

TEXTO PARA DISCUSSÃO

2955

**GERAÇÃO DE SÉRIES SINTÉTICAS
DE VELOCIDADE DO VENTO
POR MEIO DO MODELO
FATORIAL DINÂMICO**

**MÁRIO JORGE CARDOSO DE MENDONÇA
JOSÉ FRANCISCO MOREIRA PESSANHA
LUIZ ALBERTO TOSCANO MEDRANO
VICTOR ANDRADE DE ALMEIDA
AMARO OLÍMPIO PEREIRA JUNIOR
ÉRICA CARVALHO NOGUEIRA**

ipea

Instituto de Pesquisa
Econômica Aplicada

GERAÇÃO DE SÉRIES SINTÉTICAS DE VELOCIDADE DO VENTO POR MEIO DO MODELO FATORIAL DINÂMICO

MÁRIO JORGE CARDOSO DE MENDONÇA¹

JOSÉ FRANCISCO MOREIRA PESSANHA²

LUIZ ALBERTO TOSCANO MEDRANO³

VICTOR ANDRADE DE ALMEIDA⁴

AMARO OLÍMPIO PEREIRA JUNIOR⁵

ÉRICA CARVALHO NOGUEIRA⁶

1. Técnico de planejamento e pesquisa na Diretoria de Estudos e Políticas Regionais, Urbanas e Ambientais do Instituto de Pesquisa Econômica Aplicada (Dirur/Ipea). *E-mail:* mario.mendonca@ipea.gov.br.

2. Professor adjunto da Universidade do Estado do Rio de Janeiro (UERJ) e pesquisador do Centro de Pesquisa de Energia Elétrica (Cepel). *E-mail:* francisc@cepel.br.

3. Professor adjunto do Departamento de Matemática da Universidade Federal Rural do Rio de Janeiro (UFRRJ) e assistente de pesquisa na Dirur/Ipea. *E-mail:* lmedrano@ufrj.br.

4. Pesquisador do Cepel. *E-mail:* andrade@cepel.br.

5. Professor adjunto do Programa de Planejamento Energético do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia da Universidade Federal do Rio de Janeiro (Coppe/UFRJ). *E-mail:* amaro@ppe.ufrj.br.

6. Doutoranda do Programa de Planejamento Energético da Coppe/UFRJ. *E-mail:* erika@ppe.ufrj.br.

Governo Federal

Ministério do Planejamento e Orçamento

Ministra Simone Nassar Tebet

ipea Instituto de Pesquisa
Econômica Aplicada

Fundação pública vinculada ao Ministério do Planejamento e Orçamento, o Ipea fornece suporte técnico e institucional às ações governamentais – possibilitando a formulação de inúmeras políticas públicas e programas de desenvolvimento brasileiros – e disponibiliza, para a sociedade, pesquisas e estudos realizados por seus técnicos.

Presidenta

LUCIANA MENDES SANTOS SERVO

Diretor de Desenvolvimento Institucional

FERNANDO GAIGER SILVEIRA

**Diretora de Estudos e Políticas do Estado,
das Instituições e da Democracia**

LUSENI MARIA CORDEIRO DE AQUINO

Diretor de Estudos e Políticas Macroeconômicas

CLÁUDIO ROBERTO AMITRANO

**Diretor de Estudos e Políticas Regionais,
Urbanas e Ambientais**

ARISTIDES MONTEIRO NETO

**Diretora de Estudos e Políticas Setoriais,
de Inovação, Regulação e Infraestrutura**

FERNANDA DE NEGRI

Diretor de Estudos e Políticas Sociais

CARLOS HENRIQUE LEITE CORSEUIL

Diretor de Estudos Internacionais

FÁBIO VÉRAS SOARES

Chefe de Gabinete

ALEXANDRE DOS SANTOS CUNHA

Coordenador-Geral de Imprensa e Comunicação Social

ANTONIO LASSANCE

Ouvidoria: <http://www.ipea.gov.br/ouvidoria>

URL: <http://www.ipea.gov.br>

Texto para Discussão

Publicação seriada que divulga resultados de estudos e pesquisas em desenvolvimento pelo Ipea com o objetivo de fomentar o debate e oferecer subsídios à formulação e avaliação de políticas públicas.

© Instituto de Pesquisa Econômica Aplicada – **ipea** 2024

Geração de séries sintéticas de velocidade do vento por meio do modelo fatorial dinâmico / Mário Jorge Cardoso de Mendonça ... [et al.]. – Brasília, DF: IPEA, 2024.

31 p.: il., gráfs. – (Texto para Discussão ; n. 2955).

Inclui Bibliografia.

ISSN 1415-4765

1. Geração Eólica. 2. Parques Eólicos. 3. Séries Sintéticas de Vento. 4. Modelo Fatorial dinâmico. 5. MCMC. I. Mendonça, Mário Jorge Cardoso de. II. Pessanha, José Francisco Moreira, III. Medrano, Luiz Alberto Toscano. IV. Almeida, Víctor Andrade de. V. Pereira Junior, Amaro Olímpio. VI. Nogueira, Érica Carvalho. VII. Instituto de Pesquisa Econômica Aplicada.

CDD 621.4

Ficha catalográfica elaborada por Elizabeth Ferreira da Silva CRB-7/6844.

Como citar:

MENDONÇA, Mário Jorge Cardoso de; PESSANHA, José Francisco Moreira; MEDRANO, Luiz Alberto Toscano; ALMEIDA, Víctor Andrade de; PEREIRA JUNIOR, Amaro Olímpio; NOGUEIRA, Érica Carvalho. **Geração de séries sintéticas de velocidade do vento por meio do modelo fatorial dinâmico**. Brasília, DF: Ipea, jan. 2024. 31 p. (Texto para Discussão, n. 2955). DOI: <http://dx.doi.org/10.38116/td2955-port>.

JEL: Q4; Q42; Q47.

As publicações do Ipea estão disponíveis para download gratuito nos formatos PDF (todas) e ePUB (livros e periódicos).

Acesse: <http://www.ipea.gov.br/portal/publicacoes>

As opiniões emitidas nesta publicação são de exclusiva e inteira responsabilidade dos autores, não exprimindo, necessariamente, o ponto de vista do Instituto de Pesquisa Econômica Aplicada ou do Ministério do Planejamento e Orçamento.

É permitida a reprodução deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções para fins comerciais são proibidas.

SUMÁRIO

SINOPSE

ABSTRACT

1 INTRODUÇÃO	6
2 ESTADO DA ARTE DA MODELAGEM DA INCERTEZA EÓLICA NO PLANEJAMENTO DA OPERAÇÃO DO SIN	8
3 MODELO FATORIAL DINÂMICO	10
3.1 Especificação	10
3.2 Metodologia para estimação do MFD	12
3.3 Previsão do MFD com componente sazonal	15
4 RESULTADOS	15
5 CONCLUSÕES	25
REFERÊNCIAS	26
APÊNDICE A	29

SINOPSE

A geração eólica é fundamental para redução do uso de recursos fósseis e, consequentemente, da emissão de gases de efeito estufa (GEEs). Atualmente, o Sistema Interligado Nacional (SIN) conta com cerca de 9.971 aerogeradores em operação, a sexta maior rede no *ranking* mundial, sendo que no Nordeste estão localizados 80% dos parques eólicos brasileiros. No entanto, devido à intermitência da geração eólica, a entrada em larga escala da fonte eólica no planejamento de sistemas hidrotérmicos implica desafios para sua integração ao sistema. Por esse motivo, há a necessidade do tratamento da incerteza da velocidade do vento na cadeia de modelos computacionais que suportam as decisões operativas. Assim, visando contribuir com a modelagem da incerteza eólica no planejamento da operação (médio prazo), este trabalho propõe uma metodologia que consiste na aplicação do modelo fatorial dinâmico (MFD) para a geração de séries sintéticas das médias mensais da velocidade do vento nas localidades com aproveitamentos eólicos. A metodologia proposta foi aplicada em um conjunto de séries de reanálises da velocidade do vento oriundas do MERRA-2. Os resultados obtidos para os oito parques eólicos do SIN analisados mostram que a metodologia é promissora, dada a boa qualidade das previsões mensais da média mensal da velocidade do vento até dois anos à frente e, sobretudo, a boa representatividade das séries sintéticas geradas pelo modelo.

Palavras-chave: geração eólica; parques eólicos; séries sintéticas de vento; modelo fatorial dinâmico; MCMC.

ABSTRACT

Wind power generation is essential to reduce the use of fossil resources and, consequently, the emission of greenhouse gases (GHG). Currently, the National Interconnected System (SIN) has about 9971 wind turbines in operation, the sixth largest network in the world ranking, and 80% of Brazilian wind parks are located in the Northeast. However, due to the intermittency of wind generation, the large-scale entry of the wind source in the planning of hydrothermal systems implies challenges for its integration into the system. For this reason, there is a need to treat wind speed uncertainty in the chain of computational models that support operational decisions. Thus, in order to contribute to the modeling of wind uncertainty in the operation planning (medium term), this paper proposes a methodology that consists of the application of the dynamic factorial model (DFM) for the generation of synthetic series of monthly averages of wind speed in locations with wind farms. The proposed methodology is applied to a set of wind speed reanalysis series from MERRA-2. The results obtained for the eight wind farms of the SIN analyzed show that the methodology is promising, given the good quality of the monthly average wind speed forecasts up to 2 years ahead and, above all, the good representativeness of the synthetic series generated by the model.

Keywords: wind generation; wind farms; synthetic wind series; dynamic factor model; MCMC.

1 INTRODUÇÃO

A transição energética atual é motivada pelo combate às mudanças climáticas, pautada em fontes renováveis de energia (IPCC, 2022). A contribuição da geração eólica, neste sentido, é fundamental para redução do uso de recursos fósseis e, conseqüentemente, da emissão de gases de efeito estufa (GEEs). No caso brasileiro, a matriz elétrica tem uma posição privilegiada com relação à descarbonização. A participação de fontes renováveis na geração total de eletricidade foi de 78,1% em 2021, contrastando com a média mundial de 26,6% (Brasil e EPE, 2022b).

Atualmente, no Brasil, o Sistema Interligado Nacional (SIN) conta com cerca de 9.971 aerogeradores em operação, distribuídos entre 890 parques eólicos, que totalizam 28 GW de capacidade instalada (13,1% da matriz elétrica brasileira), a sexta maior no *ranking* mundial.¹ No Nordeste, estão localizados 80% dos parques eólicos brasileiros, a região tem a presença dos ventos alísios, o que favorece o aproveitamento da energia eólica (Amarante *et al.*, 2001), conferindo ao SIN um fator de capacidade médio da ordem de 43,6% no ano de 2021, um valor acima da média mundial estimada em 34%.²

A participação da geração eólica no SIN tende a aumentar nos próximos anos, inclusive com parques eólicos *offshore* (Silva, 2019), totalizando uma capacidade instalada de 30 GW em 2031 (Brasil e EPE, 2022a), um incremento de 30% em dez anos.

A expansão da geração eólica contribui para a manutenção da elevada participação das fontes renováveis na matriz elétrica brasileira – 83% em 2031 (Brasil e EPE, 2022a). No entanto, no planejamento de sistemas hidrotérmicos com elevado potencial eólico, como é o caso brasileiro, a entrada em larga escala da fonte eólica, que é intermitente, implica desafios para sua integração ao sistema. Por esse motivo, há a necessidade do tratamento da incerteza da velocidade do vento na cadeia de modelos computacionais que suportam as decisões operativas.

Em virtude da predominância hidroelétrica do SIN, o planejamento da expansão e da operação do sistema envolve a solução de um complexo problema de decisão sob incerteza com grande número de variáveis. A solução do problema é alcançada por meio da sua desagregação em horizontes de decisão específicos (Maceira *et al.*, 2023): planejamento da expansão (longo prazo), planejamento da operação no médio prazo e programação diária da operação (curto prazo). Cada um desses horizontes é resolvido através

1. Disponível em: <https://abeeolica.org.br/energia-eolica/dados-abeeolica/>. Acesso em: 7 dez. 2023.

2. Disponível em: https://abeeolica.org.br/wp-content/uploads/2022/11/2022_11_InfoVento28.pdf. Acesso em: 7 dez. 2023.

de modelos computacionais com distintos níveis de detalhes da representação do sistema e modelagem da incerteza hidráulica (Maceira *et al.*, 2002; 2018).

A elevada participação da geração eólica no SIN aponta para a necessidade da atualização dos modelos computacionais utilizados no planejamento da operação com a modelagem da incerteza eólica (Estanqueiro *et al.*, 2008). Conforme indicado por Maceira *et al.* (2023), a incerteza eólica pode ser modelada por meio de séries sintéticas das médias mensais da velocidade do vento incidentes nos aproveitamentos eólicos ou em agregados destes, denominados por parques eólicos equivalentes. Ainda segundo Maceira *et al.* (2023), a transformação das séries sintéticas de velocidade do vento em séries sintéticas de energia eólica é realizada por meio de funções de transferência ou curvas de potência obtidas a partir de dados históricos.

Assim, visando contribuir com a modelagem da incerteza eólica no planejamento da operação (médio prazo), este trabalho descreve uma abordagem bayesiana para geração de séries sintéticas das médias mensais de velocidade do vento em diferentes parques eólicos. De forma resumida, a metodologia proposta consiste na especificação e no ajuste de um modelo fatorial dinâmico (MFD) para a geração de séries sintéticas das médias mensais da velocidade do vento nas localidades com aproveitamentos eólicos. A abordagem proposta tem a capacidade de lidar com as autocorrelações temporais nas séries de velocidade do vento e com as correlações espaciais entre estas, conforme indicado por Mendonça e Medrano (2015) na previsão da receita tributária do Brasil.

O objetivo é avaliar a aplicação do MFD na geração de séries sintéticas de velocidade de vento incidentes em parques eólicos, uma contribuição para melhoria da representação do recurso eólico nos modelos de planejamento da operação. A metodologia proposta é ilustrada por meio da aplicação do modelo MFD em um conjunto de séries de reanálises da velocidade do vento, todas oriundas do MERRA-2 (Staffell e Pfenninger, 2016). A opção pelas séries de reanálise deve-se à indisponibilidade de longas séries de tempo para a velocidade de vento. Os dados são oriundos do MERRA-2 para o período de janeiro de 2000 a dezembro de 2020. A escolha do MERRA-2 deveu-se por sua facilidade de acesso, boa resolução espacial e temporal (horária) e estabilidade ao longo do tempo.

O trabalho está organizado em cinco seções, incluindo esta introdução. Na seção 2, tem-se uma descrição resumida do estado da arte das propostas para modelagem da incerteza eólica no planejamento da operação do SIN. Uma breve introdução ao MFD é apresentada na seção 3. Na sequência, na seção 4, apresentam-se os resultados obtidos pela metodologia proposta, isto é, as séries sintéticas das médias mensais da velocidade do vento para diferentes localidades do SIN. Por fim, as conclusões são apresentadas na seção 5.

2 ESTADO DA ARTE DA MODELAGEM DA INCERTEZA EÓLICA NO PLANEJAMENTO DA OPERAÇÃO DO SIN

No planejamento da operação do SIN, considera-se como horizonte de estudo um período de até cinco anos à frente com resolução temporal mensal. Dada a indisponibilidade de previsões de velocidade do vento para um horizonte tão longo, adota-se um conjunto de séries sintéticas das médias mensais da velocidade do vento, seguindo basicamente a mesma estratégia tradicionalmente considerada na modelagem das vazões afluentes aos reservatórios (Maceira *et al.*, 2023). Neste sentido, Maceira *et al.* (2023) propõem que a modelagem da incerteza eólica inicie pelo ajuste de modelos autoregressivos periódicos de ordem p – PAR(p) às séries temporais das velocidades médias mensais dos ventos incidentes nos parques eólicos equivalentes (PEEs), isto é, *clusters* de parques eólicos, como os três *clusters* identificados por Pessanha, Almeida e Melo (2020) na região Nordeste. A agregação dos parques eólicos visa evitar que o número de variáveis de estado aumente drasticamente e inviabilize a solução do problema de otimização dinâmica dual estocástica subjacente ao planejamento da operação.

Mais especificamente, na metodologia proposta por Maceira *et al.* (2023) adota-se uma abordagem univariada na modelagem da dependência temporal da velocidade do vento, a qual consiste no ajuste de um modelo PAR(p) em cada série temporal de velocidade média mensal do vento incidente em um PEE. Assim, cabe aos termos aleatórios dos modelos PAR(p) o papel de modelar as correlações espaciais entre as velocidades dos ventos incidentes nos PEEs e destas com as afluições aos reservatórios. Portanto, a abordagem proposta por Maceira *et al.* (2023) busca representar a complementaridade entre os recursos eólicos e hidráulicos (Luz, Vila e Aoki, 2023). Ademais, os termos aleatórios são modelados com distribuições de Weibull com o intuito de capturar as assimetrias nas distribuições das médias mensais das velocidades dos ventos. Assim, os termos aleatórios são modelados por distribuições de Weibull espacialmente correlacionadas. Adicionalmente, a transformação das séries sintéticas de velocidade do vento em séries sintéticas de produção eólica é realizada por meio de funções lineares previamente ajustadas por regressão linear simples.

Adicionalmente, é relevante destacar que a modelagem proposta por Maceira *et al.* (2023) foi construída para atender aos requisitos impostos pelo programa Newave, responsável pela otimização energética de médio prazo do SIN, cujo principal resultado é o custo marginal da operação (CMO), que resulta da solução de um problema de programação dinâmica dual estocástica.

Mummey (2017) também propôs uma representação estocástica da geração eólica no modelo Newave, mas na qual as usinas eólicas são tratadas como se fossem usinas

hidráulicas a fio d'água. Basicamente, a proposta de Mummey (2017) consiste em inserir os dados de produção energética das eólicas no mesmo formato dos dados das hidrelétricas.

Recentemente, Ávila *et al.* (2021) propuseram uma abordagem baseada em cópulas para simular de forma conjunta as séries temporais de vazão e de velocidade do vento, uma abordagem que leva em conta a complementaridade entre as duas fontes renováveis. Os autores demonstram a aplicação da metodologia proposta por meio de um estudo de caso com a usina hidrelétrica de Sobradinho e o parque eólico Pedra Branca, ambos localizados na Bahia.

Maçaira *et al.* (2019) apresentaram uma metodologia baseada na simulação de cadeias de Markov para a geração de séries sintéticas da velocidade do vento em cada aproveitamento eólico. As séries sintéticas de velocidade do vento são transformadas em séries de produção eólica por meio curvas de potência. Na sequência, as séries de carga dos submercados do SIN são deduzidas das séries de produção eólica simuladas a fim de obter séries de carga, líquidas da geração eólica, a serem inseridas nos modelos de otimização da operação hidrotérmica. A abordagem proposta por Maçaira *et al.* (2019) não leva conta as correlações espaciais entre as velocidades do vento e destas com as afluições aos reservatórios.

Anteriormente, Hoeltgebaum, Fernandes e Street (2018) propuseram uma abordagem multivariada com modelo generalized autoregressive score (modelo GAS) capaz de lidar simultaneamente com as correlações temporais e espaciais. Além disso, a abordagem com modelos GAS permite considerar variáveis aleatórias não normais, como é o caso da velocidade do vento. A modelagem multivariada pode trazer ganhos de eficiência, como indicado por Mendonça e Medrano (2015). O modelo proposto por Hoeltgebaum, Fernandes e Street (2018) também não incorpora as correlações entre as vazões e os ventos, ignorando a complementaridade entre os recursos hidráulicos e eólicos.

É importante ressaltar que as séries sintéticas de velocidade de vento não atendem apenas às necessidades do planejamento da operação, mas constituem informações fundamentais na avaliação da viabilidade econômica de projetos eólicos e na comercialização da energia proveniente dos aproveitamentos eólicos em operação.

3 MODELO FATORIAL DINÂMICO

3.1 Especificação

Em termos formais, o modelo fatorial dinâmico é definido da seguinte forma:

$$y_t = \beta f_t + \zeta_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma), \quad (1)$$

$$f_t = \Gamma_1 f_{t-1} + \Gamma_2 f_{t-2} + \dots + \Gamma_p f_{t-p} + w_t, \quad w_t \sim N(0, \Lambda), \quad (2)$$

em que y_t , f_t , ε_t e w_t são vetores de dimensão m , k , m e k , respectivamente, β é uma matriz $m \times k$, $\Gamma_i = \text{diag}(\gamma_{i1}, \gamma_{i2}, \dots, \gamma_{ik})$ $i = 1, \dots, p$, $\varepsilon_t \sim N(0, \Sigma)$ e $w_{it} \sim N(0, W_i)$ definido para cada parque eólico i . $\Sigma = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2)$ e $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$.

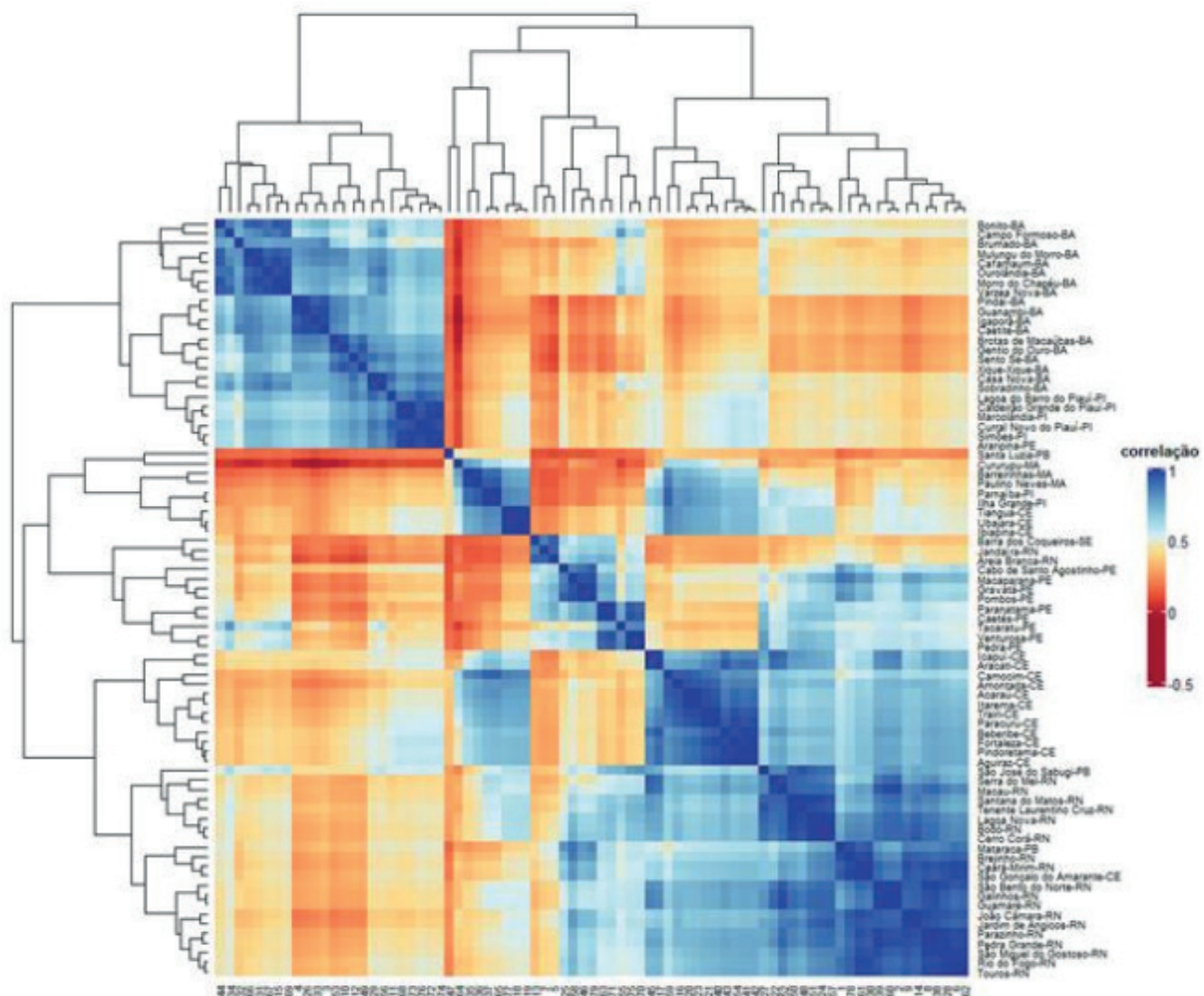
Na equação (1) y_t representa, para cada período de tempo t , o vetor de velocidade de vento para todos os m parques eólicos. Esta equação relaciona, para algum inteiro positivo específico $k \leq m$, o vetor y_t com um vetor k dimensional de variáveis aleatórias f_t , dito fatores comuns, que obedecem a uma estrutura autoregressiva de ordem p .

O fato de que $k \leq m$ implica redução de dimensionalidade. Em outras palavras, em um conjunto grande de variáveis é factível existir alta colinearidade entre elas, por existir algumas poucas fontes de informação comuns. A ideia é que a informação contida nas variáveis que são mais correlacionadas possa ser mapeada por um subconjunto menor de variáveis latentes denominados fatores. Os fatores devem ser não correlacionados, pois representam fontes independentes de informação. Dado que os fatores são ortogonais, sua matriz de covariância não condicional é diagonal. No que diz respeito à geração eólica, os fatores podem representar os diferentes agrupamentos de regimes de vento (Pessanha, Almeida e Melo, 2020).

A seguir, no gráfico 1, tem-se a matriz de correlações entre as velocidades do vento em distintas localidades do Nordeste brasileiro. As elevadas correlações indicam que as séries de velocidade do vento compartilham características em comum, isto é, as velocidades observadas são manifestações de um reduzido número de fatores latentes.

GRÁFICO 1

Matriz de correlações entre velocidade de vento em diferentes localizações do Nordeste brasileiro



Fonte: Pessanha, Almeida e Melo (2020).

Obs.: Ilustração cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

O gráfico 1 revela a existência de pelo menos dois fatores latentes, o quadrado azul no canto superior esquerdo ilustra as correlações entre as velocidades do vento no interior da Bahia, enquanto a área azul no canto inferior direito ilustra as correlações entre as velocidades do vento no litoral no Nordeste.

Conforme pode ser visto também no conjunto de equações (1)-(2), os fatores seguem uma estrutura autorregressiva, o que proporciona um melhor meio de captar as alterações de um período para outro nos regimes de vento.

O componente sazonal $S_t = (S_{1t} \ S_{2t} \ \dots \ S_{mt})^T$ (West e Harrison, 1997) é definido da seguinte forma:

$$S_{it} = \sum_{j=1}^6 \left[a_{ij} \operatorname{sen} \left(\frac{2\pi jt}{12} \right) + b_{ij} \operatorname{cos} \left(\frac{2\pi jt}{12} \right) \right] = \varphi_i X_t, \quad (3)$$

em que

$$X_t = \left(\operatorname{sen} \left(\frac{2\pi t}{12} \right) \ \operatorname{sen} \left(\frac{2\pi 2t}{12} \right) \ \dots \ \operatorname{sen} \left(\frac{2\pi 6t}{12} \right) \ \operatorname{cos} \left(\frac{2\pi t}{12} \right) \ \operatorname{cos} \left(\frac{2\pi 2t}{12} \right) \ \dots \ \operatorname{cos} \left(\frac{2\pi 6t}{12} \right) \right)^T$$

e

$$\varphi_i = (a_{i1} \ a_{i2} \ \dots \ a_{i6} \ b_{i1} \ b_{i2} \ \dots \ b_{i6}), \text{ para } i=1, \dots, m.$$

A ideia que motiva a aplicação da forma de Fourier para tratar sazonalidade está ligada ao fato de que qualquer padrão cíclico pode ser representado em termos da combinação linear de funções periódica.³

O objetivo da análise fatorial consiste em identificar os fatores latentes e, assim, reduzir a dimensão dos dados, de modo que a estrutura da matriz de correlações entre as velocidades do vento possa ser expressa de forma aproximada por um reduzido número de fatores latentes. A vantagem da abordagem MFD é que ela já incorpora a ideia de redução da dimensionalidade em sua construção. As séries dos fatores comuns (latentes) f_t revelam as fontes de informação mais importantes compartilhadas pelas séries temporais de velocidade do vento. Assim, como na análise fatorial exploratória (Pessanha, Almeida e Melo, 2020), no MFD assume-se que os fatores sejam não correlacionados, pois representam fontes independentes de informação.

A redução da dimensionalidade e a resultante da ausência de correlação aumentam a eficiência computacional. Além disso, no MFD as projeções futuras de séries de velocidade de vento de todas as PEEs podem ser geradas simultaneamente.

Conforme pode ser notado, o MFD trata endogenamente e, portanto, com maior robustez, os procedimentos *ad hoc* adotados no modelo de Maceira *et al.* (2023) para alcançar regimes genuínos da média mensal da velocidade de vento.

3.2 Metodologia para estimação do MFD

Para estimação do modelo, foram aplicados os procedimentos de inferência bayesiana – Kim e Nelson (1999), West e Harrison (1997), Carter e Kohn (1994), Gamerman

3. Uma função $g(t)$ é dita periódica se para algum inteiro $p > 1$, e todo $n, t > 0$ $g(t + np) = g(t)$.

e Lopes (2006), Sáfadi e Pena (2007) entre outros –, pois permitem resultados mais confiáveis em contextos em que a dimensão temporal da base de dados é pequena em relação ao espaço da dimensionalidade dos parâmetros, além de evitarem dificuldades, dificilmente tratáveis por meio de procedimentos “frequentistas” (Gelman *et al.*, 2003, p. 696).⁴ Em problemas práticos, especialmente para valores grandes de m , o número de fatores k será pequeno quando comparado com m , de tal modo que muito da estrutura de variância-covariância seja explicada pelos fatores comuns. As variâncias idiossincráticas, σ_i^2 , medem a variabilidade residual em cada uma das variáveis dos dados, uma vez que a contribuição dos fatores já foi considerada.

3.2.1 Restrições no modelo fatorial

Um modelo com k fatores deve apresentar alguma restrição para ser definido como um modelo único, não apresentando problemas de identificação. Há várias maneiras de identificar o modelo adicionando restrições na matriz β . Optou-se por seguir, aqui, a alternativa usada em Geweke e Zhou (1996) e Aguilar e West (2000), entre outros, ao restringir β de modo que seja uma matriz de bloco triangular inferior, assumindo ser de posto completo, com elementos da diagonal principal iguais a um, como apresentado adiante.

$$\beta = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \beta_{2,1} & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & 0 \\ \beta_{k,1} & \beta_{k,2} & \beta_{k,3} & \cdots & 1 \\ \beta_{k+1,1} & \beta_{k+1,2} & \beta_{k+1,3} & \cdots & \beta_{k+1,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \beta_{m,1} & \beta_{m,2} & \beta_{m,3} & \cdots & \beta_{m,k} \end{pmatrix}. \quad (4)$$

Tal restrição exposta anteriormente garante que a matriz β tenha posto completo k , resolvendo os problemas de identificação que surgem pela invariância do modelo sob mudanças de posição da matriz β . Ao mesmo tempo, garante que o número de parâmetros livres no tempo t na representação fatorial não exceda $m(m+1)/2$ parâmetros. Além de resolver os problemas de identificação, essa estrutura de β influencia a interpretação dos fatores e pode impactar no ajuste do modelo e na escolha do k em particular, mas não tem impacto nas previsões, pois as variâncias e covariâncias entre as séries são independentes da decisão de modelagem.

4. Para mais detalhes, ver Mendonça e Medrano (2015).

3.2.2 Especificação de distribuições *a priori*

Para completar a especificação do nosso modelo, deve-se atribuir classes de *prioris* para os parâmetros $f_0, \beta, \Gamma_p, i=1, \dots, p, \Sigma, \Lambda$ e $\varphi_p, i=1, \dots, m$. As *prioris* utilizadas são:

- Vetor de fatores comuns inicial:

$$f_0 \sim N(m_0, C_0).$$

- Elementos não nulos da matriz de cargas β :

$$\beta_{ij} \sim N(\mu_0, V_0) \text{ para } i > j.$$

- Elementos da matriz diagonal de variância-covariância Σ :

$$\sigma_i^2 \sim IG\left(\frac{n_\sigma}{2}, \frac{n_\sigma S_\sigma}{2}\right), \text{ para } i=1, \dots, m.$$

- Elementos das matrizes diagonais dos coeficientes autoregressivos $\Gamma_1, \dots, \Gamma_p$:

$$\gamma_{ij} \sim N(m_r, S_r), \text{ para } i=1, \dots, p, j=1, \dots, k.$$

- Elementos da matriz diagonal de variância-covariância Λ :

$$\lambda_j \sim IG\left(\frac{n_\lambda}{2}, \frac{n_\lambda S_\lambda}{2}\right), j=1, \dots, k.$$

- Elementos dos vetores de coeficientes sazonais $\varphi_1, \dots, \varphi_m$:

$$a_{ij} \sim N(m_s, C_s) \text{ e } b_{ij} \sim N(m_s, C_s), \text{ para } i=1, \dots, m \text{ e } j=1, \dots, k.$$

3.2.3 Método MCMC para o modelo fatorial com k fatores

O objetivo é obter as *posteriors* dos parâmetros do modelo, assim como fazer inferência sobre o processo fatorial $\{f_t: t = 1, 2, \dots, T\}$. Como uma solução exata requer o cálculo de integrais intratáveis do ponto de vista analítico, optou-se por usar o método do amostrador de Gibbs (Gamerman e Lopes, 2006), para obter uma aproximação das *posteriors* de interesse. As *distribuições condicionais completas* necessárias para implementar o algoritmo Monte Carlo Markov Chain (MCMC) encontram-se no apêndice A. Mais informações sobre como calcular essas condicionais completas no contexto de modelos fatoriais podem ser encontradas em Kim e Nelson (1999).

3.3 Previsão do MFD com componente sazonal

Nesta subseção, será descrita a metodologia para previsão h passos à frente, para um modelo de ordem de defasagem $p=1$. A extensão para $p>1$ ocorre de forma simples e encontra-se descrita no apêndice A. O interesse aqui é na densidade preditiva h passos à frente:

$$p(y_{t+h}|y) = \int p(y_{t+h}|\underline{f}_{t+h}, \beta, \varphi, \varrho) p(\underline{f}_{t+h}|\underline{f}_t, \beta, \varphi, \varrho) p(\underline{f}_t, \beta, \varphi, \varrho|y) d\underline{f}_{t+h} d\underline{f}_t d\beta d\varphi d\varrho, \quad (5)$$

em que:

$$\varrho = (\varrho, \lambda, \gamma), \varrho = (\sigma_1, \dots, \sigma_m), \gamma = (\gamma_{11}, \dots, \gamma_{1k}), \lambda = (\lambda_1, \dots, \lambda_k), y = (y_1, \dots, y_T)$$

$$(y_{t+h}|\underline{f}_{t+h}, \beta, \varphi, \varrho) \sim N(\beta \underline{f}_{t+h} + \varphi X_{t+h}, \Sigma) \text{ e } (\underline{f}_{t+h}|\underline{f}_t, \beta, \varrho) \sim N(\underline{\mu}_h, V_h),$$

com $\underline{\mu}_h = \Gamma_1^h \underline{f}_t$ e $V_h = \sum_{k=1}^h \Gamma_1^{k-1} \Lambda (\Gamma_1^{k-1})$, para $h > 0$.

Portanto, se $\{(\beta^{(1)}, \underline{f}_t^{(1)}, \varphi^{(1)}, \vartheta^{(1)}), \dots, (\beta^{(M)}, \underline{f}_t^{(M)}, \varphi^{(M)}, \vartheta^{(M)})\}$ é uma amostra de $p(\underline{f}_t, \beta, \varphi, \vartheta|y)$, basta gerar uma amostra de tamanho M de $(\underline{f}_{t+h}|\underline{f}_t, \beta, \varphi, \vartheta)$, isto é, gerar $\{(\underline{f}_{t+h}^{(1)}), \dots, (\underline{f}_{t+h}^{(M)})\}$, de modo que:

$$\hat{p}(y_{t+h}|y) = M^{-1} \sum_{j=1}^M p(y_{t+h}|\underline{f}_{t+h}^{(j)}, \beta^{(j)}, \varphi^{(j)}, \vartheta^{(j)}). \quad (6)$$

Seja uma aproximação de Monte Carlo para $p(y_{t+h}|y)$. Analogamente, uma amostra $\{y_{t+h}^{(1)}, \dots, y_{t+h}^{(M)}\}$ de $p(y_{t+h}|y)$ é obtida amostrando-se $y_{t+h}^{(j)}$ de $p(y_{t+h}|\underline{f}_{t+h}^{(j)}, \beta^{(j)}, \varphi^{(j)}, \vartheta^{(j)})$, para $j=1, \dots, M$.

4 RESULTADOS

Com o objetivo de avaliar a performance do MFD na geração de séries sintéticas de velocidade de vento, foi realizado um estudo considerando um conjunto formado por 39 séries de médias mensais de velocidades do vento em parques eólicos localizados em diferentes regiões do Brasil.

Em virtude da indisponibilidade de longas séries temporais com medições de velocidade de vento, o modelo MFD foi ajustado aos dados de reanálises oriundos do MERRA-2 (Staffell e Pfenninger, 2016) para o período de janeiro de 2000 a dezembro de 2020. Os dados de reanálise incorporam dados observacionais históricos em um modelo numérico de previsão do tempo para reconstruir o clima, gerando um conjunto completo de dados em espaço-tempo. Os dados de reanálise MERRA-2 foram escolhidos por sua facilidade de acesso, boa resolução espacial e temporal (horária)

e estabilidade ao longo das escalas de tempo. Ressalta-se que, da mesma forma que em Hoeltgebaum, Fernandes e Street (2018), neste trabalho serão consideradas apenas as séries e correlações entre as velocidades do vento.

A partir de dados da Associação Brasileira de Energia Eólica (ABEEólica), sabe-se que a produção eólica do SIN está concentrada na região Nordeste (88,7%).⁵ Maceira *et al.* (2020) agruparam os parques eólicos da região Nordeste em cinco *clusters*, cada um correspondendo a um regime de vento. Assim, o número de fatores do MFD foi fixado em cinco. Ademais, em função da persistência típica dos ventos, considerou-se um termo autoregressivo de ordem 1 (Menezes, Souza e Pessanha, 2014).

Antes de realizar a geração de séries sintéticas de velocidade de vento das estações eólicas com base no MFD, é necessário verificar a capacidade preditiva deste modelo. Isso será feito da seguinte forma: tomando por base que T é o tamanho da amostra, estima-se o MFD com uma amostra reduzida de tamanho $N < T$. A previsão é feita para $T-N$ passos à frente. Em seguida, os valores previstos são comparados com os observados. A avaliação da performance dentro da amostra pode ser implementada com base em diferentes critérios de avaliação. Serão considerados os seguintes critérios concorrentes para avaliar a capacidade preditiva do modelo: a raiz quadrada do erro quadrático médio (RMSE)⁶ que penaliza os *outliers*; o desvio absoluto médio (*mean absolute deviation* – MAD);⁷ o erro percentual absoluto médio (*mean absolute percentage error* – Mape);⁸ e o erro percentual médio (*mean percentage error* – MPE).⁹ A amostra reduzida com dados até dezembro de 2017 foi utilizada no ajuste do MFD às séries de velocidades dos ventos dos 39 parques eólicos analisados. Já a avaliação da qualidade das previsões foi realizada com base na amostra *outsample* que abrange o período de janeiro de 2018 a dezembro de 2019.

A estimação do modelo MFD foi realizado por MCMC (Markov Chain Monte Carlo), cuja convergência foi alcançada após 17 mil iterações, sendo que as primeiras 7 mil foram descartadas. As amostras foram guardadas a cada dez iterações e, portanto, os resultados estão baseados em cadeias de tamanho 1 mil. A determinação da convergência do processo foi feita por técnicas gráficas. Foram usadas os seguintes hiperparâmetros para as distribuições *a priori*: $f_0 \sim N(0,1)$, $\beta_{ij} \sim N(0; 100)$, para $i > j$; $\sigma_i^2 \sim IG(0,01; 0,01)$

5. Disponível em: https://abeeolica.org.br/wp-content/uploads/2022/11/2022_11_InfoVento28.pdf. Acesso em: 7 dez. 2023.

6. $RSME = \sqrt{N^{-1} \sum_{t=1}^N e_t^2}$, em que $e_t = Y_t - \hat{F}_t$, em que Y_t é o valor observado e \hat{F}_t o valor previsto.

7. $MAD = N^{-1} \sum_{t=1}^N |e_t|$.

8. $Mape = \left[\frac{1}{n} \sum_{j=1}^n \left| \frac{e_j}{y_j} \right| \right] \times 100$.

9. $MPE = \left[N^{-1} \sum_{t=1}^N \frac{e_t}{y_t} \right] \times 100$.

TEXTO para DISCUSSÃO

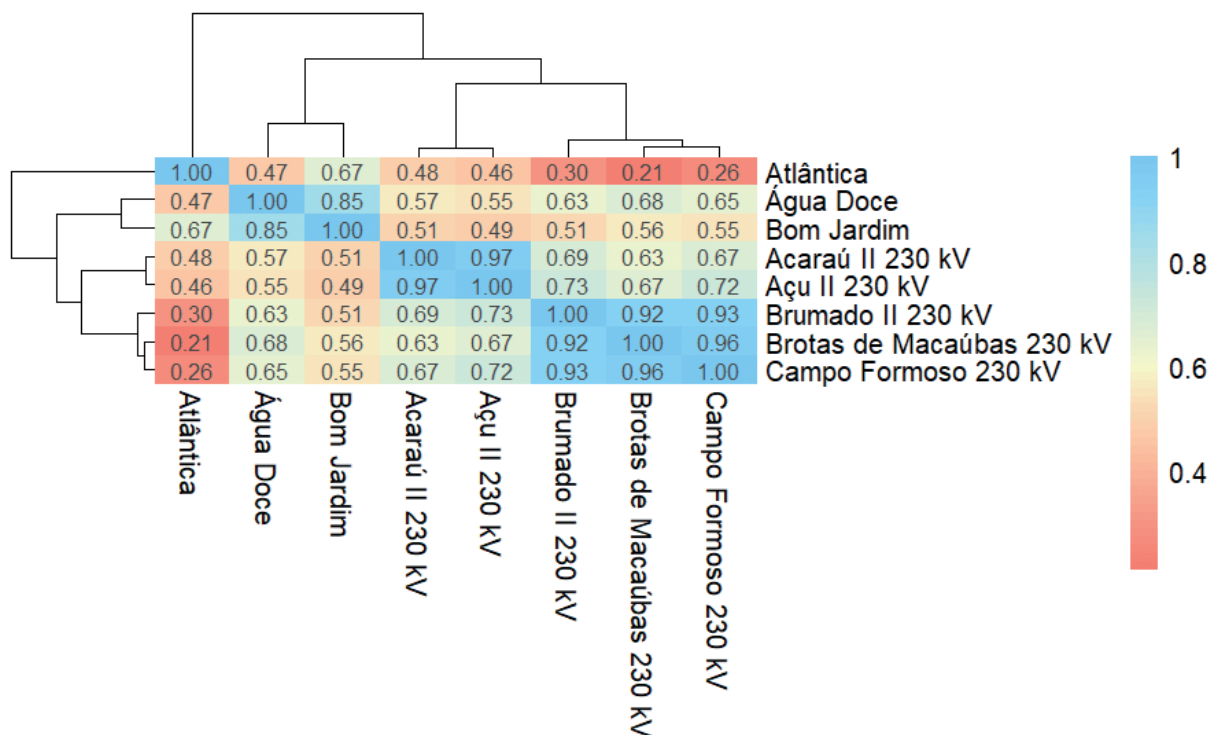
para $j = 1, \dots, m$, $\gamma_{ij} \sim N(0; 100)$ para $i = 1, \dots, p$ e $j = 1, \dots, k$, $\lambda_j \sim IG(0,01; 0,01)$ para $j = 1, \dots, k$, $a_{ij} \sim N(0; 100)$ e $b_{ij} \sim N(0; 100)$ para $i = 1, \dots, m$ e $j = 1, \dots, 6$.

O MFD pode gerar previsões para cada um dos 39 parques eólicos analisados. No entanto, em função das limitações de espaço, os resultados são apresentados para apenas oito parques eólicos selecionados, a saber: Acaraú II 230 kV (CE), Açú II 230 kV (RN), Água Doce (SC), Atlântica (RS), Bom Jardim (SC), Brotas Macaúbas 230 kV (BA), Brumado II 230 kV (BA) e Campo Formoso 230 kV (BA).

No gráfico 2, a matriz de correlações revela a existência de quatro grupos de parques eólicos: o grupo formado pelos parques eólicos Acaraú e Açú, no litoral da região Nordeste; o grupo formado pelos parques eólicos Brumado, Brotas de Macaúbas e Campo Formoso, no interior da Bahia; o grupo dos parques eólicos Água Doce e Bom Jardim, em Santa Catarina; e um grupo com apenas o parque eólico Atlântica, no Rio Grande do Sul.

GRÁFICO 2

Matriz de correlações entre as velocidades do vento nos parques eólicos selecionados



Elaboração dos autores.

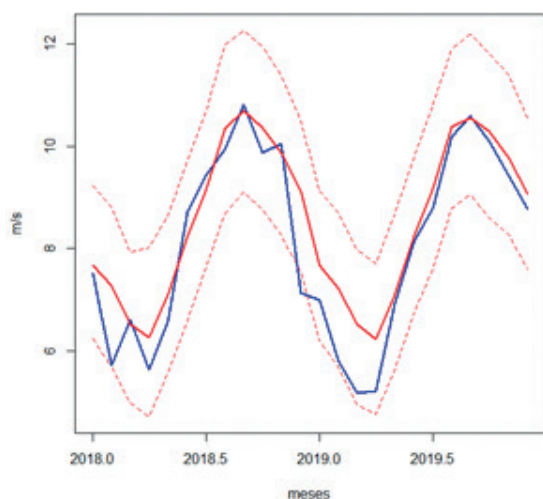
Obs.: Ilustração cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Na sequência, no gráfico 3, apresentam-se as médias mensais das velocidades do vento em cada um dos oito parques eólicos selecionados (linhas em azul) e as respectivas previsões mensais geradas pelo modelo MFD (linhas em vermelho), acompanhadas dos quantis de 5% e 95% (linhas tracejadas). Os resultados no gráfico 3 mostram a boa aderência das previsões aos valores observados no período *outsample*, com a maioria dos valores observados dentro dos intervalos de credibilidade.

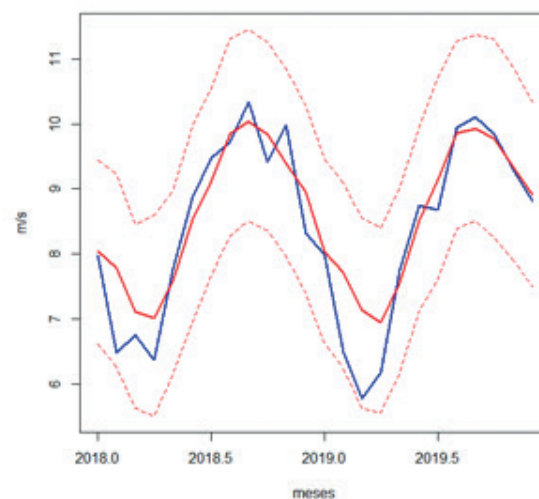
GRÁFICO 3

Valores observados, previsões e intervalos de credibilidade nos parques eólicos selecionados

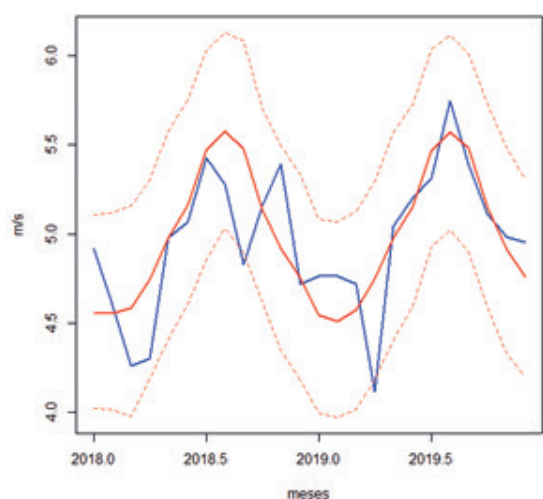
3A – Acaraú II 230 kV



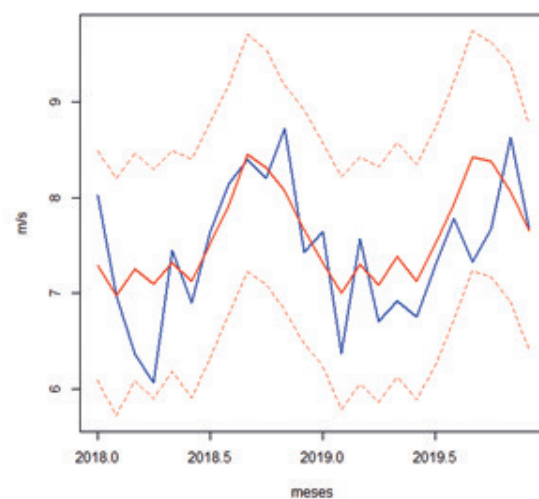
3B – Açu II 230 kV



3C – Água Doce

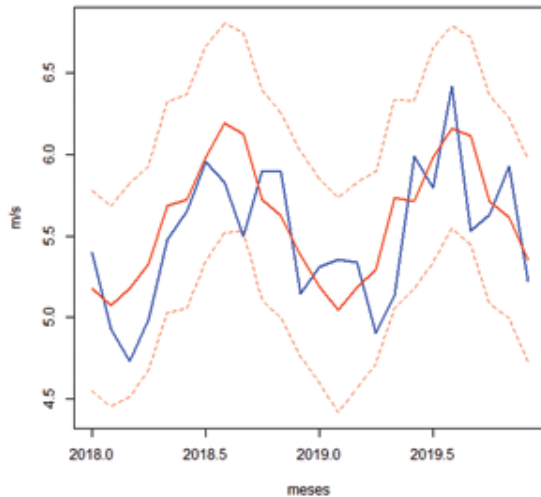


3D – Atlântica

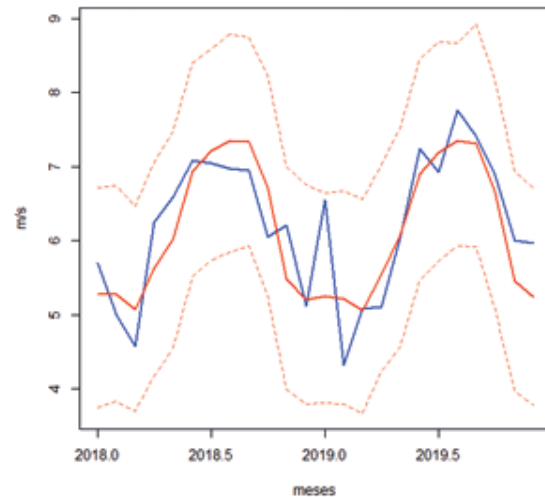


TEXTO para DISCUSSÃO

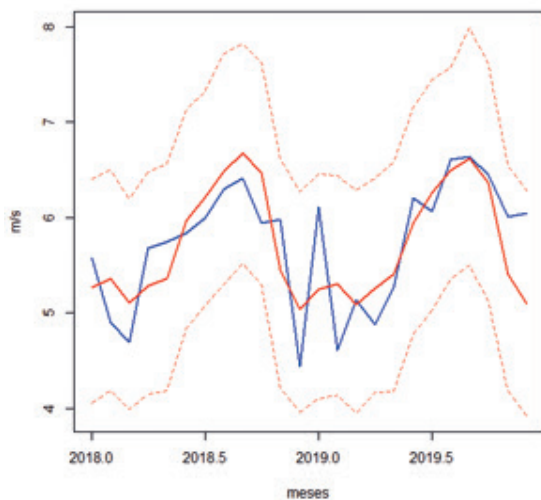
3E – Bom Jardim



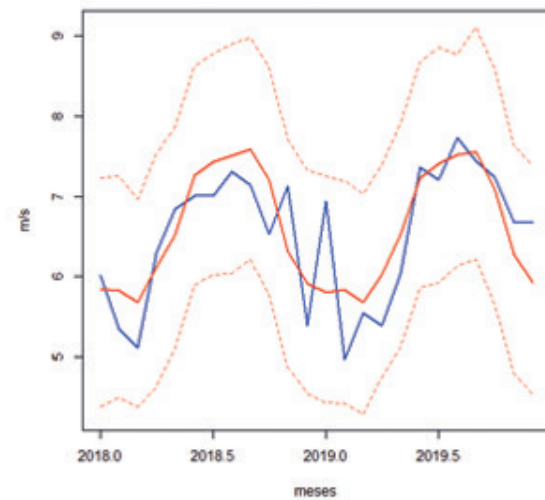
3F – Brotas de Macaúbas 230 kV



3G – Brumado II 230 kV



3H – Campo Formoso 230 kV



Elaboração dos autores.

Obs.: Ilustração cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

A boa aderência ilustrada no gráfico 3 está quantificada nas métricas de desempenho na tabela 1. Observa-se que, nos oito parques eólicos analisados, as respectivas estatísticas U-Theil são menores que a unidade, apontando que o MFD é melhor que o predictor ingênuo. As estatísticas RMSE e MAD apontam para desvios menores que 1 m/s na velocidade média mensal. Adicionalmente, a estatística MPE é negativa na maioria dos parques, indicando que o MFD tende a superestimar as médias mensais das velocidades, mas com um pequeno viés, com exceção do parque localizado em Acaraú, no qual a estatística MPE é da ordem de 7%. Em relação ao Mape, os valores variam entre 4% e 9%, sendo superiores aos encontrados na previsão da carga mensal, em geral menores

que 3%. É importante destacar que a incerteza da previsão dos recursos eólicos é superior à incerteza da previsão de carga e por esta razão as duas incertezas são tratadas distintamente no programa Newave, o modelo de otimização estocástica que suporta o planejamento da operação de médio prazo. A incerteza da carga é representada por uma previsão da carga mensal ao longo do horizonte de planejamento da operação (cinco anos), enquanto a incerteza eólica é modelada por meio de um conjunto de séries sintéticas para o mesmo período.

TABELA 1
Métricas da qualidade das previsões

Parque eólico	RMSE (m/s)	Mape (%)	MAD (m/s)	MPE (%)	Theil's U1 ¹	Theil's U2 ²
Acaraú II 230 kV	0,7516	8,1742	0,5436	-7,0805	0,0443	0,6730
Açu II 230 kV	0,5712	5,7785	0,4179	-3,5245	0,0333	0,6707
Água Doce	0,2764	4,2763	0,2039	-0,8427	0,0277	0,6833
Atlântica	0,5076	5,5603	0,3998	-2,4511	0,0337	0,6603
Bom Jardim	0,3170	5,0408	0,2727	-2,0021	0,0285	0,7602
Brotas Macaúbas 230 kV	0,5182	7,1590	0,4260	0,7910	0,0416	0,5871
Brumado II 230 kV	0,4410	6,6153	0,3656	-0,2404	0,0383	0,6076
Campo Formoso 230 kV	0,5060	6,8416	0,4267	-1,6748	0,0384	0,6293

Elaboração dos autores.

Notas: ¹ $Theil's U1 = \sqrt{\frac{\frac{1}{N} \sum_{t=1}^{N-1} (\widehat{F}_{t+1} - Y_{t+1})^2}{Y_t} / \frac{1}{N} \sum_{t=1}^{N-1} (\widehat{F}_{t+1} - Y_t)^2}$.

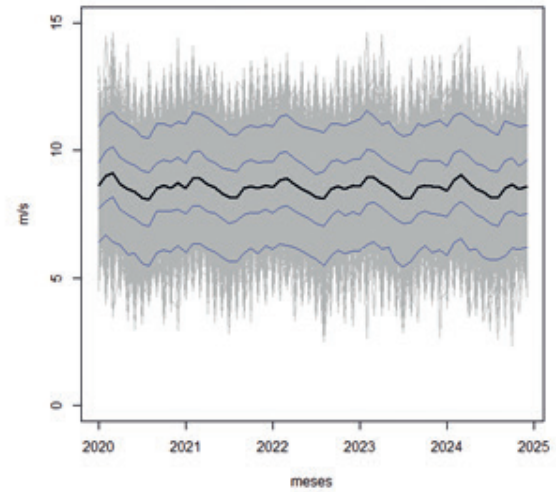
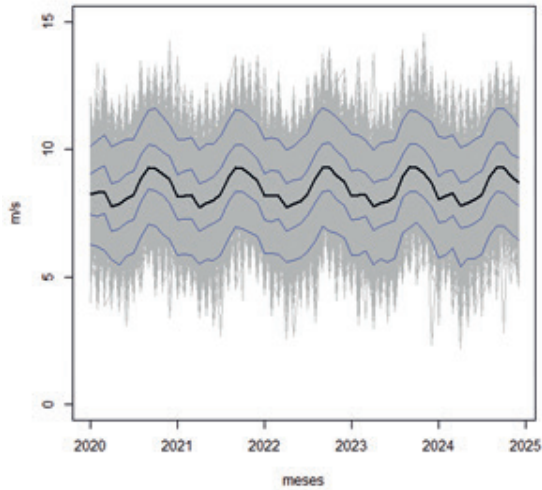
² $Theil's U2 = \sqrt{\frac{\frac{1}{N} \sum_{t=1}^{N-1} (\widehat{F}_t - Y_t)^2}{\frac{1}{N} \sum_{t=1}^N Y_t^2 + \frac{1}{N} \sum_{t=1}^N (\widehat{F}_t)^2}}$.

Os resultados atestam a boa qualidade das previsões eólicas obtidas pelo modelo MFD. Portanto, espera-se que o modelo seja capaz de gerar séries realistas das médias mensais da velocidade do vento ao longo do período de planejamento da operação, isto é, séries sintéticas coerentes com as estruturas de correlações temporais e espaciais presente nos dados de reanálises do MERRA-2. No caso do modelo MFD, as séries sintéticas são geradas a partir das distribuições *a posteriori*. A seguir, no gráfico 4, apresentam-se 1 mil séries sintéticas das médias mensais da velocidade do vento geradas pelo modelo MFD para cada um dos oito parques eólicos analisados, na qual a média das séries corresponde à linha em preto e as linhas em azul representam os quantis de 5%, 25%, 50%, 75% e 95%.

GRÁFICO 4**Séries sintéticas geradas pelo modelo MFD nos parques seleccionados (2020-2024)**

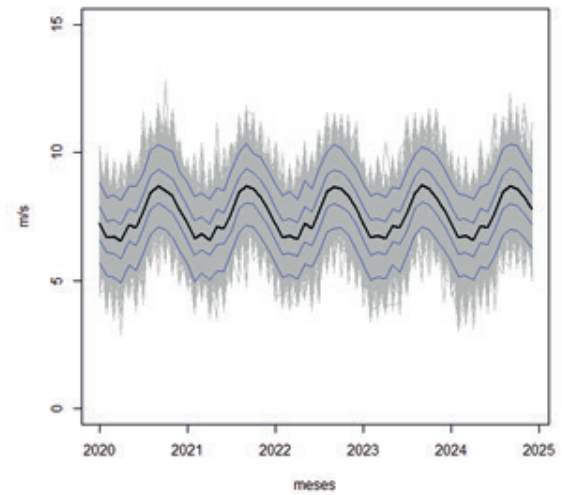
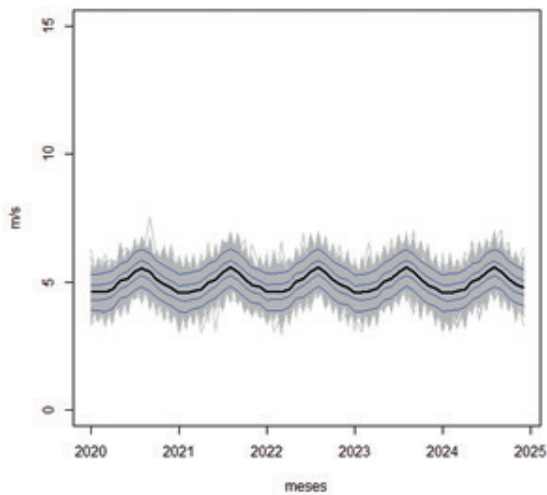
4A – Acaraú II 230 kV

4B – Açú II 230 kV

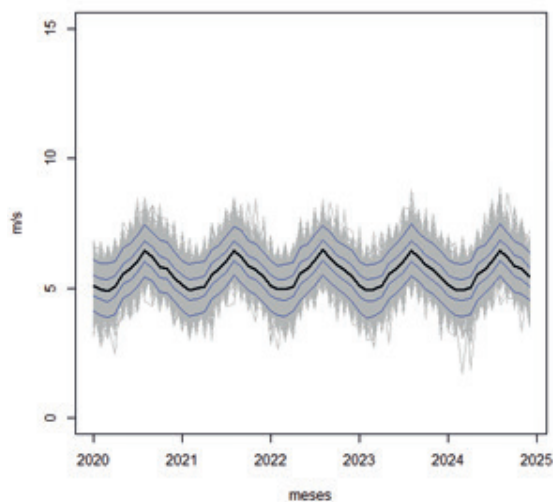


4C – Água Doce

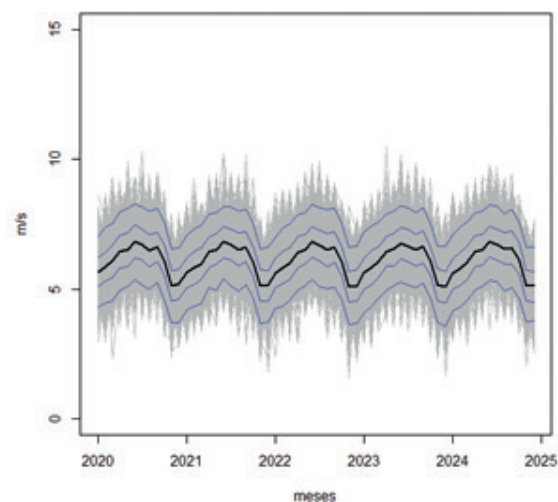
4D – Atlântica



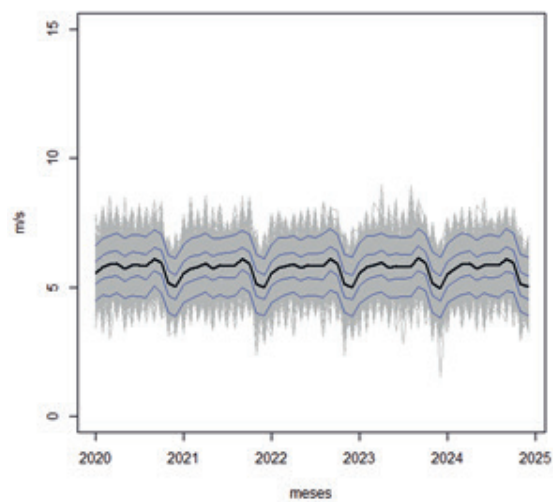
4E – Bom Jardim



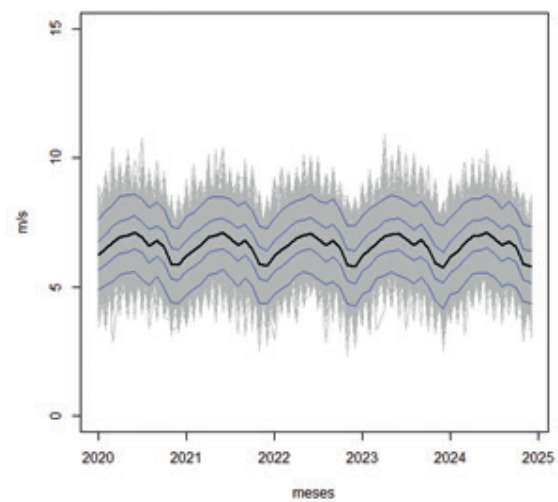
4F – Brotas de Macaúbas 230 kV



4G – Brumado II 230 kV



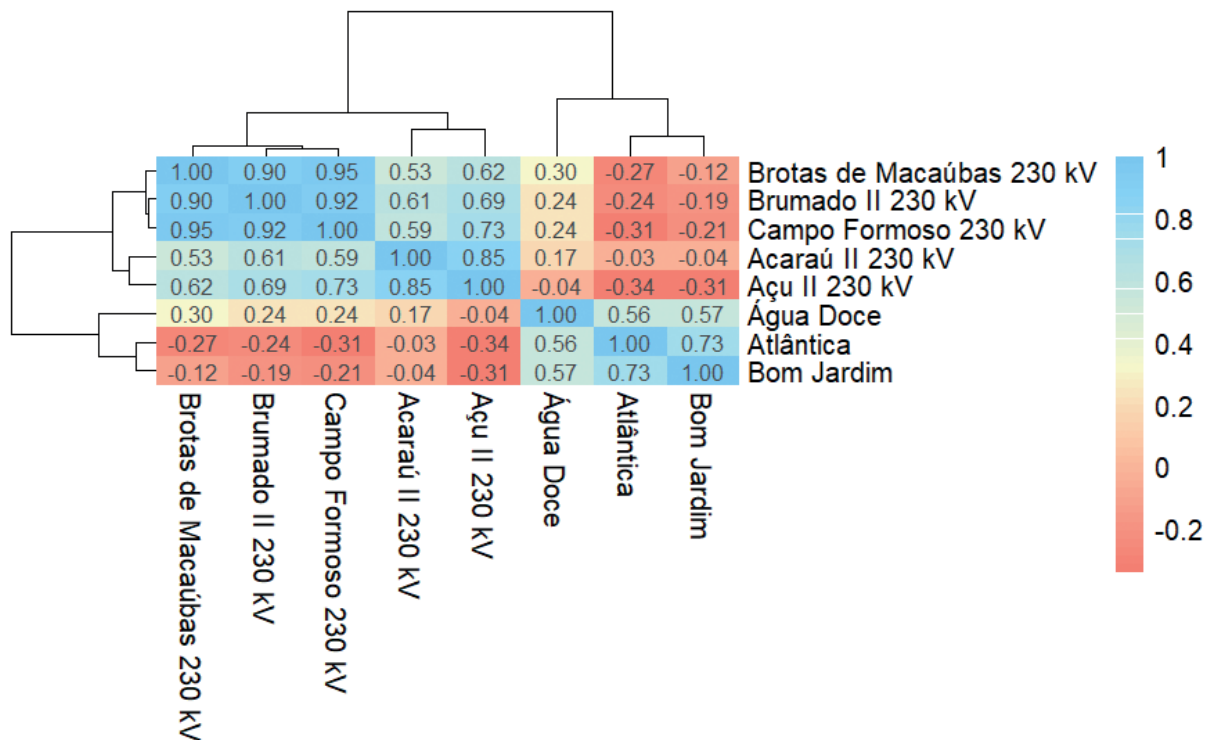
4H – Campo Formoso 230 kV



Elaboração dos autores.

Obs.: Ilustração cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

As médias e os quantis no gráfico 4 mostram que as séries sintéticas capturaram os padrões sazonais presentes nos dados de reanálise. Ademais, conforme indicado no gráfico 5, a matriz de correlações espaciais entre as séries sintéticas assemelha-se à matriz de correlações entre as reanálises no gráfico 2, embora apresente algumas correlações negativas inexistentes nas correlações entre as reanálises.

GRÁFICO 5**Matriz de correlações espaciais entre as séries sintéticas**

Elaboração dos autores.

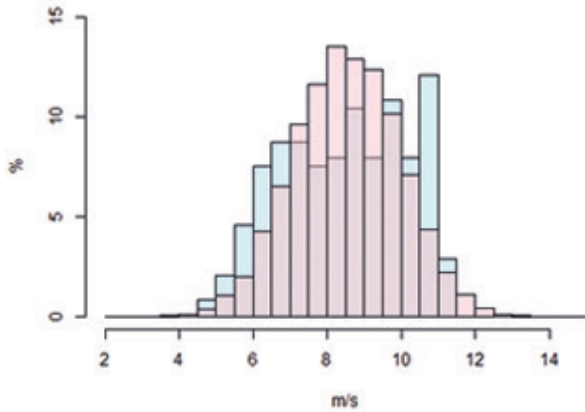
Obs.: Ilustração cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Adicionalmente, conforme ilustrado no gráfico 6, a sobreposição entre os histogramas das reanálises do período 2000-2019 (em azul) e das séries sintéticas do período 2018-2019 (em vermelho) atesta a boa qualidade das séries sintéticas geradas.

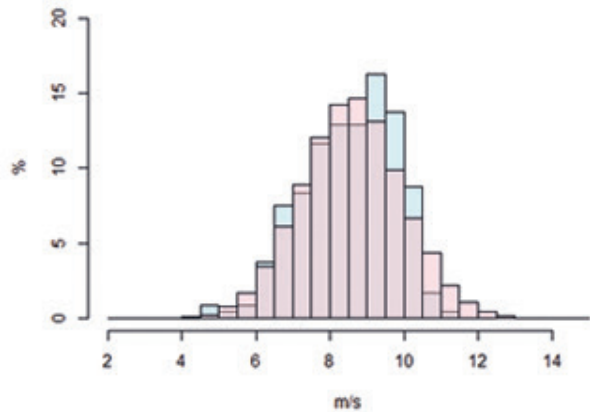
GRÁFICO 6

Histogramas das reanálises 2000-2019 (em azul) e das séries sintéticas 2018-2019 (em vermelho)

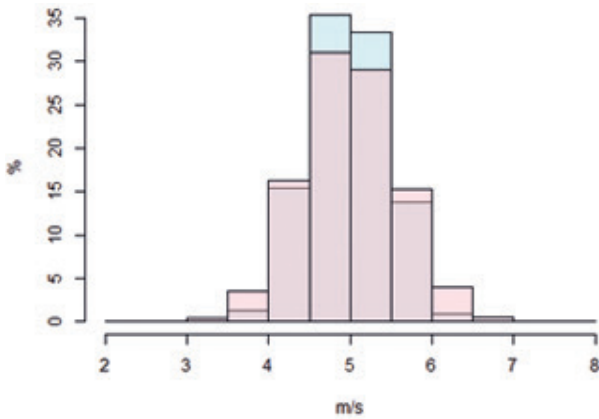
6A – Acaraú II 230 kV



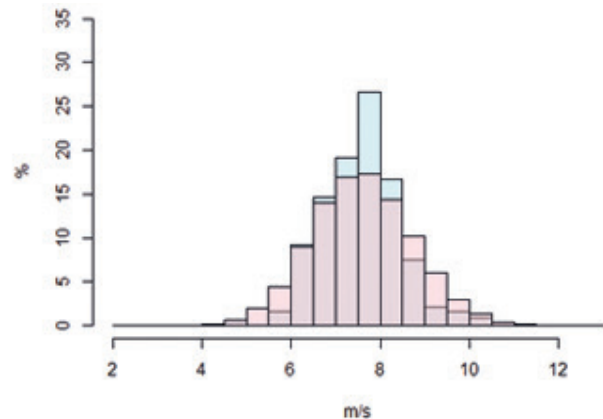
6B – Açu II 230 kV



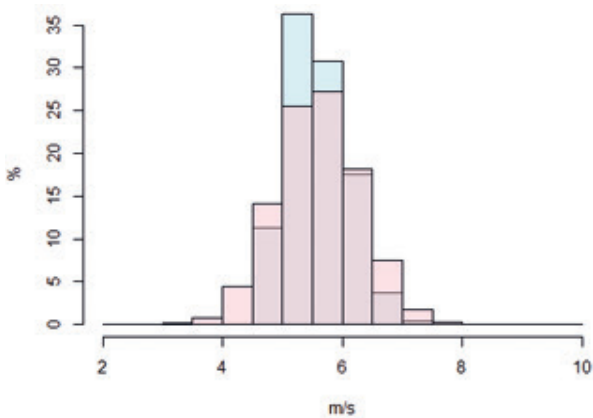
6C – Água Doce



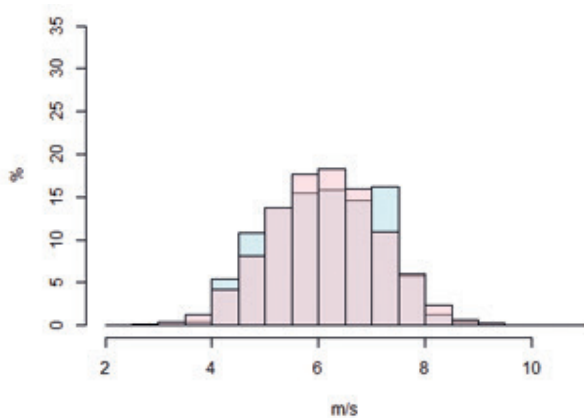
6D – Atlântica



6E – Bom Jardim

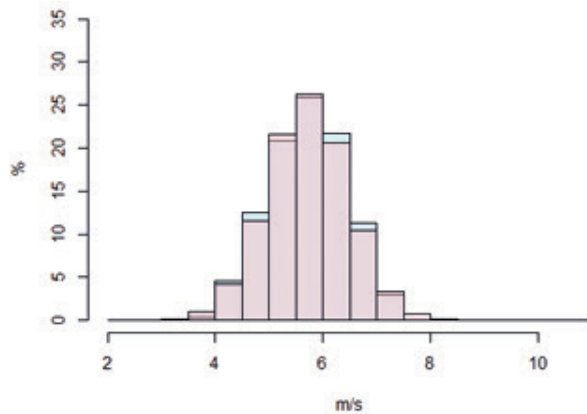


6F – Brotas de Macaúbas 230 kV

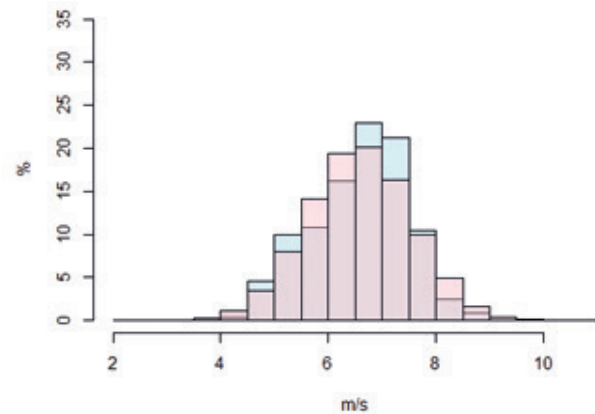


TEXTO para DISCUSSÃO

6G – Brumado II 230 kV



6H – Campo Formoso 230 kV



Elaboração dos autores.

Obs.: Ilustração cujos leiaute e textos não puderam ser padronizados e revisados em virtude das condições técnicas dos originais (nota do Editorial).

Com relação à qualidade das séries geradas, a metodologia proposta tem por objetivo gerar trajetórias plausíveis para a velocidade do vento no período de planejamento, isto é, séries sintéticas que preservam as mesmas propriedades estocásticas das séries históricas, por exemplo, média, desvio-padrão, sazonalidade e distribuição, conforme sugerido pela sobreposição dos histogramas dos valores observados e dos valores simuladas (gráfico 6). Assim, as séries sintéticas preservam a mesma incerteza observada no histórico, conforme indicado no gráfico 4 e, portanto, não é contraintuitivo que a incerteza exibida seja constante. Ao contrário, o aumento da incerteza implicaria valores irrealistas para a velocidade do vento, inclusive com velocidades bem acima dos valores históricos.

5 CONCLUSÕES

As séries sintéticas de velocidade do vento permitem modelar a incerteza eólica nos programas computacionais que suportam as decisões operativas do SIN e são informações essenciais na avaliação de risco de novos projetos eólicos. Este trabalho apresenta uma aplicação dos MFDs na geração de séries sintéticas de velocidade do vento. A metodologia proposta tem uma abordagem multivariada, o que possibilita a modelagem das correlações temporais e espaciais presentes nas séries eólicas. Adicionalmente, a metodologia fundamenta-se na inferência bayesiana, oferecendo como resposta uma sequência de distribuições de probabilidades a *posteriori*, das quais são amostradas as séries sintéticas, isto é, cenários futuros das trajetórias das médias mensais da velocidade de vento em cada parque eólico analisado. A metodologia proposta foi aplicada em um conjunto de 39 parques eólicos do SIN e, apesar de algumas correlações

especiais negativas inexistentes nos dados de reanálises, os resultados obtidos para os oito parques eólicos selecionados mostram que a metodologia é promissora, dada a boa qualidade das previsões mensais da média mensal da velocidade do vento até cinco anos à frente e, sobretudo, a boa representatividade das séries sintéticas geradas pelo modelo.

REFERÊNCIAS

AGUILAR, O.; WEST, M. Bayesian dynamic factor models and portfolio allocation. **Journal of Business and Economic Statistics**, v. 18, n. 3, p. 338-357, 2000.

AMARANTE, O. A. C. *et al.* **Atlas do potencial eólico brasileiro**. Brasília: MME, 2001.

ÁVILA, L. *et al.* Evaluation of hydro-wind complementarity in the médium-term planning of electrical power systems by joint simulation of periodic stramflow and wind speed time series: a Brazilian case study. **Renewable Energy**, v. 167, p. 685-699, abr. 2021.

BRASIL. Ministério de Minas e Energia; EPE – EMPRESA DE PESQUISA ENERGÉTICA. **Plano Decenal de Expansão de Energia 2031**. Brasília: MME; EPE, 2022a.

BRASIL. Ministério de Minas e Energia; EPE – EMPRESA DE PESQUISA ENERGÉTICA. **BEN 2022**: relatório síntese, ano-base 2021. Rio de Janeiro: MME, 2022b. Disponível em: https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-675/topico-631/BEN_Síntese_2022_PT.pdf.

CARTER, C. K.; KOHN, R. On Gibbs sampling for state space models. **Biometrika**, v. 81, n. 3, p. 541-553, 1994.

ESTANQUEIRO, A. *et al.* How to prepare a power system for 15% wind energy penetration: the Portuguese case study. **Wind Energy**, v. 11, n. 1, p. 75-84, 2008.

GAMERMAN, D.; LOPES, H. F. **Markov chain Monte Carlo**: stochastic simulation for Bayesian inference. Boca Raton: Chapman & Hall, 2006.

GELMAN, A. *et al.* **Bayesian data analysis**. 2. ed. Boca Raton: Chapman & Hall, 2003.

GEWEKE, J.; ZHOU, G. Measuring the pricing error of the arbitrage pricing theory. **The Review of Financial Studies**, v. 9, n. 2, p. 557-587, 1996.

HOELTGEBAUM, H.; FERNANDES, C.; STREET, A. Generating joint scenarios for renewable generation: the case for non-gaussian models with time-varying parameters. **IEEE Transactions on Power Systems**, v. 33, n. 6, p. 7011-7019, nov. 2018.

IPCC – INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE. **Climate Change 2022: mitigation of climate change – technical summary**. Working Group III; WMO; UNEP, 2022. Disponível em: https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_TechnicalSummary.pdf.

KIM, C. J.; NELSON, C. R. **State-space models with regime switching**. Massachusetts: The MIT Press, 1999.

LUZ, T. J.; VILA, C. U.; AOKI, A. R. Complementarity between renewable energy sources and regions: Brazilian case. **Brazilian Archives of Biology and Technology**, v. 66, 2023.

MAÇAIRA, P. M. *et al.* Including wind power generation in Brazil's long-term optimization model for energy planning. **Energies**, v. 12, n. 5, p. 826, 2019.

MACEIRA, M. E. P. *et al.* **Chain of optimization models for setting the energy dispatch and spot price in the Brazilian system**. Sevilha: PSCC, 2002.

MACEIRA, M. E. P. *et al.* **Twenty years of application of stochastic dual dynamic programming in official and agent studies in Brazil: main features and improvements on the Newave model**. Dublin: PSCC, 2018.

MACEIRA, M. E. P. *et al.* Uma abordagem para a representação das incertezas da fonte de geração eólica no modelo Newave. **Cadernos do IME – Série Estatística**, v. 48, p. 1-36, 2020.

MACEIRA, M. E. P. *et al.* **Combining monthly wind and inflow uncertainties in the stochastic dual dynamic programming: application to the Brazilian interconnected system**. Energy System, 2023.

MENDONÇA, M. J.; MEDRANO, L. A. **Aplicação do modelo fatorial dinâmico para previsão da receita tributária no Brasil**. Brasília: Ipea, 2015. (Texto para Discussão, n. 2064).

MENEZES, M. L.; SOUZA, R. C.; PESSANHA, J. F. M. Combining singular spectrum analysis and PAR(p) structures to model wind speed time series. **Journal of Systems Science and Complexity**, v. 27, p. 29-46, 2014.

MUMMEY, J. F. C. **Uma contribuição metodológica para a otimização da operação e expansão do sistema hidrotérmico brasileiro mediante a representação estocástica da geração eólica**. 2017. Tese (Doutorado) – Instituto de Energia e Ambiente, Universidade de São Paulo, 2017.

PESSANHA, J. F. M.; ALMEIDA, V. A.; MELO, A. C. G. Fator de capacidade da geração eólica na região Nordeste: um estudo com dados de reanálises oriundos do MERRA-2 (global). *In*: SIMPÓSIO BRASILEIRO DE SISTEMAS ELÉTRICOS, 8., 2020, Santo André, São Paulo. **Anais...** Santo André, 2020.

SÁFADI, T.; PENA, D. Bayesian analysis of dynamic factor models: an application to air pollution and mortality in São Paulo, Brazil. **Environmetrics**, v. 19, n. 6, p. 582-601, 2007.

SILVA, A. J. V. C. **Potencial eólico offshore no Brasil**: localização de áreas nobres através de análise multicritério. 2019. Dissertação (Mestrado) – Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa em Engenharia, Universidade Federal do Rio de Janeiro, 2019.

STAFFELL, I.; PFENNINGER, S. Using bias-corrected reanalysis to simulate current and future wind power output. **Energy**, v. 114, n. 1, p. 1224-1239, nov. 2016.

WEST, M.; HARRISON, J. **Bayesian forecasting and dynamic models**. 2. ed. Springer, 1997.

BIBLIOGRAFIA COMPLEMENTAR

LEE, M. H. L. *et al.* A comparative study of forecasting electricity consumption using machine learning models. **MDPI: Mathematics**, v. 10, n. 8, p. 1329, 2022. Disponível em: <https://doi.org/10.3390/math10081329>.

ONS – OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Desenvolvimento metodológico para previsão da geração de fonte eólica**. Rio de Janeiro: ONS, dez. 2018. (Nota Técnica, n. 151). Disponível em: <https://www.gov.br/mme/pt-br/arquivos/nota-tcnica-ons-151-2018-desenvolvimento-metodologico-para-previs-o-de-gera-o-de-fonte-e-lica.pdf>. Acesso em: 14 nov. 2021.

TORRINI, F. C. *et al.* Long term electricity consumption forecast in Brazil: a fuzzy logic approach. **Socio-Economic Planning Sciences**, v. 54, p. 18-27, jun. 2016.

ZHANG, X.; LI, R. A novel decomposition and combination technique for forecasting monthly electricity consumption. **Front. Energy Res.**, v. 9, dez. 2021. Disponível em: <https://www.frontiersin.org/articles/10.3389/fenrg.2021.792358/full>.

APÊNDICE A

A.1 DESCRIÇÃO DO ALGORITMO MCMC PARA O MODELO FATORIAL DINÂMICO

O modelo (1) pode ser representado em uma estrutura de espaço de estados, do seguinte modo:

$$(y_t - \varphi X_t) = F^E f_t^E + e_t, \quad e_t \sim N(0, \Sigma), \quad (\text{A.1})$$

$$f_t^E = G^E f_{t-1}^E + w_t^E, \quad w_t^E \sim N(0, Q^E), \quad (\text{A.2})$$

em que

$$G_{pkxp}^E = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \cdots & \Gamma_{p-1} & \Gamma_p \\ I_{kxk} & Q_{kxk} & \cdots & Q_{kxk} & Q_{kxk} \\ Q_{kxk} & I_{kxk} & \cdots & Q_{kxk} & Q_{kxk} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Q_{kxk} & Q_{kxk} & \cdots & I_{kxk} & Q_{kxk} \end{bmatrix}, \quad f_t^E = \begin{bmatrix} f_t \\ f_{t-1} \\ \vdots \\ f_{t-p+1} \end{bmatrix}, \quad w_t^E = \begin{bmatrix} w_t \\ Q_{kx1} \\ \vdots \\ Q_{kx1} \end{bmatrix}, \quad Q_{pkxp}^E = \begin{bmatrix} \Lambda_{kxk} & Q_{kxk(p-1)} \\ Q_{k(p-1)xk} & Q_{k(p-1)xk(p-1)} \end{bmatrix},$$

$$F_{mxpk}^E = [\beta_{mxk} \quad Q_{mxk} \quad Q_{mxk} \quad \cdots \quad Q_{mxk}].$$

A.2 DISTRIBUIÇÕES CONDICIONAIS COMPLETAS NECESSÁRIAS PARA A IMPLEMENTAÇÃO DO AMOSTRADOR DE GIBBS

Para $i=1, \dots, k$, escreva $\beta_i = (\beta_{i1}, \dots, \beta_{ii})^T$ que são os elementos diferentes de zero. Para os restantes das linhas $i=k+1, \dots, m$, escreva $\beta_i = (\beta_{i1}, \dots, \beta_{ik})$. Similarmente, para $i=1, \dots, k$ denote F_i a matriz $T \times i$ contendo as primeiras i colunas de F e para todo i seja y_i a i -ésima coluna de y , S_i a i -ésima coluna de $(X\varphi^T)$ com $X = (X_1^T, \dots, X_T^T)^T$ e $y_i^* = y_i - S_i$. Seja $\gamma_j = (\gamma_{1j}, \gamma_{2j}, \dots, \gamma_{pj})^T$ o vetor contendo o j -ésimo elemento da diagonal para cada uma das p matrizes $\Gamma_1, \dots, \Gamma_p$. Defina também para $j=1, \dots, k$,

$$\mathfrak{F}_j = \begin{pmatrix} f_{p,j} & f_{p-1,j} & \cdots & f_{1,j} \\ f_{p+1,j} & f_{p,j} & \cdots & f_{2,j} \\ \vdots & \vdots & \vdots & \vdots \\ f_{T-1,j} & f_{T-2,j} & \cdots & f_{T-p,j} \end{pmatrix} \mathbf{e}_{H_j} = \begin{pmatrix} f_{p+1,j} \\ f_{p+2,j} \\ \vdots \\ f_{T,j} \end{pmatrix}.$$

Para amostrar da condicional completa de f_t siga os passos a seguir.

1) Aplicar filtro de Kalman para calcular $m_t = E[f_t^E | D_t]$ e $C_t = V[f_t^E | D_t]$, em que $D_t = (y_1, \dots, y_T)$ para $t = 1, \dots, T$, do seguinte modo:

$$a_t = G^E m_{t-1}, R_t = G^E C_{t-1} G^{E^T} + Q^E, \quad (\text{A.3})$$

$$f_t = F^E a_t, Q_t = F^E R_t F^{E^T} + \Sigma, \quad (\text{A.4})$$

$$A_t = R_t F^{E^T} Q_t^{-1}, e_t = (y_t - S_t) - f_t, \quad (\text{A.5})$$

e armazenar m_t e C_t , $t = 1, \dots, T$. No final use m_t e C_t para gerar $f_T^E | \cdot \sim N(m_T, C_T)$.

2) Defina $f_t^{E^*}$ como sendo os primeiros k elementos de f_t^E , G^{E^*} , como sendo as primeiras k linhas de G^E e Q^{E^*} , como sendo o primeiro bloco $k \times k$ da matriz Q^E . Para $t = T - 1, T - 2, \dots, 1$ faça:

$$m_t^* = m_t + C_t G^{E^*T} (G^{E^*} C_t G^{E^*T} + Q^{E^*})^{-1} (f_{t+1}^{E^*} - G^{E^*} m_t), \quad (\text{A.6})$$

$$C_t^* = C_t - C_t G^{E^*T} (G^{E^*} C_t G^{E^*T} + Q^{E^*})^{-1} G^{E^*} C_t, \quad (\text{A.7})$$

e gere $f_t^E | \cdot, f_{t+1}^{E^*} \sim N(m_t^*, C_t^*)$ e mantenha somente os k primeiros elementos gerados dessa forma. Esses primeiros k elementos formam uma amostra de f_t da condicional completa.

A *posteriori* condicional completa de Σ se reduz a um conjunto de m gamas inversas independentes, com $\sigma_i^2 \sim IG\left(\frac{n_\sigma^*}{2}, \frac{n_\sigma^* S_\sigma^*}{2}\right)$, em que

$$n_\sigma^* = n_\sigma + T \quad (\text{A.8})$$

$$n_\sigma^* S_\sigma^* = n_\sigma S_\sigma + (y_i^* - F \beta_i^T)^T (y_i^* - F \beta_i^T). \quad (\text{A.9})$$

A *posteriori* condicional completa de β divide-se em:

para $i=1, \dots, k$, $\beta_i \sim N(\underline{\mu}_i, \underline{V}_i) 1(\beta_{ii} = 1)$ em que $\underline{\mu}_i = \underline{V}_i (V_0^{-1} \mu_0 1_i + \sigma_i^{-2} F_i^T y_i^*)$ e $\underline{V}_i^{-1} = V_0^{-1} I_i + \sigma_i^{-2} F_i^T F_i$

para $i=k+1, \dots, m$, $\beta_i \sim N(\underline{\mu}_i, \underline{V}_i)$ em que $\underline{\mu}_i = \underline{V}_i (V_0^{-1} \mu_0 1_k + \sigma_i^{-2} F^T y_i^*)$ e $\underline{V}_i^{-1} = V_0^{-1} I_k + \sigma_i^{-2} F^T F$.

A *posteriori* condicional completa de γ_j é dada por:

para $j=1, \dots, k$, $\gamma_j \sim N(A_j, B_j)$, em que $A_j = B_j (S_r^{-1} m_r 1_p + \lambda_j^{-1} \mathfrak{S}_j^T H_j)$ e $B_j = S_r^{-1} I_{p \times p} + \lambda_j^{-1} \mathfrak{S}_j^T \mathfrak{S}_j$.

A *posteriori* condicional completa de Λ se reduz a um conjunto de k gamas inversas independentes, com $\lambda_j \sim IG\left(\frac{n_\lambda^*}{2}, \frac{n_\lambda^* S_\lambda^*}{2}\right)$, em que

$$n_\lambda^* = n_\lambda + T \quad (\text{A.10})$$

$$n_\lambda^* S_\lambda^* = n_\lambda S_\lambda + (H_j - \mathfrak{S}_j \gamma_j)^T (H_j - \mathfrak{S}_j \gamma_j). \quad (\text{A.11})$$

A *posteriori* condicional completa de φ_i é dada por:

para $i=1, \dots, m$, $\varphi_i \sim N(C_i, D_i)$, em que $C_i = H_i (C_s^{-1} m_s 1_{12} + \sigma_i^{-1} X(y_i - F\beta_i))$ e $H_i = C_s^{-1} I_{12 \times 12} + \sigma_i^{-2} X^T X$.

A.3 PREVISÃO MODELO FATORIAL DINÂMICO COM COMPONENTE SAZONAL COM $p > 1$

Passe as equações (1) e (2) da seção 3 para a estrutura de espaço de estados definida pelas equações (A.1) e (A.2) e utilize a mesma metodologia apresentada na subseção 3.3, substituindo $\tilde{f}_t, \Lambda, \beta$ e Γ_1 por \tilde{f}_t^E, Q^E, F^E e G^E respectivamente.

Ipea – Instituto de Pesquisa Econômica Aplicada

EDITORIAL

Coordenação

Aeromilson Trajano de Mesquita

Assistentes da Coordenação

Rafael Augusto Ferreira Cardoso

Samuel Elias de Souza

Supervisão

Ana Clara Escórcio Xavier

Everson da Silva Moura

Revisão

Alice Souza Lopes

Amanda Ramos Marques Honorio

Barbara de Castro

Brena Rolim Peixoto da Silva

Cayo César Freire Feliciano

Cláudio Passos de Oliveira

Clícia Silveira Rodrigues

Olavo Mesquita de Carvalho

Regina Marta de Aguiar

Reginaldo da Silva Domingos

Jennyfer Alves de Carvalho (estagiária)

Katarinne Fabrizzi Maciel do Couto (estagiária)

Editoração

Anderson Silva Reis

Augusto Lopes dos Santos Borges

Cristiano Ferreira de Araújo

Daniel Alves Tavares

Danielle de Oliveira Ayres

Leonardo Hideki Higa

Natália de Oliveira Ayres

Capa

Aline Cristine Torres da Silva Martins

Projeto Gráfico

Aline Cristine Torres da Silva Martins

The manuscripts in languages other than Portuguese published herein have not been proofread.

Ipea – Brasília

Setor de Edifícios Públicos Sul 702/902, Bloco C

Centro Empresarial Brasília 50, Torre B

CEP: 70390-025, Asa Sul, Brasília-DF

Missão do Ipea
Aprimorar as políticas públicas essenciais ao desenvolvimento brasileiro por meio da produção e disseminação de conhecimentos e da assessoria ao Estado nas suas decisões estratégicas.