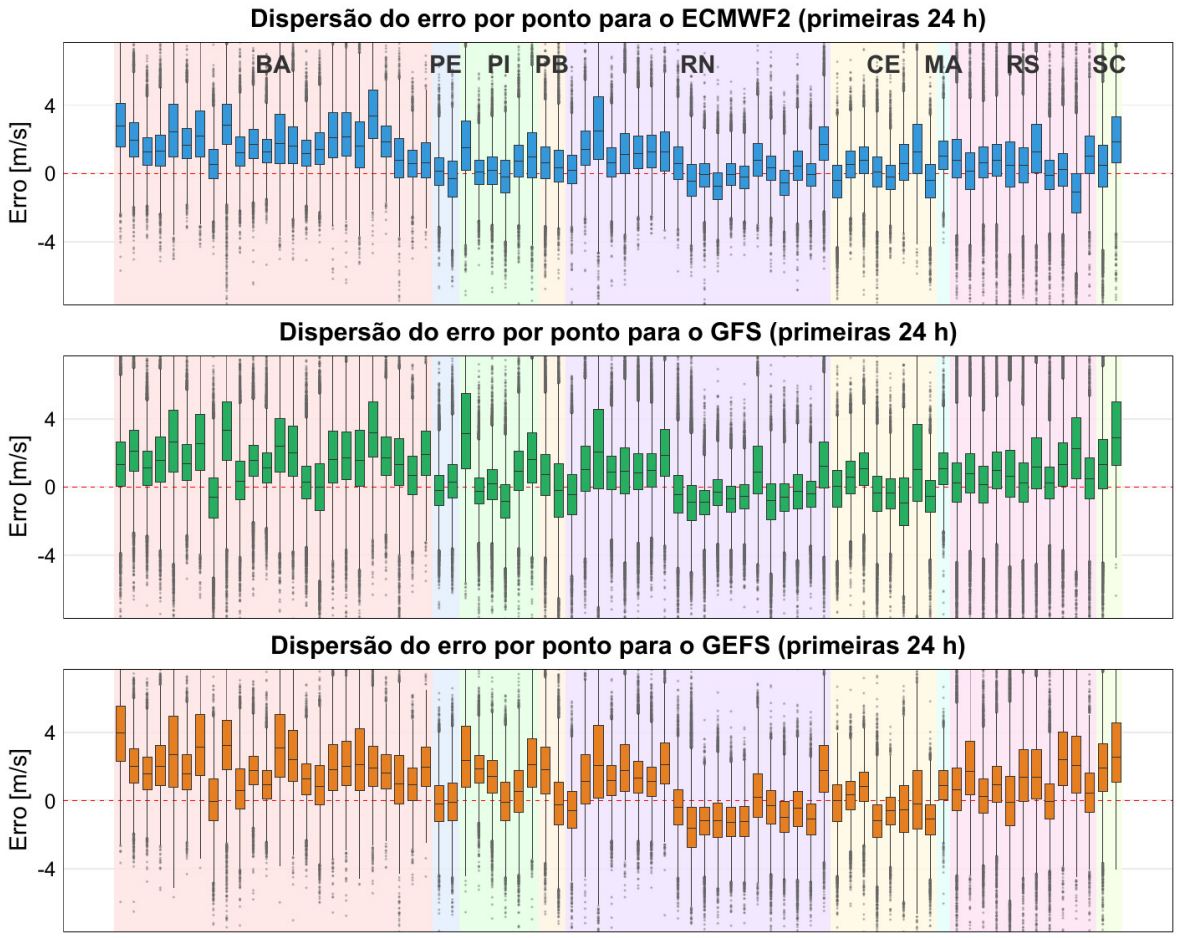
horizontes mais longos, embora a dispersão (largura das bandas) aumente, refletindo a perda gradual de previsibilidade.

5.3.2.1.2 Análise consolidada por ponto.

Os exemplos individuais apresentados anteriormente caracterizam o comportamento do erro em pontos específicos, escolhidos por exibirem padrões opostos. Para verificar se esses comportamentos são pontuais ou se refletem padrões mais amplos, a Figura 49 consolida a distribuição do erro de previsão do vento para a totalidade dos 76 pontos de análise (compreendendo mais de mil usinas), considerando os três modelos meteorológicos e o horizonte das primeiras 24 horas. Cada coluna do gráfico corresponde a um ponto e exibe um *boxplot* dos erros observados naquele ponto; os pontos são agrupados por estado, com faixas de fundo coloridas

marcando o limite de cada agrupamento e as siglas dos estados indicadas no topo do painel superior.

Figura 49 – Dispersão do erro (verificado − previsto) por ponto de análise para os três modelos meteorológicos — primeiras 24 horas de previsão.



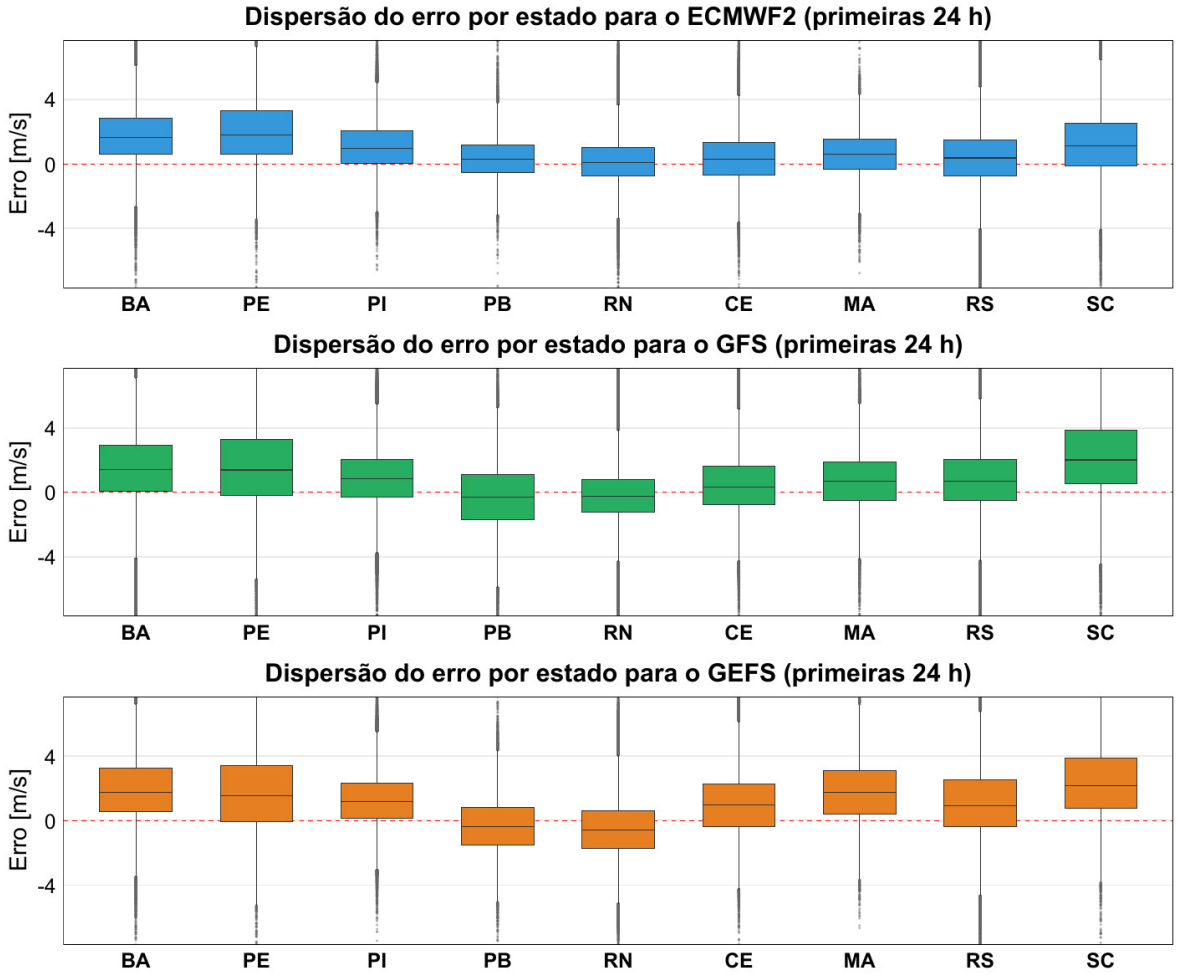
Fonte: Elaboração própria.

A figura permite identificar padrões espaciais bastante claros. Na Bahia (BA), a mediana dos erros é positiva em praticamente todos os pontos e em todos os modelos, confirmando que a subestimação observada em Brumado II 230 não é um caso isolado, mas sim um comportamento sistemático ao longo de todo o estado. No Rio Grande do Norte (RN), por sua vez, as medianas se aproximam mais de zero, com alguns pontos apresentando viés expressivo (coerente com o que se observou em Açu II 230). Os estados do Sul (RS e SC) exibem dispersões mais amplas e medianas predominantemente positivas, sobretudo em SC. Comparando os três modelos, percebe-se que os padrões observados em cada estado se preservam, isto é, estados com viés positivo em um modelo tendem a apresentar viés positivo nos demais, embora a magnitude do erro varie entre eles.

5.3.2.1.3 Análise consolidada por estado.

Para uma visão ainda mais agregada, a Figura 50 apresenta a distribuição dos erros agregados por estado, considerando-se todos os pontos pertencentes a cada estado e todos os horizontes até 24 horas. Cada *boxplot* representa, portanto, o erro consolidado do conjunto de usinas do estado.

Figura 50 – Dispersão do erro (verificado − previsto) agregada por estado para os três modelos meteorológicos — primeiras 24 horas de previsão.



Fonte: Elaboração própria.

Confirma-se nessa escala o que já havia sido sugerido pelas análises pontuais: o desempenho dos modelos varia significativamente conforme o estado. O viés positivo (subestimação) é mais marcante em BA, PE e SC; PI, CE, MA e RS apresentam viés levemente positivo; e PB e RN têm medianas mais próximas de zero, com leve tendência de superestimação. Observa-se também que a amplitude dos *boxplots* (P25–P75 e *whiskers*) varia entre estados, indicando que a previsibilidade do vento não é homogênea no território, algumas regiões apresentam erros mais concentrados, enquanto outras exibem maior dispersão. A persistência desses padrões no agregado estadual indica que o viés é de origem estrutural nos modelos meteorológicos, e não um artefato de pontos individuais.

5.3.2.1.4 Síntese e motivação da metodologia.

O conjunto de análises apresentadas evidencia que o erro de previsão dos modelos meteorológicos avaliados possui três características fundamentais: (i) varia significativamente entre pontos e estados, com regiões apresentando subestimação sistemática e outras apresentando boa aderência ao vento verificado; (ii) varia ao longo do dia, com oscilações regulares de período diário associadas ao ciclo diurno do vento; e (iii) varia entre os modelos meteorológicos, ainda que os padrões espaciais sejam qualitativamente semelhantes entre eles.

Cabe ressaltar que a dinâmica da camada limite atmosférica é apenas uma das possíveis fontes dos erros observados nas previsões de vento dos modelos meteorológicos. A topografia complexa e a heterogeneidade da superfície, por exemplo, podem gerar circulações locais de meso a microescala, em distâncias de dezenas de quilômetros, que dificilmente são resolvidas por modelos globais com resolução espacial entre 10 e 50 km. Além disso, processos de escala inferior à da grade, como a turbulência e a convecção, são representados por parametrizações, o que implica erros inerentes em relação às observações; as condições iniciais e os dados de entrada dos modelos também são fontes de incerteza. Os modelos numéricos tendem, ainda, a ser subdispersivos, isto é, a apresentar menor variabilidade do que a observada. A metodologia de correção adotada neste trabalho não busca identificar individualmente cada uma dessas causas, mas corrigir estatisticamente a componente sistemática do erro, por horário do dia e por horizonte de previsão.

Diante dessas características dos erros dos modelos meteorológicos, justifica-se a adoção de uma metodologia de correção das previsões de vento aplicada de forma individualizada por intervalo do dia e por ponto de análise, conforme descrita a seguir.

5.3.3 Metodologia utilizada para ajuste das previsões de vento

5.3.3.1 Dessazonalização das séries de vento

A análise dos dados realizada na seção anterior evidenciou que as previsões produzidas pelos modelos meteorológicos apresentam características particulares, que devem ser consideradas no processo de modelagem. Além das oscilações associadas ao ciclo diurno, é necessário levar em conta também a sazonalidade anual das séries de vento.

A Figura 51 apresenta, para o modelo GEFS e ao longo de aproximadamente cinco anos de dados, as séries temporais de vento previsto e de vento verificado no passo 90 do horizonte de previsão. Como o GEFS opera com resolução temporal de 3 horas nos primeiros 10 dias e de 6 horas no restante do horizonte, o passo 90 corresponde a 300 horas à frente do instante de emissão da previsão, ou seja, aproximadamente 12,5 dias à frente. Cada ponto das curvas representa, portanto, uma previsão emitida para a mesma hora do dia e para o mesmo horizonte, o que permite comparar a evolução das duas séries sem misturar efeitos de ciclo diário ou de diferentes horizontes de previsão. Conforme observado no painel superior da figura, o

comportamento e a magnitude das séries podem variar de forma significativa ao longo do ano. Essa variação, entretanto, não ocorre necessariamente da mesma forma para o vento previsto e para o vento verificado, que podem apresentar padrões sazonais distintos, com diferenças tanto na amplitude quanto na fase da sazonalidade anual.

Sendo assim, uma metodologia de ajuste das previsões baseada apenas em regressão linear sobre as séries originais não atende plenamente, visto que o viés do modelo meteorológico pode mudar no tempo. Para contornar essa dificuldade, buscou-se modelar a sazonalidade das séries de vento verificado (curva vermelha) e previsto (curva azul) através do ajuste de senoides (harmônico anual) [69]. Os ajustes obtidos para o vento verificado e para o previsto estão representados pelas curvas laranja e ciano, respectivamente. O painel inferior da Figura 51 mostra as curvas que representam as diferenças entre as séries originais de vento verificado e previsto em relação às suas respectivas senoides, isto é, as séries de vento após a remoção da sazonalidade, as quais seguem o equacionamento da Equação (5.3).

$$V_{iss} = V_i - (A_i \cdot \sin(\omega \cdot t + \phi_i) + B_i) \tag{5.3}$$

Em que V_{iss} é a velocidade de vento sem sazonalidade; V_i a velocidade de vento bruta; i um índice que, quando representado pela letra p, se trata de vento previsto e, pelo v, de vento verificado; ω representa a frequência angular do ciclo anual, em rad/dia, com $\omega = 2\pi/365$ (período de 365 dias); t o tempo em dias; A_i a amplitude; ϕ_i a fase; B_i o termo constante.

Após a remoção da sazonalidade, o modelo consegue representar adequadamente as séries de vento, de forma padronizada entre os dados verificados e previstos. Entretanto, essa padronização, em termos de sazonalidade, representa apenas os comportamentos divergentes entre as séries na frequência anual. Caso ainda exista alguma componente de curto prazo, essa componente é modelada pela regressão da Equação (5.4).

$$V_{ass} = \alpha_a \cdot V_{pss} + \beta_a \tag{5.4}$$

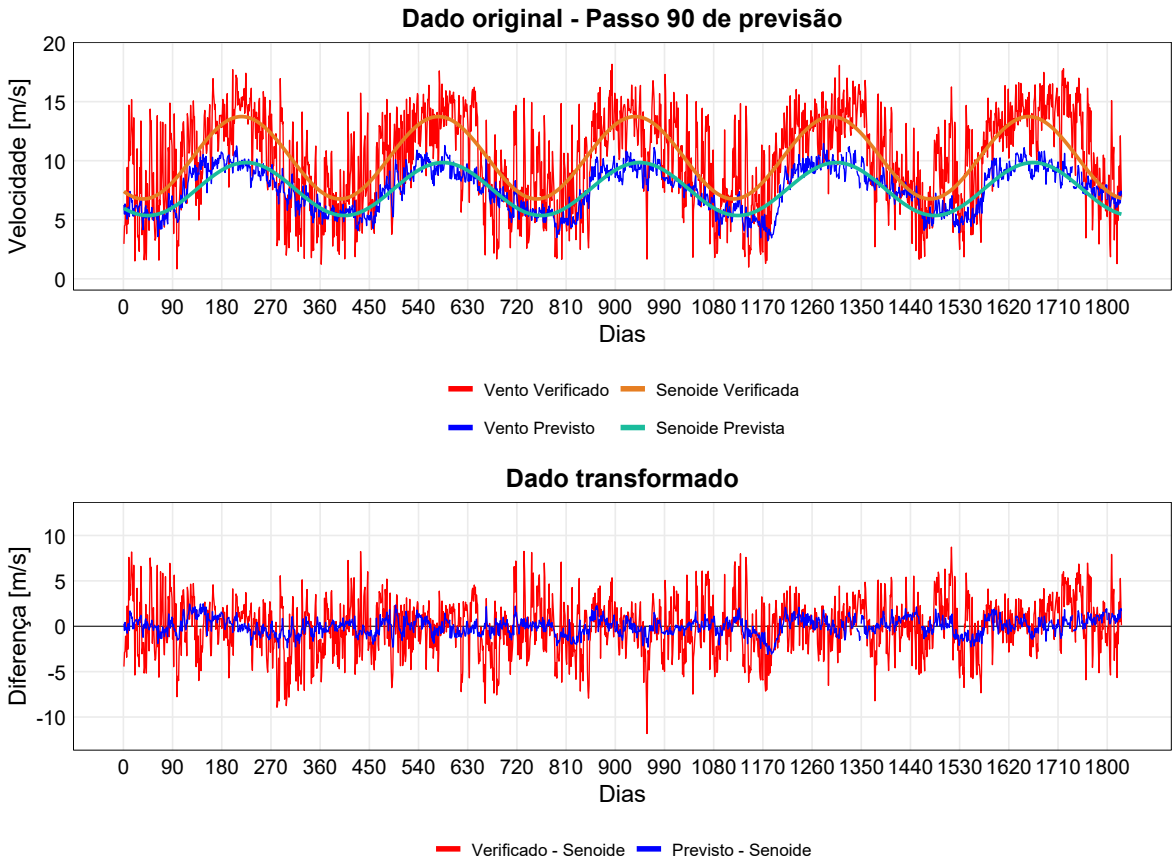
Em que V_{ass} é a velocidade de vento ajustada sem sazonalidade; V_{pss} a velocidade de vento prevista sem sazonalidade; α_a e β_a os coeficientes ajustados a partir da relação linear entre as parcelas sem sazonalidade do vento previsto e verificado, V_{pss} e V_{vss}.

Expandindo-se a Equação (5.4) utilizando a Equação (5.3) e retornando com a parcela da sazonalidade dos dados de vento verificado, tem-se a Equação (5.5).

$$V_a = \alpha_a \cdot (V_p - (A_p \cdot \sin(\omega t + \phi_p) + B_p)) + \beta_a + (A_v \cdot \sin(\omega t + \phi_v) + B_v) \tag{5.5}$$

Esse procedimento se assemelha ao cálculo de anomalias, usual na meteorologia, no qual se subtrai de uma série um ciclo anual de referência [69]. A diferença em relação às anomalias

Figura 51 – Dessazonalização de séries de vento verificado e previsto, para o modelo GEFS.



Fonte: Elaboração própria.

climáticas convencionais é que o ciclo de referência é estimado a partir do próprio período de dados, de alguns anos, e não de uma normal climatológica de longo prazo, como a de 30 anos [156]. Como esse ciclo é ajustado separadamente ao vento previsto e ao verificado, e o do verificado é reintroduzido ao final, a dessazonalização já corrige o viés sazonal do modelo, cabendo à regressão linear tratar o erro remanescente.

5.3.3.2 Modelagem linear das previsões de vento

De forma resumida, a correção da previsão do vento é realizada por meio do ajuste de regressões lineares às séries dessazonalizadas, considerando cada intervalo de meia hora do dia. Abordagem semelhante foi empregada por [144] no refinamento estatístico das previsões horárias de temperatura a 2 m geradas pelo modelo Eta para estações do Nordeste do Brasil. Essa estratégia está relacionada à metodologia MOC (*Model Output Calibration*), proposta por [157], que consiste na correção das previsões mediante o ajuste de regressões lineares multivariadas, de forma análoga às técnicas de pós-processamento do tipo MOS (*Model Output Statistics*) [69].

A correção utilizada em [144] baseou-se na correlação de cada variável do modelo com o desvio da previsão de temperatura, sendo que a correlação linear que apresentou em média maior redução dos desvios da previsão foi geralmente 0,6. A correção realizada pelo modelo

implementado é teoricamente mais simples, pois as regressões lineares são ajustadas utilizando o vento verificado e o vento previsto apenas, uma vez que não se tem disponibilidade de todas as variáveis envolvidas nos modelos numéricos de tempo. Outra particularidade utilizada em [144] é que, após o cálculo dos coeficientes de correlação e dos coeficientes da equação da regressão, são construídas vinte e quatro equações para cada hora do dia. Todavia, objetivando a previsão de geração de fonte eólica, que requer uma discretização distinta, de meia-hora, são construídas quarenta e oito equações de correção a serem aplicadas no vento previsto.

A principal vantagem dessa metodologia é sua capacidade de produzir resultados satisfatórios utilizando um histórico relativamente curto para a estimação dos modelos. Isso contrasta com técnicas mais complexas, como redes neurais artificiais, que normalmente demandam longas séries temporais de dados para alcançar desempenho adequado, o que nem sempre está disponível na prática. Em [144], por exemplo, foi adotado um período de calibração de 28 dias, sendo a correção aplicada às previsões dos 30 dias subsequentes.

No modelo implementado neste trabalho, as regressões foram ajustadas utilizando um histórico móvel de 30 dias para cada horizonte de previsão do modelo meteorológico. A correção é aplicada às previsões do dia corrente ao longo dos 35 dias de previsão (até 34 dias à frente), que corresponde ao maior horizonte considerado. Dessa forma, ao avaliar a correção para o horizonte de 35 dias, utiliza-se como referência a previsão emitida 35 dias antes da data verificada mais recente disponível. Consequentemente, o processo de calibração pode envolver previsões geradas com até 65 dias de antecedência em relação ao instante de sua avaliação.

Foram avaliados diferentes comprimentos de histórico para o ajuste das regressões. Observou-se que períodos muito curtos tornam os coeficientes estimados excessivamente sensíveis às flutuações recentes, resultando em maior variabilidade das correções. Por outro lado, históricos muito extensos tendem a incorporar condições menos representativas do comportamento atual do modelo, reduzindo a capacidade de capturar os vieses mais recentes das previsões de vento. Assim, o histórico de 30 dias mostrou-se um compromisso adequado entre estabilidade estatística e capacidade de adaptação às características mais recentes do sistema de previsão.

Caso a previsão de vento fosse perfeita, exatamente igual ao vento verificado, traçando um gráfico de dispersão entre eles tem-se uma reta de 45°, ou seja, coeficiente angular igual a um e linear igual a zero. Portanto, quanto melhor a previsão do vento, mais próximo de uma reta é a dispersão desses dados. Isso, a princípio, indicaria a escolha pela regressão linear para modelar a relação entre tais dados. A Equação (5.6) descreve os parâmetros da regressão linear.

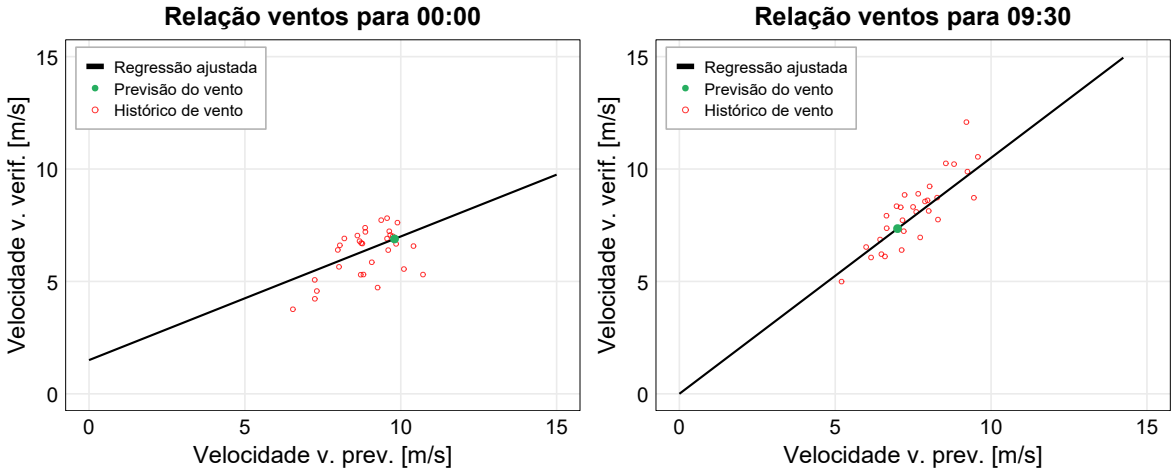
$$V_v = \alpha_a \cdot V_p + \beta_a \tag{5.6}$$

Em que α_a é o coeficiente angular da regressão ajustada; β_a o coeficiente linear da regressão ajustada; V_v corresponde aos dados de vento verificado e V_p aos dados de vento previsto.

Na prática, a relação entre o vento previsto e o verificado se aproxima dessa equação

quanto melhor for a qualidade do vento previsto. Ainda assim, poderia ser regida por alguma função não linear. Isto posto, foram testadas algumas regressões não lineares, como polinômios de vários graus, exponenciais e logísticas. Porém, os melhores resultados confirmaram a opção pelas regressões lineares simples. A Figura 52 ilustra as regressões lineares ajustadas para dois intervalos do dia, 00:00 e 09:30, da correção de um determinado dia. Como se pode observar, essas diferem um pouco da forma da Equação (5.6), mas seguem um comportamento "razoável" dos dados.

Figura 52 – Regressão ajustada para os intervalos das 00:00 e 09:30 do dia, entre vento previsto e verificado.



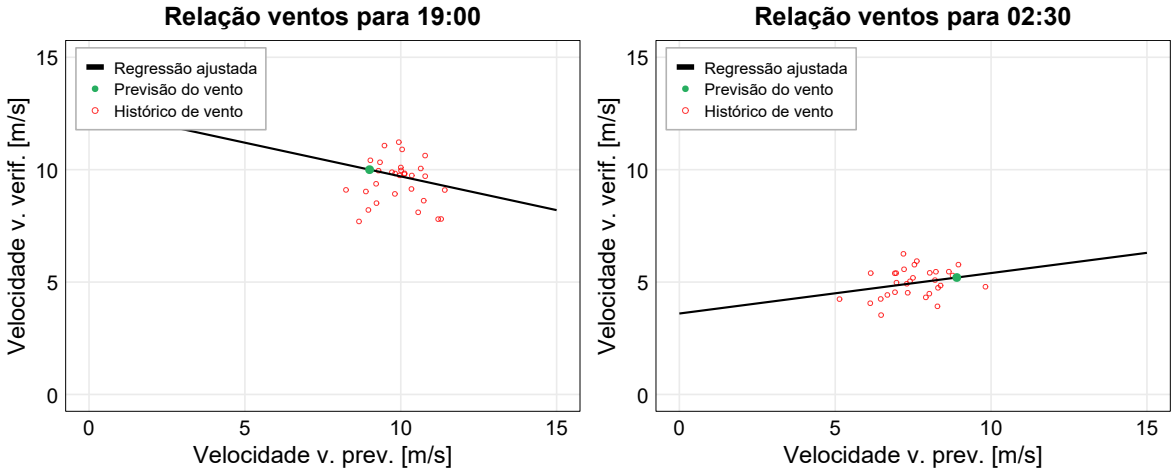
Fonte: Elaboração própria.

Os dois gráficos apresentam a dispersão do vento verificado e previsto para trinta dias (pontos em vermelho), a regressão ajustada para esses dias (reta em preto) e o vento corrigido para o dia imediatamente após os trinta dias do histórico (ponto em verde).

Em algumas situações não é possível realizar ajustes de qualidade das regressões com histórico de trinta dias. Tais situações podem ocorrer devido à baixa qualidade das previsões de vento do período ou por baixa variabilidade do vento no intervalo amostral. A Figura 53 ilustra duas situações de ajustes incoerentes, ocorridas nos intervalos das 19:00 e 02:30. No painel à esquerda da figura, a regressão estimada possui coeficiente angular negativo ($\alpha_a < 0$), o que representa um problema na correção, pois maiores valores de ventos previstos resultariam em menores ventos corrigidos. No painel à direita da mesma figura, o coeficiente angular se aproxima de zero ($\alpha_a \approx 0$), indicando que diferentes valores de vento previsto levariam a um mesmo valor de vento corrigido.

Outra dificuldade para um bom ajuste ocorre quando o coeficiente angular é muito alto, ou seja, o ângulo formado pela reta se aproxima de 90°. Essa circunstância é muito complicada do ponto de vista da correção, pois, depois de ajustada a regressão, qualquer desproporção brusca no vento previsto em relação aos dados de vento previsto que foram usados para o treinamento pode acarretar valores absurdos. Dessa forma, há necessidade de tratamento quando

Figura 53 – Regressão ajustada para os intervalos das 19:00 e 02:30 do dia, com ajuste pobre, entre vento previsto e verificado.



Fonte: Elaboração própria.

da ocorrência dessa situação. Isso ocorre quando, no período dos trinta dias, a previsão do vento para aquele determinado horário não sofre muita variação e, portanto, a dispersão não excursiona ao longo de uma reta. Quando essa situação acontece, é identificada e, através de uma heurística, o tamanho da amostra é aumentado. A heurística analisa se o valor em módulo do coeficiente angular da regressão é maior do que um equivalente para uma reta de 75° e ainda se o desvio-padrão dos dados dos últimos dias é muito pequeno, de acordo com a equação:

$$|\alpha_a| > \left|\tan\left(\frac{75\pi}{180}\right)\right| \quad \text{ou} \quad \sigma_{V_{p,i}} < 0{,}2 \tag{5.7}$$

em que $V_{p,i}$ corresponde ao conjunto de dados do vento previsto dos últimos dias de um período i e $\sigma_{V_{p,i}}$ ao desvio-padrão desses dados.

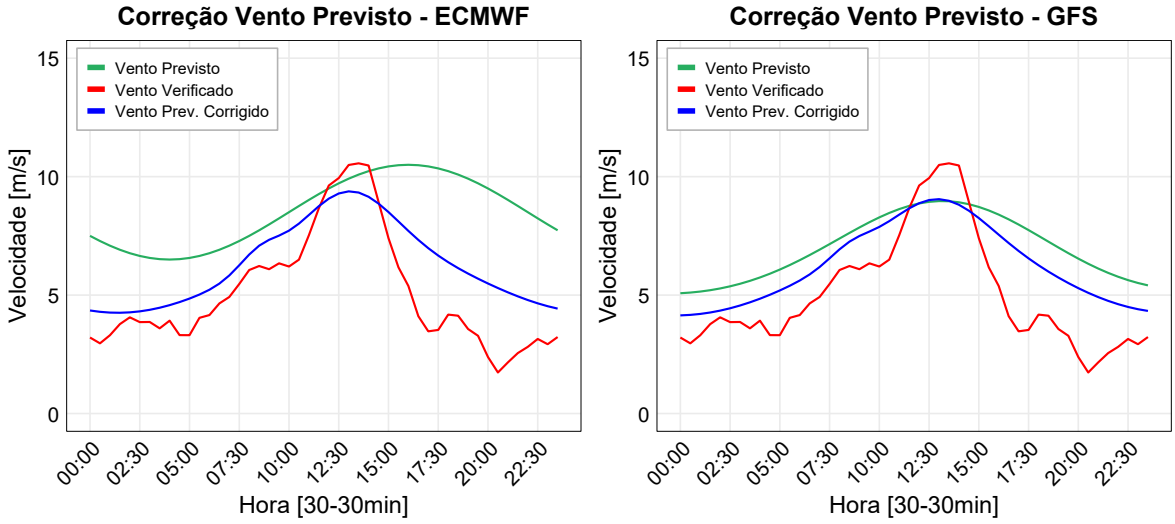
O processo de aumento do histórico se repete até que as restrições sejam atendidas ou se chegue a um limite pré-estabelecido. Esses valores foram escolhidos empiricamente, com base em alguns testes.

5.3.3.3 Exemplo de aplicação do modelo de ajuste da previsão de vento

A título ilustrativo, a Figura 54 apresenta um exemplo de aplicação do modelo de correção para um dia de previsão dos modelos ECMWF e GFS. Esse dia foi escolhido justamente por apresentar divergências entre os modelos de vento previsto. Para os dois painéis foi escolhido um ponto de conexão do subsistema Nordeste.

Observando os gráficos para o ponto selecionado, nota-se uma diferença significativa de viés entre os dois modelos meteorológicos. O ECMWF apresenta as divergências mais acentuadas em relação ao vento verificado, com perfil temporal próprio, distinto do ciclo diurno observado; o GFS, por outro lado, exibe boa aderência à observação, com viés residual de pequena magnitude.

Figura 54 – Impacto da correção de vento nas curvas de vento previsto — modelos ECMWF e GFS, para um ponto de conexão do Nordeste.



Fonte: Elaboração própria.

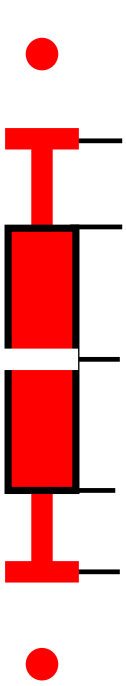
Vale ressaltar que, em ambos os casos, o modelo de correção melhorou a informação prevista, mesmo quando a previsão bruta já apresentava boa aderência ao vento verificado. Neste exemplo pontual, a previsão bruta do GFS mostra-se mais próxima da realidade, mas trata-se de uma única ocorrência: ao longo do tempo, o comportamento do viés varia, podendo cada um dos modelos apresentar maior ou menor viés em diferentes regiões e horizontes.

Um gráfico amplamente utilizado para visualizar os resultados das previsões de vento é o *boxplot*. Esse tipo de representação gráfica permite analisar a distribuição dos erros por meio de estatísticas descritivas, como a mediana, os quartis e valores extremos, possibilitando a identificação de tendências sistemáticas de superestimação ou subestimação das previsões, ou seja, de seus vieses. A Figura 55 apresenta os principais elementos e características desse tipo de gráfico.

A Figura 56 apresenta um exemplo do comportamento dos erros brutos e corrigidos das previsões de velocidade do vento produzidas pelo modelo GFS para uma determinada usina. Embora os erros tenham sido calculados para cada um dos 48 intervalos de 30 minutos que compõem o dia, apenas alguns horários foram selecionados para apresentação, de modo a preservar a clareza e a legibilidade do gráfico.

Com base nesse exemplo (Figura 56), é possível observar que a correção reduz substancialmente os erros do vento previsto. Pelos gráficos correspondentes ao vento previsto corrigido, percebe-se também que, ainda que a magnitude do viés residual seja diferente, nota-se que nos horários iniciais do dia (madrugada) e nos horários finais (noite), o vento previsto tende a ser superestimado, comportamento coerente com o ciclo diurno do erro já discutido na seção de caracterização. Uma análise mais ampla dos resultados da aplicação do modelo de correção, considerando múltiplas regiões, modelos e períodos, é apresentada no capítulo de resultados.

Figura 55 – Características do *boxplot*.



Outlier - valores maiores que 1,5 vezes o quartil superior

Valor máximo, excluindo outliers

Quartil superior, 75% dos dados são menores que esse valor

Mediana, 50% dos dados são menores do que este valor, valor do meio de um conjunto de dados

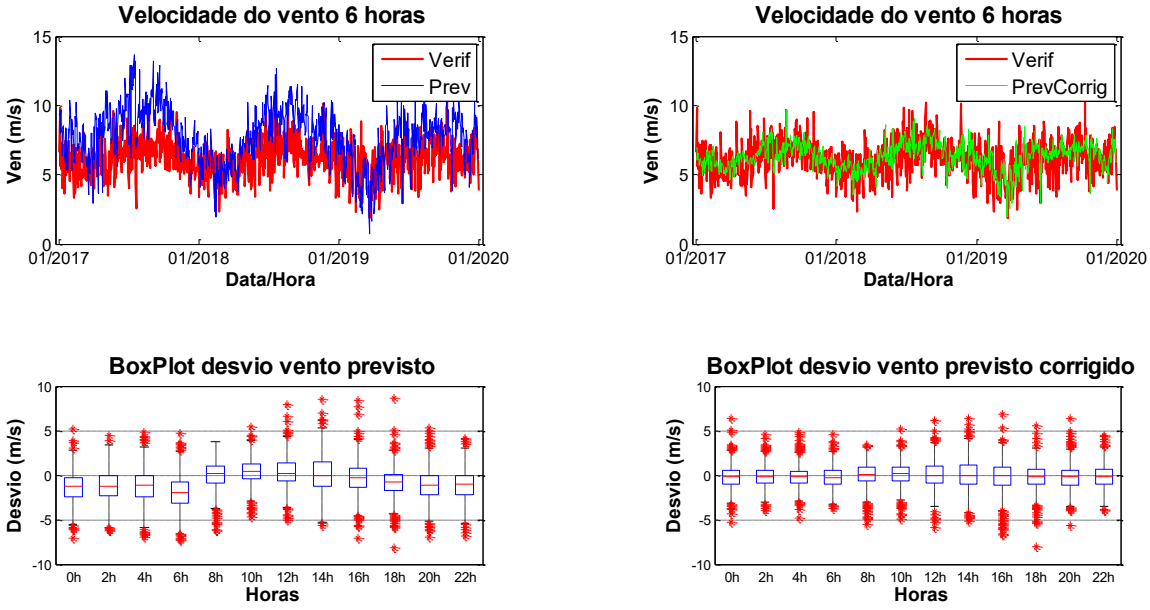
Quartil inferior, 25% dos dados são menores que esse valor

Valor mínimo, excluindo valores outliers

Outlier – valores menores que 1,5 vezes o quartil inferior

Fonte: Elaboração própria.

Figura 56 – Erros brutos do vento previsto e corrigido, pelo GFS, para uma usina.



Fonte: Elaboração própria.

De posse dos ventos previstos ajustados, é possível converter essas informações em geração e, assim, obter a previsão de geração para os próximos dias. Tal modelagem será descrita em detalhes na próxima seção.

5.4 Criação dinâmica das curvas Vento × Potência

Quanto à estratégia de previsão, os testes levaram à conclusão de que as previsões para agrupamentos de usinas apresentam resultados melhores que a previsão para usinas individualizadas. Isso ocorre porque, ao agregar alguns parques eólicos, tem-se um efeito de amortização da variabilidade da geração, o que é comumente chamado de efeito portfólio. Essa redução na variabilidade causa maior estabilidade nos modelos, visto que as previsões de vento pelos NWPs não capturam as variabilidades mais altas. Por esse motivo, a previsão de geração que será feita pelos modelos propostos neste trabalho segue agrupamentos específicos de usinas.

A forma de agrupamento (clusterização) das usinas eólicas é baseada apenas em critérios elétricos. Em geral, esses agrupamentos são determinados considerando as conexões das usinas aos pontos de entrada na rede básica. No entanto, este critério de agrupamento não possui uma justificativa técnica sólida e pode, em alguns casos, agrupar usinas que estão fisicamente muito distantes entre si, ou seja, usinas com perfis de vento possivelmente diferentes. Da mesma forma, há usinas que estão fisicamente próximas, com comportamentos de vento semelhantes, mas que podem ser agrupadas separadamente devido à sua conexão elétrica estar em outro barramento. Para aplicações no planejamento, essa forma de agrupamento pode ser considerada adequada, devido às necessidades de seu uso. No entanto, para situações mais específicas, como em casos de alto carregamento nos pontos de conexão, uma abordagem de clusterização mais apropriada poderia trazer benefícios significativos.

Nesse contexto, foi desenvolvido um método de clusterização baseado em aprendizado de máquina, adaptado ao contexto brasileiro. A abordagem metodológica empregou um Modelo de Mistura de Gaussianas (GMM, do inglês *Gaussian Mixture Model*) [152], inicializado por meio do algoritmo *K-means*, ambos modelos não supervisionados amplamente utilizados em tarefas de agrupamento. O treinamento desses modelos considerou como variáveis de entrada tanto o comportamento do vento quanto a localização geográfica das usinas eólicas, permitindo que o modelo aprendesse padrões espaciais e climáticos relevantes para a formação dos agrupamentos. Embora o modelo ainda requeira ajustes de hiperparâmetros e validações adicionais antes de ser integrado ao *pipeline* apresentado, os resultados obtidos foram promissores. Mais informações sobre o estudo podem ser encontradas em [38].

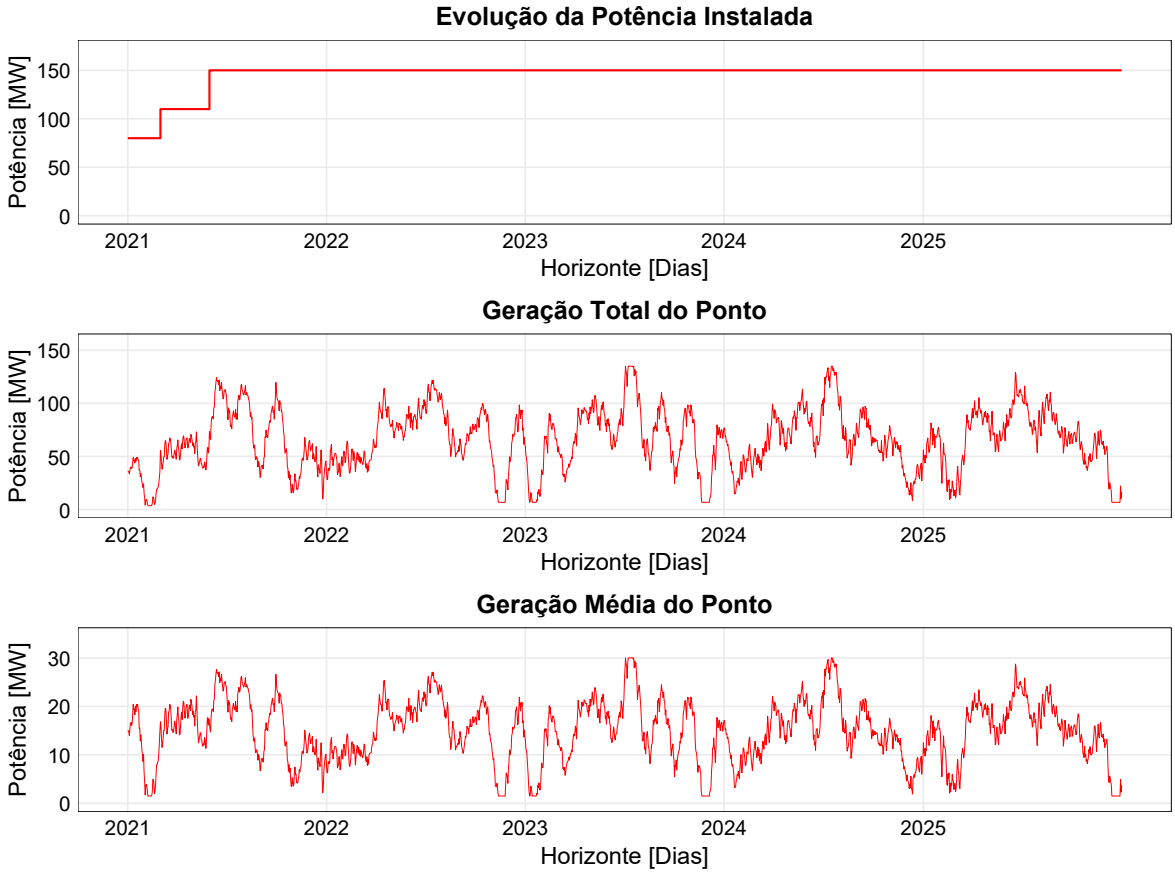
O uso de usinas agrupadas pode gerar algumas necessidades adicionais, de tratamentos e considerações das informações disponíveis para o modelo de previsão. Algumas das dificuldades encontradas serão destacadas a seguir, bem como a forma com que foram tratadas.

As séries de geração verificada nos pontos apresentam valores crescentes ao longo dos anos. Isso não é justificado pelo aumento do potencial dos ventos ao longo dos meses, mas sim pela capacidade instalada que vem aumentando nos últimos anos. Além disso, muitos pontos operam com restrição de transmissão, que eventualmente deixam de existir, ou seja, é como se a capacidade de geração desse ponto fosse alterada. Sendo assim, a calibração dos modelos deve

levar em consideração essas séries crescentes ou variáveis no tempo, cujo crescimento não segue um padrão e é diferente para cada ponto.

A Figura 57 ilustra um exemplo de um ponto em que a potência instalada não se alterou nos anos mais recentes, portanto a série histórica de geração não apresenta deformação devido à capacidade instalada. Neste caso bastaria eliminar os primeiros meses de medição, o que já é feito conforme descrito no processo de construção das curvas V $\times$ P. Nesta figura são apresentados o total e a média da geração das usinas que formam este ponto (conjunto de usinas). Essa série média é calculada considerando apenas as usinas existentes em cada momento do tempo.

Figura 57 – Evolução da capacidade instalada de um determinado ponto estabilizada, com geração total e média das usinas do ponto.

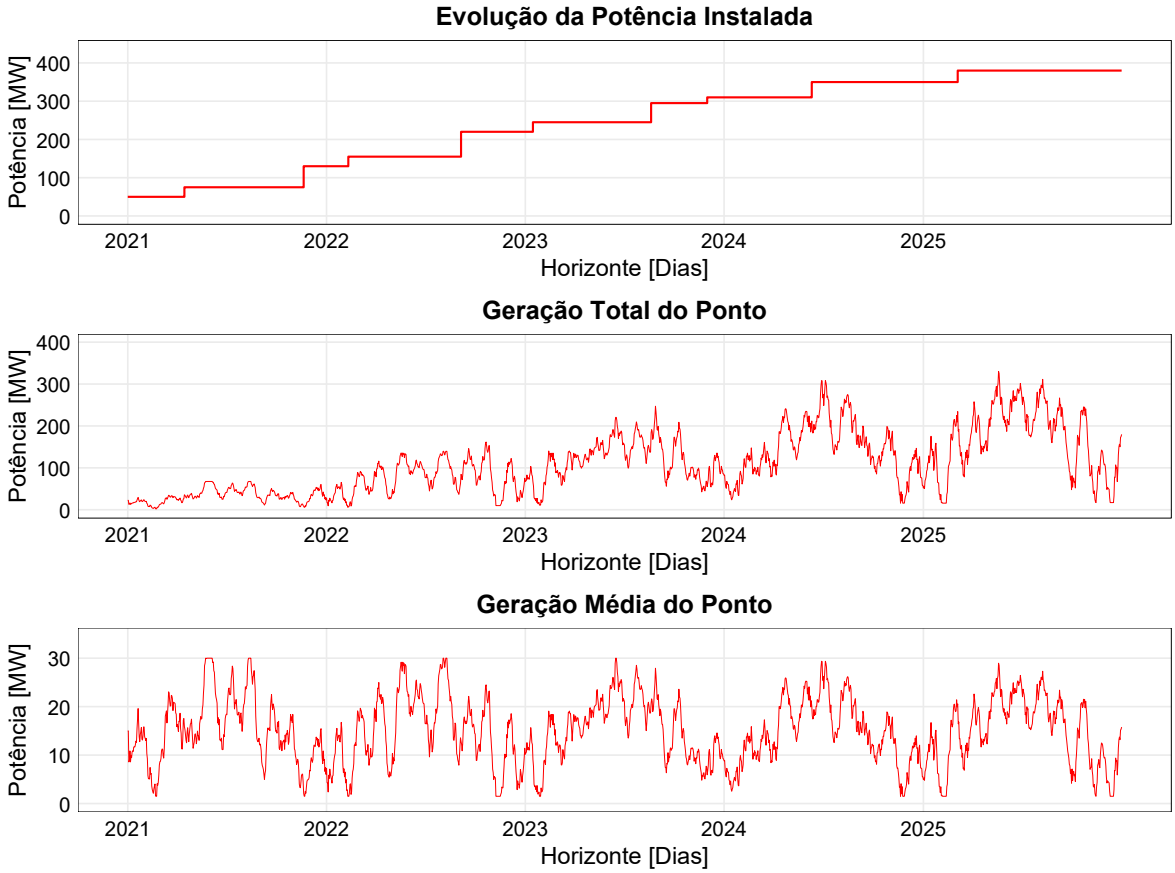


Fonte: Elaboração própria.

Entretanto, o comportamento anterior não se reflete na maioria dos pontos, principalmente na Região Nordeste, onde há entradas frequentes de usinas eólicas no sistema. A Figura 58 apresenta outro ponto de conexão em que a potência instalada vem aumentando gradativamente nos últimos anos. Dessa forma, os modelos devem considerar essa variação, dado que a característica dos ventos não segue esse aumento artificial, apresentando apenas as variações naturais.

De acordo com as Figuras 57 e 58, observa-se que as séries médias não refletem o

Figura 58 – Evolução da capacidade instalada de um determinado ponto em variação, com geração total e média das usinas do ponto.



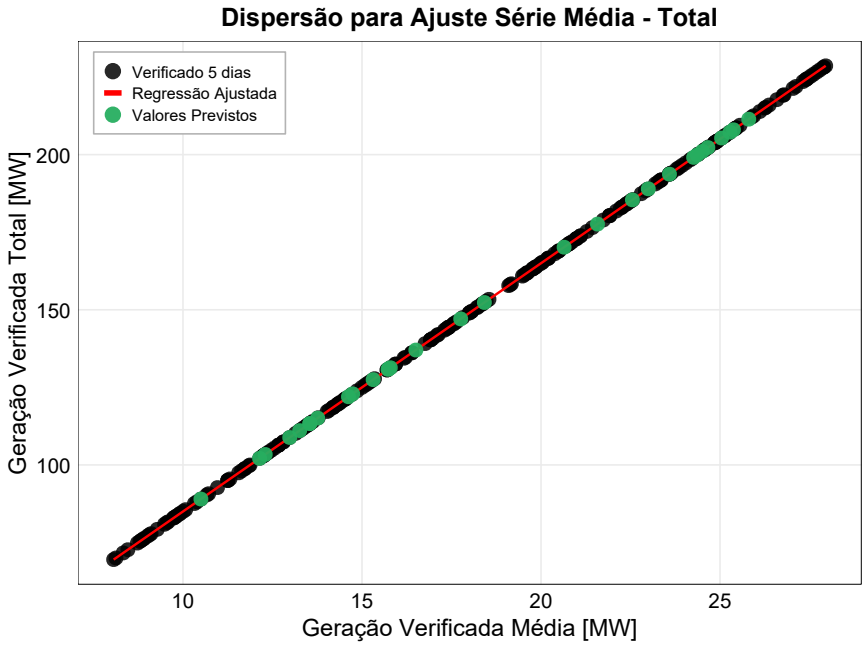
Fonte: Elaboração própria.

comportamento de crescimento do montante total de geração do ponto, tornando-se, portanto, uma alternativa adequada para utilização no modelo. Entretanto, ao transformar as séries históricas utilizadas pelo modelo de previsão em séries médias, é necessário que, ao final do processo de previsão, os resultados sejam reconvertidos para seus patamares totais, uma vez que esses correspondem aos valores de interesse. Para isso, foi ajustado um modelo de regressão linear entre a geração verificada total e a geração verificada média, utilizando os cinco dias que antecedem a data de realização da previsão, conforme ilustrado na Figura 59.

Conforme ilustrado na Figura 59, é evidente que não há praticamente nenhuma perda de informação ao utilizar as séries médias em vez das séries totais. Os pontos verificados (pretos) e previstos (verdes) se distribuem essencialmente sobre a reta de regressão (em vermelho), indicando uma relação linear extremamente bem definida entre as duas grandezas. Este exemplo é representativo de todos os pontos analisados, nos quais essa característica se mantém consistente.

O período de cinco dias foi selecionado por representar uma janela suficientemente curta para acompanhar mudanças recentes no comportamento da geração, ao mesmo tempo em que fornece uma quantidade adequada de dados para o ajuste da regressão, totalizando 240 pontos. A utilização de uma janela reduzida é particularmente importante para mitigar um problema

Figura 59 – Regressão ajustada entre geração média e geração total, usada para reconverter a previsão ao montante total do ponto.



Fonte: Elaboração própria.

recorrente na operação real das usinas eólicas: a ocorrência de restrições operacionais que não são representadas nos dados de entrada do modelo. Assim, caso uma restrição entre em vigor, a metodologia proposta passa a incorporar esse novo comportamento totalmente após cinco dias, sendo impactada apenas durante o período de transição associado ao início ou ao término da restrição.

Cabe destacar que a janela de cinco dias imediatamente anteriores à data de realização da previsão é utilizada apenas quando esse período não contém eventos restritivos, como situações de *curtailment*. Caso existam restrições nessa janela, o método retrocede no histórico até encontrar cinco dias representativos da operação normal da usina, os quais são então utilizados no ajuste da regressão.

Outras situações de restrição operativa poderiam ser tratadas de forma mais adequada caso estivessem disponíveis informações confiáveis sobre a disponibilidade de geração dos parques eólicos. Essa variável, denominada potência disponível, foi discutida no Capítulo 4 e é definida no Submódulo 2.12 dos Procedimentos de Rede [135]. Entretanto, na prática, os dados atualmente disponíveis não apresentam qualidade suficiente para permitir sua utilização em uma modelagem confiável.

Conforme discutido no Capítulo 4, foram apresentadas duas abordagens para a obtenção das curvas V $\times$ P. Embora a regressão sigmoide seja mais abrangente e capaz de representar o comportamento V $\times$ P de todas as usinas, a construção das séries médias por ponto reduz significativamente o impacto das usinas que apresentam decaimento exponencial, uma vez que

esses casos são pouco frequentes. Dessa forma, esse comportamento foi desconsiderado nesta etapa do desenvolvimento.

Em razão disso, optou-se pela utilização da regressão logística, que oferece uma formulação mais simples e adequada aos desdobramentos metodológicos apresentados nas próximas seções. Assim, o processo de formulação dinâmica das curvas V $\times$ P é composto pelas seguintes etapas:

I. **Criação das séries médias**
Inicialmente, são criadas séries de geração verificada média por agrupamento de usinas. Isso é feito para o vento verificado e o vento previsto de cada fonte meteorológica.

II. **Normalização dos dados**
Em seguida, são normalizados os dados de geração média através da geração máxima e mínima desses dados, conforme a equação:

$$NG_v = \frac{G_v - G_{\min}}{G_{\max} - G_{\min}} \tag{5.8}$$

em que NG_v é a geração normalizada; $G_{\max}$ é a geração máxima; $G_{\min}$ é a geração mínima; e G_v o valor da geração a ser normalizada.

De forma similar ao feito para a geração, normalizam-se os dados de vento médio. Porém, neste caso, a normalização é feita com o vento médio e o desvio padrão dessa série temporal, de acordo com a equação:

$$NV_v = \frac{V_v - \overline{V_v}}{\sigma_{V_v}} \tag{5.9}$$

em que NV_v é o vento normalizado; $\overline{V_v}$ é o vento médio; σ_{V_v} é o desvio-padrão do vento; e V_v o valor do vento a ser normalizado.

III. **Cálculo dos parâmetros**
Posteriormente, ajustam-se os parâmetros da seguinte equação logística:

$$NG_v = P_{\min} + \frac{P_{\max} - P_{\min}}{\left(1 + 10^{b\,(V_{mid} - NV_v)}\right)^s} \tag{5.10}$$

Em seguida, é necessário retornar os valores de geração NG_v. Assim, utilizam-se as Equações (5.8), (5.9) e (5.10), resultando em:

$$G_v = \left(P_{\min} + \frac{P_{\max} - P_{\min}}{\left(1 + 10^{\,b\,(V_{mid} - \frac{V_v - \overline{V_v}}{\sigma_{V_v}})}\right)^s} \right) \cdot (G_{\max} - G_{\min}) + G_{\min} \tag{5.11}$$

IV. **Ajuste da regressão linear entre geração média e total**

Como a equação anterior foi construída em torno de uma série de geração e vento médios, ainda é necessária a etapa de conversão em geração total. Conforme explicado no item anterior, essa etapa é realizada com um ajuste de uma regressão linear, entre os dados médios e totais de alguns dias recentes, de acordo com a seguinte equação:

$$G_{v,t} = \alpha_{mt} \cdot G_v + \beta_{mt} \tag{5.12}$$

em que $G_{v,t}$ representa a geração verificada total do ponto; α_{mt} representa o coeficiente angular da regressão; e β_{mt} o coeficiente linear.

Substituindo a Equação (5.11) na Equação (5.12) e adaptando a notação para a obtenção da geração prevista a partir do vento previsto ajustado, a equação final pode ser escrita como:

$$G_{p,t} = \alpha_{mt} \cdot \left(\left(P_{\min} + \frac{P_{\max} - P_{\min}}{\left(1 + 10^{b\,(V_{mid} - \frac{V_{pa} - \overline{V_v}}{\sigma_{V_v}})}\right)^s} \right) \cdot (G_{\max} - G_{\min}) + G_{\min} \right) + \beta_{mt} \tag{5.13}$$

em que $G_{p,t}$ representa a geração prevista total do ponto e V_{pa} o vento previsto ajustado com a metodologia apresentada anteriormente.

5.5 Modelo final de conversão de vento em potência

Apresentadas a modelagem de ajuste na previsão de vento e a construção das curvas V $\times$ P por agrupamentos, o passo final é descrever a equação resultante. Substituindo a Equação (5.5) na Equação (5.13) e realizando as devidas manipulações matemáticas e agrupamentos de parâmetros intermediários, obtém-se a Equação (5.14), que relaciona a geração prevista total do ponto com o vento previsto bruto dos modelos numéricos.

$$G_{p,t} = \frac{\alpha_{mt} \cdot (P_{\max} - P_{\min}) \cdot (G_{\max} - G_{\min})}{\left(1 + 10^{b\left(V_{mid}(\alpha_a V_p - \alpha_a A_p \sin(\omega t + \phi_p) + A_v \sin(\omega t + \phi_v) + K) - \frac{\overline{V_v}}{\sigma_{V_v}}\right)}\right)^s} + \alpha_{mt} \cdot (P_{\min} + G_{\min}) + \beta_{mt} \tag{5.14}$$

Em que A'_p é a amplitude ajustada ($\alpha_a A_p$); K é uma constante do modelo que representa parâmetros intermediários.

5.6 Modelo de combinação de previsões

5.6.1 Motivação e contexto

A combinação de previsões é uma estratégia consolidada na literatura de séries temporais e previsão de energia [142,143]. Em vez de selecionar um único modelo de previsão, a abordagem combina as saídas de múltiplos modelos individuais, explorando a diversidade entre eles para obter previsões mais robustas e com menor erro agregado.

No contexto da previsão de geração de energia eólica para o SIN, diferentes modelos meteorológicos e estatísticos podem apresentar desempenhos distintos conforme o horizonte de previsão, a época do ano ou as condições sinóticas vigentes. O modelo de combinação proposto busca ponderar adaptativamente cada modelo individual com base no seu desempenho recente, atribuindo maior peso aos modelos que apresentaram menor erro quadrático médio nas observações mais recentes.

5.6.2 Formulação geral

Considere M modelos de previsão individuais, indexados por $m \in \{1,2,\ldots,M\}$. Para cada instante de emissão i e horizonte de previsão j (coluna da matriz de previsões, correspondente a uma hora específica do horizonte), a previsão combinada $G^{\text{comb}}_{p,i,j}$ é definida como a média ponderada das previsões individuais:

$$G^{\text{comb}}_{p,i,j} = \sum_{m=1}^{M} w^{*}_{i,j,m}\, G^{(m)}_{p,i,j}, \tag{5.15}$$

em que $G^{(m)}_{p,i,j}$ é a previsão do modelo m para o instante de emissão i e horizonte j, e $w^{*}_{i,j,m}$ é o peso corrigido atribuído a esse modelo, com a restrição:

$$\sum_{m=1}^{M} w^{*}_{i,j,m} = 1, \qquad w^{*}_{i,j,m} \geq 0. \tag{5.16}$$

5.6.3 Cálculo dos pesos pelo inverso do erro quadrático médio

Os pesos são calculados com base no desempenho histórico recente de cada modelo, medido pela Raiz do Erro Quadrático Médio (RMSE, *Root Mean Square Error*).

5.6.3.1 Janela de avaliação retrospectiva

Para o cálculo do RMSE de referência na posição (i, j), utiliza-se uma janela deslizante de L instâncias de emissão anteriores. Dado que a geração verificada possui uma defasagem operacional de δ instâncias de emissão e que a coluna j corresponde ao horizonte de $d = \lfloor (j-2)/24 \rfloor$ dias, o conjunto de instâncias de avaliação válidas é:

$$\mathcal{R}_{i,j} = \{r \in \mathbb{Z} \mid r_{\max} - L + 1 \leq r \leq r_{\max}, \quad r \geq 1\}, \tag{5.17}$$

em que o limite superior da janela é:

$$r_{\max} = i - \delta - d. \tag{5.18}$$

Nesta implementação, adotou-se $L = 20$ instâncias de emissão e $\delta = 2$ instâncias, de forma a garantir que apenas dados efetivamente disponíveis no momento da previsão sejam utilizados no cômputo dos pesos, preservando a causalidade do processo.

5.6.3.2 RMSE por modelo na janela

Para cada modelo m e posição (i, j), o RMSE é calculado sobre o conjunto $\mathcal{R}_{i,j}$, desconsiderando entradas com dados faltantes (*missing data*):

$$\mathrm{RMSE}_{i,j,m} = \sqrt{\frac{1}{|\mathcal{V}_{i,j,m}|} \sum_{r \in \mathcal{V}_{i,j,m}} \left(G^{(m)}_{p,r,j} - G_{v,r,j} \right)^2}, \tag{5.19}$$

em que $G_{v,r,j}$ é a geração verificada na instância r e horizonte j, e $\mathcal{V}_{i,j,m}$ é o subconjunto de $\mathcal{R}_{i,j}$ onde tanto a previsão do modelo m quanto o verificado estão disponíveis:

$$\mathcal{V}_{i,j,m} = \left\{ r \in \mathcal{R}_{i,j} \;\middle|\; G^{(m)}_{p,r,j} \neq \mathrm{NA} \;\wedge\; G_{v,r,j} \neq \mathrm{NA} \right\}. \tag{5.20}$$

O modelo m é considerado avaliável na posição (i, j) apenas se $|\mathcal{V}_{i,j,m}| \geq n_{\min}$, onde $n_{\min} = 3$ é o número mínimo de observações exigido para garantir estabilidade estatística do RMSE.

5.6.3.3 Pesos brutos por inverso do erro quadrático médio

O peso bruto $\tilde{w}_{i,j,m}$ de cada modelo avaliável é proporcional ao inverso do quadrado do seu RMSE (i.e., ao inverso do erro quadrático médio, que incorpora tanto a variância quanto o viés do erro):

$$\tilde{w}_{i,j,m} = \begin{cases} \dfrac{1}{\mathrm{RMSE}^2_{i,j,m}}, & \text{se } m \text{ é avaliável,} \\ 0, & \text{caso contrário.} \end{cases} \tag{5.21}$$

Os pesos normalizados são então:

$$w_{i,j,m} = \frac{\tilde{w}_{i,j,m}}{\displaystyle\sum_{m'=1}^{M} \tilde{w}_{i,j,m'}}. \tag{5.22}$$

5.6.3.4 Casos degenerados

Algumas situações especiais são tratadas explicitamente:

1. **Janela insuficiente** ($r_{\max} < 1$ ou $|\mathcal{R}_{i,j}| = 0$): não há histórico suficiente para calcular o RMSE. Atribui-se peso uniforme a todos os modelos:

$$w_{i,j,m} = \frac{1}{M}, \quad \forall m. \tag{5.23}$$

2. **Nenhum modelo avaliável** ($|\mathcal{V}_{i,j,m}| < n_{\min}$ para todo m): mesma regra do caso anterior, peso uniforme.

3. **Todos os modelos avaliáveis com RMSE nulo**: indica previsão perfeita na janela. Distribui-se peso igual entre os modelos avaliáveis:

$$w_{i,j,m} = \begin{cases} \dfrac{1}{|\mathcal{A}_{i,j}|}, & \text{se } m \in \mathcal{A}_{i,j}, \\ 0, & \text{caso contrário,} \end{cases} \tag{5.24}$$

em que $\mathcal{A}_{i,j}$ é o conjunto de modelos avaliáveis em (i, j).

5.6.4 Correção dos pesos na presença de previsões ausentes

Os pesos $w_{i,j,m}$ calculados na etapa anterior são baseados exclusivamente no desempenho histórico e não levam em conta a disponibilidade da previsão na posição corrente (i, j). Quando um ou mais modelos não possuem previsão disponível em (i, j) (seja por falha operacional, atraso de transmissão ou outro motivo) é necessário redistribuir os pesos entre os modelos disponíveis.

Define-se o indicador de disponibilidade de cada modelo na posição (i, j):

$$\xi_{i,j}^{(m)} = \begin{cases} 1, & \text{se } G_{p,i,j}^{(m)} \neq \text{NA}, \\ 0, & \text{caso contrário.} \end{cases} \tag{5.25}$$

O peso corrigido $w_{i,j,m}^{*}$ é obtido zerando os pesos dos modelos indisponíveis e renormalizando:

$$w_{i,j,m}^{*} = \begin{cases} \text{NA}, & \text{se } \displaystyle\sum_{m'=1}^{M} \xi_{i,j}^{(m')} = 0 \quad \text{(nenhum modelo disponível)}, \\ \dfrac{w_{i,j,m} \cdot \xi_{i,j}^{(m)}}{\displaystyle\sum_{m'=1}^{M} w_{i,j,m'} \cdot \xi_{i,j}^{(m')}}, & \text{se } \displaystyle\sum_{m'=1}^{M} w_{i,j,m'} \cdot \xi_{i,j}^{(m')} > 0, \\ \dfrac{\xi_{i,j}^{(m)}}{\displaystyle\sum_{m'=1}^{M} \xi_{i,j}^{(m')}}, & \text{caso contrário (pesos dos disponíveis eram nulos).} \end{cases} \tag{5.26}$$

Quando nenhum modelo está disponível, a previsão combinada assume valor ausente: $G^{\text{comb}}_{p,i,j} = \text{NA}$.

5.6.5 Diagrama do fluxo metodológico

O processo completo de geração da previsão combinada pode ser resumido nas seguintes etapas sequenciais:

1. **Etapa 1, Cálculo dos pesos históricos:** Para cada posição (i, j), define-se a janela retrospectiva $\mathcal{R}_{i,j}$, calcula-se o RMSE de cada modelo comparando $G^{(m)}_{p,r,j}$ com $G_{v,r,j}$, e obtém-se os pesos normalizados $w_{i,j,m}$ pelo critério do inverso do erro quadrático médio (Equações (5.19)–(5.22)), tratando os casos degenerados conforme descrito.

2. **Etapa 2, Correção por disponibilidade:** Os pesos são corrigidos para refletir a disponibilidade operacional das previsões $G^{(m)}_{p,i,j}$ no instante corrente, gerando $w^{*}_{i,j,m}$ (Equação (5.26)).

3. **Etapa 3, Geração da previsão combinada:** Aplica-se a combinação linear ponderada para obter $G^{\text{comb}}_{p,i,j}$ conforme a Equação (5.15).

5.6.6 Propriedades do método

O método apresenta as seguintes propriedades desejáveis para aplicação operacional:

- **Causalidade**: os pesos são calculados exclusivamente com base em dados passados, comparando previsões $G^{(m)}_{p,r,j}$ com gerações verificadas $G_{v,r,j}$ já disponíveis. A defasagem δ e o deslocamento d garantem que nenhuma informação futura seja utilizada.

- **Adaptatividade**: a janela deslizante de comprimento L permite que os pesos respondam a mudanças no desempenho relativo dos modelos ao longo do tempo, sem requerer reestimação paramétrica explícita.

- **Robustez a falhas operacionais**: a etapa de correção por disponibilidade (Etapa 2) garante que $G^{\text{comb}}_{p,i,j}$ permaneça disponível mesmo na ausência de um ou mais modelos individuais.

- **Consistência nos casos extremos**: peso uniforme é atribuído quando não há histórico suficiente ou quando todos os modelos apresentam erros nulos, evitando divisões por zero e instabilidades numéricas.

5.6.7 Parâmetros do modelo

Os parâmetros adotados nesta implementação estão sumarizados na Tabela 5.

Os três modelos individuais ($M = 3$) correspondem aos modelos de previsão numérica do tempo que, à época da implementação, apresentavam maior contribuição para a qualidade da

Tabela 5 – Parâmetros do modelo de combinação por inverso do erro quadrático médio.

Símbolo	Descrição	Valor	Unidade
M	Número de modelos individuais	3	–
L	Comprimento da janela retrospectiva	20	instâncias
δ	Defasagem operacional do verificado	2	instâncias
$n_{\min}$	Mínimo de observações válidas para RMSE	3	observações

Fonte: Elaboração própria.

previsão de geração eólica no SIN, sendo os demais (como o WRF e o BAM) descartados por introduzirem ruído e não agregarem ganho ao conjunto, conforme discutido na Seção 3.1.

O comprimento da janela retrospectiva ($L = 20$) foi definido com base em testes de sensibilidade realizados com diferentes valores, buscando o equilíbrio entre adaptatividade, ou seja, a capacidade de responder a mudanças no desempenho relativo dos modelos, e estabilidade estatística do estimador de RMSE. Janelas excessivamente curtas produzem pesos instáveis e ruidosos, enquanto janelas longas incorporam regimes de vento de períodos distintos, diluindo o sinal de desempenho recente [158].

A defasagem operacional ($\delta = 2$) reflete a latência mínima do fluxo de coleta e consolidação da geração verificada no sistema operacional, garantindo estrita causalidade: apenas dados efetivamente disponíveis no momento de emissão da previsão são utilizados no cômputo dos pesos.

O limiar mínimo de observações ($n_{\min} = 3$) assegura a estabilidade numérica do estimador de RMSE, evitando pesos extremos decorrentes de amostras degeneradas. Posições com menos de três observações válidas, restritas ao período de aquecimento do modelo ou a falhas simultâneas de dados, recebem peso uniforme como alternativa conservadora.

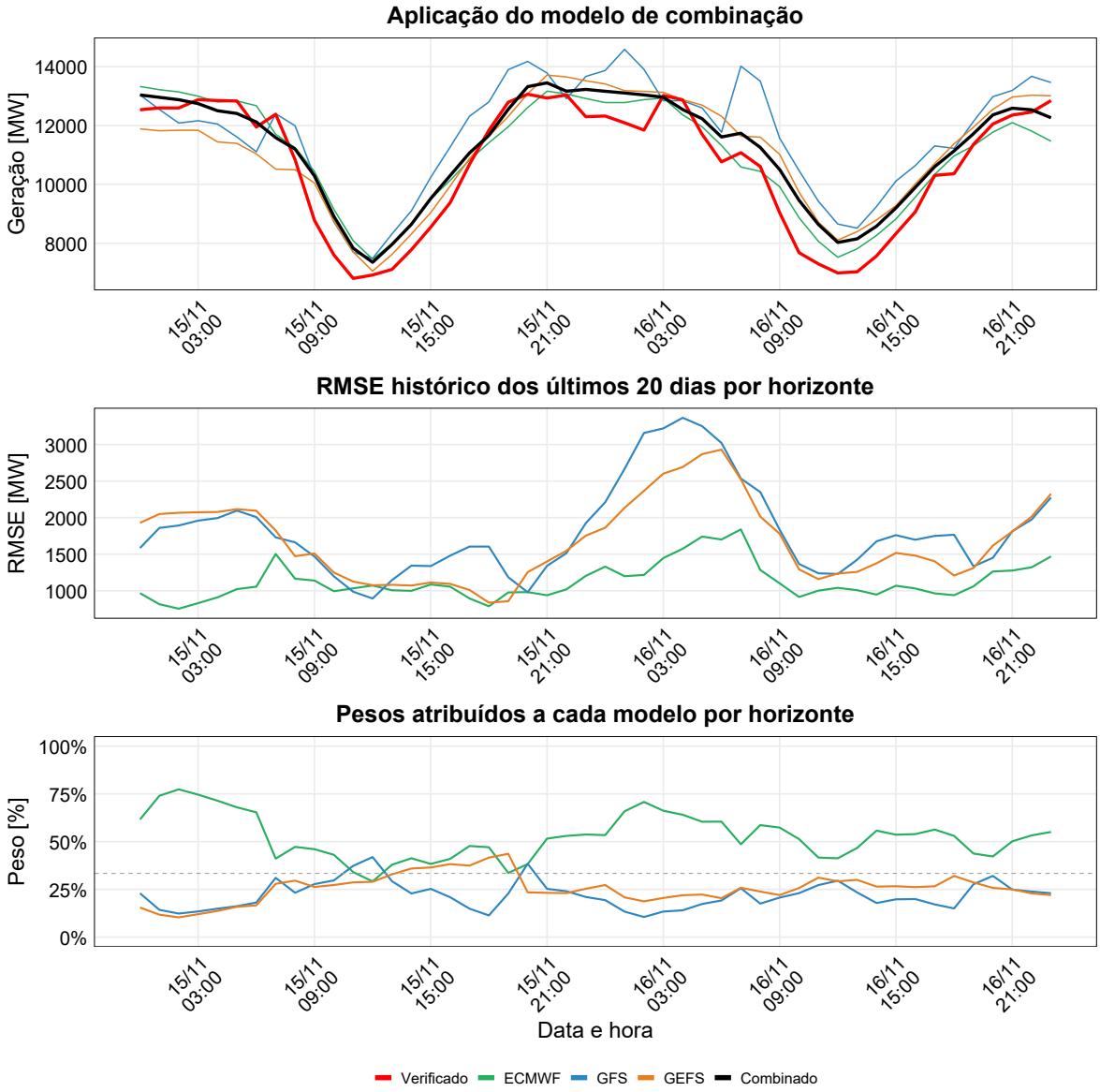
5.6.8 Aplicação do modelo

A título ilustrativo, a Figura 60 apresenta um exemplo de aplicação do modelo de combinação para um dia de previsão das primeiras 48 horas, considerando os três modelos meteorológicos avaliados (ECMWF, GFS e GEFS) e o subsistema Nordeste. O dia mostrado foi selecionado de forma a evidenciar o funcionamento da metodologia. A figura é composta por três painéis verticais.

O painel superior apresenta as séries de geração eólica do dia analisado: o vento verificado (em vermelho) e a previsão combinada (em preto) são exibidos com linhas mais espessas, em destaque, enquanto as previsões dos três modelos meteorológicos individuais (em verde, azul e laranja) aparecem com linhas mais finas, servindo de referência. Observa-se que a previsão combinada acompanha de forma consistente o comportamento verificado ao longo das 48 horas, ainda que os modelos individuais apresentem desempenhos heterogêneos entre si.

O painel central exibe o RMSE histórico de cada modelo, calculado sobre os últimos

Figura 60 – Aplicação do modelo de combinação — subsistema Nordeste (séries, RMSE histórico e pesos por horizonte).



Fonte: Elaboração própria.

20 dias de previsão, separadamente para cada um dos 48 horizontes do dia. Esse painel representa, em última instância, a informação que o modelo de combinação utiliza para decidir os pesos a serem atribuídos: modelos com RMSE histórico menor em determinado horizonte são considerados mais confiáveis naquele instante. Nota-se que a hierarquia entre os três modelos pode variar ao longo do dia, refletindo o fato de que cada modelo apresenta horários do dia em que se mostra mais ou menos aderente ao vento verificado.

O painel inferior apresenta os pesos efetivamente atribuídos a cada modelo na combinação, em função do horizonte de previsão, expressos em percentual. A linha tracejada cinza em 33,3% representa a situação de ponderação uniforme entre os três modelos. Assim, valores acima dessa referência indicam que o modelo correspondente está recebendo maior importância

na combinação, enquanto valores inferiores refletem uma penalização em relação aos demais.

Conforme esperado pela lógica da metodologia proposta, os pesos apresentados no painel inferior refletem, de forma inversa, o comportamento do RMSE histórico exibido no painel central. Em cada horizonte de previsão, o modelo que apresenta o menor RMSE histórico recebe o maior peso na combinação, enquanto modelos com erros mais elevados têm sua contribuição reduzida. Esse mecanismo estabelece uma relação direta entre o desempenho recente dos modelos e sua participação na previsão combinada, conferindo caráter adaptativo à metodologia, que ajusta dinamicamente a importância relativa de cada fonte de previsão com base em seu desempenho histórico.

Por fim, cabe destacar que, após a etapa de combinação, a série de previsão resultante pode apresentar picos ou vales artificiais. Isso ocorre porque a combinação emprega pesos definidos por horizonte (hora a hora), de modo que variações abruptas entre os pesos de horas consecutivas podem introduzir descontinuidades na série combinada. Soma-se a isso o fato de que outras etapas dos modelos podem, igualmente, gerar pequenas distorções pontuais. Para mitigar esse efeito e produzir uma série mais suave e fisicamente plausível, aplica-se, como etapa final, um filtro de suavização *Locally Estimated Scatterplot Smoothing* (LOESS) [153]. Esse método baseia-se no ajuste sucessivo de regressões locais ponderadas, a partir de dados vizinhos, atenuando as oscilações artificiais sem comprometer o comportamento geral da série prevista.

5.7 Interpolação horária das previsões de geração

Como o modelo proposto realiza previsões para horizontes de 35 dias de previsão (até 34 dias à frente), a disponibilidade dos insumos meteorológicos varia conforme o alcance temporal de cada modelo. Durante a primeira semana, as previsões são obtidas a partir dos três modelos meteorológicos utilizados neste trabalho (ECMWF, GFS e GEFS). Na segunda semana, apenas os modelos GFS e GEFS permanecem disponíveis, enquanto da terceira à quinta semana as previsões são baseadas exclusivamente no modelo GEFS, único capaz de fornecer previsões para todo esse horizonte. Assim, em cada horizonte de previsão, o modelo de combinação utiliza apenas as previsões efetivamente disponíveis.

Além disso, cada modelo meteorológico fornece previsões com uma discretização temporal própria, podendo apresentar intervalos de uma, três ou seis horas entre previsões consecutivas. Inicialmente, as previsões de geração são produzidas na mesma resolução temporal dos respectivos insumos meteorológicos. Em seguida, realiza-se uma etapa de interpolação temporal para obter previsões em discretização horária.

A interpolação empregada no modelo de previsão difere de uma simples interpolação linear. Como o intervalo entre previsões pode chegar a seis horas, especialmente no caso do modelo GEFS, uma interpolação linear tende a não representar adequadamente a evolução temporal da geração nesse período. Em razão disso, foi desenvolvida uma metodologia que

utiliza o perfil de uma curva típica para distribuir os valores intermediários, preservando o comportamento esperado da geração ao longo do tempo.

Para a construção dessa curva típica, optou-se por utilizar a série verificada de vento médio, em vez da série de geração observada. Essa escolha decorre do fato de que a geração pode sofrer reduções decorrentes de restrições operativas, como cortes de geração, o que distorceria o perfil utilizado na interpolação. Por outro lado, a série de vento médio não é influenciada por essas restrições, fornecendo uma representação mais consistente do comportamento temporal do recurso eólico e, consequentemente, uma base mais adequada para a interpolação das previsões.

5.8 Particularidades da metodologia de previsão de geração de curtíssimo prazo

Concluídas as etapas anteriores, obtém-se a modelagem final representada pela Equação 5.14, responsável por converter as previsões brutas de velocidade do vento em previsão de geração eólica para cada um dos modelos meteorológicos considerados. Na sequência, realiza-se a combinação dessas previsões, resultando em previsões horárias de geração eólica por conjuntos, com horizonte de previsão de até um mês à frente.

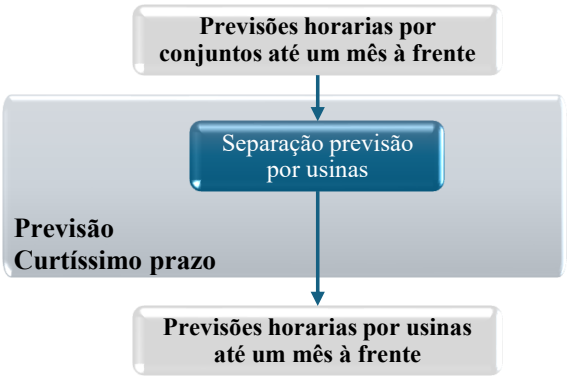
Entretanto, os modelos de planejamento de curtíssimo prazo (MPCTP), atualmente representados pelo DESSEM, requerem previsões associadas às barras elétricas da rede. Esse tipo de agrupamento é específico desse modelo e, em geral, apresenta um nível de desagregação espacial mais detalhado do que aquele adotado na etapa de previsão da geração eólica. Dessa forma, torna-se necessária a desagregação das previsões obtidas.

Além disso, existem diferentes aplicações que demandam níveis distintos de discretização espacial das previsões. Por exemplo, os operadores em tempo real necessitam de previsões no nível dos pontos de conexão das usinas; a programação diária da operação utiliza agrupamentos específicos, definidos de acordo com a representatividade das usinas para o SIN, bem como previsões por subsistema para o planejamento dos intercâmbios de energia entre regiões; por sua vez, os modelos de planejamento de curtíssimo prazo requerem previsões agregadas segundo barras elétricas específicas.

Considerando a coexistência dessas diferentes necessidades, bem como a possibilidade de surgirem novas formas de agregação no âmbito do ONS e de outros modelos energéticos, a solução mais flexível consiste na obtenção de previsões individualizadas por usina. A partir desse nível de discretização espacial, torna-se possível compor quaisquer agrupamentos de interesse, adequando as previsões à sua finalidade de utilização.

As particularidades do modelo de curtíssimo prazo estão representadas na configuração ilustrada na Figura 61. A seção seguinte apresenta a metodologia empregada para a desagregação das previsões por agrupamentos, correspondente ao bloco específico indicado nessa figura.

Figura 61 – Estrutura do modelo de curtíssimo prazo.



Fonte: Elaboração própria.

5.8.1 Metodologia de separação das previsões por usinas

Para ilustrar a metodologia proposta, considera-se, ao longo desta seção, um exemplo composto por três usinas eólicas hipotéticas (A, B e C) pertencentes a um mesmo agrupamento, neste caso um ponto de conexão, cada uma com potência instalada de 20 MW. A Figura 62 apresenta, no painel esquerdo, a geração verificada (linha azul) e a previsão do agrupamento (linha preta), correspondente ao somatório das três usinas, além da potência instalada total do ponto de conexão, representada pela linha horizontal preta em 60 MW.

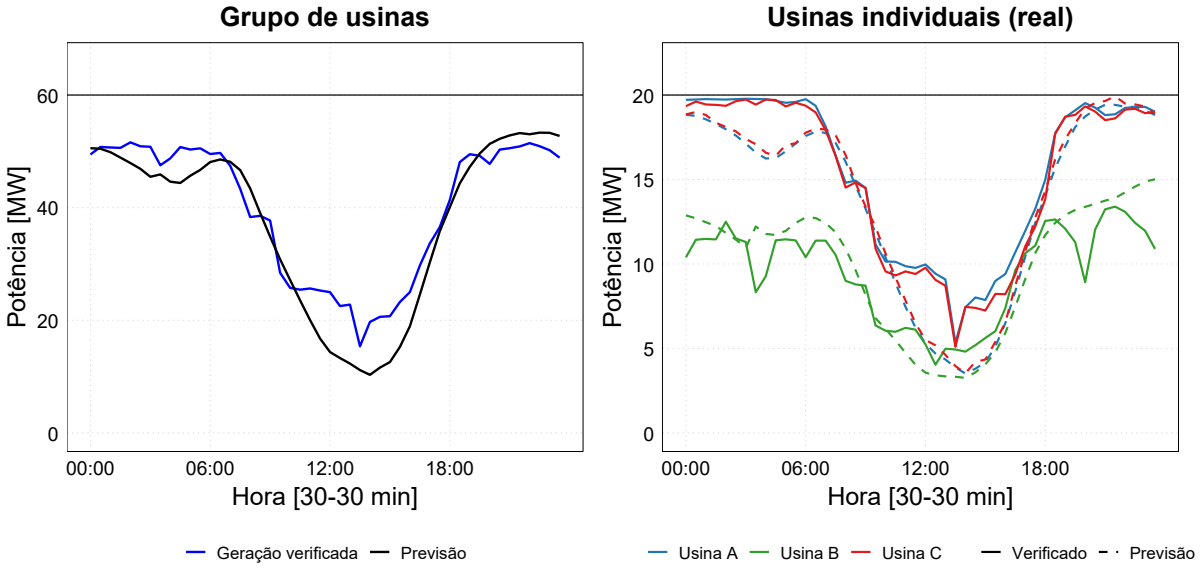
No painel direito, são apresentadas as séries temporais de geração verificada (linhas contínuas) e de geração prevista (linhas tracejadas) para cada uma das usinas individualmente. As usinas são identificadas por cores distintas: azul para a Usina A, verde para a Usina B e vermelho para a Usina C.

Ressalta-se que as previsões individuais exibidas nesse exemplo possuem caráter exclusivamente ilustrativo e são utilizadas apenas para facilitar a compreensão da metodologia. Na aplicação operacional, a previsão é realizada apenas para o agrupamento, cabendo ao procedimento descrito nas próximas seções a tarefa de desagregar a previsão agregada em previsões individualizadas para cada usina.

Para realizar a tarefa de separação das previsões por usinas, uma primeira alternativa seria a criação de fatores proporcionais à potência instalada de cada usina. Como no exemplo considerado as três usinas possuem potências instaladas iguais, esse critério atribuiria a cada usina exatamente um terço da geração ou previsão do grupo. A Figura 63 ilustra o resultado dessa estratégia: as linhas tracejadas representam a separação obtida aplicando-se um fator constante de $1/3$ tanto à série verificada do grupo (painel esquerdo) quanto à série prevista do grupo (painel direito). As linhas contínuas correspondem aos valores reais de cada usina individual.

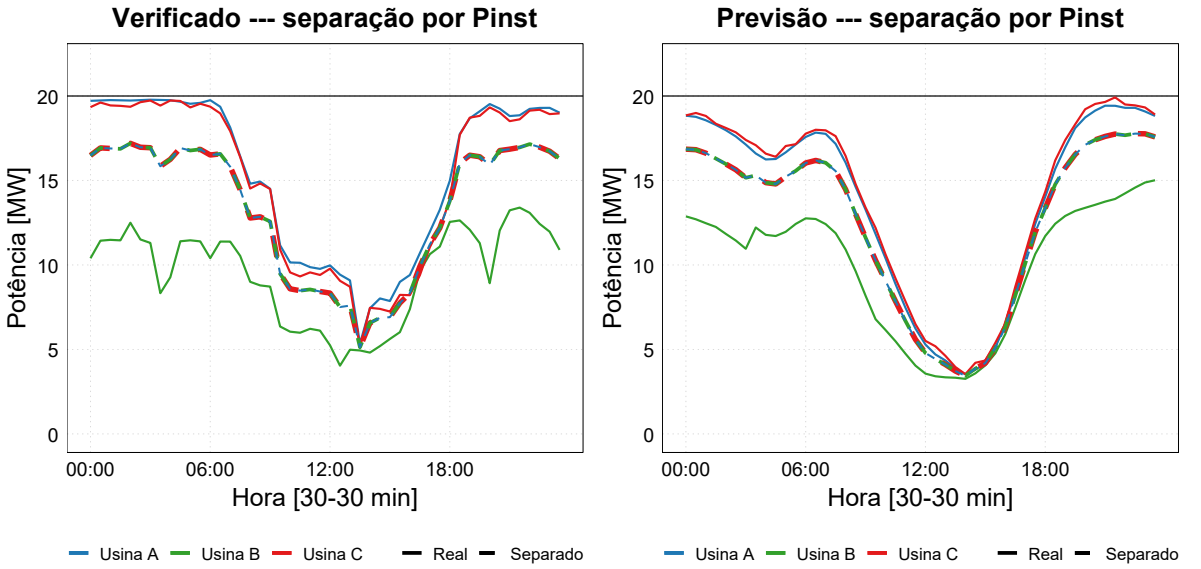
Observa-se que, como as três usinas possuem a mesma potência instalada, as linhas tracejadas correspondentes à separação ficam sobrepostas. Para que as três cores permaneçam

Figura 62 – Geração verificada e previsão para o grupo e para as usinas individuais.



Fonte: Elaboração própria.

Figura 63 – Separação das previsões por usinas utilizando fator proporcional à potência instalada.



Fonte: Elaboração própria.

visíveis, optou-se por representá-las com espessuras distintas. Comparando-se as séries reais (linhas contínuas) com as séries separadas (linhas tracejadas), fica evidente a limitação dessa abordagem: a separação proporcional à potência instalada ignora as diferenças de fator de capacidade entre as usinas. Em particular, a Usina B (em verde) apresenta gerações sistematicamente inferiores às demais, indicando um fator de capacidade menor, fato que não é capturado pela separação por potência instalada. Como consequência, a separação superestima a geração da Usina B e subestima a geração das Usinas A e C ao longo de praticamente todo o dia.

Uma alternativa mais adequada consiste em utilizar fatores de participação calculados a partir da geração verificada recente de cada usina. A premissa dessa abordagem é que a razão entre a geração observada de uma usina e a geração observada do agrupamento, em determinado período do dia, reflete a contribuição relativa dessa usina para a geração total naquele intervalo. Como esses fatores são obtidos a partir de medições recentes, eles incorporam implicitamente as características operativas e produtivas de cada usina, incluindo seu desempenho efetivo ao longo do ciclo diário e eventuais alterações em seu regime de operação. Dessa forma, a desagregação da previsão passa a refletir de maneira mais fiel a distribuição observada da geração entre as usinas do agrupamento, sem a necessidade de estimar explicitamente fatores de capacidade ou outros parâmetros adicionais.

Inicialmente, calcula-se a geração total de cada grupo de usinas nos últimos 4 dias, conforme a Equação (5.27).

$$G_{v,gr,j} = \sum_{i=1}^{N_{usi}^{gr}} G_{v,i,j} \tag{5.27}$$

Em que $G_{v,gr,j}$ é a geração verificada total nos últimos 4 dias do grupo (ponto) no intervalo j; $G_{v,i,j}$ é a geração verificada da usina i no intervalo j; e N_{usi}^{gr} é o número de usinas pertencentes ao grupo.

Posteriormente, calcula-se o fator de participação de cada usina e para cada intervalo do dia, definido como a divisão da geração média da usina i no intervalo j pela geração média do grupo no intervalo j, ambas calculadas sobre os últimos 4 dias, conforme a Equação (5.28).

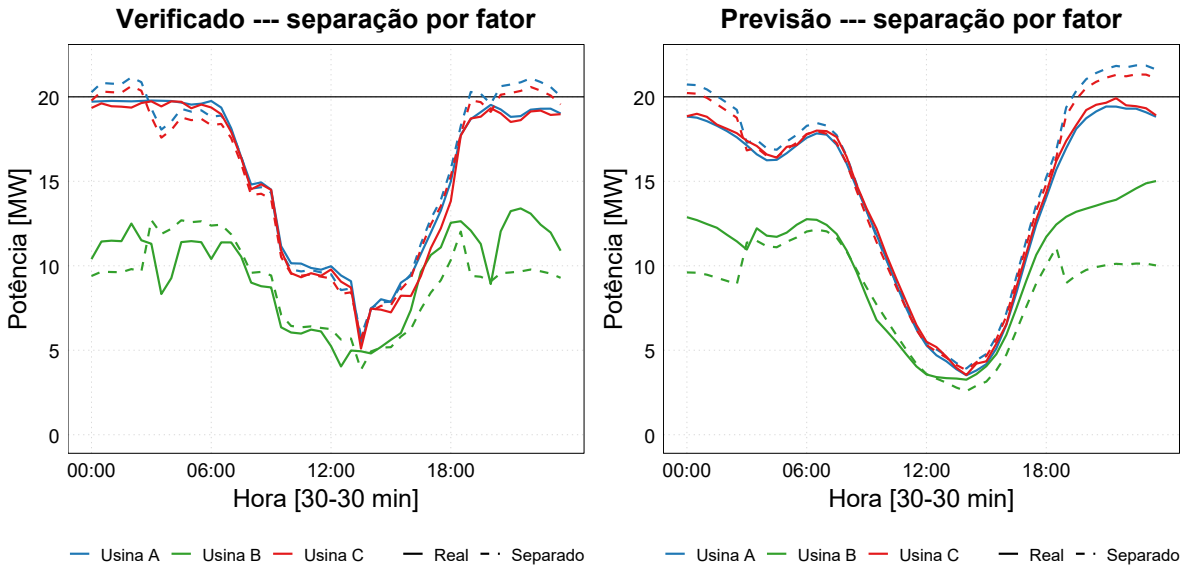
$$f_{i,j} = \frac{G_{v,i,j}}{G_{v,gr,j}} \tag{5.28}$$

Em que $f_{i,j}$ é o fator de separação da usina i no intervalo j. Sendo assim, cada usina terá 48 fatores de separação, um para cada intervalo de meia-hora do dia. Usinas com maior geração média em determinado horário recebem fatores maiores e, consequentemente, parcelas maiores da geração ou previsão do grupo naquele horário.

Na implementação prática da metodologia, algumas decisões adicionais foram adotadas para conferir maior robustez ao cálculo dos fatores. Em primeiro lugar, o cálculo da geração média na Equação (5.28) considera também intervalos vizinhos ao intervalo j, com o objetivo de suavizar a variação dos fatores entre intervalos consecutivos e reduzir a sensibilidade a oscilações pontuais. Em segundo lugar, caso a geração observada em algum intervalo da janela dos últimos 4 dias seja muito próxima de zero (tipicamente abaixo de 5% da potência instalada da usina), a janela é expandida para incluir dias passados adicionais com geração não nula, evitando que valores espúrios contaminem o cálculo. Por fim, caso a discrepância entre os fatores calculados pelas duas janelas (a janela padrão de 4 dias e a janela expandida) seja muito elevada (na implementação, superior a 150%), prevalece a janela expandida, considerada mais conservadora.

Há casos, porém, em que a aplicação direta desses fatores pode resultar em uma separação que extrapola a potência instalada de alguma usina. Isso ocorre tipicamente quando o fator de uma usina sofreu variações nos últimos dias, por exemplo, devido a paradas operativas, manutenções programadas ou indisponibilidades não previstas em outras usinas do grupo, que alteraram temporariamente a participação relativa de cada usina. A Figura 64 ilustra essa situação. Para fins de exemplificação, os fatores das Usinas A e C foram artificialmente elevados nos intervalos de início e final do dia (horários em que a geração do grupo é elevada), enquanto o fator da Usina B foi reduzido proporcionalmente para manter a soma dos fatores igual a um.

Figura 64 – Separação das previsões por usinas com violação da potência instalada.



Fonte: Elaboração própria.

Nesta figura, observa-se que as linhas tracejadas correspondentes à separação das Usinas A e C ultrapassam a linha horizontal preta (que representa a potência instalada de 20 MW) tanto no painel esquerdo, referente à série verificada, quanto no painel direito, referente à previsão. Cabe destacar que esse comportamento ocorre apenas na separação calculada, a geração verificada real das usinas (linhas contínuas) em geral não viola a potência instalada, uma vez que constitui um valor fisicamente observado e limitado pela própria capacidade do equipamento. Quando isso ocorre, as diferenças são pequenas, pois uma usina pode gerar ligeiramente acima da potência instalada cadastrada (tolerância de 10% adotada na limpeza dos dados, Seção 3.2.1).

Quando essa situação ocorre, é necessário redefinir os fatores. Para isso, é realizado um processo iterativo para que as seguintes condições sejam satisfeitas, expressas nas Equações (5.29) e (5.30).

$$G_{v,gr,k} \cdot f_{i,k} \leq P_{inst,i} \tag{5.29}$$

$$G_{p,gr,k} \cdot f_{i,k} \leq P_{inst,i} \tag{5.30}$$

Em que $G_{p,gr,k}$ é a previsão de geração do grupo no intervalo k; e $P_{inst,i}$ é a potência instalada da usina i. Em outras palavras, a separação da geração total do ponto em cada usina, assim como da previsão total do ponto em cada usina, não pode violar a potência instalada de nenhuma das usinas.

Para a primeira condição, referente à geração verificada, observa-se a existência de um intervalo k em que $G_{v,gr,k} \cdot f_{i,k} > P_{inst,i}$ e cria-se um fator de redução, de forma que o valor da geração individual calculada pelo fator seja igual à potência instalada, conforme a Equação (5.31).

$$\varphi_{v,\Omega,k} = \frac{P_{inst,\Omega}}{G_{v,gr,k} \cdot f_{\Omega,k}} \tag{5.31}$$

Em que Ω representa o conjunto de usinas cuja geração calculada viola a potência instalada; $\varphi_{v,\Omega,k}$ é o fator de redução das usinas do conjunto Ω, calculado a partir da geração verificada; $P_{inst,\Omega}$ é a potência instalada das usinas do conjunto Ω; e $f_{\Omega,k}$ é o fator de separação original do conjunto de usinas Ω no intervalo k.

Esse fator de redução é então utilizado para recalcular o fator de separação das usinas do conjunto Ω, conforme a Equação (5.32).

$$f'_{\Omega,k} = f_{\Omega,k} \cdot \varphi_{v,\Omega,k} = \frac{P_{inst,\Omega}}{G_{v,gr,k}} \tag{5.32}$$

Para as demais usinas (conjunto Ψ, cujas separações não violam a restrição), os fatores são reespecificados de modo a preservar a soma unitária, conforme a Equação (5.33).

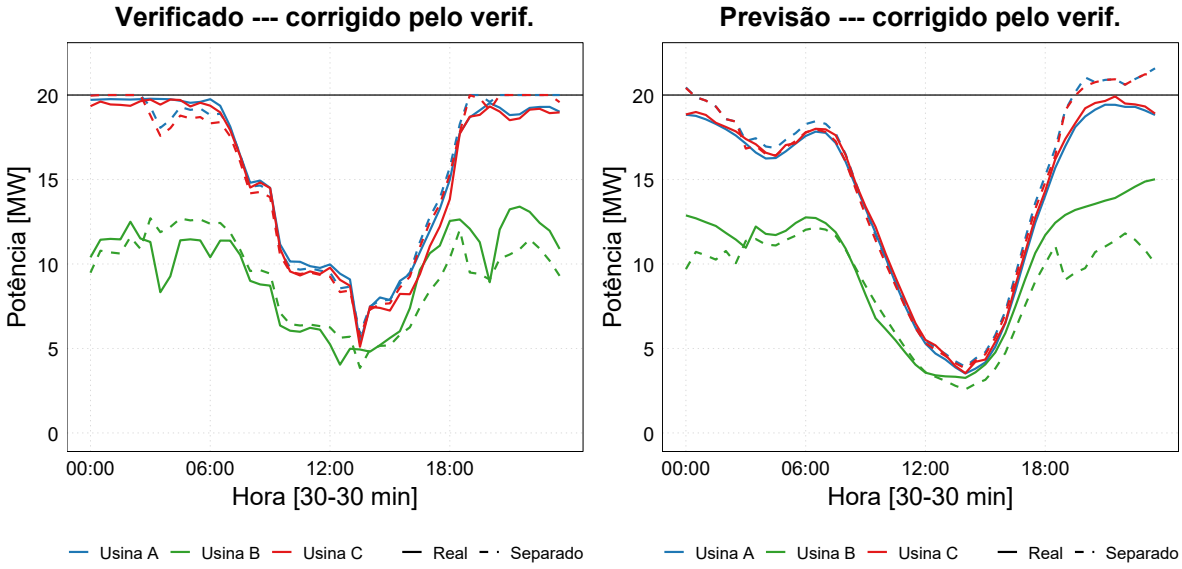
$$f'_{\Psi,k} = f_{\Psi,k} \cdot \frac{1 - \sum_{i \in \Omega} f'_{i,k}}{\sum_{i \in \Psi} f_{i,k}} \tag{5.33}$$

Em que Ψ representa o conjunto de usinas cuja geração calculada não viola a potência instalada. Por construção, $\sum_{i \in \Psi} f'_{i,k} + \sum_{i \in \Omega} f'_{i,k} = 1$, como esperado. Todavia, quando os fatores das usinas do conjunto Ω são reduzidos, ocorre um aumento dos fatores das usinas do conjunto Ψ. Podem então surgir casos em que usinas que antes não violavam passam a violar. Por essa razão, o processo é repetido iterativamente até que todas as usinas estejam dentro de seus limites.

A Figura 65 ilustra a aplicação do processo iterativo de redistribuição dos fatores, considerando apenas a primeira condição, relativa à geração verificada.

Observa-se, no painel esquerdo da figura, que após a correção a separação da geração verificada (linhas tracejadas) das três usinas respeita o limite da potência instalada. Entretanto, o painel direito mostra que a separação da previsão ainda pode violar a potência instalada

Figura 65 – Separação das previsões por usinas após a correção pela geração verificada.



Fonte: Elaboração própria.

em algumas usinas. Isso ocorre porque, embora o fator tenha sido corrigido para satisfazer a condição da geração verificada, ele é aplicado também à previsão do grupo, que pode ser maior do que o verificado em determinados intervalos.

De forma análoga, para a segunda condição, referente à previsão, verifica-se o intervalo k tal que $G_{p,gr,k} \cdot f_{i,k} > P_{inst,i}$. Cria-se um fator de redução, conforme a Equação (5.34).

$$\varphi_{p,\Omega,k} = \frac{P_{inst,\Omega}}{G_{p,gr,k} \cdot f_{\Omega,k}} \tag{5.34}$$

Em que $\varphi_{p,\Omega,k}$ é o fator de redução calculado a partir da previsão de geração do grupo. O fator das usinas em Ω é então recalculado conforme a Equação (5.35).

$$f'_{\Omega,k} = f_{\Omega,k} \cdot \varphi_{p,\Omega,k} = \frac{P_{inst,\Omega}}{G_{p,gr,k}} \tag{5.35}$$

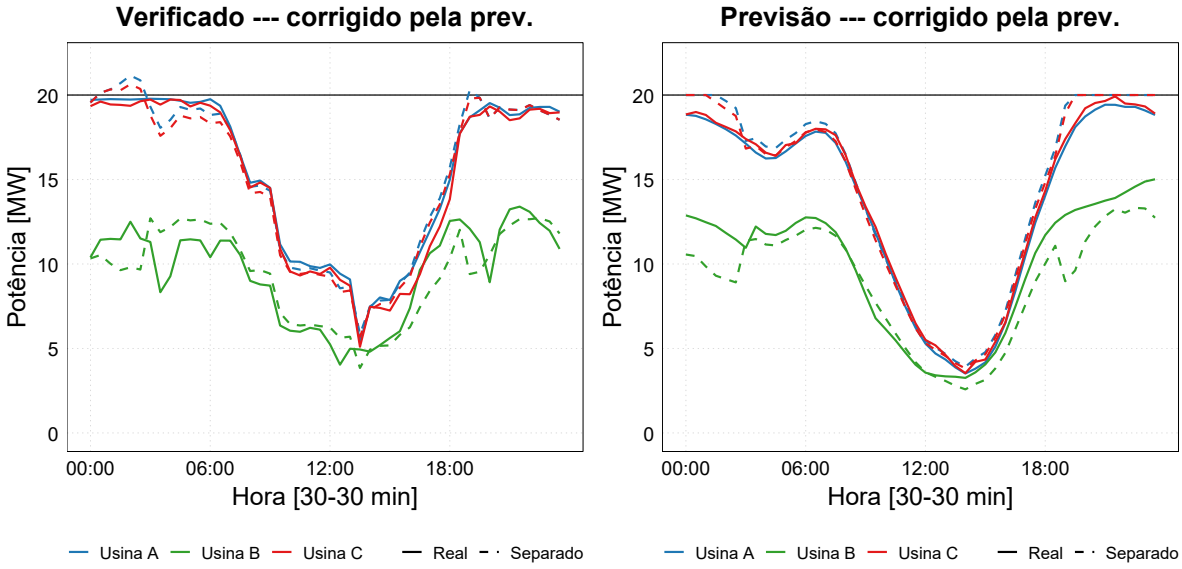
Para as demais usinas, os fatores são reespecificados conforme a Equação (5.36).

$$f'_{\Psi,k} = f_{\Psi,k} \cdot \frac{1 - \sum_{i \in \Omega} f'_{i,k}}{\sum_{i \in \Psi} f_{i,k}} \tag{5.36}$$

Como no caso anterior, o processo é iterado até que todas as usinas estejam dentro dos limites. A Figura 66 ilustra a aplicação dessa correção considerando apenas a previsão.

Nesse caso, o painel direito mostra que a separação da previsão respeita a potência instalada para as três usinas, mas o painel esquerdo evidencia que a separação da geração

Figura 66 – Separação das previsões por usinas após a correção pela geração prevista.



Fonte: Elaboração própria.

verificada ainda pode violar a potência instalada. O comportamento é simétrico ao caso anterior: a correção aplicada com base em uma das séries (verificada ou prevista) garante a aderência ao limite naquela série, mas não na outra.

Como cada correção isolada resolve apenas uma das duas restrições, pode ocorrer que a correção por previsão aumente o fator de algumas usinas e passe a violar as restrições da geração verificada, e vice-versa. Por essa razão, na metodologia adotada, as duas correções são feitas simultaneamente. Os passos anteriores serviram, portanto, apenas para ilustrar o efeito de cada correção isoladamente.

Quando existe um intervalo k tal que $G_{v,gr,k} \cdot f_{i,k} > P_{inst,i}$ e/ou $G_{p,gr,k} \cdot f_{i,k} > P_{inst,i}$, criam-se dois fatores de redução, conforme as Equações (5.37).

$$\varphi_{v,\Omega,k} = \frac{P_{inst,\Omega}}{G_{v,gr,k} \cdot f_{\Omega,k}} \qquad \text{e} \qquad \varphi_{p,\Omega,k} = \frac{P_{inst,\Omega}}{G_{p,gr,k} \cdot f_{\Omega,k}} \tag{5.37}$$

Em que Ω representa o conjunto de usinas cuja previsão e/ou geração calculada viola a potência instalada. O fator de separação é então recalculado utilizando-se o menor entre os dois fatores de redução, $\varphi_{v,\Omega,k}$ e $\varphi_{p,\Omega,k}$, conforme a Equação (5.38). A escolha pelo menor dos dois fatores de redução garante que a redução aplicada seja suficiente para satisfazer simultaneamente as duas restrições, ainda que possa ser mais conservadora do que cada correção isoladamente.

$$f'_{\Omega,k} = f_{\Omega,k} \cdot \min\left\{ \varphi_{v,\Omega,k}, \varphi_{p,\Omega,k} \right\} \tag{5.38}$$

Para as demais usinas, os fatores são reespecificados conforme a Equação (5.39).

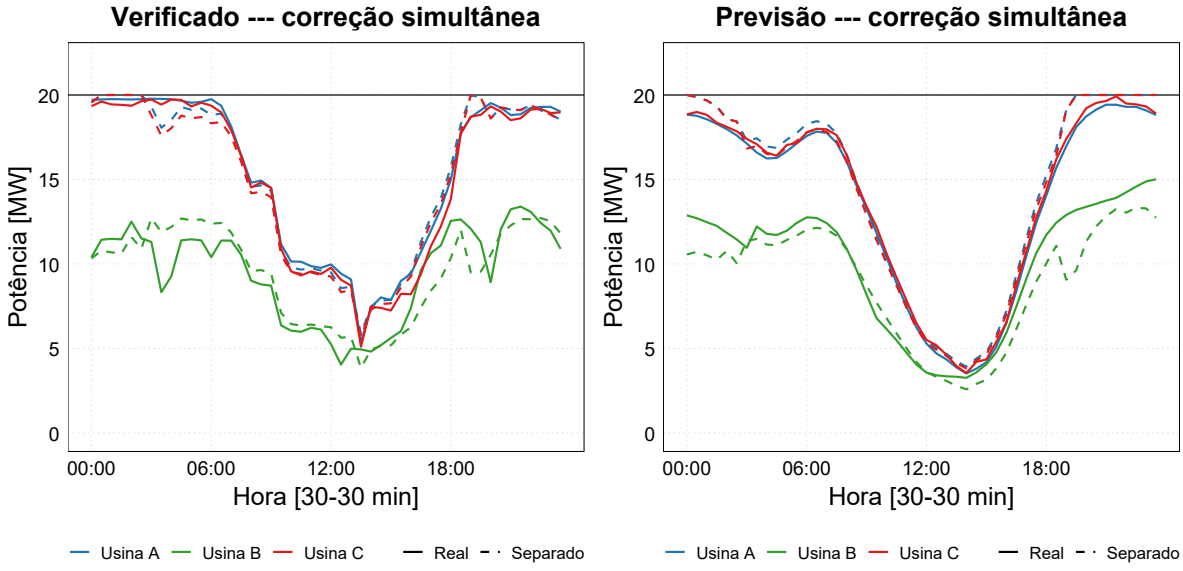
$$f'_{\Psi,k} = f_{\Psi,k} \cdot \frac{1 - \sum_{i \in \Omega} f'_{i,k}}{\sum_{i \in \Psi} f_{i,k}} \tag{5.39}$$

Em que Ψ representa o conjunto de usinas cuja previsão e/ou geração calculada não viola a potência instalada. Como nos dois casos anteriores, podem ocorrer usinas que antes não violavam e passaram a violar após a redistribuição. Por essa razão, o processo é repetido iterativamente até que todas as usinas estejam dentro dos limites para ambas as séries simultaneamente.

A convergência do processo iterativo é assegurada pelo fato de que, a cada iteração, os fatores das usinas violantes são estritamente reduzidos para valores compatíveis com a potência instalada. A soma total das violações (isto é, o excesso entre a separação calculada e a potência instalada, somado ao longo das usinas violantes) decresce monotonicamente até atingir zero em um número finito de iterações. Como salvaguarda adicional, na implementação adota-se um limite máximo de iterações por intervalo, raramente atingido na prática.

A Figura 67 ilustra o resultado da aplicação da correção simultânea.

Figura 67 – Separação das previsões por usinas após a correção simultânea pela geração verificada e pela previsão.



Fonte: Elaboração própria.

Observa-se que, após a correção simultânea, tanto a separação da geração verificada (painel esquerdo) quanto a separação da previsão (painel direito) respeitam a potência instalada das três usinas ao longo de todo o dia. Após finalizadas todas as etapas anteriores, é possível obter a previsão de curtíssimo prazo para todos os usos demandados pelos modelos de planejamento e pelo ONS, sejam elas com maiores ou menores discretizações espaciais e/ou temporais.

5.9 Particularidades da metodologia de previsão de geração de curto prazo

Conforme mencionado anteriormente, a geração eólica é representada nos modelos de planejamento de curto prazo (MPCP, atualmente o DECOMP) e médio prazo (MPMP, atualmente o NEWAVE) de forma determinística, por meio de uma estimativa baseada na REN nº 843 [29], atualizada pela REN nº 1.032/2022 da ANEEL [136]. Essa estimativa é construída a partir de médias históricas da geração verificada dos parques eólicos.

A Figura 68 apresenta a comparação entre a geração eólica média mensal estimada pela resolução e os dados verificados para o subsistema Nordeste, no período de janeiro de 2017 a dezembro de 2025. A análise da figura revela três regimes distintos de comportamento ao longo da série histórica. No primeiro período, entre 2017 e 2020, observa-se uma tendência de superestimação da geração eólica pelos critérios da resolução, indicada pelos círculos verdes na figura. Esse desvio sistemático sugeria, à época, a viabilidade de modelos de correção capazes de aproximar os valores previstos aos verificados.

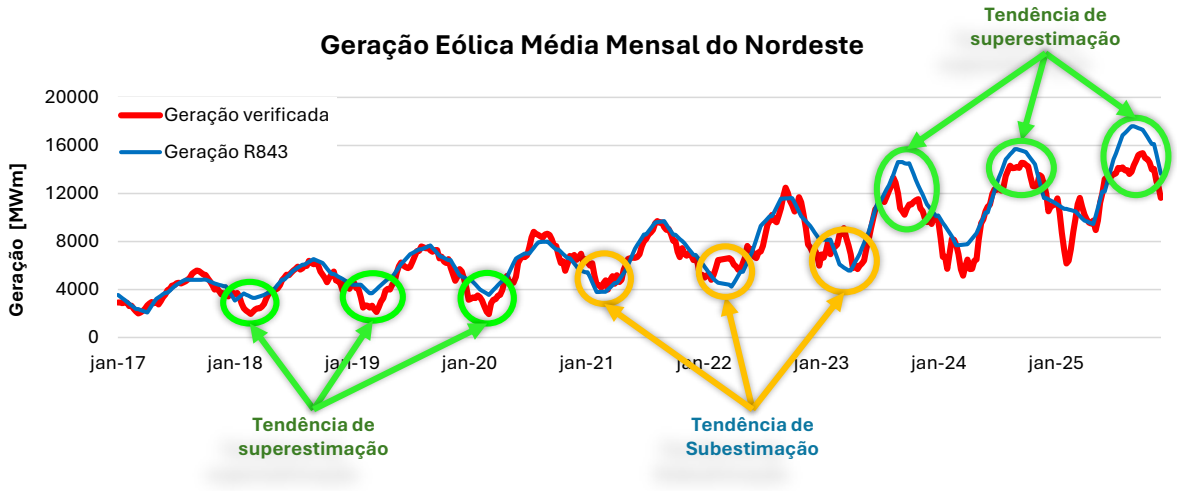
Entretanto, entre 2021 e 2023, ocorre uma inversão no comportamento do desvio: a geração eólica passa a ser subestimada pelos mesmos critérios, conforme evidenciado pelos círculos amarelos. Essa alternância entre super e subestimação compromete a construção de modelos simples de correção baseados apenas em padrões históricos do desvio. A partir de meados de 2023, a estimativa da resolução retorna a apresentar superestimação, desta vez em proporção ainda maior. Esse comportamento possui uma explicação física: nesse período, o *curtailment*, corte de geração por restrições operativas, passou a crescer expressivamente no Nordeste, fazendo com que a geração verificada se afaste sistematicamente para baixo da geração potencial. Nesse contexto, o referencial correto para a construção de modelos de previsão deixa de ser a geração verificada real e passa a ser a geração estimada sem *curtailment*, conforme apresentado anteriormente neste documento.

Dois trabalhos decorrentes desta pesquisa foram publicados e serviram como resultados iniciais para embasar a metodologia final aqui apresentada. O primeiro consistiu em uma tentativa de corrigir os desvios sistemáticos na estimativa da REN nº 843 [29] para uso em estudos prospectivos do ONS [37]. O segundo trabalho teve como foco a geração de previsões para o horizonte da primeira semana operativa, cobrindo a maioria das usinas eólicas representadas no MPCP (atualmente o DECOMP); com o objetivo de aumentar a precisão nesse período, foi proposto um modelo capaz de fornecer previsões com até uma semana de antecedência, cuja metodologia específica é detalhada em [40].

5.9.1 Estrutura básica do modelo

Antes de detalhar o modelo de previsão para o horizonte de curto prazo, é importante destacar alguns aspectos relacionados ao horizonte de curtíssimo prazo. Nesse contexto, cabe uma breve explicação sobre como os tipos de usinas são considerados nos modelos de previsão

Figura 68 – Geração eólica média mensal do Nordeste dos últimos anos, estimada pela resolução e verificada.



Fonte: Elaboração própria.

e planejamento. O MPCTP, considera como entrada de geração apenas as usinas em operação comercial e supervisionadas pelo ONS, usinas denominadas tipo I e II, definidas no Submódulo 26.2 dos Procedimentos de Rede [135]. Portanto, o modelo de previsão de curtíssimo prazo, apresentado neste estudo, engloba exclusivamente esse conjunto de usinas. As usinas do tipo III, não supervisionadas pelo ONS, são indiretamente consideradas no MPCTP, por meio da redução na carga prevista.

Atualmente, a estimativa da previsão de geração dessas usinas do tipo III é realizada com base na média aritmética da geração passada. No entanto, essa abordagem mostra um desempenho questionável devido à alta variabilidade na geração dessa fonte, que não é capturada pela simples média dos dados históricos. Como resultado desta pesquisa, foram desenvolvidas metodologias capazes de estimar, com certa precisão, a geração futura dessas usinas. Esse trabalho demonstrou resultados promissores, detalhados em [36].

Para o planejamento no horizonte de curto prazo, no entanto, as usinas tipo I, II e III e futuras (com previsão de entrada em operação dentro do horizonte do primeiro mês operativo em questão), são consideradas como fonte de geração diretamente no MPCP.

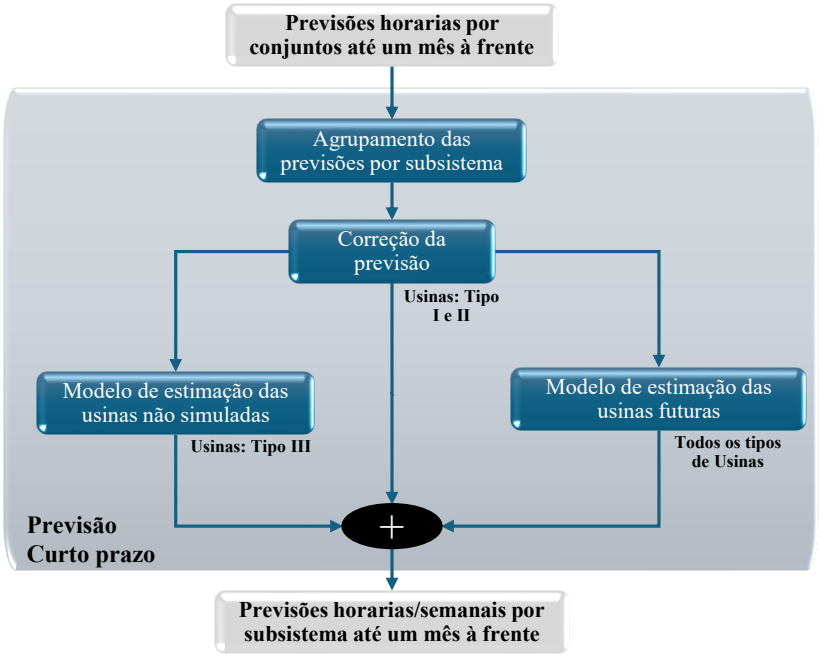
Dada a dificuldade em obter dados observados de vento das usinas do tipo III, a aplicação das curvas V $\times$ P não é viável, ao contrário do que ocorre no modelo de curtíssimo prazo. Essa mesma limitação se estende às usinas futuras, que carecem de históricos confiáveis de dados observados. Assim foi necessário desenvolver um modelo que também gera as previsões para esse montante de usinas. O modelo desenvolvido pode ser separado em três blocos, que correspondem ao montante de:

- usinas supervisionadas do tipo I e II (Bloco 1);

- usinas não supervisionadas, tipo III (Bloco 2);
- usinas novas, em operação comercial mas que ainda não possuem históricos suficientes, e futuras que entrarão durante o primeiro mês (Bloco 3).

A Figura 69 ilustra os três blocos do modelo desenvolvido para previsão de curto prazo (horizonte do primeiro mês operativo). Na figura é destacada a conexão desse modelo com o modelo de previsão de curtíssimo prazo e os modelos de planejamento de curto e médio prazo.

Figura 69 – Estrutura do modelo de curto prazo.



Fonte: Elaboração própria.

5.9.2 Metodologia proposta para o bloco 1

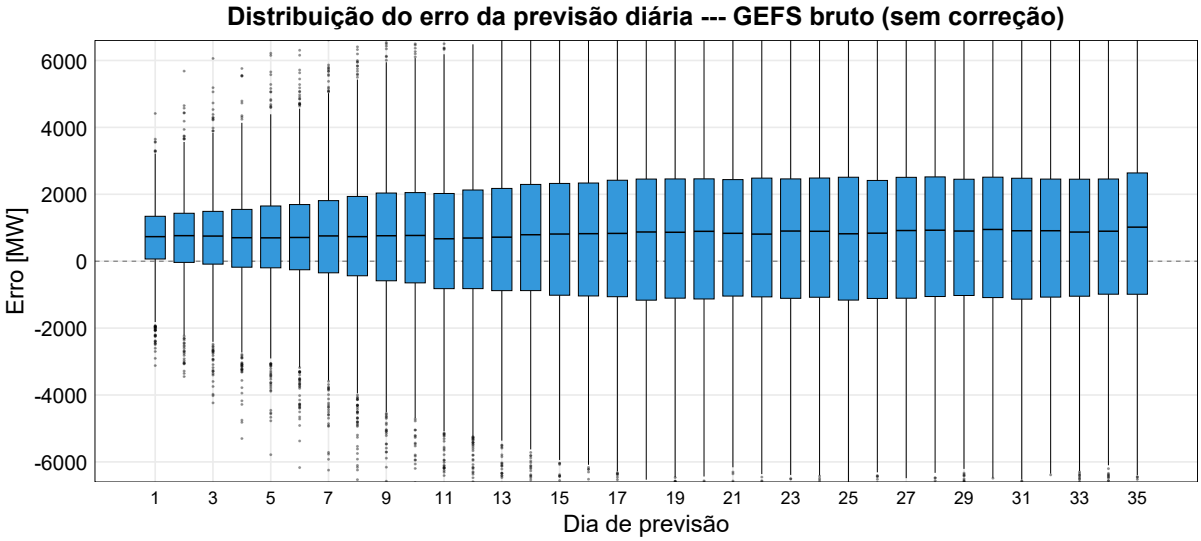
Conforme mencionado anteriormente, para aplicação na cadeia de modelos energéticos de curto prazo, as previsões de geração devem ser agregadas espacialmente em nível de subsistema e temporalmente em semanas operativas. Além disso, as previsões semanais precisam ser agregadas de acordo com os patamares de carga pesada, média e leve, de forma compatível com a representação adotada pelos modelos energéticos. Os intervalos de duração de cada patamar de carga utilizados pela cadeia de modelos podem ser consultados em [159].

Ao realizar o agrupamento espacial em nível de subsistema, verificou-se, de forma distinta do esperado, a presença de um viés sistemático significativo nas previsões produzidas pelo modelo de geração em relação aos valores verificados, conforme ilustrado na Figura 70. Essa figura apresenta a distribuição dos erros diários de previsão (verificado − previsto) ao longo dos 35 dias do horizonte de previsão, considerando a geração agregada do subsistema Nordeste

obtida a partir do modelo GEFS em sua forma bruta, ou seja, sem qualquer procedimento de correção.

Observa-se que, para todos os horizontes de previsão, a mediana dos erros permanece sistematicamente acima de zero, evidenciando uma tendência consistente de subestimação da geração prevista em relação à geração verificada. Além disso, nota-se que a magnitude dos erros aumenta à medida que o horizonte de previsão se amplia, comportamento compatível com o crescimento da incerteza inerente às previsões meteorológicas em horizontes mais longos.

Figura 70 – Distribuição do erro da previsão diária — GEFS bruto (sem correção).



Fonte: Elaboração própria.

A presença desse viés sistemático na previsão de geração é inicialmente contraintuitiva, uma vez que a etapa de correção do vento previsto descrita anteriormente foi especificamente construída para minimizar vieses na variável de entrada do modelo de conversão. A análise das possíveis causas dessa discrepância permitiu identificar a não-linearidade da curva de conversão V $\times$ P como a principal fonte do problema, complementada por outros fatores secundários discutidos a seguir.

A causa principal está associada à desigualdade de Jensen [160], que estabelece que, para uma função não-linear φ e uma variável aleatória V, em geral $E[\varphi(V)] \neq \varphi(E[V])$. No caso da geração eólica, a função φ corresponde à curva V $\times$ P, tipicamente sigmoide ou logística, com regiões de *cut-in*, transição não-linear, saturação e *cut-out*. Mesmo que a previsão do vento tenha viés nulo em média, $E[V_p] = E[V_v]$, a aplicação da curva V $\times$ P a cada série pode resultar em valores médios distintos de geração quando as duas séries possuem variâncias diferentes.

Esse efeito é particularmente relevante porque as previsões meteorológicas, mesmo após a correção do viés, tendem a apresentar uma variância menor que a série verificada. Isso ocorre porque os modelos meteorológicos representam médias sobre células de grade de dimensões espaciais expressivas, o que naturalmente filtra parte da variabilidade temporal observada no

vento real. Como consequência, os picos da série de vento previsto são atenuados em relação aos da série verificada, assim como os vales, ou seja, a previsão produz uma série mais "suavizada".

O efeito dessa suavização sobre a geração não é simétrico, devido ao formato característico da curva V $\times$ P. Em ventos baixos, próximos da região de *cut-in* (tipicamente em torno de 3 m/s), pequenas variações no vento produzem mudanças abruptas na geração: abaixo do *cut-in* a turbina não gera, e acima dela passa a gerar segundo a região não-linear da curva, onde a potência depende aproximadamente do cubo da velocidade do vento ($P \propto V^3$). Assim, quando a previsão suaviza vales de vento abaixo do *cut-in*, "elevando-os" artificialmente para valores acima do *cut-in*, a curva V $\times$ P passa a gerar potência onde, na realidade, a geração foi nula. Esse efeito tende a produzir uma superestimação da geração nesses momentos.

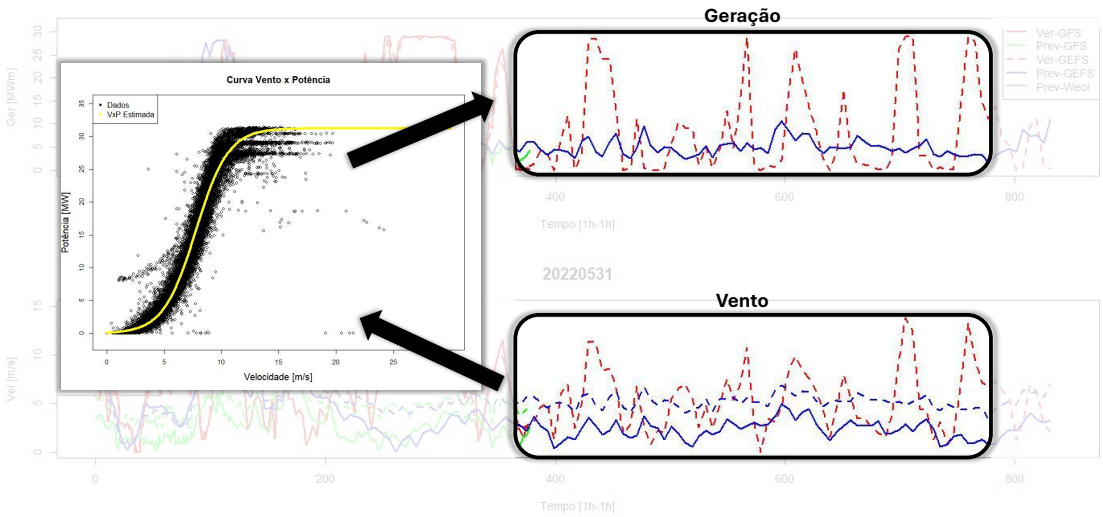
Na direção oposta, em ventos elevados, a curva V $\times$ P satura em torno da potência nominal das turbinas, e variações no vento têm pouco efeito sobre a geração. Quando a previsão atenua picos de vento, a perda associada à curva V $\times$ P é pequena, mas existe e tende a produzir subestimação da geração nesses instantes.

O efeito líquido sobre o viés agregado depende da distribuição dos ventos da usina e da forma específica da curva V $\times$ P. No caso dos dados analisados para o subsistema Nordeste, a subestimação observada na Figura 70 é compatível com um regime em que a maior parte dos ventos opera na região ascendente da curva (próxima da saturação), onde a perda nos picos atenuados não é totalmente compensada pelo ganho na "elevação" dos vales. Cabe ressaltar, no entanto, que em outras discretizações espaciais ou regimes de vento o efeito pode ser inverso, com predominância da superestimação, o sinal do viés líquido não é universal, mas o mecanismo subjacente da não-linearidade da curva é o mesmo.

A Figura 71 ilustra esse mecanismo. No painel inferior, a linha azul cheia corresponde à previsão de vento bruta, e a linha azul tracejada à previsão de vento já ajustada pelo modelo de correção de viés do vento; a linha vermelha representa a velocidade verificada. Observa-se que a previsão ajustada, apesar de ter média alinhada à verificada, apresenta variabilidade visivelmente menor, comportamento consistente com o efeito de suavização discutido anteriormente. No painel superior, são apresentados os perfis de geração previsto e verificado correspondentes, evidenciando os picos significativos da geração verificada que não são reproduzidos pela geração prevista. Na figura central, a curva V $\times$ P do conjunto de usinas analisado é destacada: observa-se que, ao dobrar a velocidade de 5 m/s para 10 m/s, a geração aumenta aproximadamente cinco vezes, expressão direta da não-linearidade que produz os picos de potência verificados e que, na média semanal, contribuem para o viés sistemático observado.

Além do efeito principal associado à não-linearidade da curva, dois fatores secundários podem contribuir para o viés observado. O primeiro refere-se à calibração da própria curva V $\times$ P, que é estimada empiricamente a partir do histórico de operação de cada conjunto de usinas. Eventuais desatualizações dessa curva (decorrentes da entrada de novas turbinas, do envelhecimento natural dos equipamentos (com a consequente redução gradual de eficiência),

Figura 71 – Perfis da geração e velocidade do vento verificadas e previstas, e curva V $\times$ P do conjunto de usinas analisado.



Fonte: Elaboração própria.

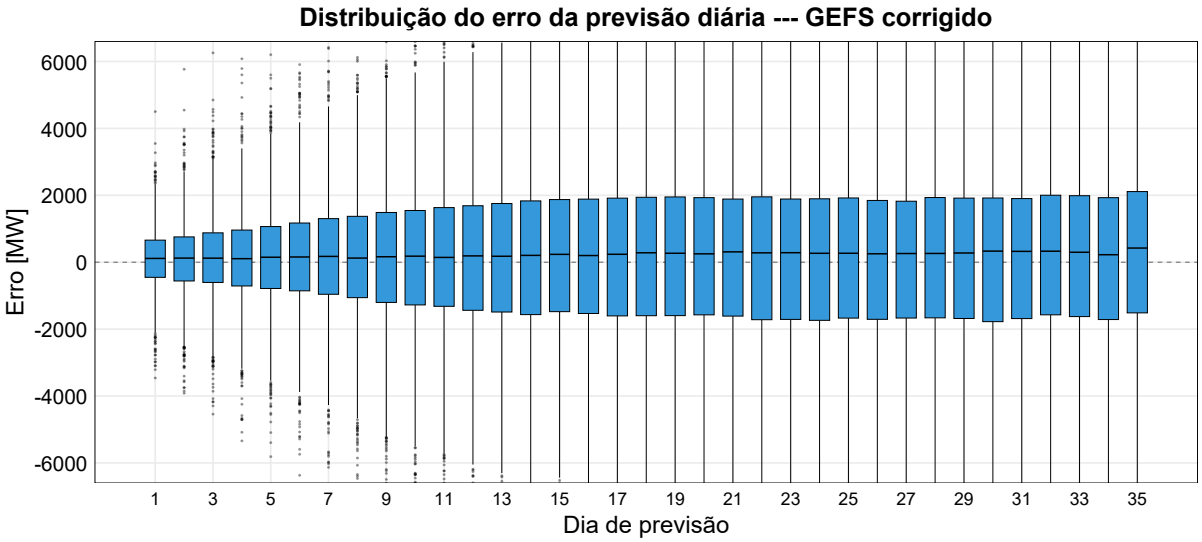
de mudanças na configuração do parque ou de variações de densidade do ar não consideradas) podem introduzir viés sistemático mesmo com previsão de vento perfeita. O segundo refere-se a indisponibilidades e restrições operacionais, como manutenções programadas ou restrições de transmissão (*curtailment*), que reduzem a geração efetiva em relação ao que a curva V $\times$ P indicaria a partir do vento previsto. Esses fatores tendem a produzir superestimação da previsão, em direção oposta ao viés agregado observado nesta análise, mas podem ser relevantes em discretizações espaciais mais finas, onde indisponibilidades localizadas têm maior impacto relativo.

Para mitigar o viés sistemático identificado, foi implementada uma etapa adicional de ajuste, baseada em regressões lineares aplicadas às previsões de geração agregadas a nível de subsistema. Essa correção opera diretamente sobre a saída do modelo de conversão e captura o viés residual remanescente após a correção do vento. A Figura 72 apresenta a distribuição dos erros diários da previsão de geração após a aplicação dessa correção. Observa-se que as medianas dos erros passam a se distribuir mais próximas de zero ao longo de todo o horizonte de previsão, com redução também na dispersão dos erros, evidenciando a efetividade da correção proposta na mitigação do viés sistemático.

5.9.3 Estimativa da geração eólica para o bloco 2

Pela indisponibilidade de vento verificado das usinas do tipo III, não é possível levantar as curvas V $\times$ P dessas usinas, dessa forma, não é possível utilizar a metodologia anterior. Todavia, os valores de geração dessas usinas são observados pelo SMF da CCEE, logo essas

Figura 72 – Distribuição do erro da previsão diária — GEFS após a correção do viés.



Fonte: Elaboração própria.

informações são utilizadas para criar um modelo de estimação da previsão dessas usinas tipo III.

A estimativa da previsão de geração para usinas tipo III é calculada através de uma regressão linear, obtida relacionando-se dados da CCEE do montante de geração verificada para todas as usinas (tipos I, II e III) de cada subsistema com o montante de geração verificado de todas as usinas simuladas pelo modelo de curtíssimo prazo (tipos I e II), conforme a Equação (5.40).

$$G_{v,CCEE,gr\,1,2,3} = \alpha_{gr} \cdot G_{v,CCEE,gr\,1,2} + \beta_{gr} \tag{5.40}$$

Em que $G_{v,CCEE,gr\,1,2,3}$ representa o montante de geração verificada para todas as usinas dos tipos I, II e III; $G_{v,CCEE,gr\,1,2}$ o montante de geração verificado de todas as usinas dos tipos I e II; e α_{gr} e β_{gr} os coeficientes angular e linear da regressão que relaciona os dois grupos, respectivamente.

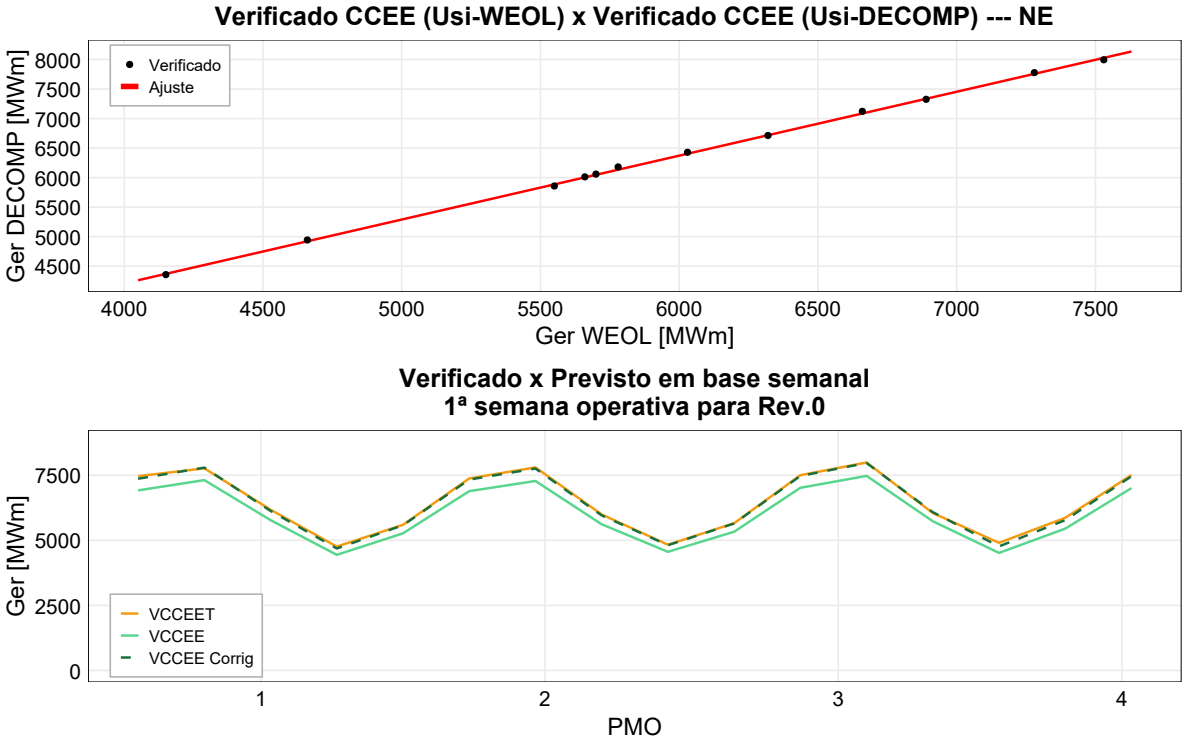
Ressalta-se que são utilizados no ajuste das retas os dados de geração do SMF considerando o período de um ano, sendo que os dados são agregados em base mensal e por patamar de carga. Os coeficientes estimados para um subsistema são únicos para todos os patamares de carga e são aplicados às previsões do Bloco 1 de cada semana.

A Figura 73 ilustra a aplicação da metodologia para o subsistema Nordeste. No painel superior, os pontos pretos representam os pares de geração verificada CCEE considerando apenas as usinas simuladas pelo modelo de curtíssimo prazo (eixo horizontal, usinas dos tipos I e II) e considerando o total das usinas em operação no subsistema (eixo vertical, usinas dos tipos I, II e III). A linha vermelha corresponde à reta de regressão ajustada conforme a Equação (5.40). Observa-se que os pontos verificados se distribuem essencialmente sobre a reta, indicando que a relação entre as duas grandezas é bem capturada por um ajuste linear.

No painel inferior, são apresentadas três séries em base semanal: a geração verificada

total da CCEE (curva em laranja, denominada VCCEET), a geração verificada apenas das usinas simuladas pelo modelo de curtíssimo prazo (curva em verde claro, denominada VCCEE) e a geração corrigida obtida pela aplicação da regressão linear sobre a curva VCCEE (curva em verde escuro tracejada, denominada VCCEE Corrig). Observa-se que a curva VCCEE Corrig praticamente se sobrepõe à curva VCCEET, evidenciando a qualidade do ajuste proposto na reconstituição do montante total de geração a partir da parcela simulada.

Figura 73 – Ajuste da regressão linear para inclusão da geração das usinas do tipo III — Subsistema Nordeste.



Fonte: Elaboração própria.

O mesmo processo é utilizado para estimar a geração prevista das usinas eólicas não simuladas pelo modelo de curtíssimo prazo, isto é, a regressão linear ajustada é aplicada ao montante previsto para as usinas do Bloco 1 que foi agregado em base semanal por patamar de carga, resultando em uma previsão total da geração das usinas em operação consideradas no MPCP. A parcela não simulada pode ser calculada como a diferença entre a previsão total e a parcela correspondente ao Bloco 1.

5.9.4 Estimativa da geração eólica para o bloco 3

A previsão de geração correspondente ao conjunto de usinas futuras e de usinas que entraram recentemente em operação comercial, e que ainda não possuem histórico suficiente para serem incorporadas à metodologia do Bloco 1, é realizada por meio de uma abordagem baseada em similaridade com usinas já em operação. Na metodologia proposta, são identificadas

as cinco usinas geograficamente mais próximas e calculados fatores de proporcionalidade entre a potência instalada da usina em análise e a potência instalada de cada uma dessas usinas de referência.

As previsões semanais das usinas existentes mais próximas já são conhecidas, uma vez que são usinas simuladas no bloco 1. Dessa forma, a estimativa de previsão da usina futura é obtida através da equação:

$$G_{pfut} = \frac{1}{5} \cdot \sum_{i=1}^{5} f_i \cdot GP_i \tag{5.41}$$

em que G_{pfut} é a previsão semanal de geração eólica por patamar de carga para dada usina futura; GP_i é a previsão semanal de geração por patamar de carga da i-ésima usina mais próxima; f_i é o fator que relaciona a potência instalada da usina futura e da i-ésima usina mais próxima.

Por fim, as previsões consideradas no MPCP consistem no somatório das parcelas previstas para as usinas do tipo I e II (bloco 1), somadas às usinas do tipo III (bloco 2) e às usinas futuras com entrada nas semanas operativas do primeiro mês (bloco 3). Ressalta-se mais uma vez que os valores disponibilizados para o MPCP são as previsões semanais por patamar de carga e agregadas por subsistema.

O MPCP (atualmente o DECOMP), desempenha um papel crucial no Planejamento da Operação de Curto Prazo, projetando dois meses à frente, com uma discretização semanal para o primeiro mês e mensal para o segundo. A metodologia proposta neste trabalho traz melhorias para o primeiro mês, enquanto o segundo mês é estimado conforme a REN nº 1.032/2022 [136].

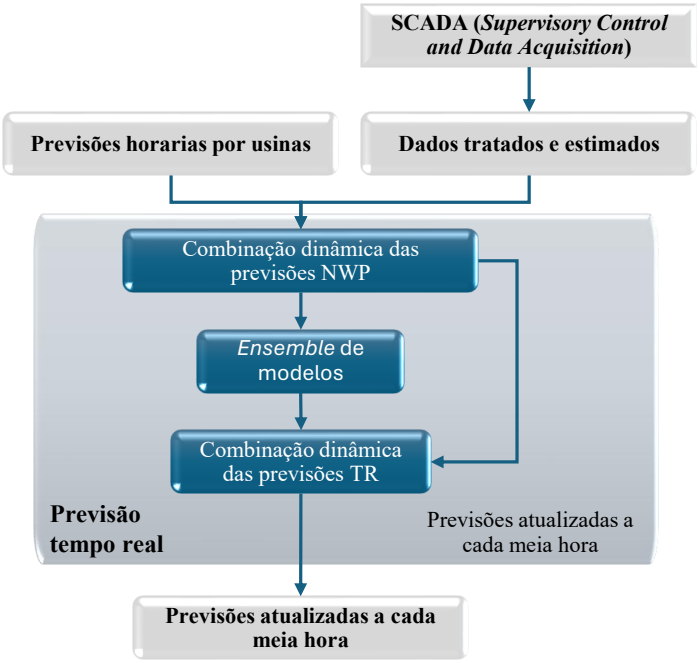
5.10 Modelo de previsão de geração em tempo real

As previsões horárias e semanais de geração, seja em nível de usina, grupo ou subsistema, são produzidas apenas uma vez por dia. Entretanto, a operação em tempo real do SIN exige previsões com maior frequência de atualização, capazes de incorporar rapidamente as condições mais recentes da geração observada e, consequentemente, refletir com maior precisão o estado atual do sistema.

Para atender a esse requisito, foi desenvolvido um modelo de previsão em tempo real, cujo fluxo metodológico é apresentado na Figura 74 e descrito nas seções a seguir.

O modelo recebe como entrada as previsões horárias por usinas, com horizonte de até um mês à frente, geradas pelo modelo de previsão de curto prazo, além da geração verificada adquirida em tempo real por meio do SCADA. A partir dessas entradas, o modelo é estruturado em três etapas encadeadas. Na primeira, é realizada uma combinação dinâmica das previsões dos modelos numéricos de previsão do tempo (NWP), na qual os pesos atribuídos a cada modelo meteorológico são continuamente recalculados com base em seu desempenho recente frente à geração observada, produzindo uma previsão de referência denominada NWP combinado.

Figura 74 – Fluxograma do modelo de previsão de geração eólica em tempo real.



Fonte: Elaboração própria.

Na segunda etapa, um conjunto de modelos de previsão de naturezas distintas, o *ensemble* de modelos, é executado em paralelo. Esse conjunto reúne, deliberadamente, representantes das principais famílias da literatura de previsão eólica: persistências, estatísticos clássicos (ARIMA e suavização exponencial), lineares (regressão), de aprendizado de máquina por *bagging* (*Random Forest*) e *boosting* (*XGBoost*), redes neurais (MLP) e um modelo de controle puramente temporal (SARIMA). Cada modelo gera, a cada emissão, uma previsão para todo o horizonte de tempo real.

Na terceira etapa, as previsões do *ensemble* são combinadas dinamicamente, atribuindo-se maior peso aos modelos que vêm apresentando melhor desempenho recente por horizonte, resultando na previsão final de tempo real. Essa combinação foi desenvolvida em duas variantes: uma baseada em pesos proporcionais ao inverso do erro absoluto médio (1/MAE) e outra baseada em um filtro de Kalman adaptativo. O resultado é um conjunto de previsões atualizadas a cada meia hora, que alimentam diretamente os processos de operação em tempo real do SIN.

As subseções a seguir descrevem as características gerais do modelo, os modelos que compõem o *ensemble* e as estratégias de combinação. O equacionamento completo de cada modelo e das combinações é apresentado no Apêndice A.

5.10.1 Características principais

Algumas premissas foram adotadas para o desenvolvimento do modelo de tempo real, tomando como objetivo atender às necessidades dos operadores na sala de controle para a

operação do SIN. Resumidamente, o propósito é fornecer aos operadores algum indicativo de como será a previsão nas próximas horas. Assim sendo, os principais pontos considerados foram:

(i) o horizonte de previsão pode se estender vários dias à frente, pois algumas decisões, como a programação de manutenção, podem ser tomadas pelos operadores junto aos agentes operadores dos parques eólicos; adota-se um horizonte de até 24 horas à frente, discretizado em 48 passos de 30 minutos;

(ii) o intervalo de atualização é de 30 minutos, utilizando informações da geração verificada proveniente do SCADA;

(iii) mesmo havendo mais rodadas dos NWPs por dia, será realizada apenas uma atualização das previsões meteorológicas por dia, pois esse modelo se alimenta das previsões de curtíssimo e curto prazo, e essas são rodadas uma única vez ao dia;

(iv) a forma de distribuição das previsões é por ponto de conexão da geração eólica à rede de transmissão, pois é nessa discretização espacial que os operadores observam o SIN.

Dado o exposto, o modelo de tempo real combina o poder preditivo das previsões meteorológicas, herdado dos modelos de curtíssimo e curto prazo, com a informação da geração verificada mais recente. Diferentemente de um único modelo de correção, adotou-se uma abordagem de *ensemble*, na qual diferentes modelos (de persistência, estatísticos e de aprendizado de máquina, em sua maioria associados às previsões de NWP e, portanto, de natureza híbrida) são executados em paralelo e posteriormente combinados. Essa estratégia cobre o espectro de famílias de modelos da literatura e permite que a combinação final aproveite, dinamicamente, o modelo mais adequado a cada horizonte e condição operativa. As subseções seguintes descrevem os modelos que compõem o *ensemble* e as estratégias de combinação.

5.10.2 Protocolo de treinamento e simulação

Os modelos foram avaliados por meio de uma simulação que reproduz a operação em tempo real. A cada emissão, realizada a cada 30 minutos, cada modelo dispõe apenas das informações disponíveis até aquele instante: a geração verificada até o instante de emissão e a rodada vigente do NWP, que é atualizada uma vez por dia, às 9 h (horário de Brasília); antes desse horário, utiliza-se a rodada do dia anterior. Como série verificada, adota-se a série tratada conforme o Capítulo 3, de modo que a simulação não reproduz as falhas de telemetria que ocorreriam em operação. A simulação abrange o período de 1 a 28 de janeiro de 2025.

Os modelos calibráveis são treinados de forma semanal e causal (*walk-forward*): no início de cada semana da simulação, cada modelo é reestimado apenas com dados anteriores a esse instante, e seus parâmetros permanecem fixos ao longo da semana. A janela de treinamento é de 365 dias para a regressão linear, o *Random Forest*, o *XGBoost* e a MLP, e de 30 dias para

o ARIMA, a suavização exponencial e o SARIMA, que modelam a dinâmica recente da série. Nestes três últimos, o estado interno do modelo é atualizado a cada emissão com os dados mais recentes, sem reestimar os parâmetros. Os alvos utilizados no treinamento terminam antes do instante de corte, de modo que nenhuma informação futura é empregada.

Os hiperparâmetros foram definidos de forma empírica, a partir de testes em alguns agrupamentos e períodos, nos quais diferentes valores foram avaliados quanto ao desempenho, e mantidos fixos, com a mesma configuração, para todos os agrupamentos. Não foi utilizado um conjunto de validação específico nem parada antecipada: o desempenho é avaliado integralmente fora da amostra, no período de simulação. O controle de sobreajuste decorre da limitação da complexidade dos modelos (profundidade e número de árvores, taxa de aprendizado reduzida, número reduzido de neurônios e decaimento de pesos), da janela de treinamento de um ano, da reestimação semanal e da saturação da previsão entre zero e a geração máxima histórica do agrupamento. Os valores adotados encontram-se no Apêndice A.

Como limitações, os hiperparâmetros não foram otimizados de forma exaustiva nem específica para cada agrupamento, e o treinamento envolve componentes aleatórios (reamostragem das florestas e inicialização das redes), de modo que execuções repetidas podem apresentar pequenas variações. Além disso, os modelos de aprendizado de máquina utilizam como insumos apenas a série de geração do ponto e a previsão do ECMWF. Testes exploratórios com outras abordagens e variáveis adicionais, realizados para um grupo de usinas, não indicaram ganhos expressivos e generalizados em relação à configuração adotada.

5.10.3 Modelos do *ensemble*

O *ensemble* é composto por modelos pertencentes a diferentes famílias, descritos de forma sucinta a seguir. O detalhamento matemático de cada um deles é apresentado no Apêndice A. De maneira geral, os modelos podem ser divididos em dois grupos: aqueles que utilizam exclusivamente a série histórica de geração e aqueles que incorporam previsões provenientes de modelos de previsão numérica do tempo (NWP). Os modelos do segundo grupo partem da previsão de vento de NWP, de base física, e a convertem ou corrigem por meio de técnicas estatísticas e de aprendizado de máquina, caracterizando, conforme a classificação apresentada na Seção 2.5.1, abordagens híbridas.

Nos modelos de aprendizado de máquina e nos modelos estatísticos baseados na modelagem dos resíduos, optou-se por utilizar como insumo meteorológico as previsões do modelo ECMWF, por apresentar, de forma consistente, um dos melhores desempenhos individuais entre os modelos NWP avaliados. Embora, em princípio, a previsão combinada dos modelos NWP represente a melhor estimativa disponível, sua utilização nesses modelos apresenta uma limitação importante. Como os pesos da combinação são recalculados dinamicamente a cada instante de previsão, a série combinada não constitui uma referência única e estável para o cálculo e a modelagem dos resíduos ao longo do tempo.

Por outro lado, os modelos baseados em persistência utilizam como referência a previsão combinada dos modelos NWP. Nessa classe de modelos, a variabilidade temporal dos pesos da combinação não compromete o desempenho, uma vez que não há uma etapa explícita de treinamento de parâmetros. Dessa forma, a previsão combinada constitui a melhor estimativa meteorológica disponível durante a janela de simulação, sendo, portanto, a referência mais adequada para esse tipo de abordagem.

5.10.3.1 Persistência pura

Modelo de referência (*baseline*), que projeta para todo o horizonte o último valor de geração verificada. Não utiliza previsões meteorológicas e degrada rapidamente à medida que o horizonte aumenta, servindo como limite inferior de desempenho. Participa da combinação apenas como candidato, sendo frequentemente excluída pelo limiar à medida que o horizonte aumenta.

5.10.3.2 Persistência inteligente

Evolução da persistência pura, combina a geração verificada recente com a previsão do NWP combinado por meio de etapas sucessivas: correção do viés recente, incorporação de uma componente autorregressiva estimada a partir do histórico e convergência gradual para a previsão meteorológica ao longo do horizonte, controlada por fatores de decaimento exponencial. Trata-se de um *baseline* robusto, que integra a combinação final. Os fatores de decaimento e os horizontes de atuação de cada etapa são fixos e comuns a todos os agrupamentos, definidos empiricamente, e os coeficientes autorregressivos são estimados uma vez por agrupamento, com até um ano de histórico anterior ao período de simulação.

5.10.3.3 Persistência com derivada

Variante da persistência que incorpora a tendência recente da geração, estimada como uma média ponderada das últimas variações observadas, projetada com decaimento ao longo do horizonte e posteriormente mesclada à previsão do NWP combinado.

5.10.3.4 Regressão linear

Modelo linear ajustado por horizonte e por meia-hora do dia de emissão, que relaciona a geração futura a um conjunto de variáveis explicativas (*features*): geração atual e defasada, tendência recente, erro pontual do NWP, termos de sazonalidade diária e a previsão do ECMWF para o horizonte alvo. Os coeficientes são estimados por mínimos quadrados a partir de uma janela de 365 dias. Como cada regressão dispõe de cerca de uma amostra por dia da janela e de poucos preditores, o risco de sobreajuste é limitado, mas o modelo não captura relações não lineares.

5.10.3.5 ARIMA

Modelo estatístico da classe *Box–Jenkins* [85] aplicado à série de erro do NWP (erro = $G_v - G_{\mathrm{ECMWF}}$), e não à geração diretamente. A ordem de diferenciação é forçada em zero ($d = 0$), de modo que a série de erro, estacionária, seja tratada como reversível à média; a previsão final soma o erro projetado à previsão do ECMWF. É calibrado em janela de 30 dias, com ordem selecionada automaticamente, com reaproveitamento dos parâmetros e atualização do estado a cada emissão com os erros dos últimos sete dias. Por tratar o erro como estacionário, o modelo reverte rapidamente à média do erro nos horizontes mais longos.

5.10.3.6 Suavização exponencial com reversão à média

Modelo da família de suavização exponencial (espaço de estados), também aplicado à série de erro do NWP. Como a suavização exponencial pura projeta o nível recente de forma constante, sem reverter à média, adotou-se uma variante que combina o nível exponencialmente suavizado com um termo explícito de reversão à média global do erro. Os parâmetros de suavização e de reversão são otimizados por busca em grade, minimizando o erro absoluto médio de previsão do erro sobre a própria janela de treinamento de 30 dias.

5.10.3.7 *Random Forest*

Modelo de aprendizado de máquina por *bagging* de árvores de decisão, implementado em duas variantes: por horizonte (*per step*), em que um modelo distinto é treinado para cada passo, e de múltiplas saídas (*multi-output*), em que um único modelo prevê todos os passos simultaneamente. Ambas utilizam as *features* de base acrescidas da previsão do ECMWF e janela de treino de 365 dias. A complexidade é controlada por um número moderado de árvores e por um tamanho mínimo de nó, e o conjunto de hiperparâmetros é apresentado no Apêndice A.

5.10.3.8 *XGBoost*

Modelo de aprendizado de máquina por *boosting* de árvores, com um modelo independente para cada passo de previsão, com o mesmo conjunto de *features*, janela de treino do *Random Forest*. São consideradas duas configurações de hiperparâmetros: a configuração por horizonte e uma configuração mais regularizada, rotulada como *multi-output* nos resultados, com árvores mais rasas, taxa de aprendizado menor e maior número de rodadas, e que, na implementação, também consiste em um modelo por passo. A variante por horizonte está entre os modelos individuais de melhor desempenho e integra a combinação final.

5.10.3.9 Rede neural (MLP)

Rede neural *multilayer perceptron* (MLP), com uma rede dedicada a cada horizonte. Adota-se uma topologia com camada de entrada de 14 variáveis, uma camada oculta de 8

neurônios com função de ativação sigmoide e um neurônio de saída linear, com regularização por decaimento de pesos. A padronização *z-score* das variáveis de entrada e da resposta é essencial para a convergência. Trata-se do modelo individual de melhor desempenho. Cada rede é treinada com um número máximo de iterações; a pequena dimensão da camada oculta e o decaimento de pesos limitam o sobreajuste.

5.10.3.10 SARIMA

Modelo *Seasonal Autoregressive Integrated Moving Average* (SARIMA) aplicado diretamente à série de geração, com sazonalidade diária. Diferentemente dos demais, não utiliza previsões meteorológicas, funcionando como modelo de controle puramente temporal; por isso, não integra a combinação final. É calibrado em janela de 30 dias, com atualização do estado a cada emissão a partir dos últimos 14 dias de geração.

5.10.4 Estratégias de combinação

Dado que vários modelos atuam em paralelo, e que a operação requer uma previsão única, é necessário combiná-los. Os candidatos à combinação são todos os modelos individuais do *ensemble*, incluindo o NWP combinado, com exceção do SARIMA, que, por utilizar apenas a série histórica de geração, é mantido como modelo de controle. Combinar todos os candidatos sem qualquer critério de exclusão tende a piorar o resultado, pois modelos de desempenho consistentemente inferior, ou fortemente correlacionados entre si, diluem a contribuição dos melhores. Por isso, em vez de se definir previamente um subconjunto fixo, a seleção é feita de forma dinâmica: a cada emissão e para cada horizonte, os modelos de pior desempenho são excluídos e os demais são ponderados conforme seu desempenho recente. Essa combinação seletiva de fontes complementares reduz o erro em relação às fontes isoladas, como se observa, para a previsão com os diferentes modelos meteorológicos, na Figura 104.

Os critérios de exclusão e a regra de ponderação diferem entre as duas estratégias desenvolvidas, descritas a seguir. Ambas utilizam apenas informações já verificadas no instante de cada emissão e são avaliadas fora da amostra no protocolo de simulação apresentado na Seção 5.10.2. A configuração final foi definida com base no desempenho observado nos experimentos realizados, e o comportamento resultante dos pesos é analisado na Seção 6.5.3. Foram desenvolvidas duas estratégias de combinação.

5.10.4.1 Combinação por inverso do erro (1/MAE)

Atribui a cada modelo, por horizonte, um peso proporcional ao inverso do seu erro absoluto médio recente, calculado em uma janela deslizante de sete dias com ponderação temporal crescente (observações mais recentes recebem maior peso). Um limiar exclui da combinação, em cada horizonte, os modelos cujo erro supera um múltiplo do erro do melhor modelo, e os pesos são renormalizados considerando apenas os modelos com previsão válida.

O valor adotado para o limiar, de 1,5 vez o menor erro, foi definido empiricamente em testes preliminares, e, caso todos os modelos de um horizonte ultrapassem o limite, todos são reativados, evitando que a combinação fique sem componentes.

5.10.4.2 Combinação por filtro de Kalman

Estratégia adaptativa que trata os pesos da combinação como um vetor de estado que evolui no tempo, com um filtro de Kalman independente por horizonte. O estado (pesos) segue um passeio aleatório e é atualizado recursivamente a cada nova observação de geração verificada. Para garantir estabilidade, evitando a explosão dos pesos por multicolinearidade entre modelos correlacionados, a cada passo os pesos são projetados no simplex (não negativos e com soma unitária) e restritos ao subconjunto de modelos elegíveis, que é redefinido dinamicamente por horizonte a partir de uma validação sobre a janela recente, e não por um limiar fixo. Por tratar os pesos como um estado atualizado recursivamente e por redefinir continuamente o conjunto elegível, o Kalman tende a apresentar pesos mais variáveis do que o método 1/MAE.

Essa combinação apresentou o melhor desempenho global, com ganhos que se tornam mais expressivos nos horizontes de previsão mais longos. O equacionamento completo do filtro encontra-se no Apêndice A.

5.10.5 Síntese dos modelos

A Tabela 6 resume as principais características dos modelos que compõem o *ensemble* e das estratégias de combinação, evidenciando a cobertura das diferentes famílias de modelos da literatura de previsão eólica.

A coluna *Técnica* indica o método de modelagem empregado, enquanto a coluna *Abordagem* classifica o modelo segundo a taxonomia apresentada na Seção 2.5.1; assim, um modelo de série temporal aplicado ao erro de um NWP é estatístico quanto à técnica e híbrido quanto à abordagem.

Vale salientar que o desempenho do modelo de tempo real depende diretamente da qualidade das previsões dos modelos meteorológicos, herdadas dos modelos de curtíssimo e curto prazo.

5.11 Considerações finais do capítulo

Este capítulo apresentou os modelos de previsão de geração eólica para os horizontes de tempo real, curtíssimo e curto prazo, contemplando desde o tratamento das previsões meteorológicas até a conversão vento–geração e a combinação dinâmica das previsões.

A Figura 75 consolida a arquitetura do modelo proposto, integrando os módulos desenvolvidos nos Capítulos 3, 4 e 5. O processo inicia-se com os **dados de entrada**, que reúnem informações técnicas das usinas, medições de geração e vento, disponibilidade dos aerogeradores

Tabela 6 – Síntese dos modelos do *ensemble* de tempo real e das combinações.

Modelo	**Técnica**	**Abordagem**	**Fonte NWP**	**O que modela**
Persistência pura	Persistência	Referência	—	Repete o último valor verificado
Persistência inteligente	Persistência	Híbrida	Combinado	Converge para o NWP em etapas (nível, AR, *blend*)
Persistência com derivada	Persistência	Híbrida	Combinado	Persistência mais tendência recente
Regressão linear	Regressão	Híbrida	ECMWF	Regressão sobre *features* e NWP do horizonte, por meia-hora do dia
ARIMA	Série temporal (Box–Jenkins)	Híbrida	ECMWF (resíduo)	Erro do NWP com $d = 0$ (reversão à média)
Suavização exponencial	Série temporal (espaço de estados)	Híbrida	ECMWF (resíduo)	Nível suavizado mais reversão à média
Random Forest	Aprendizado de máquina (*bagging*)	Híbrida	ECMWF	*Features* de base mais NWP
XGBoost	Aprendizado de máquina (*boosting*)	Híbrida	ECMWF	*Features* de base mais NWP
Rede neural (MLP)	Aprendizado de máquina (redes neurais)	Híbrida	ECMWF	Rede $14 \to 8 \to 1$ por horizonte
SARIMA	Série temporal	Estatística	—	Geração com sazonalidade diária
NWP combinado	Combinação de NWP	Estatística	ECMWF/ GFS/ GEFS	Pesos 1/MAE dos modelos NWP
Combinação 1/MAE	Combinação dinâmica	Estatística	—	Pesos 1/MAE por horizonte
Combinação Kalman	Combinação dinâmica	Estatística	—	Pesos como estado (simplex)

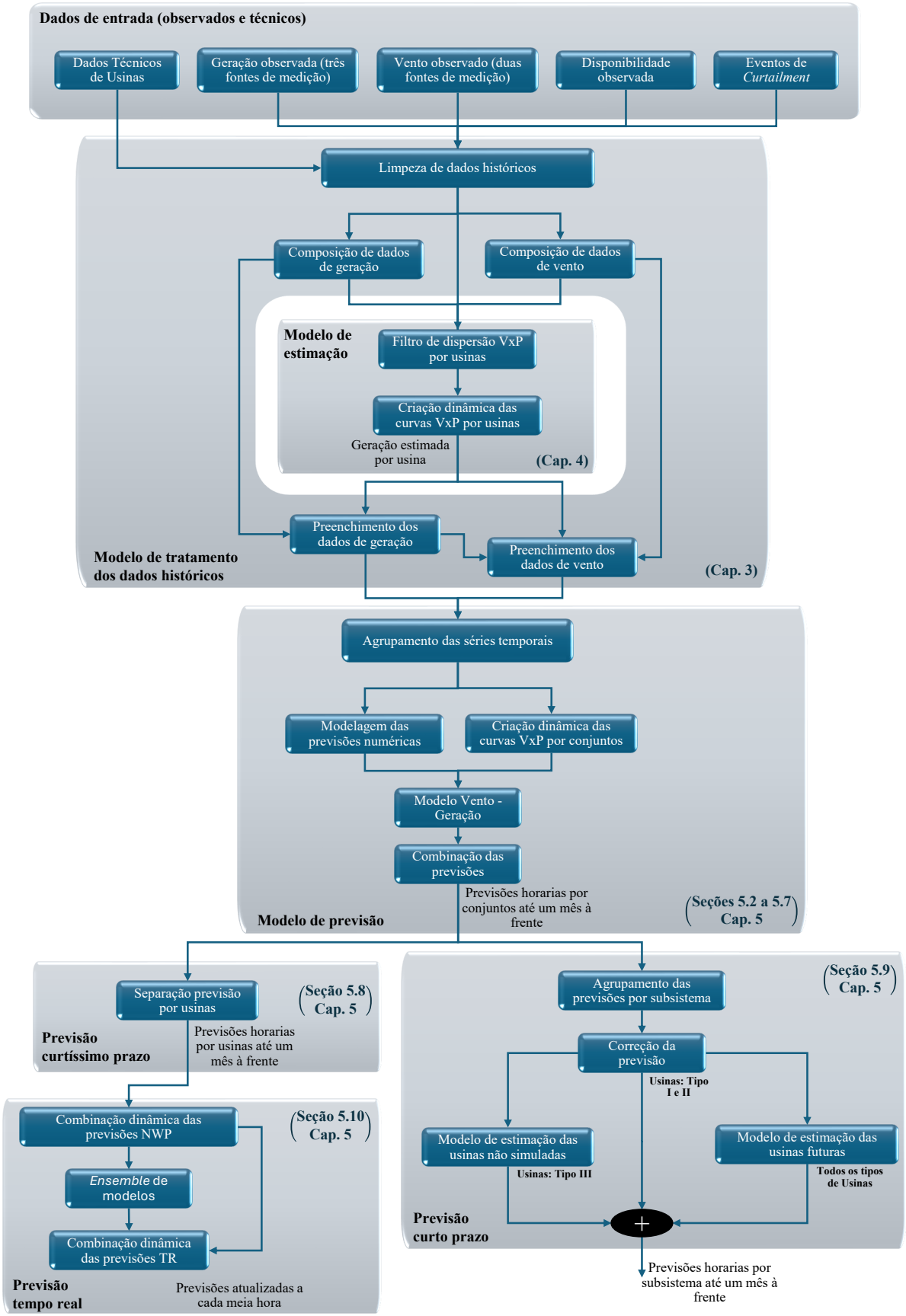
Fonte: Elaboração própria.

e registros de eventos de *curtailment*. Esses dados alimentam o **modelo de tratamento dos dados históricos** (Capítulo 3), responsável pela limpeza, composição e consolidação das séries históricas.

O **modelo de estimação** (Capítulo 4) atua sobre essas séries por meio do filtro de dispersão e da construção dinâmica das curvas V $\times$ P por usina, produzindo estimativas tratando os efeitos do *curtailment* que complementam e refinam a base histórica. Na sequência, o **modelo de previsão** (Capítulo 5) realiza o agrupamento das usinas, o tratamento das previsões meteorológicas, a construção das curvas V $\times$ P por conjuntos e a geração das previsões horárias com alcance de até um mês. A partir desse processo, são obtidas as previsões para os horizontes de tempo real, curtíssimo e curto prazo, cada uma com os tratamentos específicos requeridos pela operação. O fluxograma evidencia, assim, a integração entre os modelos de tratamento, estimação e previsão, sintetizando a metodologia desenvolvida ao longo desta tese.

Os principais conceitos e resultados aqui apresentados foram consolidados em [32], artigo que descreve o modelo de previsão de geração eólica de curto prazo desenvolvido e atualmente em operação no ONS, com horizontes de até um mês à frente. Naquele trabalho, são detalhados o arcabouço regulatório, a experiência internacional de operadores de sistema, a modelagem das previsões numéricas de tempo e a análise de desempenho em função do horizonte

Figura 75 – Fluxograma do modelo proposto, em sua versão detalhada.



Fonte: Elaboração própria.

de previsão e do modelo meteorológico utilizado (GFS e GEFS). Em relação a esse artigo, mais sucinto, a abordagem desenvolvida nesta tese amplia o detalhamento metodológico de cada etapa do modelo, integra a previsão às etapas anteriores de tratamento de dados e de estimação, e aprofunda a discussão dos diferentes horizontes temporais, incluindo o modelo de previsão em tempo real.

Além do artigo que consolida a metodologia, diversas publicações decorrentes desta pesquisa marcaram a evolução dos modelos de previsão aqui apresentados. A metodologia para o horizonte de curto prazo teve como pontos de partida os trabalhos [37] e [33], que abordaram, respectivamente, a correção dos desvios sistemáticos da estimativa vigente para estudos prospectivos do ONS e os avanços metodológicos na previsão eólica para o planejamento de curto prazo. A previsão para o horizonte da primeira semana operativa, voltada às usinas representadas no MPCP (atualmente o DECOMP), foi detalhada em [40]. As ideias para a representação das usinas do tipo III, modeladas no contexto da previsão de curto prazo, originaram-se em [36]. O modelo de previsão em tempo real, por sua vez, teve início no trabalho [34].

No que se refere ao agrupamento de usinas, o trabalho [38] representou uma evolução na forma de definir os conjuntos de parques. Embora os testes e resultados desta tese tenham adotado os grupos atualmente utilizados pelo ONS, a abordagem de clusterização proposta constitui uma alternativa mais eficiente e um aprimoramento potencial do processo. Por fim, a definição metodológica que privilegia os modelos numéricos de tempo como principal insumo foi embasada nas análises comparativas entre métodos clássicos de séries temporais e técnicas de *machine learning* apresentadas em [42]. Em conjunto, esses trabalhos evidenciam a maturação progressiva dos modelos de previsão consolidados nesta tese.

6 RESULTADOS

Este capítulo apresenta os principais resultados obtidos com as metodologias desenvolvidas ao longo do trabalho, organizados de forma a privilegiar as análises em níveis mais agregados do sistema, isto é, por subsistema e em panoramas nacionais. Os resultados mais desagregados espacialmente, por conjuntos de usinas e por usinas individuais, são apresentados no Apêndice C, de modo a manter o capítulo conciso sem prejuízo da completude da avaliação. As métricas de erro e os critérios de desempenho empregados nas análises a seguir estão definidos no Apêndice B.

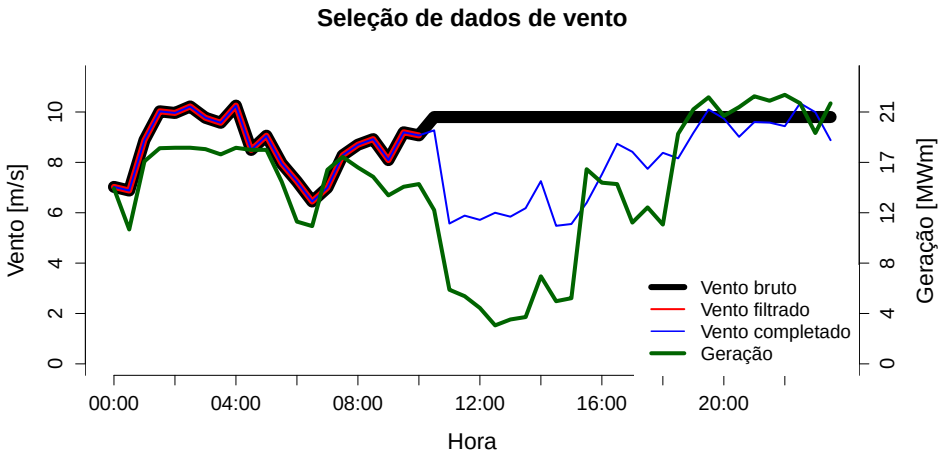
6.1 Resultados do processo de tratamento dos dados observados

Conforme descrito no Capítulo 3, os dados das séries temporais de vento foram submetidos a rigoroso controle de qualidade e avaliação de consistência. Esse processo inclui a remoção de medições fisicamente inconsistentes, a identificação automática de sensores com comportamento anômalo e a integração hierárquica de múltiplas fontes de dados observados. Além disso, lacunas temporais remanescentes foram preenchidas utilizando informações de parques eólicos geograficamente próximos, selecionados com base em critérios espaciais e estatísticos. O preenchimento de lacunas é realizado apenas quando um alto nível de consistência entre as séries é observado, preservando assim os dados originalmente medidos e garantindo rastreabilidade completa das informações utilizadas.

A Figura 76 apresenta um exemplo da aplicação dos procedimentos de seleção, tratamento e completamento de dados de vento, bem como sua relação com a geração eólica observada. A série de vento bruta corresponde às medições originalmente disponíveis, enquanto a série filtrada resulta da aplicação de critérios de consistência. Por sua vez, a série de vento completada incorpora o preenchimento das lacunas de dados remanescentes, produzindo um sinal mais adequado para uso em modelos de estimação. A geração eólica observada é mostrada no eixo secundário, permitindo uma comparação direta entre as variações relativas da velocidade do vento e da potência produzida, o que facilita a avaliação da coerência física entre os dados de vento processados e a resposta de geração eólica.

A Figura 77 apresenta o percentual de falhas identificadas nas séries de geração de energia para cada usina eólica, ordenadas de forma crescente pelo total de dados inválidos na série bruta. Três estágios do processo de tratamento de dados são representados. Os marcadores pretos correspondem ao percentual total de registros inválidos ou ausentes detectados na série original bruta da base do ONS, obtidos após a etapa de limpeza descrita na Seção 3.2.1 do Capítulo 3. Os marcadores vermelhos indicam as lacunas remanescentes após a consolidação de fontes redundantes de medição disponíveis para uma mesma usina, ou seja, após o preenchimento com os dados medidos da base da CCEE, de 5 minutos e horária, etapa que já proporciona uma

Figura 76 – Dados de vento brutos, filtrados e completados, juntamente com a geração eólica.

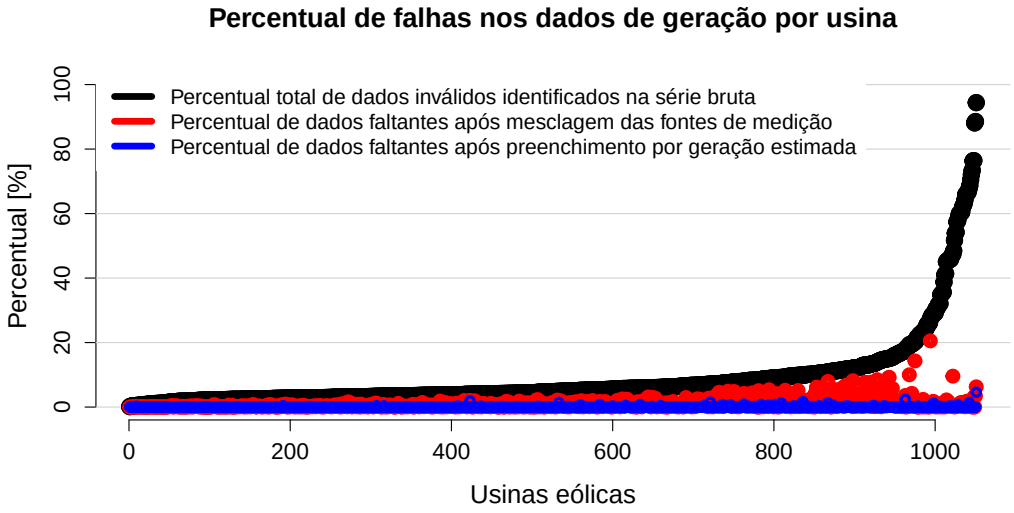


Fonte: Elaboração própria.

redução expressiva para diversas instalações e que corresponde à composição dos históricos apresentada na Seção 3.3.1.

Por fim, os marcadores azuis representam o percentual de dados faltantes após a aplicação de um modelo de estimação de geração, etapa de substituição por estimação, na qual os valores são obtidos por modelo e não por medição, que constitui a penúltima etapa de preenchimento, detalhada na Seção 3.3.2. A etapa final, correspondente ao preenchimento estatístico das lacunas residuais, não está representada na figura, uma vez que, após sua aplicação, o percentual de falhas remanescentes é reduzido a zero.

Figura 77 – Percentual de falhas nas séries de geração de energia por usina em cada etapa do processo de tratamento de dados.



Fonte: Elaboração própria.

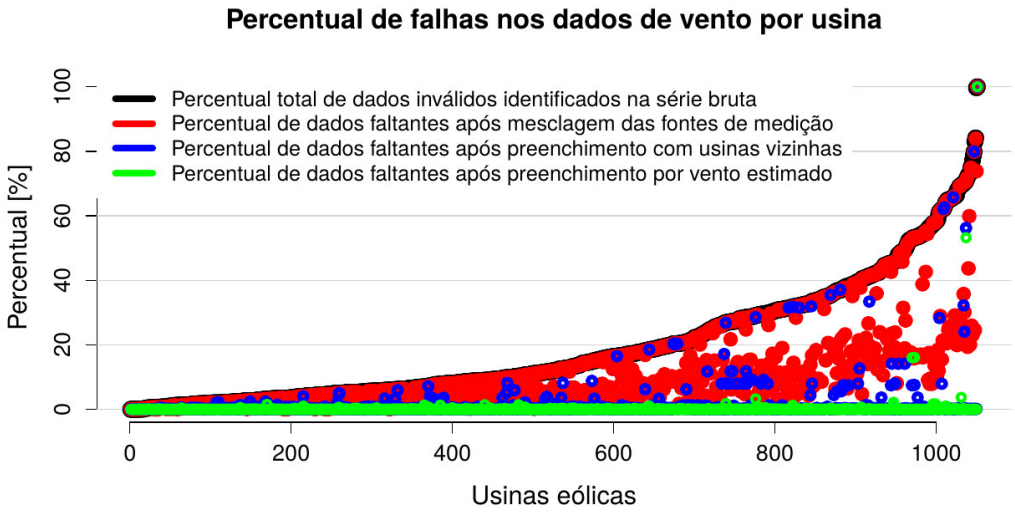
Diferentemente do tratamento aplicado às séries de velocidade de vento, não é empregada uma etapa intermediária de preenchimento espacial baseada em usinas vizinhas, uma vez que a

geração depende não apenas das condições de vento, mas também de características físicas e operativas específicas de cada instalação. Ainda assim, o modelo de estimação mostra-se eficaz em reduzir os dados faltantes a níveis aceitáveis para a maioria das usinas.

A Figura 78 apresenta uma análise análoga para as séries de velocidade de vento, nas quais quatro estágios do processo de tratamento são ilustrados. Os marcadores pretos representam o percentual total de dados inválidos identificados na série bruta da base do ONS. Os marcadores vermelhos mostram as lacunas restantes após a consolidação das fontes de medição, ou seja, após o preenchimento com os dados medidos da base da EPE (Seção 3.4.1). Os marcadores azuis indicam os dados faltantes residuais após um procedimento de preenchimento espacial que explora séries correlacionadas de usinas vizinhas, conforme a Seção 3.4.2.

Os marcadores verdes, por sua vez, exibem o percentual de falhas que persiste após a aplicação de um modelo de estimação de velocidade de vento, etapa de substituição por estimação, na qual os valores são obtidos por modelo e não por medição, que representa a penúltima etapa de preenchimento, descrita na Seção 3.4.2. A etapa final, correspondente ao preenchimento estatístico das lacunas residuais, não está representada na figura, uma vez que, após sua aplicação, o percentual de falhas remanescentes é reduzido a zero.

Figura 78 – Percentual de falhas nas séries de velocidade de vento por usina em cada etapa do processo de tratamento de dados.



Fonte: Elaboração própria.

A redução progressiva observada do preto ao verde ilustra a eficácia cumulativa de cada etapa sucessiva do tratamento, demonstrando que a metodologia proposta é capaz de reduzir substancialmente a indisponibilidade de dados ao longo de todo o parque eólico. Uma comparação entre as Figuras 77 e 78 revela que as séries brutas de velocidade de vento apresentam, em geral, percentuais de dados inválidos superiores aos das séries de geração, evidenciando a importância de um arcabouço robusto de tratamento de dados para os registros de velocidade de vento.

Conforme ilustrado nas Figuras 77 e 78, o tratamento progressivo dos dados resultou em uma redução expressiva no percentual de falhas remanescentes nas séries temporais das 1.051 usinas eólicas analisadas. Na série bruta de geração, 28 usinas apresentavam mais de 50% de falhas e 212 superavam o patamar de 10%. Já para as séries de vento, a situação inicial era ainda mais crítica: 89 usinas possuíam mais de 50% de dados faltantes, e 576 ultrapassavam 10%. Essa diferença decorre de fatores técnicos e econômicos, discutidos a seguir.

Em primeiro lugar, a medição de grandezas elétricas é mais consolidada e menos sujeita a falhas do que a medição do vento. A tensão e a corrente são medidas por transformadores de instrumento e medidores, equipamentos mais estáveis, com menos peças móveis e calibração mais previsível, e os dados de geração contam ainda com estimadores de estado. A medição do vento, por sua vez, depende de anemômetros, sensores mecânicos ou ultrassônicos expostos às intempéries, sujeitos a desgaste, descalibração e falhas mecânicas, e não dispõe de um mecanismo equivalente de reconstituição dos valores. Em segundo lugar, há uma razão econômica: a geração está diretamente ligada à receita e à liquidação do agente gerador, o que cria incentivo para manter sua medição em boas condições, ao passo que, para o vento, a consequência comercial é mais indireta. Somam-se a isso os problemas de comunicação, as falhas de equipamentos, as medições congeladas e a baixa qualidade histórica da infraestrutura de medição, descritos na Seção 3.2, que afetam de forma mais severa as séries de vento.

Após a aplicação sucessiva das técnicas de preenchimento baseadas em estimações, o volume de falhas foi drasticamente reduzido. Na série tratada até a etapa de estimação, apenas 10 usinas na geração e 20 no vento ainda apresentavam mais de 1% de dados faltantes, um limiar consideravelmente mais baixo do que os utilizados para caracterizar a série bruta. Por fim, a etapa final de tratamento, de preenchimento estatístico das lacunas residuais (média de dias adjacentes, no caso da geração, pontos válidos próximos, no caso do vento, ou o método estatístico descrito na Seção 3.6), aplicada apenas após o esgotamento das fontes complementares e das estimações, permitiu preencher as lacunas residuais, completando o processo de reconstituição das séries.

6.2 Resultados da aplicação do modelo de estimação por subsistema

6.2.1 Premissas e organização dos resultados

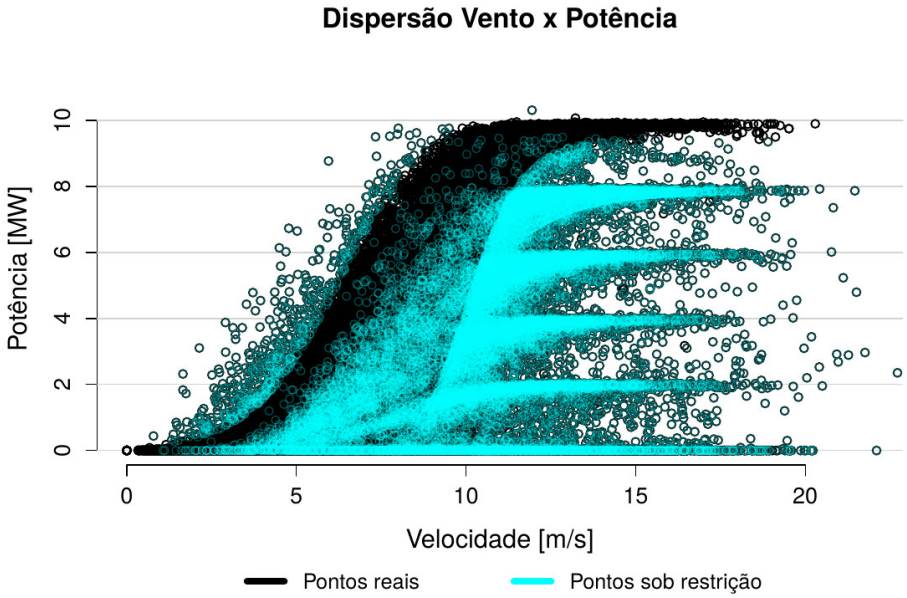
O modelo de estimação, cuja construção das curvas Vento $\times$ Potência foi detalhada na Seção 4.5 do Capítulo 4, é aplicado de forma individualizada por usina. No entanto, os cortes de geração são realizados por conjuntos de usinas, de modo que não é possível afirmar com certeza absoluta se uma usina específica foi submetida a corte, uma vez que a decisão de quais usinas cortar dentro do conjunto cabe ao agente operador responsável. Para efeitos de avaliação, adota-se a premissa de que, quando o conjunto está sob corte, todas as usinas que o compõem também estão. Diversos resultados do modelo de previsão serão apresentados considerando agrupamentos específicos de usinas; sendo assim, o modelo de estimação tem seus resultados avaliados por usina individual, por esses agrupamentos específicos e, também,

de forma agregada por subsistema. As avaliações por usina individual e por agrupamentos, de caráter mais desagregado, são apresentadas no Apêndice C; neste capítulo, o foco recai sobre os resultados agregados por subsistema. As figuras a seguir, ainda nesta subseção, ilustram aspectos gerais do método de identificação de restrições operativas, comum a todos os níveis de análise.

Além dos cortes de geração determinados por solicitação do ONS, outros fatores operativos nas usinas eólicas comprometem seu desempenho máximo, como a operação com número de máquinas inferior ao total instalado. Nem sempre os dados de potência disponível, que refletem o número de máquinas em operação, são informados corretamente. Assim, para efeito de avaliação dos resultados por usina e por ponto de medição, foi desenvolvido um mecanismo de identificação desses períodos de operação atípica, permitindo o expurgo desses intervalos das séries históricas verificadas e estimadas utilizadas na avaliação.

A Figura 79 apresenta a relação entre a velocidade do vento (m/s) e a potência gerada (MW). Os pontos em preto representam os dados reais de operação, enquanto os pontos em ciano indicam os registros obtidos sob condições de restrição operacional. Observa-se que, em condições normais, a potência tende a crescer em direção à região de saturação com o aumento da velocidade do vento, seguindo o comportamento esperado da curva de potência do aerogerador. Já os pontos sob restrição apresentam valores de potência inferiores ao esperado para a respectiva faixa de velocidade, evidenciando perdas associadas a limitações operacionais, como *curtailment*, manutenções ou restrições impostas pelo operador do sistema, cujo tratamento foi apresentado nas Seções 4.1 e 4.2 do Capítulo 4. Essa distinção é fundamental para a correta avaliação do desempenho do parque eólico e para a estimativa da energia desperdiçada.

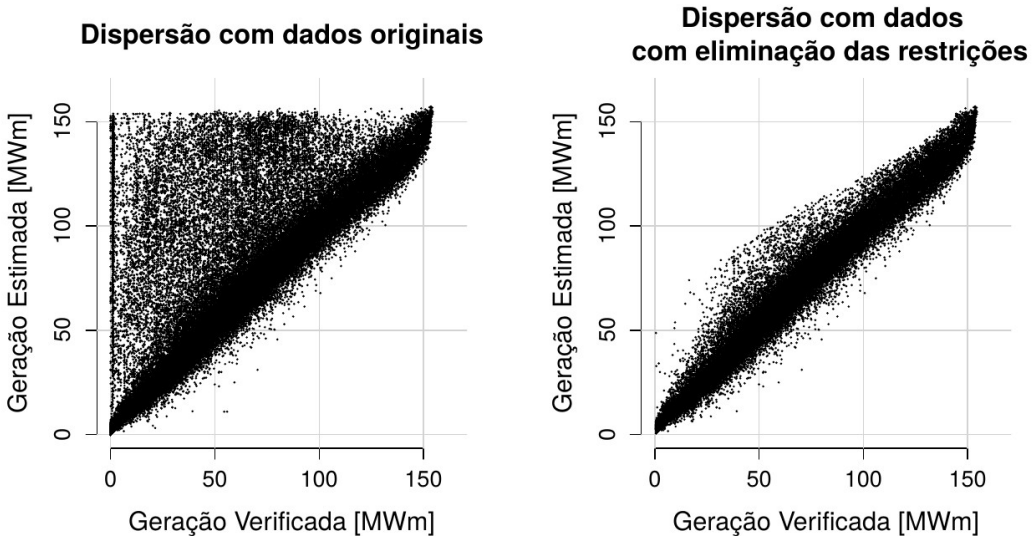
Figura 79 – Dispersão entre velocidade do vento e potência gerada, com os dados reais de operação e os registros sob restrição operacional.



Fonte: Elaboração própria.

A Figura 80 apresenta a dispersão entre a geração verificada e a geração estimada para um agrupamento de usinas em que o mecanismo de identificação de restrições operacionais apresentou bom desempenho. O gráfico à esquerda exibe os dados originais, sem qualquer tratamento, onde se observa a presença de pontos com geração verificada significativamente inferior à estimada, formando uma dispersão alongada ao longo do eixo horizontal. Esses pontos correspondem a períodos em que a usina operou sob restrições, como cortes de geração, manutenções ou redução no número de máquinas disponíveis. O gráfico à direita apresenta os mesmos dados após a aplicação do mecanismo de eliminação dos períodos de operação atípica. Nota-se que a remoção desses registros resulta em uma dispersão mais concentrada em torno da reta identidade, indicando boa aderência entre os valores estimados e verificados. Esse resultado evidencia que, para este agrupamento, o procedimento de filtragem foi eficaz na identificação e no expurgo dos períodos comprometidos por restrições operacionais.

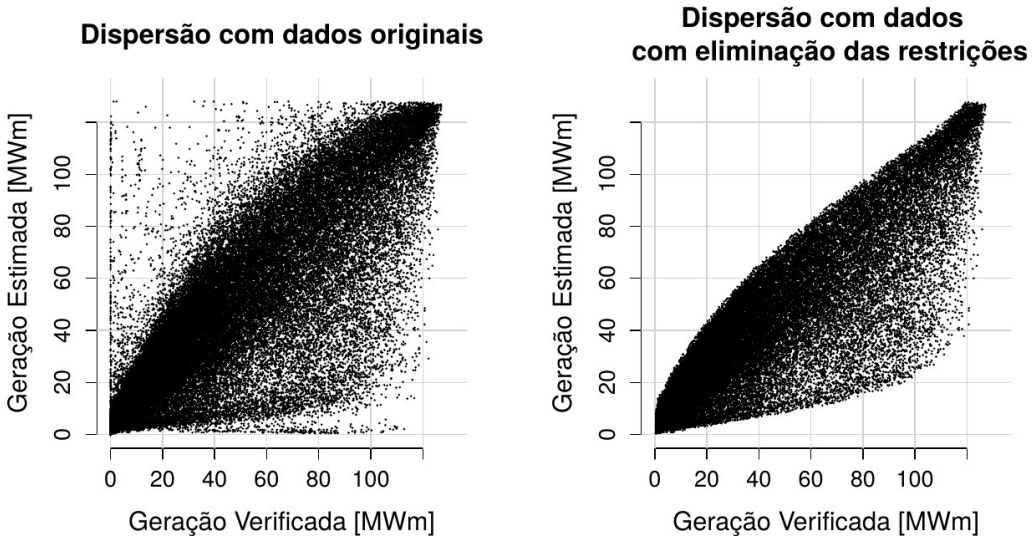
Figura 80 – Dispersão entre geração verificada e geração estimada com boa identificação de restrições.



Fonte: Elaboração própria.

A Figura 81 apresenta o mesmo tipo de análise para um agrupamento de usinas em que o mecanismo de identificação de restrições não obteve resultado satisfatório. No gráfico à esquerda, com os dados originais, observa-se novamente a presença de pontos dispersos associados a períodos de operação restrita. No entanto, no gráfico à direita, após a aplicação do tratamento, ainda permanecem pontos afastados da reta identidade, indicando que o mecanismo não foi capaz de identificar adequadamente todos os períodos de restrição operacional. Essa limitação pode estar associada a fatores como a imprecisão nos dados de potência disponível informados pelo agente, a ocorrência de restrições parciais de difícil detecção ou a presença de anomalias operacionais não contempladas pelos critérios de filtragem adotados. Esse caso ilustra a necessidade de aprimoramento contínuo do mecanismo de identificação de restrições, bem como a importância de se considerar essas limitações na interpretação dos resultados do modelo de estimação.

Figura 81 – Dispersão entre geração verificada e geração estimada com identificação insatisfatória de restrições.



Fonte: Elaboração própria.

Embora o modelo vise estimar a geração durante períodos com corte, sua validação requer intervalos de tempo não afetados por esse fenômeno. Adicionalmente, os dados de geração observada incorporam outras fontes de incerteza e limitações operacionais que não são totalmente refletidas nos dados reportados ao ONS, utilizados nesse trabalho. Embora parte dessas inconsistências seja abordada durante o pré-processamento dos dados, algumas ainda persistem nas séries temporais, impactando os indicadores de desempenho de formas difíceis de quantificar.

Considerando o período de análise de cinco anos, as 1.051 usinas eólicas avaliadas ao longo de sua operação comercial entre 2021 e 2025 compõem um conjunto de dados de aproximadamente 73,9 milhões de registros de geração. Desse total, cerca de 19% não puderam ser aproveitados nas análises devido à identificação de restrições operativas nas usinas, de modo que aproximadamente 59,6 milhões de registros permanecem disponíveis para a validação dos resultados do modelo de estimação.

No que se refere aos agrupamentos, para o mesmo período e considerando 76 grupos de parques eólicos avaliados ao longo de seu comissionamento e operação entre 2021 e 2025, o conjunto de dados compreende aproximadamente 5,9 milhões de registros de geração. Desse total, cerca de 29% foram descartados das análises em razão da identificação de restrições operativas nos agrupamentos, resultando em aproximadamente 4,2 milhões de registros disponíveis para a validação dos resultados do modelo de estimação.

Em síntese, o modelo apresenta boa aderência aos dados verificados, com erros que não se concentram em nenhuma região específica, e viés sistemático reduzido, tipicamente inferior a 5% na maioria dos agrupamentos. Os vieses residuais observados em alguns casos são atribuídos a períodos de restrição operativa não integralmente filtrados e à entrada recente de novas usinas.

A seguir, os resultados são discutidos em nível de subsistema.

6.2.2 Resultados por subsistema

Em consonância com a abordagem adotada neste capítulo, os resultados em nível de subsistema elétrico fornecem uma visão abrangente do comportamento do sistema, sendo particularmente relevantes no contexto brasileiro, caracterizado pela elevada participação das regiões Sul e Nordeste na capacidade eólica instalada. Tal agregação é importante para a gestão do corte associado à transmissão de potência entre esses subsistemas. Para avaliar o comportamento geral do corte no país, agregações ainda mais amplas são necessárias, possibilitando análises em nível estratégico. Esses resultados são essenciais para aprimorar o planejamento da operação do sistema, reduzir perdas econômicas, orientar investimentos em infraestrutura de transmissão e armazenamento de energia e aumentar a utilização efetiva de energia limpa.

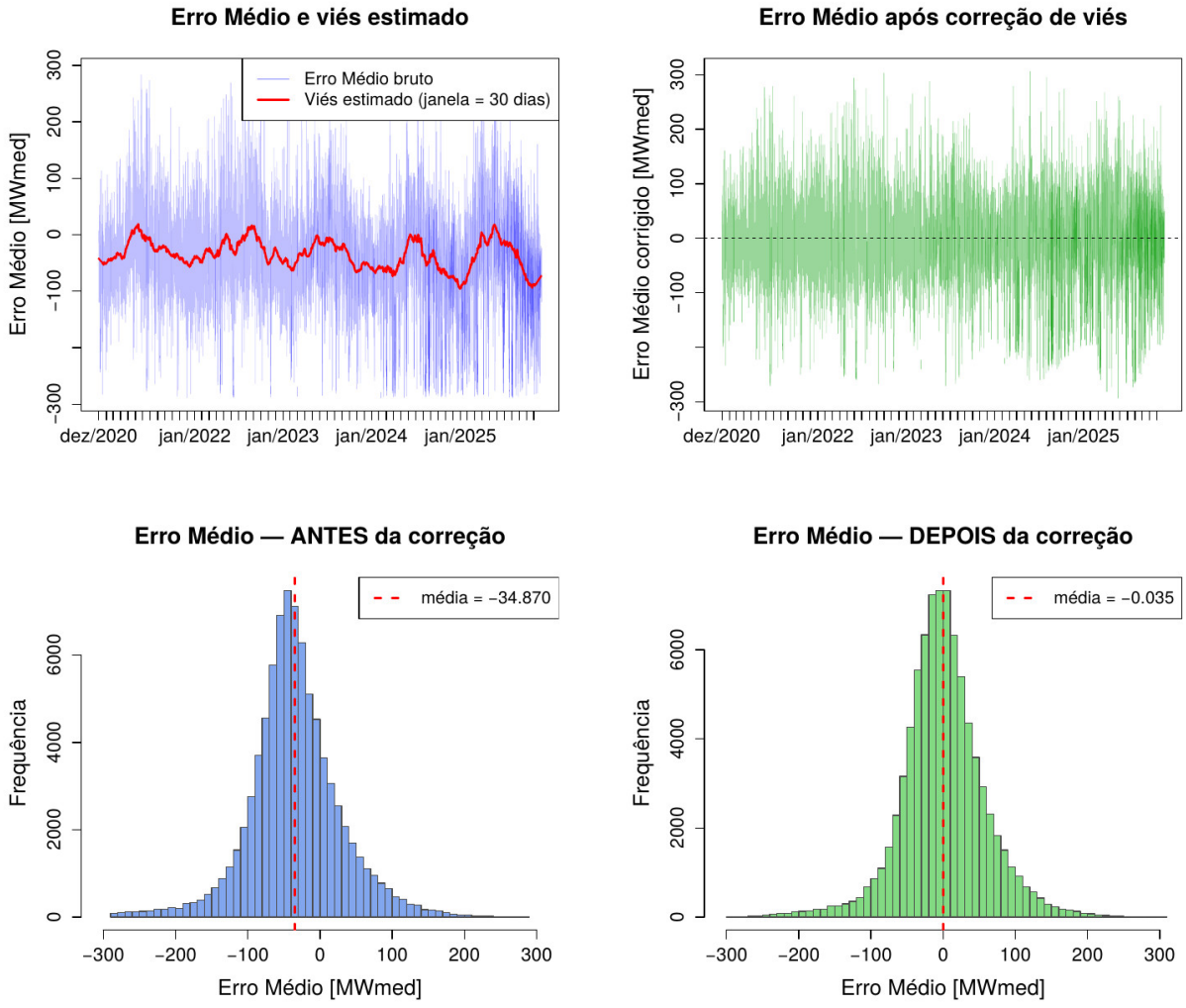
A validação do modelo de estimação requer períodos livres de restrições operativas, uma vez que, durante os cortes, a geração verificada não reflete a geração potencial que o modelo busca estimar; a inclusão desses períodos confundiria o erro próprio do modelo com o efeito das restrições, comprometendo a avaliação. Nos resultados por usinas individualizadas e por agrupamentos de usinas, apresentados no Apêndice C, foi possível identificar os momentos específicos em que os parques operavam sob restrições operativas e eliminar esses registros das séries históricas para efeito de avaliação. No entanto, para a análise no nível de subsistema, essa abordagem não é aplicável, uma vez que, em praticamente todos os instantes de tempo, há pelo menos uma usina ou agrupamento submetido a algum tipo de restrição. A eliminação desses registros resultaria na remoção de um volume excessivo de dados, comprometendo a representatividade da amostra. Dessa forma, optou-se por remover das análises apenas registros sabidamente não representativos de condições normais de operação, como algumas poucas falhas pontuais e eventos atípicos, a exemplo do ocorrido em 15 de agosto de 2023.

Ao se avaliar os resultados agregados por subsistema, observa-se a presença de um viés residual. Esse comportamento é atribuído, em grande parte, a restrições operativas que não caracterizam propriamente eventos de *curtailment*, tais como a redução da quantidade de aerogeradores disponíveis para operação ou limitações de potência que não são adequadamente representadas pelos dados de entrada do modelo. Como consequência, parte dessas condições operativas não é capturada durante o processo de estimação, resultando em desvios (erros) sistemáticos entre a geração estimada e a observada.

Para mitigar esse viés residual, foi adotada uma abordagem complementar baseada na correção da geração estimada agregada por subsistema. A metodologia consiste na estimação de um viés variante no tempo, obtido por meio de regressão linear aplicada a uma janela móvel do erro médio entre a geração estimada e a geração verificada. O viés estimado é então removido da série de geração estimada, produzindo uma série corrigida com menor componente sistemática de erro e, consequentemente, maior aderência aos valores observados.

A Figura 82 apresenta o resultado do procedimento de correção de viés para o subsistema Sul. O gráfico superior esquerdo exibe a série temporal do erro médio bruto (em azul) entre a geração estimada e a verificada, juntamente com o viés estimado pela janela móvel (em vermelho). O gráfico superior direito apresenta a série temporal do erro médio após a correção, onde se observa que os valores passam a oscilar de forma mais simétrica em torno de zero. Os gráficos inferiores apresentam os histogramas do erro médio antes e depois da correção. No subsistema Sul, devido à menor capacidade instalada, o viés original é relativamente pequeno, da ordem de 35 MWm. Ainda assim, a correção mostrou-se eficiente, centralizando a distribuição dos erros em torno de zero e reduzindo a assimetria observada antes do ajuste.

Figura 82 – Correção de viés da geração estimada para o subsistema Sul: séries temporais e histogramas do erro médio antes e depois da correção.

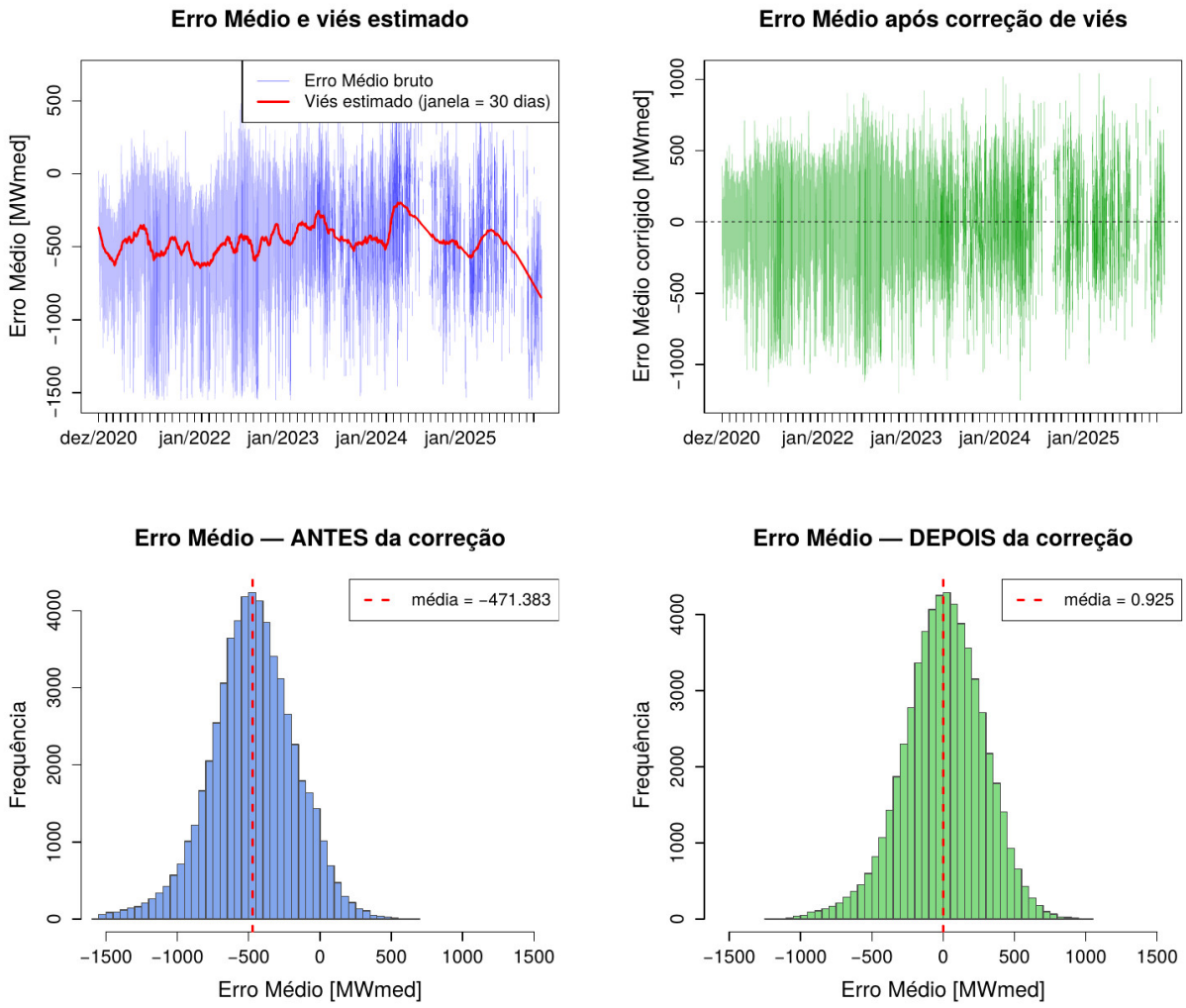


Fonte: Elaboração própria.

A Figura 83 apresenta a mesma análise para o subsistema Nordeste. Neste caso, o viés original é substancialmente maior, da ordem de 471 MWm, refletindo a maior capacidade instalada e a maior incidência de restrições operativas na região. No gráfico superior esquerdo, observa-se que a série do viés estimado acompanha de forma consistente a tendência do erro

médio bruto, capturando tanto a componente de longo prazo quanto as variações sazonais associadas ao regime de ventos do Nordeste. Após a aplicação da correção (gráfico superior direito), a série do erro médio passa a oscilar em torno de zero de forma satisfatória. Os histogramas inferiores confirmam a eficácia do procedimento, com a distribuição dos erros deslocando-se de uma posição claramente enviesada para uma configuração centrada em zero. A magnitude do viés corrigido no Nordeste reforça a importância dessa etapa de ajuste para a obtenção de estimativas confiáveis de energia não fornecida por *curtailment* neste subsistema.

Figura 83 – Correção de viés da geração estimada para o subsistema Nordeste: séries temporais e histogramas do erro médio antes e depois da correção.



Fonte: Elaboração própria.

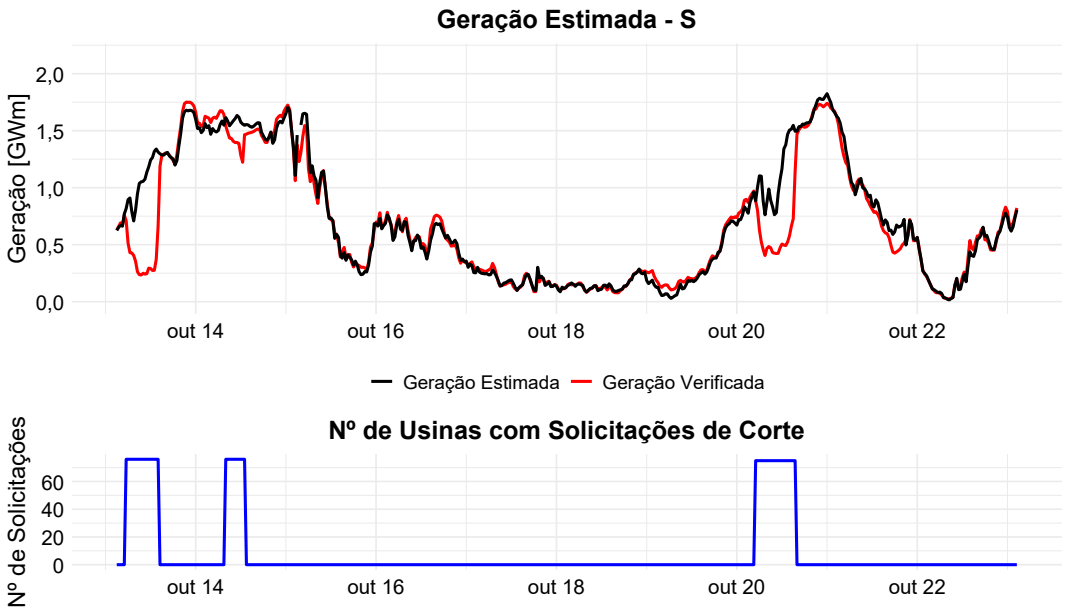
As Figuras 84 e 85 apresentam, para os subsistemas Sul e Nordeste respectivamente, a comparação entre a geração verificada e a geração estimada pelo modelo, bem como o número de usinas com solicitações de corte de geração, para o período de 13 a 22 de outubro de 2024. Cada figura é composta por dois gráficos dispostos verticalmente. O gráfico superior exibe as séries temporais de geração verificada (em vermelho) e geração estimada (em preto), em intervalos de meia hora, com valores expressos em GWm. O gráfico inferior apresenta o número de usinas

submetidas a solicitações de corte de geração ao longo do mesmo período.

No subsistema Sul (Figura 84), observa-se um alto nível de concordância entre as duas séries durante os períodos sem corte, destacando a capacidade do modelo de reproduzir consistentemente o comportamento real da geração eólica sob condições normais de operação. Essa concordância é um aspecto fundamental para a aplicação do modelo em análises contrafactuais, pois garante que as diferenças observadas durante eventos de restrição possam ser atribuídas com maior confiança aos efeitos do corte de geração, e não a deficiências na modelagem da geração potencial. Nos intervalos em que o número de solicitações de corte aumenta, a geração verificada apresenta valores inferiores à geração estimada, permitindo quantificar a energia não fornecida em decorrência das restrições operativas.

No subsistema Nordeste (Figura 85), o comportamento é análogo, com a geração estimada acompanhando adequadamente a geração verificada nos períodos de operação livre. As diferenças entre as duas séries nos momentos de corte evidenciam o impacto das restrições impostas pelo operador do sistema sobre a produção efetiva dos parques eólicos. A geração estimada, por não estar sob efeito dessas restrições, representa o potencial de geração que seria alcançado em condições normais de operação. A análise conjunta dos dois subsistemas demonstra a capacidade do modelo de estimação em capturar o comportamento esperado da geração eólica e, simultaneamente, evidenciar os períodos em que a operação foi comprometida por fatores externos à disponibilidade do recurso eólico.

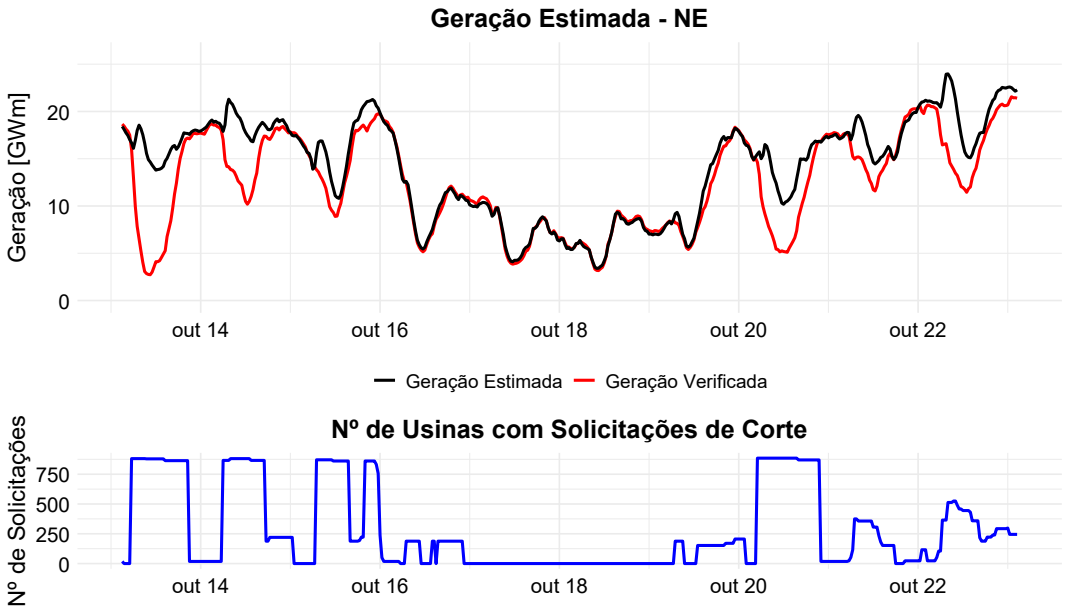
Figura 84 – Geração estimada e verificada com indicação de cortes para o subsistema Sul.



Fonte: Elaboração própria.

Para possibilitar uma avaliação mais abrangente do desempenho do modelo para todo o sistema elétrico brasileiro, as métricas apresentadas no Apêndice B foram aplicadas. Dife-

Figura 85 – Geração estimada e verificada com indicação de cortes para o subsistema Nordeste.



Fonte: Elaboração própria.

rentemente da normalização por amplitude utilizada na avaliação de desempenho para grupos de parques eólicos, na análise a seguir os erros brutos foram normalizados pela capacidade instalada em cada passo de tempo multiplicada pelo fator de capacidade médio ao longo do período analisado. Em termos interpretativos, essa referência equivale a considerar a geração eólica média do SIN. Essa abordagem de normalização foi adotada porque, ao agregar dados para todo o SIN, existem diferenças significativas na capacidade instalada entre o início e o final do período analisado que levam a variações na amplitude da série ao longo do tempo, conforme ilustrado na Figura 125 do Apêndice C. Portanto, essa normalização permite comparações de erros sob diferentes condições operacionais e escalas de geração, reduzindo a influência do crescimento da capacidade instalada na análise.

As Tabelas 7 e 8 apresentam as métricas de erro normalizadas do modelo de estimação de geração eólica para os subsistemas Sul, Nordeste e para o SIN, antes e após a aplicação da correção de viés, respectivamente. Além das métricas de erro, são apresentadas a faixa de potência instalada ao longo do período de análise (2021 a 2025) e o fator de capacidade médio de cada subsistema, fornecendo contexto sobre a dimensão e o aproveitamento do recurso eólico em cada região. As métricas de erro avaliadas são o erro médio normalizado (nME), o erro médio absoluto normalizado (nMAE), a raiz do erro quadrático médio normalizado (nRMSE) e o erro percentual absoluto médio normalizado (nMAPE).

Na Tabela 7, observa-se que todos os subsistemas apresentam valores negativos de nME, indicando uma tendência sistemática de superestimação da geração pelo modelo antes da correção. O subsistema Nordeste apresenta o maior viés em termos absolutos, seguido pelo SIN

e pelo Sul. As demais métricas de dispersão (nMAE e nRMSE) são comparativamente maiores no subsistema Sul, refletindo a maior variabilidade do regime de ventos e o menor número de usinas nessa região, característica discutida na Seção 2.1 do Capítulo 2. Nota-se que o Nordeste concentra a ampla maioria da capacidade instalada, tendo mais que dobrado ao longo do período avaliado, e apresenta fator de capacidade ligeiramente superior ao do Sul, o que é consistente com a maior qualidade do recurso eólico na região.

Tabela 7 – Métricas de erro normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, por subsistema antes da correção de viés.

Subsistema	P_{inst} (MW)	FC	nME	nMAE	nRMSE	nMAPE (%)
S	1.943 – 2.246	0,34	−0,0550	0,0872	0,1183	8,72
NE	14.146 – 31.464	0,40	−0,0653	0,0666	0,0842	6,66
SIN	16.089 – 33.709	0,39	−0,0638	0,0650	0,0810	6,50

Fonte: Elaboração própria.

Após a aplicação da correção de viés (Tabela 8), o nME é reduzido a valores praticamente nulos em todos os subsistemas, confirmando a eficácia do procedimento na eliminação da tendência sistemática. Além da remoção do viés, observa-se uma melhora expressiva nas demais métricas. Em todos os subsistemas, os valores de nMAE, nRMSE e nMAPE apresentam reduções significativas, com destaque para o Nordeste e o SIN, que atingem os menores patamares de erro após a correção. No agregado do SIN, os resultados corrigidos apresentam os menores valores de erro entre todos os níveis de análise, evidenciando o benefício da agregação espacial na compensação de erros individuais. O subsistema Sul, embora também apresente melhora consistente, mantém valores de erro relativamente superiores aos demais, comportamento coerente com as características de maior variabilidade e menor número de usinas da região.

Tabela 8 – Métricas de erro normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, por subsistema após a correção de viés.

Subsistema	P_{inst} (MW)	FC	nME	nMAE	nRMSE	nMAPE (%)
S	1.943 – 2.246	0,34	0,0000	0,0619	0,0846	6,19
NE	14.146 – 31.464	0,40	0,0001	0,0278	0,0359	2,78
SIN	16.089 – 33.709	0,39	0,0001	0,0265	0,0343	2,65

Fonte: Elaboração própria.

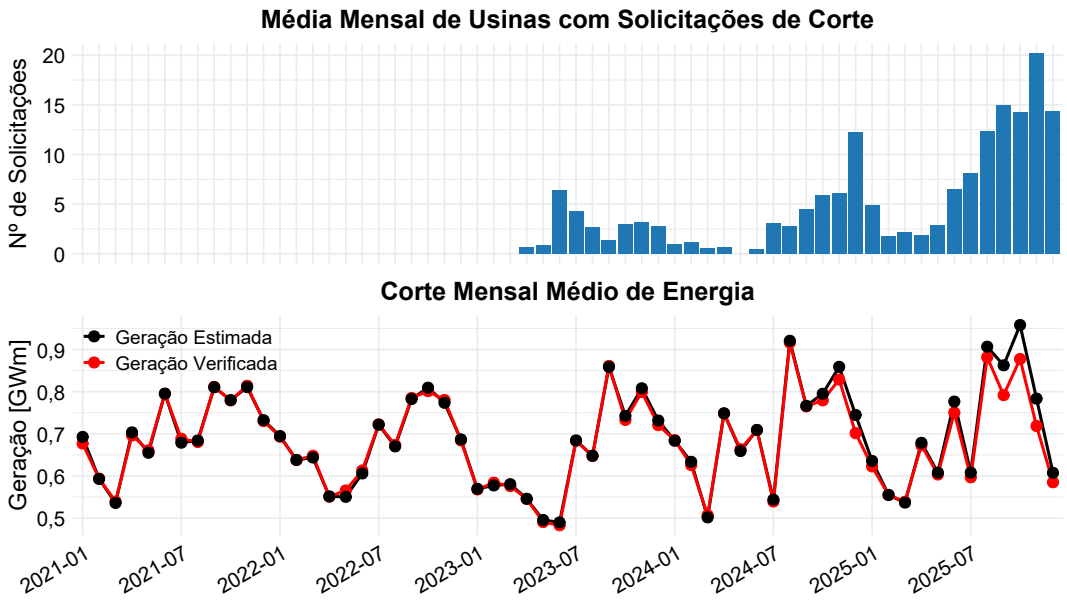
Finalmente, a agregação mensal da geração estimada por subsistema elétrico fornece uma visão consolidada da geração potencial e das perdas de energia associadas a eventos de corte ao longo do tempo. Essa análise constitui um insumo relevante para o planejamento energético de médio e longo prazo, bem como para a avaliação de políticas e estratégias voltadas à integração de fontes de energia renovável no sistema elétrico brasileiro.

As Figuras 86, 87 e 88 apresentam, para os subsistemas Sul, Nordeste e para o SIN respectivamente, a evolução mensal da geração eólica estimada e verificada, bem como o número médio mensal de usinas com solicitações de corte de geração, ao longo do período de 2021 a

2025. Cada figura é composta por dois gráficos dispostos verticalmente. O gráfico superior exibe, em barras, a média mensal do número de usinas submetidas a solicitações de corte. O gráfico inferior apresenta as séries mensais de geração verificada (em vermelho) e geração estimada (em preto), expressas em GWm, permitindo visualizar a diferença entre o potencial de geração e a produção efetivamente realizada.

No subsistema Sul (Figura 86), observa-se que as séries de geração estimada e verificada apresentam valores próximos na maior parte dos meses. No entanto, devido à maior variabilidade do regime de ventos e à menor capacidade instalada na região, as divergências entre os valores estimados e verificados tornam-se mais perceptíveis em termos relativos, embora o modelo mantenha boa qualidade de estimação. Diferentemente do subsistema Nordeste, não se observa no Sul uma sazonalidade tão aparente na energia perdida por restrições operativas, indicando que os cortes de geração nesta região ocorrem de forma um pouco mais distribuída ao longo do ano, sem concentração em períodos específicos.

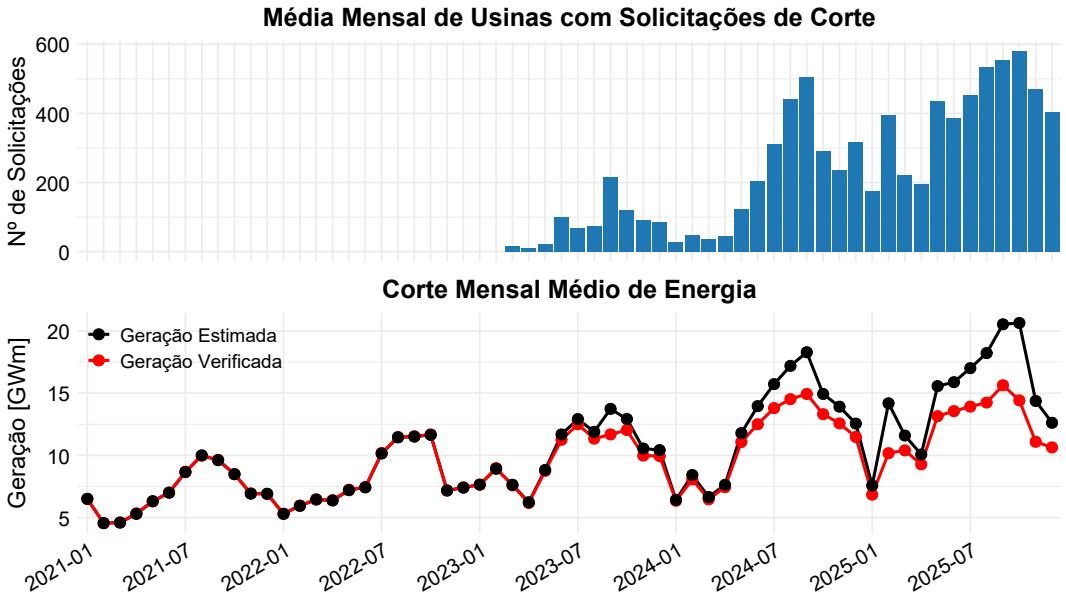
Figura 86 – Evolução mensal da geração estimada e verificada com indicação de cortes para o subsistema Sul.



Fonte: Elaboração própria.

No subsistema Nordeste (Figura 87), nota-se o crescimento progressivo da capacidade instalada ao longo dos anos, acompanhado de um aumento correspondente nas perdas energéticas por *curtailment*. Além disso, devido à sazonalidade natural do recurso eólico na região, observam-se perdas mais expressivas no segundo semestre de cada ano, período característico de ventos mais intensos no Nordeste. Essa combinação entre o crescimento da capacidade instalada e a concentração sazonal dos ventos resulta em uma demanda crescente por cortes de geração nos meses de maior produção eólica, evidenciando o desafio operacional de acomodar volumes cada vez maiores de geração eólica no sistema.

Figura 87 – Evolução mensal da geração estimada e verificada com indicação de cortes para o subsistema Nordeste.



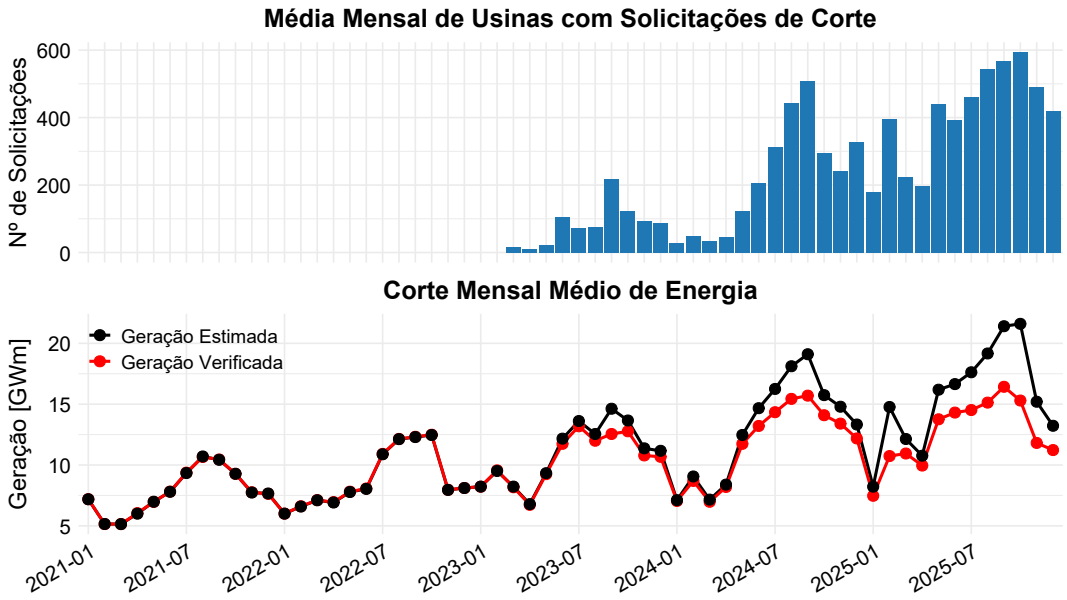
Fonte: Elaboração própria.

No agregado do SIN (Figura 88), o comportamento observado reflete a composição dos subsistemas, com predominância da contribuição do Nordeste tanto em termos de geração quanto de cortes. A análise agregada permite dimensionar o montante total de energia não fornecida ao sistema em decorrência das restrições operativas ao longo do período avaliado. A boa aderência entre as séries estimada e verificada nos meses com baixa incidência de cortes reforça a confiabilidade do modelo de estimação e sua adequação para a quantificação das perdas associadas aos cortes de geração em escala nacional.

Embora os resultados estimados ajustados apresentem excelente aderência à geração verificada nas agregações mensais, especialmente nos períodos sem *curtailment*, quando a análise é realizada em resolução semi-horária podem ser observadas variações mais significativas entre as duas séries. Esse comportamento é esperado, uma vez que nas agregações mensais as variações positivas e negativas tendem a se compensar, resultando em um erro médio reduzido. A Figura 89 ilustra essa característica por meio dos gráficos de dispersão entre a geração verificada e a geração estimada, em resolução semi-horária, para os subsistemas Sul e Nordeste. Em ambos os casos, foram excluídos da análise os instantes em que havia solicitações de corte de geração, de modo que os pontos representados correspondem exclusivamente a períodos de operação sem *curtailment*.

No subsistema Sul (gráfico à esquerda), observa-se uma dispersão mais volumosa em torno da reta identidade. Esse comportamento é atribuído à maior variabilidade do regime de ventos na região e ao número significativamente menor de usinas que compõem o subsistema, o que reduz o efeito de compensação entre sub e superestimações individuais. No subsistema

Figura 88 – Evolução mensal da geração estimada e verificada com indicação de cortes para o SIN.



Fonte: Elaboração própria.

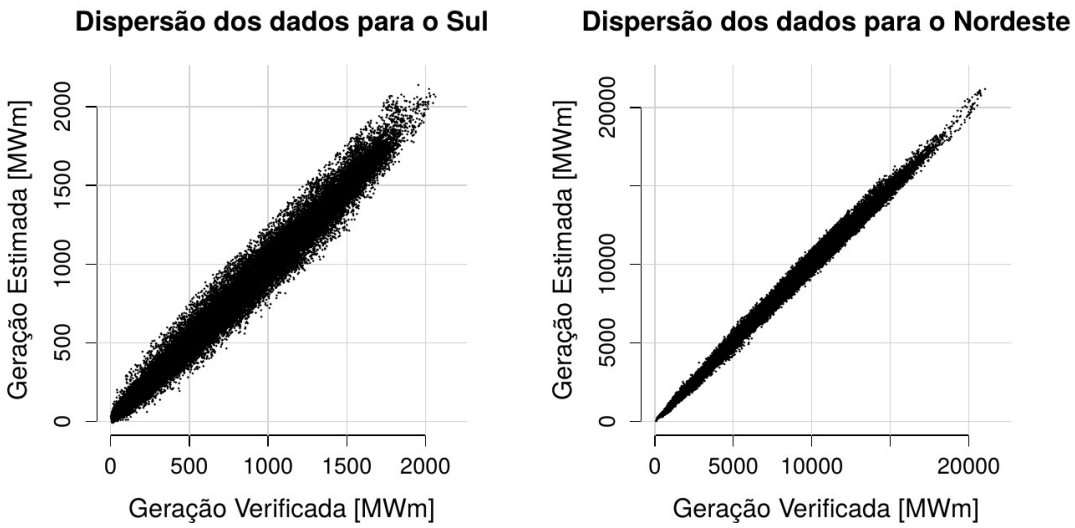
Nordeste (gráfico à direita), apesar da maior capacidade instalada e das faixas de geração mais amplas em valores absolutos, a dispersão apresenta-se proporcionalmente mais concentrada. Isso se deve à maior estabilidade das séries de dados na região e ao número expressivamente maior de usinas, que proporciona um efeito de compensação mais eficaz entre os erros individuais de estimação. Em ambos os subsistemas, a nuvem de pontos apresenta-se razoavelmente simétrica em relação à reta identidade, confirmando a ausência de viés sistemático após a etapa de correção. A dispersão observada na resolução semi-horária é inerente à natureza estocástica do recurso eólico e às incertezas associadas ao modelo de estimação, sendo atenuada quando os resultados são avaliados em horizontes temporais mais longos.

Com base na aplicação do modelo a períodos afetados por corte, as perdas médias anuais de geração eólica foram estimadas em aproximadamente 4,3% em 2023, 9,7% em 2024 e 18,9% em 2025, com indicações de tendência de alta contínua ao longo de 2026. Esses resultados refletem a frequência e severidade crescentes das restrições operacionais associadas à rápida expansão da capacidade instalada, particularmente em regiões que enfrentam limitações na transmissão de energia para outras regiões.

Para quantificar o desempenho do modelo de estimação, foi utilizado o coeficiente de eficiência de Nash-Sutcliffe (NSE), proposto por [161] e amplamente empregado na avaliação de modelos preditivos, definido pela Equação (B.7) do Apêndice B, com a geração estimada no lugar da prevista. Valores mais próximos de 100% indicam maior aderência entre as séries estimada e verificada.

A avaliação numérica dos resultados confirma a robustez da metodologia proposta. Ao

Figura 89 – Dispersão entre geração verificada e geração estimada em resolução semi-horária para os subsistemas Sul e Nordeste.



Fonte: Elaboração própria.

considerar exclusivamente períodos sem ocorrência de corte de geração, o modelo alcança, na resolução semi-horária, eficiência (NSE) de 98,27% para o subsistema Sul, 99,38% para o Nordeste e 99,35% para o SIN. Quando avaliado em base agregada mensal, os valores elevam-se para 99,50% no Sul, 99,99% no Nordeste e 99,99% no SIN, evidenciando que a compensação de erros nas agregações temporais mais longas contribui significativamente para a melhoria do desempenho do modelo em todos os subsistemas.

Esses resultados demonstram a elevada qualidade do processo de estimação, evidenciando sua utilidade em diferentes aplicações. Além de possibilitar uma análise mais precisa dos impactos energéticos decorrentes de eventos de *curtailment*, os resultados atestam a capacidade do modelo para subsidiar processos de compensação financeira dos agentes afetados. Adicionalmente, as estimativas produzidas constituem um importante insumo para procedimentos de correção e preenchimento de dados e para os modelos de previsão de geração eólica, contribuindo para o aumento da precisão e da confiabilidade desses processos.

6.3 Resultados da aplicação da metodologia de ajuste nas previsões de vento

Embora a etapa de modelagem do vento previsto esteja intrinsecamente relacionada ao modelo de previsão de geração eólica, a eficácia dessa abordagem pode ser analisada isoladamente a partir de seu impacto sobre as previsões meteorológicas brutas. Assim, nesta seção são apresentados os resultados da aplicação da metodologia de correção de vento descrita na Seção 5.3.3 do Capítulo 5, destacando seus efeitos na redução dos erros e na melhoria da qualidade das previsões de velocidade do vento.

6.3.1 Análise da magnitude do vento previsto

Antes de apresentar os resultados do processo de ajuste das previsões de vento, é importante avaliar o comportamento das previsões brutas, de modo a estabelecer uma referência para a análise dos ganhos obtidos com a metodologia proposta. As Figuras 90 e 91 apresentam o vento médio nas primeiras 24 horas de previsão, para cada ponto de análise, considerando o vento verificado e as previsões dos três modelos meteorológicos avaliados (ECMWF, GFS e GEFS). Cada ponto é colorido segundo a magnitude da velocidade média do vento, em uma escala comum aos quatro painéis de cada figura, permitindo comparação direta entre verificação e previsões.

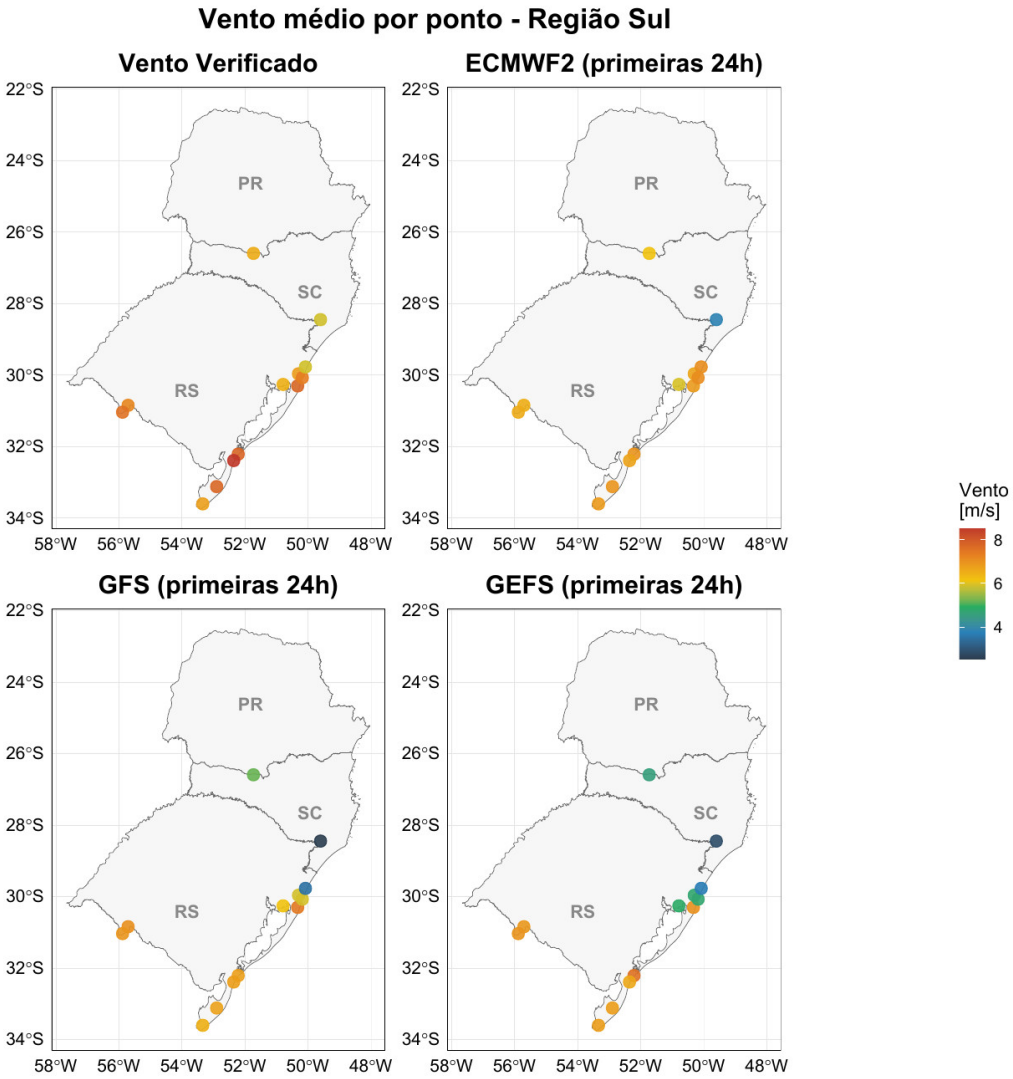
Cabe esclarecer, conforme apresentado na Seção 5.3.1 do Capítulo 5, que o ECMWF disponibiliza dois produtos distintos: o ECMWF-HRES (*High Resolution*), de alta resolução e caráter determinístico, e o ECMWF-Ens (*Ensemble*), de caráter probabilístico. Ao longo deste trabalho, utiliza-se o produto *Ensemble*. Para simplificar a notação, adota-se ECMWF1 para o HRES e ECMWF2 para o *Ensemble*. No texto deste capítulo, todas as menções ao "ECMWF" referem-se ao *Ensemble*; a sigla ECMWF2 é mantida apenas no interior de algumas figuras.

Observa-se, na Região Sul, que o vento verificado apresenta intensidades mais elevadas nos pontos próximos ao litoral do Rio Grande do Sul e nas regiões mais ao sul do estado, enquanto os pontos no interior de Santa Catarina apresentam velocidades inferiores. Ao comparar com as previsões dos modelos, percebe-se uma tendência geral de subestimação, os tons nos painéis dos modelos são, em geral, mais frios (menores velocidades) do que no painel da verificação. Essa subestimação se mostra mais acentuada nos modelos GFS e GEFS, sobretudo nos pontos do litoral catarinense, enquanto o ECMWF apresenta um comportamento mais aderente, com magnitudes próximas das verificadas em diversos pontos.

Na Região Nordeste, o vento verificado é caracterizado por velocidades elevadas no litoral do Rio Grande do Norte, Ceará e Piauí, e por pontos com intensidades intermediárias no interior da Bahia. Os modelos meteorológicos reproduzem qualitativamente esse padrão espacial, com ventos mais fortes no litoral e mais fracos no interior, mas apresentam diferenças importantes de magnitude. Nota-se que os modelos tendem a subestimar significativamente os ventos nos pontos do interior da Bahia e do interior do Piauí, ao passo que reproduzem com mais fidelidade as velocidades observadas nos pontos litorâneos do Nordeste setentrional.

As Figuras 92 e 93 apresentam, agora, o vento médio nas primeiras 24 horas após a aplicação da correção do vento bruto, para as Regiões Sul e Nordeste, respectivamente. Diferentemente das Figuras 90 e 91 (que apresentavam os quatro painéis correspondentes ao vento verificado e às previsões dos três modelos meteorológicos, nas quais se observavam divergências de magnitude no vento médio entre eles), no caso das previsões corrigidas as diferenças entre os três modelos tornam-se tão reduzidas que as cores associadas a cada ponto não se alteram perceptivelmente quando se troca o modelo. Por essa razão, optou-se por exibir,

Figura 90 – Vento médio por ponto na Região Sul — verificação e previsões (primeiras 24 horas).

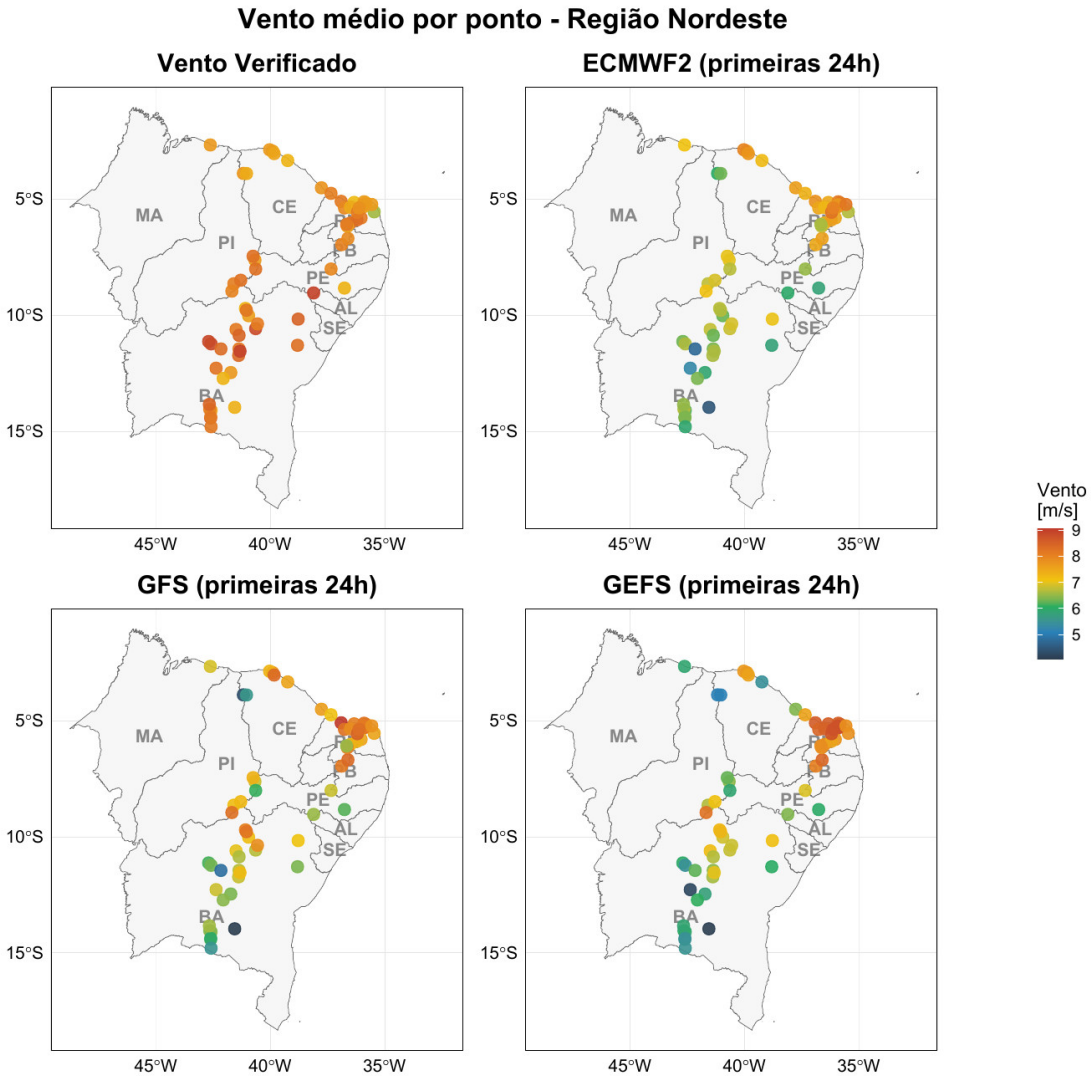


Fonte: Elaboração própria.

no painel à direita, a média simples das previsões corrigidas dos três modelos, representação que evidencia a aderência conjunta do *ensemble* corrigido ao vento verificado, sem a redundância visual de mostrar quatro painéis quase idênticos.

A comparação direta entre os painéis das figuras revela aderência espacial consistente entre o vento verificado e a média dos modelos corrigidos, tanto na Região Sul quanto na Região Nordeste. Os padrões espaciais característicos de cada região (ventos mais intensos no litoral do Rio Grande do Sul, gradiente decrescente em direção ao interior catarinense, regime intenso no litoral do Nordeste setentrional e velocidades intermediárias no interior da Bahia) aparecem fielmente reproduzidos pelas previsões corrigidas. Quando contrastado com o padrão observado nas figuras das previsões brutas, no qual havia subestimação visível em diversos pontos, em particular no litoral catarinense e no interior da Bahia, fica evidente o efeito da correção em aproximar a magnitude prevista da magnitude verificada.

Figura 91 – Vento médio por ponto na Região Nordeste — verificação e previsões (primeiras 24 horas).



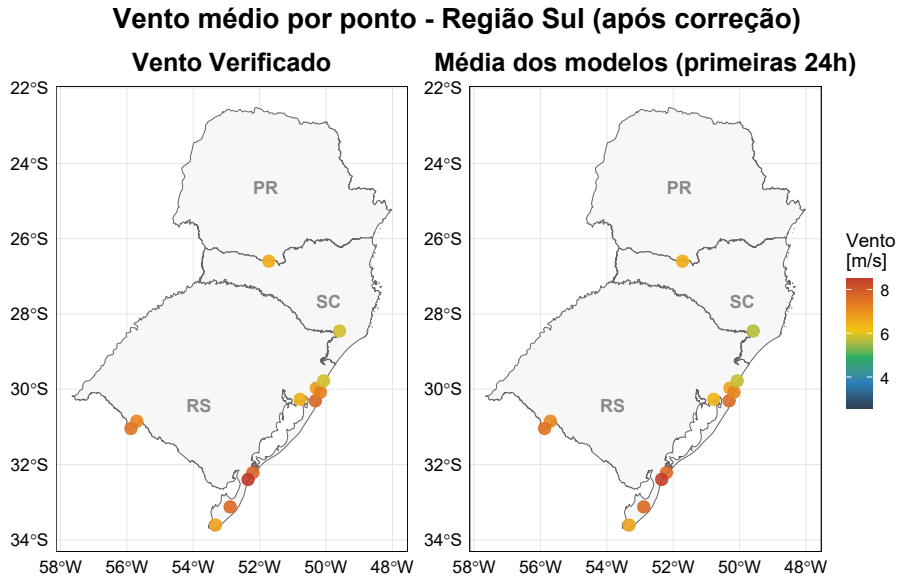
Fonte: Elaboração própria.

6.3.2 Panorama nacional do viés médio

Por fim, para fornecer um panorama mais amplo do desempenho dos modelos meteorológicos sobre todo o território de interesse, as Figuras 94 e 95 apresentam o viés médio do vento previsto pelo modelo GEFS, calculado sobre os primeiros oito dias de previsão (horizonte de 192 horas) e agregado por estado, antes e após a aplicação da correção. Cada estado é colorido segundo a magnitude e o sinal do viés médio: tons em azul indicam superestimação do vento previsto em relação ao verificado, tons em vermelho indicam subestimação, e a intensidade da cor é proporcional à magnitude do viés. A escala de cores é mantida idêntica entre as duas figuras para permitir comparação visual direta.

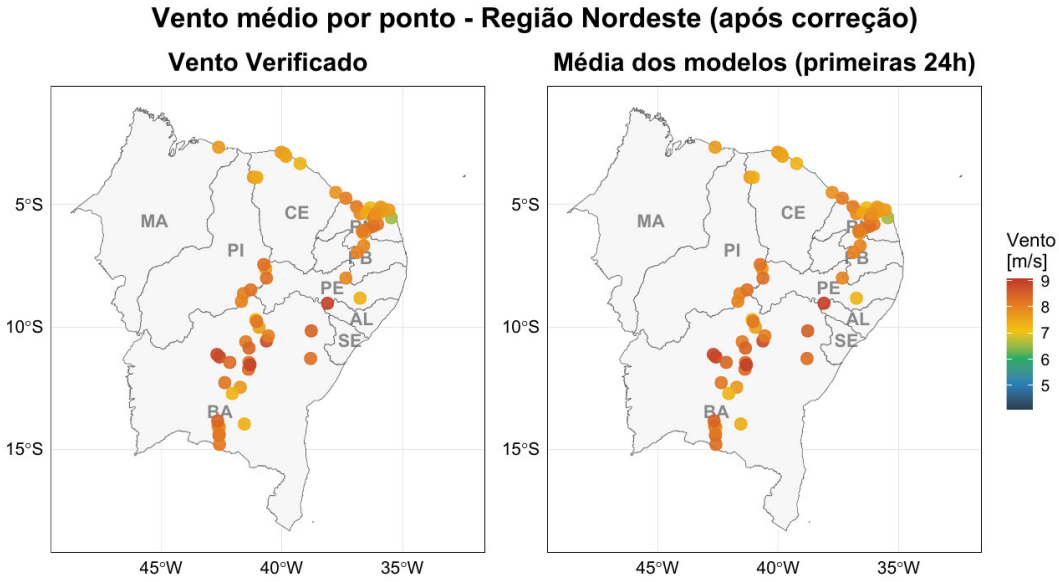
A Figura 94 permite identificar padrões regionais claros nas previsões brutas. Na Região Sul, predomina o viés positivo, sendo Santa Catarina o estado com a maior subestimação observada. Na Região Nordeste, o padrão é heterogêneo: Bahia e Pernambuco apresentam o

Figura 92 – Vento médio por ponto na Região Sul — verificação e média corrigida (primeiras 24 horas).



Fonte: Elaboração própria.

Figura 93 – Vento médio por ponto na Região Nordeste — verificação e média corrigida (primeiras 24 horas).

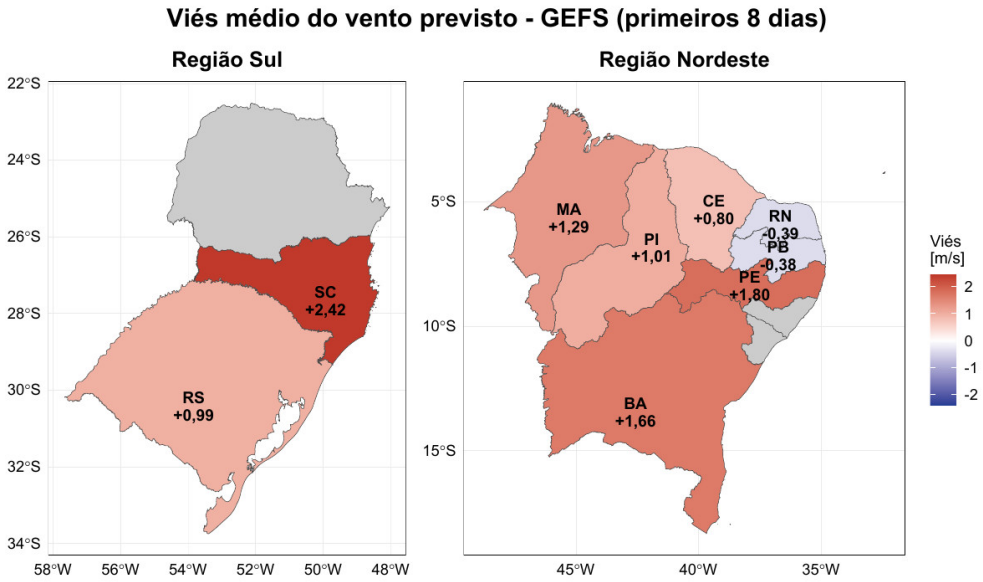


Fonte: Elaboração própria.

viés positivo mais marcante; Piauí, Ceará e Maranhão apresentam viés levemente positivo; e Rio Grande do Norte e Paraíba apresentam medianas mais próximas de zero, com leve tendência de superestimação (pequeno viés negativo). Estados sem capacidade eólica instalada relevante (e, portanto, sem pontos de análise associados) aparecem em cinza.

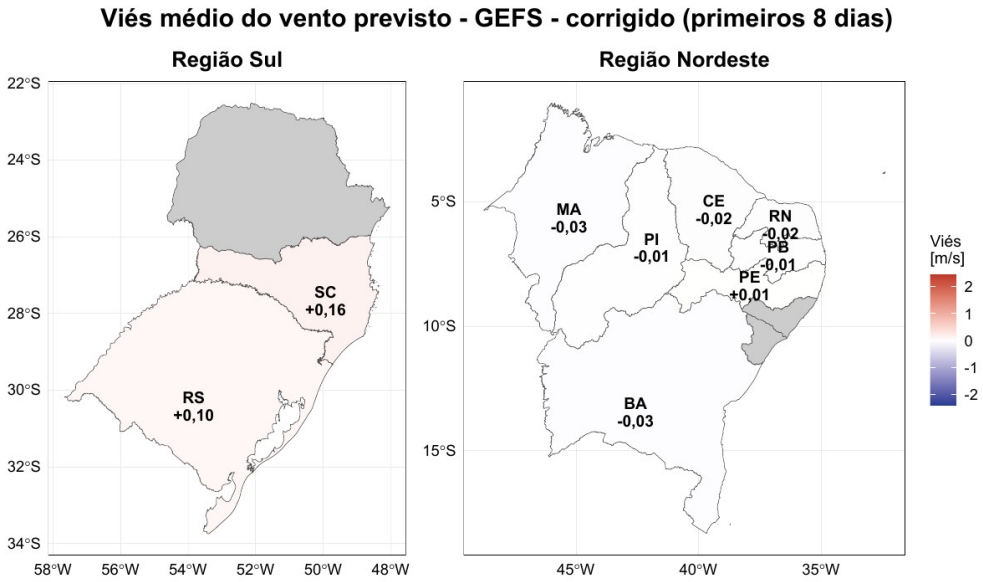
O contraste com a Figura 95, que apresenta o panorama após a correção, é expressivo. Após a aplicação da metodologia, todos os estados convergem para tons muito próximos do branco, com valores de viés residual de pequena magnitude e com sinal alternando ligeira-

Figura 94 – Viés médio do vento do GEFS por estado (primeiros 8 dias) — previsões brutas.



Fonte: Elaboração própria.

Figura 95 – Viés médio do vento do GEFS por estado (primeiros 8 dias) — previsões corrigidas.



Fonte: Elaboração própria.

mente entre positivo e negativo. Esse resultado evidencia que a correção, aplicada de forma individualizada por intervalo do dia e por ponto de análise, é eficaz na remoção do viés sistemático identificado nas previsões brutas, ao mesmo tempo que preserva a coerência espacial das previsões em todo o conjunto de estados analisados.

Em síntese, considerando os resultados apresentados nesta seção e as análises desagregadas por ponto de análise e por estado apresentadas no Apêndice C, observa-se que a metodologia de correção proposta atinge os objetivos esperados em três dimensões complementares: (i) aproxima a magnitude do vento previsto da magnitude do vento verificado, conforme se observa

nos mapas de pontos; (ii) remove o viés sistemático ao longo do horizonte de previsão, atenuando inclusive as oscilações horárias associadas ao ciclo diurno do vento; e (iii) reduz a dispersão dos erros tanto nos quantis centrais quanto nos extremos. Os ganhos são consistentes entre os três modelos meteorológicos avaliados e se preservam desde as análises mais detalhadas até os resultados agregados, indicando robustez da metodologia para aplicação no previsor eólico em escala do SIN.

6.4 Resultados do modelo de previsão de geração por subsistema

Esta seção tem como objetivo apresentar e analisar os resultados obtidos com o modelo de previsão de geração eólica, cuja arquitetura foi apresentada na Seção 5.2 do Capítulo 5. O desempenho desse tipo de modelo depende da qualidade dos dados observados utilizados em seu ajuste e validação, bem como da correta representação da relação entre a velocidade do vento e a potência gerada. Contudo, a principal fonte de incerteza está associada às previsões meteorológicas. Mesmo com curvas V $\times$ P devidamente calibradas, conforme descrito na Seção 5.5, desvios entre as velocidades de vento previstas e as efetivamente observadas comprometem a precisão das previsões de geração. Assim, a avaliação dos resultados apresentada nesta seção permite analisar o desempenho conjunto das etapas de previsão meteorológica, correção de vento e conversão de vento em potência.

Como os resultados apresentados nesta seção são analisados em nível de subsistema, é útil destacar inicialmente algumas conclusões obtidas a partir das análises mais desagregadas por agrupamentos de usinas, apresentadas no Apêndice C, que auxiliam na interpretação dos comportamentos observados em cada subsistema. Essas análises mostram que, diferentemente do modelo de estimação, os erros de previsão apresentam relação mais clara com a localização geográfica: os menores erros concentram-se em agrupamentos do Nordeste e os maiores no subsistema Sul, comportamento coerente com a maior variabilidade do regime de ventos na região Sul e a maior previsibilidade dos padrões atmosféricos no Nordeste. Observa-se ainda que as distribuições de erro permanecem aproximadamente centradas em zero, sem viés sistemático significativo, porém com dispersão superior à observada na estimação, refletindo a incerteza inerente à previsão meteorológica. Com esse contexto, os resultados são apresentados a seguir em nível de subsistema.

6.4.1 Previsões para o primeiro dia à frente

De forma análoga à análise apresentada para o modelo de estimação por subsistema, são discutidos a seguir os resultados equivalentes para o modelo de previsão, permitindo uma comparação direta entre as duas abordagens. Nas análises iniciais, são apresentados os desvios de previsão para o primeiro dia à frente, com o objetivo de fornecer uma visão geral do desempenho do modelo e de sua dependência em relação aos dados históricos verificados. Posteriormente, são

exploradas as características do modelo de previsão para os demais horizontes, contemplando os diversos dias à frente.

As Figuras 96 e 97 apresentam, para os subsistemas Sul e Nordeste respectivamente, a comparação entre a geração verificada real, a geração verificada tratada e a geração prevista pelo modelo, bem como o número de usinas com solicitações de corte de geração, para o período de 13 a 22 de outubro de 2024. Cada figura é composta por dois gráficos dispostos verticalmente. O gráfico superior exibe as séries temporais de geração verificada real (em vermelho), geração verificada tratada (em laranja) e geração prevista (em preto), em intervalos de meia hora, com valores expressos em GWm. A geração verificada tratada corresponde à série na qual os efeitos das restrições operativas por *curtailment* foram removidos, representando o comportamento que seria observado sob condições de maior normalidade de operação. O gráfico inferior apresenta o número de usinas submetidas a solicitações de corte de geração ao longo do mesmo período.

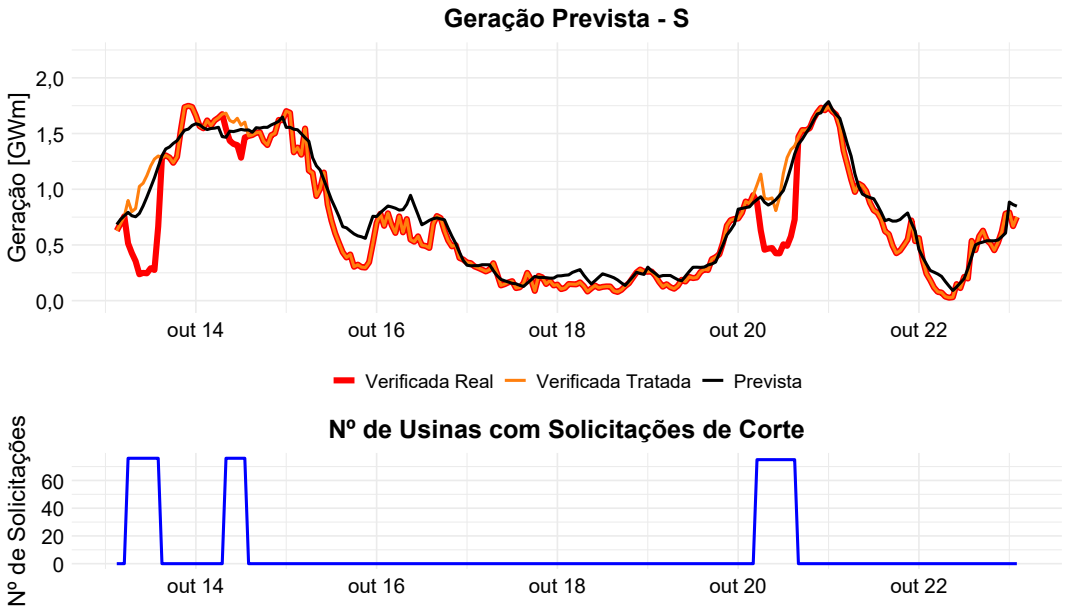
No subsistema Sul (Figura 96), observa-se um alto nível de concordância entre a geração prevista e a geração verificada tratada durante todo o período avaliado, destacando a capacidade do modelo de antecipar adequadamente o comportamento esperado da geração eólica. Essa concordância é um aspecto fundamental para a aplicação operacional do modelo de previsão, pois garante que as decisões de operação fundamentadas nas previsões possam ser tomadas com maior confiabilidade. Nos intervalos em que o número de solicitações de corte aumenta, a geração verificada real apresenta valores inferiores tanto à geração prevista quanto à geração verificada tratada, permitindo quantificar a energia não fornecida em decorrência das restrições operativas e avaliar o impacto desses cortes sobre a aderência da previsão à operação efetiva.

No subsistema Nordeste (Figura 97), o comportamento é análogo, com a geração prevista acompanhando adequadamente a geração verificada tratada nos períodos de operação livre. As diferenças entre a geração verificada real e as demais séries nos momentos de corte evidenciam o impacto das restrições impostas pelo operador do sistema sobre a produção efetiva dos parques eólicos. A geração prevista, por não incorporar essas restrições, representa o potencial de geração antecipado para condições normais de operação. A análise conjunta dos dois subsistemas demonstra a capacidade do modelo de previsão em antecipar o comportamento esperado da geração eólica e, simultaneamente, evidenciar os períodos em que a operação foi comprometida por fatores externos à disponibilidade do recurso eólico.

Para possibilitar uma avaliação mais abrangente do desempenho do modelo de previsão para todo o sistema elétrico brasileiro, as métricas apresentadas no Apêndice B foram aplicadas tomando como referência a geração verificada tratada, de forma a avaliar a qualidade da previsão isoladamente, sem a interferência dos efeitos de *curtailment*.

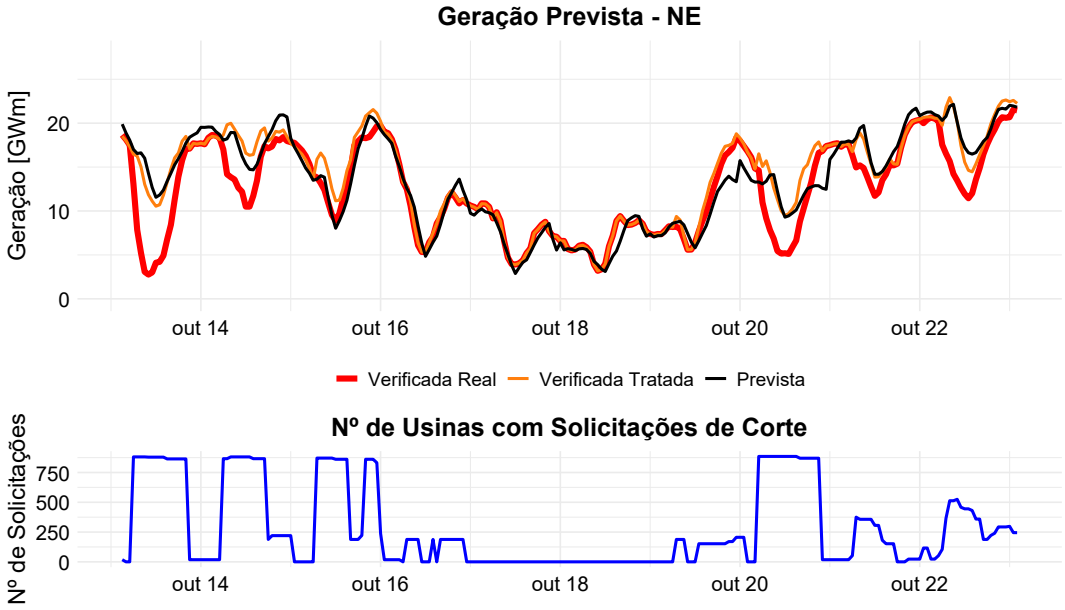
De forma análoga ao realizado nas análises por subsistema para o modelo de estimação, a avaliação apresentada a seguir emprega uma estratégia de normalização distinta daquela utilizada para grupos de parques eólicos. Enquanto, nas análises por agrupamento, os erros foram normalizados pela amplitude da série de geração, nesta avaliação os erros absolutos foram

Figura 96 – Geração prevista, verificada tratada e verificada real com indicação de cortes para o subsistema Sul.



Fonte: Elaboração própria.

Figura 97 – Geração prevista, verificada tratada e verificada real com indicação de cortes para o subsistema Nordeste.



Fonte: Elaboração própria.

normalizados pela capacidade instalada disponível em cada instante, multiplicada pelo fator de capacidade médio observado ao longo do período analisado.

Em termos práticos, essa referência corresponde aproximadamente à geração eólica média do SIN, conferindo maior uniformidade à interpretação dos resultados. A adoção dessa abordagem justifica-se pelo significativo crescimento da capacidade instalada de geração eólica ao longo do horizonte de análise, conforme ilustrado na Figura 125 do Apêndice C. Quando os dados são agregados para todo o SIN, esse crescimento provoca alterações substanciais na amplitude da série temporal, tornando menos adequada a utilização de uma única referência baseada na amplitude dos dados.

Dessa forma, a normalização proposta permite comparar os erros de previsão sob diferentes condições operativas e níveis de geração, reduzindo a influência das mudanças estruturais associadas à expansão da capacidade instalada e proporcionando uma avaliação mais consistente do desempenho do modelo ao longo do tempo.

A Tabela 9 apresenta as métricas de erro normalizadas do modelo de previsão de geração eólica para os subsistemas Sul, Nordeste e para o SIN. Além das métricas de erro, são apresentadas a faixa de potência instalada ao longo do período de análise (2021 a 2025) e o fator de capacidade médio de cada subsistema, fornecendo contexto sobre a dimensão e o aproveitamento do recurso eólico em cada região. As métricas de erro avaliadas são o erro médio normalizado (nME), o erro médio absoluto normalizado (nMAE), a raiz do erro quadrático médio normalizado (nRMSE) e o erro percentual absoluto médio normalizado (nMAPE).

Tabela 9 – Métricas de erro normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio do modelo de previsão por subsistema.

Subsistema	P_{inst} (MW)	FC	nME	nMAE	nRMSE	nMAPE (%)
S	$1.943 - 2.246$	0,34	$-0{,}0139$	0,1682	0,2265	16,82
NE	$14.146 - 31.464$	0,39	$-0{,}0034$	0,0944	0,1230	9,44
SIN	$16.089 - 33.709$	0,39	$-0{,}0049$	0,0885	0,1153	8,85

Fonte: Elaboração própria.

Observa-se que os valores de nME são praticamente nulos em todos os subsistemas, indicando ausência de viés sistemático nas previsões. As métricas de dispersão (nMAE, nRMSE e nMAPE) são comparativamente maiores no subsistema Sul, refletindo a maior variabilidade do regime de ventos e o menor número de usinas nessa região, característica discutida na Seção 2.1 do Capítulo 2. O Nordeste e o SIN atingem os menores patamares de erro, sendo que o agregado do SIN apresenta os menores valores entre todos os níveis de análise, o que evidencia o benefício da agregação espacial na compensação de erros individuais de previsão. Nota-se ainda que o Nordeste concentra a ampla maioria da capacidade instalada, tendo mais que dobrado ao longo do período avaliado, e apresenta fator de capacidade ligeiramente superior ao do Sul, comportamento consistente com a maior qualidade do recurso eólico na região. O subsistema Sul, embora apresente bom desempenho, mantém valores de erro relativamente superiores aos

demais, comportamento coerente com as características de maior variabilidade e menor número de usinas da região.

Finalmente, a agregação mensal da geração prevista por subsistema elétrico fornece uma visão consolidada da geração potencial antecipada e das perdas de energia associadas a eventos de corte ao longo do tempo. Essa análise constitui um insumo relevante para o planejamento energético de curto e médio prazo, bem como para a avaliação de políticas e estratégias voltadas à integração de fontes de energia renovável no sistema elétrico brasileiro.

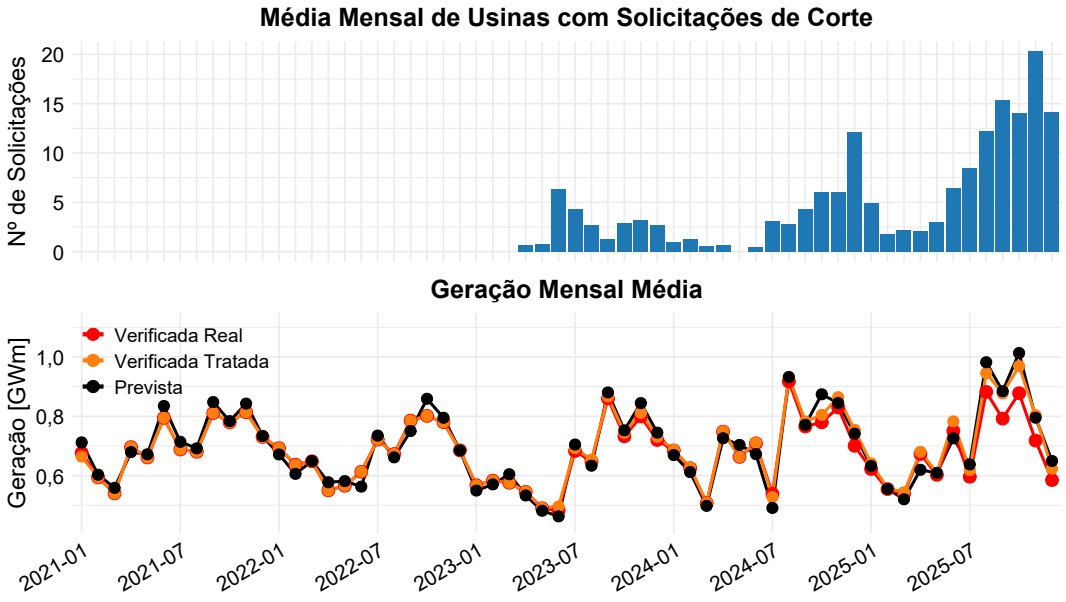
As Figuras 98, 99 e 100 apresentam, para os subsistemas Sul, Nordeste e para o SIN respectivamente, a evolução mensal da geração eólica prevista, verificada tratada e verificada real, bem como o número médio mensal de usinas com solicitações de corte de geração, ao longo do período de 2021 a 2025. Cada figura é composta por dois gráficos dispostos verticalmente. O gráfico superior exibe, em barras, a média mensal do número de usinas submetidas a solicitações de corte. O gráfico inferior apresenta as séries mensais de geração verificada real (em vermelho), geração verificada tratada (em laranja) e geração prevista (em preto), expressas em GWm, permitindo visualizar a diferença entre o potencial de geração antecipado pelo modelo, o potencial efetivamente disponível e a produção realizada após as restrições operativas.

No subsistema Sul (Figura 98), observa-se que as séries de geração prevista e verificada tratada apresentam valores próximos na maior parte dos meses. No entanto, devido à maior variabilidade do regime de ventos e à menor capacidade instalada na região, as divergências entre os valores previstos e verificados tornam-se mais perceptíveis em termos relativos, embora o modelo mantenha boa qualidade de previsão. Diferentemente do subsistema Nordeste, não se observa no Sul uma sazonalidade tão aparente na energia perdida por restrições operativas, indicando que os cortes de geração nesta região ocorrem de forma um pouco mais distribuída ao longo do ano, sem concentração em períodos específicos.

No subsistema Nordeste (Figura 99), nota-se o crescimento progressivo da capacidade instalada ao longo dos anos, acompanhado de um aumento correspondente nas perdas energéticas por *curtailment*, evidenciado pela diferença crescente entre as séries de geração verificada tratada e verificada real. Além disso, devido à sazonalidade natural do recurso eólico na região, observam-se perdas mais expressivas no segundo semestre de cada ano, período característico de ventos mais intensos no Nordeste. Essa combinação entre o crescimento da capacidade instalada e a concentração sazonal dos ventos resulta em uma demanda crescente por cortes de geração nos meses de maior produção eólica, evidenciando o desafio operacional de acomodar volumes cada vez maiores de geração eólica no sistema. A geração prevista mantém boa aderência à geração verificada tratada ao longo de todo o período, demonstrando a capacidade do modelo de antecipar adequadamente o comportamento esperado mesmo diante do crescimento expressivo da capacidade instalada.

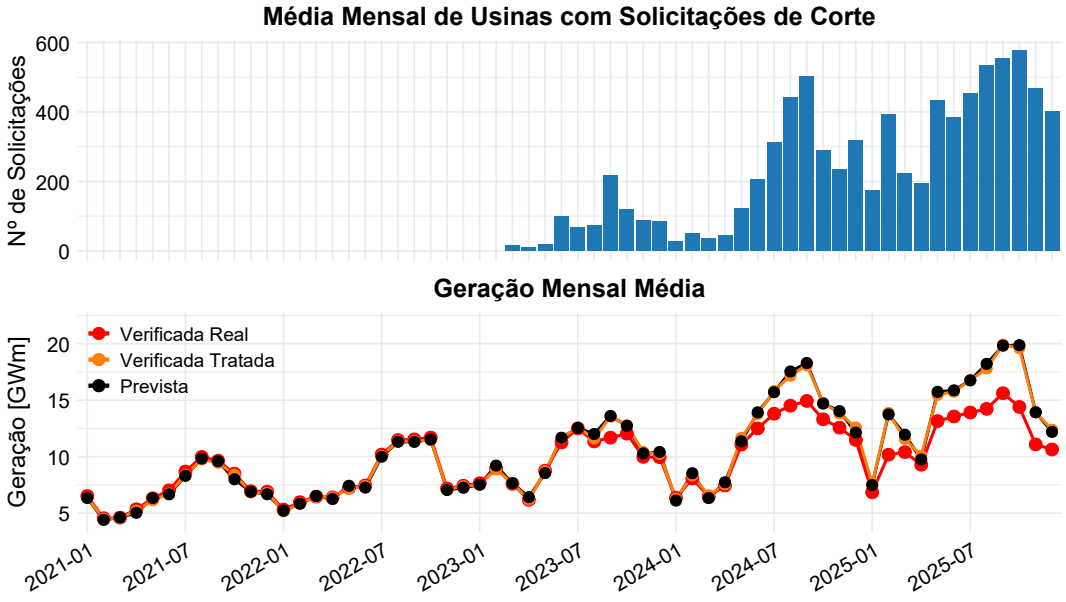
No agregado do SIN (Figura 100), o comportamento observado reflete a composição dos subsistemas, com predominância da contribuição do Nordeste tanto em termos de geração quanto

Figura 98 – Evolução mensal da geração prevista, verificada tratada e verificada real com indicação de cortes para o subsistema Sul.



Fonte: Elaboração própria.

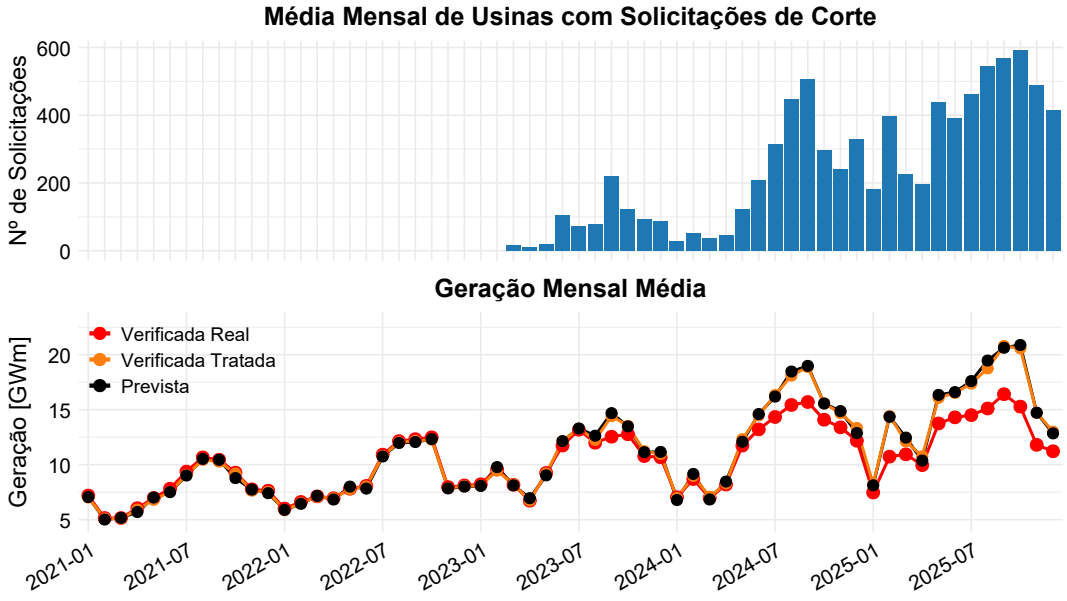
Figura 99 – Evolução mensal da geração prevista, verificada tratada e verificada real com indicação de cortes para o subsistema Nordeste.



Fonte: Elaboração própria.

de cortes. A análise agregada permite dimensionar o montante total de energia não fornecida ao sistema em decorrência das restrições operativas ao longo do período avaliado. A boa aderência entre as séries prevista e verificada tratada nos meses com baixa incidência de cortes reforça a confiabilidade do modelo de previsão e sua adequação para a antecipação do comportamento da geração eólica em escala nacional, fornecendo subsídios para a operação e o planejamento do sistema elétrico.

Figura 100 – Evolução mensal da geração prevista, verificada tratada e verificada real com indicação de cortes para o SIN.



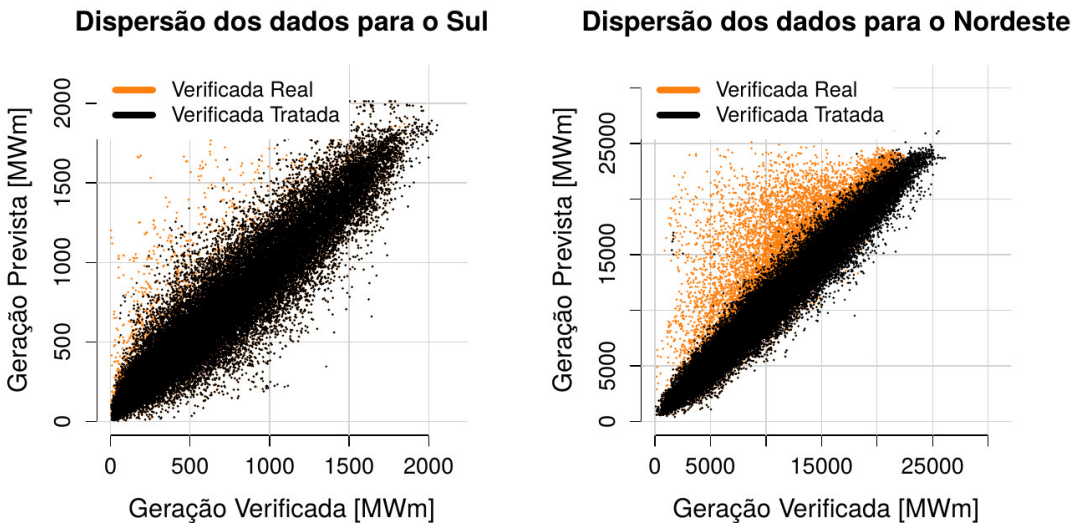
Fonte: Elaboração própria.

Embora os resultados previstos ajustados apresentem excelente aderência à geração verificada tratada nas agregações mensais, especialmente nos períodos sem *curtailment*, quando a análise é realizada em resolução semi-horária podem ser observadas variações mais significativas entre as duas séries. Esse comportamento é esperado, uma vez que nas agregações mensais as variações positivas e negativas tendem a se compensar, resultando em um erro médio reduzido. A Figura 101 ilustra essa característica por meio dos gráficos de dispersão entre a geração verificada tratada e a geração prevista, em resolução semi-horária, para os subsistemas Sul e Nordeste.

Em cada gráfico, são representadas duas nuvens de pontos: os pontos em preto correspondem à dispersão entre a geração verificada tratada e a geração prevista, enquanto os pontos em laranja representam a dispersão entre a geração verificada real e a geração prevista. Essa dupla representação permite avaliar simultaneamente o desempenho intrínseco do modelo de previsão (frente à geração que efetivamente ocorreria sem restrições) e o desvio adicional introduzido pelos eventos de *curtailment* sobre a operação real do sistema.

No subsistema Sul (gráfico à esquerda), observa-se uma dispersão mais volumosa em torno da reta identidade. Esse comportamento é atribuído à maior variabilidade do regime de

Figura 101 – Dispersão entre geração verificada tratada e geração prevista em resolução semi-horária para os subsistemas Sul e Nordeste.



Fonte: Elaboração própria.

ventos na região e ao número significativamente menor de usinas que compõem o subsistema, o que reduz o efeito de compensação entre sub e superestimações individuais da previsão. No subsistema Nordeste (gráfico à direita), apesar da maior capacidade instalada e das faixas de geração mais amplas em valores absolutos, a dispersão em preto apresenta-se proporcionalmente mais concentrada. Isso se deve à maior estabilidade das séries de dados na região e ao número expressivamente maior de usinas, que proporciona um efeito de compensação mais eficaz entre os erros individuais de previsão.

Já a dispersão em laranja, que confronta a previsão com a geração verificada real, evidencia o impacto dos cortes operativos sobre o desempenho aparente do modelo. No subsistema Nordeste, onde os índices de perdas por *curtailment* são mais elevados, observa-se um descolamento perceptível da nuvem de pontos em laranja em relação à reta identidade, com tendência sistemática de valores previstos superiores aos verificados reais. Esse comportamento não decorre de deficiências do modelo de previsão, mas sim das restrições operativas impostas pelo operador do sistema, que reduzem a geração efetivamente entregue ao SIN em relação ao potencial disponível. No subsistema Sul, em que as perdas por *curtailment* são proporcionalmente menores, esse descolamento é menos pronunciado, com as duas nuvens de pontos apresentando comportamento mais semelhante.

Em ambos os subsistemas, a nuvem de pontos em preto apresenta-se razoavelmente simétrica em relação à reta identidade, confirmando a ausência de viés sistemático do modelo de previsão. A dispersão observada na resolução semi-horária é inerente à natureza estocástica do recurso eólico e às incertezas associadas ao modelo de previsão, sendo atenuada quando os resultados são avaliados em horizontes temporais mais longos.

Para quantificar o desempenho do modelo de previsão, foi utilizado o coeficiente de

eficiência de Nash-Sutcliffe (NSE), previamente empregado na avaliação do modelo de estimação e definido pela Equação (B.7) do Apêndice B, com as devidas adaptações de notação para o contexto da previsão de geração eólica.

A avaliação numérica dos resultados confirma a robustez da metodologia proposta. Ao considerar exclusivamente períodos sem ocorrência de corte de geração, o modelo de previsão alcança, na resolução semi-horária, eficiência (NSE) de 87,84% para o subsistema Sul, 91,41% para o Nordeste e 91,49% para o SIN. Quando avaliado em base agregada mensal, os valores elevam-se para 89,48% no Sul, 99,48% no Nordeste e 99,48% no SIN, evidenciando que a compensação de erros nas agregações temporais mais longas contribui significativamente para a melhoria do desempenho do modelo em todos os subsistemas, com exceção do subsistema Sul, em que o ganho proporcionado pela agregação temporal é menos pronunciado. A capacidade do modelo de previsão em antecipar o potencial de geração permite, ainda, fornecer subsídios para a quantificação prévia dessas perdas, contribuindo para o planejamento operacional e a tomada de decisão.

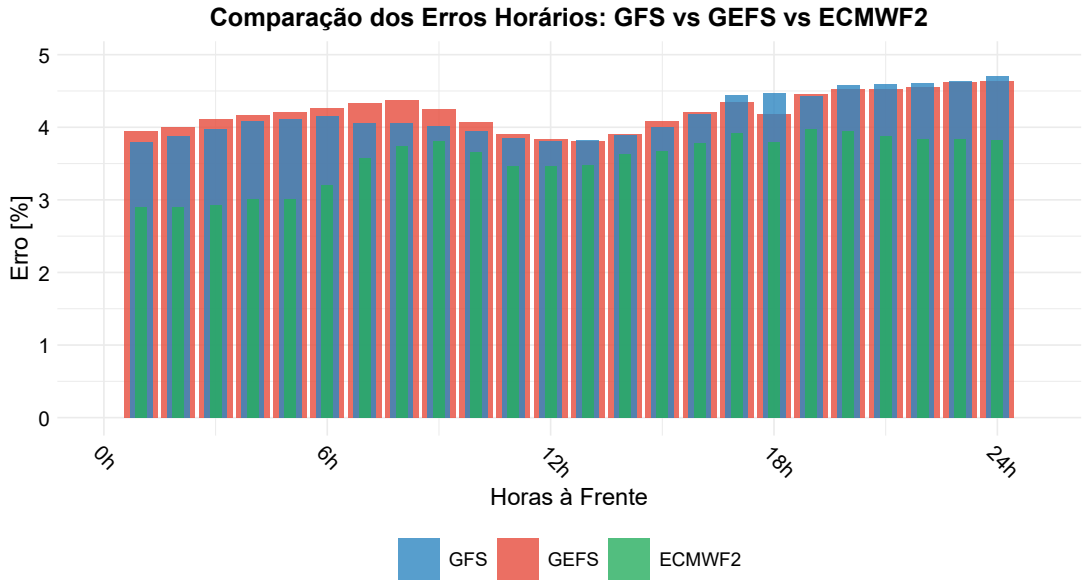
6.4.2 Previsões horárias para o horizonte de 35 dias

Nas seções anteriores, em especial na Subseção 6.4.1, foram apresentadas avaliações de desempenho considerando exclusivamente as previsões para o primeiro dia à frente, com o objetivo de caracterizar o comportamento dos modelos frente aos dados e de modo a possibilitar uma comparação mais direta com os resultados do modelo de estimação. A partir desta subseção, a análise é estendida para o horizonte completo considerado neste trabalho, contemplando previsões de 35 dias (até 34 dias à frente). Esse escopo ampliado permite avaliar não apenas a acurácia pontual dos modelos, mas também a forma como o erro evolui ao longo do horizonte de previsão, evidenciando a degradação progressiva da qualidade preditiva à medida que se afasta do instante de inicialização e fornecendo subsídios para o uso desses modelos em estudos de planejamento de curto prazo da operação eletroenergética.

A Figura 102 apresenta a comparação dos erros horários percentuais das previsões de geração eólica obtidas a partir dos modelos GFS, GEFS e ECMWF (primeiras 24 horas). O eixo das abscissas representa o horizonte de previsão, em horas à frente, enquanto o eixo das ordenadas indica o erro percentual associado a cada hora prevista. A representação em barras sobrepostas, com larguras decrescentes, foi adotada de modo a permitir a visualização simultânea dos três modelos para um mesmo instante, facilitando a comparação direta entre seus desempenhos.

Os resultados evidenciam que o modelo ECMWF apresenta, de forma sistemática, os menores erros ao longo de praticamente todo o horizonte de previsão, com valores variando aproximadamente entre 2,9% e 4,0%. Os modelos GFS e GEFS, por sua vez, exibem comportamento semelhante entre si, com erros tipicamente entre 3,8% e 4,7%, indicando desempenho inferior em relação ao modelo europeu. Observa-se ainda uma tendência geral de crescimento

Figura 102 – Comparação dos erros horários percentuais, normalizados pela potência instalada, das previsões de geração eólica obtidas pelos modelos GFS, GEFS e ECMWF ao longo das primeiras 24 horas de previsão.



Fonte: Elaboração própria.

do erro à medida que o horizonte de previsão aumenta, o que é consistente com a propagação da incerteza inerente aos modelos numéricos de previsão atmosférica em horizontes mais longos.

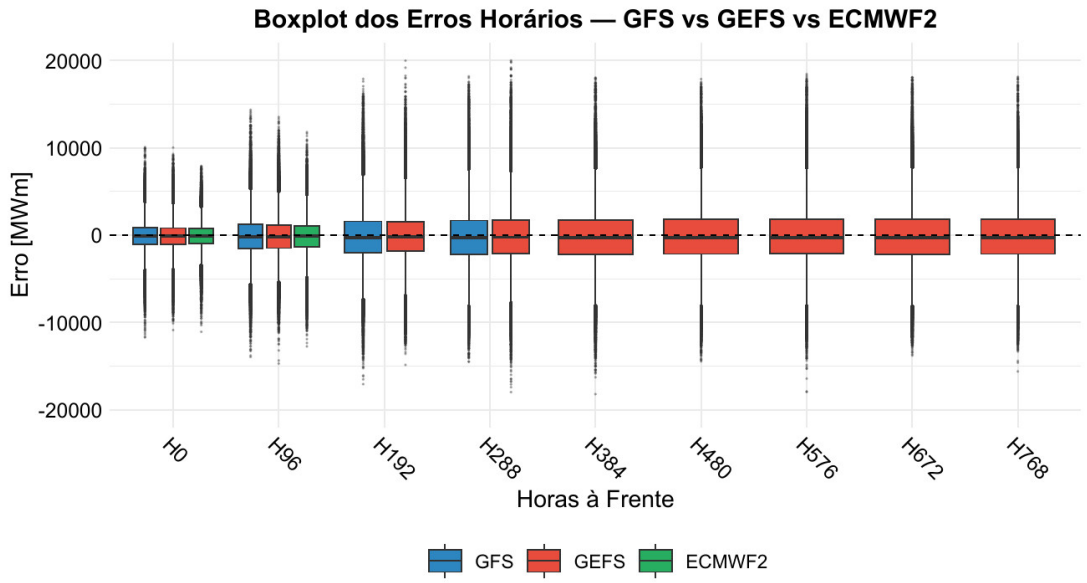
Um aspecto particularmente relevante diz respeito ao comportamento dos erros ao longo do ciclo diário. Verifica-se uma tendência de redução dos erros de previsão nos períodos mais próximos ao meio-dia, com mínimos locais observados, sobretudo para os modelos GFS e GEFS, em torno de 11 h a 14 h à frente. Esse comportamento está diretamente associado à natureza física da geração eólica no SIN, cuja maior parcela provém da região Nordeste do Brasil. Nesse período, o desenvolvimento pleno da camada limite atmosférica, somado à intensificação consistente dos ventos alísios e à atuação da brisa marítima, confere ao regime de ventos uma maior estabilidade e maior previsibilidade. Como consequência, a geração eólica torna-se menos suscetível a flutuações de curta escala, reduzindo o erro associado às previsões nessas horas.

Em contrapartida, nos períodos de transição (amanhecer e entardecer) e durante a madrugada, a variabilidade do vento aumenta de maneira significativa, em função de fenômenos de mesoescala como o jato noturno de baixos níveis, típico do Nordeste brasileiro, e da reorganização da camada limite atmosférica. Essa maior variabilidade explica os picos de erro observados, em especial após as 17 h à frente, onde os três modelos apresentam seus maiores valores percentuais de erro.

Complementarmente à análise dos erros médios horários, a Figura 103 apresenta a distribuição dos erros de previsão, em MWm, para algumas horas específicas à frente de previsão, para os três modelos avaliados. A representação em *boxplots* permite visualizar não apenas a tendência central do erro, mas também a sua dispersão, a presença de assimetrias e a ocorrência de

valores extremos, aspectos particularmente relevantes quando se avaliam previsões em horizontes mais distantes do instante de inicialização.

Figura 103 – Distribuição dos erros horários, em MWm, das previsões de geração eólica obtidas pelos modelos GFS, GEFS e ECMWF, agrupados por faixas do horizonte de previsão.



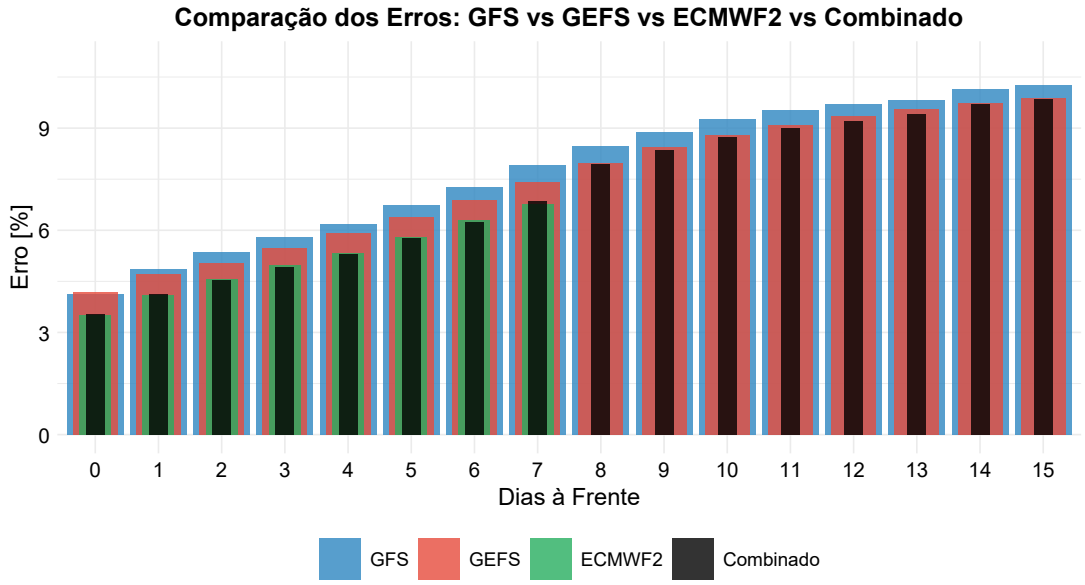
Fonte: Elaboração própria.

Observa-se que, para todos os modelos, a mediana dos erros permanece próxima de zero ao longo do horizonte de previsão, indicando ausência de viés sistemático relevante. No entanto, a amplitude interquartil e o alcance dos *whiskers* aumentam de forma progressiva à medida que o horizonte se afasta do instante de inicialização, evidenciando o aumento da dispersão dos erros nas horas mais distantes. Esse comportamento reflete a perda natural de previsibilidade dos modelos numéricos de previsão atmosférica em horizontes mais longos e reforça a necessidade de tratamento probabilístico das previsões em estudos de planejamento de médio prazo, em que a magnitude da incerteza passa a ser tão relevante quanto o próprio valor previsto.

A Figura 104 apresenta uma comparação dos erros médios diários percentuais das previsões de geração eólica obtidas a partir dos modelos meteorológicos GFS, GEFS e ECMWF, considerando os primeiros 15 dias do horizonte de previsão. Além desses modelos, é avaliada a previsão resultante da metodologia de combinação descrita na Seção 5.6 do Capítulo 5. Essa abordagem foi desenvolvida com o objetivo de explorar a complementaridade existente entre os diferentes modelos meteorológicos, buscando reduzir os erros de previsão observados quando cada modelo é utilizado individualmente.

No gráfico, o eixo das abscissas representa o horizonte de previsão, expresso em dias à frente, enquanto o eixo das ordenadas apresenta o erro percentual médio correspondente a cada horizonte. Essa análise permite avaliar não apenas o desempenho individual dos modelos meteorológicos, mas também os ganhos obtidos por meio da combinação das previsões ao longo do horizonte analisado.

Figura 104 – Comparação dos erros médios diários percentuais, normalizados pela potência instalada, das previsões de geração eólica obtidas pelos modelos GFS, GEFS, ECMWF e pela combinação *Comb*, ao longo dos primeiros 15 dias do horizonte de previsão.

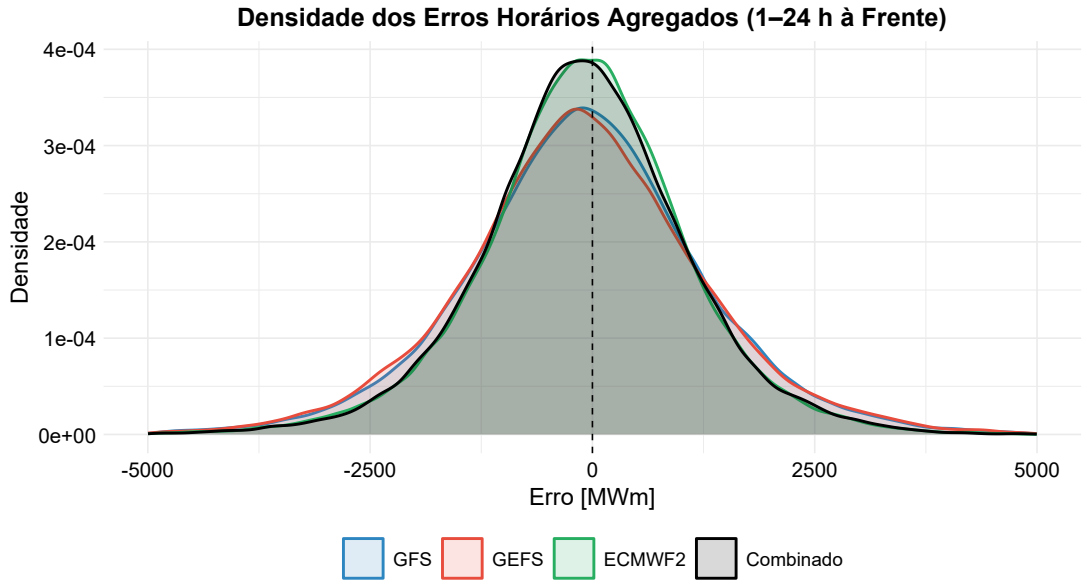


Fonte: Elaboração própria.

Observa-se um crescimento monotônico do erro à medida que o horizonte de previsão se distancia do instante de inicialização, comportamento esperado em decorrência da degradação progressiva da previsibilidade atmosférica em escalas mais longas. Nos primeiros dias, o modelo ECMWF apresenta desempenho consistentemente superior ao dos modelos GFS e GEFS, reafirmando a tendência já identificada na análise horária. Essa superioridade se mantém ao longo dos horizontes em que o modelo está disponível. Entre os modelos do NCEP, o GEFS tende a apresentar erros ligeiramente inferiores aos do GFS a partir do segundo dia de previsão, comportamento compatível com o caráter probabilístico do modelo de *ensemble*, cuja média tende a suavizar parte da incerteza associada aos membros individuais. Os resultados também mostram que a combinação explora de forma eficiente as informações complementares contidas nos diferentes modelos meteorológicos, produzindo previsões mais robustas ao longo de todo o horizonte analisado.

A série combinada apresenta desempenho competitivo ao longo de todo o horizonte considerado, posicionando-se tipicamente próxima ao melhor modelo disponível em cada faixa de previsão. Nos primeiros dias, a combinação alcança resultados comparáveis aos do ECMWF; nos horizontes subsequentes, mantém erros inferiores aos do GFS e comparáveis ou ligeiramente inferiores aos do GEFS. Esse resultado evidencia o benefício da combinação de previsões, prática consolidada na literatura, que tende a reduzir a variância do erro ao agregar informações provenientes de modelos com diferentes estruturas e parametrizações físicas. Esse efeito pode ser observado de forma direta na Figura 105, que apresenta a densidade dos erros horários agregados ao longo das 24 primeiras horas do horizonte de previsão.

Figura 105 – Densidade dos erros horários, em MWm, para os modelos GFS, GEFS, ECMWF e Combinado, considerando o conjunto agregado das 24 primeiras horas do horizonte de previsão.



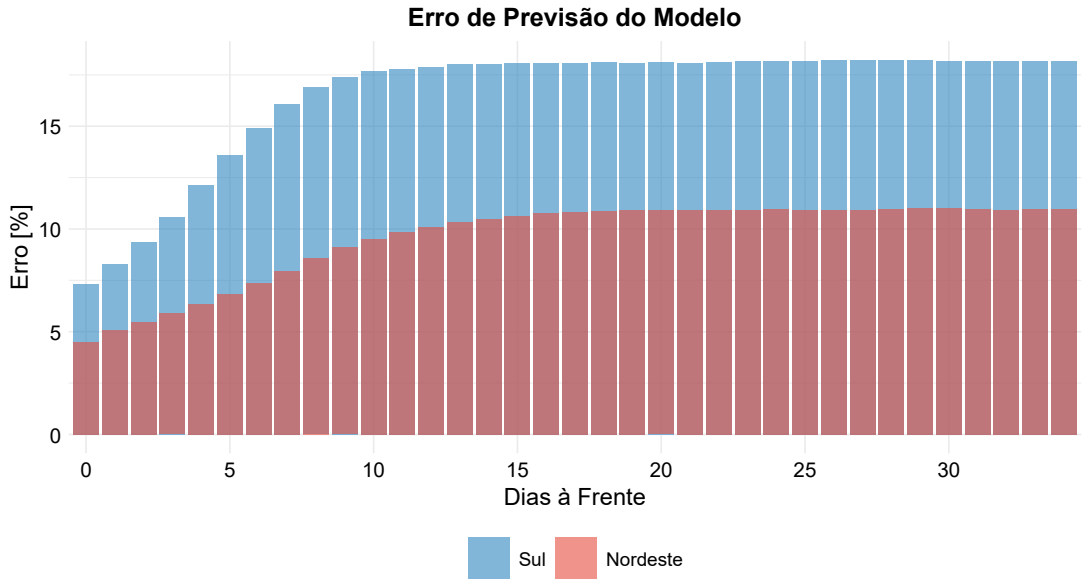
Fonte: Elaboração própria.

Verifica-se que a distribuição associada ao modelo combinado apresenta-se mais concentrada em torno do zero e com caudas menos pronunciadas do que as dos modelos individuais, indicando menor probabilidade de ocorrência de erros de grande magnitude. Esse comportamento decorre do fato de os erros dos modelos GFS, GEFS e ECMWF apresentarem correlação positiva, porém imperfeita: ao agregar previsões cujas estruturas físicas e parametrizações diferem, as componentes não correlacionadas do erro tendem a se cancelar parcialmente, resultando em uma série combinada de menor variância. Para a operação do SIN, esse comportamento é particularmente relevante, pois sugere que estratégias de combinação podem ser empregadas como alternativa robusta nos casos em que algum dos modelos individuais não esteja disponível ou apresente degradação pontual de desempenho, além de implicar menor probabilidade de ocorrência de erros extremos, que tipicamente impactam de modo mais significativo a programação da operação eletroenergética.

Por fim, a Figura 106 apresenta o erro de previsão do modelo proposto, desagregado por subsistema, considerando os subsistemas Sul e Nordeste ao longo do horizonte completo de 35 dias de previsão. Essa análise é particularmente relevante para a operação do SIN, dado que esses dois subsistemas concentram a quase totalidade da capacidade instalada de geração eólica do país e apresentam regimes de vento e características climatológicas significativamente distintas.

Os resultados evidenciam diferenças expressivas no desempenho do modelo entre os dois subsistemas. O subsistema Nordeste apresenta erros sistematicamente inferiores aos do subsistema Sul ao longo de todo o horizonte de previsão, com diferença que se amplia à medida que se avança no tempo de antecedência. Esse comportamento é coerente com a maior

Figura 106 – Erro de previsão do modelo, em termos percentuais e normalizado pela potência instalada, para os subsistemas Sul e Nordeste ao longo do horizonte de previsão de 35 dias.



Fonte: Elaboração própria.

previsibilidade do regime de ventos na região Nordeste, fortemente associada à atuação dos ventos alísios e à influência da circulação de larga escala sobre o Atlântico tropical, que conferem ao recurso eólico nessa região uma maior estabilidade sazonal e de curto prazo.

No subsistema Sul, por outro lado, o regime de ventos é fortemente modulado pela passagem de sistemas sinóticos transientes, como frentes frias, ciclones extratropicais e linhas de instabilidade, cuja previsibilidade decai mais rapidamente com o aumento do horizonte de previsão. Como consequência, o erro associado às previsões para o Sul cresce de forma mais acentuada nos primeiros dias e atinge patamares mais elevados que os observados no Nordeste em todo o horizonte avaliado.

Outro aspecto relevante é a tendência de estabilização do erro a partir de aproximadamente duas semanas à frente, situando-se entre o décimo quinto e o vigésimo dia de previsão nos dois subsistemas analisados. Esse comportamento é característico de previsões em horizontes mais longos, nos quais a perda gradual de previsibilidade determinística faz com que o erro convirja para um patamar associado à variabilidade climatológica do recurso eólico. Do ponto de vista operacional, esse resultado indica que, para horizontes superiores a aproximadamente duas semanas, a utilização das previsões deve privilegiar abordagens probabilísticas e sazonalizadas. Nessas condições, tais abordagens podem fornecer informações mais representativas para o planejamento eletroenergético de curto prazo do que previsões determinísticas voltadas à programação da operação.

A fim de complementar a análise gráfica apresentada anteriormente, são compiladas, nas Tabelas 10, 11 e 12, as principais métricas de erro do modelo combinado, em base diária, para os subsistemas Sul, Nordeste e para o agregado do SIN, respectivamente. As métricas reportadas

são o erro médio normalizado (nME), o erro absoluto médio normalizado (nMAE), a raiz do erro quadrático médio normalizada (nRMSE) e o erro percentual absoluto médio normalizado, este apresentado em duas variantes ($\text{nMAPE}_{\text{fatcap}}$ e $\text{nMAPE}_{\text{Pinst}}$), ao longo do horizonte de previsão de até 21 dias à frente (D0 a D21). Todas essas métricas são normalizadas pela potência instalada, em cada instante, multiplicada pelo fator de capacidade médio do período, à exceção do $\text{nMAPE}_{\text{Pinst}}$, normalizado apenas pela potência instalada. As três primeiras (nME, nMAE e nRMSE) são adimensionais, enquanto as duas variantes do nMAPE são expressas em termos percentuais.

A análise comparativa entre os três conjuntos evidencia, de forma quantitativa, as diferenças de previsibilidade já observadas nas figuras anteriores. O subsistema Sul (Tabela 10) apresenta os maiores erros entre os casos avaliados, refletindo a maior complexidade de previsão associada ao seu regime de ventos. O nMAE cresce de aproximadamente 0,17 em D0 para cerca de 0,53 em D21, enquanto o $\text{nMAPE}_{\text{Pinst}}$ varia entre 5,8% e 18,2% no mesmo horizonte. Observa-se ainda uma tendência de aumento da superestimação da geração à medida que o horizonte de previsão se amplia, com o nME atingindo valores próximos de $-0{,}07$ a partir de D14. Esse comportamento é consistente com a forte influência de sistemas sinóticos transientes sobre o regime eólico da região, o que reduz a previsibilidade em horizontes mais longos.

Tabela 10 – Métricas de erro do modelo combinado para o subsistema Sul, por horizonte de previsão (D0 a D21), normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, exceto o $\text{nMAPE}_{\text{Pinst}}$, normalizado pela potência instalada.

Métrica	**D0**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
nME	0,00	0,00	0,00	−0,01	−0,01	−0,02	−0,03	−0,04	−0,05	−0,05	−0,06
nMAE	0,17	0,20	0,24	0,27	0,32	0,37	0,41	0,46	0,49	0,51	0,51
nRMSE	0,22	0,27	0,31	0,35	0,41	0,47	0,52	0,57	0,60	0,62	0,63
$\text{nMAPE}_{\text{fatcap}}$	16,85	20,24	23,59	27,23	31,88	36,96	41,30	46,00	49,17	50,75	51,35
$\text{nMAPE}_{\text{Pinst}}$	5,81	6,98	8,13	9,38	10,99	12,74	14,23	15,85	16,94	17,49	17,70
Métrica	**D11**	**D12**	**D13**	**D14**	**D15**	**D16**	**D17**	**D18**	**D19**	**D20**	**D21**
nME	−0,05	−0,06	−0,06	−0,07	−0,07	−0,07	−0,07	−0,07	−0,07	−0,07	−0,07
nMAE	0,52	0,52	0,52	0,52	0,53	0,53	0,53	0,53	0,53	0,53	0,53
nRMSE	0,63	0,63	0,64	0,64	0,64	0,64	0,64	0,64	0,64	0,64	0,64
$\text{nMAPE}_{\text{fatcap}}$	51,68	52,05	52,33	52,43	52,67	52,67	52,68	52,87	52,69	52,80	52,71
$\text{nMAPE}_{\text{Pinst}}$	17,81	17,93	18,03	18,07	18,15	18,15	18,15	18,22	18,16	18,19	18,16

Fonte: Elaboração própria.

Em contraste, o subsistema Nordeste (Tabela 11) apresenta, de forma geral, o melhor desempenho do modelo combinado. O nMAE evolui de aproximadamente 0,09 em D0 para cerca de 0,24 em D21, enquanto o $\text{nMAPE}_{\text{Pinst}}$ varia entre 3,8% e 10,9%. Além disso, o nME permanece próximo de zero ao longo de todo o horizonte de previsão, com valores ligeiramente negativos, indicando apenas uma leve tendência de superestimação e ausência de viés sistemático relevante.

A comparação entre os dois subsistemas evidencia a maior previsibilidade do recurso eólico no Nordeste em relação ao Sul, resultado associado às características climáticas mais persistentes da circulação atmosférica na região nordestina e à menor influência de sistemas meteorológicos de rápida evolução.

Tabela 11 – Métricas de erro do modelo combinado para o subsistema Nordeste, por horizonte de previsão (D0 a D21), normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, exceto o $\text{nMAPE}_{\text{Pinst}}$, normalizado pela potência instalada.

Métrica	**D0**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
nME	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01
nMAE	0,09	0,10	0,11	0,12	0,13	0,14	0,15	0,16	0,19	0,20	0,21
nRMSE	0,11	0,13	0,14	0,15	0,16	0,18	0,19	0,21	0,24	0,26	0,27
$\text{nMAPE}_{\text{fatcap}}$	8,53	9,96	10,95	11,89	12,72	13,86	14,93	16,38	18,98	20,07	20,98
$\text{nMAPE}_{\text{Pinst}}$	3,80	4,43	4,88	5,29	5,67	6,17	6,65	7,30	8,46	8,94	9,35
Métrica	**D11**	**D12**	**D13**	**D14**	**D15**	**D16**	**D17**	**D18**	**D19**	**D20**	**D21**
nME	−0,01	−0,02	−0,02	−0,02	−0,02	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01
nMAE	0,22	0,22	0,23	0,23	0,24	0,24	0,24	0,24	0,24	0,24	0,24
nRMSE	0,28	0,28	0,29	0,29	0,30	0,31	0,31	0,31	0,31	0,31	0,31
$\text{nMAPE}_{\text{fatcap}}$	21,65	22,10	22,59	23,22	23,64	24,02	24,14	24,25	24,32	24,32	24,39
$\text{nMAPE}_{\text{Pinst}}$	9,64	9,84	10,06	10,34	10,53	10,70	10,75	10,80	10,83	10,83	10,86

Fonte: Elaboração própria.

Para o agregado do SIN (Tabela 12), as métricas apresentam-se em patamares próximos aos observados no subsistema Nordeste, refletindo o fato de que a maior parcela da geração eólica do SIN provém dessa região. O nMAE varia de aproximadamente 0,08 em D0 a 0,23 em D21, enquanto o $\text{nMAPE}_{\text{Pinst}}$ oscila entre 3,5% e 10,0% no mesmo intervalo. O nME, semelhante ao do Nordeste (entre $-0{,}01$ e $-0{,}02$), indica uma leve tendência de superestimação do agregado, com pequena contribuição do subsistema Sul.

Vale destacar a diferença de magnitude entre as duas variantes do nMAPE: enquanto a versão normalizada apenas pela potência instalada ($\text{nMAPE}_{\text{Pinst}}$) traduz o erro como uma fração da capacidade nominal, a versão normalizada pela potência instalada multiplicada pelo fator de capacidade ($\text{nMAPE}_{\text{fatcap}}$) traduz o erro como fração da geração média esperada, sendo, portanto, sistematicamente maior. A diferença é particularmente expressiva no subsistema Sul, em virtude do fator de capacidade mais baixo ($f_c \approx 0{,}35$), em contraste com o subsistema Nordeste ($f_c \approx 0{,}45$). Esse comportamento reforça a importância de explicitar o critério de normalização ao se reportarem métricas de erro de previsão de geração eólica, uma vez que diferentes convenções podem conduzir a percepções distintas a respeito do mesmo desempenho.

Por fim, observa-se em todos os subsistemas a tendência de saturação das métricas a partir, aproximadamente, do décimo quinto dia de previsão, com os valores se aproximando de um patamar associado à variabilidade climatológica do recurso eólico. Esse comportamento é consistente com o esperado para previsões em horizontes de médio prazo e reforça a recomenda-

Tabela 12 – Métricas de erro do modelo combinado para o agregado do SIN, por horizonte de previsão (D0 a D21), normalizadas pela capacidade instalada multiplicada pelo fator de capacidade médio, exceto o $\text{nMAPE}_{\text{Pinst}}$, normalizado pela potência instalada.

Métrica	**D0**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
nME	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01	−0,01	−0,02	−0,02	−0,02
nMAE	0,08	0,09	0,10	0,11	0,12	0,13	0,14	0,16	0,18	0,19	0,20
nRMSE	0,10	0,12	0,13	0,14	0,15	0,17	0,18	0,20	0,23	0,24	0,25
$\text{nMAPE}_{\text{fatcap}}$	8,07	9,39	10,33	11,24	12,09	13,19	14,22	15,59	18,00	18,97	19,80
$\text{nMAPE}_{\text{Pinst}}$	3,53	4,11	4,52	4,91	5,29	5,77	6,22	6,82	7,87	8,29	8,66
Métrica	**D11**	**D12**	**D13**	**D14**	**D15**	**D16**	**D17**	**D18**	**D19**	**D20**	**D21**
nME	−0,02	−0,02	−0,02	−0,02	−0,02	−0,01	−0,02	−0,01	−0,01	−0,01	−0,01
nMAE	0,20	0,21	0,21	0,22	0,22	0,23	0,23	0,23	0,23	0,23	0,23
nRMSE	0,26	0,26	0,27	0,28	0,28	0,29	0,29	0,29	0,29	0,29	0,29
$\text{nMAPE}_{\text{fatcap}}$	20,40	20,84	21,31	21,95	22,31	22,63	22,75	22,84	22,91	22,90	22,96
$\text{nMAPE}_{\text{Pinst}}$	8,92	9,11	9,32	9,60	9,76	9,89	9,95	9,99	10,02	10,01	10,04

Fonte: Elaboração própria.

ção, já apontada nas análises anteriores, de que o uso das previsões em horizontes superiores a aproximadamente duas semanas seja conduzido sob uma abordagem probabilística, mais adequada ao planejamento eletroenergético do que à programação determinística da operação.

6.4.3 Previsões semanais para o horizonte até 5 semanas à frente

A análise comparativa do desempenho do modelo proposto em relação ao modelo de referência (REN nº 843, Resolução Normativa da ANEEL nº 843 [29]), adotado como *benchmark*, é apresentada a seguir, inicialmente sob a perspectiva da distribuição dos erros percentuais por meio de *boxplots* para os subsistemas Sul e Nordeste, e, em seguida, por meio dos erros médios percentuais (nMAPE, Equação (B.5) do Apêndice B) por patamar de carga, organizados por passos semanais à frente, considerando os subsistemas Sul, Nordeste e o SIN.

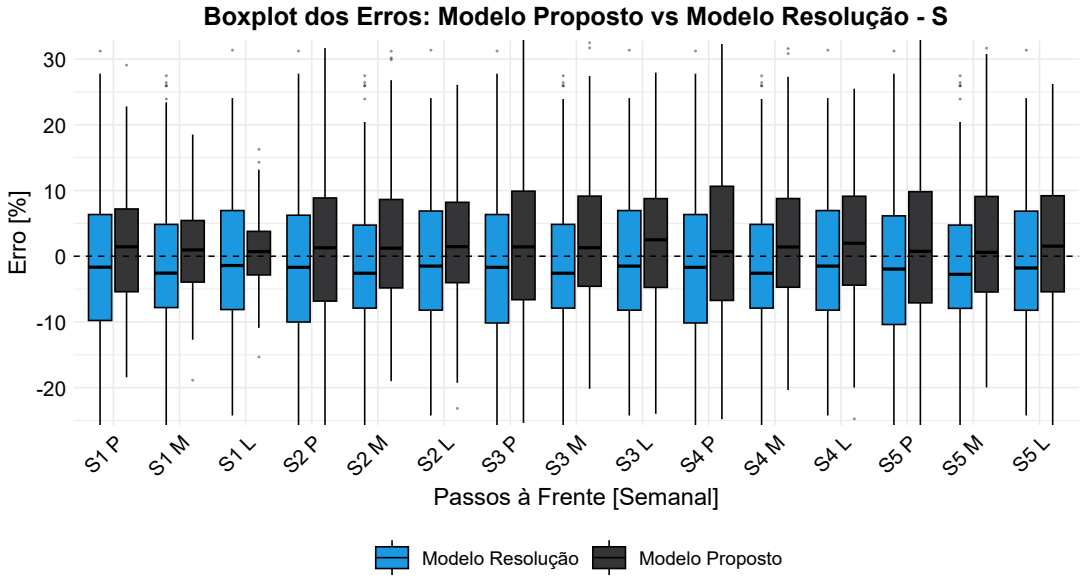
A Figura 107 apresenta a distribuição dos erros percentuais para o subsistema Sul, comparando o modelo proposto e o modelo de resolução, organizados pelos cinco passos semanais à frente (S1 a S5) e pelos três patamares de carga (Pesada, Média e Leve).

Observa-se que o modelo proposto apresenta, de forma sistemática, intervalos interquartis mais estreitos e medianas mais próximas de zero do que as obtidas pelo modelo de resolução, evidenciando desempenho consistentemente superior ao do *benchmark* em todos os passos e patamares avaliados. Verifica-se ainda que a dispersão dos erros tende a aumentar à medida que se avança no horizonte de previsão, comportamento esperado em decorrência da degradação progressiva da previsibilidade em horizontes mais longos.

A Figura 108 apresenta a mesma análise para o subsistema Nordeste.

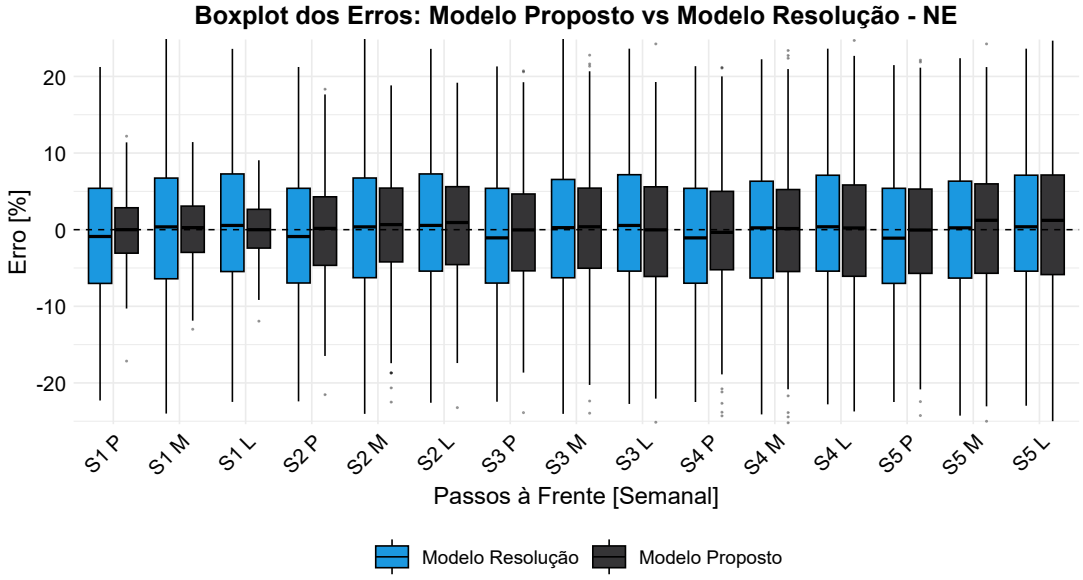
De maneira análoga ao observado no subsistema Sul, o modelo proposto apresenta

Figura 107 – Distribuição dos erros percentuais, normalizados pela potência instalada, por passo semanal e patamar de carga para o subsistema Sul, comparando o modelo proposto e o modelo de resolução.



Fonte: Elaboração própria.

Figura 108 – Distribuição dos erros percentuais, normalizados pela potência instalada, por passo semanal e patamar de carga para o subsistema Nordeste, comparando o modelo proposto e o modelo de resolução.

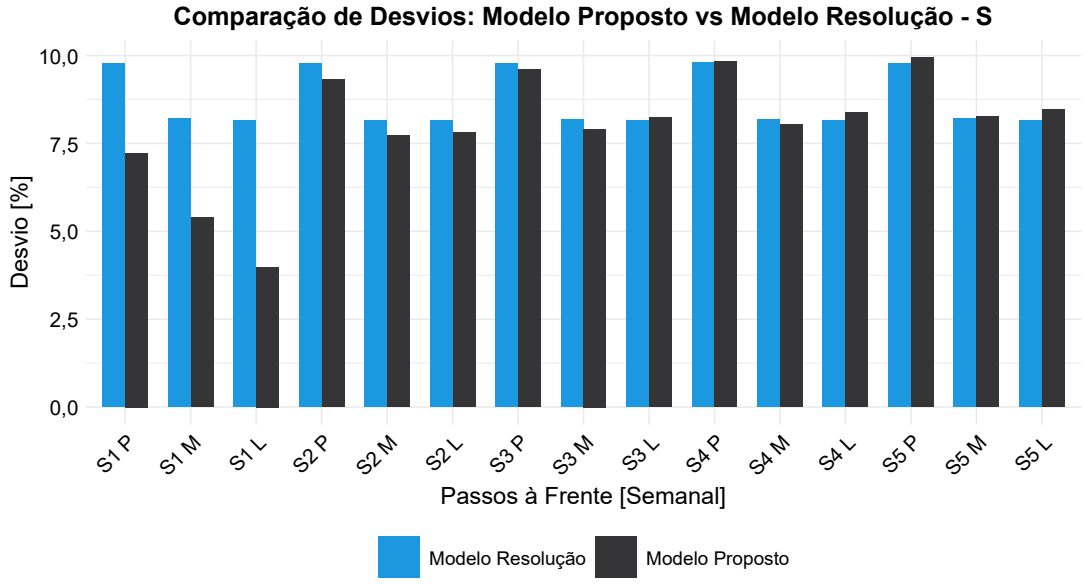


Fonte: Elaboração própria.

desempenho superior ao do modelo de resolução em todos os passos e patamares avaliados, com intervalos interquartis mais estreitos e medianas mais próximas de zero. Comparativamente, o subsistema Nordeste exibe erros de menor magnitude e menor dispersão do que os verificados no subsistema Sul, comportamento coerente com a maior estabilidade do regime de ventos na região Nordeste, conforme caracterizado na Seção 2.1 do Capítulo 2. Já o subsistema Sul, sujeito à atuação mais frequente de sistemas sinóticos transientes, apresenta naturalmente maior variabilidade do recurso eólico e, consequentemente, maior dispersão dos erros de previsão.

Apresentada a análise da distribuição dos erros, passa-se à comparação dos erros médios percentuais por patamar de carga, segregados por passo semanal à frente, para os subsistemas Sul, Nordeste e SIN. A Figura 109 apresenta os resultados para o subsistema Sul.

Figura 109 – Comparação dos erros médios percentuais (nMAPE, normalizado pela potência instalada) por passo semanal e patamar de carga para o subsistema Sul, entre o modelo proposto e o modelo de resolução.



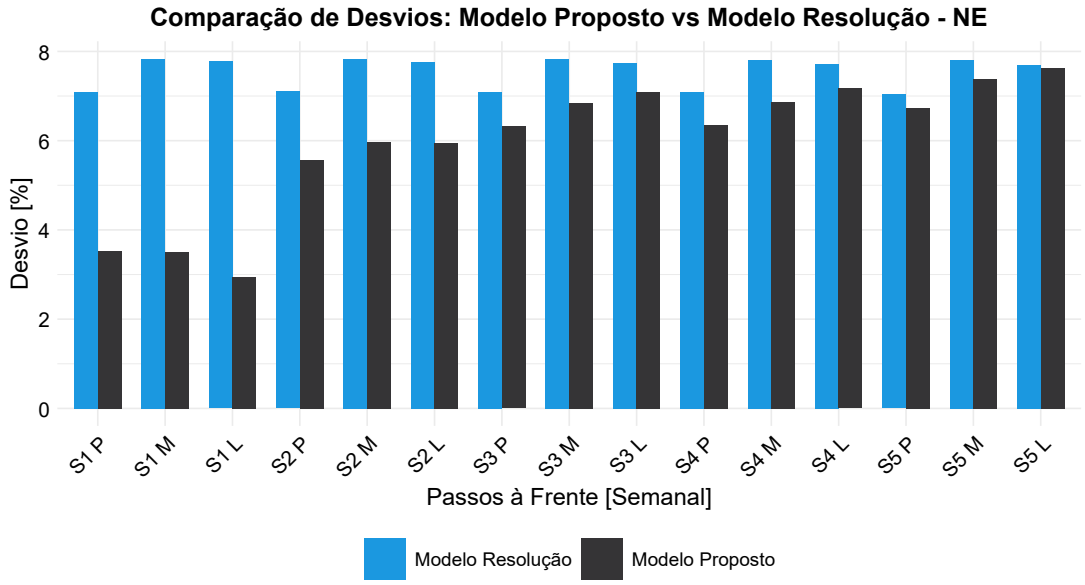
Fonte: Elaboração própria.

Verifica-se que, no primeiro passo semanal, o modelo proposto apresenta erros médios sensivelmente inferiores aos do modelo de resolução em todos os patamares de carga, com destaque para o patamar leve, onde o ganho relativo é mais pronunciado. Nos passos subsequentes, a diferença entre os dois modelos tende a diminuir, refletindo a perda natural de previsibilidade em horizontes mais distantes. Ainda assim, o modelo proposto mantém-se em geral em patamares próximos ou inferiores aos do modelo de resolução, indicando ganho de desempenho consistente ao longo do horizonte avaliado.

A Figura 110 apresenta a mesma análise para o subsistema Nordeste.

No subsistema Nordeste, o ganho de desempenho do modelo proposto em relação ao modelo de resolução é particularmente expressivo no primeiro passo semanal, com reduções relevantes em todos os patamares de carga e, novamente, com maior destaque para o patamar leve.

Figura 110 – Comparação dos erros médios percentuais (nMAPE, normalizado pela potência instalada) por passo semanal e patamar de carga para o subsistema Nordeste, entre o modelo proposto e o modelo de resolução.



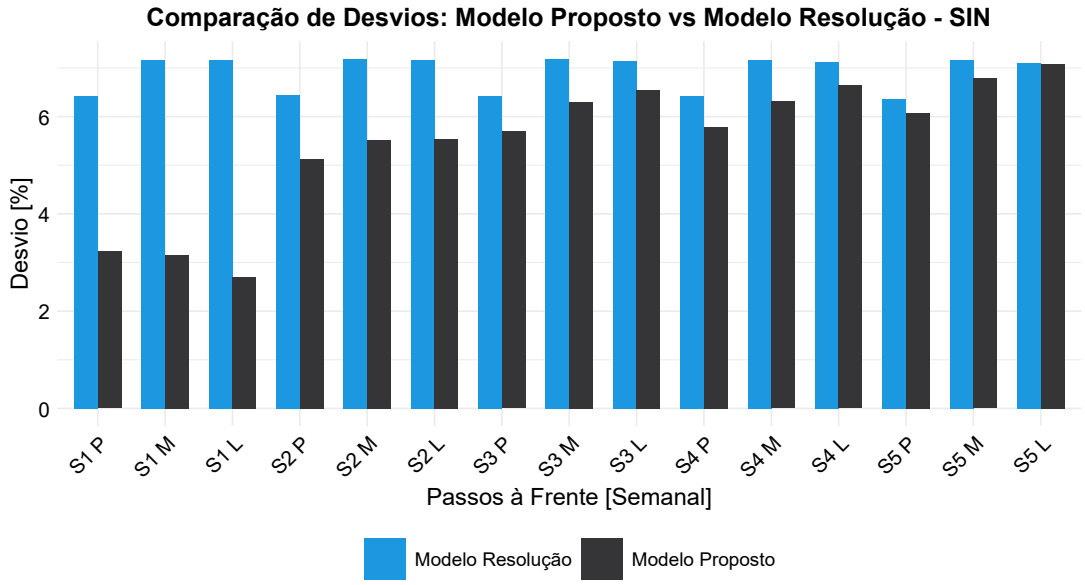
Fonte: Elaboração própria.

Nos passos seguintes, a vantagem relativa do modelo proposto se reduz progressivamente, embora ele continue exibindo desempenho compatível com o do modelo de resolução, frequentemente em patamares inferiores. Esse comportamento é coerente com a maior previsibilidade do regime de ventos no Nordeste, que favorece o desempenho do modelo no curto prazo, mas que, em horizontes mais longos, tende a se aproximar do limite associado à variabilidade climatológica do recurso eólico.

Por fim, a Figura 111 apresenta a comparação dos erros médios percentuais para o agregado do SIN.

Para o agregado do SIN, o ganho proporcionado pelo modelo proposto é claramente evidenciado no primeiro passo semanal, no qual os erros médios situam-se em torno de 3,2%, 3,2% e 2,7% para os patamares pesado, médio e leve, respectivamente, frente a valores próximos de 6,4%, 7,2% e 7,2% obtidos pelo modelo de resolução. Tal resultado representa, em termos práticos, uma redução de aproximadamente metade do erro observado no *benchmark* no horizonte de uma semana à frente. À medida que o horizonte avança, a diferença entre os dois modelos diminui, com o modelo proposto apresentando, no quinto passo semanal, erros em torno de 6,1%, 6,8% e 7,1%, contra valores próximos de 6,4%, 7,2% e 7,1% do modelo de resolução. Esse comportamento confirma, no agregado do SIN, a tendência observada nos subsistemas individualmente: ganhos expressivos no curto prazo e convergência progressiva dos desempenhos em horizontes mais longos, em virtude da perda natural de previsibilidade.

Figura 111 – Comparação dos erros médios percentuais (nMAPE, normalizado pela potência instalada) por passo semanal e patamar de carga para o SIN, entre o modelo proposto e o modelo de resolução.



Fonte: Elaboração própria.

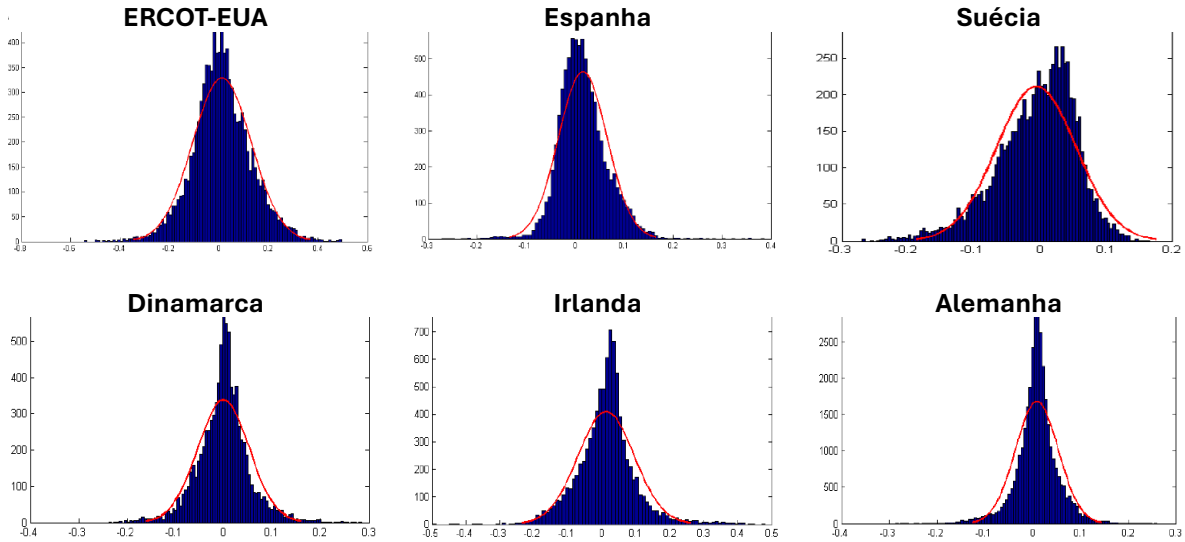
6.4.4 Discussões complementares

Como não há *benchmarking* na literatura para aplicações em escala sistêmica, os resultados apresentados não foram comparados com nenhum outro trabalho existente além do modelo anteriormente vigente da REN nº 843. Devido à baixa autocorrelação nas séries de geração eólica para horizontes de previsão do dia seguinte a vários dias à frente, modelos de previsão baseados apenas no histórico da série não são adequados, sendo este um dos principais benefícios alcançados com o modelo proposto. Essa conclusão foi destacada nas análises comparativas realizadas no trabalho [42], que decorre desta pesquisa.

No entanto, existem estudos que abordam o comportamento dos erros de previsão de geração eólica em nível sistêmico, como o trabalho [139], que analisa o desempenho da previsão eólica em diversos países. Comparar os resultados obtidos com o modelo proposto com os de outros países não seria totalmente apropriado, uma vez que a dificuldade de previsão pode variar significativamente entre regiões, refletindo-se na dimensão dos erros observados. Apesar de não ser uma comparação direta, a Figura 112 apresenta os erros de geração eólica, normalizados pela capacidade instalada, de alguns países, proporcionando uma compreensão qualitativa dos resultados. A figura também mostra o comportamento da distribuição normal, calculada com base na média e no desvio padrão dos erros de previsão.

Conforme se pode notar, em várias situações, a distribuição dos erros difere significativamente de uma distribuição normal com média zero, como seria esperado, exibindo, em alguns casos, caudas pesadas e/ou vieses. Para efeitos de comparação com o modelo proposto, a Figura 113 apresenta os histogramas dos erros horários normalizados pela potência instalada,

Figura 112 – Distribuição dos erros de previsão eólica de alguns países.



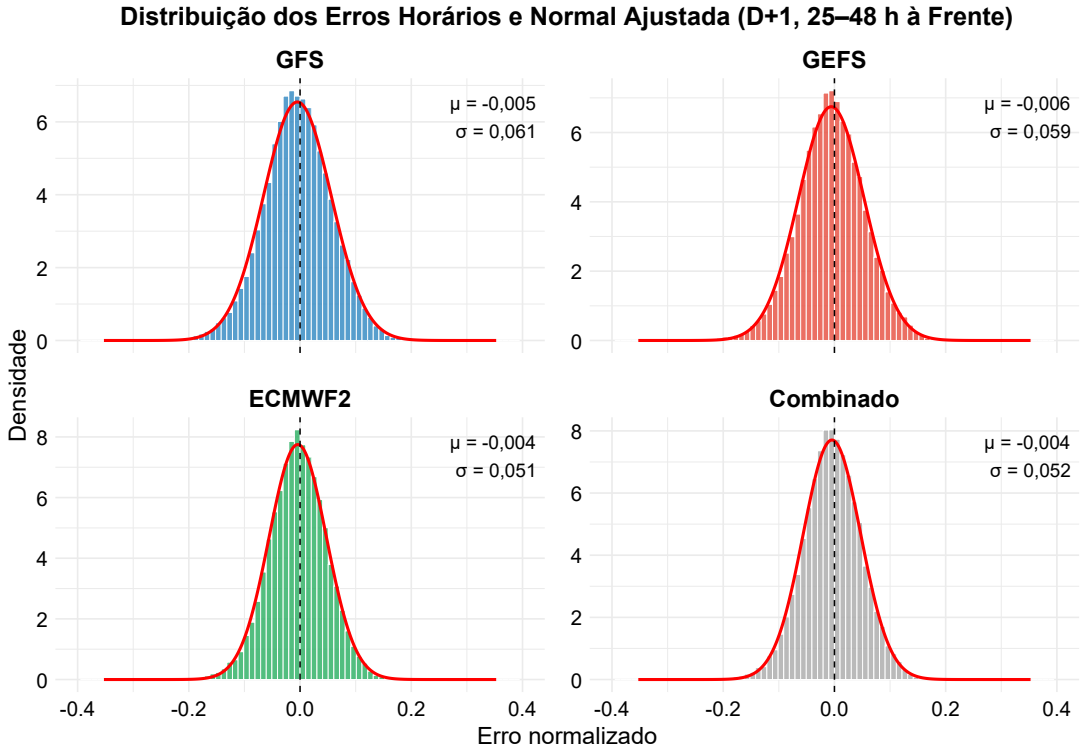
Fonte: Adaptado de [139].

para os modelos GFS, GEFS, ECMWF e Combinado, considerando o horizonte D+1 (de 25 a 48 horas à frente). A cada painel é sobreposta a curva da distribuição Normal ajustada aos respectivos dados, parametrizada pela média (μ) e pelo desvio-padrão (σ) amostrais, indicados na própria figura, em formato análogo ao adotado pela referência utilizada para os demais países.

Observa-se que, para os quatro modelos, as distribuições dos erros apresentam-se aproximadamente simétricas e centradas próximas ao zero, com médias amostrais de pequena magnitude, indicando ausência de viés sistemático relevante no horizonte avaliado. A curva Normal ajustada acompanha de forma razoável a parte central das distribuições empíricas, mas evidencia caudas mais pesadas do que as previstas pelo modelo gaussiano, comportamento típico de erros de previsão associados a fenômenos meteorológicos, sujeitos à ocorrência ocasional de eventos extremos, comportamento qualitativamente compatível com o observado nas distribuições reportadas por [139] para diferentes países.

A comparação entre os painéis permite ainda identificar diferenças relevantes na dispersão dos erros entre os modelos. O modelo Combinado apresenta um dos menores desvio-padrão, com histograma visivelmente mais concentrado em torno do zero, ratificando o efeito de redução de variância proporcionado pela combinação dos modelos individuais, conforme já discutido anteriormente. O modelo ECMWF, por sua vez, exibe a menor dispersão entre os modelos individuais, em conformidade com o desempenho superior verificado nas demais análises, enquanto os modelos GFS e GEFS apresentam dispersões semelhantes entre si, ligeiramente mais elevadas.

Figura 113 – Distribuição dos erros horários normalizados e Normal ajustada para os modelos GFS, GEFS, ECMWF e Combinado, considerando o horizonte D+1 (25 a 48 horas à frente).



Fonte: Elaboração própria.

Ao analisar os resultados alcançados com o modelo proposto para previsões em nível do SIN, pode-se afirmar que são competitivos em comparação com os desempenhos apresentados para os países mencionados. É importante ressaltar, mais uma vez, que essa comparação é apenas qualitativa e não deve ser usada para conclusões definitivas. Por exemplo, ao comparar as previsões para o Sul (Figura 109) e para o Nordeste (Figura 110), elaboradas com o mesmo modelo, os desempenhos são completamente distintos, refletindo as diferenças entre os regimes de vento de cada subsistema, conforme caracterizado na Seção 2.1 do Capítulo 2.

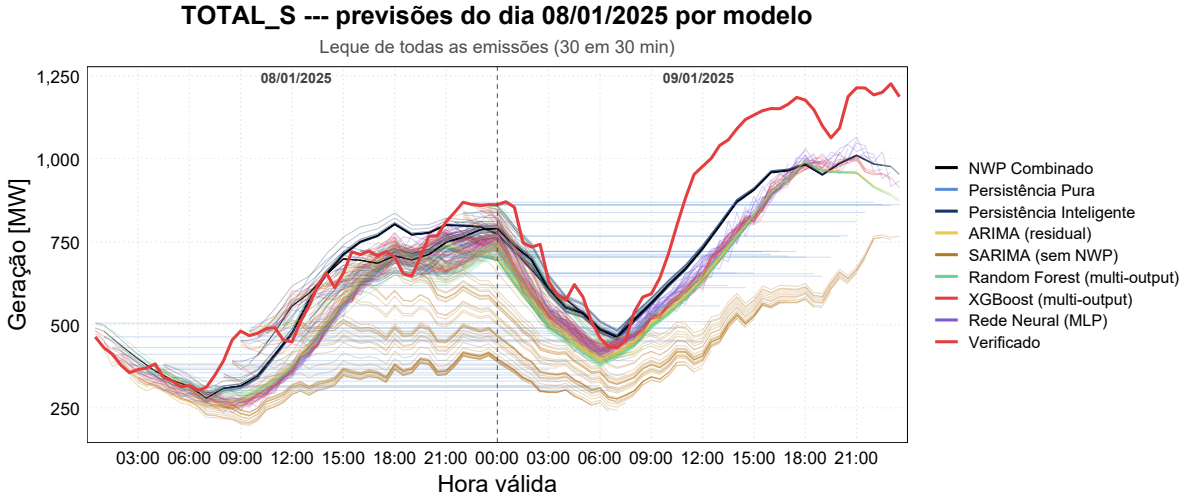
6.5 Resultados dos modelos de previsão em tempo real por subsistema

A avaliação dos modelos de tempo real, cujo *ensemble* e estratégias de combinação foram descritos na Seção 5.10 do Capítulo 5, é apresentada nesta seção em nível de subsistema. Como referência para as análises, utiliza-se inicialmente o dia representativo de 08/01/2025, empregado tanto nos resultados agregados aqui apresentados quanto nas análises desagregadas do Apêndice C, permitindo relacionar os comportamentos observados em ambas as escalas. Em seguida, são apresentados os resultados consolidados por horizonte de previsão e as distribuições dos erros para horizontes selecionados nos subsistemas Sul e Nordeste.

6.5.1 Visão conjunta dos modelos

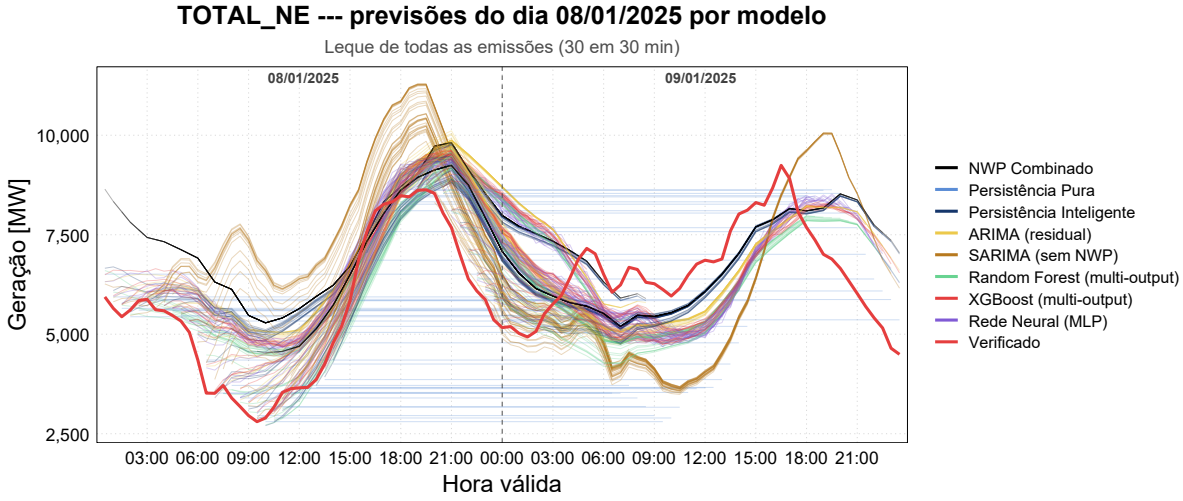
As Figuras 114 e 115 apresentam o leque das emissões dos modelos considerados para o dia representativo. Diferentemente do caso por agrupamentos individuais, em que as flutuações locais são visíveis com maior amplitude, a agregação por subsistema suaviza parte da variabilidade em virtude do efeito de suavização espacial: as flutuações não correlacionadas entre os agrupamentos tendem a se cancelar na soma, restando o sinal comum ao subsistema como um todo. Como consequência, o verificado apresenta um perfil diário mais regular, e o espalhamento entre os leques dos modelos com NWP fica mais estreito.

Figura 114 – Comportamento conjunto de múltiplos modelos no subsistema Sul para o dia 08/01/2025.



Fonte: Elaboração própria.

Figura 115 – Comportamento conjunto de múltiplos modelos no subsistema Nordeste para o dia 08/01/2025.



Fonte: Elaboração própria.

A partir das Figuras 114 e 115, algumas observações podem ser destacadas sobre o comportamento dos modelos ao longo do horizonte. Os modelos baseados em previsão numérica de tempo permanecem coesos em torno do NWP Combinado, que fornece um sinal

de referência mais estável do que qualquer fonte individual. A construção dessa combinação, a partir das previsões do ECMWF, GFS e GEFS, é detalhada no Apêndice C, onde se observa seu posicionamento central em relação às fontes individuais.

Os modelos de tempo real agregam valor ao corrigir a previsão de referência nas primeiras horas, ancorando-se na geração observada mais recente, e convergem gradualmente para o NWP Combinado à medida que o horizonte de previsão aumenta. Entre as abordagens avaliadas, as combinações dinâmicas por inverso do erro (1/MAE) e por filtro de Kalman destacam-se por apresentarem os menores erros e por reduzirem o viés observado no NWP Combinado.

Já os modelos utilizados como comparação apresentam comportamentos distintos ao longo do horizonte de previsão. O modelo de persistência pura, por exemplo, é utilizado principalmente como referência visual e de *benchmark*, uma vez que seu desempenho tende a se deteriorar rapidamente à medida que o horizonte aumenta, conforme esperado. O modelo SARIMA, construído exclusivamente a partir da série histórica de geração, mostra-se mais competitivo em horizontes muito curtos ou em regiões com menor variabilidade e ciclo diário de geração mais regular. Entretanto, sua capacidade preditiva se deteriora em situações de transição de regime, justamente nos períodos em que a previsão se torna mais desafiadora e a informação meteorológica desempenha papel fundamental. Análises mais detalhadas sobre o comportamento individual dos modelos e suas características específicas são apresentadas no Apêndice C.

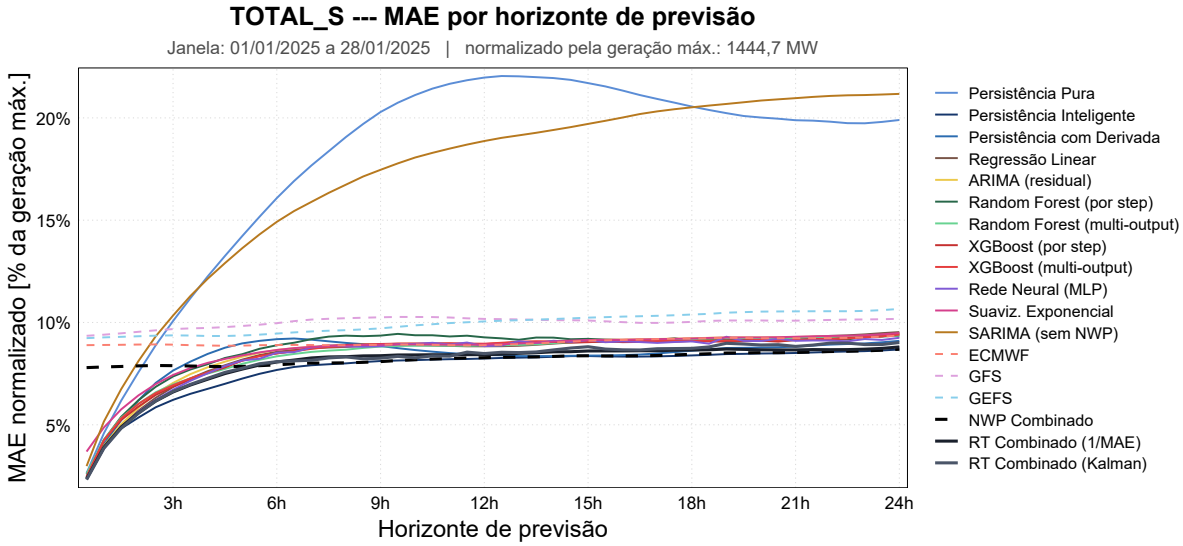
6.5.2 Erro médio absoluto por horizonte

As Figuras 116 e 117 consolidam o MAE normalizado em função do horizonte de previsão para os dois subsistemas, sobre toda a janela de avaliação. Em relação aos resultados por agrupamentos, o patamar geral de erro reduz-se em ambos os casos, o que é consistente com a suavização espacial: a agregação reduz a variabilidade proporcionalmente mais que reduz o sinal médio, favorecendo os indicadores relativos de desempenho. As curvas mantêm as posições relativas entre os modelos observadas no caso por agrupamentos, o que reforça a robustez das conclusões: as combinações de tempo real (RT Combinado $1/\text{MAE}$ e Kalman) apresentam os menores erros na maior parte do horizonte, seguidas pelos modelos NWP e suas combinações; a Persistência Pura mantém a curva característica em forma de corcova; e o SARIMA aparece como um dos piores, em contraste com os modelos que utilizam a informação meteorológica.

No subsistema Sul (Figura 116), a Persistência Pura apresenta a curva característica em forma de corcova, ultrapassando 20% da geração máxima do período em torno da metade do horizonte, enquanto o SARIMA sobe monotonicamente até atingir aproximadamente esse mesmo patamar em h+24. As previsões numéricas individuais (linhas tracejadas) mantêm-se em uma faixa entre 9 e 11% ao longo do horizonte, ao passo que o NWP Combinado (linha tracejada preta) permanece próximo de 8%. As combinações de tempo real e o conjunto dos modelos baseados em NWP evoluem em uma faixa entre aproximadamente 4 e 8% ao longo do horizonte, partindo de valores próximos a 3% nos primeiros passos, reflexo da forte autocorrelação de

curtíssimo prazo, e chegando ligeiramente abaixo do NWP Combinado nos horizontes mais longos.

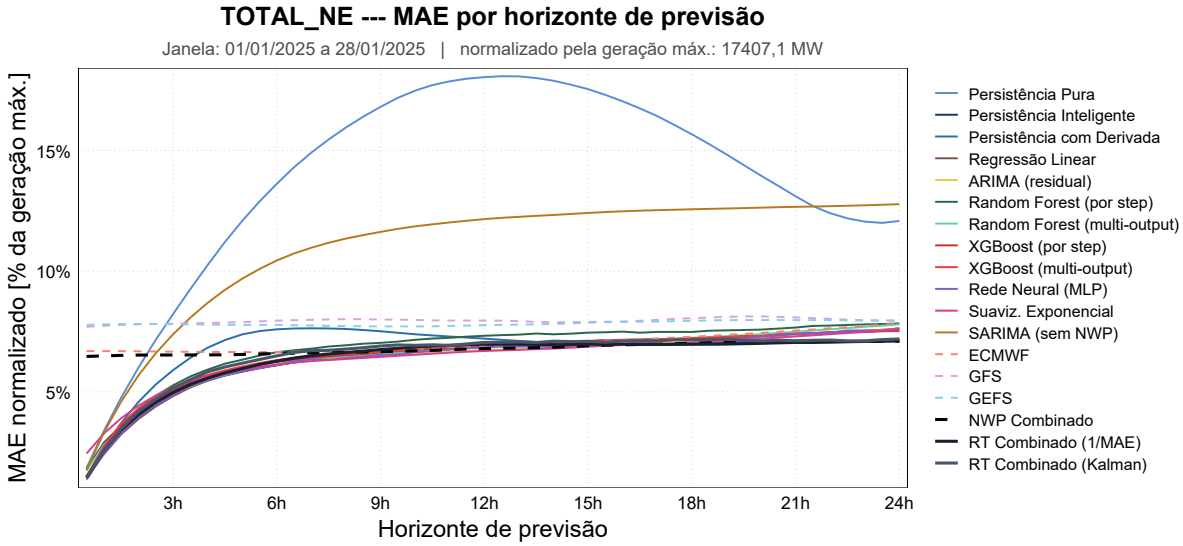
Figura 116 – MAE normalizado pela geração máxima do período em função do horizonte de previsão no subsistema Sul.



Fonte: Elaboração própria.

No subsistema Nordeste (Figura 117), o comportamento qualitativo dos modelos é análogo, mas o patamar geral de erro é sensivelmente menor: a Persistência Pura atinge seu máximo próximo de 18% (contra 22% no Sul), o SARIMA fica em torno de 13% ao final do horizonte, e as combinações de tempo real, junto com os modelos apoiados em NWP, situam-se em uma faixa próxima de 2 a 7% ao longo do horizonte de previsão. As posições relativas entre os modelos são as mesmas nos dois subsistemas, o que reforça a robustez das conclusões.

Figura 117 – MAE normalizado pela geração máxima do período em função do horizonte de previsão no subsistema Nordeste.

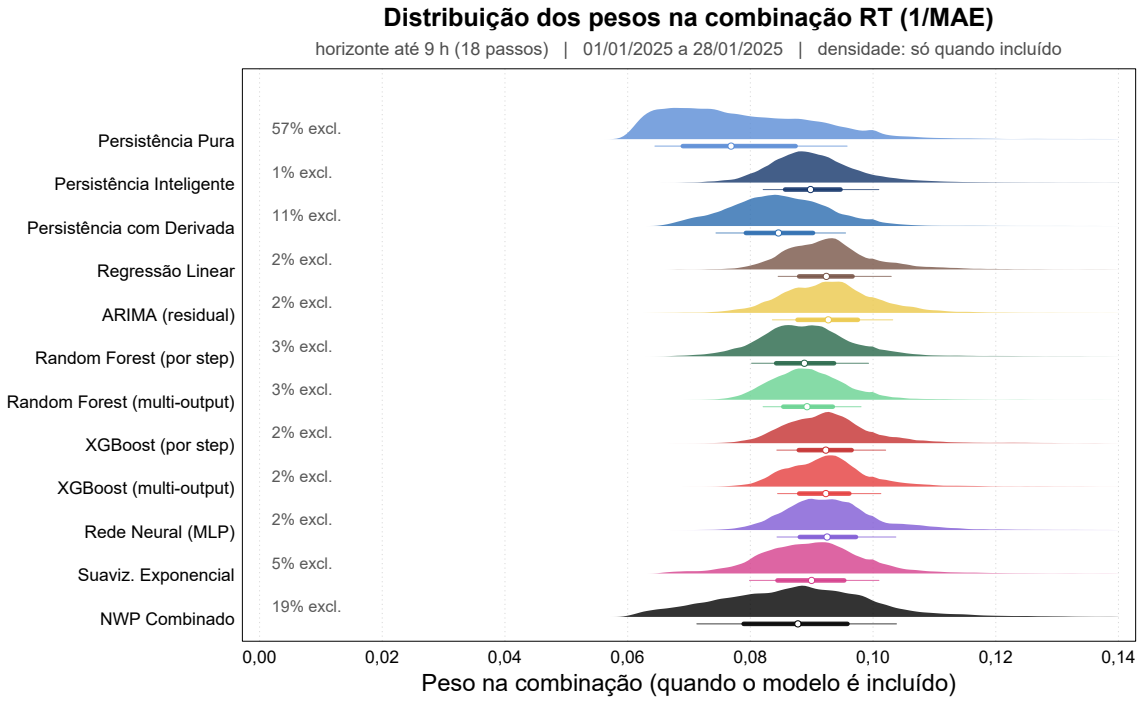


Fonte: Elaboração própria.

6.5.3 Comportamento dos pesos da combinação

A combinação dos modelos em tempo real atribui, a cada emissão e para cada horizonte, um peso a cada modelo, e a análise desses pesos permite entender quais modelos contribuem, em quais horizontes e por quê. Nesta seção, essa análise é feita para a combinação por inverso do erro (1/MAE), pois nela o peso é uma função explícita do erro recente e a exclusão segue uma regra simples de limiar, o que permite associar de forma direta a posição e a frequência de exclusão de cada modelo ao seu desempenho por horizonte. A Figura 118 apresenta a distribuição dos pesos recebidos por cada modelo e a Figura 119 o peso médio ao longo do horizonte, considerando os agrupamentos de usinas do SIN e o período de 1 a 28 de janeiro de 2025. O horizonte analisado, de até 9 horas, corresponde à faixa em que os modelos já estão convergindo para a previsão do NWP combinado.

Figura 118 – Distribuição dos pesos dos modelos na combinação por inverso do erro (1/MAE) e frequência de exclusão de cada modelo, para os agrupamentos de usinas do SIN e o horizonte de até 9 horas.



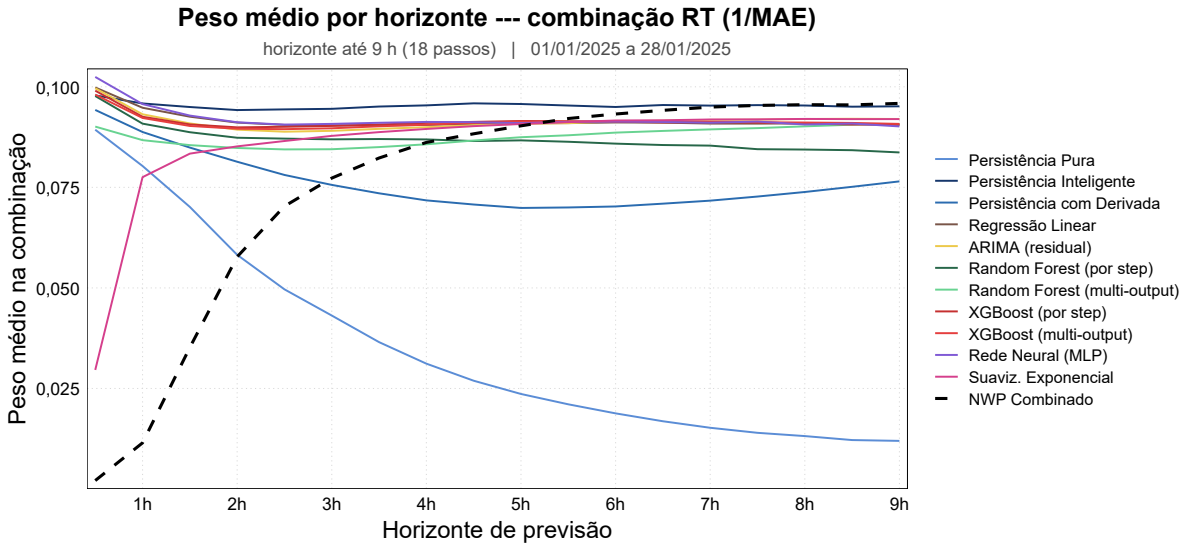
Fonte: Elaboração própria.

Na Figura 118, a densidade de cada modelo é estimada apenas sobre os pesos positivos, pois o peso nulo corresponde a um evento distinto, a exclusão do modelo pelo limiar, que é reportado separadamente pela frequência indicada à esquerda. A barra fina cobre o intervalo de 10% a 90%, a barra grossa o intervalo interquartílico e o ponto branco a mediana, e os modelos estão ordenados pelo peso médio quando incluídos. As distribuições dos modelos competitivos são semelhantes e concentradas em torno de 0,08 a 0,10, com baixa frequência de exclusão, entre 1% e 5%. Como há doze modelos candidatos, pesos próximos de 0,083 indicam que, ao longo das primeiras horas, esses modelos têm desempenho comparável e que a ponderação se aproxima de uma média entre eles; o ganho da combinação decorre, portanto, mais da exclusão

dos modelos de pior desempenho e da troca entre fontes ao longo do horizonte do que de pesos muito díspares.

Dois modelos destoam do conjunto. A persistência pura é excluída em cerca de 57% do tempo e, mesmo quando incluída, recebe pesos menores, o que mostra que, a partir de poucas horas, ela deixa de ser competitiva, como esperado, já que apenas repete o último valor observado. A persistência com derivada segue o mesmo padrão, de forma mais branda, com cerca de 11% de exclusão. O NWP combinado apresenta exclusão intermediária, de cerca de 19%, e uma distribuição mais larga, pois tende a ser descartado nos primeiros passos, em que o estado atual é mais informativo, e a ganhar peso nos horizontes mais longos da análise.

Figura 119 – Peso médio de cada modelo na combinação por inverso do erro (1/MAE) em função do horizonte de previsão, para os agrupamentos de usinas do SIN.



Fonte: Elaboração própria.

A Figura 119 evidencia a transição entre dois grupos de modelos ao longo do horizonte, com o peso médio calculado incluindo as exclusões. No primeiro passo, o NWP combinado parte de peso praticamente nulo e a suavização exponencial de peso baixo, enquanto os demais modelos, baseados no estado atual, partem de pesos próximos de 0,09 a 0,10. À medida que o horizonte aumenta, o peso da persistência pura cai de forma acentuada e contínua, de cerca de 0,09 para pouco mais de 0,01 em 9 horas, e o da persistência com derivada cai de modo mais suave. Em sentido inverso, o peso do NWP combinado cresce de forma monotônica e passa a ser comparável ao dos melhores modelos a partir de cerca de 6 horas, enquanto a suavização exponencial se aproxima rapidamente do patamar dos demais. Os modelos estatísticos e de aprendizado de máquina mantêm pesos estáveis em torno de 0,09 ao longo de todo o horizonte, atuando como um núcleo robusto da combinação. O cruzamento entre as curvas do NWP combinado e das persistências, entre cerca de 2 e 4 horas, marca a fronteira em que a combinação transfere a confiança dos modelos baseados no estado atual para a previsão meteorológica.

Esse comportamento justifica o uso de uma combinação dinâmica por horizonte com

exclusão por limiar: não há um único modelo ótimo para todos os horizontes, e modelos que seriam prejudiciais nos horizontes mais longos da análise, como a persistência pura, contribuem nos primeiros passos e são descartados de forma automática quando deixam de ser competitivos, sem a necessidade de removê-los previamente do conjunto.

A combinação por filtro de Kalman persegue o mesmo objetivo por um mecanismo distinto. Em vez de calcular os pesos diretamente a partir do erro sobre uma janela, ela os trata como um estado atualizado recursivamente a cada nova observação verificada e redefine dinamicamente, por horizonte, o subconjunto de modelos que entra na combinação, em vez de aplicar um limiar fixo. Em decorrência, seus pesos tendem a ser mais variáveis, com distribuições mais largas e caudas mais pronunciadas, e a composição do conjunto ativo tende a oscilar mais ao longo do tempo. Essa maior variabilidade é o custo, e potencialmente a vantagem, da adaptatividade: um combinador mais reativo pode acompanhar melhor mudanças de regime, ao preço de pesos menos estáveis e de leitura menos direta. Por essa razão, a análise dos pesos é apresentada para o 1/MAE, cujo comportamento é mais estável e interpretável.

6.5.4 Distribuição do erro por horizonte

As Figuras 120 e 121 apresentam a distribuição do erro (verificado menos previsto, normalizado pela geração máxima) para os subsistemas Sul e Nordeste, respectivamente, nos horizontes de 30 minutos (h+1), 3 horas (h+6) e 24 horas (h+48). Em ambos os casos, a dispersão do erro cresce com o horizonte, mantendo o comportamento observado no caso por agrupamentos. Nos horizontes curtos, todos os modelos concentram-se em faixas estreitas em torno de zero; no horizonte de 24 horas, os modelos que dependem exclusivamente da série histórica de geração apresentam distribuições visivelmente mais dispersas que os modelos apoiados em NWP.

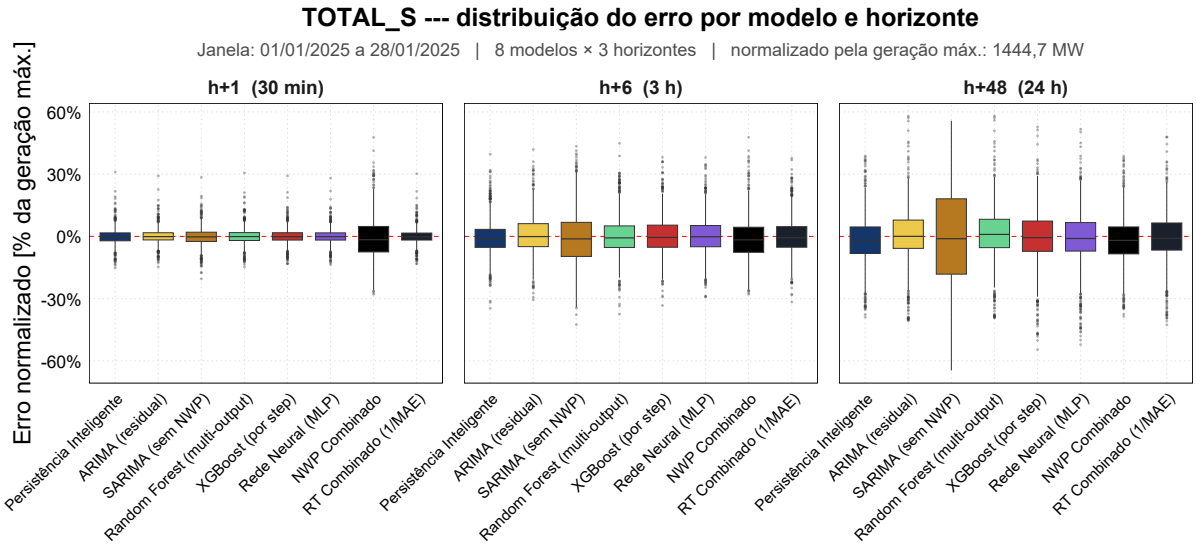
No subsistema Sul (Figura 120), o SARIMA se destaca negativamente no horizonte de 24 horas, com dispersão e amplitude interquartílica muito superiores às dos demais modelos, reforçando a conclusão da subseção anterior de que modelos sem NWP são inadequados para horizontes longos em sítios com variabilidade significativa. A combinação de tempo real (RT Combinado $1/\text{MAE}$) mantém-se entre as distribuições mais estreitas em todos os horizontes, corrigindo, no horizonte h+1, o pequeno viés negativo do NWP Combinado.

No subsistema Nordeste (Figura 121), o padrão qualitativo é análogo, com o SARIMA apresentando novamente a maior dispersão em h+48 e a RT Combinado $1/\text{MAE}$ figurando entre as distribuições mais compactas.

6.5.5 Síntese dos resultados por subsistema

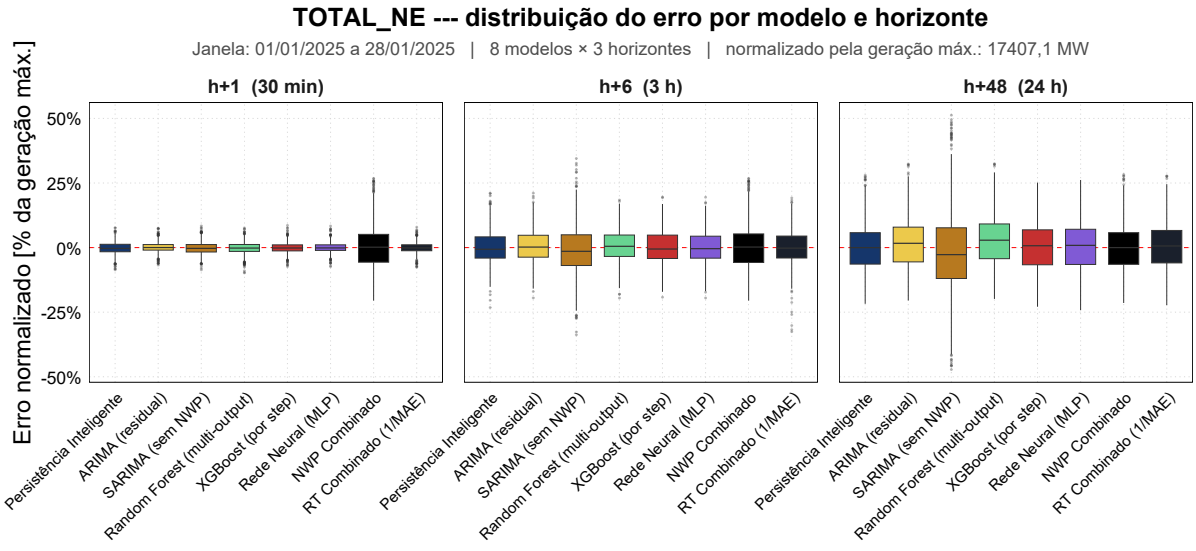
A avaliação por subsistema confirma, em nível agregado, as conclusões obtidas por agrupamento individual: as combinações de tempo real (RT Combinado $1/\text{MAE}$ e Kalman) apresentam consistentemente os menores erros e as distribuições mais compactas; os modelos baseados exclusivamente em série histórica (Persistência Pura e SARIMA) são inadequados

Figura 120 – Distribuição do erro normalizado por modelo nos horizontes de 30 minutos, 3 horas e 24 horas no subsistema Sul.



Fonte: Elaboração própria.

Figura 121 – Distribuição do erro normalizado por modelo nos horizontes de 30 minutos, 3 horas e 24 horas no subsistema Nordeste.



Fonte: Elaboração própria.

fora de horizontes muito curtos; e a informação meteorológica, na forma do NWP Combinado, permanece como base indispensável das previsões. A agregação por subsistema reduz o patamar geral de erro por conta da suavização espacial, mas preserva a hierarquia relativa entre os modelos, o que evidencia a estabilidade da metodologia proposta em diferentes escalas de agregação. Essa avaliação é completa para ambos os subsistemas, consolidando as conclusões obtidas em nível de agrupamento.

6.6 Considerações finais do capítulo

Este capítulo apresentou os resultados obtidos em cada uma das etapas que compõem o trabalho (tratamento de dados, estimação e previsão de geração), permitindo avaliar de forma integrada o desempenho das metodologias propostas nos Capítulos 3, 4 e 5.

Quanto ao tratamento de dados, os resultados demonstraram que o processo de tratamento progressivo é capaz de reduzir substancialmente o percentual de falhas remanescentes nas séries históricas de vento e geração das usinas analisadas. As séries brutas de velocidade do vento apresentaram, em geral, percentuais de dados inválidos superiores aos das séries de geração, o que evidencia a importância de um arcabouço robusto de tratamento, especialmente para os registros anemométricos. A aplicação sucessiva das etapas de composição de fontes, preenchimento espacial e estimação reduziu as lacunas a níveis aceitáveis para a maioria das usinas.

Em relação à estimação da geração, os resultados confirmaram a robustez da metodologia de construção dinâmica das curvas V $\times$ P. A avaliação realizada em diferentes níveis de agregação indicou elevada capacidade de representação da geração observada, com erros médios reduzidos e baixo viés sistemático. Observou-se ainda que as análises em grupos de parques e em nível de subsistema tendem a apresentar desempenho superior ao observado em usinas individuais, conforme detalhado no Apêndice C, consequência da redução da variabilidade aleatória proporcionada pelo efeito de portfólio. Os casos de maior desvio mostraram-se associados, em grande parte, não a deficiências do modelo, mas a inconsistências nos dados de geração observada e a condições operacionais atípicas, como restrições de disponibilidade e eventos de *curtailment*, que nem sempre são fielmente refletidos nos dados reportados ao operador. Vale ressaltar, conforme discutido também na literatura decorrente desta pesquisa, a dificuldade de estabelecer comparações diretas (*benchmarking*) com outros trabalhos, dada a inexistência de bases de dados e estudos equivalentes, no contexto brasileiro, com escala comparável em número de usinas e grupos operacionais analisados.

No que tange à previsão de geração, os resultados evidenciaram ganhos consistentes de acurácia em relação à metodologia anteriormente vigente, com maior redução dos erros nas primeiras semanas do horizonte de previsão, período em que a previsibilidade meteorológica é naturalmente mais elevada. Diferentemente da estimação, a etapa de previsão é desenvolvida para conjuntos agregados de usinas, o que permite explorar diretamente o efeito de portfólio e reduzir parte da variabilidade associada à geração individual. Os resultados confirmaram ainda

a maior complexidade da previsão na região Sul em comparação ao Nordeste, em razão da maior variabilidade dos sistemas atmosféricos que influenciam o regime de ventos da região. As análises de sensibilidade aos modelos meteorológicos e a evolução dos vieses ao longo do horizonte demonstraram a capacidade da metodologia proposta de absorver e compensar parte dos vieses sistemáticos presentes nas previsões de vento, resultando em previsões de geração mais precisas e consistentes.

Em síntese, os resultados das três etapas demonstram a coerência e a complementaridade das metodologias desenvolvidas: a qualidade do tratamento de dados sustenta a acurácia da estimação, que por sua vez fornece insumos consistentes, como as curvas V $\times$ P, para os modelos de previsão. O conjunto de resultados, aplicado em larga escala a mais de mil parques eólicos do SIN, evidencia a adequação das abordagens propostas para uso operacional e para o planejamento eletroenergético em sistemas com alta penetração de fontes renováveis variáveis.

7 CONSIDERAÇÕES FINAIS

Este trabalho propôs um conjunto integrado de metodologias e modelos para o tratamento de dados, a estimação e a previsão da geração eólica, voltado à operação e ao planejamento do SIN. Retomando os objetivos estabelecidos no Capítulo 1, considera-se que todos foram atingidos, conforme se detalha a seguir. Desenvolveu-se uma metodologia de tratamento e recomposição das séries históricas de vento e geração, apresentada no Capítulo 3, capaz de integrar múltiplas fontes de medição, identificar falhas e dados espúrios e produzir séries consistentes para os modelos subsequentes. A partir dessas séries, propôs-se um modelo de estimação da geração baseado na construção automática e dinâmica de modelos de conversão de vento em potência, descrito no Capítulo 4, aplicável tanto à reconstrução de séries quanto à quantificação das perdas associadas a eventos de corte de geração. No âmbito da previsão, detalhado no Capítulo 5, foram desenvolvidos os modelos de pós-processamento das previsões meteorológicas, os modelos de previsão da geração para os horizontes de tempo real, curtíssimo e curto prazo, e os modelos dinâmicos de combinação das previsões provenientes de diferentes fontes. Adicionalmente, foram conduzidas análises comparativas entre métodos clássicos de séries temporais e técnicas de aprendizado de máquina, e caracterizou-se o comportamento da geração eólica no Brasil, em especial as diferenças entre os regimes das regiões Sul e Nordeste, evidenciadas ao longo dos resultados apresentados no Capítulo 6 e no Apêndice C. As considerações a seguir detalham as principais constatações relativas a cada uma dessas frentes, na mesma ordem em que os modelos são desenvolvidos ao longo do trabalho.

A Tabela 13 sintetiza o atendimento aos objetivos, relacionando cada um às contribuições da tese, apresentadas em detalhe no Capítulo 1, às suas principais evidências e às publicações associadas. Ressalta-se que as contribuições aqui apresentadas consolidam e aprofundam resultados previamente publicados, complementados por desenvolvimentos adicionais realizados ao longo desta pesquisa.

Tabela 13 – Síntese das contribuições e evidências associadas ao atendimento dos objetivos da tese.

Objetivo	**Contribuição da tese**	**Principal evidência ou resultado**	**Publicação decorrente**
Caracterizar a operação da geração eólica no Brasil em larga escala	Panorama do comportamento operacional eólico a partir de dados reais de mais de mil usinas, com as diferenças de regime entre Sul e Nordeste e os impactos das restrições operativas	Caracterização regional e sub-regional apresentada no Capítulo 6 e no Apêndice C	[31], [32]
Avaliar os modelos numéricos de previsão do tempo no contexto brasileiro	Análise e comparação dos principais modelos meteorológicos aplicados às localidades dos parques nacionais, caracterizando limitações e viés por horizonte	Comparação entre ECMWF, GFS e GEFS ao longo dos horizontes, nos Capítulos 5 e 6	[32]

continua na próxima página

(continuação da Tabela 13)

Objetivo	**Contribuição da tese**	**Principal evidência ou resultado**	**Publicação decorrente**
Desenvolver uma metodologia de tratamento e composição das séries de vento e geração	Metodologia que integra e trata dados de múltiplas fontes (ONS, CCEE, EPE, ANEEL), detectando falhas e dados espúrios e compondo séries consistentes por parque	Redução progressiva do percentual de falhas nas séries tratadas, no Capítulo 6	[30], [39]
Estimar a geração potencial das usinas eólicas	Construção automática e dinâmica das curvas de conversão de vento em potência, com remoção de pontos sob restrição e estimação da geração na ausência de corte	Estimação em escala sistêmica e quantificação das perdas por *curtailment*, no Capítulo 6 e no Apêndice C	[31], [35], [36], [41]
Ajustar as previsões de vento dos modelos meteorológicos	Modelagem e correção do viés de vento de forma específica por modelo e por horizonte, com representação da sua sazonalidade	Redução do viés e da dispersão das previsões de vento, no Capítulo 6 e no Apêndice C	[32], [42]
Aprimorar a previsão e a representação eólica nos horizontes de curtíssimo e de curto prazo	Modelo único de previsão, horário e com alcance de até um mês, fundamentado nas previsões numéricas do tempo, em vez de se basear no comportamento passado da geração observada, e com representação conjunta de múltiplos parques, do qual derivam os produtos de curtíssimo e de curto prazo	Superioridade frente a abordagens clássicas e ganhos ante a metodologia anterior, nos Capítulos 5 e 6	[32], [33], [36], [37], [38], [40], [42]
Desenvolver modelos de previsão para a operação em tempo real	*Ensemble* de modelos de tempo real combinados dinamicamente, integrando previsões meteorológicas e correções baseadas na geração verificada mais recente	Correção dos desvios da previsão nos primeiros horizontes, com base na geração verificada mais recente, e convergência suave à previsão de referência ao final, no Capítulo 6 e no Apêndice C	[34]

Nota: a relação entre contribuições e publicações não é de um para um.

Fonte: Elaboração própria.

Os parágrafos a seguir detalham cada uma das contribuições sintetizadas na Tabela 13, discutindo as evidências que sustentam o atendimento aos objetivos. A metodologia proposta para tratamento e composição das séries temporais atendeu de forma eficaz ao objetivo. O modelo de tratamento de dados apresenta boa capacidade de identificação dos dados faltantes (*missing data*) e dados espúrios (*outliers*) das séries temporais de geração e de vento observados, considerando a aplicação do método em dados com diferentes graus de severidade em termos de falhas de medição. Embora os princípios fundamentais da metodologia proposta possam ser aplicados a outros tipos de séries temporais, o modelo desenvolvido é específico para a fonte eólica e é voltado para aplicações que envolvem a consideração simultânea de múltiplos parques. O principal diferencial da metodologia é sua capacidade de absorver e tratar dados (de geração e/ou vento) de várias fontes, incluindo diversas instituições do setor e outros parques eólicos, para identificar falhas e gerar as séries temporais tratadas.

Na maioria dos casos, o filtro identificou corretamente as falhas de medição. No entanto, como qualquer modelo baseado em métodos estatísticos e computacionais, também é esperado que haja identificação/substituição de falsos dados espúrios. Essas identificações incorretas ocorrem esporadicamente, o que mantém a qualidade geral do modelo. Dessa forma, é possível concluir que o modelo proposto tem o potencial de oferecer uma contribuição substancial para os processos de estimação e de previsão de geração eólica, que posteriormente serão utilizados nos modelos de planejamento e operação do ONS. Quanto à aplicabilidade prática, esse modelo pode ser facilmente incorporado aos processos do ONS devido à sua baixa complexidade algorítmica e à rapidez de processamento.

No que se refere ao modelo de estimação da geração, os resultados confirmaram a eficácia da construção automática e dinâmica dos modelos de conversão de vento em potência a partir de dados medidos. Embora existam diversas técnicas disponíveis para essa finalidade, o modelo proposto consegue capturar automaticamente e com boa qualidade os pontos sob restrição operativa e outros problemas relacionados ao ajuste desse tipo de relação, por meio de filtros e controles automáticos das curvas V $\times$ P, o que viabiliza sua aplicação em larga escala, cobrindo a totalidade das usinas eólicas do SIN sem necessidade de ajuste manual caso a caso. Cabe destacar dois papéis fundamentais dessa etapa. O primeiro é a sua contribuição para a composição das séries históricas de geração: a geração estimada é utilizada no preenchimento de lacunas remanescentes das séries observadas, complementando o processo de tratamento de dados descrito anteriormente. O segundo é a estimação da geração potencial, isto é, daquela que ocorreria na ausência de corte, essencial para a quantificação da energia perdida por *curtailment* e para fornecer um referencial adequado ao ajuste dos modelos de previsão, uma vez que a geração verificada, nos períodos de corte, deixa de refletir o recurso eólico disponível.

Embora parcialmente abordadas durante a etapa de construção dos modelos de conversão de vento em potência, situações de restrição na operação dos parques podem ter impactos negativos nos resultados do modelo proposto. Nesse sentido, a metodologia desenvolvida [31,35], resultante desta pesquisa, permite a consideração da disponibilidade de potência para corrigir a estimação das curvas V $\times$ P. Isso resulta em ganhos significativos ao reduzir a dispersão dos pontos considerados, contribuindo para um ajuste mais preciso da curva que representa a capacidade de geração das usinas eólicas. Essa abordagem é especialmente relevante quando os parques estão sujeitos a restrições físicas que limitam sua capacidade de geração, o que proporciona melhorias diretas no desempenho da estimação e das previsões.

No que tange ao desempenho dos modelos meteorológicos avaliados, os resultados evidenciaram diferenças relevantes entre o ECMWF, o GFS e o GEFS, além de comportamentos distintos entre as regiões Sul e Nordeste. De modo geral, o ECMWF apresentou o melhor desempenho na maior parte dos horizontes. Um aspecto que merece destaque é o desempenho do GEFS em relação ao GFS: mesmo possuindo resolução espacial e temporal inferiores à do GFS, o GEFS tende a produzir melhores resultados a partir do segundo dia de previsão, comportamento

coerente com sua natureza de *ensemble*, cuja média entre os membros atenua parte da incerteza individual de cada previsão. Essa constatação reforça o valor das abordagens probabilísticas de previsão numérica do tempo à medida que o horizonte se estende.

No que se refere ao modelo de previsão da geração, uma etapa fundamental é a modelagem das previsões numéricas de vento. O modelo de ajuste desenvolvido consegue aprimorar essas previsões de forma simples e eficaz, capturando como o viés se manifesta de maneira distinta em cada intervalo do dia e ao longo do horizonte, e corrigindo-o de forma individualizada. Uma das evoluções metodológicas mais impactantes dessa etapa é a identificação e representação da sazonalidade do viés na previsão de vento dos modelos meteorológicos. A partir das previsões de vento ajustadas, a geração é obtida por meio dos modelos de conversão de vento em potência, resultando em um modelo único de previsão com alcance de até um mês à frente. Entre os principais benefícios alcançados, destaca-se a capacidade de representar conjuntamente múltiplos parques eólicos, minimizando os efeitos da alta variabilidade na velocidade do vento que não são capturados pelos modelos meteorológicos, bem como a aplicabilidade da metodologia em larga escala, com a construção automática dos modelos de conversão de vento em potência para todo o parque eólico do SIN. Cabe ressaltar, ainda, que a abordagem metodológica adotada não explora a autocorrelação das variáveis, ou seja, o modelo não está vinculado a atrasos temporais das séries temporais, o que é particularmente relevante quando se trata de séries com características estocásticas tão específicas, como a fonte eólica no horizonte de previsão considerado, do dia seguinte a vários dias à frente.

Desse modelo único derivam dois blocos que tratam as particularidades dos horizontes de curtíssimo e de curto prazo. No curtíssimo prazo, cujas previsões alimentam o MPCTP (atualmente o DESSEM), destaca-se a capacidade do modelo de separar as previsões dos conjuntos de usinas em previsões por usina individual, permitindo o posterior reagrupamento nos diferentes formatos espaciais exigidos pelos diversos usos da operação, respeitando os limites de potência instalada de cada usina. Cabe ressaltar a boa qualidade do modelo de previsão desenvolvido: resultados preliminares desta pesquisa [42] já indicavam que, para os horizontes de interesse, os modelos baseados em previsões numéricas de tempo superam abordagens clássicas de séries temporais e técnicas de aprendizado de máquina que se apoiam apenas no histórico da geração. Essa constatação é reforçada pelos resultados apresentados no Capítulo 6: mesmo dentro do primeiro dia de previsão, a partir de meados desse horizonte, o modelo baseado nos NWPs já se mostra superior às demais abordagens, e essa vantagem se amplia nos horizontes seguintes, que são justamente os de maior interesse para a programação diária da operação e para o horizonte de até sete dias do MPCTP.

No curto prazo, cujas previsões alimentam o MPCP (atualmente o DECOMP), a principal particularidade é a necessidade de modelar também os montantes de geração de usinas sem observação disponível, isto é, as usinas não supervisionadas e as usinas futuras, que ainda não dispõem de históricos confiáveis. Para essas usinas, não sendo viável a aplicação direta dos modelos de

conversão de vento em potência, foram desenvolvidas metodologias específicas de estimação, permitindo compor a previsão total do subsistema. Ressalta-se, ainda, a compatibilização dos dados de geração eólica entre os horizontes de curtíssimo e curto prazo, o que aproxima as etapas da programação diária da operação e do planejamento eletroenergético, com maior assertividade das previsões eólicas em todo o primeiro mês. Cabe destacar que essas previsões são fornecidas em base semanal ao MPCP e, adicionalmente, o montante do primeiro mês é utilizado como previsão do primeiro mês no MPMP (atualmente o NEWAVE). Em relação ao desempenho, verificou-se uma maior aproximação entre os montantes de geração verificados e previstos para as semanas operativas do primeiro mês do MPCP, em comparação com o modelo anteriormente em vigor [29], com os maiores benefícios na primeira semana, seguidos de melhorias mais modestas nas semanas subsequentes.

Para o horizonte da operação em tempo real, foi desenvolvido um modelo baseado em um *ensemble* de modelos de previsão de geração combinados dinamicamente, que integra desde adaptações da persistência até modelos estatísticos e de aprendizado de máquina, tendo como referência a combinação das previsões de geração baseadas nos modelos meteorológicos. A combinação é adaptativa, com pesos recalculados continuamente segundo o desempenho recente de cada modelo em cada horizonte, por meio de duas estratégias: a ponderação pelo inverso do erro absoluto médio e a combinação por filtro de Kalman. As atualizações são realizadas a cada meia hora, incorporando as informações de geração verificada mais recentes. O principal benefício da metodologia é sua capacidade de se adaptar aos dados verificados mais recentes, corrigindo o viés das previsões meteorológicas nos primeiros horizontes e convergindo a elas de forma suave ao final do horizonte de tempo real. Cabe ressaltar, contudo, que, por se apoiar em correções sobre previsões já realizadas, o modelo não tem por objetivo antecipar fenômenos de rápida duração, como as rampas de geração.

O modelo de tempo real, por se basear em dados de geração adquiridos em tempo real, pode gerar previsões contaminadas sempre que há falha de medição ou o parque opera sob restrição operativa. Um caso cada vez mais comum, dada a crescente penetração da fonte eólica no SIN, são as solicitações de redução de geração eólica por comando do ONS em função de razões elétricas e/ou energéticas. O momento de liberação do corte é um desafio para os operadores nas salas de controle, uma vez que o patamar para o qual a geração eólica irá saltar ao final da restrição é desconhecido. Uma alternativa para esse problema, explorada nesta pesquisa, consiste em utilizar não a geração em tempo real, mas a velocidade do vento verificada em tempo real, estimando a geração que estaria sendo verificada a despeito dos cortes.

Em síntese, este trabalho propôs um conjunto de metodologias de estimação e de previsão de geração eólica adaptadas a sistemas elétricos de grande porte com alta penetração dessa fonte. Embora desenvolvidas considerando os requisitos específicos do operador do sistema brasileiro (ONS), as abordagens demonstram potencial de aplicabilidade a operadores de sistema em outros contextos, bem como à comunidade acadêmica dedicada ao avanço de técnicas preditivas

para fontes renováveis variáveis. Contribui para essa transferibilidade o fato de os modelos de tratamento de dados, estimação e previsão serem retreinados de forma automática e dinâmica, permitindo sua adaptação contínua às mudanças observadas no sistema elétrico e nos padrões climáticos. Além disso, os modelos de conversão de vento em potência são continuamente recalibrados, possibilitando acompanhar o envelhecimento dos parques eólicos e as alterações graduais na relação entre vento e potência gerada ao longo do tempo.

Os resultados obtidos até o momento incentivam a continuidade dos estudos voltados ao aprimoramento do processo de previsão da geração eólica, tanto para a Programação Diária da Operação quanto para o planejamento de curto prazo e a operação em tempo real. O aumento da qualidade dos dados, dos processos de tratamento e das previsões de geração possibilita uma utilização mais eficiente dos recursos energéticos, reduzindo, por exemplo, a necessidade de manutenção de reservas operativas ou de despacho adicional de geração termelétrica além daquele definido pelo mérito econômico. Como consequência, essas melhorias podem contribuir para a redução dos custos operacionais do sistema.

Além dos benefícios econômicos, os avanços obtidos contribuem para o aumento da segurança operativa, ao reduzir as incertezas associadas à geração eólica nas decisões de programação e de operação em tempo real. Em particular, os modelos de estimação e previsão em tempo real constituem importantes ferramentas para a gestão dos cortes de geração e sua posterior liberação. A estimação em tempo real pode fornecer ao operador uma indicação do patamar de geração para o qual as usinas tendem a convergir após a remoção de uma restrição operativa. Essa informação é relevante porque uma liberação realizada sem o devido controle ou sob elevado grau de incerteza pode resultar em uma elevação da geração superior à esperada, ocasionando, por exemplo, eventos de sobrefrequência no sistema elétrico.

Cabe ressaltar, contudo, que a quantificação direta do benefício operacional associado à redução dos erros de previsão não foi objeto deste trabalho, uma vez que ainda não existe metodologia consolidada para traduzir ganhos de acurácia em ganhos operacionais. Entre as principais dificuldades, destacam-se a elevada não linearidade da operação do sistema, na qual um mesmo erro de previsão pode ter efeitos muito distintos conforme as condições de carga, hidrologia, disponibilidade de geração e restrições da rede, e a impossibilidade de reproduzir exatamente as decisões operativas tomadas no passado, que dependeram de informações e condições específicas de cada momento. Por essa razão, o impacto costuma ser estimado de forma indireta, por meio de seus efeitos sobre o despacho, sobre o custo marginal de operação e sobre os resultados dos modelos energéticos. Assim, os ganhos de acurácia apresentados nesta tese devem ser interpretados como indicativos do potencial de benefício operacional, e não como uma medida direta desse benefício.

7.1 Trabalhos futuros

A partir dos desenvolvimentos apresentados neste trabalho, vislumbram-se diversas linhas de continuidade capazes de ampliar o alcance e o desempenho das metodologias propostas. As principais perspectivas de trabalhos futuros são descritas a seguir, em uma sequência sugerida para o seu desenvolvimento, embora algumas possam ser conduzidas em paralelo:

I. Ampliação do portfólio de modelos numéricos de previsão do tempo, com a incorporação de previsões de alta resolução do modelo IFS/ECMWF (ECMWF-HRES), cuja avaliação preliminar indicou ganhos de desempenho, especialmente nos primeiros dias de previsão, bem como o aprofundamento da compreensão sobre as vantagens das abordagens de *ensemble* na previsão de vento, a exemplo do melhor desempenho do GEFS em relação ao GFS apesar de sua menor resolução espacial e temporal;

II. Extensão das metodologias de tratamento de dados, de estimação e de previsão para outras fontes renováveis variáveis, em especial a geração solar fotovoltaica, e desenvolvimento de abordagens integradas e multivariadas que modelem de forma simultânea as gerações eólica e solar, inclusive em usinas híbridas, explorando sua complementaridade e a dependência entre seus erros de previsão, o que beneficia um melhor dimensionamento da reserva de potência, bem como a incorporação de outros elementos do sistema, como a carga e o armazenamento;

III. Incorporação explícita da disponibilidade dos aerogeradores na construção e na aplicação das curvas V $\times$ P, refletindo de forma mais fiel as condições operacionais reais dos parques;

IV. Evolução para um sistema de previsão híbrido, integrando as previsões produzidas centralizadamente pelo operador às previsões fornecidas pelos agentes de mercado, o que demandaria adaptações no arcabouço regulatório brasileiro;

V. Aprimoramento das técnicas de pós-processamento das previsões de geração e avaliação de técnicas de *machine learning* em diferentes horizontes temporais;

VI. Avaliação da incorporação de metodologias de previsão e de geração de cenários sintéticos para os horizontes de médio e longo prazo, fornecendo uma caracterização mais abrangente dos riscos associados às restrições de transmissão, à penetração renovável e à evolução da demanda;

VII. Desenvolvimento de metodologias para a avaliação e a quantificação econômica dos ganhos decorrentes da melhoria da previsão de geração eólica, por exemplo por meio de simulações, com os modelos energéticos e com ferramentas que reproduzam a operação real do sistema, que comparem o despacho e o custo marginal de operação obtidos com previsões de diferentes níveis de acurácia, traduzindo os aprimoramentos de acurácia em benefícios operacionais e financeiros para o sistema;

VIII. Desenvolvimento de modelos dedicados à previsão de rampas de geração, voltados à antecipação de variações abruptas da geração eólica, que representam um dos maiores desafios para a operação em tempo real;

IX. Integração de abordagens probabilísticas aos modelos de previsão e de estimação, em linha com a transição do setor elétrico de previsões determinísticas para uma operação orientada à quantificação do risco, que subsidia decisões como o dimensionamento de reservas e a gestão de cortes, permitindo a representação explícita das incertezas associadas à geração eólica e a avaliação de cenários de corte de geração sob diferentes condições de operação e expansão, bem como a geração de trajetórias consistentes entre usinas e entre horizontes de previsão;

X. Investigação das causas físicas dos erros das previsões dos modelos meteorológicos, incluindo os efeitos da topografia, da heterogeneidade da superfície, das circulações de mesoescala e dos processos de escala inferior à da grade, por meio de modelos de mesoescala ou de maior resolução, de modo a subsidiar melhorias na correção das previsões de vento;

XI. Articulação junto aos órgãos nacionais de meteorologia e ao ONS para a evolução das previsões de vento dos modelos numéricos operados em território nacional, com ênfase em aplicações eólicas, de modo que possam, no futuro, ser incorporados aos modelos de combinação e integrados ao processo operacional de previsão de geração;

XII. Evolução do modelo estatístico de preenchimento de falhas para abordagens multivariadas, nas quais diferentes variáveis, como geração, velocidade e direção do vento, temperatura e irradiância, sejam modeladas de forma conjunta, explorando as correlações entre elas. Adicionalmente, podem ser investigados modelos de aprendizado de máquina mais sofisticados para essa finalidade, como as redes adversárias generativas (*Generative Adversarial Networks*, GANs);

XIII. Organização e disponibilização pública dos modelos desenvolvidos nesta tese em repositório aberto (por exemplo, GitHub), favorecendo a reprodutibilidade, o reúso e a evolução colaborativa das metodologias;

XIV. Articulação junto ao ONS para a criação de um laboratório aberto de modelagem, com problemas e dados reais do sistema, fomentando o desenvolvimento colaborativo e a validação de novas metodologias pela comunidade técnica e acadêmica.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] IEA. *World Energy Outlook 2024*. International Energy Agency. Paris, France, 2024. Acesso em 15 agosto 2025. Disponível em: <https://iea.blob.core.windows.net/assets/140a0470-5b90-4922-a0e9-838b3ac6918c/WorldEnergyOutlook2024.pdf>.

[2] REN21. *Global Status Report Collection: Energy Supply*. Renewable Energy Policy Network for the 21st Century, 2024. Acesso em 15 agosto 2025. Disponível em: <https://www.ren21.net/>.

[3] EIA. *Independent Statistics and Analysis*. U.S. Energy Information Administration, 2024. Acesso em 10 agosto 2025. Disponível em: <https://www.eia.gov/>.

[4] IEA. *Analysis and Forecast to 2030*. International Energy Agency. Paris, France, 2024. Acesso em 15 agosto 2025. Disponível em: <https://iea.blob.core.windows.net/assets/493a4f1b-c0a8-4bfc-be7b-b9c0761a3e5e/Oil2024.pdf>.

[5] EPE. *Balanço energético nacional 2024*. Empresa de Pesquisa Energética. Rio de Janeiro, RJ, Brazil, 2024. Acesso em 15 agosto 2025. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-819/topico-723/BEN2024.pdf>.

[6] GWEC. *Global Wind Report 2025*. Global Wind Energy Council (GWEC), 2025. Acesso em 21 maio 2026. Disponível em: <https://www.gwec.net/reports/globalwindreport/2025>.

[7] RIBEIRO, L.; STREET, A.; VALLADÃO, D.; FREIRE, A. C.; BARROSO, L. Technical and economical aspects of wholesale electricity markets: An international comparison and main contributions for improvements in Brazil. *Electric Power Systems Research*, Elsevier, v. 220, p. 109364, 2023.

[8] TOLMASQUIM, M. T.; CORREIA, T. de B.; PORTO, N. A.; KRUGER, W. Electricity market design and renewable energy auctions: The case of Brazil. *Energy Policy*, Elsevier, v. 158, p. 112558, 2021.

[9] SANTOS, T.; DINIZ, A.; SABOIA, C.; CABRAL, R.; CERQUEIRA, L. Hourly pricing and day-ahead dispatch setting in Brazil: The DESSEM model. *Electric Power Systems Research*, Elsevier, v. 189, p. 106709, 2020.

[10] CEPEL. *Manual do Usuário do Modelo DESSEM*. Centro de Pesquisas de Energia Elétrica, 2023. Acesso em 6 julho 2025. Disponível em: <https://www.cepel.br/wp-content/uploads/2022/05/DESSEM_ManualUsuario_v19.0.24.3.pdf>.

[11] CEPEL. *Manual de Referência do Modelo DECOMP*. Centro de Pesquisas de Energia Elétrica, 2021. Acesso em 6 julho 2025. Disponível em: <https://www.cepel.br/wp-content/uploads/2022/05/DECOMP_ManualUsuario_v31-Dez2021.pdf>.

[12] CEPEL. *Manual de Referência do Modelo NEWAVE*. Centro de Pesquisas de Energia Elétrica, 2023. Acesso em 4 julho 2025. Disponível em: <https://www.cepel.br/wp-content/uploads/2024/10/ManualUsuario-2.pdf>.

[13] IEA. *Integrating Solar and Wind*. International Energy Agency. Paris, France, 2024. Acesso em 15 agosto 2025. Disponível em: <https://iea.blob.core.windows.net/assets/4e495603-7d8b-4f8b-8b60-896a5936a31d/IntegratingSolarandWind.pdf>.

[14] YOUSUF, M. U.; AL-BAHADLY, I.; AVCI, E. Current perspective on the accuracy of deterministic wind speed and power forecasting. *IEEE Access*, IEEE, v. 7, p. 159547–159564, 2019.

[15] GIZ. *International Best Practices in Solar and Wind Power Forecasting*. Deutsche Gesellschaft für Internationale Zusammenarbeit, 2023. Acesso em 15 agosto 2025. Disponível em: <https://www.get-transform.eu/wp-content/uploads/2024/01/GET.transform-Brief_VRE-Forecasting-Solar-Wind.pdf>.

[16] ELTOHAMY, M. S.; MOTELEB, M. S. A.; TALAAT, H. E.; MEKHEMER, S. F.; OMRAN, W. Analyzing wind power ramps for high penetration of variable renewable generation. In: *International Middle East Power Systems Conference (MEPCON)*. [S.l.: s.n.], 2019. p. 768–775.

[17] VIEIRA, G.; SILVA, W.; LOURENÇO, L.; MONARO, R. Characterizing wind and solar curtailment in Brazil: An evidence-based analysis of operational drivers. *Electric Power Systems Research*, v. 254, p. 109684, 2026.

[18] EA. *Review of Forecasting Provisions for Intermittent Generators in the Spot Market*. Electricity Authority - Te Mana Hiko. New Zealand, 2024. Acesso em 15 agosto 2025. Disponível em: <https://www.ea.govt.nz/documents/5244/Review_of_forecasting_provisions_for_intermittent_generators_in_the_spot_marke_se5mcdm.pdf>.

[19] KUNDUR, P. *Power System Stability and Control*. USA: McGraw-Hill, 1994.

[20] ZSIBORÁCS, H.; PINTÉR, G.; VINCZE, A.; BARANYAI, N. H. Wind power generation scheduling accuracy in Europe: an overview of ENTSO-E countries. *Sustainability*, MDPI, v. 14, n. 24, p. 16446, 2022.

[21] HANIFI, S.; LIU, X.; LIN, Z.; LOTFIAN, S. A critical review of wind power forecasting methods: past, present and future. *Energies*, MDPI, v. 13, n. 15, p. 3764, 2020.

[22] BAZIONIS, I. K.; GEORGILAKIS, P. S. Review of deterministic and probabilistic wind power forecasting: Models, methods, and future research. *Electricity*, MDPI, v. 2, n. 1, p. 13–47, 2021.

[23] BAZIONIS, I. K.; KARAFOTIS, P. A.; GEORGILAKIS, P. S. A review of short-term wind power probabilistic forecasting and a taxonomy focused on input data. *IET Renewable Power Generation*, Wiley Online Library, v. 16, n. 1, p. 77–91, 2022.

[24] PIOTROWSKI, P.; RUTYNA, I.; BACZYŃSKI, D.; KOPYT, M. Evaluation metrics for wind power forecasts: A comprehensive review and statistical analysis of errors. *Energies*, MDPI, v. 15, n. 24, p. 9657, 2022.

[25] JOSHI, M.; INSKEEP, S. *Institutional Framework of Variable Renewable Energy forecasting in India*. [S.l.], 2023. Acesso em 15 agosto 2025. Disponível em: <https://docs.nrel.gov/docs/fy23osti/83949.pdf>.

[26] MAGO, N. *WIND AND SOLAR FORECASTING PRACTICES AT ERCOT*. 2022. Electric Reliability Council of Texas. Acesso em 15 agosto 2025. Disponível em: <https://www.esig.energy/download/wind-and-solar-forecasting-practices-at-ercot-nitika-mago/#>.

[27] MOTLEY, A. *CAISO: Advances in the Use of Wind and Solar Forecasting*. California Independent System Operator. Folsom, CA, USA, 2023. Webinar presentation, G-PST/ESIG Webinar Series. Disponível em: <https://www.esig.energy/event/g-pst-esig-webinar-series-advances-in-the-use-of-wind-and-solar-forecasting/>.

[28] AEMO. *Australian Wind Energy Forecasting System (AWEFS) and Australian Solar Energy Forecasting System (ASEFS)*. Australian Energy Market Operator, 2023. Acesso em 14 agosto 2025. Disponível em: <https://aemo.com.au/-/media/files/electricity/nem/security_and_reliability/dispatch/policy_and_process/australian-wind-energy-forecasting-solar-energy-forecasting-system.pdf?la=en>.

[29] ANEEL. *Resolução Normativa Nº 843/2019*. Agência Nacional de Energia Elétrica, 2019. Acesso em 30 março 2025. Disponível em: <http://www2.aneel.gov.br/cedoc/ren2019843.pdf>.

[30] NASCIMENTO, P. S. C.; MARCATO, A. L. M. A statistical approach for missing data imputation and outlier detection in wind power time series for large-scale power systems. In: *Proceedings of the 61st International Universities Power Engineering Conference (UPEC)*. [S.l.: s.n.], 2026.

[31] NASCIMENTO, P. S. C.; COSSICH, W.; ARAUJO, L.; SANTOS, I.; ALMEIDA, K.; MARCATO, A. L. M. An integrated methodology for assessing wind power curtailment using anemometric measurements and operational data in the Brazilian context. *Atmosphere*, v. 17, n. 4, p. 333, 2026. DOI: 10.3390/atmos17040333.

[32] NASCIMENTO, P. S. C.; DEOTTI, L. M. P.; MARCATO, A. L. M.; FARIAS, W. C. M. Wind power forecasting for independent system operators: Modeling approaches, practical applications, and operational challenges in the Brazilian context. *IEEE Access*, v. 14, p. 27895–27914, 2026.

[33] NASCIMENTO, P. S. C.; ARAUJO, L. F.; ALVES, R. J. M.; FARIAS, W. C. M.; JUNIOR, C. A. A.; MARCATO, A. L. M. Avanços metodológicos na previsão eólica para o planejamento de curto prazo do ONS. In: *Brazil Windpower*. São Paulo: [s.n.], 2024.

[34] NASCIMENTO, P. S. C.; ARAUJO, L. F.; MUNDSTOCK, T. F. G.; MARCATO, A. L. M.; NASCIMENTO, G. S. V. Modelo de correção das previsões de geração eólica em tempo real. In: *Brazil Windpower*. São Paulo: [s.n.], 2022.

[35] NASCIMENTO, P. S. C.; ALVES, R. J. M.; MUNDSTOCK, T. F. G.; JR, A. J. I. A.; MARCATO, A. L. M. Impacto da disponibilidade de potência na previsão e estimação da geração eólica. In: *Brazil Windpower*. São Paulo: [s.n.], 2022.

[36] ARAUJO, L. F.; NASCIMENTO, P. S. C.; GONCALVES, G. A.; MARCATO, A. L. M.; MADEIRA, F.; COSSICH, W. Modelo de estimação de geração eólica: Uma proposta no contexto de recursos energéticos distribuídos. In: *Brazil Windpower*. São Paulo: [s.n.], 2022.

[37] NASCIMENTO, P. S. C.; ARAUJO, L. F.; MARCATO, A. L. M.; JUNIOR, C. A. A.; BARROS, A. M. L. Modelo de previsão eólica semanal para estudos prospectivos de curto prazo do ONS. In: *LIV Simpósio Brasileiro de Pesquisa Operacional*. Juiz de Fora: [s.n.], 2022. Disponível em: <https://proceedings.science/sbpo/sbpo-2022/trabalhos/modelo-de-previsao-eolica-semanal-para-estudos-prospectivos-de-curto-prazo-do-on?lang=pt-br>.

[38] NASCIMENTO, P. S. C.; KHENAYFIS, L. S.; ALVES, R. J. M.; MARCATO, A. L. M.; FONSECA, C. M. S.; SILVA, P. S. O. Desenvolvimento de um modelo de clusterização de usinas eólicas para atendimento dos modelos de previsão do ONS. In: *IX Simpósio Brasileiro de Sistemas Elétricos*. Santa Maria: [s.n.], 2022.

[39] NASCIMENTO, P. S. C.; MENDES, E. L.; MARCATO, A. L. M.; ARAUJO, L. F.; KHENAYFIS, L. S. Statistical model for treatment of missing and outliers in time series for wind power forecast. In: *Proceedings of the XIV Latin-American Congress on Electricity Generation and Transmission (CLAGTEE)*. Rio de Janeiro, Brazil: Even3 Publicações, 2022. Paper no. B-5.3-4. Disponível em: <https://www.feg.unesp.br/Home/Eventos/clagtee/topic-5---energy-planning-and-management.zip>. Acesso em 20 maio 2025.

[40] NASCIMENTO, P. S. C.; ARAUJO, L. F.; MADEIRA, F.; KHENAYFIS, L. S.; AZAMBUJA, R. R. Desenvolvimento de um modelo de previsão eólica semanal para planejamento de curto prazo. In: *Brazil Windpower*. São Paulo: [s.n.], 2021.

[41] ARAUJO, L. F.; NASCIMENTO, P. S. C.; MADEIRA, F.; BARROS, A. M. L.; ALVES, R. J. M. Modelo de estimação de geração eólica a partir de dados de vento observado e curvas vento x potência. In: *Brazil Windpower*. São Paulo: [s.n.], 2021.

[42] NASCIMENTO, P. S. C.; FRANCA, A. C. G.; KHENAYFIS, L. S.; DUARTE, V. S.; JUNIOR, H. C. Desenvolvimento e análise de modelos clássicos e de machine learning na previsão de geração de energia eólica. In: *Brazil Windpower*. São Paulo: [s.n.], 2020.

[43] ONS. *Plano da Operação Energética*. Operador Nacional do Sistema Elétrico, 2023. Acesso em 30 março 2024. Disponível em: <https://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/ONS_PEN_2023.pdf>.

[44] ONS. *Relatório de Análise de Perturbação*. Operador Nacional do Sistema Elétrico, 2023. Acesso em 30 março 2024. Disponível em: <https://www.ons.org.br/AcervoDigitalDocumentosEPublicacoes/RAP%202023.08.15%2008h030min%20vers%C3%A3o%20final.pdf>.

[45] EPE. *Roadmap Eólica Offshore Brasil*. Empresa de Pesquisa Energética, 2020. Acesso em 4 julho 2025. Disponível em: <https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-456/Roadmap_Eolica_Offshore_EPE_versao_R2.pdf>.

[46] Concept. *Intermittent generation forecasting arrangements – review of international jurisdictions*. Concept Consulting Group Ltd, 2023.

[47] USAID. Grid integration series: Variable renewable energy forecasting. *United States Agency for International Development*, 2020.

[48] ESMAP. *Using Forecasting Systems to Reduce Cost and Improve Dispatch of Variable Renewable Energy*. ESMAP / The World Bank Group. Washington, DC, 2019. Disponível em: <http://dx.doi.org/10.1596/32072>.

[49] AHLQVIST, V.; HOLMBERG, P.; TANGERÅS, T. A survey comparing centralized and decentralized electricity markets. *Energy Strategy Reviews*, Elsevier, v. 40, p. 100812, 2022.

[50] MITRA, I.; HEINEMANN, D.; RAMANAN, A.; KAUR, M.; SHARMA, S. K.; TRIPATHY, S. K.; ROY, A. Short-term PV power forecasting in India: recent developments and policy analysis. *International Journal of Energy and Environmental Engineering*, Springer, v. 13, n. 2, p. 515–540, 2022.

[51] THEMA. A study on balancing and redispatching strategies. *THEMA Consulting Group*, 2019.

[52] YE, X.; LU, Z.; QIAO, Y.; MIN, Y.; O'MALLEY, M. Identification and correction of outliers in wind farm time series power data. *IEEE Transactions on Power Systems*, v. 31, p. 4197–4205, 2016.

[53] ANDRADE, P. H. M. de; VILLANUEVA, J. M. M.; BRAZ, H. D. de M. An outliers processing module based on artificial intelligence for substations metering system. *IEEE Transactions on Power Systems*, v. 35, p. 3400–3409, 2020.

[54] HARVEY, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press, 1990.

[55] SILVA, H. F. da. *Um Sistema Integrado de Monitoração e Previsão de Carga Elétrica de Curto Prazo*. Tese (Doutorado) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2001.

[56] MENDES, E. L. *Tratamento de Dados Faltantes e Aberrantes em Séries Temporais de Carga Minuto a Minuto*. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.

[57] LYNCH, C.; O'MAHONY, M. J.; SCULLY, T. Simplified method to derive the Kalman filter covariance matrices to predict wind speeds from a NWP model. *Energy Procedia*, v. 62, p. 676–685, 2014.

[58] CAO, M.; QIU, Y.; FENG, Y.; WANG, H.; LI, D. Study of wind turbine fault diagnosis based on unscented Kalman filter and SCADA data. *Energies*, v. 9, p. 847, 2016.

[59] CATALINA, A.; ALAÍZ, C. M.; DORRONSORO, J. Combining numerical weather predictions and satellite data for PV energy nowcasting. *IEEE Transactions on Sustainable Energy*, 2019.

[60] GIRALDO, F. X.; RESTELLI, M.; LÄUTER, M. Semi-implicit formulations of the Navier–Stokes equations: Application to nonhydrostatic atmospheric modeling. *SIAM Journal on Scientific Computing*, p. 3394–3425, 2010.

[61] XU, W.; NING, L.; LUO, Y. Wind speed forecast based on post-processing of numerical weather predictions using a gradient boosting decision tree algorithm. *Atmosphere*, p. 738, 2020.

[62] SANDU, I.; BECHTOLD, P.; NUIJENS, L.; BELJAARS, A.; BROWN, A. *On the Causes of Systematic Forecast Biases in Near-Surface Wind Direction over the Oceans*. [S.l.], 2020. ECMWF Technical Memo.

[63] MARAUN, D. Bias correcting climate change simulations: A critical review. *Current Climate Change Reports*, p. 211–220, 2016.

[64] COSTOYA, X.; ROCHA, A.; CARVALHO, D. Using bias-correction to improve future projections of offshore wind energy resource: A case study on the Iberian peninsula. *Applied Energy*, 2020.

[65] MONAHAN, A. H. The Gaussian predictability of wind speeds. *Journal of Geophysical Research*, 2012.

[66] HOLTHUIJZEN, M. F.; BECKAGE, B.; CLEMINS, P. J.; HIGDON, D.; WINTER, J. M. Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile-mapping method. *Theoretical and Applied Climatology*, p. 863–882, 2022.

[67] MICHEL, C.; SORTEBERG, A. Future projections of EURO-CORDEX raw and bias-corrected daily maximum wind speeds over Scandinavia. *Journal of Geophysical Research: Atmospheres*, 2023.

[68] FANG, Y.; WU, Y.; WU, F.; YAN, Y.; LIU, Q.; LIU, N.; XIA, J. Short-term wind speed forecasting bias correction in the Hangzhou area of China based on a machine learning model. *Atmospheric and Oceanic Science Letters*, v. 16, n. 4, p. 100339, 2023.

[69] WILKS, D. S. *Statistical Methods in the Atmospheric Sciences*. 4. ed. Amsterdam: Elsevier, 2019.

[70] LI, D.; FENG, J.; XU, Z.; YIN, B.; SHI, H.; QI, J. Statistical bias correction for simulated wind speeds over CORDEX-East Asia. *Earth and Space Science*, 2019.

[71] NIU, D.; LIANG, Y.; HONG, W.-C. Wind speed forecasting based on EMD and GRNN optimized by FOA. *Energies*, v. 10, p. 2001, 2017.

[72] ALENCAR, D. B.; AFFONSO, C. M.; OLIVEIRA, R. C. L.; FILHO, J. C. R. Hybrid approach combining SARIMA and neural networks for multi-step ahead wind speed forecasting in Brazil. *IEEE Access*, p. 55986–55994, 2018.

[73] COSTA, A.; CRESPO, A.; NAVARRO, J.; LIZCANO, G.; MADSEN, H.; FEITOSA, E. A review on the young history of the wind power short-term prediction. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 12, n. 6, p. 1725–1744, 2008.

[74] FOLEY, A. M.; LEAHY, P. G.; MARVUGLIA, A.; MCKEOGH, E. J. Current methods and advances in forecasting of wind power generation. *Renewable energy*, Elsevier, v. 37, n. 1, p. 1–8, 2012.

[75] LEI, M.; SHIYAN, L.; CHUANWEN, J.; HONGLING, L.; YAN, Z. A review on the forecasting of wind speed and generated power. *Renewable and sustainable energy reviews*, Elsevier, v. 13, n. 4, p. 915–920, 2009.

[76] FRANÇA, C. G.; FACEROLI, S. T.; NASCIMENTO, P. S. C.; KLIGERMAN, A. S. Análise comparativa de modelos de previsão de geração de energia eólica baseados em machine learning. *Revista de Sistemas e Computação (RSC)*, v. 9, n. 2, p. 348–358, 2019. Disponível em: <https://revistas.unifacs.br/index.php/rsc/article/view/6432>.

[77] ECMWF. *IFS Documentation – Cy49r1*. European Centre for Medium-Range Weather Forecasts, 2023. Disponível em: <https://www.ecmwf.int/en/publications/ifs-documentation>.

[78] NCEP. *The Global Forecast System (GFS) Documentation*. National Centers for Environmental Prediction, 2025. Acesso em 6 agosto 2025. Disponível em: <https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php>.

[79] POWERS, J. G.; KLEMP, J. B.; SKAMAROCK, W. C.; DAVIS, C. A.; DUDHIA, J.; GILL, D. O.; COEN, J. L.; GOCHIS, D. J.; AHMADOV, R.; PECKHAM, S. E.; GRELL, G. A.; MICHALAKES, J.; TRAHAN, S.; BENJAMIN, S. G.; ALEXANDER, C. R.; DIMEGO, G. J.; WANG, W.; SCHWARTZ, C. S.; ROMINE, G. S.; LIU, Z.; SNYDER, C.; CHEN, F.; BARLAGE, M. J.; YU, W.; DUDA, M. G. The weather research and forecasting model: Overview, system efforts, and future directions. *Bulletin of the American Meteorological Society*, v. 98, n. 8, p. 1717–1737, 2017. DOI: 10.1175/bams-d-15-00308.1.

[80] WU, Y.-K.; SU, P.-E.; WU, T.-Y.; HONG, J.-S.; HASSAN, M. Y. Probabilistic wind-power forecasting using weather ensemble models. *IEEE Transactions on Industry Applications*, IEEE, v. 54, n. 6, p. 5609–5620, 2018.

[81] GIEBEL, G.; BROWNSWORD, R.; KARINIOTAKIS, G.; DENHARD, M.; DRAXL, C. *The State-Of-The-Art in Short-Term Prediction of Wind Power, A Literature Overview, 2nd Edition*. [S.l.], 2011. Disponível em: <https://hal.science/hal-03713576>.

[82] LANGE, M.; FOCKEN, U. *Physical Approach to Short-Term Wind Power Prediction*. Berlin, Heidelberg: Springer, 2006.

[83] SHEN, L.; YAN, B.; CHOW, F. K. High-fidelity simulations and field measurements for characterizing wind fields in a utility-scale wind farm. *Applied Energy*, Elsevier, v. 278, p. 115733, 2020. DOI: 10.1016/j.apenergy.2020.115733.

[84] YATIYANA, E.; RAJAKARUNA, S.; GHOSH, A. Wind speed and direction forecasting for wind power generation using ARIMA model. In: *Australasian Universities Power Engineering Conference (AUPEC)*. [S.l.]: IEEE, 2017. p. 1–6.

[85] BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time Series Analysis: Forecasting and Control*. [S.l.]: Wiley, 2015.

[86] ABDEL-NASSER, M.; MAHMOUD, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing and Applications*, p. 2727–2740, 2019.

[87] GENSLER, A.; HENZE, J.; SICK, B.; RAABE, N. Deep learning for solar power forecasting: an approach using AutoEncoder and LSTM neural networks. In: *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.]: IEEE, 2016. p. 002858–002865.

[88] HE, J.; XU, J. Ultra-short-term wind speed forecasting based on support vector machine with combined kernel function and similar data. *EURASIP Journal on Wireless Communications and Networking*, Springer, v. 2019, n. 1, p. 248, 2019.

[89] WANG, J.; HU, J.; WANG, J.; CHEN, Z.; WANG, Y. Short-term wind power prediction based on extreme learning machine with error correction. *Protection and Control of Modern Power Systems*, Springer, v. 1, n. 1, p. 1–8, 2016.

[90] SHI, K.; QIAO, Y.; ZHAO, W.; WANG, Q.; LIU, M.; LU, Z. An improved random forest model of short-term wind-power forecasting to enhance accuracy, efficiency, and robustness. *Wind energy*, Wiley Online Library, v. 21, n. 12, p. 1383–1394, 2018.

[91] WANG, H.-Z.; LI, G.-Q.; WANG, G.-B.; PENG, J.-C.; JIANG, H.; LIU, Y.-T. Deep learning based ensemble approach for probabilistic wind power forecasting. *Applied Energy*, p. 56–70, 2017.

[92] SOMAN, S.; ZAREIPOUR, H.; MALIK, O. A review of wind power and wind speed forecasting methods with different time horizons. *IEEE PES General Meeting*, p. 1–8, 2010.

[93] ZHANG, Y.; WANG, J.; WANG, X. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, Elsevier, v. 32, p. 255–270, 2014.

[94] WAN, C.; XU, Z.; PINSON, P.; DONG, Z. Y.; WONG, K. P. Probabilistic forecasting of wind power generation using extreme learning machine. *IEEE transactions on power systems*, IEEE, v. 29, n. 3, p. 1033–1044, 2013.

[95] PINSON, P. Wind energy: Forecasting challenges for its operational management. *Statistical Science*, v. 28, n. 4, p. 564–585, 2013.

[96] GONZÁLEZ, G.; DÍAZ-GUERRA, B.; SOTO, F.; LÓPEZ, S.; SÁNCHEZ, I.; USAOLA, J.; ALONSO, M.; LOBO, M. G. SIPREÓLICO - wind power prediction tool for the Spanish Peninsular power system. In: *Proceedings of the CIGRÉ 40th General Session*. Paris, France: [s.n.], 2004. p. 1–8. Session Paper C2-101.

[97] LANDBERG, L.; WATSON, S. J. Short-term prediction of local wind conditions. *Boundary-Layer Meteorology*, Springer, v. 70, n. 1, p. 171–195, 1994.

[98] FOCKEN, U.; LANGE, M.; WALDL, H.-P. Previento-a wind power prediction system with an innovative upscaling algorithm. In: *Proceedings of the European Wind Energy Conference, Copenhagen, Denmark*. [S.l.: s.n.], 2001. v. 276, p. 826–829.

[99] JØRGENSEN, J. U.; MOEHRLEN, C.; GALLAGHOIR, B.; SATTLER, K.; MCKEOGH, E. HIRPOM: Description of an operational numerical wind power prediction model for large scale integration of on-and offshore wind power in Denmark. In: *Poster on the Global Windpower Conference and Exhibition, Paris, France*. [S.l.: s.n.], 2002. p. 1–5.

[100] LANDBERG, L. Short-term prediction of local wind conditions. *Journal of Wind Engineering and Industrial Aerodynamics*, Elsevier, v. 89, n. 3-4, p. 235–245, 2001.

[101] TSAI, W.-C.; HONG, C.-M.; TU, C.-S.; LIN, W.-M.; CHEN, C.-H. A review of modern wind power generation forecasting technologies. *Sustainability*, MDPI, v. 15, n. 14, p. 10757, 2023.

[102] GIEBEL, G.; LANDBERG, L.; NIELSEN, T. S.; MADSEN, H. The Zephyr project–the next generation prediction system. In: *Proc. of the 2001 European Wind Energy Conference, EWEC'01, Copenhagen, Denmark*. [S.l.: s.n.], 2001. p. 777–780.

[103] XU, Q.; HE, D.; ZHANG, N.; KANG, C.; XIA, Q.; BAI, J.; HUANG, J. A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining. *IEEE Transactions on sustainable energy*, IEEE, v. 6, n. 4, p. 1283–1291, 2015.

[104] KARINIOTAKIS, G.; MAYER, D.; MOUSSAFIR, J.; CHEVALLAZ-PERRIER, R.; USAOLA, J.; SANCHEZ, I.; MARTI, I.; MADSEN, H.; NIELSEN, T. S.; LAC, C.; FRAYSSINET, P.; WALDL, H.; HALLIDAY, J.; GIEBEL, G.; KALLOS, G.; OTTAVI, J.; FOCKEN, U.; LANGE, M.; HEINEMANN, D.; ANCIN, J. K.; TOEFTING, J.; O'DONNEL, P.; COY,

D. M.; COLLMANN, M.; GIGANDIDOU, A.; GONZALES-MORALES, G.; BARQUERO, C.; CRUZ, I.; HATZIARGYRIOU, N. D. Anemos: development of a next generation wind power forecasting system for the large-scale integration of onshore & offshore wind farms. In: *Proceedings EWEC 2003*. Madrid, Spain: [s.n.], 2003. p. CD–ROM.

[105] LANDBERG, L.; GIEBEL, G.; NIELSEN, H. A.; NIELSEN, T.; MADSEN, H. Short-term prediction: an overview. *Wind Energy*, Wiley Online Library, v. 6, n. 3, p. 273–280, 2003.

[106] COELHO, I. *Estudo sobre a Aplicação de Redes Neurais Recorrentes para Previsão da Geração Eólica e do Preço de Liquidação das Diferenças*. Dissertação (Mestrado) — Universidade Federal de Santa Catarina, 2018.

[107] SANTOS, P. H. C. d. *Previsão da Geração de Potência Eólica Utilizando Redes Neurais Artificiais*. Monografia (Bacharelado em Engenharia de Computação) — Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, 2014. Disponível em: <https://www.ecomp.poli.br/ListaTCC/20141/TCC%20-%20Pedro%20Henrique%20Carvalho%20dos%20Santos.pdf>.

[108] HIRUMA, F. S. *Previsão da Geração de Energia Eólica com Redes Neurais Artificiais*. Dissertação (Mestrado) — Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental, Curitiba, 2016.

[109] SOUZA, I. M. M. d. *Um Estudo Comparativo para Previsão da Cotação de Ações da BM&FBovespa Utilizando Redes Neurais Artificiais*. Monografia (Bacharelado em Engenharia de Computação) — Escola Politécnica de Pernambuco, Universidade de Pernambuco, Recife, 2012.

[110] ALENCAR, D. B. de. *Modelo Híbrido Baseado em Séries Temporais e Redes Neurais para Previsão da Geração de Energia Eólica*. Tese (Doutorado) — Universidade Federal do Pará (UFPA), Belém, 2018.

[111] WEIGEND, A. S. *Time Series Prediction: Forecasting the Future and Understanding the Past*. [S.l.]: Routledge, 2018.

[112] YASUDA, Y.; BIRD, L.; CARLINI, E. M.; ERIKSEN, P. B.; ESTANQUEIRO, A.; FLYNN, D.; VRANA, T. K. CE (curtailment–energy share) map: An objective and quantitative measure to evaluate wind and solar curtailment. *Renewable and Sustainable Energy Reviews*, v. 160, p. 112212, 2022.

[113] BIRD, L.; LEW, D.; MILLIGAN, M.; CARLINI, E. M.; ESTANQUEIRO, A.; FLYNN, D.; MILLER, J. Wind and solar energy curtailment: A review of international experience. *Renewable and Sustainable Energy Reviews*, v. 65, p. 577–586, 2016.

[114] BIRD, L.; COCHRAN, J.; WANG, X. *Wind and Solar Energy Curtailment: Experience and Practices in the United States*. [S.l.], 2014. Acesso em 5 fevereiro 2026. Disponível em: <http://www.nrel.gov/docs/fy14osti/60983.pdf>.

[115] JACOBSEN, H. K.; SCHRÖDER, S. T. Curtailment of renewable generation: Economic optimality and incentives. *Energy Policy*, v. 49, p. 663–675, 2012.

[116] JOOS, M.; STAFFELL, I. Short-term integration costs of variable renewable energy: Wind curtailment and balancing in Britain and Germany. *Renewable and Sustainable Energy Reviews*, v. 86, p. 45–65, 2018.

[117] KIM, D. Y.; KIM, B. S. Exploring wind curtailment effects and economic implications in the growing variable renewable energy penetration. *Journal of Renewable and Sustainable Energy*, v. 16, p. 023310, 2024.

[118] MORENO, B. O. Renewable compensation policies and conventional energy investment: A theoretical model. *Heliyon*, v. 10, p. e33971, 2024.

[119] VIEIRA, G.; SILVA, W.; SALLES, M.; LOURENÇO, L. Assessment of constrained-off in renewable energy integration in Brazil. In: *2025 International Conference on Clean Electrical Power*. Villasimius, Italy: IEEE, 2025. p. 121–127. Acesso em 19 março 2026. Disponível em: <http://dx.doi.org/10.1109/ICCEP65222.2025.11143733>.

[120] SILVA, W.; VIEIRA, G.; SIMONE, L.; LOURENÇO, L.; SALLES, M. Data-driven analysis of wind curtailment energy imbalance and distributed generation in Brazil. In: *2025 International Conference on Clean Electrical Power*. Villasimius, Italy: IEEE, 2025. p. 728–734. Acesso em 19 março 2026. Disponível em: <http://dx.doi.org/10.1109/ICCEP65222.2025.11143737>.

[121] WEN, Z.; BAI, Y.; WANG, Y. Research on automatic calculating methods of wind curtailment based on measured data. In: *The Proceedings of the 11th Frontier Academic Forum of Electrical Engineering (FAFEE2024)*. Singapore: Springer Nature, 2024. p. 107–115. https://doi.org/10.1007/978-981-97-8780-7_11.

[122] SIMANKOV, V.; BUCHATSKIY, P.; TEPLOUKHOV, S.; ONISHCHENKO, S.; KAZAK, A.; CHETYRBOK, P. Review of estimating and predicting models of the wind energy amount. *Energies*, v. 16, p. 5926, 2023. https://doi.org/10.3390/en16165926.

[123] REN, G.; WAN, J.; LIU, J.; YU, D. Spatial and temporal correlation analysis of wind power between different provinces in China. *Energy*, v. 191, p. 116514, 2020.

[124] LI, H. Short-term wind power prediction via spatial temporal analysis and deep residual networks. *Frontiers in Energy Research*, v. 10, p. 920407, 2022.

[125] WIDODO, D. A.; IKSAN, N. Machine learning-driven wind energy mapping enhanced by natural neighbor interpolation. *Journal of Energy Systems*, v. 8, p. 193–206, 2024.

[126] ZHAO, W.; ZHONG, Y.; LI, Q.; LI, M.; LIU, J.; TANG, L. Comparison and correction of IDW based wind speed interpolation methods in urbanized Shenzhen. *Frontiers in Earth Science*, v. 16, p. 798–808, 2022.

[127] CAI, C.; SHI, Q.; JIN, Y.; HUA, M.; TAO, Y.; HOU, S. Offshore wind power forecasting with wind-regime clustering and multi-scale feature learning. *International Journal of Electrical Power and Energy Systems*, v. 174, p. 111553, 2026.

[128] CELLURA, M.; CIRRINCIONE, G.; MARVUGLIA, A.; MIRAOUI, A. Wind speed spatial estimation for energy planning in Sicily: A neural kriging application. *Renewable Energy*, v. 33, p. 1251–1266, 2008.

[129] SOHONI, V.; GUPTA, S. C.; NEMA, R. K. A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems. *Journal of Energy*, v. 2016, p. 8519785, 2016.

[130] MUSHTAQ, K.; ZOU, R.; WARIS, A.; YANG, K.; WANG, J.; IQBAL, J.; JAMEEL, M. Multivariate wind power curve modeling using multivariate adaptive regression splines and regression trees. *PLoS ONE*, v. 18, p. e0290316, 2023.

[131] LIU, T.; LV, K.; CHEN, F.; GOH, H. H.; KURNIAWAN, T. A.; HU, R.; JIANG, M.; ZHANG, D. Wind power curve model combining smoothed spline with first-order moments and density-adjusted wind speed strategy. *Energy*, v. 313, p. 133628, 2024.

[132] LI, L.-L.; ZHAO, X.; TSENG, M.-L.; TAN, R. R. Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm. *Journal of Cleaner Production*, v. 242, p. 118447, 2020.

[133] SINGH, U.; RIZWAN, M.; ALARAJ, M.; ALSAIDAN, I. A machine learning-based gradient boosting regression approach for wind power production forecasting. *Energies*, v. 14, p. 5196, 2021.

[134] ZHENG, H.; WU, Y. A XGBoost model with weather similarity analysis and feature engineering for short-term wind power forecasting. *Applied Sciences*, v. 9, p. 3019, 2019.

[135] ONS. *Procedimentos de Rede*. Operador Nacional do Sistema Elétrico, 2024. Acesso em 25 março 2025. Disponível em: <https://www.ons.org.br/paginas/sobre-o-ons/procedimentos-de-rede/vigentes>.

[136] ANEEL. *Resolução Normativa ANEEL Nº 1.032*. Agência Nacional de Energia Elétrica, 2022. Acesso em 25 março 2024. Disponível em: <https://in.gov.br/en/web/dou/-/resolucao-normativa-aneel-n-1.032-de-26-de-julho-de-2022-419531086>.

[137] IUNG, A. M.; OLIVEIRA, F. L. C.; MARCATO, A. L.; PEREIRA, G. A. Evaluating the potential of copulas for modeling correlated scenarios for hydro, wind, and solar energy. *Forecasting*, MDPI, v. 7, n. 1, p. 7, 2025.

[138] IUNG, A. M.; OLIVEIRA, F. L. C.; MARCATO, A. L. M. A review on modeling variable renewable energy: Complementarity and spatial–temporal dependence. *Energies*, MDPI, v. 16, n. 3, p. 1013, 2023.

[139] HODGE, B.-M.; LEW, D.; MILLIGAN, M.; HOLTTINEN, H.; SILLANPÄÄ, S.; GÓMEZ-LÁZARO, E.; SCHARFF, R.; SÖDER, L.; LARSÉN, X. G.; GIEBEL, G.; FLYNN, D.; DOBSCHINSKI, J. Wind power forecasting error distributions: An international comparison. In: *11th International Workshop on Large-Scale Integration of Wind Power into Power Systems*. Lisbon, Portugal: [s.n.], 2012. p. 1–7. NREL/CP-5500-56130.

[140] ROGUS, R.; CASTRO, R.; SOŁTYSIK, M. Comparative analysis of wind energy generation forecasts in Poland and Portugal and their influence on the electricity exchange prices. *Inventions*, MDPI, v. 5, n. 3, p. 35, 2020.

[141] GODINHO, M.; CASTRO, R. Comparative performance of AI methods for wind power forecast in Portugal. *Wind Energy*, Wiley Online Library, v. 24, n. 1, p. 39–53, 2021.

[142] BATES, J. M.; GRANGER, C. W. J. The combination of forecasts. *Operational Research Quarterly*, v. 20, n. 4, p. 451–468, 1969.

[143] TIMMERMANN, A. Forecast combinations. In: ELLIOTT, G.; GRANGER, C. W. J.; TIMMERMANN, A. (Ed.). *Handbook of Economic Forecasting*. Amsterdam: Elsevier, 2006. v. 1, cap. 4, p. 135–196.

[144] CHOU, S. C.; SOUZA, C. R.; GOMES, J. L.; EVANGELISTA, E. F. D.; OSÓRIO, C.; CATALDI, M. Refinamento estatístico das previsões horárias de temperatura a 2 m do modelo Eta em estações do Nordeste do Brasil. *Revista Brasileira de Meteorologia*, v. 22, p. 287–296, 2007.

[145] ROLO, M. N. M. *Previsão de Produção Eólica com Modelização de Incertezas*. Dissertação (Mestrado) — Faculdade de Engenharia da Universidade do Porto, Porto, Portugal, 2014. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/75803/2/31592.pdf>. Acesso em 20 maio 2025.

[146] ZHAO, Y.; YE, L.; WANG, W.; SUN, H.; JU, Y.; TANG, Y. Data-driven correction approach to refine power curve of wind farm under wind curtailment. *IEEE Transactions on Sustainable Energy*, v. 9, p. 95–105, 2018.

[147] ONS. *Dados de Restrição de Operação por Constrained-off de Usinas Eólicas - Detalhamento por Usinas*. Operador Nacional do Sistema Elétrico, 2026. Acesso em 19 março 2026. Disponível em: <https://dados.ons.org.br/dataset/restricao_coff_eolica_detail>.

[148] CCEE. *Geração*. Câmara de Comercialização de Energia Elétrica, 2026. Acesso em 19 março 2026. Disponível em: <https://www.ccee.org.br/dados-e-analises/dados-geracao>.

[149] ANEEL. *SIGA - Sistema de Informações de Geração da ANEEL*. Agência Nacional de Energia Elétrica, 2026. Acesso em 19 março 2026. Disponível em: <https://dadosabertos.aneel.gov.br/dataset/siga-sistema-de-informacoes-de-geracao-da-aneel>.

[150] CPTEC/INPE. *Sistema de Transferência de Dados*. Centro de Previsão de Tempo e Estudos Climáticos, 2026. Acesso em 19 março 2026. Disponível em: <https://dataserver.cptec.inpe.br/dataserver_modelos/>.

[151] DURBIN, J.; KOOPMAN, S. J. A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, v. 89, p. 603–616, 2002.

[152] BISHOP, C. M. *Pattern Recognition and Machine Learning*. [S.l.]: Springer, 2006. Corr. 2nd Printing 2011.

[153] CLEVELAND, W. S. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, v. 74, n. 368, p. 829–836, 1979.

[154] ANEEL. *Resolução Normativa ANEEL Nº 1.073, de 12 de setembro de 2023*. Agência Nacional de Energia Elétrica, 2023. Acesso em 20 março 2024. Disponível em: <https://www2.aneel.gov.br/cedoc/ren20231073.html>.

[155] ANEEL. *Resolução Normativa da ANEEL Nº 1.030*. Agência Nacional de Energia Elétrica, 2022. Acesso em 19 abril 2024. Disponível em: <https://www2.aneel.gov.br/cedoc/ren20221030.pdf>.

[156] World Meteorological Organization. *WMO Guidelines on the Calculation of Climate Normals*. Geneva, 2017.

[157] MAO, Q.; MCNIDER, R. T.; MUELLER, S. F.; JUANG, H. An optimal model output calibration algorithm suitable for objective temperature forecasting. *Weather and Forecasting*, v. 14, p. 190–202, 1999.

[158] GENRE, V.; KENNY, G.; MEYLER, A.; TIMMERMANN, A. Combining expert forecasts: Can anything beat the simple average? *International Journal of Forecasting*, v. 29, n. 1, p. 108–121, 2013.

[159] CEPEL. *Atualização dos Intervalos de Duração dos Patamares de Carga para a Operação Energética 2023–2027*. Centro de Pesquisas de Energia Elétrica (CEPEL), 2022.

[160] JENSEN, J. L. W. V. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, v. 30, p. 175–193, 1906. DOI: 10.1007/BF02418571.

[161] NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, v. 10, n. 3, p. 282–290, 1970. DOI: 10.1016/0022-1694(70)90255-6.

[162] HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting*, v. 22, p. 679–688, 2006.

[163] BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001.

[164] CHEN, T.; GUESTRIN, C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. [S.l.: s.n.], 2016. p. 785–794.

[165] KALMAN, R. E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, v. 82, n. 1, p. 35–45, 1960.

[166] MESSNER, J. W.; PINSON, P.; BROWELL, J.; BJERREGÅRD, M. B.; SCHICKER, I. Evaluation of wind power forecasts: An up-to-date view. *Wind Energy*, v. 23, p. 1461–1481, 2020.

[167] GONZÁLEZ-SOPEÑA, J. M.; PAKRASHI, V.; GHOSH, B. An overview of performance evaluation metrics for short-term statistical wind power forecasting. *Renewable and Sustainable Energy Reviews*, v. 138, p. 110515, 2021.

[168] ONS. *NT 0151-2018: Desenvolvimento Metodológico para Previsão de Geração de Fonte Eólica*. Operador Nacional do Sistema Elétrico, 2018. Acesso em 6 julho 2025. Disponível em: <https://sintegre.ons.org.br/sites/8/103/105/Produtos/439>.

[169] PINTO, L.; LIMA, F.; MARTINS, F.; PEREIRA, E. Análise de agrupamento na otimização de futuras aplicações de modelagem numérica na estimativa e previsão de recurso eólico. *Revista Brasileira de Geografia Física*, v. 10, n. 6, p. 1698–1711, 2017. DOI: 10.26848/rbgf.v10.6.p1698-1711.

[170] GURGEL, A. d. R. C.; SALES, D. C.; LIMA, K. C. Wind power density in areas of Northeastern Brazil from regional climate models for a recent past. *PLOS ONE*, v. 19, n. 7, p. e0307641, 2024. DOI: 10.1371/journal.pone.0307641.

APÊNDICE A – DETALHAMENTO DOS MODELOS DE TEMPO REAL

Este apêndice apresenta o equacionamento completo dos modelos que compõem o *ensemble* de previsão em tempo real, descritos de forma resumida na Seção 5.10 do Capítulo 5, bem como das estratégias de combinação. Mantém-se a notação adotada ao longo do trabalho, com as seguintes convenções específicas deste apêndice:

- t denota o instante atual (instante de emissão da previsão) e k o passo de previsão à frente, em passos de 30 minutos; o horizonte total é $H = 48$ passos (24 horas). Por conveniência, define-se $h_k = k/2$ como o horizonte correspondente, em horas.
- $G_{v,t}$ é a geração verificada no instante t; $G_{p,t+k}$ é a geração prevista para o instante $t+k$.
- $G_{\mathrm{NWP},t+k}$ é a previsão do NWP combinado e $G_{\mathrm{ECMWF},t+k}$ a previsão do modelo ECMWF, ambas para o horizonte k. Tal como no restante do trabalho, as menções ao ECMWF referem-se ao produto *Ensemble* (ECMWF2).
- $\mathrm{clamp}(x) = \min(\max(x, 0), G_{\max})$ é o operador de saturação, que limita a previsão ao intervalo fisicamente admissível entre zero e a geração máxima histórica do agrupamento, $G_{\max}$. Por clareza, esse operador é omitido nas equações intermediárias e aplicado apenas à previsão final de cada modelo.

Quanto à fonte de dados meteorológicos, os modelos repartem-se em dois grupos. As persistências utilizam o NWP combinado como destino de convergência, por ser a melhor estimativa pronta na janela de simulação. Já os modelos estatísticos baseados em resíduo (ARIMA e suavização exponencial) e os de aprendizado de máquina utilizam o modelo ECMWF como fonte fixa. Há duas razões para isso. A primeira é prática: o treinamento desses modelos exige um histórico longo, ao passo que o NWP combinado só existe para a janela de simulação. A segunda, e mais importante do ponto de vista metodológico, é que o NWP combinado é uma série temporal *mutante no tempo*: seus pesos são recalculados continuamente conforme o desempenho recente de cada modelo meteorológico, de modo que a mesma entrada teria significado estatístico distinto em momentos diferentes. Isso comprometeria o ajuste dos modelos (sobretudo os estatísticos, que pressupõem uma série de erro estacionária) e violaria o princípio de que treino e previsão devem compartilhar a mesma distribuição de entrada. O ECMWF bruto, por ser uma fonte estável e disponível em todo o histórico, preserva essa coerência.

A Tabela 14 resume a janela de treinamento e os principais hiperparâmetros dos modelos calibráveis, todos reestimados semanalmente conforme a Seção 5.10.2. Os valores foram definidos de forma empírica, em testes com diferentes valores em alguns agrupamentos e períodos, e mantidos fixos para todos os agrupamentos.

Tabela 14 – Janela de treinamento e principais hiperparâmetros dos modelos calibráveis.

Modelo	**Janela**	**Principais hiperparâmetros**
Regressão linear	365 dias	Mínimos quadrados; 8 preditores e intercepto; uma regressão por horizonte e meia-hora do dia
ARIMA	30 dias	Seleção automática de ordem; $d = 0$; $p, q \leq 5$; estado atualizado a cada emissão com sete dias
Suavização exponencial	30 dias	$\alpha \in \{0{,}1; 0{,}2; 0{,}3; 0{,}4\}$ e $\phi \in \{0{,}80; 0{,}85; 0{,}90; 0{,}93\}$ por busca em grade
Random Forest	365 dias	100 árvores; tamanho mínimo de nó igual a 5
XGBoost (por horizonte)	365 dias	Erro quadrático; 100 rodadas; taxa de aprendizado de 0,1; profundidade máxima 4
XGBoost (*multi-output*)	365 dias	Erro quadrático; 200 rodadas; taxa de aprendizado de 0,05; profundidade máxima 3
MLP	365 dias	8 neurônios na camada oculta; decaimento de pesos de 0,01; número máximo de 200 iterações; padronização *z-score*
SARIMA	30 dias	$s = 48$; $d = 0$ e $D = 1$; $p, q \leq 3$; $P, Q \leq 1$; estado atualizado a cada emissão com 14 dias

Fonte: Elaboração própria.

A.1 Persistência pura

A persistência pura constitui o modelo de referência (*baseline*) mais simples, amplamente adotado na literatura de previsão eólica como referência de comparação [81, 92], assumindo que a condição física momentânea do vento se mantém ao longo de todo o horizonte. Assim, projeta-se o último valor de geração verificada para todos os passos:

$$G_{p,t+k} = G_{v,t}, \qquad k = 1, \ldots, H \tag{A.1}$$

Esse modelo é preciso para o primeiro passo, mas degrada-se rapidamente à medida que o horizonte aumenta, em razão da alta variabilidade da geração eólica. Por exemplo, se a geração verificada atual é de 40 MW, o modelo prevê 40 MW para todos os 48 passos das próximas 24 horas, ignorando qualquer informação meteorológica. Por isso, serve como limite inferior de desempenho.

A.2 Persistência inteligente

A persistência inteligente realiza uma transição suave entre dois regimes: no curtíssimo prazo, confia no estado atual (geração recente, com correção de viés e dinâmica autorregressiva); no longo prazo, converge para o NWP combinado, que detém habilidade preditiva meteorológica. Essa transição é controlada por decaimentos exponenciais com constantes de tempo distintas. Define-se o fator de decaimento associado a uma constante de tempo ε (em horas) por:

$$\alpha_k^{(\varepsilon)} = e^{-h_k/\varepsilon}, \qquad h_k = k/2 \tag{A.2}$$

Esse fator vale aproximadamente 1 para horizontes muito menores que ε (predomínio do termo associado) e tende a zero para horizontes muito maiores. A constante ε é, portanto, o horizonte (em horas) no qual o termo cai a $1/e \approx 37\%$ do seu valor inicial. Adotam-se quatro constantes de

tempo, calibradas empiricamente: $\varepsilon_L = 20\,\text{h}$ (correção de viés do NWP), $\varepsilon_{MB} = 7\,\text{h}$ (ancoragem no valor atual), $\varepsilon_{AR} = 0{,}75\,\text{h}$ (peso da componente autorregressiva) e $\varepsilon_{AJ} = 0{,}4\,\text{h}$ (correção fina ancorada no verificado, atuante apenas nos primeiros passos). O modelo é construído em quatro etapas encadeadas, sendo cada constante o único parâmetro que controla a sua etapa.

A.2.1 Etapa 1, correção de viés de longo prazo

Estima-se o viés recente do NWP como o erro médio real observado no mesmo trecho passado, isto é, a média da diferença entre a geração verificada e a previsão que o NWP combinado havia feito para aqueles mesmos instantes. Sejam $G_{\text{NWP},t-j}$ a previsão histórica do NWP combinado (da mesma rodada) para o instante passado $t-j$ e n_b o número de meias-horas consideradas ($n_b = 12$, correspondentes a 6 horas):

$$\delta_1 = \frac{1}{n_b} \sum_{j=0}^{n_b-1} \left(G_{v,t-j} - G_{\text{NWP},t-j} \right) \tag{A.3}$$

Ou seja, δ_1 mede o quanto, em média, o NWP combinado vinha subestimando ($\delta_1 > 0$) ou superestimando ($\delta_1 < 0$) a geração nas últimas 6 horas. Caso não haja a previsão histórica do combinado disponível, adota-se $\delta_1 = 0$ (sem correção, de forma conservadora). Esse viés é então aplicado à previsão futura do NWP, decaindo lentamente (constante ε_L):

$$G^{(1)}_{p,t+k} = \alpha_k^{(\varepsilon_L)} \left(G_{\text{NWP},t+k} + \delta_1 \right) + \left(1 - \alpha_k^{(\varepsilon_L)} \right) G_{\text{NWP},t+k} = G_{\text{NWP},t+k} + \alpha_k^{(\varepsilon_L)}\, \delta_1 \tag{A.4}$$

A forma simplificada à direita evidencia o efeito: parte-se da previsão do NWP e adiciona-se a correção de viés, que decai lentamente com o horizonte. Como $\varepsilon_L = 20\,\text{h}$ é elevado, a correção persiste por muitas horas. Por exemplo, se nas últimas 6 horas o NWP subestimou a geração em média 5 MW ($\delta_1 = 5$ MW), em $h_k = 2\,\text{h}$ a correção ainda vale $\alpha_4^{(20)} \cdot 5 \approx 0{,}90 \cdot 5 = 4{,}5$ MW; em $h_k = 20\,\text{h}$, cai para $1/e$ desse valor.

A.2.2 Etapa 2, ancoragem no valor atual

Corrige-se a diferença entre o valor atual $G_{v,t}$ e a previsão do primeiro passo da etapa anterior, com decaimento mais rápido (constante ε_{MB}). Para evitar instabilidade em regiões de geração muito baixa, o fator só é aplicado quando ambos os valores superam 1 MW:

$$\delta_2 = \begin{cases} G_{v,t} - G^{(1)}_{p,t+1}, & \text{se } G_{v,t} > 1 \text{ e } G^{(1)}_{p,t+1} > 1 \\ 0, & \text{caso contrário} \end{cases} \tag{A.5}$$

$$G^{(2)}_{p,t+k} = \alpha_k^{(\varepsilon_{MB})} \left(G^{(1)}_{p,t+k} + \delta_2 \right) + \left(1 - \alpha_k^{(\varepsilon_{MB})} \right) G_{\text{NWP},t+k} \tag{A.6}$$

No primeiro passo, $G^{(2)}_{p,t+1} \approx G_{v,t}$ (a previsão fica ancorada no presente); conforme k cresce, $\alpha_k^{(\varepsilon_{MB})}$ tende a zero e a previsão converge para G_{NWP}.

A.2.3 Etapa 3, componente autorregressiva

Captura-se a dinâmica de curtíssimo prazo por meio de um modelo autorregressivo de primeira ordem, condicionado à hora do dia. Para cada passo $i = 1, \ldots, n_{pa}$ (com $n_{pa} = 8$, isto é, até 4 horas), a geração é propagada recursivamente:

$$G_i^{\mathrm{AR}} = b_0^{(i)} + b_1^{(i)}\, G_{i-1}^{\mathrm{AR}}, \qquad G_0^{\mathrm{AR}} = G_{v,t} \tag{A.7}$$

Os coeficientes $\left(b_0^{(i)}, b_1^{(i)}\right)$ são obtidos por regressões lineares pré-calculadas, uma para cada combinação de posição na meia-hora do dia (*offset* o) e passo i. Cada regressão é ajustada sobre os pares de valores verificados separados por i passos, observados sempre no mesmo horário do dia (espaçamento de 48 passos):

$$G_{v,t+i} = b_0^{(i)} + b_1^{(i)}\, G_{v,t} + \varepsilon \tag{A.8}$$

Em forma matricial, para um dado par (o, i), sejam $\mathbf{y}$ o vetor dos valores verificados em $t+i$ e $\mathbf{X}$ a matriz de regressão contendo uma coluna de uns e os valores verificados em t, ao longo de todas as ocorrências históricas daquele horário:

$$\mathbf{y} = \mathbf{X}\,\mathbf{b}^{(i)} + \boldsymbol{\varepsilon}, \qquad \mathbf{X} = \begin{bmatrix} 1 & G_{v,t_1} \\ 1 & G_{v,t_2} \\ \vdots & \vdots \\ 1 & G_{v,t_m} \end{bmatrix}, \quad \mathbf{b}^{(i)} = \begin{bmatrix} b_0^{(i)} \\ b_1^{(i)} \end{bmatrix} \tag{A.9}$$

cuja solução por mínimos quadrados é $\mathbf{b}^{(i)} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1} \mathbf{X}^\top \mathbf{y}$. Impõe-se a restrição $\left|b_1^{(i)}\right| \le 1$: caso a estimativa exceda esse limite, fixa-se $b_1^{(i)}$ no intervalo $[-1,\, 1]$ e recalcula-se $b_0^{(i)} = \overline{\mathbf{y}} - b_1^{(i)}\, \overline{G_{v,t}}$. Essa restrição preserva a correlação (inclusive negativa, associada ao ciclo diurno), mas impede o crescimento explosivo da previsão a cada passo recursivo.

Para os passos além de n_{pa}, a componente autorregressiva é estendida mantendo-se o último valor calculado:

$$G_k^{\mathrm{AR,ext}} = \begin{cases} G_k^{\mathrm{AR}}, & k \le n_{pa} \\ G_{n_{pa}}^{\mathrm{AR}}, & k > n_{pa} \end{cases} \tag{A.10}$$

A transição entre a componente autorregressiva e a previsão da Etapa 2 é feita de forma contínua, com peso que decai desde o primeiro passo (constante ε_{AR}), evitando descontinuidades:

$$G_{p,t+k}^{(3)} = \alpha_k^{(\varepsilon_{AR})}\, G_k^{\mathrm{AR,ext}} + \left(1 - \alpha_k^{(\varepsilon_{AR})}\right) G_{p,t+k}^{(2)} \tag{A.11}$$

Como $\varepsilon_{AR} = 0{,}75\,\mathrm{h}$ é pequeno, o peso da componente autorregressiva decai rapidamente: em $h_k = 0{,}5\,\mathrm{h}$ ainda vale aproximadamente 0,51; em $h_k = 1\,\mathrm{h}$, cai para 0,26; em $h_k = 3\,\mathrm{h}$, já é praticamente nulo ($\approx 0{,}02$). Assim, a componente AR tem influência apenas nos primeiros passos, e a partir de aproximadamente 3 horas a Etapa 3 é dominada pela previsão da Etapa 2. Essa é a razão pela qual a constante ε_{AR} é o único parâmetro que controla a intensidade e o alcance temporal da componente autorregressiva.

A.2.4 Etapa 4, correção fina ancorada no verificado atual

A última etapa aplica uma correção fina ancorada no valor verificado atual $G_{v,t}$, atuando apenas nos primeiros passos e desaparecendo rapidamente com o horizonte (constante ε_{AJ}). O resultado é diretamente a previsão final:

$$G_{p,t+k} = \text{clamp}\left[\alpha_k^{(\varepsilon_{AJ})}\, G_{v,t} + \left(1 - \alpha_k^{(\varepsilon_{AJ})}\right) G_{p,t+k}^{(3)}\right] \tag{A.12}$$

Com $\varepsilon_{AJ} = 0{,}4\,\text{h}$, o peso do verificado atual em $h_k = 0{,}5\,\text{h}$ vale $\alpha_1^{(0,4)} \approx 0{,}29$; em $h_k = 1\,\text{h}$, cai para $\approx 0{,}08$; em $h_k = 2\,\text{h}$, é praticamente nulo. Portanto, a Etapa 4 atua como um ajuste fino que ancora a previsão no valor observado no instante da emissão, sem influenciar a evolução do modelo além do curtíssimo prazo. A partir daí, a previsão fica inteiramente determinada por $G_{p,t+k}^{(3)}$, que por sua vez tende a $G_{p,t+k}^{(2)}$ (via decaimento de $\alpha_k^{(\varepsilon_{AR})}$) e, em horizonte longo, a $G_{\text{NWP},t+k}$ (via decaimento de $\alpha_k^{(\varepsilon_{MB})}$ e $\alpha_k^{(\varepsilon_L)}$).

Um aspecto importante do desenho é a separação entre as Etapas 3 e 4, cada uma controlada por uma constante de tempo própria e com papel não redundante: ε_{AR} controla exclusivamente o peso da componente autorregressiva e ε_{AJ} controla exclusivamente a correção fina ancorada no verificado atual. Essa separação evita a dupla modulação da componente autorregressiva, que ocorreria caso um mesmo mecanismo atuasse simultaneamente sobre o peso do termo AR e sobre a mesclagem final. Uma dupla modulação desse tipo tenderia a produzir um pico local de erro em torno do horizonte de 4 horas (um "dente" na curva de MAE), na região em que nem o valor observado atual nem o NWP dominam, decorrente do viés de reversão à média introduzido pela componente autorregressiva quando o horizonte se estende além do intervalo em que ela tem correlação útil. Com as constantes desacopladas, esse efeito é evitado e a persistência inteligente converge de forma monotônica ao NWP combinado ao longo do horizonte.

A Tabela 15 resume o regime dominante em cada faixa de horizonte.

Tabela 15 – Regime dominante da persistência inteligente por faixa de horizonte.

Horizonte	**Termo dominante**	**Constante**
0–1 h	Valor atual e componente autorregressiva	$\varepsilon_{AJ} = 0{,}4$; $\varepsilon_{AR} = 0{,}75$
1–7 h	Ancoragem no valor atual corrigido	$\varepsilon_{MB} = 7$
7–20 h	Correção de viés do NWP	$\varepsilon_L = 20$
> 20 h	NWP combinado puro	—

Fonte: Elaboração própria.

A.3 Persistência com derivada

Esta variante incorpora a tendência recente da geração. Estima-se a tendência τ_t (em MW por passo) como média ponderada das últimas n variações da geração verificada, com pesos

ρ_j crescentes para as observações mais recentes:

$$\tau_t = \sum_{j=1}^{n} \rho_j \left(G_{v,t-j+1} - G_{v,t-j} \right), \qquad \rho_j = \frac{e^{j/n}}{\sum_{l=1}^{n} e^{l/n}} \tag{A.13}$$

A tendência é projetada com decaimento ao longo do horizonte (constante ε_d) e o resultado é mesclado à previsão do NWP combinado:

$$G^{\text{tend}}_{t+k} = G_{v,t} + \tau_t \, k \, e^{-h_k/\varepsilon_d} \tag{A.14}$$

$$G_{p,t+k} = \text{clamp}\left[e^{-h_k/(3\varepsilon_d)} \, G^{\text{tend}}_{t+k} + \left(1 - e^{-h_k/(3\varepsilon_d)} \right) G_{\text{NWP},t+k} \right] \tag{A.15}$$

Por exemplo, se a geração vem subindo de forma consistente ($\tau_t > 0$), o modelo projeta essa subida nos primeiros passos, atenuando-a progressivamente até convergir para a previsão meteorológica.

A.4 Regressão linear

Para cada horizonte k, ajusta-se um modelo de regressão linear que relaciona a geração futura a um conjunto de variáveis explicativas. O modelo é estimado por mínimos quadrados sobre uma janela de 365 dias, com coeficientes específicos para cada par (*offset*, k):

$$\begin{aligned} G_{p,t+k} = {} & \beta_0 + \beta_1 G_{v,t} + \beta_2 G_{v,t-1} + \beta_3 G_{v,t-3} + \beta_4 \left(G_{v,t} - G_{v,t-1} \right) \\ & + \beta_5 G_{\text{ECMWF},t+k} + \beta_6 \left(G_{v,t} - G_{\text{ECMWF},t} \right) + \beta_7 \sin\frac{2\pi o}{48} + \beta_8 \cos\frac{2\pi o}{48} \end{aligned} \tag{A.16}$$

em que o é a posição na meia-hora do dia (*offset*). A equação reúne quatro tipos de componentes. Os termos $\beta_1 G_{v,t}$, $\beta_2 G_{v,t-1}$ e $\beta_3 G_{v,t-3}$ formam a parcela autorregressiva, que relaciona a geração futura aos valores recentes da própria série. $\beta_1 G_{v,t}$, em particular, representa a contribuição da geração atual. O termo $\beta_4 \left(G_{v,t} - G_{v,t-1} \right)$ é a componente de tendência (derivada de curto prazo), que capta se a geração vem subindo ou descendo. O termo $\beta_5 G_{\text{ECMWF},t+k}$ é a componente meteorológica, que injeta a previsão do NWP para o exato horizonte-alvo, enquanto $\beta_6 \left(G_{v,t} - G_{\text{ECMWF},t} \right)$ representa o erro pontual do NWP no instante atual, permitindo ao modelo aprender a corrigir vieses correntes da previsão meteorológica. Por fim, os termos $\beta_7 \sin(2\pi o/48)$ e $\beta_8 \cos(2\pi o/48)$ constituem a componente sazonal diária, que modela o ciclo característico de 24 horas. Dessa forma, o modelo combina, de maneira linear, a inércia recente da série, sua tendência, a informação meteorológica do horizonte previsto e o padrão diário.

A.5 ARIMA sobre o resíduo do NWP

O modelo ARIMA é ajustado não à geração diretamente, mas à série de erro do modelo ECMWF, definida por:

$$\eta_t = G_{v,t} - G_{\text{ECMWF},t} \tag{A.17}$$

Modelar o erro tem a vantagem de produzir uma série estacionária (média estável e autocorrelação que decai), em contraste com a geração, fortemente não estacionária. De modo geral, o modelo ARIMA(p,d,q) [84, 85] descreve a série de erro pela forma compacta com operador de retardo:

$$\phi_p(B)\,(1-B)^d\,\eta_t = \theta_q(B)\,a_t \tag{A.18}$$

em que B é o operador de retardo ($B\,\eta_t = \eta_{t-1}$), $\phi_p(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ é o polinômio autorregressivo, $\theta_q(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ o polinômio de médias móveis e a_t um ruído branco.

Para facilitar a interpretação da notação polinomial, com $d = 0$ a Equação (A.18) pode ser escrita explicitamente como uma regressão do erro sobre seus próprios valores passados (termos autorregressivos) e sobre os ruídos passados (termos de média móvel):

$$\eta_t = \phi_1\,\eta_{t-1} + \phi_2\,\eta_{t-2} + \cdots + \phi_p\,\eta_{t-p} + a_t + \theta_1\,a_{t-1} + \cdots + \theta_q\,a_{t-q} \tag{A.19}$$

Por exemplo, um ARIMA$(1,0,0)$ reduz-se a $\eta_t = \phi_1\,\eta_{t-1} + a_t$: o erro atual é uma fração do erro anterior mais um ruído. Como $|\phi_1| < 1$, a previsão de k passos à frente, $\hat{\eta}_{t+k} = \phi_1^k\,\eta_t$, decai geometricamente para zero (a média do erro), o que caracteriza o comportamento reversível à média.

As ordens p e q são selecionadas automaticamente por minimização do critério de informação de Akaike corrigido (AICc) (com $p, q \le 5$). Um aspecto determinante é a imposição de diferenciação nula, $d = 0$. Com $d = 1$, a seleção automática trataria o erro como um passeio aleatório, projetando o último valor de forma constante; com $d = 0$, a série estacionária é tratada como reversível à média, de modo que a previsão do erro converge para a média do erro, corrigindo o viés do NWP. A previsão final recompõe a geração somando o erro previsto à previsão do ECMWF:

$$G_{p,t+k} = \mathrm{clamp}\big(G_{\mathrm{ECMWF},t+k} + \hat{\eta}_{t+k}\big) \tag{A.20}$$

O modelo é calibrado em janela de 30 dias; a cada emissão, os parâmetros são mantidos e apenas o estado interno é atualizado com os erros dos últimos sete dias.

A.6 Suavização exponencial com reversão à média

Pertencente à família de suavização exponencial (modelos de espaço de estados) [54, 162], este modelo é também aplicado à série de erro η_t do ECMWF. A suavização exponencial pura projeta o nível recente de forma constante e não reverte à média, inadequado para o erro do NWP, que é estacionário. Adota-se, portanto, uma variante com reversão explícita à média. O nível l_t é atualizado recursivamente por suavização exponencial:

$$l_t = \alpha\,\eta_t + (1-\alpha)\,l_{t-1} \tag{A.21}$$

e a previsão do erro reverte à média global μ do erro, com fator de amortecimento $\phi \in (0,1)$:

$$\hat{\eta}_{t+k} = \mu + \phi^{\,k}\,(l_t - \mu) \tag{A.22}$$

A previsão final é $G_{p,t+k} = \text{clamp}\big(G_{\text{ECMWF},t+k} + \hat{\eta}_{t+k}\big)$. Intuitivamente, l_t resume o nível recente do erro; a previsão parte desse nível e o puxa de volta para a média global μ a uma velocidade controlada por ϕ (quanto menor ϕ, mais rápida a reversão). Por exemplo, se o nível recente do erro é $l_t = 8$ MW, a média global é $\mu = 2$ MW e $\phi = 0{,}9$, então a previsão do erro em $k = 4$ passos é $\hat{\eta}_{t+4} = 2 + 0{,}9^4\,(8-2) \approx 2 + 0{,}656 \cdot 6 \approx 5{,}9$ MW. Os parâmetros (α, ϕ) são determinados por busca em grade ($\alpha \in \{0{,}1;\ 0{,}2;\ 0{,}3;\ 0{,}4\}$ e $\phi \in \{0{,}80;\ 0{,}85;\ 0{,}90;\ 0{,}93\}$), minimizando o erro absoluto médio de previsão do erro, em horizontes selecionados (1, 6, 12, 24 e 48 passos), sobre a própria janela de treinamento de 30 dias; a cada emissão, o nível é recalculado com os erros dos últimos sete dias. O termo de reversão à média é o ingrediente essencial para corrigir o erro do NWP nesta família de modelos.

A.7 *Random Forest* e *XGBoost*

Ambos os modelos aprendem uma relação não linear entre um vetor de variáveis explicativas (*features*) e a geração futura. O vetor de *features* de base, comum às duas variantes, é composto por 13 elementos, todos disponíveis no instante de emissão t, descritos na Tabela 16.

Tabela 16 – *Features* de base dos modelos de aprendizado de máquina (instante t).

#	*Feature*	Descrição
1	$G_{v,t}$	Geração verificada atual
2–7	$G_{v,t-1}, \ldots, G_{v,t-6}$	Seis defasagens (até 3 h atrás)
8	$G_{v,t} - G_{v,t-2}$	Tendência de 1 h
9	$G_{v,t} - G_{v,t-6}$	Tendência de 3 h
10	$G_{v,t} - G_{\text{ECMWF},t}$	Erro pontual do NWP no instante atual
11	$\sin(2\pi o/48)$	Sazonalidade diária (seno)
12	$\cos(2\pi o/48)$	Sazonalidade diária (cosseno)
13	$G_{v,t}^2$	Termo quadrático da geração atual

Fonte: Elaboração própria.

Na variante por horizonte (*per step*), um modelo distinto $\mathcal{F}_k$ é treinado para cada passo k. Ele recebe as 13 *features* de base, agrupadas no vetor $\mathbf{x}_t$, acrescidas de uma 14ª *feature* (a previsão do ECMWF para o exato horizonte-alvo, $G_{\text{ECMWF},t+k}$) e prevê a geração verificada nesse horizonte:

$$\widehat{G_{v,t+k}} = \mathcal{F}_k\big(\mathbf{x}_t,\ G_{\text{ECMWF},t+k}\big) \ \Rightarrow\ G_{p,t+k} = \text{clamp}\big(\widehat{G_{v,t+k}}\big) \tag{A.23}$$

Em outras palavras, o alvo (*target*) do modelo do passo k é a geração verificada $G_{v,t+k}$, e as entradas combinam o estado recente da série (geração atual e defasada, tendências, sazonalidade) com a previsão meteorológica para aquele horizonte. Na variante de múltiplas saídas (*multi-output*) do *Random Forest*, um único modelo recebe as 13 *features* de base e o vetor completo de previsões do ECMWF, $\big[G_{\text{ECMWF},t+1}, \ldots, G_{\text{ECMWF},t+H}\big]$, e prevê simultaneamente todos os H valores $\big[G_{v,t+1}, \ldots, G_{v,t+H}\big]$. No *XGBoost*, a configuração rotulada como *multi-output* é implementada como H modelos independentes, um por passo, com as mesmas *features* da variante por horizonte e hiperparâmetros mais regularizados (Tabela 14).

A diferença entre os dois algoritmos está na forma de construir o estimador $\mathcal{F}$. O *Random Forest* [163] agrega, por média (*bagging*), um conjunto de árvores de decisão treinadas em reamostragens aleatórias dos dados, reduzindo a variância. O *XGBoost* [164] constrói as árvores de forma sequencial (*boosting* de gradiente), em que cada nova árvore corrige os erros residuais das anteriores, reduzindo o viés. Ambos empregam janela de treino de 365 dias.

No *Random Forest*, utilizam-se 100 árvores com tamanho mínimo de nó igual a 5. No *XGBoost*, minimiza-se o erro quadrático, com 100 rodadas, taxa de aprendizado de 0,1 e profundidade máxima 4 na configuração por horizonte, e 200 rodadas, taxa de 0,05 e profundidade 3 na configuração *multi-output*. A limitação da profundidade e do número de árvores, a taxa de aprendizado reduzida e a janela de um ano atuam como controle de sobreajuste, uma vez que não é utilizado conjunto de validação nem parada antecipada.

A.8 Rede neural (MLP)

Para cada horizonte k, ajusta-se uma rede *multilayer perceptron* [152] com uma camada oculta. A camada de entrada recebe 14 variáveis: as 13 *features* de base da Tabela 16 (todas no instante t) mais a previsão do ECMWF para o horizonte-alvo, $G_{\mathrm{ECMWF},t+k}$. O alvo (*target*) é a geração verificada nesse horizonte, $G_{v,t+k}$. Concretamente, para prever a geração daqui a k passos, a rede recebe a geração atual $G_{v,t}$, suas seis defasagens $G_{v,t-1},\ldots,G_{v,t-6}$, as tendências recentes, o erro corrente do NWP, os termos de sazonalidade e a previsão meteorológica $G_{\mathrm{ECMWF},t+k}$, e produz a estimativa de $G_{v,t+k}$.

As 14 entradas e a resposta são padronizadas por *z-score* (isto é, subtraída a média e dividida pelo desvio-padrão, estimados no treino), etapa indispensável para a convergência do treinamento. Denotando por z_i a i-ésima *feature* padronizada, a saída da rede (ainda padronizada) é:

$$\widetilde{G}_{p,t+k} = \gamma_0 + \sum_{j=1}^{N_h} \gamma_j \, \sigma\left(\omega_{0j} + \sum_{i=1}^{14} \omega_{ij} \, z_i \right) \tag{A.24}$$

em que $\sigma(u) = 1/(1+e^{-u})$ é a função de ativação sigmoide, $N_h = 8$ é o número de neurônios da camada oculta, e ω_{ij} e γ_j são os pesos das camadas oculta e de saída. A topologia pode ser resumida como $14 \to 8\,(\text{sigmoide}) \to 1\,(\text{linear})$. O treinamento minimiza o erro quadrático com regularização por decaimento de pesos (*weight decay* de 0,01), que penaliza pesos elevados e reduz o sobreajuste. A previsão final é obtida desfazendo a padronização da saída e aplicando a saturação:

$$G_{p,t+k} = \mathrm{clamp}\left(\widetilde{G}_{p,t+k} \cdot s_y + \bar{y} \right) \tag{A.25}$$

em que $\bar{y}$ e s_y são a média e o desvio-padrão da resposta no treino. Treina-se uma rede independente para cada um dos 48 horizontes. Cada rede é treinada com até 200 iterações. A inicialização dos pesos é aleatória, de modo que execuções repetidas podem apresentar pequenas variações.

A.9 SARIMA

O SARIMA é aplicado diretamente à série de geração, incorporando a sazonalidade diária de período $s = 48$ (um ciclo de 24 horas em passos de 30 minutos). De forma compacta, o modelo SARIMA$(p,d,q)(P,D,Q)_s$ [85] é descrito por:

$$\phi_p(B)\,\Phi_P(B^s)\,(1-B)^d\,(1-B^s)^D\,G_{v,t} = \theta_q(B)\,\Theta_Q(B^s)\,a_t \tag{A.26}$$

em que Φ_P e Θ_Q são os polinômios sazonais autorregressivo e de médias móveis, e as demais notações seguem a Equação (A.18).

A diferença em relação ao ARIMA é a inclusão de termos que relacionam o instante t a instantes distantes $s = 48$ passos (um dia antes). Por exemplo, um termo autorregressivo sazonal de primeira ordem ($P = 1$) introduz uma dependência explícita da forma $\Phi_1\,G_{v,t-48}$, capturando o fato de que a geração em um dado horário tende a se assemelhar à do mesmo horário do dia anterior. Adota-se diferenciação não sazonal nula ($d = 0$), de modo a obter comportamento reversível à média; a diferenciação sazonal é fixada em $D = 1$, de modo a garantir a persistência sazonal (mesmo horário do dia anterior) em todos os agrupamentos, e as demais ordens são selecionadas automaticamente, com $p,q \leq 3$ e $P,Q \leq 1$.

O modelo é calibrado em janela de 30 dias e, a cada emissão, seu estado é atualizado com os últimos 14 dias de geração. Por não utilizar qualquer informação meteorológica, o SARIMA funciona como modelo de controle puramente temporal, demonstrando o teto de desempenho atingível sem previsões NWP.

A.10 Combinação por inverso do erro (1/MAE)

A combinação de previsões provenientes de modelos distintos é uma estratégia consagrada para reduzir o erro frente aos modelos individuais [142, 143]. Nesta variante, para cada horizonte k, calcula-se o erro absoluto médio ponderado de cada modelo m sobre uma janela deslizante de sete dias, com pesos temporais f_τ crescentes (observações mais recentes recebem maior peso):

$$\mathrm{MAE}_{m,k} = \sum_{\tau} f_\tau \left| G_{v,\tau} - G^{(m)}_{p,\tau} \right|, \qquad f_\tau = \frac{\tau}{\sum_{\tau'} \tau'} \tag{A.27}$$

Os pesos de combinação são proporcionais ao inverso do erro. Um limiar de elegibilidade exclui, em cada horizonte, os modelos cujo erro supera 1,5 vezes o menor erro; os pesos são então renormalizados considerando apenas os modelos elegíveis com previsão válida:

$$w_{m,k} = \frac{1/\mathrm{MAE}_{m,k}}{\sum_{m'} 1/\mathrm{MAE}_{m',k}}, \qquad G_{p,t+k} = \sum_{m} w_{m,k}\,G^{(m)}_{p,t+k} \tag{A.28}$$

A título de exemplo, se em um dado horizonte três modelos elegíveis apresentam erros ponderados de 10, 12 e 20 MW, os pesos não normalizados são $1/10$, $1/12$ e $1/20$; normalizados, resultam em aproximadamente 0,43, 0,36 e 0,21, ou seja, o modelo mais preciso recebe o maior peso.

A.11 Combinação por filtro de Kalman

A combinação adaptativa trata o vetor de pesos $\mathbf{w}_t$ atribuídos aos modelos como um vetor de estado que evolui no tempo, estimado por um filtro de Kalman [57, 165] independente para cada horizonte. O modelo em espaço de estados é:

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \tag{A.29}$$

$$G_{v,t} = \mathbf{f}_t^{\top}\, \mathbf{w}_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, R) \tag{A.30}$$

em que a equação de estado modela os pesos como um passeio aleatório, com covariância $\mathbf{Q}$, e a equação de observação relaciona a geração verificada à combinação das previsões dos modelos. O vetor $\mathbf{f}_t$ reúne as previsões dos M modelos no instante t (atuando como matriz de observação) e R é a variância do ruído de observação. As equações recursivas de predição e atualização são:

$$\mathbf{w}_t^{-} = \mathbf{w}_{t-1}, \qquad \mathbf{P}_t^{-} = \mathbf{P}_{t-1} + \mathbf{Q} \tag{A.31}$$

$$S_t = \mathbf{f}_t^{\top}\, \mathbf{P}_t^{-}\, \mathbf{f}_t + R \tag{A.32}$$

$$\mathbf{K}_t = \frac{\mathbf{P}_t^{-}\, \mathbf{f}_t}{S_t} \tag{A.33}$$

$$\mathbf{w}_t = \mathbf{w}_t^{-} + \mathbf{K}_t \left(G_{v,t} - \mathbf{f}_t^{\top}\, \mathbf{w}_t^{-} \right) \tag{A.34}$$

$$\mathbf{P}_t = \mathbf{P}_t^{-} - \mathbf{K}_t\, \mathbf{f}_t^{\top}\, \mathbf{P}_t^{-} \tag{A.35}$$

em que $\mathbf{P}_t$ é a covariância do estado, S_t a variância da inovação e $\mathbf{K}_t$ o ganho de Kalman. O termo $\left(G_{v,t} - \mathbf{f}_t^{\top}\, \mathbf{w}_t^{-} \right)$ é a inovação: a diferença entre a geração verificada e a previsão combinada com os pesos vigentes; o ganho $\mathbf{K}_t$ determina o quanto os pesos são corrigidos em função dessa inovação. O Algoritmo 1 resume o procedimento.

Algorithm 1 Combinação por filtro de Kalman (um filtro por horizonte k).

Inicializar pesos $\mathbf{w}_0$ (uniformes nos modelos elegíveis) e covariância $\mathbf{P}_0$
for cada instante de emissão t **do**
  $\mathbf{w}^{-} \leftarrow \mathbf{w}_{t-1}$; $\mathbf{P}^{-} \leftarrow \mathbf{P}_{t-1} + \mathbf{Q}$ {predição}
  if há geração verificada $G_{v,t}$ disponível (validade já transcorrida) **then**
    $S \leftarrow \mathbf{f}_t^{\top}\, \mathbf{P}^{-}\, \mathbf{f}_t + R$
    $\mathbf{K} \leftarrow \mathbf{P}^{-}\, \mathbf{f}_t / S$
    $\mathbf{w}_t \leftarrow \mathbf{w}^{-} + \mathbf{K}\,(G_{v,t} - \mathbf{f}_t^{\top}\, \mathbf{w}^{-})$ {atualização}
    $\mathbf{P}_t \leftarrow \mathbf{P}^{-} - \mathbf{K}\, \mathbf{f}_t^{\top}\, \mathbf{P}^{-}$
  else
    $\mathbf{w}_t \leftarrow \mathbf{w}^{-}$; $\mathbf{P}_t \leftarrow \mathbf{P}^{-}$
  end if
  Restringir $\mathbf{w}_t$ ao subconjunto de modelos elegíveis no horizonte
  Projetar $\mathbf{w}_t$ no simplex: $w_m \geq 0$ e $\sum_m w_m = 1$
  $G_{p,t+k} \leftarrow \mathbf{f}_t^{\top}\, \mathbf{w}_t$ {previsão combinada}
end for

A projeção no simplex, componentes não negativas e soma unitária, é decisiva para a estabilidade: uma regressão dinâmica livre, com intercepto, tende a divergir em razão da

multicolinearidade entre modelos correlacionados, ao passo que o simplex sacrifica flexibilidade em favor da estabilidade. A atualização é estritamente causal, utilizando apenas verificações cujo instante de validade já transcorreu. Essa estratégia apresentou o melhor desempenho global, com ganho crescente nos horizontes mais longos.

Na implementação, a variância-base do ruído de estado é $Q = 10^{-4}$ e a do ruído de observação é $R = 300$, da ordem do quadrado do erro absoluto médio do resíduo. Ambas variam com o horizonte: R cresce até quatro vezes o valor do primeiro passo e Q decresce até um quinto, de modo que o filtro é mais reativo nos horizontes curtos e os pesos são mais estáveis nos longos. O conjunto de modelos elegíveis é redefinido uma vez por dia, por horizonte, por seleção dinâmica de *pool*: sobre a janela de sete dias, avaliam-se três tamanhos de *pool* aninhados, formados pelos 6, 5 e 3 modelos de menor erro entre os 12 candidatos, e escolhe-se o de menor erro em um retroteste de origem móvel, em que os dois últimos dias da janela são utilizados para validação. Para evitar trocas por ganhos marginais, a composição só é alterada se o ganho no retroteste superar 3%.

APÊNDICE B – MÉTRICAS DE VALIDAÇÃO E DESEMPENHO

A validação tem por objetivo demonstrar que as estimativas são robustas, estáveis e fisicamente consistentes sob diferentes condições operacionais e meteorológicas. Para isso, adotam-se critérios baseados em métricas de viés, de erros absolutos e relativos e de capacidade de explicar a variabilidade, levando em conta de forma explícita as incertezas observacionais, as associadas ao modelo e as de natureza aleatória [166]. Como uma métrica isolada raramente caracteriza por completo o desempenho de um modelo, é prática usual empregar um conjunto de métricas complementares.

Para avaliar o desempenho dos modelos de previsão, aplicam-se diversas métricas, entre elas o Erro Médio (ME), o Erro Absoluto Médio (MAE), a Raiz do Erro Quadrático Médio (RMSE) e o Erro Percentual Absoluto Médio (MAPE). O ME serve sobretudo para identificar viés sistemático; o MAE quantifica a magnitude média do erro, sendo menos sensível a *outliers* do que o RMSE; o RMSE penaliza erros de grande magnitude; e o MAPE expressa o erro percentual entre estimativas e observações [162].

Essas métricas de erro também podem ser normalizadas, dando origem ao nME, nMAE, nMAPE e nRMSE, em geral pela capacidade instalada, pela amplitude da série temporal ou por outros fatores de escala, o que viabiliza comparações entre parques eólicos de características distintas [167], [168]. Além das métricas baseadas em erro, emprega-se o coeficiente de eficiência de Nash-Sutcliffe (NSE) para quantificar o desempenho do modelo em relação à média dos valores observados.

O MAPE, contudo, não é recomendado quando os valores observados se aproximam de zero, pois sua normalização pode inflar artificialmente o erro mesmo em modelos de bom desempenho. Por essa razão, análises de geração eólica costumam recorrer ao Erro Percentual Absoluto Médio normalizado (nMAPE).

Cada métrica apresenta vantagens e limitações, de modo que a mais apropriada depende da análise em questão. As equações a seguir apresentam as métricas utilizadas. A primeira é o erro médio (*mean error*):

$$ME = \frac{1}{N}\sum_{t=1}^{N}\left(G_{v,t} - G_{p,t}\right) \tag{B.1}$$

em que $G_{v,t}$ é a geração verificada no instante t; $G_{p,t}$ é a geração prevista no instante t; e N o número de elementos da amostra.

O erro absoluto médio (*mean absolute error*) é calculado como segue:

$$MAE = \frac{1}{N}\sum_{t=1}^{N}\left|G_{v,t} - G_{p,t}\right| \tag{B.2}$$

O erro percentual absoluto médio (*mean absolute percentage error*):

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{G_{v,t} - G_{p,t}}{G_{v,t}}\right| \cdot 100 \tag{B.3}$$

Outra métrica relevante é a raiz do erro quadrático médio (*root mean square error*), descrita na equação:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(G_{v,t} - G_{p,t}\right)^2} \tag{B.4}$$

Emprega-se ainda o erro percentual absoluto médio normalizado (*normalized mean absolute percentage error*), normalizado pela potência instalada:

$$nMAPE_{Pinst} = \frac{1}{N \cdot P_{inst}^{tot}}\sum_{t=1}^{N}\left|G_{v,t} - G_{p,t}\right| \cdot 100 \tag{B.5}$$

Também é utilizado o erro percentual absoluto médio normalizado pela potência instalada ponderada pelo fator de capacidade:

$$nMAPE_{fatcap} = \frac{1}{N \cdot P_{inst}^{tot} \cdot fc}\sum_{t=1}^{N}\left|G_{v,t} - G_{p,t}\right| \cdot 100 \tag{B.6}$$

em que fc é o fator de capacidade do conjunto de usinas considerado. Essa normalização utiliza como referência a geração média esperada, e não a potência instalada total, o que torna o índice mais representativo do desempenho relativo, sobretudo em regiões cujo fator de capacidade é baixo.

Por fim, o coeficiente de eficiência de Nash-Sutcliffe (NSE) [161], dado por:

$$NSE = \left(1 - \frac{\sum_{t=1}^{N}\left(G_{v,t} - G_{p,t}\right)^2}{\sum_{t=1}^{N}\left(G_{v,t} - \overline{G_v}\right)^2}\right) \times 100\% \tag{B.7}$$

O NSE varia de $-\infty$ a 100%, em que 100% indica ajuste perfeito entre as séries prevista e verificada, 0% indica que o modelo possui poder preditivo equivalente ao da média observada, e valores negativos indicam que a média observada seria um preditor superior ao modelo.

Cada uma dessas métricas evidencia um aspecto distinto do desempenho, o que justifica seu uso complementar. O ME, por preservar o sinal do erro, é a métrica adequada para identificar viés sistemático, embora erros de sinais opostos possam se cancelar em sua média. O MAE quantifica a magnitude média do erro absoluto na mesma unidade da grandeza, mostrando-se mais robusto à presença de *outliers*. O RMSE, ao elevar os erros ao quadrado, penaliza com mais intensidade os erros de grande magnitude, sendo por isso mais sensível a erros elevados e útil para avaliar a capacidade do modelo de evitá-los.

O MAPE e sua variante normalizada (nMAPE) expressam o erro em termos percentuais, o que facilita a comparação entre usinas e regiões de diferentes portes; as formas normalizadas

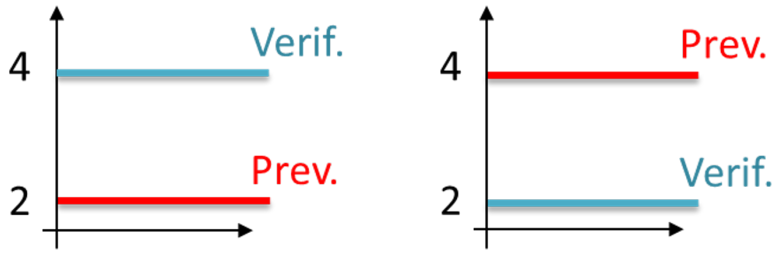
pela potência instalada (Equação (B.5)) ou por outro fator de escala contornam a instabilidade do MAPE quando a geração verificada se aproxima de zero. Já o NSE compara o erro quadrático do modelo com a variabilidade dos dados observados em torno de sua média, indicando se o modelo supera o uso da média observada como previsão. Como uma única métrica costuma ser insuficiente para caracterizar plenamente o desempenho, recomenda-se a análise conjunta de um conjunto complementar [166, 167].

Cabe destacar que, embora as métricas apresentadas tenham sido descritas no contexto da avaliação da previsão de geração, elas são igualmente empregadas na avaliação dos resultados da estimação de geração. Em ambos os casos, o cálculo é análogo, substituindo-se apenas o valor previsto pelo valor estimado no confronto com a geração verificada.

O MAPE, definido na Equação (B.3), é bastante útil na análise de erros de previsão, mas exige cautela, pois se trata de um somatório de frações cujo denominador é a geração verificada. Como a usina eólica pode apresentar geração muito baixa, esse denominador reduzido tende a superestimar o valor do somatório. Por isso, o MAPE não é indicado para usinas individuais ou pequenos conjuntos. Já para conjuntos agregados maiores, como a avaliação de erros por ponto de conexão, por estado ou por região, seu uso pode até ser viável.

Outra limitação do MAPE é que um mesmo erro absoluto pode resultar em erros percentuais muito diferentes. A Figura 122 apresenta duas previsões hipotéticas, em dias distintos. O gráfico à esquerda representa um dia em que o valor previsto foi 2 e o verificado foi 4; o gráfico à direita, um dia em que o previsto foi 4 e o verificado foi 2.

Figura 122 – Exemplo 1 de aplicação do MAPE e nMAPE, com dois dias hipotéticos de mesmo erro absoluto.



Fonte: Elaboração própria.

Em ambas as situações o erro absoluto é de 2 unidades; no entanto, o MAPE difere de forma acentuada entre elas, conforme calculado a seguir:

$$Erro_1 = \left| \frac{Verif_1 - Prev_1}{Verif_1} \right| = \left| \frac{4-2}{4} \right| = 50\% \tag{B.8}$$

$$Erro_2 = \left| \frac{Verif_2 - Prev_2}{Verif_2} \right| = \left| \frac{2-4}{2} \right| = 100\% \tag{B.9}$$

No nMAPE esse inconveniente não ocorre, pois a referência é a mesma nas duas situações. Tomando, hipoteticamente, o valor 4 como referência:

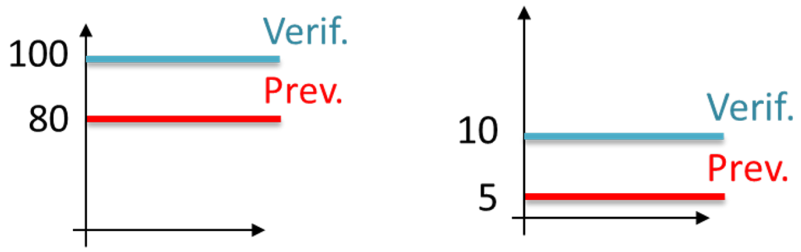
$$Erro_1 = \frac{1}{4} |Verif_1 - Prev_1| = \frac{1}{4} |4-2| = 50\% \tag{B.10}$$

$$Erro_2 = \frac{1}{4}|Verif_2 - Prev_2| = \frac{1}{4}|2-4| = 50\% \tag{B.11}$$

Comparando os resultados, verifica-se que os índices do nMAPE apresentam maior concordância entre as duas situações do que os do MAPE.

O MAPE distingue-se ainda do nMAPE quanto à gravidade que seu valor sugere. Na Figura 123, o gráfico à esquerda resulta em MAPE de 20% para um erro de 20 unidades, enquanto o da direita resulta em 50% para um erro de apenas 5 unidades. Assim, um MAPE mais elevado não indica necessariamente um erro maior em MW, que é, de fato, a grandeza mais relevante.

Figura 123 – Exemplo 2 de aplicação do MAPE e nMAPE, com MAPE maior para um erro absoluto menor.

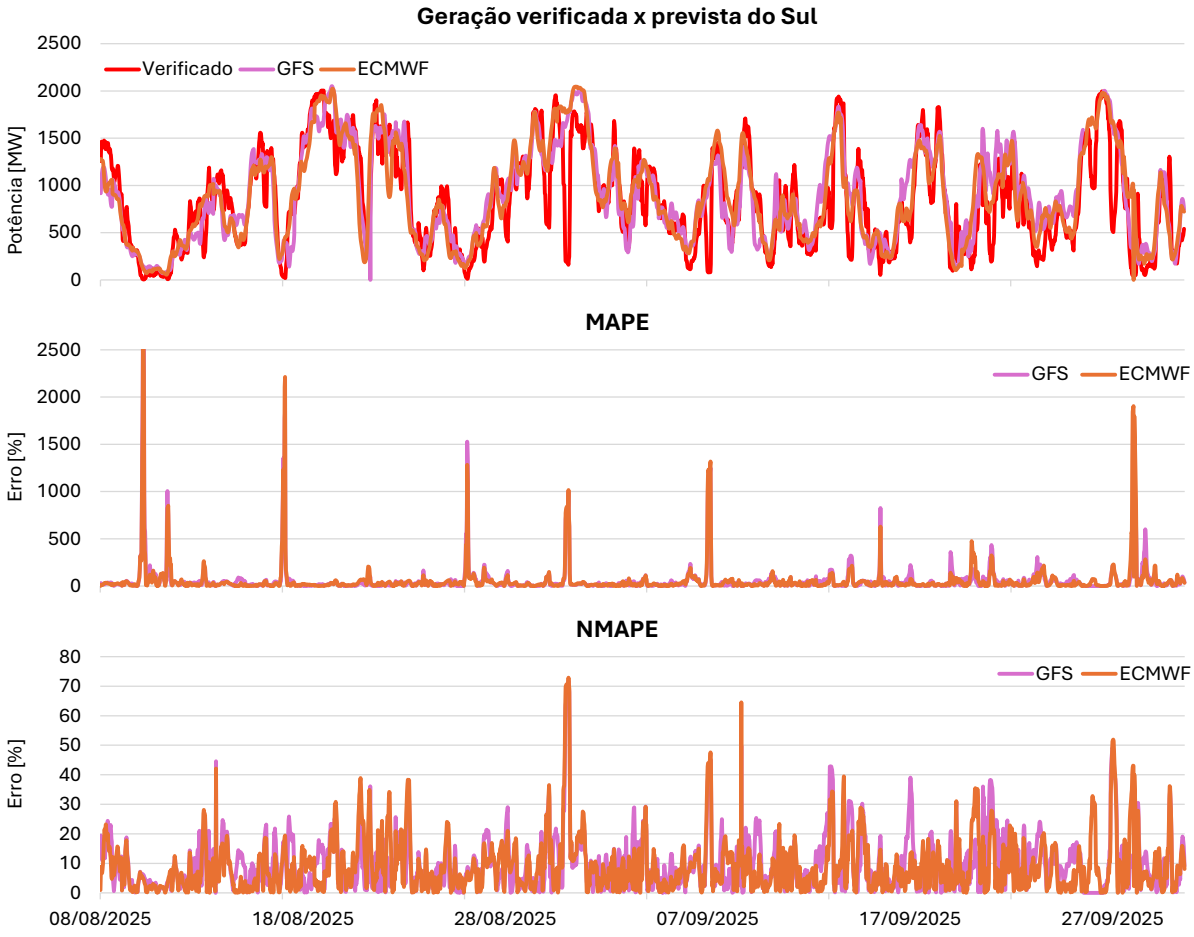


Fonte: Elaboração própria.

Para a análise dos erros do subsistema Nordeste, o MAPE pode ser considerado, ainda que com as ressalvas apontadas. No subsistema Sul, porém, seu uso não é recomendado, embora neste trabalho a métrica seja apresentada. Isso porque a geração eólica na região se aproxima de zero com muita frequência, o que produz erros percentuais extremamente elevados no MAPE mesmo quando os erros em MW são baixos. A Figura 124 ilustra um caso ocorrido entre agosto e setembro de 2025. Observam-se erros superiores a 2000% pelo MAPE, ao passo que o nMAPE resulta em erro baixo, condizente com a realidade, uma vez que o erro em MW é muito pequeno.

Diante do exposto, conclui-se que o uso de uma única métrica pode ser inadequado. O ideal é analisá-las em conjunto, observando o que cada uma indica; a preferência por uma ou outra dependerá do objetivo de aplicação da previsão de geração eólica.

Figura 124 – Exemplo de erro para um período do Sul, com erros percentuais elevados pelo MAPE apesar de erros baixos em MW.



Fonte: Elaboração própria.

APÊNDICE C – ANÁLISES DESAGREGADAS DOS RESULTADOS

Este apêndice reúne os resultados apresentados em níveis mais desagregados espacialmente, por conjuntos (agrupamentos) de usinas e por usinas individuais, complementando as análises agregadas por subsistema e os panoramas nacionais discutidos no Capítulo 6. A organização segue a mesma ordem do capítulo: estimação da geração, ajuste das previsões de vento, previsão de geração de curtíssimo e curto prazo e previsão em tempo real.

C.1 Comportamento da geração eólica no Brasil

Além da concentração espacial nas regiões Nordeste e Sul, ilustrada na Figura 3 do Capítulo 2, existem diferenças estruturais substanciais nos regimes de vento entre essas duas regiões, que resultam em padrões distintos de geração eólica, com implicações significativas para o planejamento e a operação do sistema elétrico.

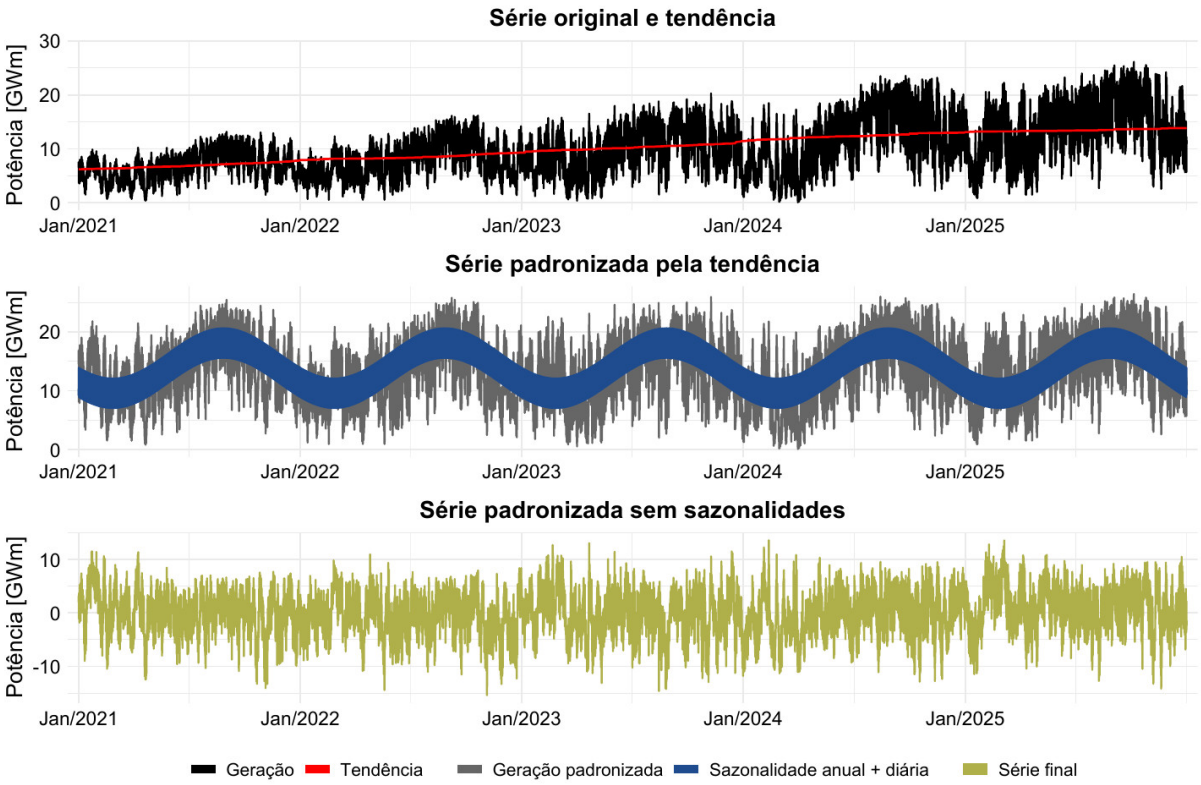
No Nordeste, os padrões de vento são impulsionados pela interação de diversos sistemas de circulação atmosférica com escalas espaciais e temporais variadas [169]. Entre os contribuintes mais notáveis estão a Zona de Convergência Intertropical (ZCIT), os Vórtices Ciclônicos de Altos Níveis (VCAN), os sistemas frontais e as circulações locais de mesoescala, todos levando a um padrão sazonal bem definido na intensidade média do vento. Velocidades de vento mais altas são tipicamente observadas entre o inverno e a primavera austrais, com intensidades menores ocorrendo durante o outono austral [170]. A presença de ventos alísios, derivados da circulação geral da atmosfera, também contribui para um perfil de geração eólica mais contínuo e previsível nessa região, com fatores de capacidade frequentemente superiores a 40%.

A Figura 125 apresenta a evolução histórica e o comportamento horário da geração eólica verificada no Nordeste. O primeiro gráfico mostra a série temporal bruta juntamente com sua tendência de crescimento, ajustada com base na capacidade instalada e no fator de capacidade médio. O segundo gráfico exibe a série normalizada em relação à tendência, enquanto o terceiro remove tanto as componentes sazonais anuais quanto diárias, revelando a série residual e sua variabilidade intrínseca.

Em contraste, a região Sul exibe um regime de vento mais instável, fortemente influenciado por sistemas sinóticos transientes, como frentes frias e ciclones extratropicais. Isso leva a mudanças abruptas na direção e na intensidade do vento, aumentando a intermitência da geração e reduzindo a previsibilidade. Tais condições demandam modelos de previsão mais robustos e estratégias operacionais adaptativas para garantir a segurança e a confiabilidade do sistema elétrico. A Figura 126 ilustra a geração de energia eólica no Sul, seguindo a mesma estrutura de decomposição aplicada ao Nordeste.

Uma comparação entre as Figuras 125 e 126 destaca as diferenças nos padrões de geração entre as duas regiões. No Nordeste, os ciclos sazonais anuais e diários são mais consistentes,

Figura 125 – Comportamento da geração de energia eólica na região Nordeste, decomposta em tendência, sazonalidades e série residual.



Fonte: Elaboração própria.

enquanto o Sul é caracterizado por flutuações de maior amplitude. O intervalo interquartil (IIQ), normalizado pela capacidade instalada, enfatiza ainda mais esse contraste: aproximadamente 17% no Nordeste ante 34% no Sul. Esses resultados evidenciam a variabilidade significativamente maior da geração eólica no Sul, indicando desafios adicionais na previsão, na modelagem e na integração ao sistema.

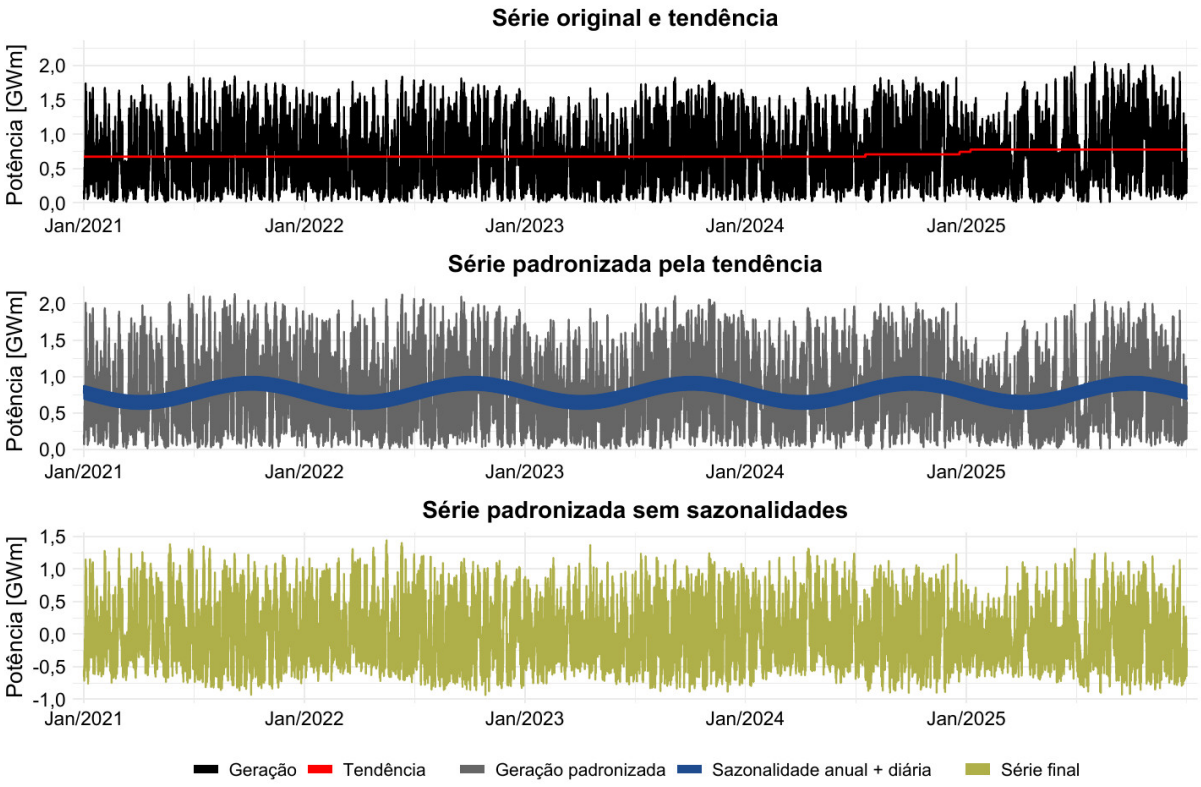
A caracterização do comportamento típico e da variabilidade do recurso eólico em cada região, apresentada nesta seção, fornece a base interpretativa para as análises desagregadas que se seguem. As diferenças de regime de vento e de sazonalidade entre Nordeste e Sul ajudam a explicar os padrões observados nos resultados da estimação e da previsão da geração, permitindo uma leitura mais completa do desempenho dos modelos em cada usina, agrupamento e sub-região.

C.2 Resultados da estimação por usinas e agrupamentos

C.2.1 Resultados por usinas individuais

Conforme descrito no Apêndice B, para uma avaliação adequada e uma interpretação mais abrangente do desempenho do modelo proposto, é recomendável utilizar mais de uma

Figura 126 – Comportamento da geração de energia eólica na região Sul, decomposta em tendência, sazonalidades e série residual.



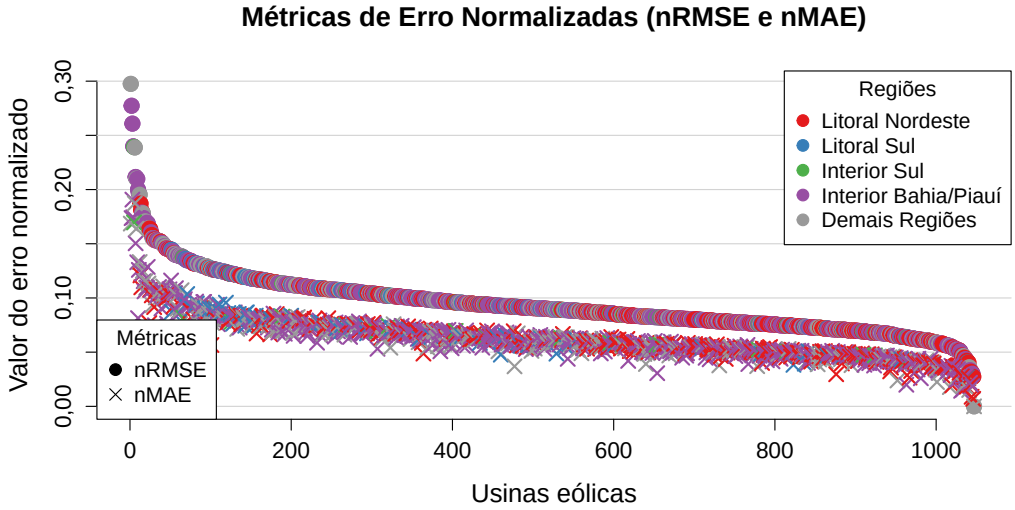
Fonte: Elaboração própria.

métrica de desempenho. A adoção de métricas complementares permite que diferentes aspectos do erro de estimação sejam analisados e proporciona uma visão mais robusta da qualidade dos resultados obtidos.

A Figura 127 apresenta as métricas de erro normalizadas, nRMSE e nMAE, calculadas para cada usina eólica individualmente. As usinas estão ordenadas de forma crescente pelo valor do nRMSE, e a diferenciação por cores indica a região geográfica à qual cada usina pertence: Litoral Nordeste, Litoral Sul, Interior Sul, Interior Bahia/Piauí e Demais Regiões, conforme a distribuição apresentada na Figura 3 do Capítulo 2. Os círculos preenchidos representam o nRMSE, enquanto os marcadores em forma de cruz representam o nMAE. Observa-se que a maioria das usinas apresenta valores de nRMSE e nMAE em faixas relativamente baixas, indicando boa aderência do modelo de estimação aos dados verificados. Nota-se também que as usinas com maiores erros não estão concentradas em uma única região, sugerindo que o desempenho do modelo não é determinado exclusivamente pela localização geográfica, mas também por fatores locais como a qualidade dos dados operacionais e a complexidade do regime de ventos.

A Figura 128 apresenta a distribuição do erro médio normalizado (nME) calculado para as usinas individualizadas, na forma de histograma. O eixo horizontal representa o valor do nME,

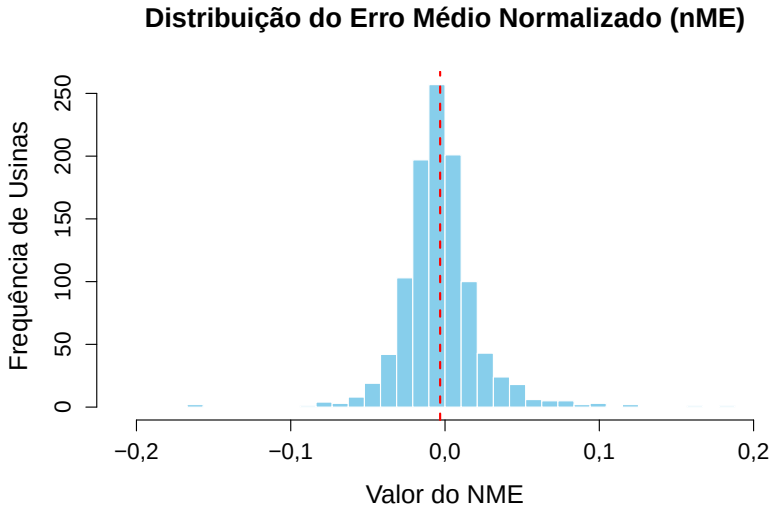
Figura 127 – Métricas de erro normalizadas pela amplitude da série (nRMSE e nMAE) por usina eólica, ordenadas pelo nRMSE.



Fonte: Elaboração própria.

com escala simétrica em torno de zero, enquanto o eixo vertical indica a frequência de usinas em cada faixa de erro. A linha tracejada em vermelho indica o valor médio do nME sobre o conjunto de usinas. Observa-se que a distribuição é aproximadamente centrada em zero, o que indica ausência de viés sistemático significativo no modelo de estimação quando avaliado no nível de usinas individuais. A concentração das ocorrências em torno de valores próximos de zero reforça a capacidade do modelo em reproduzir, na média, os volumes de geração verificados. Eventuais usinas com nME mais afastado de zero podem estar associadas a particularidades operacionais ou a deficiências nos dados de entrada que merecem investigação complementar.

Figura 128 – Distribuição do erro médio normalizado (nME) por usina eólica, normalizado pela amplitude da série.



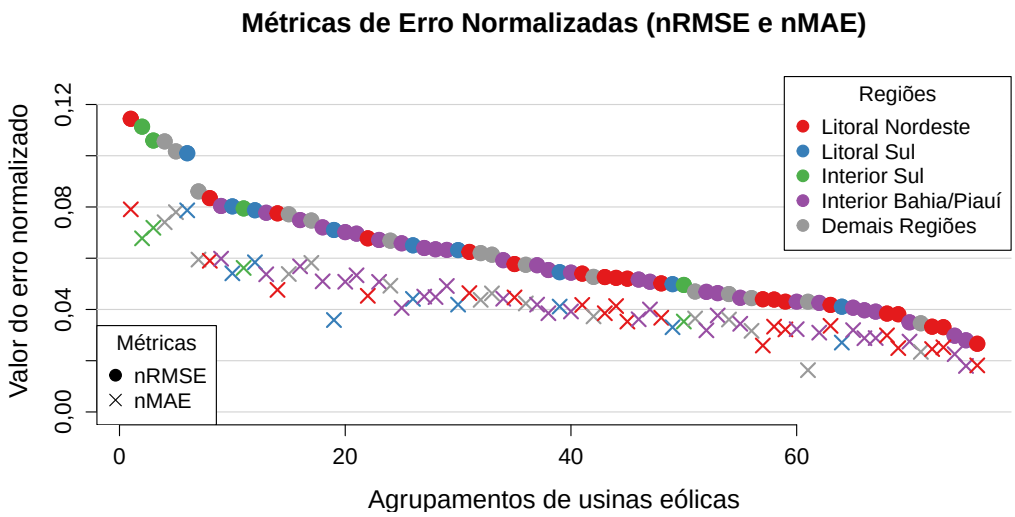
Fonte: Elaboração própria.

C.2.2 Resultados por agrupamentos de usinas

A Figura 129 apresenta os valores das métricas nRMSE e nMAE calculados para os diferentes grupos de parques eólicos. A normalização foi necessária porque os grupos possuem capacidades de geração significativamente diferentes; para tornar os erros comparáveis, cada série temporal foi normalizada por sua amplitude, definida como a diferença entre seus valores máximo e mínimo. Cada ponto no gráfico corresponde ao desempenho de um grupo, permitindo uma comparação direta do erro relativo de estimação de geração entre regiões com características operacionais distintas.

Os círculos preenchidos representam os valores de nRMSE, enquanto os símbolos em forma de "x" representam os valores de nMAE, possibilitando a visualização simultânea do comportamento de ambas as métricas. As cores dos pontos indicam a região geográfica à qual cada grupo pertence, conforme mostrado na legenda superior. Essa categorização permite observar que, embora existam padrões regionais bem definidos nos regimes de vento (como diferenças entre áreas costeiras e interioranas), esses padrões não se traduzem diretamente em diferenças nos erros do modelo de estimação. Isso ocorre porque os dados de vento observados utilizados no modelo já capturam adequadamente as características aerodinâmicas e climáticas locais. Como resultado, fatores como qualidade de medição, disponibilidade de anemômetros e consistência das informações operacionais fornecidas ao operador do sistema exercem influência muito maior no desempenho da estimação do que a localização geográfica.

Figura 129 – Desempenho do modelo para cada grupo de parques eólicos, normalizado pela amplitude da série.



Fonte: Elaboração própria.

É importante destacar que essa observação se restringe ao contexto da estimação de geração. Em modelos de previsão de energia eólica, a variabilidade meteorológica entre diferentes regiões exerce um impacto significativamente mais relevante no desempenho dos modelos, influenciando diretamente a qualidade das previsões de velocidade do vento e, por consequência,

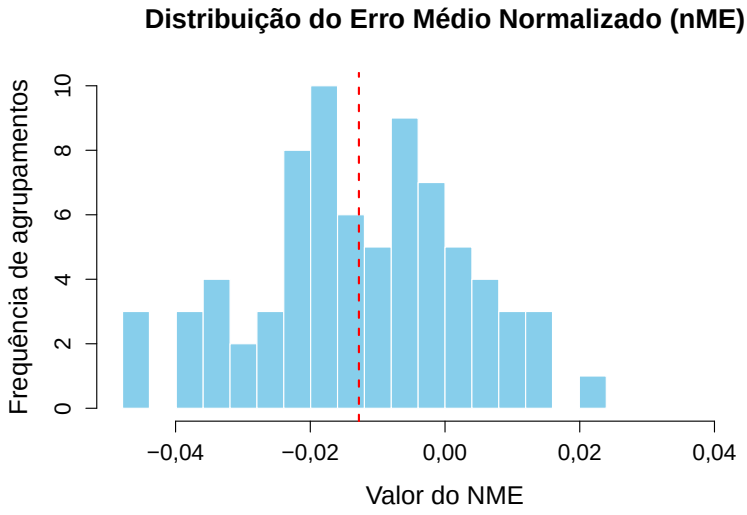
das previsões de geração eólica, conforme evidenciado em [32]. Esse comportamento será demonstrado nas seções dedicadas aos resultados do processo de previsão.

Finalmente, o uso combinado de nRMSE e nMAE proporciona uma avaliação mais abrangente do erro do modelo: enquanto o nRMSE atribui maior peso a grandes erros, o nMAE é mais robusto a *outliers*. A comparação de ambas as métricas oferece, portanto, uma visão consistente e equilibrada da precisão da estimação de geração para cada grupo analisado.

Complementarmente, a Figura 130 apresenta a distribuição do erro médio normalizado (nME) calculado para os agrupamentos de usinas eólicas utilizados na avaliação do modelo de estimação. Cada agrupamento corresponde a um conjunto de usinas cuja geração é agregada para fins de operação no SIN. O eixo horizontal representa o valor do nME, com escala simétrica em torno de zero, enquanto o eixo vertical indica a frequência de agrupamentos em cada faixa de erro. A linha tracejada em vermelho indica o valor médio do nME sobre o conjunto de agrupamentos.

Observa-se que a distribuição não está totalmente concentrada em torno de zero, indicando a existência de viés em determinados agrupamentos. Esse comportamento pode ser atribuído ao fato de que, embora as restrições operativas sejam identificadas e removidas do conjunto de dados, o processo de filtragem não é perfeito. Como consequência, períodos residuais de operação sob restrição podem permanecer nas séries históricas, comprometendo a avaliação do desempenho do modelo. Adicionalmente, a análise por agrupamentos é realizada sobre subconjuntos com número relativamente reduzido de amostras, o que aumenta a sensibilidade das métricas e pode acentuar a percepção de vieses observados.

Figura 130 – Distribuição do erro médio normalizado (nME) por agrupamento de usinas, normalizado pela amplitude da série.



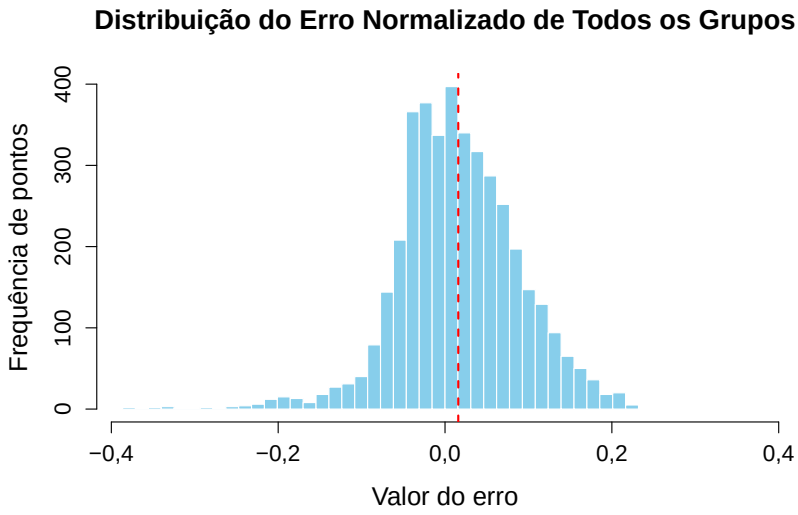
Fonte: Elaboração própria.

A Figura 131 apresenta a distribuição do erro considerando todas as amostras em con-

junto, associadas a todos os agrupamentos de usinas, oferecendo uma visão mais granular do desempenho do modelo. Diferentemente da figura anterior, em que cada observação corresponde ao nME de um agrupamento, nesta figura cada observação corresponde ao erro de um ponto temporal individual dentro de cada agrupamento de usinas, resultando em uma amostra significativamente maior. A linha tracejada em vermelho indica o valor médio do erro sobre todos os pontos.

A distribuição neste caso apresenta-se mais concentrada em torno de zero, confirmando a ausência de viés sistemático neste nível de desagregação. Esse resultado é explicado pelo fato de que, embora alguns agrupamentos apresentem individualmente tendências mais acentuadas, tais vieses foram identificados em uma distribuição com número de amostras muito menor. Ao considerar todos os pontos em conjunto, os erros positivos e negativos tendem a se distribuir, evidenciando que o modelo não possui viés sistemático global. Contudo, observa-se uma maior dispersão em relação à figura anterior, o que é esperado, uma vez que a análise por pontos individuais captura variabilidades locais que são atenuadas quando os resultados são agregados por agrupamento. Essa análise complementar permite identificar pontos específicos com desempenho discrepante que podem requerer investigação adicional quanto à qualidade dos dados de entrada ou às condições operacionais locais.

Figura 131 – Distribuição do erro normalizado de todos os pontos dos agrupamentos de usinas.

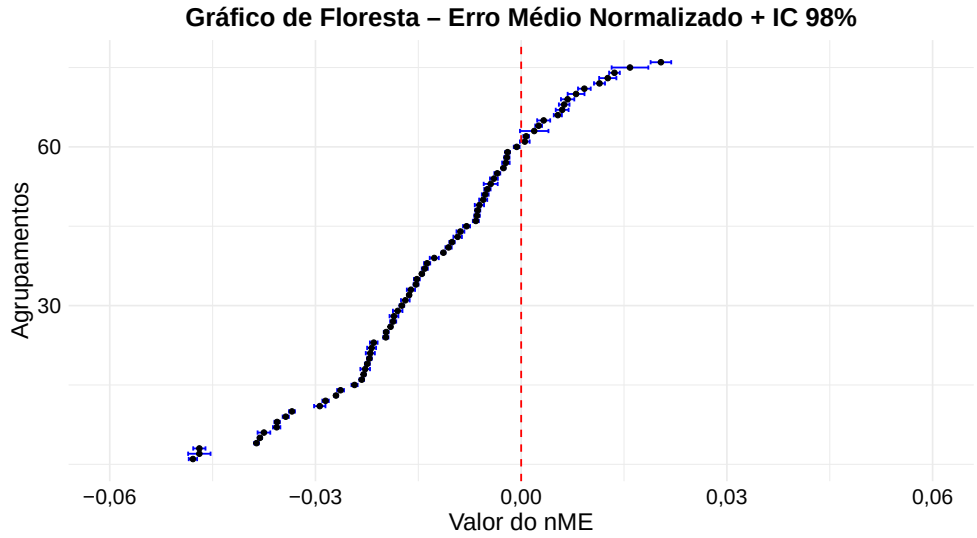


Fonte: Elaboração própria.

A Figura 132 apresenta uma análise complementar, um gráfico de floresta (*forest plot*) mostrando os erros médios normalizados para cada grupo analisado, juntamente com seus respectivos intervalos de confiança de 98%. Cada ponto representa o valor médio do erro normalizado obtido para o grupo, enquanto as barras horizontais indicam a incerteza associada a essa estimativa. A linha tracejada vertical em zero corresponde ao cenário ideal de ausência de viés. De uma perspectiva estatística, o intervalo de confiança é calculado a partir do erro médio amostral, do desvio padrão amostral e do tamanho da amostra. Como a variância populacional

é desconhecida e os tamanhos amostrais são finitos, a distribuição t de Student é empregada, fornecendo um valor crítico apropriado para cada grau de liberdade.

Figura 132 – Gráfico de floresta dos erros médios dos grupos de parques eólicos, com seus intervalos de confiança.



Fonte: Elaboração própria.

Conforme mostrado na Figura 132, os valores médios do erro normalizado (verificado − estimado) revelam uma tendência de viés negativo em alguns grupos, indicando que, nesses casos, os valores estimados são, em média, ligeiramente superiores aos valores observados. Conforme mencionado anteriormente, isso ocorre porque, em determinados momentos, alguns parques eólicos operam sob restrições operacionais internas que nem sempre são fielmente refletidas nos dados submetidos ao operador do sistema. Consequentemente, parte desse efeito de limitação operacional é interpretada pelo modelo como erro de estimação, resultando em viés negativo.

Por outro lado, alguns grupos apresentam viés positivo, indicando que a estimativa foi inferior à geração observada. Esse comportamento está geralmente associado à entrada recente de novas usinas no grupo, o que altera sua capacidade total de geração. Como o modelo de estimação depende de um conjunto mínimo de dados históricos para calibração e treinamento, pode ocorrer uma defasagem temporal entre a expansão da frota eólica e a disponibilidade de dados suficientes para configurar adequadamente o modelo, resultando em subestimação temporária. Apesar desses comportamentos isolados, a vasta maioria dos grupos apresenta erros médios muito pequenos, tipicamente abaixo de 5%, sejam positivos ou negativos. Isso indica que, de modo geral, o modelo apresenta bom desempenho na estimação da geração dos grupos, com viés sistemático reduzido.

Os resultados em nível de grupo de parques eólicos são particularmente relevantes para a avaliação da contabilização de geração e são essenciais nos procedimentos de compensação para agentes geradores afetados. Além disso, esses resultados são importantes para a operação em tempo real, pois os valores de geração estimada podem servir como referência para a liberação de

restrições operacionais sem comprometer a segurança do sistema. Isso porque, se um operador libera uma restrição sem uma estimativa confiável da geração eólica, pode ocorrer excesso de geração, potencialmente levando a eventos de sobrefrequência e problemas operacionais mais graves.

C.3 Resultados do ajuste das previsões de vento por ponto e por estado

A análise é estruturada em duas escalas espaciais complementares (ponto individual e estado) e, em cada uma delas, são apresentados lado a lado os comportamentos das previsões brutas e das previsões corrigidas, permitindo a comparação direta entre os dois cenários. Essa organização tem por objetivo facilitar a percepção visual do impacto da correção em cada escala, bem como evidenciar a consistência dos ganhos obtidos ao longo das diferentes dimensões de análise. O panorama nacional do viés médio, de caráter agregado, é apresentado no Capítulo 6.

C.3.1 Análise por ponto de análise

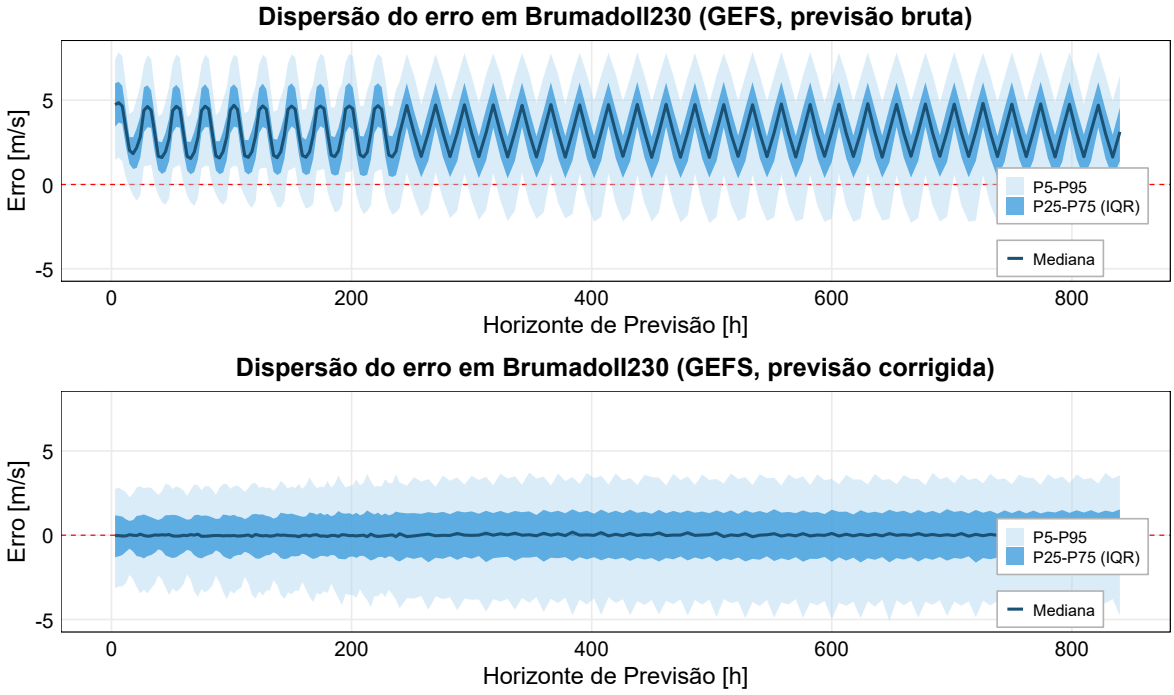
Para detalhar o comportamento do erro em escala pontual, esta subseção apresenta análises em dois níveis complementares: primeiro, os dois pontos representativos discutidos na Seção 5.3.2.1 do Capítulo 5, Brumado II 230 e Açu II 230, são revisitados, agora exibindo o erro antes e após a correção em uma mesma figura; em seguida, a distribuição do erro é consolidada para os 76 pontos de análise por meio de *boxplots*, contemplando os três modelos meteorológicos.

C.3.1.1 Comportamento ao longo do horizonte para pontos representativos

Nesta seção, o erro é definido como a diferença entre o vento verificado e o vento previsto (verificado − previsto); assim, valores positivos indicam subestimação e valores negativos, superestimação do vento pelos modelos. As Figuras 133 e 134 apresentam a dispersão do erro ao longo do horizonte de previsão, antes (painel superior) e após (painel inferior) a aplicação da correção, para os pontos Brumado II 230 e Açu II 230. Por questões de espaço e clareza, os gráficos exibem apenas o modelo GEFS; entretanto, os comportamentos observados e as conclusões obtidas estendem-se de forma análoga aos modelos ECMWF e GFS.

Conforme caracterizado na Seção 5.3 do Capítulo 5, o viés observado nas previsões brutas varia conforme a hora do dia, apresentando oscilações regulares de período aproximado de 24 horas ao longo do horizonte de previsão. Esse comportamento reflete o fato de que os modelos meteorológicos representam de forma imperfeita o ciclo diurno do vento, em alguns horários do dia tendem a superestimar a velocidade e, em outros, a subestimá-la. Como a metodologia de correção é aplicada de forma individualizada para cada um dos quarenta e oito intervalos do dia, essa variabilidade horária do viés é capturada e ajustada intervalo a intervalo.

Figura 133 – Dispersão do erro de vento ao longo do horizonte para o ponto Brumado II 230, antes e após a correção.



Fonte: Elaboração própria.

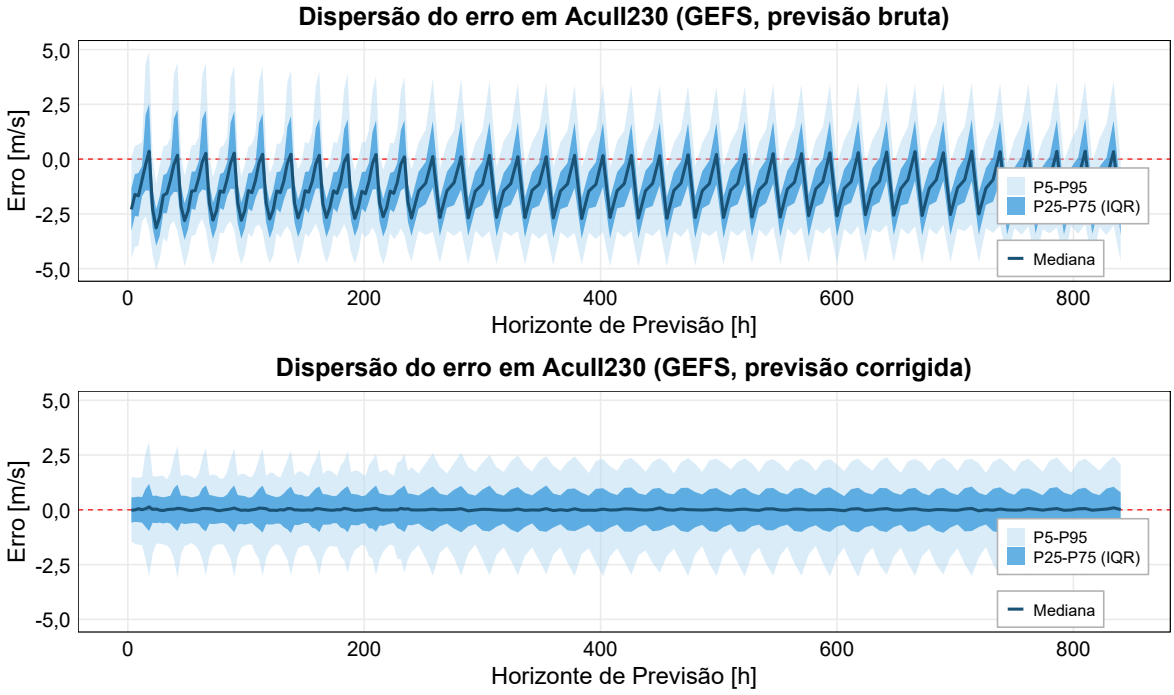
A análise dos painéis inferiores das Figuras 133 e 134 confirma esse comportamento. Em ambos os pontos, a mediana do erro após a correção mantém-se próxima de zero ao longo de todo o horizonte de previsão, com as oscilações de período diário substancialmente atenuadas em relação aos painéis superiores (previsões brutas). No ponto Brumado II 230, onde a previsão bruta apresentava viés positivo persistente (subestimação sistemática do vento), observa-se a remoção desse viés sistemático. No ponto Açu II 230, em que o viés bruto já era negativo (leve superestimação), mas em magnitude menor, a correção promove a redução da amplitude residual das oscilações horárias. Em ambos os casos, observa-se também redução da dispersão das bandas P25–P75 e P5–P95, indicando que a correção, além de remover o viés médio, reduz a magnitude dos erros nos quantis extremos.

C.3.1.2 Análise consolidada para o conjunto dos 76 pontos

Para verificar se o comportamento observado nos pontos representativos se generaliza para o conjunto completo de pontos de análise, as Figuras 135, 136 e 137 apresentam, para cada modelo meteorológico, a distribuição dos erros das previsões brutas (painel superior) e corrigidas (painel inferior) para os 76 pontos de análise, considerando o horizonte das primeiras 24 horas. Os pontos são agrupados por estado, com faixas de fundo coloridas marcando os limites de cada agrupamento e as siglas dos estados indicadas no topo do painel superior.

A inspeção dos painéis superiores das três figuras revela que os padrões já identificados em Brumado II 230 e Açu II 230 se generalizam para o conjunto dos 76 pontos. Na Bahia (BA),

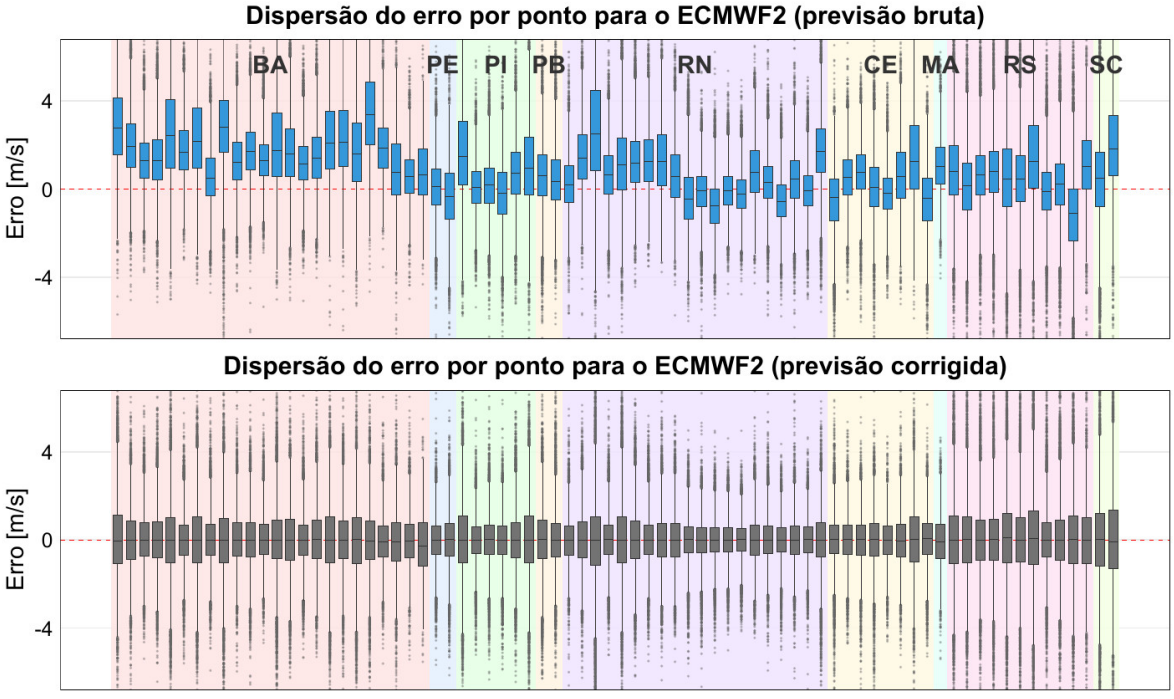
Figura 134 – Dispersão do erro de vento ao longo do horizonte para o ponto Açu II 230, antes e após a correção.



Fonte: Elaboração própria.

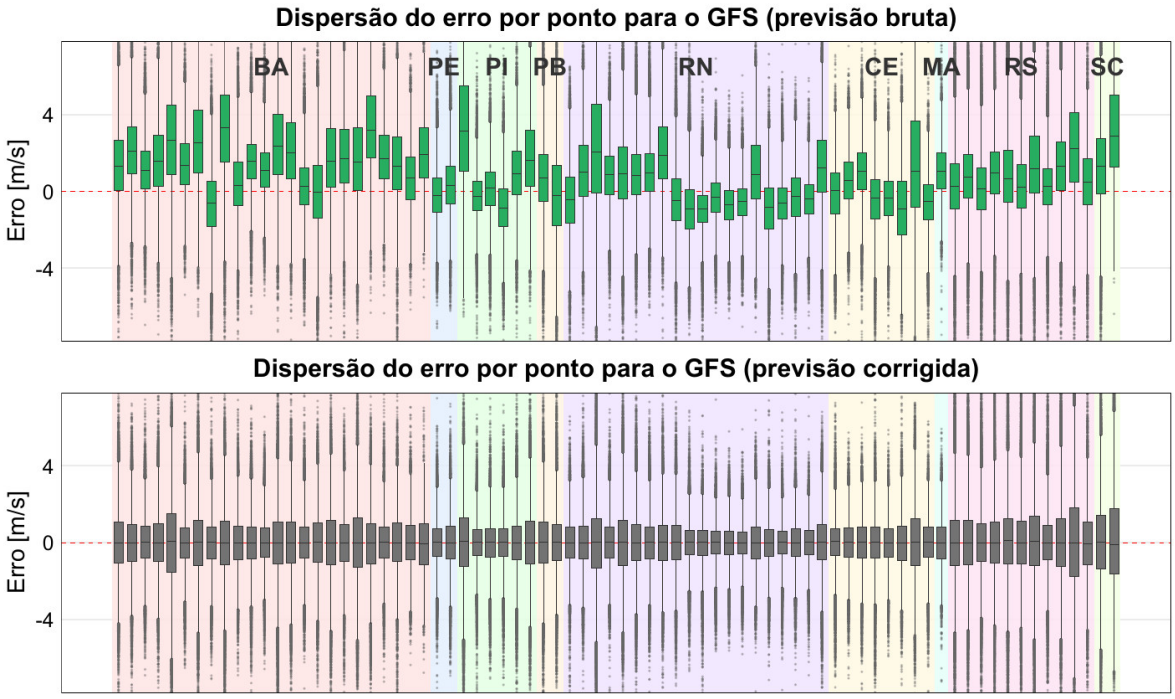
as medianas dos erros são predominantemente positivas em todos os modelos, confirmando a subestimação sistemática nessa região; no Rio Grande do Norte (RN), as medianas são mais

Figura 135 – Dispersão do erro de vento do ECMWF por ponto de análise, antes e após a correção.



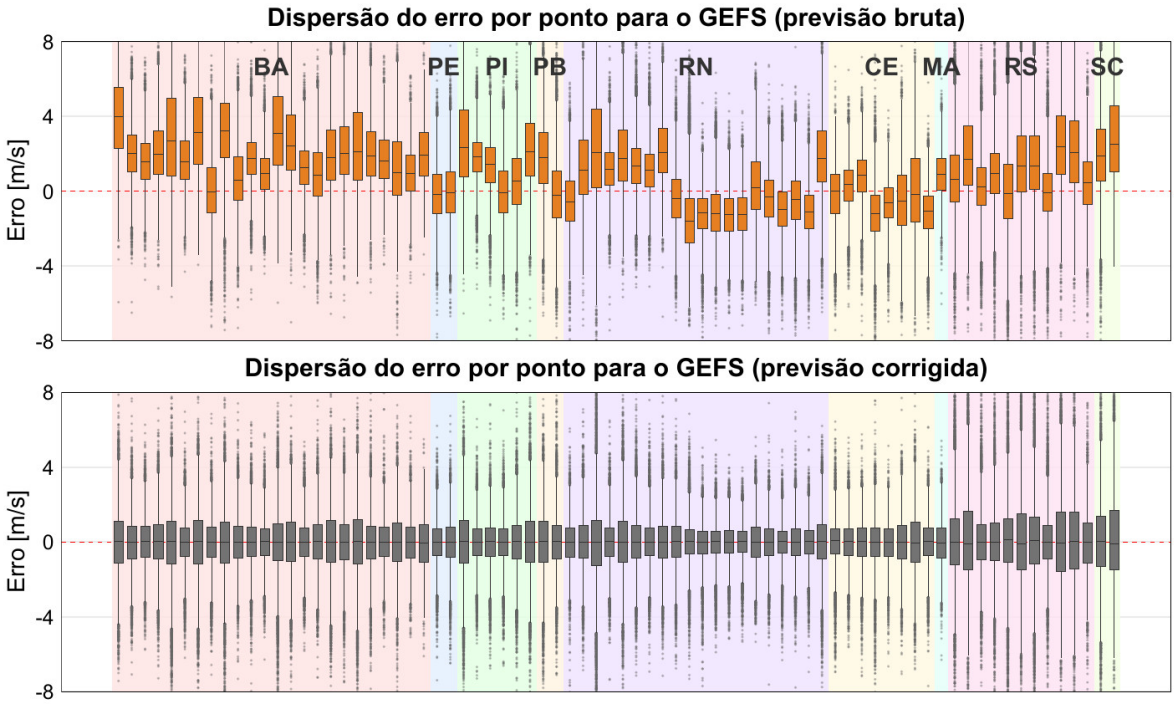
Fonte: Elaboração própria.

Figura 136 – Dispersão do erro de vento do GFS por ponto de análise, antes e após a correção.



Fonte: Elaboração própria.

Figura 137 – Dispersão do erro de vento do GEFS por ponto de análise, antes e após a correção.



Fonte: Elaboração própria.

próximas de zero em grande parte dos pontos; e estados como SC e BA apresentam dispersões mais amplas, refletindo maior dificuldade dos modelos em representar o regime de vento nessas regiões. Entre os três modelos avaliados, o GEFS é o que apresenta a maior amplitude do erro nos extremos (conforme se observa pela escala vertical da Figura 137, que excede a dos demais

modelos), ao passo que ECMWF e GFS exibem distribuições mais concentradas em valores menores.

Os painéis inferiores, correspondentes às previsões corrigidas, evidenciam o impacto da metodologia em escala granular. Em todos os três modelos, observa-se: (i) deslocamento das medianas para próximo de zero, com remoção dos vieses sistemáticos por estado; (ii) redução significativa da amplitude dos *boxplots* (P25–P75) e dos extremos (*whiskers*); e (iii) homogeneização do comportamento entre estados, isto é, após a correção, os pontos de estados anteriormente bem distintos passam a apresentar dispersões de magnitude semelhante. Esses ganhos são observados de forma consistente nos três modelos meteorológicos, indicando robustez da metodologia.

C.3.2 Análise por estado

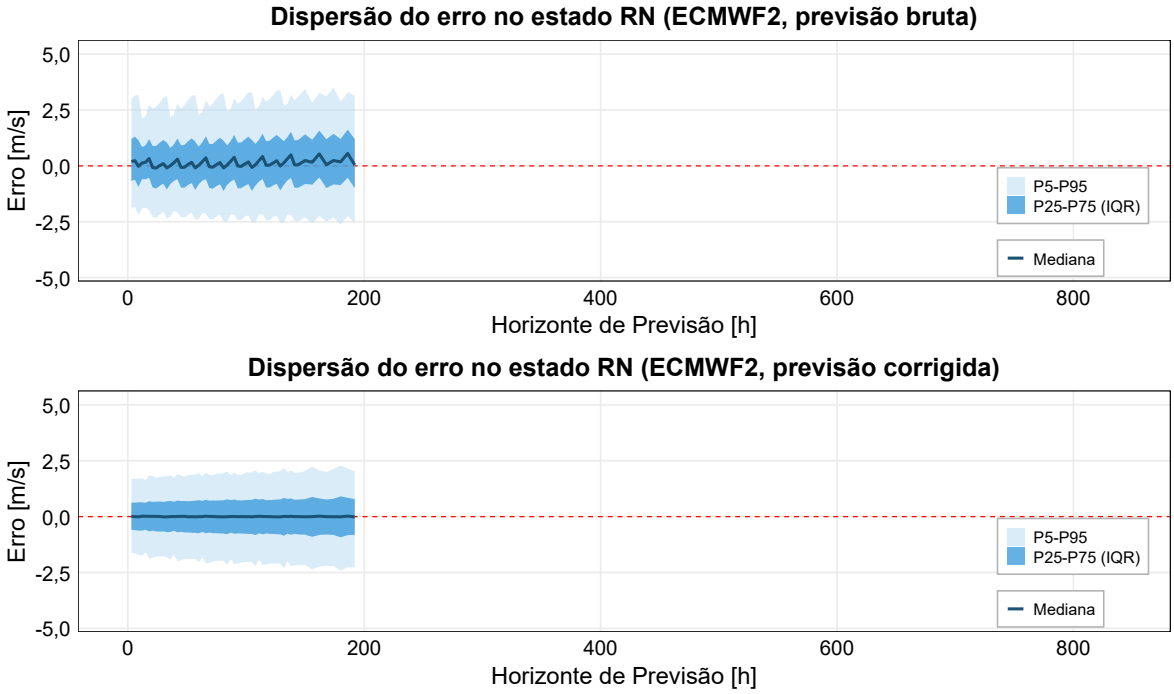
Em escala estadual, a análise considera todos os erros dos pontos pertencentes a um mesmo estado, agregados por horizonte de previsão. Esta subseção apresenta, primeiro, a evolução temporal do erro agregado para os dois estados com maior capacidade instalada de energia eólica no SIN, Rio Grande do Norte e Bahia, antes e após a correção. Em seguida, apresenta-se a distribuição consolidada do erro por estado, para todos os estados analisados e para os três modelos meteorológicos.

C.3.2.1 Comportamento ao longo do horizonte para estados representativos

Nas figuras desta subseção, cada figura reúne, para um dado estado e um dado modelo meteorológico, dois painéis: o painel superior apresenta a dispersão do erro das previsões brutas e o painel inferior, a das previsões corrigidas, ambos ao longo de todo o horizonte de previsão. Essa disposição permite avaliar diretamente, em uma única figura, o efeito da correção sobre o comportamento temporal do erro agregado por estado.

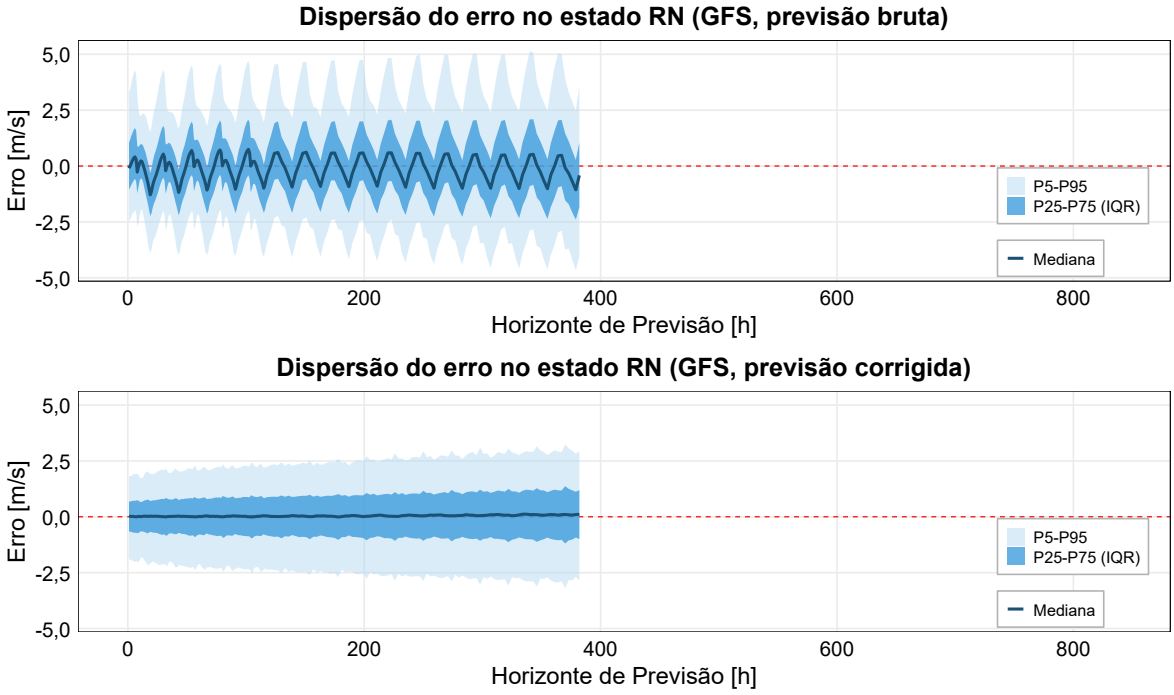
As Figuras 138 a 140 apresentam a dispersão do erro agregada para o Rio Grande do Norte, considerando, respectivamente, os modelos ECMWF, GFS e GEFS. Nas previsões brutas (painéis superiores), a mediana do erro mantém-se próxima de zero para o ECMWF e o GFS e levemente negativa para o GEFS, e observam-se oscilações regulares de período aproximado de 24 horas, associadas à representação imperfeita do ciclo diurno do vento pelos modelos meteorológicos. Essas oscilações preservam-se ao longo de todo o horizonte, confirmando que o ciclo diurno do erro é uma característica robusta do desempenho dos modelos sobre a região, e não um artefato pontual de uma única usina. Nas previsões corrigidas (painéis inferiores), a mediana permanece próxima de zero em todos os modelos e horizontes, as oscilações diárias são substancialmente atenuadas e a amplitude das bandas de dispersão (P25–P75 e P5–P95) é sensivelmente reduzida em relação ao cenário sem correção.

Figura 138 – Dispersão do erro de vento do ECMWF agregada para o Rio Grande do Norte, antes e após a correção.



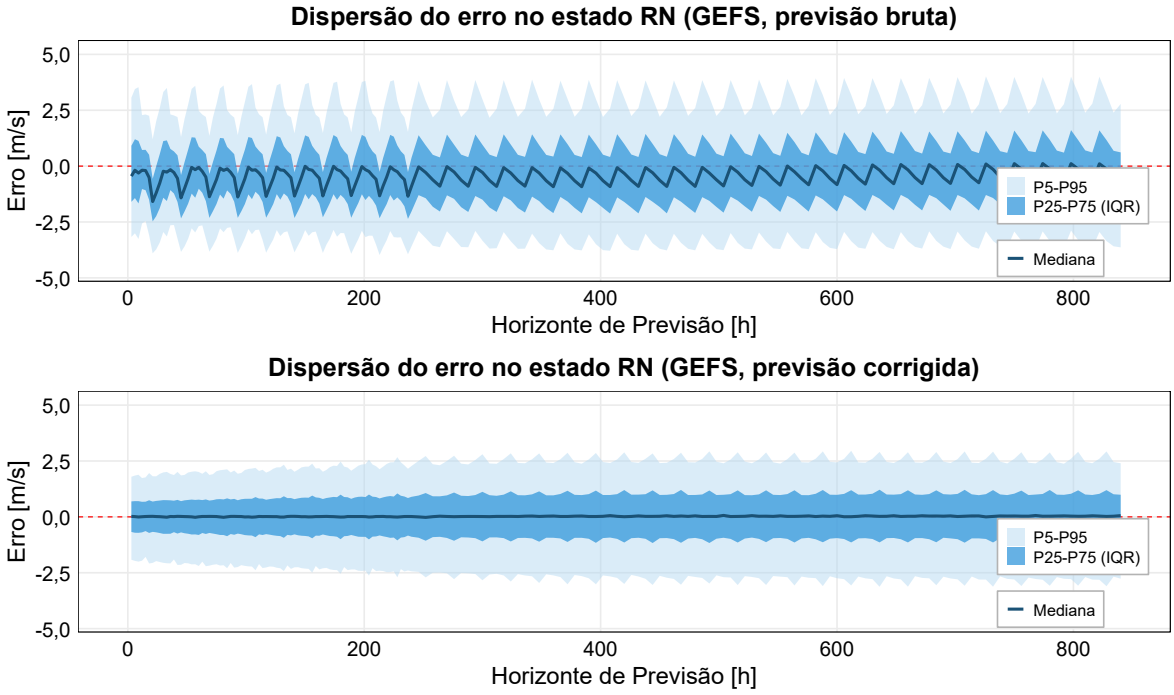
Fonte: Elaboração própria.

Figura 139 – Dispersão do erro de vento do GFS agregada para o Rio Grande do Norte, antes e após a correção.



Fonte: Elaboração própria.

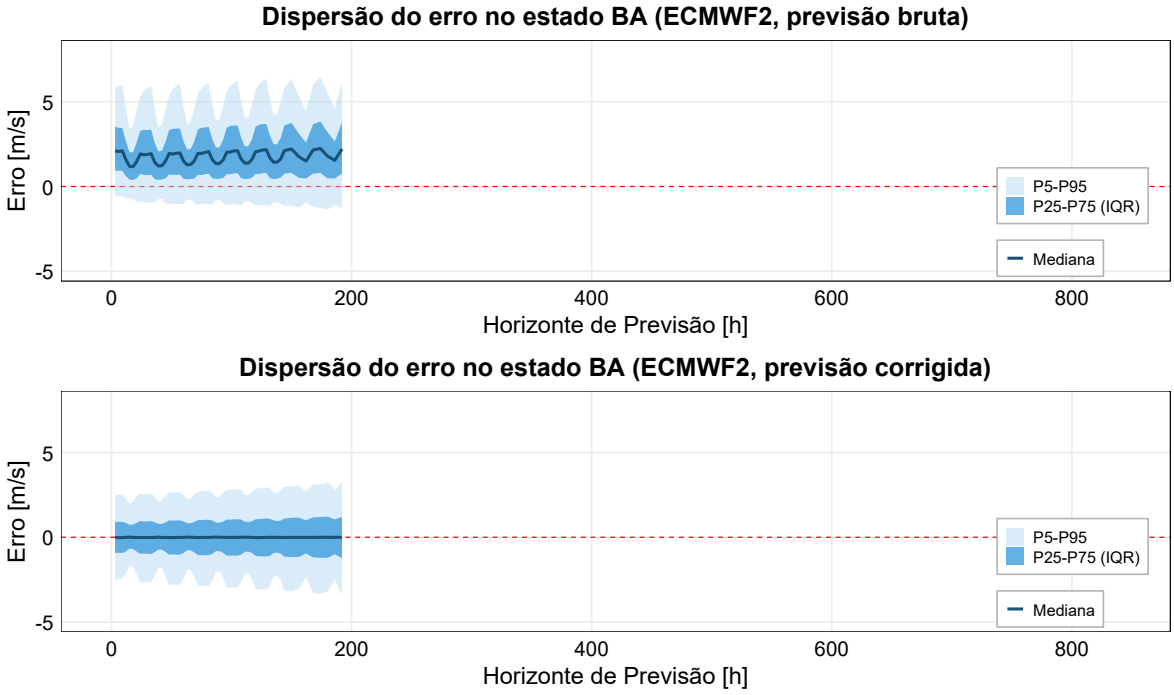
Figura 140 – Dispersão do erro de vento do GEFS agregada para o Rio Grande do Norte, antes e após a correção.



Fonte: Elaboração própria.

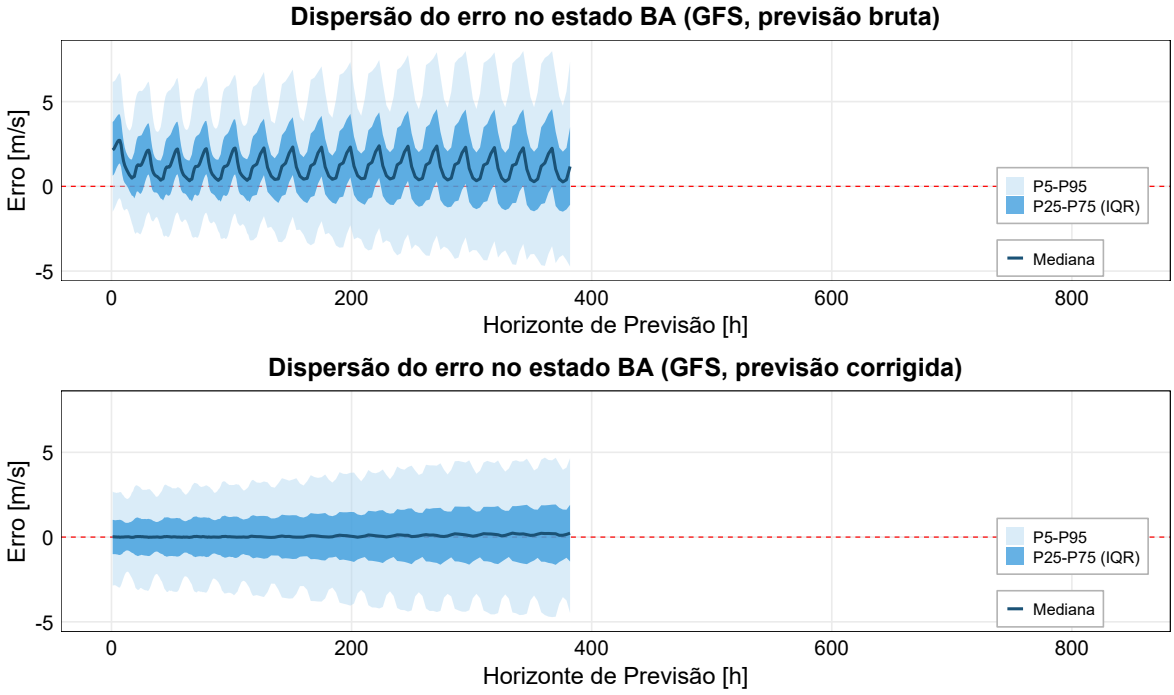
As Figuras 141 a 143 apresentam os mesmos agregados para a Bahia, novamente para os modelos ECMWF, GFS e GEFS. Nas previsões brutas (painéis superiores), observa-se um viés positivo predominante, indicando subestimação sistemática do vento previsto, comportamento notadamente distinto do verificado no Rio Grande do Norte. As oscilações de período diário também estão presentes e, no caso do GEFS, a amplitude das bandas é a mais elevada entre os três modelos. Nas previsões corrigidas (painéis inferiores), a correção remove o viés sistemático: a mediana passa a oscilar em torno de zero, e a magnitude dos extremos (P5–P95) reduz-se expressivamente. Como esperado, a dispersão das bandas tende a aumentar de forma gradual com o alongamento do horizonte de previsão, refletindo a perda natural de previsibilidade; ainda assim, o erro corrigido permanece em uma faixa significativamente menor do que o erro bruto em todos os horizontes, confirmando que o ganho proporcionado pela correção se mantém mesmo nos horizontes mais longos. Cabe destacar que estados com comportamentos brutos bastante distintos — o Rio Grande do Norte, próximo da neutralidade, e a Bahia, com forte subestimação — passam, após a correção, a exibir distribuições de erro muito semelhantes, evidenciando a capacidade da metodologia de equalizar o desempenho dos modelos meteorológicos em diferentes regimes regionais.

Figura 141 – Dispersão do erro de vento do ECMWF agregada para a Bahia, antes e após a correção.



Fonte: Elaboração própria.

Figura 142 – Dispersão do erro de vento do GFS agregada para a Bahia, antes e após a correção.

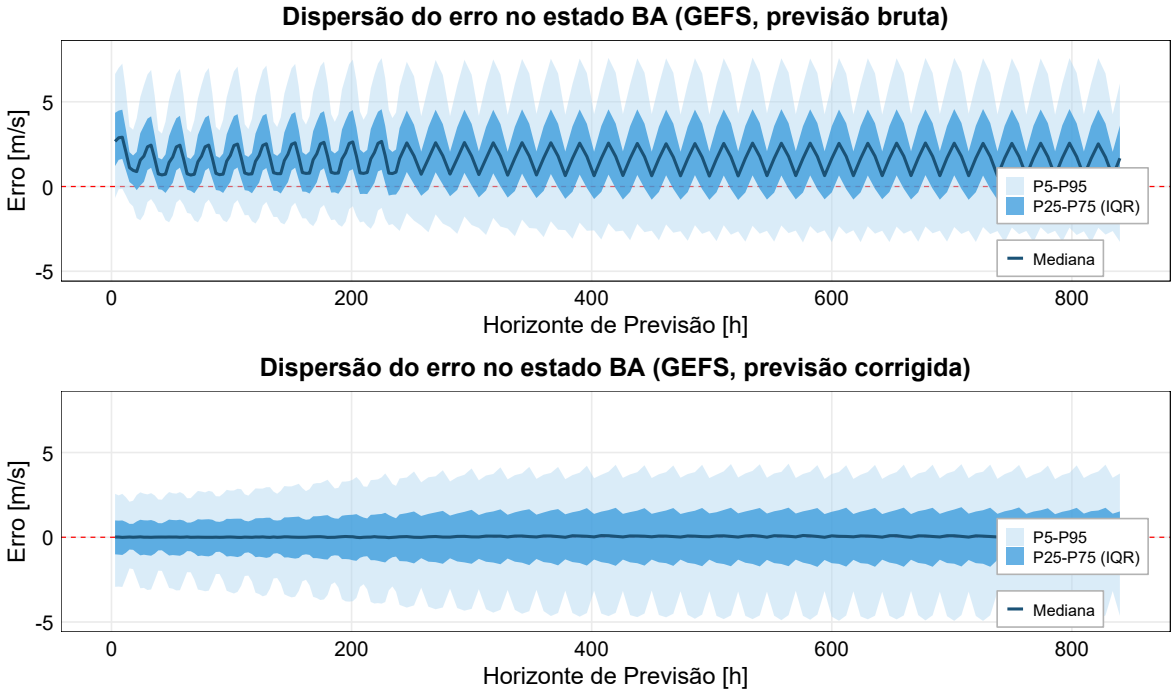


Fonte: Elaboração própria.

C.3.2.2 Análise consolidada para todos os estados

Para verificar se os ganhos observados em RN e BA se replicam nos demais estados, as Figuras 144, 145 e 146 apresentam, para cada modelo meteorológico, a distribuição agregada

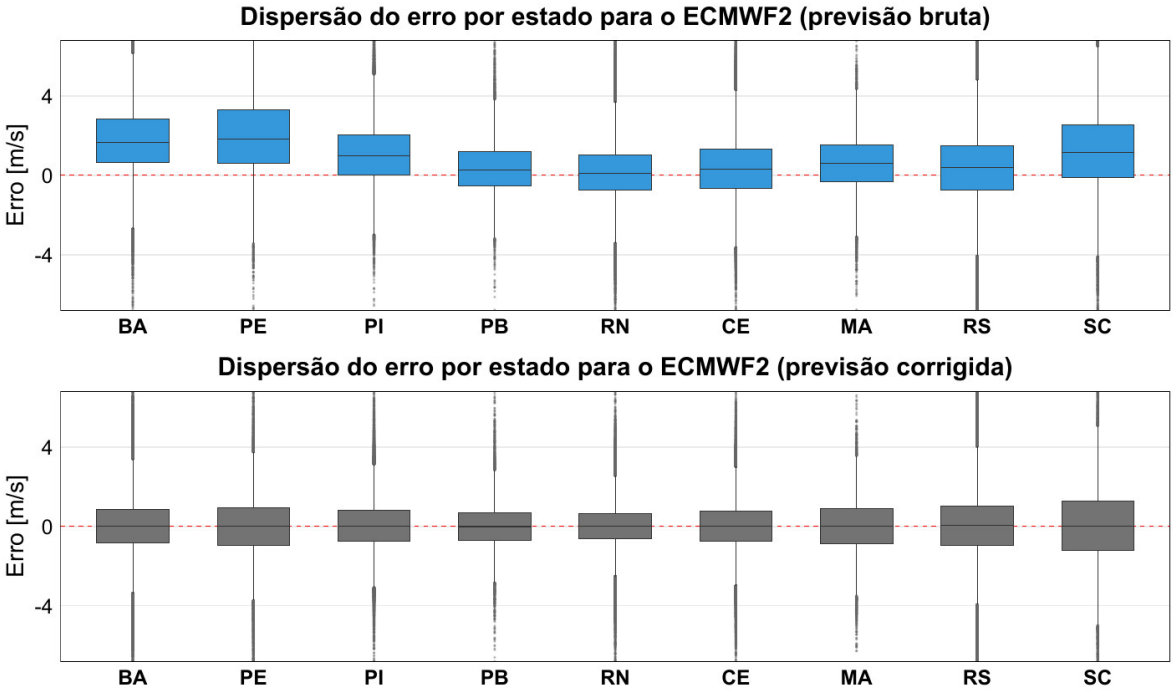
Figura 143 – Dispersão do erro de vento do GEFS agregada para a Bahia, antes e após a correção.



Fonte: Elaboração própria.

do erro por estado, antes (painel superior) e após (painel inferior) a correção, considerando o horizonte das primeiras 24 horas.

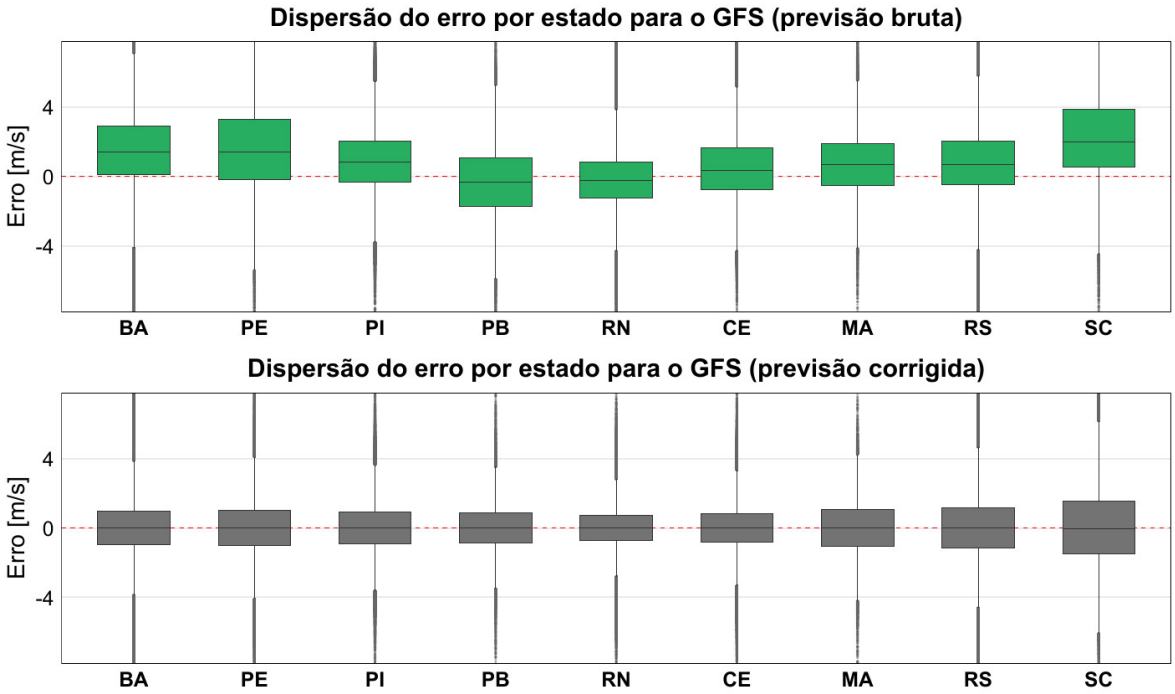
Figura 144 – Dispersão do erro de vento do ECMWF agregada por estado, antes e após a correção.



Fonte: Elaboração própria.

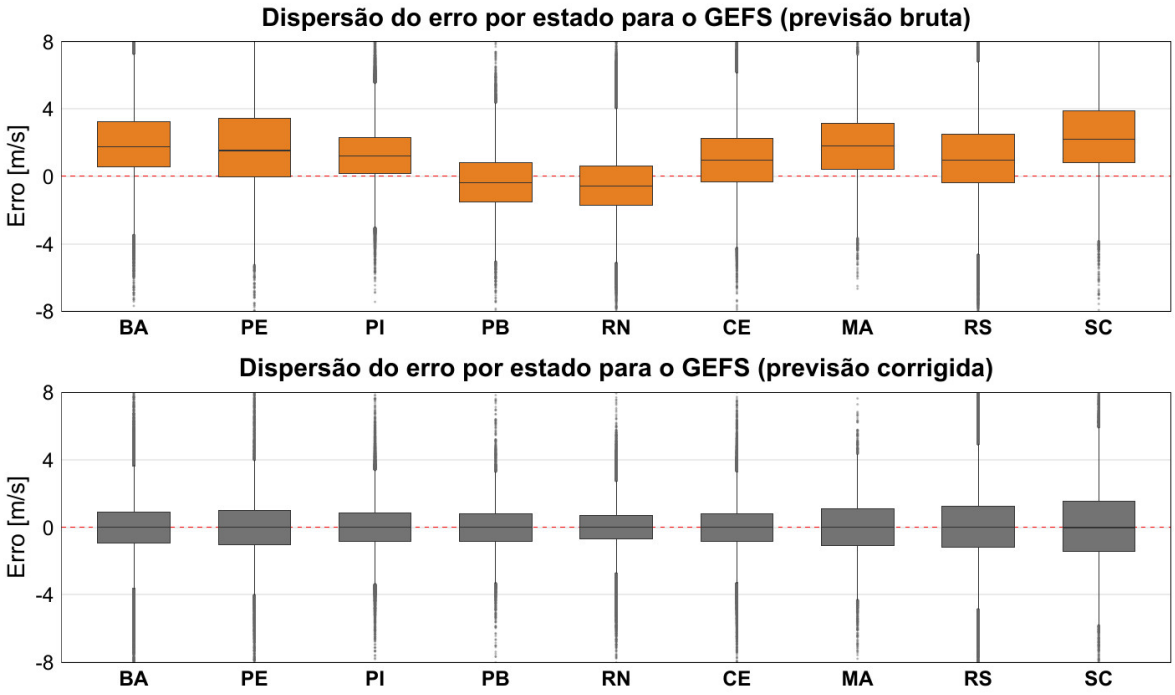
Os painéis superiores confirmam o comportamento heterogêneo entre estados nas pre-

Figura 145 – Dispersão do erro de vento do GFS agregada por estado, antes e após a correção.



Fonte: Elaboração própria.

Figura 146 – Dispersão do erro de vento do GEFS agregada por estado, antes e após a correção.



Fonte: Elaboração própria.

visões brutas: BA, PE e SC exibem o viés positivo mais marcante, sinalizando subestimação sistemática; PI, CE, MA e RS apresentam viés levemente positivo; e RN e PB concentram suas medianas próximas de zero, com leve tendência de superestimação. PI, CE e RS apresentam, ainda, dispersões mais amplas. Entre os modelos, o GEFS novamente se destaca pela maior

amplitude do erro, conforme indicado pela escala vertical da Figura 146.

Nos painéis inferiores, observam-se três efeitos consistentes em todos os três modelos: as medianas dos estados convergem para próximo de zero, removendo o viés sistemático; a amplitude dos *boxplots* é substancialmente reduzida em comparação com o cenário sem correção; e a heterogeneidade entre estados é atenuada, com a magnitude dos erros se tornando muito mais homogênea ao longo do território. Esse último ponto é particularmente relevante: estados que, antes da correção, apresentavam comportamentos bastante distintos, como BA com forte subestimação e RN próximo da neutralidade, passam a exibir distribuições muito similares após a correção, evidenciando a capacidade da metodologia em equalizar o desempenho dos modelos meteorológicos em diferentes regimes regionais.

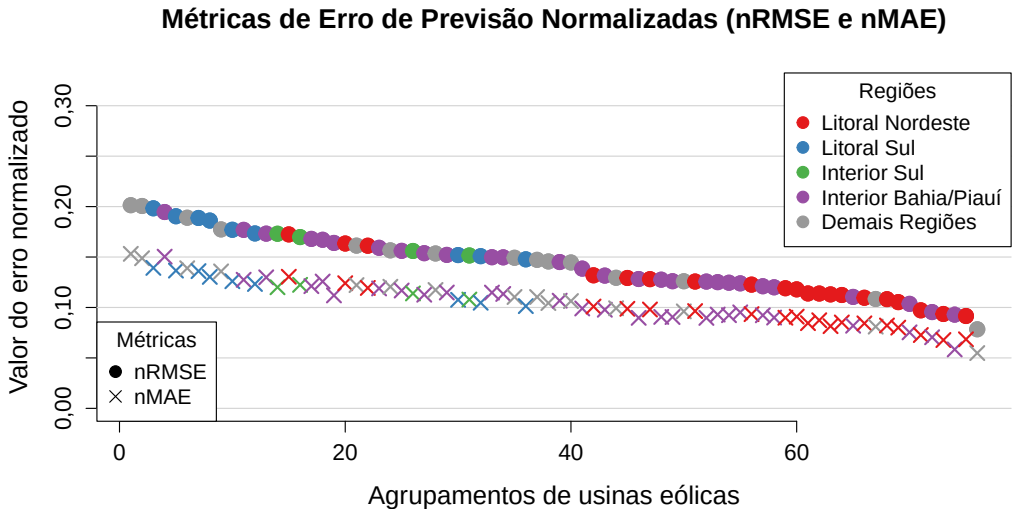
C.4 Resultados da previsão de geração por agrupamentos

Os resultados do modelo de previsão de geração eólica são apresentados a seguir. Para possibilitar a comparação entre os agrupamentos de usinas, que possuem capacidades de geração significativamente diferentes, os erros foram normalizados pela amplitude de cada série temporal, definida como a diferença entre seus valores máximo e mínimo. Essa normalização torna os erros comparáveis entre grupos de diferentes portes e características operacionais. Além disso, assim como na avaliação do modelo de estimação em nível de grupos de usinas, os erros foram calculados considerando exclusivamente os períodos em que não havia restrição operativa, uma vez que essas condições não representam o comportamento padrão de operação e sua inclusão comprometeria a análise de desempenho do modelo de previsão.

A Figura 147 apresenta as métricas de erro normalizadas (nRMSE e nMAE) da previsão de geração para o primeiro dia de horizonte, utilizando o modelo meteorológico ECMWF, para todos os agrupamentos de usinas considerados. Os agrupamentos estão ordenados de forma decrescente pelo valor do nRMSE e diferenciados por cores de acordo com a sub-região geográfica a que pertencem, conforme a distribuição apresentada na Figura 3 do Capítulo 2. Diferentemente dos resultados apresentados na Figura 129 para o modelo de estimação da geração, que não evidenciava uma relação direta entre o erro médio de cada grupo e sua localização geográfica, para a previsão observa-se uma relação mais clara com a região. Nota-se uma tendência de erros maiores em agrupamentos localizados no subsistema Sul, enquanto os menores erros se concentram em agrupamentos situados na região Nordeste. Esse comportamento é consistente com a maior variabilidade do regime de ventos na região Sul e com a maior previsibilidade dos padrões atmosféricos no Nordeste.

A Figura 148 apresenta uma comparação entre o desempenho do modelo de estimação, que utiliza o vento verificado como entrada, e o modelo de previsão de geração, que utiliza a previsão meteorológica do ECMWF para o primeiro dia de horizonte. A métrica nRMSE é exibida para cada agrupamento de usinas, com os valores do modelo de estimação representados em cinza e os da previsão em salmão. Conforme esperado, os erros da previsão de geração são

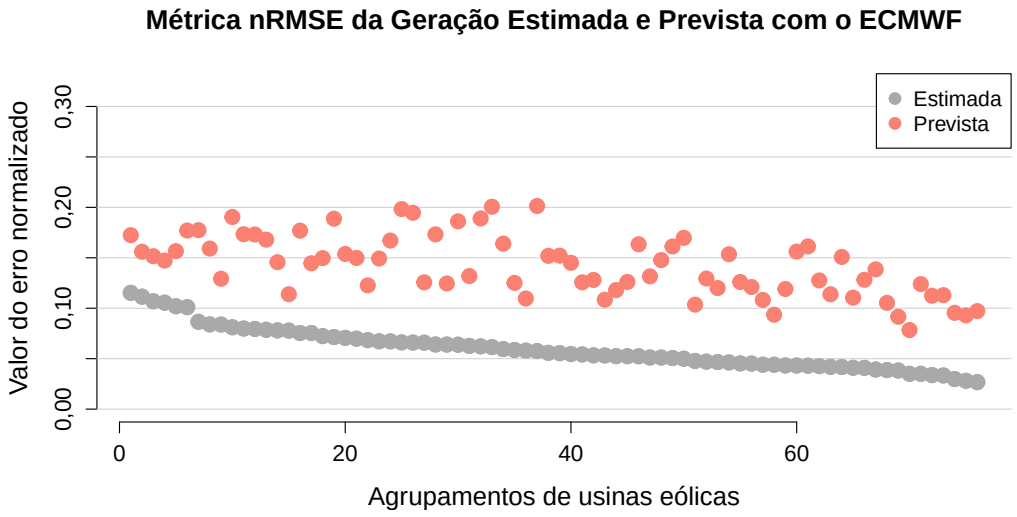
Figura 147 – Métricas de erro normalizadas pela amplitude da série (nRMSE e nMAE) da previsão de geração com o ECMWF para o primeiro dia de previsão, por agrupamento de usinas.



Fonte: Elaboração própria.

consideravelmente superiores aos do modelo de estimação, uma vez que incorporam a incerteza inerente à previsão meteorológica. No entanto, é importante observar que, mesmo quando o vento verificado é conhecido, o modelo de estimação ainda apresenta erros não desprezíveis, representados pelos valores em cinza. Esses erros residuais são atribuídos ao próprio processo de conversão de vento em potência, que envolve simplificações na modelagem aerodinâmica, efeitos de esteira e variabilidades locais não capturadas integralmente pelo modelo. Essa parcela do erro constitui um limite inferior de desempenho que independe da qualidade da previsão meteorológica.

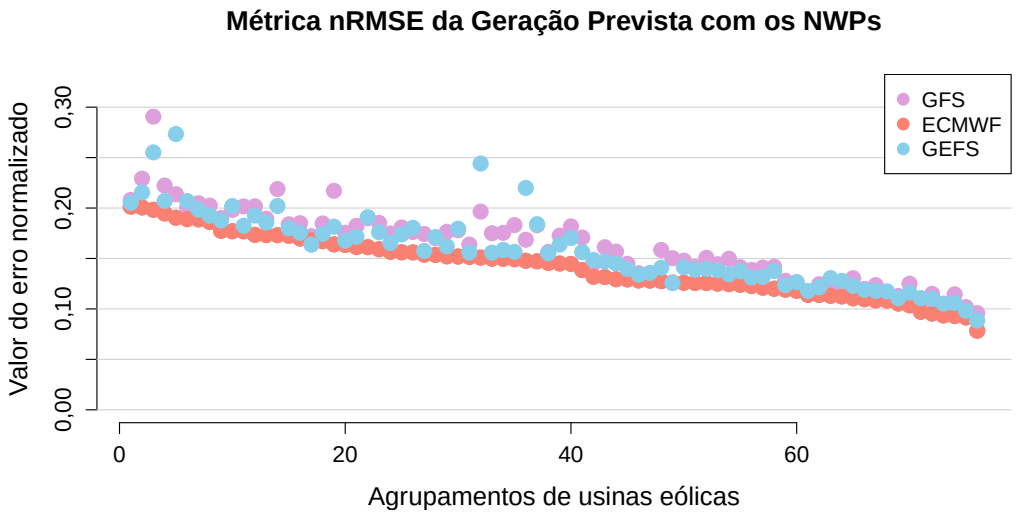
Figura 148 – Comparação do nRMSE, normalizado pela amplitude da série, entre o modelo de estimação e a previsão de geração com o ECMWF para o primeiro dia de previsão.



Fonte: Elaboração própria.

A Figura 149 apresenta a comparação do nRMSE da previsão de geração para o primeiro dia de horizonte utilizando os três modelos meteorológicos avaliados: GFS, ECMWF e GEFS. Os agrupamentos de usinas estão ordenados de forma decrescente pelo nRMSE do ECMWF. A análise permite identificar, para cada agrupamento, qual modelo meteorológico fornece a previsão de geração com menor erro. Observa-se que o ECMWF tende a apresentar desempenho superior ou equivalente aos demais modelos na maioria dos agrupamentos, embora em casos específicos o GFS ou o GEFS possam apresentar resultados competitivos. A comparação entre os três modelos evidencia que a escolha do modelo meteorológico tem impacto relevante na qualidade da previsão de geração e que esse impacto varia conforme as características de cada agrupamento.

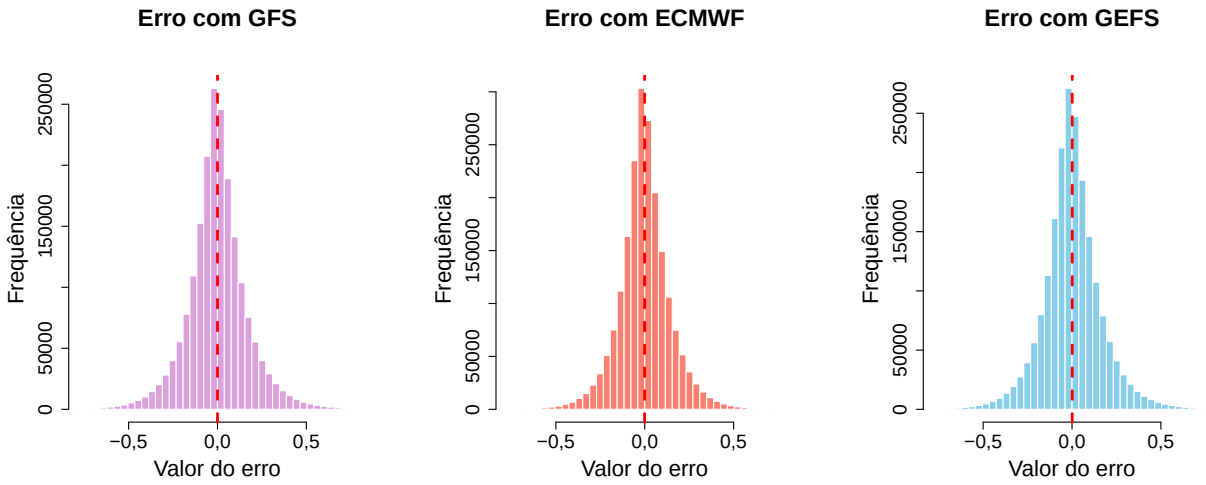
Figura 149 – Comparação do nRMSE, normalizado pela amplitude da série, da previsão de geração entre os modelos GFS, ECMWF e GEFS para o primeiro dia de previsão.



Fonte: Elaboração própria.

A Figura 150 apresenta os histogramas do erro normalizado de previsão, calculado em resolução semi-horária, considerando todos os agrupamentos de usinas de forma conjunta, para cada um dos três modelos meteorológicos. O erro é definido como a diferença entre a geração verificada e a geração prevista, normalizada pela amplitude. A linha tracejada em vermelho indica o valor médio do erro para cada modelo. As três distribuições apresentam-se aproximadamente centradas em torno de zero, indicando ausência de viés sistemático significativo nos modelos de previsão quando avaliados de forma agregada. A dispersão dos histogramas reflete a incerteza inerente à previsão meteorológica na resolução semi-horária, sendo esta consideravelmente maior do que a observada para o modelo de estimação. A comparação visual entre os três painéis permite identificar diferenças na dispersão e na simetria das distribuições de erro entre os modelos, contribuindo para a avaliação comparativa de seu desempenho.

Figura 150 – Distribuição do erro normalizado de previsão semi-horário para os modelos GFS, ECMWF e GEFS.



Fonte: Elaboração própria.

C.5 Resultados dos modelos de tempo real por agrupamentos

A avaliação dos modelos de tempo real, cujo *ensemble* e estratégias de combinação foram descritos na Seção 5.10 do Capítulo 5, é apresentada para dois agrupamentos representativos de comportamentos distintos do recurso eólico: **Acaraú II (69 kV)**, um sítio litorâneo com ciclo diário de brisa marcado e variabilidade moderada, e **Atlântica**, de geração mais volátil e sujeita a mudanças de regime mais bruscas.

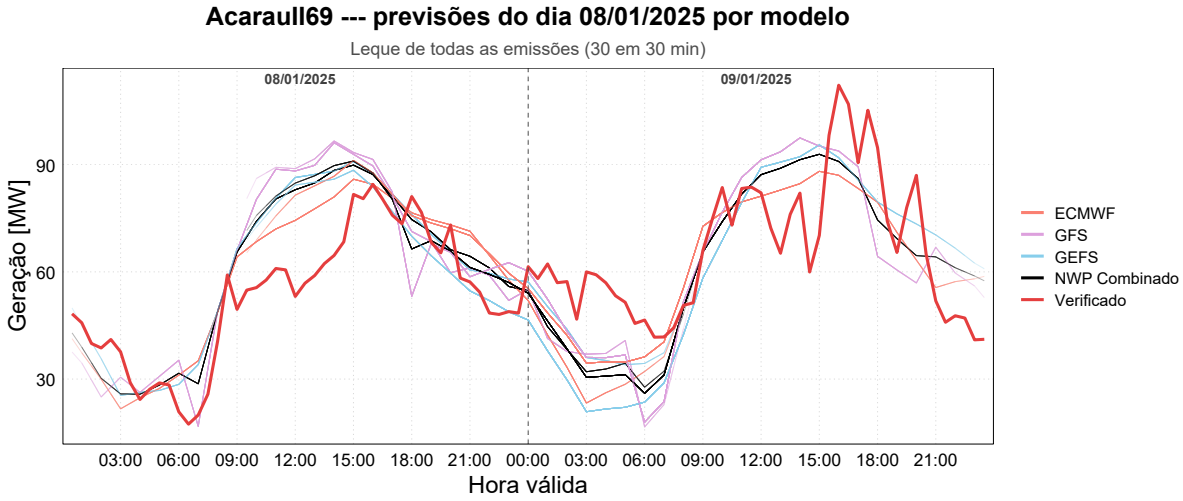
Para todos os modelos, o horizonte de previsão é de 24 horas (48 passos de 30 minutos) e a janela de avaliação compreende o mês de janeiro de 2025. As figuras de leque diário ilustram o dia 08/01/2025, escolhido por conter uma transição de regime; as métricas agregadas (MAE por horizonte e distribuição do erro) consolidam toda a janela. O erro é normalizado pela geração máxima observada no período em cada agrupamento, de modo a permitir a comparação direta entre sítios de portes diferentes.

C.5.1 A combinação das previsões de geração baseadas nos modelos NWP

A base sobre a qual os modelos de tempo real operam é a combinação das previsões de geração baseadas nos diferentes modelos de previsão numérica de tempo (NWP). A Figura 151 apresenta, para o dia representativo selecionado, as três fontes individuais (ECMWF, GFS e GEFS) e a combinação resultante (NWP Combinado) para o ponto Acaraú II. Observa-se que a combinação se posiciona de forma central em relação às fontes individuais, suavizando os desvios particulares de cada modelo e reproduzindo de maneira mais estável o ciclo diário da geração. Em nenhum instante a combinação acompanha os extremos isolados de um único NWP, o que reduz a sensibilidade a erros idiossincráticos de qualquer fonte e fornece um sinal de referência mais robusto, que constitui o insumo corrigido pelos modelos de tempo real, conforme

discutido na Seção 6.5 do Capítulo 6.

Figura 151 – Previsões de geração baseadas nos modelos NWP individuais (ECMWF, GFS, GEFS) e sua combinação (NWP Combinado) no ponto Acaraú II para o dia 08/01/2025.

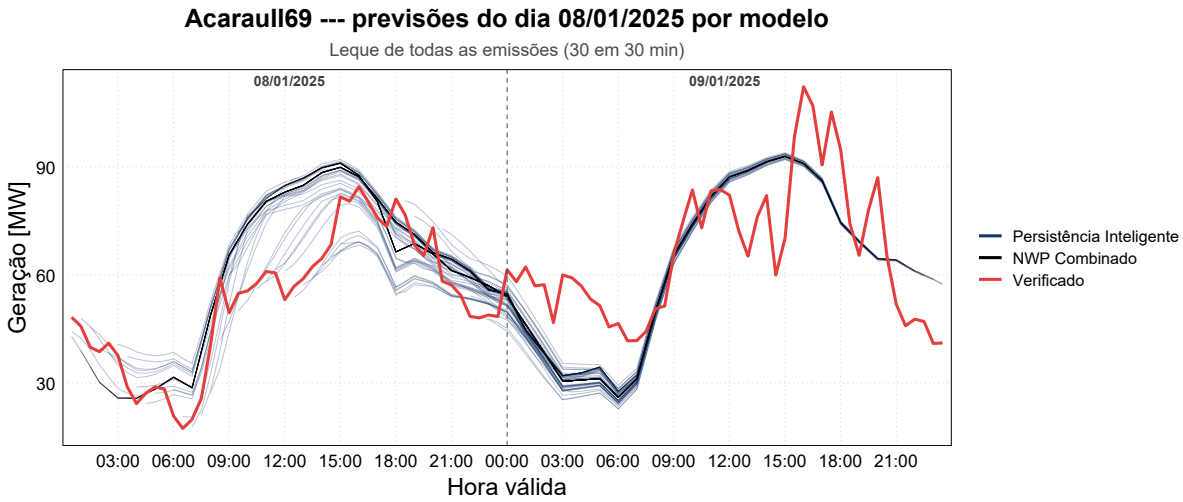


Fonte: Elaboração própria.

C.5.2 Comportamento dos modelos ao longo do horizonte

Um aspecto central do projeto dos modelos de tempo real é a transição entre a informação observada recentemente, dominante nos primeiros passos, e a previsão de geração baseada nos modelos NWP, dominante nos horizontes mais longos. As Figuras 152 e 153 ilustram esse comportamento em dois casos. Na Figura 152, o leque de emissões da Persistência Inteligente parte de correções ancoradas na geração recente e, à medida que o horizonte avança, converge para o NWP Combinado, sobre o qual as curvas se sobrepõem quase integralmente na porção final de cada emissão.

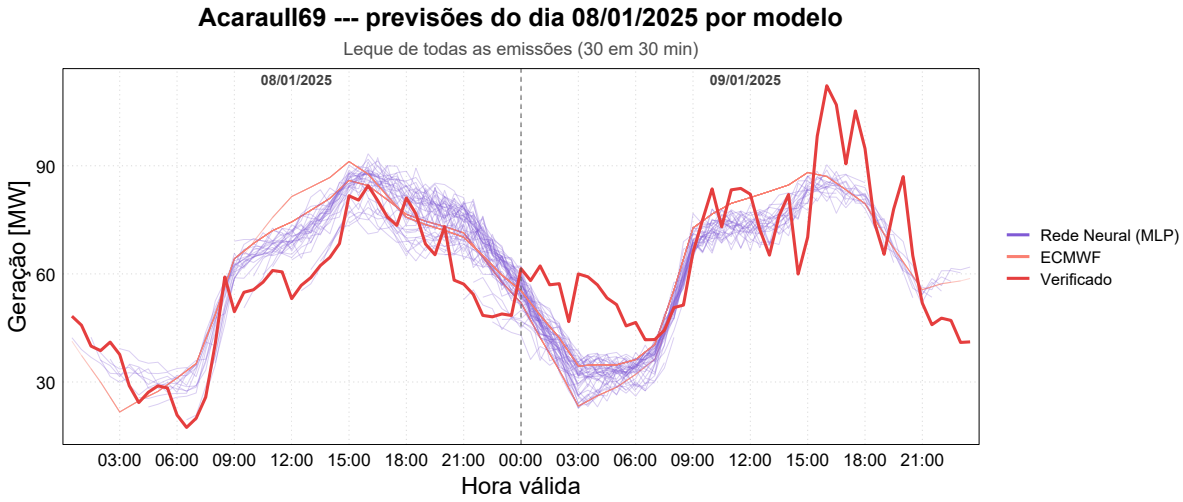
Figura 152 – Leque de previsões da Persistência Inteligente convergindo ao NWP Combinado no ponto Acaraú II para o dia 08/01/2025.



Fonte: Elaboração própria.

Na Figura 153, a Rede Neural (MLP) apresenta dinâmica análoga, porém tendendo à previsão do ECMWF nos horizontes finais. Esse comportamento reflete o aprendizado do modelo de que, na ausência de informação observada útil, a previsão de geração baseada nos modelos NWP é o melhor estimador disponível.

Figura 153 – Leque de previsões da Rede Neural (MLP) tendendo ao ECMWF nos horizontes finais no ponto Acaraú II para o dia 08/01/2025.



Fonte: Elaboração própria.

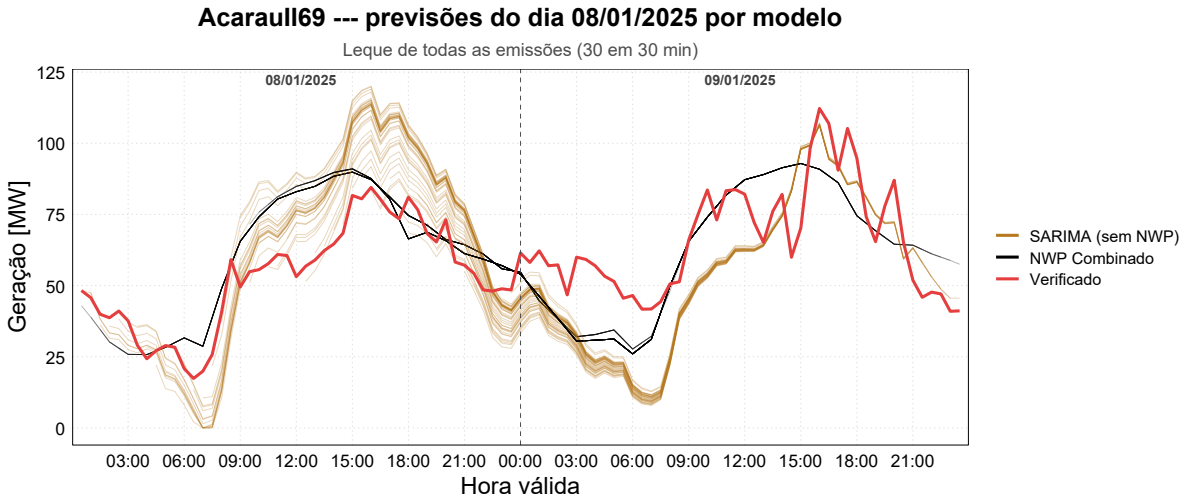
Vale salientar que ambas as figuras evidenciam que os modelos de tempo real não têm como objetivo substituir a previsão meteorológica de referência, mas sim corrigir seus desvios nas primeiras horas do horizonte de previsão. À medida que o horizonte avança, a contribuição da correção é gradualmente reduzida, fazendo com que a previsão converja de forma suave para a previsão original. Esse comportamento é observado de maneira consistente em todos os modelos que utilizam previsões NWP como variáveis explicativas, preservando a coerência da informação meteorológica no final do horizonte ao mesmo tempo em que aprimoram a qualidade das previsões nos horizontes iniciais.

C.5.3 O modelo de controle sem NWP (SARIMA)

Para isolar o papel da informação meteorológica proveniente dos modelos NWP, foi incluído um modelo de controle puramente univariado: um SARIMA ajustado apenas à série histórica de geração, sem qualquer informação meteorológica. As Figuras 154 e 155 contrastam seu desempenho nos dois agrupamentos. Em Acaraú II (Figura 154), onde o ciclo diário é regular, o SARIMA reproduz a forma geral do dia, ainda que superestime a amplitude nos horários de pico. Seu desempenho é aceitável apenas nos horizontes muito curtos e em sítios com sazonalidade marcada e baixa variabilidade.

Em Atlântica (Figura 155), a limitação fica evidente: na transição de regime do dia, em que a geração verificada se eleva a um novo patamar, o SARIMA, ancorado apenas no padrão sazonal e nos valores recentes, não acompanha a mudança e subestima sistematicamente

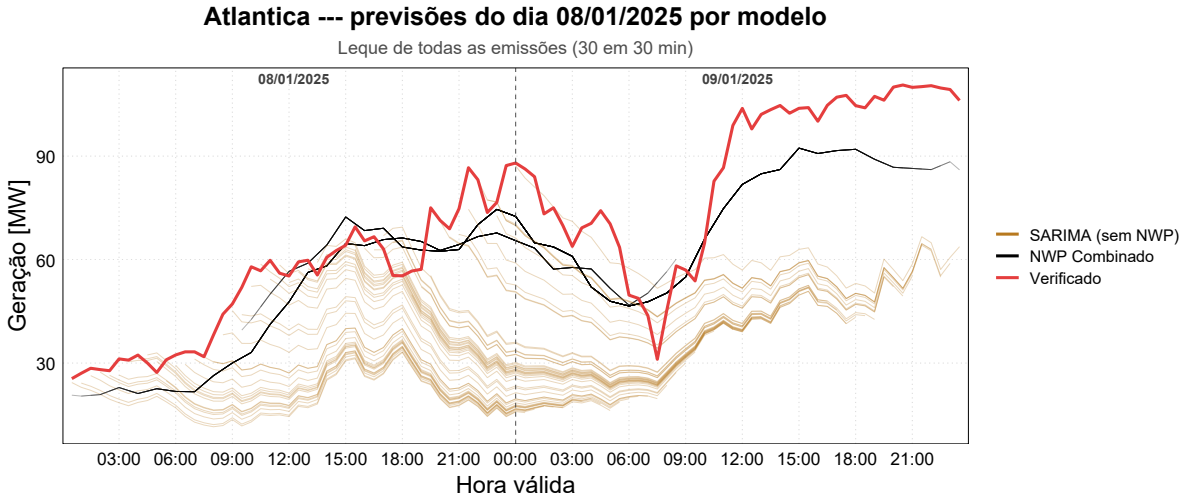
Figura 154 – Previsão do SARIMA em comparação ao NWP Combinado e ao verificado no ponto Acaraú II para o dia 08/01/2025.



Fonte: Elaboração própria.

a geração. Conclui-se que um modelo sem NWP só é competitivo quando o padrão diário é repetível e a variabilidade é baixa, ou em horizontes muito curtos; fora dessas condições, a previsão numérica é indispensável.

Figura 155 – Previsão do SARIMA em comparação ao NWP Combinado e ao verificado no ponto Atlântica para o dia 08/01/2025.



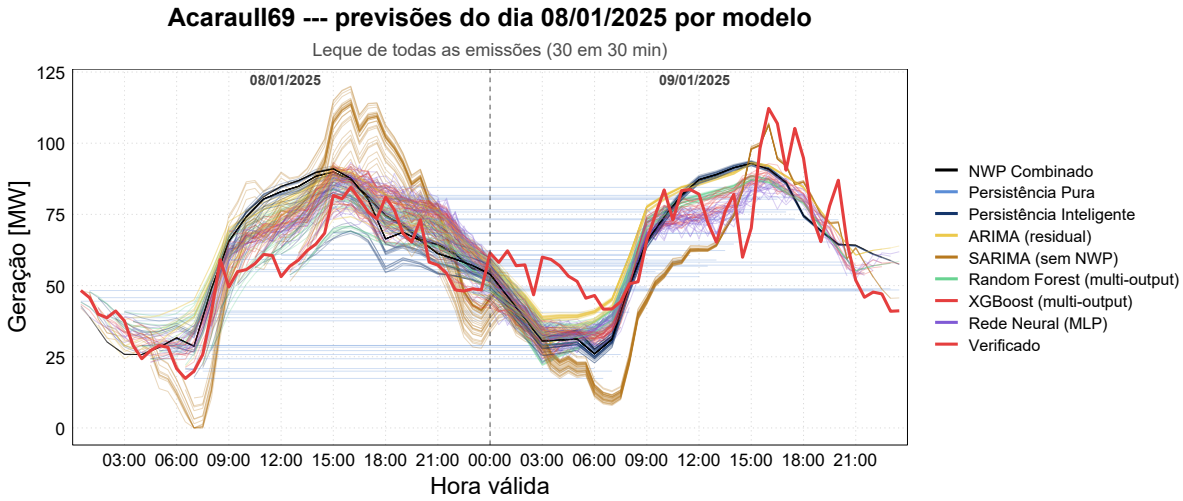
Fonte: Elaboração própria.

C.5.4 Visão conjunta dos modelos

As Figuras 156 e 157 sobrepõem simultaneamente o leque de emissões de um amplo conjunto de modelos, permitindo observar o comportamento coletivo. Em ambos os agrupamentos, os modelos baseados em NWP (híbridos, com técnicas estatísticas e de aprendizado de máquina) formam um agrupamento coeso em torno do NWP Combinado e acompanham a evolução do verificado, com diferenças sobretudo de amplitude e de defasagem nos instantes

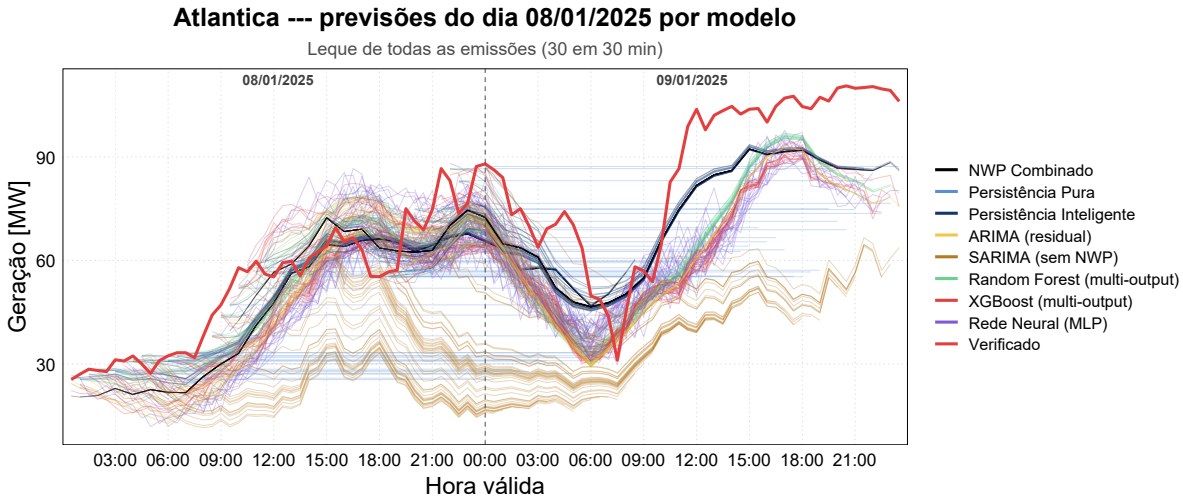
de virada. Destacam-se como exceções a Persistência Pura, que aparece como segmentos horizontais por manter constante o último valor observado, e o SARIMA, cujo leque se afasta do conjunto, subindo demais em Acaraú II e permanecendo abaixo do verificado em Atlântica. As conclusões gerais sobre o comportamento conjunto dos modelos, também observáveis em nível de subsistema, são discutidas na Seção 6.5 do Capítulo 6.

Figura 156 – Comportamento conjunto de múltiplos modelos no ponto Acaraú II para o dia 08/01/2025.



Fonte: Elaboração própria.

Figura 157 – Comportamento conjunto de múltiplos modelos no ponto Atlântica para o dia 08/01/2025.



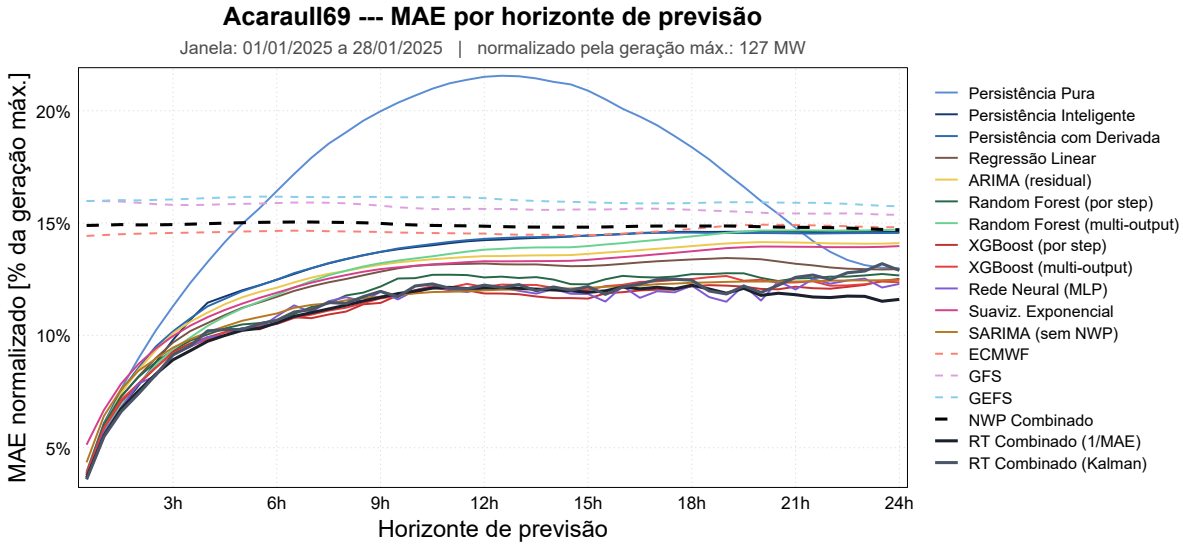
Fonte: Elaboração própria.

C.5.5 Erro médio absoluto por horizonte

As Figuras 158 e 159 apresentam o MAE normalizado em função do horizonte de previsão, consolidando toda a janela de avaliação. Em Acaraú II (Figura 158), a Persistência Pura exibe a curva característica em forma de corcova, com erro crescente até cerca de 22% da geração máxima por volta da metade do horizonte e posterior redução, consequência de o valor mantido voltar a coincidir com o mesmo horário do dia em 24 horas. As previsões de geração baseadas

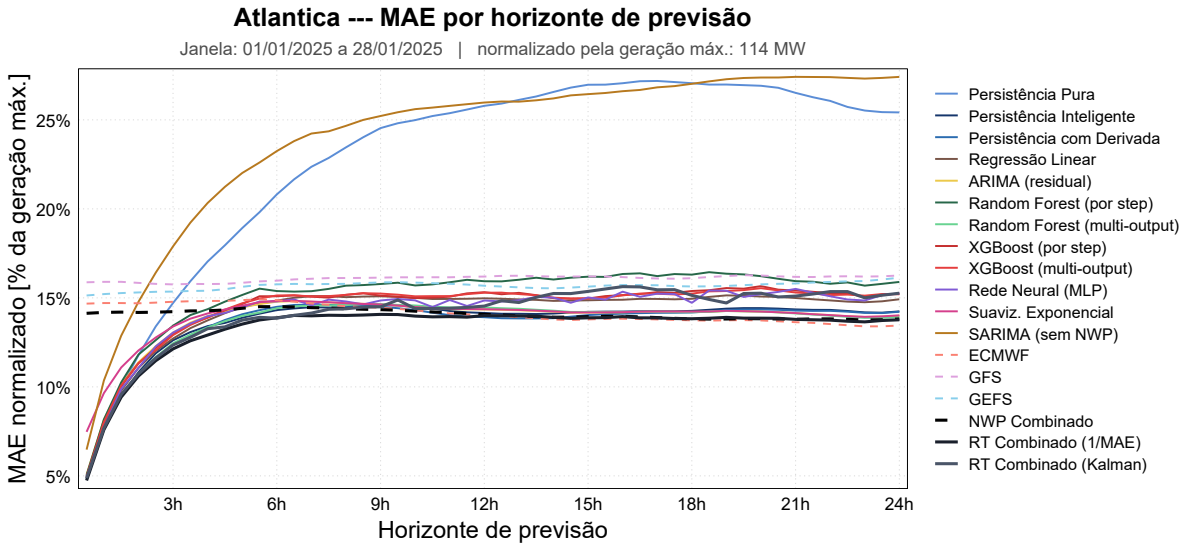
nos modelos NWP (linhas tracejadas) mantêm-se aproximadamente constantes em torno de 14 a 16%, ao passo que as combinações de tempo real (RT Combinado 1/MAE e Kalman) apresentam os menores erros em praticamente todo o horizonte ($\approx$ 12 a 13%), superando tanto as fontes individuais de NWP quanto a sua combinação direta. Nos primeiros passos, todos os modelos convergem a erros baixos ($\approx$ 3 a 5%), refletindo a forte autocorrelação nas primeiras meias horas.

Figura 158 – MAE normalizado pela geração máxima do período em função do horizonte de previsão no ponto Acaraú II.



Fonte: Elaboração própria.

Figura 159 – MAE normalizado pela geração máxima do período em função do horizonte de previsão no ponto Atlântica.



Fonte: Elaboração própria.

Em Atlântica (Figura 159), o patamar geral de erro é mais elevado, coerente com a maior variabilidade do sítio, e o SARIMA sobressai como um dos piores modelos, com erro crescente

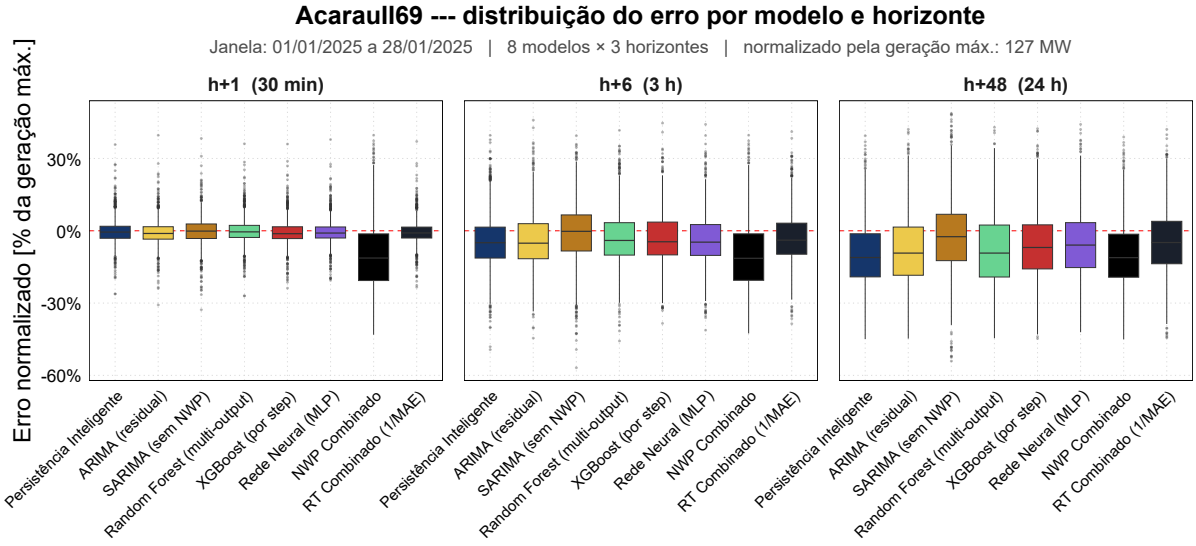
até cerca de 27% da geração máxima, em linha com a Persistência Pura e em contraste com seu desempenho aceitável em Acaraú II. Novamente, as combinações de tempo real situam-se entre os menores erros, confirmando o ganho da estratégia de combinação dinâmica em relação aos modelos individuais.

C.5.6 Distribuição do erro por horizonte

As Figuras 160 e 161 detalham a distribuição do erro (verificado menos previsto, normalizado pela geração máxima) em três horizontes representativos: 30 minutos (h+1), 3 horas (h+6) e 24 horas (h+48). Em ambos os agrupamentos, a dispersão do erro cresce com o horizonte, como esperado. No menor horizonte, a maioria dos modelos concentra-se de forma estreita em torno de zero, indicando baixo viés e pequena dispersão.

Destaca-se, contudo, que o NWP Combinado apresenta viés negativo (mediana abaixo de zero), sobretudo em Acaraú II, evidenciando tendência sistemática de superestimação. Os modelos de tempo real corrigem esse viés ao incorporar a geração verificada recente; consequentemente, a combinação de tempo real (RT Combinado $1/\mathrm{MAE}$), por consolidar esses modelos, apresenta o viés minimizado, com mediana próxima de zero, e mantém-se entre as distribuições mais estreitas, evidenciando o duplo efeito de reduzir tanto o viés quanto a dispersão.

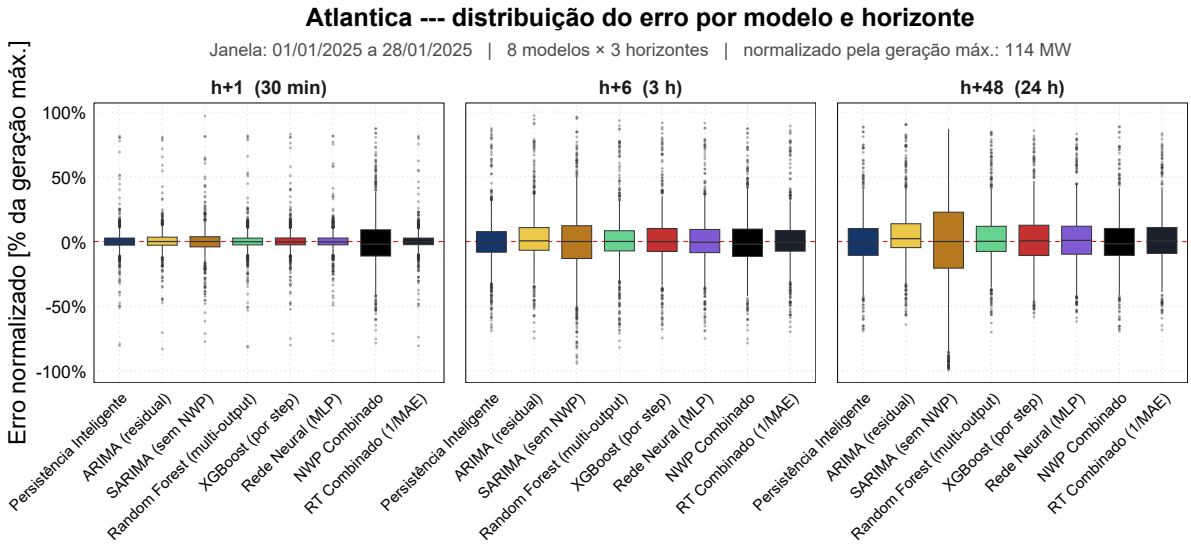
Figura 160 – Distribuição do erro normalizado por modelo nos horizontes de 30 minutos, 3 horas e 24 horas no ponto Acaraú II.



Fonte: Elaboração própria.

O comportamento do SARIMA reaparece de forma marcante: em Atlântica, no horizonte de 24 horas (Figura 161), sua distribuição é nitidamente a mais larga, com dispersão e amplitude interquartílica muito superiores às dos demais modelos, confirmando a instabilidade do modelo sem NWP em sítios voláteis. Em Acaraú II (Figura 160), as distribuições são mais homogêneas entre os modelos, mas mantêm-se a superioridade das combinações de tempo real em termos de viés e dispersão.

Figura 161 – Distribuição do erro normalizado por modelo nos horizontes de 30 minutos, 3 horas e 24 horas no ponto Atlântica.



Fonte: Elaboração própria.

C.5.7 Síntese dos resultados por agrupamentos

O conjunto de resultados sustenta três conclusões principais. Em primeiro lugar, a combinação das previsões de geração baseadas nos modelos NWP fornece um sinal de referência mais estável que qualquer fonte individual. Em segundo lugar, os modelos de tempo real agregam valor ao corrigir essa previsão nas primeiras horas, convergindo a ela no final do horizonte de tempo real, e as combinações dinâmicas (1/MAE e Kalman) atingem os menores erros, além de reduzir o viés sistemático do NWP Combinado. Por fim, um modelo sem informação meteorológica (SARIMA) só é competitivo em condições restritas, como horizontes muito curtos ou sítios de sazonalidade regular e baixa variabilidade; ele falha justamente onde a previsão é mais difícil e onde, portanto, o valor da informação numérica de tempo é maior.