

Previsão de velocidade do vento no nordeste brasileiro através do modelo ARIMAX

Henrique do Nascimento Carmelo¹, Paulo Sérgio Lucio², João Bosco Verçosa Leal Junior³, Paulo Cesar Marques de Carvalho⁴

¹Mestre, Professor do Instituto Federal do Ceará - IFCE, henriquecamelo13@yahoo.com.br

²Doutor, Professor da Universidade Federal do Rio Grande do Norte - UFRN, pslucio@ccet.ufrn.br

³Doutor, Professor da Universidade Estadual do Ceará -UECE, jblealjr@uece.br

⁴Doutor, Professor da Universidade Federal do Ceará -UFC, carvalho@dee.ufc.br

Resumo – O presente artigo utiliza o modelo Auto-Regressivo Integrado de Médias Móveis com entradas Exógenas (ARIMAX), o qual representa uma tipo de modelagem da metodologia Box–Jenkins, para fornecer previsões de médias mensais de séries temporais da velocidade do vento em regiões do nordeste brasileiro, ao nível de 10 m de altura com o uso do software livre R. Como variáveis exógenas optou-se pela utilização de pressão atmosférica e temperatura ambiente, as quais possuem características físicas que podem ser associadas à velocidade do vento. Os resultados da estatística de erros mostram a eficácia do modelo, com erro percentual próximo de 9 %, comparando-se as séries observadas e ajustadas. Além disso, foi possível identificar que as séries previstas conseguem acompanhar o perfil das séries observadas, através de semelhanças de valores mínimos e máximos da velocidade do vento. Esse resultado pode evidenciar o fato de que o modelo ARIMAX proposto para as regiões de estudo consegue captar a existência de sazonalidades. Esse trabalho poderá ter utilidade como uma possível ferramenta para auxiliar em tomadas de decisões nas diversas áreas do setor eólico, configurando-se em mais uma tentativa de garantias para reforçar a escolha de uma região na exploração de seu potencial eólico.

Palavras-chave: Energia eólica, séries temporais, variáveis exógenas, software R.

Abstract – The present paper uses the ARIMAX, which represents a kind of modeling of the Box-Jenkins methodology, to provide forecasts of monthly averages of wind-time time series in northeastern regions Brazil, at a level of 10 m in height with the use of free software R. As exogenous variables we opted for the use of atmospheric pressure and ambient temperature, which have physical characteristics that can be associated with the wind speed. The results of the error statistics show the efficacy of the model, with a percentage error close to 9%, comparing the observed and adjusted series. In ad-

dition, it was possible to identify that the predicted series can follow the profile of the observed series, through similarities of minimum and maximum values of the wind speed. This result may highlight the fact that the ARIMAX model proposed for the study regions can capture the existence of seasonality. This work could be useful as a possible tool to assist in decision making in the various areas of the wind sector, becoming another attempt of guarantees to reinforce the choice of a region in the exploration of its wind potential.

Keywords: Wind power, time series, exogenous variables, software R.

I. INTRODUÇÃO

As fontes renováveis de energias estão desempenhando papel fundamental no que se refere a redução de poluentes atmosféricos, entretanto, é necessário, ainda mais, conscientizar a sociedade sobre seus benefícios, em especial, no que se refere a capacidade que possuem de atenuar efeitos do aquecimento global em larga escala [1]. O Painel Intergovernamental sobre Mudanças Climáticas (IPCC¹), em seu mais recente relatório lançado 2015, apontou que o aquecimento global é uma realidade e a contribuição do ser humano é significativa para a ocorrência de fenômenos ligados às mudanças climáticas². Por isso, devemos agir imediatamente em escala global para reverter o que for possível. Nessa perspectiva muitos cientistas apontam como uma estratégia real uma maior inserção de fontes renováveis em matrizes energéticas de diversas nações [2].

A energia eólica é uma das fontes limpa e renovável que vem se tornando economicamente viável para geração de energia elétrica. No Brasil de acordo com Associação Brasileira de Energia Eólica (ABEEólica³), no tocante a capacidade instalada de geração eólica no mês de dezembro de 2016, o país contava com a potência instalada de aproximadamente 10,6 GW, o que representa um aumento significativo quando comparado ao ano de 2005, cuja potência instalada representava aproximadamente 27 MW. Ainda sobre esse aumento a ABEEólica⁴

¹De sua denominação em inglês *Intergovernmental Panel on Climate Change*. Maiores detalhes em sua página eletrônica < <http://www.ipcc.ch/> >.

² Consultado em (11/01/2017) na página eletrônica do IPCC em < <https://www.ipcc.ch/> >.

³Maiores detalhes podem ser obtidos em <<http://www.abeeolica.org.br/noticias/brasil-ja-tem-106-gw-de-capacidade-instalada-de-energia-eolica>>.

⁴<<http://www.abeeolica.org.br/wp-content/uploads/2016/12/Dados-Mensais-ABEEolica-Dezembro-2016.pdf>>.

também estima que até 2020 o país poderá ter algo em torno de aproximadamente 18 GW de potência instalada. Ressaltando que o Brasil possui grande capacidade de produzir eletricidade a partir de fontes renováveis de energia. Exemplo disso, em sua matriz energética a principal fonte é proveniente de hidrelétricas com aproximadamente 70% de energia elétrica gerada a partir desta fonte de acordo com o Balanço Energético Brasileiro de 2014 [3].

Esse aumento da capacidade instalada de geração eólica no Brasil, certamente é resultado de muitas pesquisas que fornecem garantias em termos econômicos e ambientais para instalação de parques eólicos em muitas partes do território nacional. Uma linha de pesquisa que poderá contribuir ainda mais nesse crescimento, trata-se da previsão de séries temporais da velocidade do vento, por exemplo, em Kavasseri e Seetharaman [4], os autores utilizam o modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA), o qual também é conhecido por metodologia Box-Jenkins [5], para a previsão da velocidade do vento no horizonte de 24 h em quatro localidades da região de Dakota do Norte nos EUA.

Outra metodologia que vem sendo estudada para previsão de velocidade do vento é a utilização de Redes Neurais Artificiais (RNA), como em Cadenas e Rivera [6] os autores realizam previsão a curto prazo horária na região de La Venta, Oaxaca, México. Ainda outro modelo que vem sendo estudado para previsão de velocidade do vento é a Suavização Exponencial Simples (SES), a qual trabalha com médias ponderadas no sentido de fornecer pesos maiores às observações mais recentes, por exemplo, em Cadenas *et al.* [7] os autores mostram que a SES fornece menores erros quando comparado aos modelos ARIMA e RNA aplicados em região litorânea do México. Ainda é possível encontrar trabalhos com a combinação dos modelos ARIMA e RNA para previsão de séries temporais, ou seja, são os chamados modelos híbridos como em Liu *et al.* [8] os autores analisam a utilização desse tipo de modelagem na previsão da velocidade do vento em fazendas eólicas da china.

Conforme comentado provavelmente o sucesso da geração eólica em muitos lugares do mundo, como é o caso do Brasil, é resultado de muitas pesquisas que garantem a viabilidade dessa fonte na geração de energia elétrica para atender a demanda. E para que esse crescimento tenha, cada vez mais expressão, e inserido na matriz energética nacional, é necessário a continuidade de pesquisas do setor eólico em especial de forma regionalizada, levando em consideração as características das condições meteorológicas locais. Nesse sentido o presente trabalho pretende mostrar metodologia derivada de Box-Jenkins, através do modelo Auto-Regressivo Integrado de Médias Móveis com entrada Exógenas (ARIMAX), capaz de realizar previsões da velocidade do vento, em termos de médias mensais um estudo de caso

para regiões do nordeste brasileiro. Além disso será verificado se ocorrem mudanças nas previsões obtidas, com a inclusão ou não da precipitação local no modelo proposto.

II. MATERIAIS E MÉTODOS

Nessa seção comenta-se sobre os dados das regiões de estudo, e também sobre o modelo de previsão proposto. Ainda será mostrado os métodos estatísticos de erros utilizados para identificar a acurácia das séries temporais ajustadas produzidas pelo modelo quando comparado as séries temporais observadas. Todos os cálculos produzidos nesse estudo e da mesma forma a parte gráfica foram executados pelo *software* livre R [9], o qual vem sendo muito utilizado em diversas atividades acadêmicas.

A. Região de estudo

As regiões de estudo nesse trabalho tratam-se das localidades de Fortaleza (lat: -3,770 e lon: -38,530), Parnaíba (lat:-2,540 e lon:-41,460) e Natal (lat:-5,750 e lon:-35,350) todas no nordeste do Brasil. Os dados meteorológicos usados nesse trabalho são: velocidade do vento (variável dependente, medida em m/s), pressão, temperatura e precipitação (variáveis independentes que também são chamadas de variáveis exógenas, as quais serão explicadas adiante, medidas em hpa, °C e mm, respectivamente). Todos esses dados são constituídos de médias mensais com 144 medições que começam em janeiro de 2003 e terminam em dezembro de 2014, sendo que todos esses dados servirão de entrada no modelo proposto, o qual será executado no *software* R. Os dados da velocidade do vento são originalmente medidos ao nível de 10 m de altura, e foram coletados de torres anemométricas dos aeroportos das respectivas localidades, e da mesma forma também os dados de pressão, temperatura e precipitação. Os dados citados foram doados pelo projeto Airmetar que surgiu da necessidade de ferramentas meteorológicas para consultar e analisar informações dos aeroportos do Brasil através da consulta e decodificação dos dados METAR⁵ fornecidos pela aeronáutica do Brasil.

B. Modelagem Box - Jenkins

Para realizar previsões da velocidade do vento de nossas regiões de estudo foi utilizado modelagem do tipo Box-Jenkins, a qual refere-se ao método sistemático de identificação, ajuste, checagem e uso de modelos Auto-Regressivos Integrados à Média Móvel, cuja a abreviatura é comumente dada por ARIMA⁶. O uso de modelos ARIMA é uma abordagem poderosa na solução de muitos problemas de previsão, pois pode proporcionar previsões, extremamente acuradas, de séries temporais [10]. Como uma série temporal tem dados coletados sequencialmente ao longo do tempo, espera-se que ela

⁵Maiores detalhes conferir na página eletrônica <<http://www.airmetar.com.br/>>.

⁶Do inglês (*Auto Regressive Integrated Moving Average*).

apresente correlação seriada no tempo. Os modelos ARIMA são modelos matemáticos que visam captar o comportamento da correlação seriada ou autocorrelação entre os valores da série temporal, e, com base nesse comportamento, realizar previsões futuras [11].

A modelagem ARIMA é comumente aplicada em processos lineares não estacionários, supõem que as séries temporais não são estacionárias (que há movimentos de tendência seja de crescimento ou decaimento de qualquer natureza). Para torná-las estacionárias, deve-se aplicar um número determinado de diferenças entre seus dados de ordem (d), geralmente uma ou duas [12]. Para esses processos o modelo é comumente representado por ARIMA(p, d, q). O número de vezes que a série original tem que ser diferenciada antes de resultar uma série estacionária é denominado “ordem de homogeneidade”. Estes processos são descritos pelos modelos Auto-Regressivos – Integrados (número de diferenciações) – Médias Móveis de ordens p, d, e q, ARIMA(p, d, q) que podem ainda, ter incluído um operador sazonal, representado então por SARIMA(p, d, q)(P, D, Q), sendo P, D, e Q os termos encontrados da parte sazonal da série temporal em questão, ou seja, semelhante aos descritos para o ARIMA, maiores detalhes podem ser obtidos em [12].

O modelo propriamente utilizado nesse trabalho trata-se do ARIMAX, o qual é derivado do ARIMA. A diferença entre os dois é que o modelo Auto-Regressivo (Auto-Regressive) Integrado (Integrated) de Médias Móveis (Moving Average) e Entradas Exógenas (Exogenous inputs) possui, além dos parâmetros auto-regressivos e de médias móveis, a entrada exógena [13]. O modelo ARIMAX pode ser entendido como a combinação dos modelos Auto-Regressivo AR(p), Integrado (d), Média-Móvel MA(q) e Exógeno X(r), e que pode então ser simbolizado por ARIMAX(p, d, q, r). Uma forma simplificada de representar matematicamente esse modelo está descrito na Equação 1:

$$y_t = \rho + \sum_{i=1}^p \beta_i y_{t-i} + \sum_{j=1}^r \omega_j w_j + \sum_{j=1}^q (\theta_j \varepsilon_{t-j}) + \varepsilon_t, \quad (1)$$

em que y_t é variável dependente no tempo t e que se pretende projetar (ou seja, a velocidade do vento); ρ é uma constante; y_{t-i} é a variável dependente (que também é a velocidade do vento) defasada por i passos de tempo; β_i é o coeficiente de y_{t-i} ; p é o número máximo de intervalos de tempo; w_j representa as variáveis exógenas (nesse caso foram incluídas de duas formas no modelo: I - pressão e temperatura; e II - pressão, temperatura e precipitação) do modelo; ω_j representa os coeficientes das variáveis exógenas; r é o número máximo de variáveis exógenas; θ_j é o coeficiente do termo de ε_{t-j} que por sua vez representa o erro no tempo t defasado de j. E ε_t é a componente de erro do modelo, com $\varepsilon_t \sim N(0, \sigma^2)$. Os coeficientes dos modelos são estimados por regressão, maiores detalhes sobre a Equação (1) [14]. Ressalta-se que a estratégia de utilizar duas formas distintas de variáveis exógenas no modelo ARIMAX, tem como obje-

tivo identificar se ocorrem mudanças nas previsões obtidas, com ou sem a inclusão da precipitação local no modelo.

Para avaliar se os modelos ARIMAX usados nesse trabalho possuem viabilidade para a previsão da velocidade do vento foi realizado uma análise de resíduos. Conforme em Shumway e Stoffer [15], para que o modelo ARIMAX seja viável no tocante aos ajustes dos dados observados, o termo de erro ε_t , deste modelo deve se comportar como um ruído branco, ou seja, possuir média zero, variância constante (são homocedásticos) e ser não correlacionado, ou seja, os erros são independentes. E ainda o termo ε_t deve seguir uma distribuição normal. Para identificação desses pressupostos existem alguns testes possíveis de serem aplicados sobre os resíduos. Dentre eles destaca-se o teste de verificação de normalidade dos resíduos a partir do Teste de Shapiro-Wilk, maiores detalhes [16], teste de independência dos resíduos a partir do Teste de Durbin-Watson, maiores detalhes [17], e o teste que verifica a igualdade da variância dos resíduos, ou seja, hipótese de homoscedasticidade a partir do Teste de Breusch-Pagan, maiores detalhes em Gujarati e Porter [18]. Na Tabela 1, tem-se a representação resumida dos testes utilizados, sendo que os mesmos foram aplicados ao nível de significância de $\alpha = 0,05$.

Tabela 1- Apresentação dos testes que identificam ruído branco do modelo ARIMAX.

Teste de Shapiro-Wilk
<p>H_0: amostra provém de uma população Normal. H_1: amostra não provém de uma população Normal. Tomar decisão: se o p-valor for maior do que α, ou seja, $p > 0,05$ (não rejeita H_0).</p>
Teste Durbin-Watson
<p>H_0: os resíduos são independentes. H_1: os resíduos não são independentes. Tomar decisão: se o p-valor for maior do que α, ou seja, $p > 0,05$ (não rejeita H_0).</p>
Teste Breusch-Pagan
<p>H_0: os resíduos possuem homoscedasticidade. H_1: os resíduos possuem heterocedasticidade. Tomar decisão: se o p-valor for maior do que α, ou seja, $p > 0,05$ (não rejeita H_0).</p>

Ainda sobre a escolha das variáveis exógenas aplicadas sobre o modelo ARIMAX, é importante falar que a velocidade do vento está relacionada com deslocamentos de massas de ar que, por sua vez, esses deslocamentos ocorrem em virtude de diferenças de pressão na atmosfera [19]. O movimento sempre acontece das zonas de alta para as zonas de baixa pressão e determinam, assim, a dinâmica geral da circulação atmosférica e as formações dos diferentes tipos climáticos [20]. Ressalta-se que massas de ar podem ser definidas como porções ou

volumes da atmosfera que possuem praticamente as mesmas características de pressão, temperatura e umidade por causa de sua localização e são bastante espessas e homogêneas [20]. É importante dizer também que o ar é composto principalmente de nitrogênio, oxigênio e argônio, que juntos constituem a maior parte dos gases da atmosfera. Os demais gases incluem gases do efeito estufa como vapor de água, dióxido de carbono, metano, óxido nitroso e ozônio. Variáveis como pressão, volume e temperatura são consideradas grandezas intrínsecas de quaisquer gases [21], dessa forma, pela importância das variáveis pressão e temperatura no tocante ao deslocamento de massas de ar, optou-se nesse trabalho pela relação de ambas com a velocidade do vento na aplicação do modelo proposto.

Para o cálculo das séries temporais ajustadas foi necessário a utilização da função `arimax(...)`, como exemplo: `arimax(dados-velocidade, order=c(2, 2, 1), seasonal=list(order=c(2,0,0), period=12), include.mean=T, xreg=dados-exógenos)`. Essa função faz parte do pacote (`forecast`) do software R, e foi desenvolvida por Hyndman e Khandakar [22]. De forma que para a utilização do modelo ARIMAX nesse trabalho, leva-se em consideração a existência de sazonalidade da série temporal em questão uma vez que trata-se de séries temporais meteorológicas. A ordem do modelo ARIMAX, ou seja, os números que dão origem ao funcionamento da função foram encontrados através de outra função, a `auto.arima(...)` que também foi criada por Hyndman e Khandakar em [22] que leva em consideração uma série de critérios para encontrar o modelo que irá fornecer o melhor ajuste como, por exemplo, o critério Akaike (cuja abreviatura é AIC, do inglês Akaike Information Criterion) [23]. Esse critério é baseado na identificação do menor valor para AIC, o qual teoricamente resulta no melhor modelo para ajusta-se aos dados observados [10].

Para aplicação dos dados citados para previsão no modelo de previsão ARIMAX foram seguidas as etapas: (I) fase de ajuste – consiste em introduzir os dados observados (velocidade de vento – variável dependente, pressão, temperatura e precipitação – variáveis exógenas) nos modelos de previsão para seja possível fornecer os respectivos ajustes, utilizando o período de janeiro de 2003 até dezembro de 2014; (II) fase da qualidade dos ajustes – essa averiguação é realizada mediante a utilização de métodos de estatística de erros (os quais serão apresentados adiante); (III) fase de previsão – através dos melhores modelos ajustados, os quais foram validados pela estatística de erros foi possível realizar as previsões para o ano seguinte aos dados observados, ou seja, relativo aos meses de 2015. (IV) todas as etapas mencionadas foram realizadas de duas formas, com e sem a inclusão da precipitação local com o propósito de verificar sua influência na previsão da velocidade do vento.

⁷O significado é oriundo da língua inglesa - *Mean Absolute Error*.

Na literatura é possível encontrar alguns trabalhos sobre previsões de séries temporais em diversas áreas com o uso do modelo ARIMAX. Yang *et al.* em [24] propuseram a utilização desse modelo para fornecer previsões horárias de carga para um dia à frente de sistemas de energia em Taiwan. Para investigar o desempenho da abordagem ARIMAX, os autores utilizaram diferentes tipos de dados de cargas horárias relacionadas as estações do ano e também usando variáveis meteorológicas como, temperatura, umidade relativa, velocidade do vento e precipitação. Li *et al.* em [25] utilizam o modelo ARMAX, ou seja, modelo derivado do ARIMAX para séries temporais estacionárias, a fim de fornecer previsões da potência elétrica de saída de grade sistema fotovoltaico ligado à rede. O modelo sugerido leva em consideração a entrada das variáveis exógenas temperatura, precipitação, duração de insolação e umidade. Os resultados das previsões mostram que o modelo ARIMAX melhora consideravelmente a precisão das previsões comparando-se com um modelo univariado ARIMA, ou seja, o qual leva em consideração apenas uma série temporal de entrada. Wangdi *et al.* em [26] buscam entender a incidência da doença malária através do modelo ARIMAX. De acordo com esse trabalho a malária continua a ser um problema de saúde pública em alguns distritos do país Butão, apesar da marcada redução dos casos nos últimos anos. Para fortalecer as medidas de prevenção e controle do país, este estudo foi desenvolvido para realizar previsões da incidência de malária nos distritos endêmicos do Butão.

C. Estatística de erros

Para que seja possível identificar a acurácia do modelo de previsão proposto é necessário averiguar os modelos de cálculo de erros. O Erro Médio Absoluto, mede o valor médio de erro entre as séries observadas e ajustadas. Nessa análise essa variável será representada por MAE⁷, cuja representação matemática é dada por,

$$MAE = \frac{1}{n} \sum_{i=1}^n |v_{ajustada} - v_{observado}|, \quad (2)$$

em que $v_{ajustada}$ representa valor individual da série temporal da velocidade do vento prevista e $v_{observado}$ representa valor individual da série temporal da velocidade do vento observada, e n é a ordem das séries. O MAE possui a unidade de medida neste estudo de metros por segundo (m/s). A Raiz do Erro Quadrático Médio (simbolizado por RMSE⁸), representa as diferenças individuais quadráticas entre as séries temporais da velocidade do vento relativas as observações e ajustes.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_{ajustada} - v_{observado})^2} \quad (3)$$

Da mesma forma o RMSE na Equação 3 possui unidade de m/s. É importante ressaltar que o RMSE pode ainda ser interpretado da seguinte maneira: se houver grandes valores desta variável, os mesmos representam grandes

⁸ Essa abreviatura do inglês, *Root Mean Squared Error*.

erros nas variáveis ajustadas, e valores próximos de zero indicam um ajuste quase perfeito. Maiores detalhes sobre RMSE podem ser obtidos em Cochran [27]. Outra forma de medida de erro é a Média do Erro Absoluto Percentual (simbolizado por MAPE⁹). A grande vantagem de utilizar essa expressão está na sua representação em termos percentuais (%) que fornece um rápido entendimento. Já uma desvantagem que deve ser considerada está no sentido de que se, por acaso, o valor do observado for muito pequeno, qualquer discrepância faz o MAPE “explodir”. A expressão utilizada com essa variável é representada por,

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{v_{ajustada} - v_{observado}}{v_{observado}} \right| \times 100. \quad (4)$$

Maiores detalhes sobre estatística de erros, em especial as Equações 2, 3, e 4 [27], [28].

III. RESULTADOS

Para a previsão da velocidade do vento em nossas regiões de estudo, a partir do modelo ARIMAX, conforme já comentado anteriormente, ressalta-se que a escolha pelas variáveis exógenas temperatura e pressão foram motivadas pelas possíveis relações físicas existentes de ambas com a velocidade do vento. E para a escolha da precipitação a motivação ocorreu para identificar, qual a influência desta variável aplicada ao modelo na previsão da velocidade do vento, uma vez que existem diversos trabalhos na literatura que mostram diferenças na intensidade dos ventos sobre a região nordeste do Brasil nos distintos períodos chuvoso e seco [29], [30], [31].

A Tabela 2, mostra os resultados da estatística de erros para identificar acurácia do modelo na comparação entre as séries temporais ajustadas e observadas, sendo que *Erro(1)* é referente a utilização do ARIMAX com as variáveis exógenas pressão e temperatura, e *Erro(2)* trata-se da utilização do ARIMAX com as variáveis exógenas pressão, temperatura e precipitação. Todos os cálculos foram realizados pelo pacote do software R (*forecast*), a partir da função *accuracy* (observado, ajustado). Os resultados mostram que a partir da inclusão da precipitação local os erros são menores, por exemplo, respectivamente em Fortaleza, Natal e Parnaíba, os valores do MAPE(2) foram 8,48%, 8,47% e 10,40%, enquanto para o MAPE(1) são 8,91%, 9,06%, e 10,64%. Esses resultados podem fornecer indicativo de que para minimizar os erros da previsão da velocidade do vento, a precipitação local deverá ser incluída no modelo. Ressaltando que quanto mais reduzido for os valores da estatística de erros, a expectativa é de que exista cada vez mais semelhanças entre as séries temporais ajustadas e observadas, mostrando dessa forma precisão na estimativa das séries temporais previstas.

Tabela 2 – Estatística de erros para identificar acurácia do modelo ARIMAX. Erro(1) representa os erros sem precipitação, e Erro(2) com precipitação.

<i>Erro(1) / Local</i>	<i>Fortaleza</i>	<i>Natal</i>	<i>Parnaíba</i>
MAE(1)	0,40	0,39	0,55
RMSE(1)	0,52	0,50	0,72
MAPE(1)	8,91	9,06	10,64
<i>Erro(2) / Local</i>	<i>Fortaleza</i>	<i>Natal</i>	<i>Parnaíba</i>
MAE(2)	0,37	0,37	0,54
RMSE(2)	0,48	0,45	0,71
MAPE(2)	8,48	8,47	10,40

A partir da análise de alguns trabalhos da literatura que são semelhantes ao exposto nesse artigo sobre previsão de média mensal da velocidade do vento, é possível afirmar que os erros encontrados na Tabela 2 são similares como, por exemplo, Hu *et al.* em [32] avaliam a previsão para três fazendas eólicas localizadas no nordeste da China. Os resultados obtidos revelam valor de MAPE da ordem de aproximadamente 13,60% para o período entre janeiro de 2001 até dezembro de 2006. Cadenas e Rivera em [33] comparam a performance dos modelos SARIMA e RNA na tentativa de realizar previsões de médias mensais da velocidade do vento em região costeira do México. De acordo com os autores o modelo SARIMA apresentou maior sensibilidade ao ajuste e previsão da velocidade do vento para este caso em particular, o valor de MAPE encontrado foi de aproximadamente 13,40%. No entanto, é provável que com o aumento do número de vetores de treinamento para o modelo RNA, seu desempenho melhore o ajuste fornecendo erros estatísticos menores do que é apresentado para o MAPE da ordem de 20,70%. Ressalta-se que em termos de previsão da velocidade do vento com o uso específico do modelo ARIMAX que leve em consideração as características do presente trabalho, até o presente momento não pôde ser identificado trabalhos na literatura para que pudesse ser comparado com o mesmo.

Para que de fato o modelo ARIMAX seja viável no tocante aos ajustes sobre os dados observados, o termo de erro ϵ_t (também chamado de resíduo), deve se comportar como ruído branco, ou seja, atendendo os pressupostos de distribuição normal (através do Teste Shapiro-Wilk), independência (através do Teste de Durbin-Watson), e por fim o teste que verifica a igualdade da variância, ou seja, hipótese de homoscedasticidade (através do Teste de Breusch-Pagan).

⁹ Do inglês *Mean Absolute Percentage Error*.

Conforme a Tabela 3, tem-se a representação resumida dos testes, Teste(1) sem e Teste(2) com, a inclusão da precipitação. Os valores mostrados são relativos ao (p-valor) de cada teste em cada localidade, sendo que esses testes foram aplicados com nível de significância de $\alpha = 0,05$. As análises dos resíduos para todas as localidades, mostram que todos os pressupostos são atendidos nos dois casos, ou seja, possuem homoscedasticidade; independência; e distribuição normal, com todos os (p-valor) acima do nível de significância de 0,05. Vale destacar que esses resultados são indicativos de aderência do modelo aos dados observados, ou seja, que de fato existe viabilidade para a utilização da modelagem ARIMAX em nossas regiões de estudo para previsão da velocidade do vento. No software R, houve a necessidade de instalação do pacote (car), para viabilizar o uso dos testes citados sobre os resíduos a partir das seguintes funções resumidamente: shapiro.test(resíduos), dwtest(resíduos) e bptest(resíduos), para respectivamente Shapiro-Wilk, Durbin-Watson, e Breusch-Pagan.

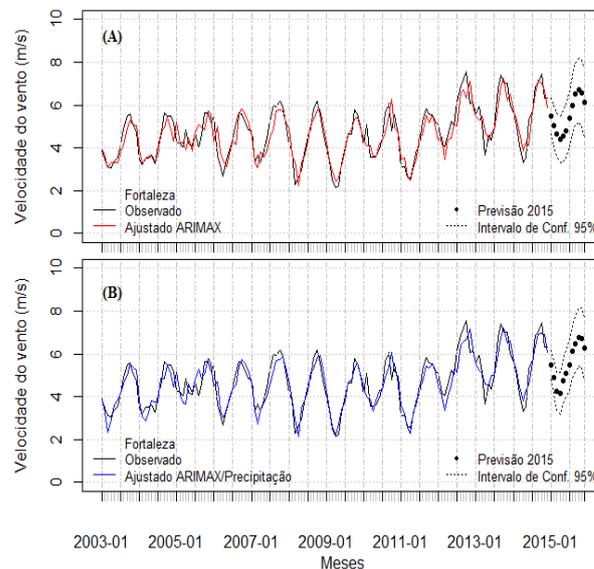
Tabela 3 - P-valores dos testes aplicados sobre os resíduos para identificar se os pressupostos do modelo ARIMAX são atendidos.

Local / Teste(1)	Shapiro Wilk(1)	Durbin Watson(1)	Breusch Pagan(1)
Fortaleza	0,119	0,985	0,375
Natal	0,446	0,460	0,590
Parnaíba	0,140	0,504	0,684
Local / Teste(2)	Shapiro Wilk(2)	Durbin Watson(2)	Breusch Pagan(2)
Fortaleza	0,371	0,622	0,515
Natal	0,531	0,811	0,512
Parnaíba	0,252	0,455	0,618

A Figura 1 é relativa a localidade de Fortaleza, a qual representa as médias mensais das séries temporais: Figura 1A (variáveis exógenas pressão e temperatura); Figura 1B (variáveis exógenas pressão, temperatura e precipitação); observada (linha contínua cor preta); ajustada ARIMAX (linha contínua cor vermelha); ajustada ARIMAX/Precipitação (linha contínua cor azul); e prevista (círculos sólidos na cor preta) para o ano seguinte das observações, ou seja, para o ano de 2015. A finalidade dessa representação se faz necessária para que seja possível identificar a existência de semelhanças entre as séries. No tocante a comparação entre as séries temporais ajustadas (ARIMAX e ARIMAX/Precipitação) com a observação, é possível identificar a existência de semelhanças de máximos e mínimos da velocidade do vento na maioria dos meses nos dois casos, e conforme foi mostrado na Tabela 2, em que os erros Erro(1) e Erro(2) são muito similares, por esse motivo é possível esperar que nas Figuras 1A e 1B, quase não se identifica diferenças como de fato se percebe. Porém os valores

mostram a existência de diferenças, por exemplo, em junho de 2007 os valores ajustados sem e com precipitação foram de 3,88 m/s e 4,08 m/s respectivamente, enquanto para o valor observado no mesmo período foi de 4,44 m/s.

Figura 1 - Comparação entre médias mensais da velocidade do vento em Fortaleza.



Fonte: Acervo do autor.

Ainda sobre a Tabela 2, os ajustes aos dados observados somente ocorrem devido os baixos valores de erros encontrados, esse resultado certamente teve importância na precisão da previsão da velocidade do vento para os meses do ano de 2015, o que de fato é identificado, ou seja, a série prevista consegue acompanhar o perfil da série observada, com ou sem a inclusão da precipitação. É importante comentar que para as previsões (ARIMAX e ARIMAX/Precipitação), em ambas ocorrem aumento no intervalo de confiança (representado na linha seccionada cor preta em 95%) em função dos meses, sendo que isso poderá ser um indicativo de que os modelos propostos possuem dificuldades em prever valores distantes da série observada. Porém esse aumento na Figura 1B é menor na situação de considerar a inclusão da precipitação, ou seja, um resultado que poderá indicar que a partir da inclusão desta variável ocorre maior precisão na velocidade do vento.

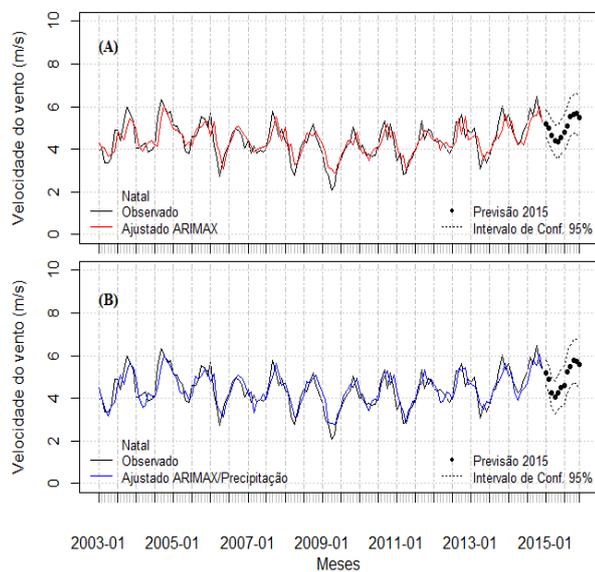
De acordo com estudo realizado a respeito da normal climatológica em todo território brasileiro pelo Instituto Nacional de Meteorologia (INMET¹⁰), para a cidade de Fortaleza os meses de maior precipitação ou também comumente chamados de períodos chuvosos estão entre fevereiro e maio, com pico no mês de abril. Nesse período segundo Camelo *et al.* [34], a média mensal da intensidade da velocidade do vento em Fortaleza

¹⁰Consultar na página eletrônica: <<http://www.inmet.gov.br/portal/index.php?r=clima/normaisclimatologicas>>. Consultado em 11/05/2016.

é menor quando comparado ao período de baixa precipitação ou também chamado de período seco entre os meses de outubro até dezembro. Ainda nas Figuras 1A e 1B, em Fortaleza é possível identificar que o modelo ARIMAX consegue representar características sazonais da velocidade do vento dessa localidade, ou seja, menores intensidades no período chuvoso em especial no mês de abril, e também maiores intensidades para o período seco como é o caso do mês de outubro. Esse resultado poderá ser indicativo de que a velocidade do vento sofre influência da precipitação local.

A Figura 2 mostra as séries temporais para Natal, as análises são semelhantes obtidas para Fortaleza, ou seja, na maioria dos meses as séries temporais ajustadas (linha cor vermelha ARIMAX, e linha azul ARIMAX/Precipitação) possui semelhanças com a série temporal observada (linha cor preta), principalmente no tocante a coincidências de máximos e mínimos da velocidade do vento. Para as séries temporais previstas em 2015 (círculos sólidos na cor preta) nos dois casos, conseguem acompanhar o perfil dos dados observados de Natal. Em especial no que se refere aos valores de mínimo e máximo, os quais são coincidentes na ocorrência em termos de mesmo período, ou seja, acontecendo nos meses de abril e outubro respectivamente. Esse resultado trata-se de um fator positivo do modelo, isso referente ao fato de que o modelo consegue distinguir períodos de baixa e alta intensidade da velocidade do vento.

Figura 2 – Comparação entre médias mensais da velocidade do vento em Natal.



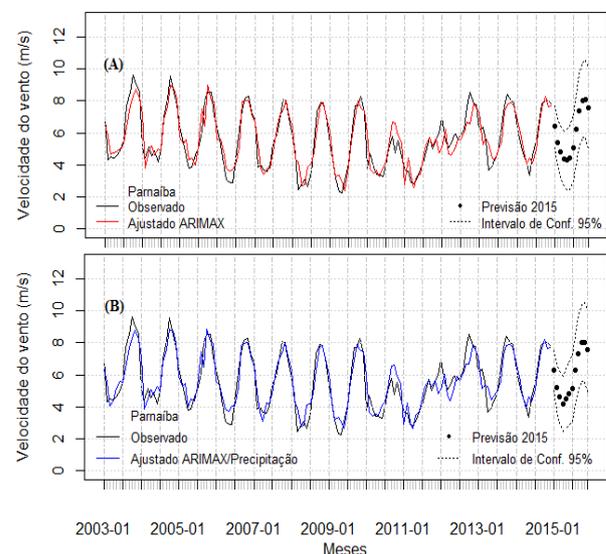
Fonte: Acervo do autor.

Ainda de acordo com INMET para a cidade de Natal, os meses entre março e julho são considerados os de maiores precipitações, sendo em abril o mais elevando, assim, é possível sugerir que nesses meses estão associados com menores valores da velocidade do vento. E os valores mais intensos da velocidade do vento poderão ser associados aos períodos de baixa precipitação como

é o caso dos meses de setembro até dezembro. Portanto os resultados fornecidos pelos modelos (ARIMAX) na Figura 2A e (ARIMAX/Precipitação) na Figura 2B, na representação da velocidade do vento em Natal estão em conformidade com a normal climatológica dessa região.

A Figura 3 representa as séries temporais para Parnaíba, e conforme já comentado na estatística de erros na Tabela 2 que mostrou baixos valores, em função disso ocorre o esperado, ou seja, os ajustes (linha cor vermelha ARIMAX, e linha cor azul ARIMAX/Precipitação) aos dados observados (linha cor preta) é satisfatório no tocante as similaridades entre as séries. Os ajustes fornecidos pelos modelos conseguem representar com bastante eficácia os valores de mínimo e máximo da velocidade do vento de Parnaíba nos dois casos. De acordo com o INMET, para Parnaíba os meses que apresentam maiores valores de precipitações estão entre janeiro e maio, sendo o mês de abril com o pico da precipitação. E os meses de menores valores de precipitações estão entre agosto e novembro, sendo que em outubro apresenta o mínimo. Para as séries temporais previstas em 2015 (círculo sólido na cor preta) nos dois casos, é possível identificar que o menor valor da velocidade de vento ocorre em abril em aproximadamente 4,0 m/s, e o maior valor ocorre em outubro em aproximadamente 8,0 m/s, estes valores estão associados com a máxima e mínima precipitação respectivamente, ou seja, revelando uma possível prova de que o modelo proposto consegue captar séries temporais com existência de sazonalidades.

Figura 3 – Comparação entre médias mensais da velocidade do vento em Parnaíba.



Fonte: Acervo do autor.

Pelo exposto acerca dos resultados da previsão da velocidade do vento em nossas regiões de estudo, os modelos (ARIMAX e ARIMAX/Precipitação), conseguem apresentar de forma satisfatória ajustes aos dados observados, ou seja, mostrando muitas equivalências entre as duas séries, sendo que as previsões também seguem esse mesmo perfil. Ressalta-se que a inclusão da variável

exógena precipitação consegue reduzir ainda mais os erros estatísticos, fornecendo ajustes mais próximos aos dados observados em todas as localidades. Esse trabalho poderá auxiliar como uma possível ferramenta a ser utilizada em diversas áreas, por exemplo, no setor eólico, configurando-se em uma tentativa a mais de garantias aos tomadores de decisão do setor, tendo em vista, a possibilidade de realizar previsões da intensidade da velocidade do vento local, ou seja, conhecer o regime de ventos no futuro. É importante enfatizar que muito embora os resultados para as previsões da velocidade do vento nesse trabalho foram apresentados ao nível de 10 m de altura, e mencionando o fato de que na atualidade os geradores eólicos no Brasil estão sendo implantados acima de 100 m de altura, conforme ABEEólica¹¹, mas a metodologia apresentada para as previsões não possui dependência de alturas, ou seja, pode ser implantada para qualquer nível, para isso basta apenas entrar com as séries temporais observadas nos modelos ARIMAX para fornecer as possíveis projeções.

IV. CONCLUSÕES

O uso dos modelos (ARIMAX com variáveis exógenas pressão e temperatura) e (ARIMAX com variáveis exógenas pressão, temperatura e precipitação), propostos nesse trabalho e executados no *software* livre R, para realizar previsões de médias mensais da velocidade do vento em regiões do nordeste brasileiro, foram eficientes sendo essa afirmação baseada em baixos valores encontrados para a estatística de erros, e a inclusão da variável exógena precipitação consegue reduzir ainda mais esses erros, fornecendo assim ajustes mais próximos aos dados observados. Além disso, foi possível identificar que as séries temporais previstas por estes modelos conseguem acompanhar o perfil das séries temporais observadas das regiões de estudo, através de semelhanças de valores mínimos e máximos da velocidade do vento, esse resultado indica que os modelos propostos conseguem captar existências de sazonalidades das séries temporais. Esse trabalho poderá ter utilidade como uma possível ferramenta em diversas áreas como, por exemplo, no setor eólico, configurando-se em uma tentativa a mais de garantias aos tomadores de decisão do setor, tendo em vista, a possibilidade de realizar previsões da intensidade da velocidade do vento da região de estudo em questão, ou seja, conhecer o regime de ventos no futuro, e a partir disso ser possível atender a demanda de energia elétrica local proveniente de parques eólicos. E para trabalhos futuros poderá ser testado outros modelos de séries temporais em nossas regiões de estudo com a finalidade de tentar encontrar cada vez mais modelos que possam fornecer precisão nas previsões da velocidade do vento.

AGRADECIMENTOS

Os autores gostariam de agradecer à Agência Nacional de Petróleo (ANP) / PRH-51 pela suporte financeiro.

REFERÊNCIAS

- [1] MCGLADE, C.; EKINS, P. *The geographical distribution of fossil fuels unused when limiting global warming to 2 °C*. Nature, v. 517, n. 7533, p. 187-190, 2015.
- [2] PANWAR, N. L., KAUSHIK, S. C., KOTHARI, S. *Role of renewable energy sources in environmental protection: A review*. Renewable & Sustainable Energy Reviews, v. 15, p. 1513–1524, 2011.
- [3] EPE – Empresa de Pesquisa Energética. 2014. Disponível em: <<http://www.epe.gov.br>>. Consultado em 12/05/2015.
- [4] KAVASSERI, R., SEETHARAMN, K. *Day-ahead wind speed forecasting using f-ARIMA models*. Renewable Energy, v. 34, p., 2009.
- [5] BOX, G. P. E., JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. Edição revista. São Francisco: Holden Day, p. 645, 1978.
- [6] CADENAS, E., RIVERA, W. *Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks*. Renewable Energy, v. 34, n. 1, p. 274–278, 2009.
- [7] CADENAS, E., JARAMILLO, O. A. RIVERA, W. *Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method*. Renewable Energy, v. 35, n. 5, p. 925–930, 2010.
- [8] LIU, H.; TIAN, H.; LI, Y. *Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction*. Applied. Energy, v. 98, p. 415-424, 2012.
- [9] R-PROJECT. (2016). The R Project for Statistical Computing. Disponível em: <<http://www.r-project.org/>>. Acesso em junho de 2016.
- [10] BROCKWELL, P. J., DAVIS, R. A. **Time Series: Theory and Methods**. 2ª ed. New York: Springer Verlag, p. 518, 1991.
- [11] PINDYCK, R. S.; RUBINFELD, D. L. **Econometria: modelos & previsões**. Rio de Janeiro: Elsevier, p. 726, 2004.
- [12] MAKRIDAKIS, S. G., WHEELWRIGHT, S. C., HYNDMAN, R. J. **Forecasting: methods and applications**. 3ª ed. New York: Wiley, 656 p., 1998.

¹¹<http://www.portalabeeolica.org.br/noticias/5283-torres-mais-altas-elevam-em-6-vezes-potencial-elico.html>. Consultado em 11/01/2017.

- [13] BENNETT, C.; STEWART, R. A.; LU, J. *Auto-regressive with exogenous variables and neural network short-term load forecast models for residential low voltage distribution networks*. *Energies*, v. 7, n. 5, p. 2938-2960, 2014.
- [14] YOGARAJAH, B.; ELANKUMARAN, C.; VIGNESWARAN, R. *Application of ARIMAX model for forecasting paddy production in Trincomalee district in Sri Lanka*. Proceedings of the Third International Symposium, SEUSL: 6-7 July 2013, Oluvil, Sri Lanka.
- [15] SHUMWAY, R. H.; STOFFER, D. S. **Time series analysis and its applications with R examples**. New York: Springer, 609 p., 2006.
- [16] ROYSTON, J. B. *Some techniques for assessing multivariate based on the Shapiro-Wilk W*. *Journal of Applied Statistics*, v. 32, n. 2, p. 121-133, 1983.
- [17] MONTGOMERY, D. C., PECK, E. A., VINING, G. G. **Introduction to Linear Regression Analysis**. 3ª ed. New York: John Wiley & Sons, 470 p., 2001.
- [18] GUJARATI, D. N., PORTER, D. C. **Basic Econometrics**. 4ª ed. New York: McGraw-Hill, 922 p., 2009.
- [19] BURTON, T., SHARPE, D., JENKINS, N., & BOSSANVI, E. **Wind energy handbook**. Chichester: John Wiley & Sons, 642 p., 2001.
- [20] AHRENS, C. D. **Meteorology today: an introduction to weather, climate, and the environment**. 9ª ed. Boston: Cengage Learning, 621 p., 2012.
- [21] TSONIS, ANASTASIOS A. **An introduction to atmospheric thermodynamics**. Cambridge University Press, 199 p., 2002.
- [22] HYNDMAN, R.J.; KHANDAKAR, Y. *Automatic Time Series Forecasting: The forecast Package for R*. *Journal of Statistical Software*, v. 27, n. 3, p. 1-22, 2008.
- [23] AKAIKE, H. *Information theory and an extension of the maximum likelihood principle*. Proc. 2ª Inter. Symposium on Information Theory, p. 267-281, Budapest, 1973.
- [24] YANG, H.; HUANG, CHAO, M. *A new short-term load forecasting approach using self-organizing fuzzy ARMAX models*. *IEEE Transactions on Power Systems*, v. 13, n. 1, p. 217-225, 1998.
- [25] LI, Y.; SU, Y.; SHU, L. *An ARMAX model for forecasting the power output of a grid connected photovoltaic system*. *Renewable Energy*, v. 66, p. 78-89, 2014.
- [26] WANGDI, K., SINGHASIVANON, P., SILAWAN, T., LAWPOOLSRI, S., WHITE, N. J., KAEWKUNGWAL, J. *Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan*. *Malaria Journal*, v. 9, n. 1, p. 1, 2010.
- [27] COCHRAN, W. G. **Sampling techniques**. 3ª ed. New York: John Wiley & Sons, 428 p., 1977.
- [28] MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. **Introduction to Time Series Analysis and Forecasting**. 6ª ed. New York: Wiley-Interscience, 469 p., 2008.
- [29] SEINFRA – Secretaria de Infraestrutura do Estado do Ceará. 2001. Atlas do Potencial Eólico do Estado do Ceará. Fortaleza – CE, Brasil. Disponível em <<http://www.seinfra.ce.gov.br/index.php/downloads/category/6-energia>>. Consultado em 12/01/2015.
- [30] AMARANTE, O. A.; BROWER, M.; ZACK, J.; EOLICA, C. S. E.; SOLUTIONS, T. *Atlas do potencial eólico brasileiro*. In: Atlas do potencial eólico brasileiro. Ministério de Minas e Energia Eletronbras, 2001.
- [31] AMARANTE, O. D., DA SILVA, F. J. L., RIOS FILHO, L. G., & VIDAL, M. (2003). Atlas do potencial eólico do estado do Rio Grande do Norte. Camargo Schubert Engenharia Eólica, CO-SERN/Iberdrola Empreendimentos do Brasil, 47. Disponível em: <http://www.cresesb.cepel.br/publicacoes/download/atlas_eolico/atlas_eolico_RN.pdf>.
- [32] HU, J.; WANG, J.; ZENG, G. *A hybrid forecasting approach applied to wind speed time series*. *Renewable Energy*, v. 60, p. 185-194, 2013.
- [33] CADENAS, E.; RIVERA, W. *Wind speed forecasting in the south coast of Oaxaca, Mexico*. *Renewable energy*, v. 32, n. 12, p. 2116-2128, 2007.
- [34] CAMELO, H. N., CARVALHO, P. C. M., LEAL JUNIOR, J. B. V., ACCIOLY FILHO, J. B. P. *Análise estatística da velocidade de vento do estado do Ceará*. *Revista Tecnologia*, v. 29, n. 2, p. 211-223, 2008.