São José dos Campos - SP
LV SBPO
LV Simpósio Brasileiro de Pesquisa Operacional
São José dos Campos, SP - 6 a 9 de novembro de 2023

Modelos de Simulação Estocástica de Cenários de Velocidade do Vento com Incorporação de Variáveis Climáticas

Rafael Couto
Pontifícia Universidade Católica do Rio de Janeiro - Departamento de Engenharia Industrial
Gávea, Rio de Janeiro
rafael.arcouto@gmail.com

Paula Maçaira
Pontifícia Universidade Católica do Rio de Janeiro - Departamento de Engenharia Industrial
Gávea, Rio de Janeiro
paulamacaira@puc-rio.br

Fernando Cyrino
Pontifícia Universidade Católica do Rio de Janeiro - Departamento de Engenharia Industrial
Gávea, Rio de Janeiro
cyrino@puc-rio.br

RESUMO

A energia eólica no Brasil está em um contínuo crescimento nos últimos anos. Com o objetivo de potencializá-la, é necessário atentar às mudanças climáticas pois a sua geração é altamente dependente do clima. Nesse sentido, é necessária a introdução de variáveis climáticas exógenas para a modelagem de séries eólicas, servindo como fator de diminuição de incertezas. Para cumprir essa função, modelos Periódicos Autorregressivos com variáveis exógenas (PARX) podem ser utilizados. Quando comparados com o modelo vigente do NEWAVE (PAR), a modelagem nos estados do Rio Grande do Norte e Rio Grande do Sul utilizando o PARX com a variável exógena ENSO, verifica-se que os modelos aqui propostos possuem um desempenho superior na modelagem das séries de velocidade do vento.

PALAVRAS CHAVE. Velocidade do Vento, PARX, ENSO.

TÓPICOS. Estatística e Modelos Probabilísticos; PO na Área de Energia, Petróleo e Gás.

ABSTRACT

Wind energy in Brazil has been in continuous growth in recent years. In order to enhance it, it is necessary to pay attention to climate change, as its generation is highly dependent on the climate. In this sense, it is necessary to introduce exogenous climatic variables for the modeling of wind series, serving as a factor to reduce uncertainties. To fulfill this function, Periodic Autoregressive models with exogenous variables (PARX) can be used. When compared with the current model of NEWAVE (PAR), the modeling in the states of Rio Grande do Norte and Rio Grande do Sul using PARX with the exogenous variable ENSO, it appears that the models proposed here have a superior performance in modeling the wind speed series.

KEYWORDS. Wind Speed, PARX, ENSO.

TOPICS. Statistics and Probabilistic Models; OM in the Energy, Oil and Gas Area.

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

1. Introdução

Atualmente, a eletricidade no Brasil é produzida principalmente por fontes renováveis, que correspondem por mais de 80% do total da capacidade de geração de eletricidade, com destaque para as hidrelétricas, cuja proporção na matriz elétrica do país chega a mais de 65% [EPE, 2021]. No entanto, fontes térmicas ainda são necessárias para atender parte da demanda interna, principalmente em períodos de estiagem, que podem afetar os reservatórios hídricos, sendo necessário, para suprir essa carência, acionar as usinas termelétricas movidas a combustíveis fósseis em modo contínuo e em sua capacidade máxima [Mendes, 2014]. Para suprir essa demanda de forma renovável é importante ampliar outras fontes além da térmica, como a eólica, no qual há um crescimento exponencial da instalação de novas usinas nos últimos anos, onde a capacidade instalada de geração eólica cresceu mais de 300%, isto é, saltou de 178 GW em 2010 para cerca de 734 GW em 2020 [IRENA, 2021].

Devido à incerteza associada ao planejamento energético no Brasil, exige-se a modelagem estocástica das séries de energia da maneira mais eficiente possível. Atualmente, tal processo é realizado pelo NEWAVE a partir de um modelo Periódico Autorregressivo (PAR) para geração integrada de séries sintéticas de velocidades de ventos correlacionadas às afluências dos reservatórios das hidrelétricas, com base em Maceira et al. [2022]. Entretanto, como este artigo trabalhará somente com velocidade do vento, será considerada a aplicação do PAR para a velocidade, ajustado com base nos parâmetros estimados do histórico da mesma. Esse modelo de geração de cenários assume que as séries eólicas são estacionárias, lineares e seguem distribuição Normal, dentre outras características [Noakes et al., 1985]. Além disso, é um modelo univariado, isto é, não considera qualquer informação exógena que possa afetar os regimes de vento e, consequentemente, a produção de energia.

Dessa forma, com o objetivo de potencializar a energia eólica, é necessário atentar às mudanças climáticas atuais, pois a geração de energia elétrica a partir de fontes renováveis é altamente dependente do clima, cujas condições podem afetar a produção em diversas escalas como horária, mensal ou sazonal. Nesse sentido, estudos apontam que o uso de variáveis climáticas na modelagem de séries de velocidade do vento pode servir como fator de diminuição de incertezas, tanto no Brasil [do Nascimento Camelo et al., 2018; de Mattos Neto et al., 2021], como no mundo [Yang et al., 2005; Zhang et al., 2016]. As informações climáticas mais utilizadas são pressão, temperatura e precipitação, entretanto o fenômeno ENSO também pode ser uma alternativa, uma vez que Arpe [2020] identificou a sua forte relação com a velocidade do vento e Maçaira [2018] realizou uma profunda análise entre ENSO e um outro recurso primário (vazão), onde encontrou sucesso.

Junto à isso, Maçaira et al. [2018] realizaram uma Revisão Sistemática da literatura e identificou que Modelos de Regressão, Redes Neurais, ARIMAX, *Support Vector Machine* e Modelos Estruturais figuram entre os mais utilizados em estudos que desejam a incorporação de variáveis exógenas. Para cumprir essa função, esse artigo propõe uma alternativa, que já obteve sucesso com Maçaira [2018] para vazão, que tem a utilização de modelos Periódicos Autorregressivos com variáveis exógenas (PARX), com a expectativa de que a abordagem tenha capacidade de modelagem e simulação superior a atualmente empregada.

Portanto, o objetivo desse artigo é a proposição de uma abordagem metodológica que seja capaz de incorporar variáveis climáticas de forma a aperfeiçoar a modelagem e simulação por meio da adição de uma variável explicativa ao modelo PAR, utilizando o modelo PARX para dados de reanálise de velocidade do vento no Brasil, no qual foram escolhidos para o estudo, os estados do Rio Grande do Norte (RN) e Rio Grande do Sul (RS).

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

Este trabalho está organizado em quatro capítulos. Após este Capítulo 1 introdutório, a metodologia aplicada é apresentada no Capítulo 2. O Capítulo 3 traz os resultados e suas implicações. Por fim, no Capítulo 4 serão descritas as conclusões obtidas nesta pesquisa.

2. Metodologia

Este capítulo se inicia com a apresentação dos dados e seus tratamentos, seguindo para as diferentes modelagens aqui trabalhadas e fechando com processamento dos resultados obtidos na modelagem.

2.1. Pré-Processamento

Para as análises deste artigo será utilizada a velocidade do vento porque a transformação do vento em energia depende de vários fatores como o número de turbinas podendo ocultar correlações com variáveis climáticas. Foram selecionadas os estados do RN e RS, onde os dados dessas regiões foram obtidos a partir do banco de dados de reanálise MERRA-2, sobre os quais foram calculadas as médias mensais de alguns pontos geográficos de cada estado para representar as séries históricas de velocidade do vento para altura de 100 m, no período de janeiro de 1980 a dezembro de 2020.

Já para os dados de ENSO, no qual o *El Niño* se refere ao aquecimento das águas do Oceano Pacífico e *La Niña*, ao resfriamento, estes foram obtidos a partir do NOAA [2023], desde o período de 1931 até 2022, tendo a data de início variando entre os índices do ENSO. Entretanto, já que os dados de vento se iniciaram em 1980, os dados de ENSO seguiram esta data. Em cima da base de dados foram criadas duas novas variáveis, sendo a primeira o índice acumulado ao longo do tempo pois é interessante investigar também se as séries acumuladas apresentam algum tipo de tendência, o que pode indicar um aumento ou não da temperatura do mar; a segunda foi a identificação do períodos do ENSO: *El Niño*, *La Niña* ou Neutro.

2.2. Modelagem

2.2.1. Modelo Periódico Autorregressivo (PAR)

A metodologia sobre o modelo PAR é explicada de modo aprofundado em Maçaira [2018]. Em suma, de acordo com Hipel & McLeod [1994], ajustar um modelo PAR a uma série sazonal é similar a ajustar um modelo autorregressivo (AR) para cada período do ano, isto é, uma série com periodicidade mensal será modelada de forma que cada mês possua seu próprio modelo AR. Por isso, o PAR também pode ser chamado de PAR(p), onde p corresponde a ordem do modelo.

Seguindo a notação comumente utilizada ao se referir ao modelo PAR, seja Z uma série com S períodos e N número de anos, então $Z = [z_{(1,1)}, z_{(1,2)}, \ldots, z_{(1,S)}, \ldots, z_{(N,S)}]$. O modelo PAR da série Z no período m está matematicamente descrito na Equação 1.

$$\left(\frac{z_{(t,m)} - \mu_m}{\sigma_m}\right) = \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}}\right) + a_{t,m}, \tag{1}$$

onde μ_m é a média do período m, σ_m é o desvio padrão do período m, $\varphi_i^{(m)}$ é o i-ésimo coeficiente autorregressivo do período m, p_m é a ordem do operador autorregressivo do período m e $a_{t,m}$ é a série de ruídos independentes com média 0 e desvio padrão σ_m^a.

Existem, basicamente, duas maneiras de determinar a ordem autorregressiva de cada mês: a partir da análise da função de autocorrelação periódica (PeACF) e da função de autocorrelação parcial periódica (PePACF) [McLeod, 1995]; ou através dos critérios de informação, como o *Akaike's Information Criterion* (AIC) e o *Bayesian Information Criterion* (BIC), que são funções penalizadoras para a seleção de modelos e medidas relativas da qualidade do ajuste, com base no número de parâmetros e na função de verossimilhança [Schwarz, 1978].

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

Por conta do BIC ser melhor na seleção de um modelo dentre um grupo este será o utilizado para selecionar as ordens de cada período das séries de velocidade do vento do modelo PAR, ou seja, a ordem escolhida de cada período será aquela que apresentar menor BIC.

$$BIC = \ln(n)k - 2\ln(\hat{L}) \tag{2}$$

Após definir a ordem do modelo é necessário realizar a estimação dos parâmetros $\varphi_i^{(m)}$. Seja $\beta_m = (\varphi_1^{(m)}, \ldots, \varphi_{p_m}^{(m)})$ o vetor de parâmetros autorregressivos do período m. Um estimador assintoticamente eficiente, $\hat{\beta}_m$, pode ser obtido ao resolver as Equações 3,

$$\gamma_l^{(m)} = \sum_{i=1}^{p_m} \hat{\varphi}_i^{(m)} \gamma_{l-i}^{(m-i)}, l = 1, \ldots, p_m. \tag{3}$$

2.2.2. Modelo Periódico Autorregressivo com uma variável exógena (PARX)

Seja Z a série periódica já definida anteriormente e X a variável exógena na modelagem de Z, com o mesmo número de observações ($N \times S$) e com a mesma periodicidade (S) de Z. Segundo Ursu & Pereau [2017] e Silveira et al. [2017], o modelo periódico autorregressivo com uma variável exógena (PARX) da variável dependente Z e da variável exógena X pode ser matematicamente descrito como:

$$\left(\frac{z_{(t,m)} - \mu_m}{\sigma_m}\right) = \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}}\right) + \sum_{j=0}^{v_m} \theta_j^{(m)} \left(\frac{x_{(t,m-j)} - \mu_{m-j}^{(x)}}{\sigma_{m-j}^{(x)}}\right) + a_{t,m} \tag{4}$$

onde, μ_m é a média da variável dependente Z do período m, σ_m é o desvio padrão da variável dependente Z do período m, $\varphi_i^{(m)}$ é o i-ésimo coeficiente autorregressivo da variável dependente Z do período m, p_m é a ordem do operador autorregressivo da variável dependente Z do período m. $\mu_m^{(x)}$ é a média da variável independente X do período m, $\sigma_m^{(x)}$ é o desvio padrão de X do período m, $\theta_j^{(m)}$ é o j-ésimo coeficiente autorregressivo da variável exógena X do período m, v_m é a ordem do operador autorregressivo da variável exógena X do período m, e $a_{t,m}$ é a série de ruídos independentes com média 0 e desvio padrão σ_m^a.

Para definição das ordens autorregressivas da variável dependente e da variável exógena para cada período (p_m, v_m) será utilizado novamente o critério BIC. No caso do modelo PARX, para cada período, o BIC obtido será referente ao conjunto (p_m,v_m), isto é, o conjunto de parâmetros que obtiver menor BIC será o escolhido.

A estimação dos parâmetros do modelo, assim como para o modelo PAR, é realizada via Mínimos Quadrados Ordinários [Ursu e Pereau, 2017]. Considere $Y_{ns+m} = \left(\frac{z_{(t,m)} - \mu_m}{\sigma_m}\right)$ e $X_{ns+m} = \left(\frac{x_{(t,m-j)} - \mu_{m-j}^{(x)}}{\sigma_{m-j}^{(x)}}\right)$, onde $n = 0, \ldots, N-1$ e $m = 1, \ldots, s$, com tamanho Ns. Seja $\mathbf{w}_m = [Y_m, Y_{m+s}, \ldots, Y_{(N-1)s+m}]$ e $\mathbf{a}_m = [a_m, a_{m+s}, \ldots, a_{(N-1)s+m}]$ vetores de dimensão $(N \times 1)$ com T sendo o operador de transposição e $\mathbf{W}_m = [\mathbf{Y}_m \mathbf{X}_m]$ a matriz com dimensão $N \times (p_m + 1 + v_m)$, onde $\mathbf{Y}_m$ e $\mathbf{X}_m$ são descritos abaixo

$$\mathbf{Y}_m = \begin{bmatrix} Y_{m-1} & Y_{m-2} & \cdots & Y_{m-p_m} \\ Y_{s+m-1} & Y_{s+m-2} & \cdots & Y_{s+m-p_m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{(N-1)s+m-1} & Y_{(N-1)s+m-2} & \cdots & Y_{(N-1)s+m-p_m} \end{bmatrix}; \mathbf{X}_m = \begin{bmatrix} X_m & X_{m-1} & \cdots & X_{m-v_m} \\ X_{s+m} & X_{s+m-1} & \cdots & X_{s+m-v_m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{(N-1)s+m} & X_{(N-1)s+m-1} & \cdots & X_{(N-1)s+m-v_m} \end{bmatrix} \tag{5}$$

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

Seja

$$\beta_m = \left(\varphi^{(m)}, \theta^{(m)}\right)^T$$

o vetor paramétrico, onde

$$\varphi^{(m)} = \left(\varphi_1^{(m)}, \ldots, \varphi_{p_m}^{(m)}\right)^T; \theta^{(m)} = \left(\theta_1^{(m)}, \ldots, \theta_{v_m}^{(m)}\right)^T.$$

Dado que a Equação 4 é um modelo linear, pode ser escrita na forma de modelo de regressão:

$$\mathbf{w}_m = \mathbf{W}_m \beta_m + \mathbf{a}_m, m = 1, \ldots, s.$$

A matriz de variância-covariância do vetor aleatório $\mathbf{a}_m$ é $\sigma_m^2 \mathbf{I}_N$, onde $\mathbf{I}_N$ é a matriz identidade de tamanho N. O estimador de mínimos quadrados de β_m é obtido ao minimizar a Equação 6.

$$S(\beta) = \sum_{m=1}^{s} \mathbf{a}_m^T \mathbf{a}_m = \sum_{n=0}^{N-1} \sum_{m=1}^{s} \left(Y_{ns+m} - \sum_{i=1}^{p_m} \varphi_i^{(m)} Y_{ns+m-i} - \sum_{i=1}^{v_m} \theta_j^{(m)} X_{ns+m-j} \right)^2 \tag{6}$$

Finalmente, ao aplicar o operador de diferença na Equação 6 tem-se os estimadores via mínimos quadrados para $\hat{\beta}_m = \left(\varphi^{(m)}, \theta^{(m)}\right)^T$.

$$\hat{\beta}_m = \left\{ \mathbf{W}_m^T \mathbf{W}_m \right\}^{-1} \mathbf{W}_m^T \mathbf{w}_m.$$

2.3. Pós-Processamento

2.3.1. Métricas de Erro

Serão utilizadas duas métricas de avaliação de erro nesse trabalho, a primeira é a *Mean Absolute Relative Error* (MARE) e a segunda métrica é chamada de *Mean Square Relative Error* (MSRE), dadas por

$$\text{MARE} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{f_t - y_t}{y_t} \right|; \text{MSRE} = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{f_t - y_t}{y_t} \right)^2 \tag{7}$$

onde y_t é a velocidade do vento observada no momento t, f_t é a velocidade do vento prevista pelo modelo no tempo t, $\bar{y}$ é a media observada e T é o tamanho do horizonte de previsão.

2.3.2. Geração de Cenários Sintéticos de Velocidade do Vento

Após a seleção do melhor modelo serão construídos cenários sintéticos de velocidade do vento. Portanto, o objetivo do procedimento a ser descrito nesta seção é dado um dos modelos PAR ou PARX ajustado à série original, aproximar o comportamento estocástico e, sinteticamente, gerar tantas novas séries temporais quanto se queira, diferentes do histórico original, mas igualmente possíveis do ponto de vista estatístico. A geração detalhada de cenários pode ser vista em Maçaira [2018].

3. Resultados

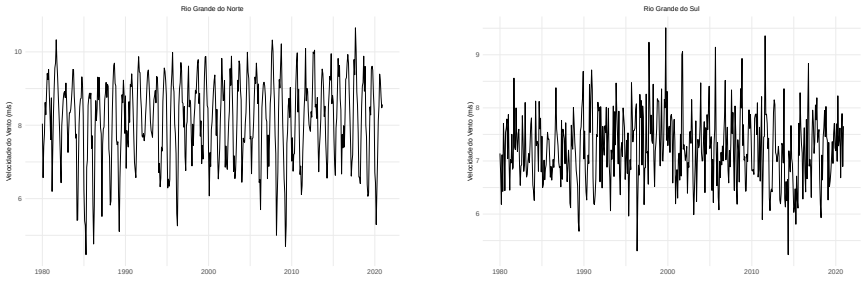
Este capítulo começa com uma análise descritiva dos dados de velocidade do vento e ENSO, tal como a relação existente entre ambas. Em seguida o passo a passo feito na modelagem do PAR e do PARX é apresentado, junto a análise da modelagem por medidas de erro, levando a obtenção de cenários sintéticos de velocidade do vento com o melhor modelo PARX para cada estado.

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

3.1. Análise descritiva dos dados

3.1.1. Velocidade do Vento

Inicialmente, a Figura 1 apresenta as séries temporais de velocidade média do vento para os estados do Rio Grande do Norte (RN) e Rio Grande do Sul (RS) no período de janeiro de 1980 a dezembro de 2020.



(a) RN (b) RS

Figura 1: Séries Temporais de Velocidade do Vento por Estado. Fonte: Elaborado pelo autor.

Observando os gráficos, percebe-se que, no geral, os estados apresentam um comportamento sazonal bem marcado, com destaque para RN. É importante também analisar as estatísticas descritivas das séries temporais, mostradas na Tabela 1. Observa-se que apresentam o coeficiente de variação similar. Quanto à assimetria, todos os valores localizam-se no intervalo [-1, +1], típico de distribuições levemente assimétricas. Já em relação à curtose, note que, ambos estados apresentam curtose em torno de três, mostrando que tais distribuições possuem uma curva de frequência próxima à distribuição normal. No que tange às medidas de centralidade, verifica-se a proximidade entre os valores de média e mediana para os estados. Além disso, a maior média é observada, conforme esperado, no RN (8,18 m/s) localizado no Nordeste.

Estado	Média	Mediana	Desvio Padrão	Coeficiente de Variação	Assimetria	Curtose
Rio Grande do Norte	8,18	8,41	1,14	0,14	-0,60	2,96
Rio Grande do Sul	7,24	7,23	0,65	0,09	0,20	3,36

Tabela 1: Estatísticas Descritivas das Séries de Velocidade do Vento por Estado. Fonte: Elaborado pelo autor.

Para comprovar a estacionariedade ou não das séries, aplica-se, o teste de Dickey-Fuller Aumentado [1984]. Como os p-valores são todos inferiores a um nível de significância de 5%, há evidências estatísticas suficientes para rejeitar a hipótese nula de não-estacionariedade em ambos os testes para os dois estados.

3.1.2. ENSO

O indicador do ENSO mais antigo que é o *Southern Oscillation Index* (SOI) e tem seu cálculo baseado na diferença de pressão atmosférica ao nível do mar entre as regiões do Tahiti, no Pacífico Ocidental, e Darwin, na Austrália (Pacífico Oriental). O fato do SOI se basear na pressão do nível do mar em apenas dois pontos mostra que pode ser afetado por flutuações de curto prazo, dia a dia ou semana a semana, não relacionadas ao ENSO. Outra limitação do SOI é que tanto o Taiti quanto Darwin estão localizados ao sul da linha do Equador, enquanto o fenômeno ENSO se

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

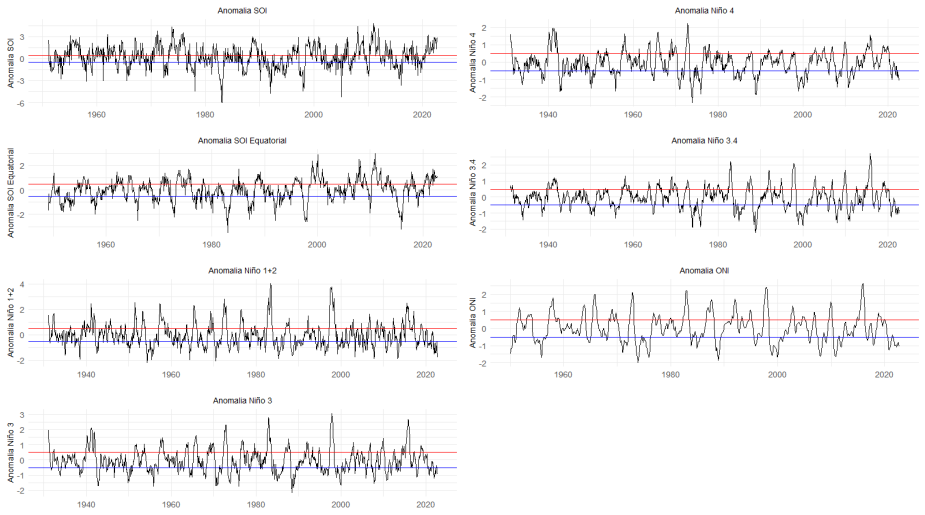


Figura 2: Séries Temporais das Anomalias dos Índices do ENSO. Fonte: Elaborado pelo autor.

concentra mais próximo à mesma linha. Tendo essa situação em vista, o SOI Equatorial supera tal limitação, uma vez que calcula a diferença da média de pressão atmosférica ao nível do mar entre duas regiões centradas na linha do Equador: a Indonésia e o Pacífico Oriental. Já as bases SST apresentam informações sobre a temperatura da superfície do mar (*Sea Surface Temperature*, SST) no Oceano Pacífico equatorial ao longo do tempo, pois é o principal indício da ocorrência do fenômeno [Bjerknes, 1969]. Inicialmente, certas regiões foram definidas para as medições: Niño 1, Niño 2 (combinado em Niño 1+2), Niño 3 e Niño 4. Mais tarde, uma área chamada Niño 3.4 foi identificada como sendo a mais representativa do ENSO [Barnston et al., 1997]. Localizada entre (e sobreposto com) Niño 3 e Niño 4, esta é a região cuja anomalia de temperatura é refletida pelo *Oceanic Niño Index* (ONI), com a diferença de que o ONI busca remover a tendência de aquecimento existente na região. Para constatar a ocorrência de um fenômeno ENSO são utilizadas as séries de anomalia dos índices, isto é, a diferença de valores entre o período anterior e o período atual, com a diferença que para o ONI, as suas séries de anomalias são calculadas através da média móvel de três meses em cima da diferença de valores entres os períodos. Em relação ao SOI, períodos consecutivos de índices positivos com cinco valores mensais acima de 0,5° evidenciam a ocorrência do fenômeno *La Niña*. Já períodos negativos com cinco valores mensais consecutivos abaixo de -0,5° evidenciam a presença da *El Niño*. No caso da SST, os períodos consecutivos de índices positivos com cinco valores mensais acima de 0,5° evidenciam a ocorrência do fenômeno *El Niño* e os períodos negativos com cinco valores mensais consecutivos abaixo de -0,5° evidenciam a presença da *La Niña*. Para todos os índices do ENSO, foram feitos os gráficos da Figura 2, apresentando os históricos mensais das variáveis de 1931-2022, onde a sequência de pontos acima da linha vermelho indicam a ocorrência de eventos *La Niña* ; *El Niño* e a sequência de pontos abaixo da linha azul, de eventos *El Niño* ; *La Niña*, seguindo as diretrizes mencionadas.

Já as séries acumuladas podem ser observadas na Figura 3, mostrando que, segundo os índices SOI e SOI Equatorial, há tendência de crescimento da pressão atmosférica na superfície do mar tanto entre as regiões do Taiti e Darwin, como entre a Indonésia e o Pacífico Oriental recentemente. Também na Figura 3 são apresentados os índices acumulados para SST e ONI. Note que após 1980, todos os índices, com exceção do ONI, apresentam tendência de crescimento.

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

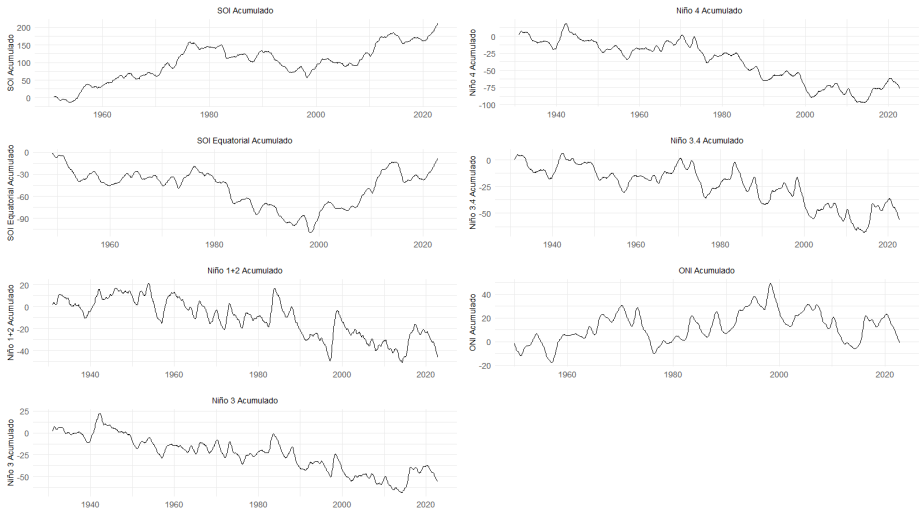


Figura 3: Séries Temporais as Anomalias Acumuladas dos Índices do ENSO. Fonte: Elaborado pelo autor.

3.1.3. Relação Velocidade do Vento e ENSO

A fim de avaliar estatisticamente as diferenças das distribuições, foi realizado o Teste de Kruskal-Wallis [1952] que é utilizado para comparar mais de dois grupos independentes, de tamanhos iguais ou não, com variável resposta quantitativa. O teste dispensa a pressuposição de normalidade e homoscedasticidade. As hipóteses do teste são as seguintes:

H_0: As k amostras são oriundas de uma mesma população.
H_1: Pelo menos uma das amostras vem de uma população distinta das demais.

Foi utilizado um nível de significância de 5%. Para a resposta dada pelo teste foi definido que ***Sim*** representaria que as fases possuem a mesma distribuição e, portanto, a hipótese nula não é rejeitada; ***Não*** representaria que as fases não possuem a mesma distribuição e, dessa forma, a hipótese nula é rejeitada.

Estado	SOI	SOI Equatorial	Niño 1+2	Niño 3	Niño 4	Niño 3.4	ONI
Rio Grande do Norte	0,03	0,88	0	0,9	0,19	0,6	0,8
	Não	Sim	Não	Sim	Sim	Sim	Sim
Rio Grande do Sul	0,22	0,03	0,19	0,03	0,55	0,3	0,08
	Sim	Não	Sim	Não	Sim	Sim	Sim

Tabela 2: P-Valor e Respostas do Teste de Kruskal-Wallis. Fonte: Elaborado pelo autor.

É possível notar que ambos estados possuem ao menos um índice comprovando que as distribuições entre as fases são distintas. Dessa forma, prova-se estatisticamente que o fenômeno climático ENSO influencia o comportamento da velocidade do vento no RN e RS.

3.2. Modelagem

Primeiramente os modelos propostos serão testados a partir do seu poder de previsão e os cenários de velocidade do vento serão gerados apenas com o modelo selecionado. Para verificar a performance de cada modelo, as bases de dados foram separadas em seis janelas de ajuste e previsão, descritas na Tabela 3.

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

Janela	***In-sample***	***Out-of-sample***
1	jan/1980-dez/2010	jan/2011-dez/2015
2	jan/1980-dez/2011	jan/2012-dez/2016
3	jan/1980-dez/2012	jan/2013-dez/2017
4	jan/1980-dez/2013	jan/2014-dez/2018
5	jan/1980-dez/2014	jan/2015-dez/2019
6	jan/1980-dez/2015	jan/2016-dez/2020

Tabela 3: Janelas de Ajuste e Previsão. Fonte: Elaborado pelo autor.

A janela 1 será utilizada para definir a escolha dos parâmetros $p(v)$ e $m(v)$ do PARX. Isto é, para cada estado e para cada combinação dos parâmetros deve-se realizar o passo-a-passo a seguir, considerando somente a primeira janela.

1. Ajustar a série de velocidade do vento para o período *in-sample*;

2. Simular o período *out-of-sample*;

3. Comparar os valores previstos (média dos cenários) com o observado;

4. Guardar os erros obtidos.

O parâmetro escolhido para cada estado e modelo será aquele que obtiver menor erro. Esses parâmetros serão utilizados nas janelas 2 a 6 para o cálculo das métricas de erro repetindo os passos da janela 1, porém somente para a melhor combinação de parâmetros na primeira janela. Os valores de erro apresentados nesta seção são obtidos ao realizar a média dos erros em cada uma das cinco janelas (2 a 6) com o objetivo de verificar de forma mais robusta o poder de previsão dos modelos, onde serão avaliadas pelas métricas MARE e MSRE.

O modelo PARX necessita da variável climática especificada para o período *out-of-sample*, porém, por conta da complexidade de previsão das séries dos índices ENSO, foram utilizados os próprios valores realizados no período *out-of-sample* para essas variáveis. Este ponto é uma limitação do artigo, porém o foco não é a previsão de variáveis climáticas e sim a inserção de tais variáveis nos modelos de previsão e/ou simulação, para tanto, além da utilização dos próprios valores observados sugere-se a utilização da previsão de órgãos oficiais como o próprio NOAA.

Na Figura 4 é apresentada a performance, segundo os critérios MARE e MSRE, de todos os 15 modelos testados para cada estado (em cinza), assim como a métrica para o modelo PAR (em vermelho) e o modelo PARX que obteve a menor métrica utilizando o fenômeno ENSO (em verde). É possível concluir que para os dois estados foi possível encontrar um modelo PARX que apresente erro menor que a metodologia PAR ao utilizar um índice que represente o El Niño e o La Niña. A Tabela 4 apresenta os modelos que obtiveram menor erro segundo as métricas MARE e MSRE, tal como a melhoria obtida em relação ao modelo PAR. Note que para o estado do RN, ambas medidas de erro indicaram o uso do índice Niño 1+2 Acumulado, chegando a uma melhoria de -13,04%, já para o estado do RS, as duas medidas de erro indicaram o uso do índice ONI Acumulado, com uma melhoria de até -7,05%.

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931



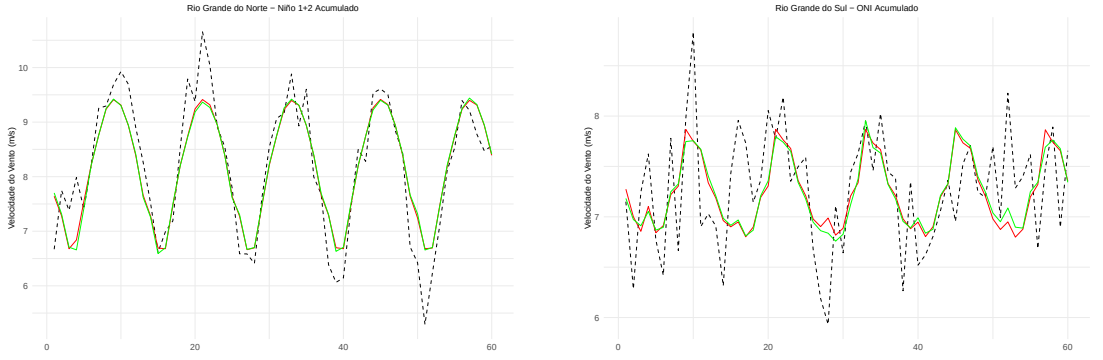
(a) MARE (b) MSRE

Figura 4: MARE e MSRE obtido para todos os 15 modelos (cinza), com o melhor modelo PARX que usa um índice ENSO (verde) e o PAR (vermelho) para cada estado. Fonte: Elaborado pelo autor.

Estado	PAR	Melhor modelo	Melhor PARX	Melhoria (%)
Rio Grande do Norte - MARE	0,051	Niño 1+2 Acumulado	0,0482	- 5,49
Rio Grande do Norte - MSRE	0,0046	Niño 1+2 Acumulado	0,004	- 13,04
Rio Grande do Sul - MARE	0,0718	ONI Acumulado	0,0678	- 5,57
Rio Grande do Sul - MSRE	0,008	ONI Acumulado	0,0074	- 7,5

Tabela 4: Melhores modelos segundo a métrica MARE e MRSE. Fonte: Elaborado pelo autor.

Na Figura 5 são apresentadas as velocidades do vento observadas durante o período de validação da janela 6 (jan/2016 a dez/2020) em preto, a previsão obtida ao utilizar o PAR em vermelho e com o melhor modelo PARX proposto com um índice ENSO em verde.



(a) RN (b) RS

Figura 5: Velocidade do Vento observada (preto), previsão obtida via modelo PAR (vermelho) e melhor modelo com índice ENSO (verde). Fonte: Elaborado pelo autor.

4. Conclusão

Devido o crescimento exponencial da energia eólica no Brasil [IRENA, 2021] frente às mudanças climáticas atuais, o objetivo desse artigo é a proposição de uma abordagem metodológica que seja capaz de incorporar variáveis climáticas de forma a aperfeiçoar a modelagem e simulação

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

por meio da adição de uma variável explicativa ao modelo Periódico Autorregressivo (PAR), que é o modelo atualmente utilizado no NEWAVE [Maceira et al., 2022], por meio do modelo Periódico Autorregressivo com variável exógena (PARX) para dados de reanálise de velocidade do vento do Rio Grande do Norte (RN) e Rio Grande do Sul (RS). Quando comparados com o modelo vigente (PAR) verifica-se que os modelos aqui propostos possuem um desempenho superior na modelagem das séries de velocidade do vento, concluindo que a inclusão de variáveis climáticas influencia significativamente na velocidade do vento dos estados brasileiros analisados e consequentemente a geração de energia do país. Em relação aos modelos que envolveram a utilização de variáveis climáticas, conclui-se que o Niño 1+2 Acumulado deve ser inserido para modelagem do estado do RN e o ONI Acumulado para o estado do RS. Enxerga-se como continuação deste trabalho a ampliação da análise para todos os estados brasileiros.

5. Agradecimentos

Este trabalho foi apoiado por: Coordenação Brasileira de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) [número 001]; o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) com Bolsas [números 422470/2021-0, 307084/2022-1 e 311519/2022-9], a Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) com Bolsas [números 210.618/2019, 211.086/2019, 211.645/2021, 201.243/2022, 201.348/2022 e 210.041/2023] e Programa de P&D da ANEEL e Grupo CTG BRASIL sob outorga nº 10381-0322/2022 (Portfolio Diversification with a View to Optimizing Revenue from Electricity Sales).

Referências

Arpe, K., Molavi-Arabshahi, M., e Leroy, S. A. G. (2020). Wind variability over the caspian sea, its impact on caspian seawater level and link with enso. *International Journal of Climatology*, 40 (14):6039–6054.

Barnston, A., Chelliah, M., e Goldenberg, S. (1997). Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmosphere-Ocean*, 35:367–383.

Bjerknes, J. (1969). Atmospheric teleconnections from the equatorial Pacific. *Journal of Physical Oceanography*, 97:163–172.

de Mattos Neto, P. S., de Oliveira, J. F., Júnior, D. S. d. O. S., Siqueira, H. V., Marinho, M. H., e Madeiro, F. (2021). An adaptive hybrid system using deep learning for wind speed forecasting. *Information Sciences*, 581:495–514.

do Nascimento Camelo, H., Lucio, P. S., Junior, J. B. V. L., e de Carvalho, P. C. M. (2018). A hybrid model based on time series models and neural network for forecasting wind speed in the brazilian northeast region. *Sustainable Energy Technologies and Assessments*, 28:65–72.

EPE (2021). Disponível em: https://www.epe.gov.br/pt/publicacoes-dadosabertos/publicacoes/balanco-energetico-nacional-ben. Acessado: 10-05-2023.

Hipel, K. e McLeod, A. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier.

IRENA (2021). URL https://www.irena.org/newsroom/pressreleases/2021/Apr/World-Adds-Record-New-Renewable-Energy-Capacity-in-2020. Acessado:10-05-2023.

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931

Kruskal, W. H. e Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260):583–621.

Maçaira, P., Thomé, A., Cyrino Oliveira, F., e de Almeida, F. (2018). Time series analysis with explanatory variables: A systematic literature review. *Environmental Modelling & Software*, 107:199—209.

Maceira, M., Melo, A., Pessanha, J., Cruz, C., Almeida, V., e Justino, T. (2022). Wind uncertainty modeling in long-term operation planning of hydro-dominated systems. In *2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, p. 1–6. IEEE.

Maçaira, P. M. M. (2018). *Modelos de Simulação Estocástica de Cenários de Vazão com Incorporação de Variáveis Climáticas*. PhD thesis, PUC-Rio.

McLeod, A. (1995). Diagnostic Checking Periodic Autoregression Models With Application. *The Journal of Time Series Analysis*, 15:221–233.

Mendes, L. F. R. (2014). Considerações acerca da geração hidroelétrica fluminense a partir da crise hídrica do estado de são paulo. *Boletim do Observatório Ambiental Alberto Ribeiro Lamego*, 8 (1):35–49.

NOAA (2023). Disponível em: `https://www.ncei.noaa.gov/access/monitoring/enso/`. Acesso em: 05 de Fevereiro de 2023.

Noakes, D., McLeod, A., e Hipel, K. (1985). Forecasting monthly riverflow time series. *International Journal of Forecasting*, 1(2):179–190. ISSN 01692070.

Said, S. e D, D. (1984). Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order. *Biometrika*, 71:599–607.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6:461–464.

Silveira, C., Alexandre, A., de Souza Filho, F., Vasconcelos Junior, F., e Cabral, S. (2017). Monthly streamflow forecast for National Interconnected System (NIS) using Periodic Auto-regressive Endogenous Models (PAR) and Exogenous (PARX) with climate information. *Brazilian Journal of Water Resources*, 22.

Ursu, E. e Pereau, J. (2017). Estimation and identification of periodic autoregressive models with one exogenous variable. *Journal of the Korean Statistical Society*, 46:629–640.

Yang, X., Xiao, Y., e Chen, S. (2005). Wind speed and generated power forecasting in wind farm. *Proceedings-Chinese Society of Electrical Engineering*, 25(11):1.

Zhang, C., Wei, H., Zhao, J., Liu, T., Zhu, T., e Zhang, K. (2016). Short-term wind speed forecasting using empirical mode decomposition and feature selection. *Renewable Energy*, 96:727–737.

DOI: 10.59254/sbpo-2023-174931
http://live-proceedings.pantheonsite.io/p/174931