

## MODELAGEM BOX – JENKINS APLICADA A PREVISÃO DE VELOCIDADE DO VENTO EM REGIÕES DO NORDESTE BRASILEIRO PARA FINS DE GERAÇÃO EÓLICA

Modeling Box - Jenkins applied a prediction of wind speed in the regions of the Brazilian northeast for the fins of wind generation

<sup>1</sup>Henrique do Nascimento Camelo, <sup>1</sup>Paulo Sérgio Lucio, <sup>2</sup>João Bosco Verçosa Leal Junior e <sup>2</sup>Paulo Cesar Marques de Carvalho

<sup>1</sup>Universidade Federal do Rio Grande do Norte, Brasil

<sup>2</sup>Universidade Estadual do Ceará, Brasil

### Resumo

*No presente trabalho foi realizado um estudo de previsão de média mensal de velocidade do vento em regiões do nordeste brasileiro. Para isso foi utilizado a metodologia de modelagem Box – Jenkins, aplicado aos dados de velocidade do vento a 10 m de altura entre janeiro de 2010 até dezembro de 2013. A previsão em todas as localidades de estudo foi realizada para o ano de 2014, através do modelo SARIMA, o qual predominou em praticamente todas as regiões, ou seja, um indicativo de que o modelo de previsão ideal necessariamente deve ser introduzido a componente sazonal. A previsão mostrou-se eficiente em algumas regiões como, por exemplo, em Aracaju foi possível encontrar um erro de MAPE no valor de 4,66%. Nas localidades de Aracaju e Salvador é possível identificar que a série prevista tende a possuir comportamento semelhante a série observada no tocante a semelhança de máximos e mínimos de velocidade de vento. Esse trabalho poderá ser utilizado como uma ferramenta de previsão de velocidade do vento, para estudo e avanço da geração eólica em diversas regiões, proporcionando aos tomadores de decisão a exploração eólica local, uma vez que será possível estimar o regime de vento no futuro.*

**Palavras-chaves:** Séries temporais. ARIMA. Software R

### Abstract

*In the present work, a study was carried out to predict monthly average wind speed in regions of the Brazilian Northeast. For this purpose, the Box - Jenkins modeling methodology was applied to the 10 m high wind speed data from January 2010 to December 2013. The forecast in all study locations was for the year 2014 through of the SARIMA model, which predominated in practically all regions, that is, an indication that the ideal forecast model must necessarily be introduced to the seasonal component. The prediction was efficient in some regions, for example, in Aracaju it was possible to find a MAPE error of 4.66%. In the localities of Aracaju and Salvador it is possible to identify that the predicted series tend to have similar behavior to the observed series regarding the similarity of maximum and minimum wind speed. This work could be used as a wind speed prediction tool to study and advance wind generation in several regions, providing decision makers with local wind exploitation, since it will be possible to estimate the wind regime in the future.*

**Keywords:** Time series. ARIMA. Software R

## 1 Introdução

O desafio energético mundial é conseguir satisfazer a crescente demanda de energia, proporcionar o uso racional das fontes energéticas atuais e a busca por novas formas de geração de eletricidade que produzam baixa ou nenhuma emissão de gases geradores do efeito estufa. Assim, as fontes de energias renováveis como a energia solar e eólica, são apresentadas como alternativas à substituição das fontes convencionais de produção de eletricidade como, as que utilizam combustão. O caráter renovável atribuído a estas fontes de energia se deve ao argumento de que estes recursos estão presentes na natureza de forma abundante, o que caracteriza uma fonte praticamente inesgotável.

O Brasil pode ser considerado como um dos que mais possui capacidade de produzir eletricidade a partir de fontes renováveis de energia. Exemplo disso é que em sua matriz energética a principal fonte é proveniente de hidráulica com aproximadamente 65%, de acordo com o Balanço Energético Brasileiro de 2016 (AQUILA *et al.* 2017). Em termos de potencial eólico, se estima que o Brasil possa gerar cerca de 143 mil MW de potência eólica, sendo que 51% está sobre a região Nordeste do país, e em grande parte do litoral do nordeste a velocidade do vento, a 50 m de altura, é da ordem de 8,5 m/s (AQUILA *et al.* 2017), o que tem feito com que essa região receba bastante atenção de empresas do setor eólico na atualidade. Esta média anual de velocidade do vento é bastante significativa quando comparada com as médias encontradas no Atlas Europeu do Vento, em especial na Alemanha, onde a velocidade média dos ventos é da ordem de 5 a 6 m/s no mesmo nível de altura (TROEN e PETERSEN, 1989).

Nos últimos anos a previsão de séries temporais de velocidade média do vento para fins de geração de energia, vem sendo muito estudada por diversos pesquisadores pelo mundo afora. Vários métodos têm sido utilizados para poder chegar ao menor erro possível entre dados observados (medição *in loco*) e dados previstos. Nesse sentido é possível destacar os seguintes métodos como: Utilização de Redes Neurais Artificiais (RNA); utilização de modelagem numérica de mesoescala por meio de programas de modelagem atmosférica, exemplo disso, o modelo *Weather Research and Forecasting* (WRF) e ainda modelagem Box-Jenkins e Suavização Exponencial (SE) ou Holt-Winters (HW).

Um modelo de previsão de séries temporais que deve ser destacado, devido a sua importância e utilidade é o modelo de Box – Jenkins (BOX e JENKINS, 1978). Essa metodologia envolve o modelo Autorregressivo Integrado de Médias Móveis (ARIMA), que por sua vez, dá origem aos modelos, Autorregressivo de Médias Móveis (ARMA) e Sazonal Autorregressivo Integrado de Médias Móveis (SARIMA). Os modelos ARIMA possuem características flexíveis, permitindo ajustes e adaptações em seus parâmetros, contribuindo para que diferentes tipos de séries temporais sejam representados, assumindo assim, uma estrutura de correlação linear entre os valores das séries temporais (MORETTIN e TOLOI, 2006).

Cadenas e Rivera (2010) identificam a viabilidade de utilização da modelagem ARIMA e também RNA para a previsão de séries temporais de velocidade do vento em localidade do México. Utilizam ainda estatística de erros como, por exemplo, o erro médio absoluto (cujo valor encontrado foi de 1,47 m/s) para mostrar a melhor configuração dos modelos mencionados. Vale ressaltar que a utilização de modelagem ARIMA é muito útil quando se trabalha com séries temporais Não-Estacionárias, ou seja, quando há variações bruscas da série em questão. Em Kavasseri e Seetharaman (2009) analisam a velocidade dos ventos de Dakota do Norte nos EUA para examinar a possibilidade de previsão com a modelagem ARIMA no sentido de identificar valores para 24 h e 48 h adiante.

Esse trabalho tem como objetivo utilizar o modelo Auto Regressivo Integrado de Médias Móveis (ARIMA) e modelos derivados deste como, o ARMA e o SARIMA, para realizar previsões de séries temporais de média mensal de velocidade do vento em regiões do nordeste brasileiro. Dessa forma, pretende-se avaliar se os modelos propostos oferecem menores erros em relação aos dados observados de velocidade do vento, identificando assim, uma possível ferramenta que poderá ser amplamente utilizada para o estudo e avanço da potência eólica em todo país.

## 2 Fundamentação Teórica

Nesta seção serão abordadas características das séries temporais de estudo, e também uma breve revisão dos conceitos empregados na metodologia de trabalho como, por exemplo, características da modelagem ARIMA e uma forma de validação deste modelo.

### 2.1. Região de estudo e dados coletados

Os dados de velocidade de vento, utilizados nesse trabalho são constituídos de médias horárias para um período de quatro anos, ou seja, de janeiro de 2010 até dezembro de 2014 e possuem unidade de metro por segundo (m/s). Para esse período foram utilizados dados de localidades de nove estados da região nordeste do Brasil, relativos as torres meteorológicas instaladas em aeroportos, conforme Tabela 1, abaixo. Vale ressaltar que as medidas de velocidade de vento são realizadas a uma altura de 10 m em relação ao solo.

Tabela 1 – Localização dos dados de velocidade de vento

Localidade	Latitude	Longitude
Fortaleza	-3,77°	-38,52°
Natal	-5,92°	-35,25°
Parnaíba	-2,92°	-41,75°
São Luís	-2,6°	-44,22°
Recife	-8,13°	-34,92°
Aracaju	-10,97°	-37,07°
João Pessoa	-7,12°	-34,95°
Salvador	-12,9°	-38,32°
Maceió	-9,52°	-35,77°

Para a realização da previsão com os dados citados em cada localidade foram utilizadas três etapas: (1) aplicação dos modelos ARIMA aos dados de velocidade do vento no período de janeiro de 2010 até dezembro de 2013. Essa etapa consiste em fazer com que o modelo proposto possa captar tendência, nível e sazonalidade da série temporal observada; (2) previsão dos dados de velocidade do vento no período de janeiro de 2014 até dezembro de 2014; e (3) comparação da série temporal prevista na etapa (2) com a série temporal observada no mesmo período de previsão.

Os dados de velocidade de vento, citados acima foram doados pelo Projeto Airmetar que foi desenvolvido pela kSoftwares e surgiu da necessidade de ferramentas meteorológicas para consultar e analisar informações dos aeroportos do Brasil através da consulta e decodificação dos dados METAR fornecidos pela aeronáutica do Brasil. Vale ressaltar que esse projeto não possui fins lucrativos e estão disponíveis gratuitamente no endereço eletrônico (<<http://www.airmetar.com.br/>>) consultado em: 02/02/16.

## 2.2 Modelagem Box – Jenkins

A metodologia de Box–Jenkins (BOX e JENKINS, 1978), também conhecida como modelo Auto Regressivo Integrado de Médias Móveis (ARIMA), vem sendo muito utilizada para previsão de séries temporais como, pode ser observado em (CADENAS e RIVERA, 2010; GOMES, 2014; VARGAS e PESSANHA, 2014). O modelo ARIMA possui características flexíveis, permitindo ajustes e adaptações em seus parâmetros, contribuindo para que diferentes tipos de séries temporais sejam representados, assumindo assim, uma estrutura de correlação linear entre os valores das séries temporais (MORETTIN e TOLOI, 2006). Os modelos ARIMA podem descrever duas classes de processos:

1) Os processos lineares estacionários, em que uma série temporal é dita dessa forma quando apresenta ao longo do tempo, média constante ou ainda variando em torno da média (GUJARATI, 2010). Para esses processos é possível utilizar basicamente três tipos de modelos:

Processo autorregressivo de ordem ( $p$ ), representado por  $AR(p)$ ;

Processo de médias móveis de ordem ( $q$ ), representado por  $MA(q)$ ;

Processo autorregressivo e de médias móveis de ordem ( $p$ ) e ( $q$ ), representado por  $ARMA(p,q)$ .

2) 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 (MORETTIN e TOLOI, 2006). Para esses processos o modelo é comumente representado por  $ARIMA(p, d, q)$ .

De acordo com Pindyck e Rubinfeld (2004), o número de vezes que a série original tem que ser diferenciada antes de resultar uma série estacionária é denominada “ordem de homogeneidade”. Estes processos são descritos pelos modelos autorregressivos – 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  $SARIMA(p, d, q)(P, D, Q)$  (MORETTIN e TOLOI, 2006).

Em (FAVA, 2000), explicando de outra maneira, afirma que “os modelos ARIMA resultam da combinação de três componentes também denominados filtros”: o componente autorregressivo (AR), o filtro de integração (I) e o componente de médias móveis (MA). Esses modelos, ainda segundo a autora, permitem prever valores futuros explorando a correlação serial que existe geralmente entre os valores da série temporal. Segundo Pindyck e Rubinfeld (2004), séries não-estacionárias podem ser transformadas em séries estacionárias quando suas observações são diferenciadas uma ou mais vezes. A Equação (1), abaixo descreve a primeira diferenciação dos dados:

$$Z_t = \check{X}_t = Y_t - Y_{t-1} \quad (1)$$

em que:

$Y_t$  é a observação  $Y$ , no período  $t$  da série  $Y_t$  sem diferenciação;

$Y_{t-1}$  é a observação  $Y$ , no período  $t-1$  da série  $Y_t$  sem diferenciação;

$\Delta Y_t = Z_t$  é a observação  $Z$ , no período  $t$ , pertencente a série  $Z_t$  com os dados da série  $Y_t$  diferenciados a primeira vez. De acordo com Morettin e Toloi (2006), a construção de um modelo ARIMA é baseada em um ciclo com as seguintes etapas:

**Identificação:** nesta etapa, encontram-se os filtros ARIMA e a ordem do modelo, representados pelos valores de  $p$ ,  $d$ , e  $q$  do modelo;

**Estimação:** logo após, estimam-se os parâmetros dos termos autorregressivos e de médias móveis que serão incluídos no modelo, por meio do Método de Mínimos Quadrados – MMQ;

**Verificação:** nesta fase, verifica-se se o modelo escolhido ajusta-se adequadamente aos dados, isto é, analisa-se se os resíduos estimados se comportam como um ruído branco, em caso positivo, passa-se para a quarta etapa, e em caso negativo, deve-se recomençar;

**Previsão:** esta é a fase final do processo, em que é estimada a previsão para a série temporal, em geral para um horizonte de curto prazo.

Com relação ao item de **Identificação**, definido a ordem de integração  $d$ , passa-se a trabalhar com a série estacionária  $Y_t$ . Inicia-se, em seguida, a fase de identificação dos filtros AR e MA, ou seja, dos valores das ordens  $p$  e  $q$  do modelo ARIMA. Essa fase será realizada observando-se as características de dois instrumentos: a função de autocorrelação (ACF) e a função de autocorrelação parcial (PACF), cuja a escolha foi realizada conforme a Tabela 2. A (ACF) mede a correlação serial de ordem  $k$ , ou seja, a autocorrelação entre  $Y_t$  e  $Y_{t-k}$ . A primeira correlação (Corr1), então, mede a relação de  $Y_t$  com todos os outros valores de  $Y$  na série, o valor de (Corr2) mede a relação de  $Y_2$  com todos os outros valores de  $Y$  na série, e assim sucessivamente. Já a (PACF) mede a correlação entre  $Y_t$  e  $Y_{t-1}$ , depois que a influência de  $Y_{t-1}$ ,  $Y_{t-2}$ ,  $Y_{t-3}$ , ...,  $Y_{t-k+1}$  sobre  $Y_t$  foi descontada.

Tabela 2 – Padrões de FAC e FACP.

Tipo de Modelo	Padrão Típico de ACF	Padrão Típico de PACF
AR (p)	Declina exponencialmente ou com padrão de onda senoidal, ou ambos.	Picos significativos através das defasagens p.
MA (q)	Picos significativos através das defasagens q.	Declina exponencialmente.
ARMA (p, q)	Declínio exponencial.	Declínio exponencial.

Para a estimação dos parâmetros do modelo no item (2) parágrafo anterior, conforme Fava (2000), determinado os valores de  $p$ , e  $q$ , pode-se estimar os  $p$  parâmetros  $\varphi$  (ou seja, coeficiente relativo ao filtro de médias móveis), os  $q$  parâmetros  $\theta$  (ou seja, coeficiente relativo ao filtro autorregressivo estacionário), e o valor da variância dos erros do modelo  $\varepsilon_t$ , conforme a Equação (2):

$$W_t = \varphi_1 W_{t-1} + \varphi_2 W_{t-2} + \dots + \varphi_p W_{t-p} + \dots \\ \dots + \hat{a}_t - \hat{e}_1 \hat{a}_{t-1} - \hat{e}_2 \hat{a}_{t-2} - \dots - \hat{e}_q \hat{a}_{t-q} \quad (2)$$

Maiores detalhes sobre a Equação (2), podem ser encontrados em (MORETTIN e TOLOI, 2006). A seleção do melhor modelo ARIMA para cada localidade foi realizado através da utilização do critério Akaike (cuja abreviatura é AIC, do inglês *Akaike Information Criterion*), (AKAIKE, 1998). Esse critério é baseado na identificação do menor valor para AIC, o qual teoricamente resultado no melhor modelo para ajusta-se aos dados observados.

### 2.3. Validação do modelo

Para avaliar se os modelos ARIMA encontrados nesse trabalho possuem viabilidade para a previsão de média mensal de velocidade do vento foram utilizados dois procedimentos a seguir:

#### (a) Análise dos Resíduos

Para que o modelo ARIMA representada na Equação (2) seja viável no tocante ao ajuste dos dados observados, o termo de erro  $\varepsilon_t$ , deste modelo deve se comportar como ruído branco, ou seja, possuir média zero, variância constante (são homocedásticos) e ser não correlacionado, ou seja, os erros são independentes (GUJARATI, 2010). E também o termo  $\varepsilon_t$  deve seguir uma distribuição normal.

Para identificação desses pressupostos existem alguns testes na literatura possíveis de serem aplicados sobre os resíduos. Dentre eles destaca-se o teste de verificação de normalidade dos resíduos através do Teste de Shapiro-Wilk, maiores detalhes

em (ROYSTON, 1983), teste de independência dos resíduos através do Teste de Durbin-Watson, maiores detalhes em (DOANE e SEWARD, 2014), e o teste que verifica a igualdade da variância dos resíduos, ou seja, hipótese de homoscedasticidade através do Teste de Breusch-Pagan, maiores detalhes em (GUJARATI, 2010). Conforme a Tabela 3, tem-se a representação resumida dos testes utilizados, sendo que os mesmos foram aplicados ao nível de significância de  $\alpha = 0,05$ .

Tabela 3 – Apresentação dos testes para identificar ruído branco do modelo ARIMA.

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

(b) Estatística de Erros

O Erro Médio Absoluto mede o valor médio de erro entre as séries observadas e previstas. Nessa análise essa variável será representada por MAE (do inglês *Mean Absolute Error*), cuja representação matemática é dada por,

$$MAE = \frac{1}{o} \sum_{i=1}^o \left| v_p - v_r \right| \quad (3)$$

em que  $v_p$  representa valor individual da série prevista e  $v_r$  representa valor individual da série real (ou observada), e  $o$  é a ordem das séries. O MAE possui a unidade de medida neste estudo de metros por segundo, ou m/s.

A Raiz do Erro Quadrático Médio (simbolizado por RMSE, sendo essa abreviatura do inglês *Root Mean Squared Error*) representa as diferenças individuais quadráticas entre as séries temporais observadas e previstas, que pode ser definido matematicamente por,

$$RMSE = \sqrt{\frac{1}{o} \sum_{i=1}^o (v_p - v_r)^2} \quad (4)$$

Da mesma forma o RMSE na Equação (4) 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 erros nas variáveis previstas, e valores próximos de zero indicam uma previsão quase perfeita. Maiores detalhes sobre a Raiz do Erro Quadrático Médio podem ser obtidos em (COCHRAN, 1977).

Outra forma de medida de erro é a Média do Erro Absoluto Percentual (simbolizado por MAPE, do inglês *Mean Absolute Percentage Error*). 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 real for muito pequeno, qualquer discrepância faz o MAPE “explodir”. A expressão utilizada com essa variável é representada por,

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{v_p - v_r}{v_r} \right| \times 100 \quad (5)$$

Maiores detalhes sobre estatística de erros, em especial às Equações (3), (4), e (5) que serão usadas nesse trabalho, podem ser obtidos em (MONTGOMERY *et al.*, 2008). Todos os cálculos realizados nesse estudo e além da parte gráfica, foram produzidos pelo software livre *R*, (R CORE TEAM, 2017). O *R* fornece uma ampla variedade (modelagem linear e não-linear, testes estatísticos clássicos, análise de séries temporais, testes paramétricos e não-paramétricos, suavização, etc.) e técnicas gráficas, e é altamente extensível. Um dos pontos fortes do *R* é a facilidade com que pode ser produzida a edição, com qualidade, de gráficos, incluindo símbolos e fórmulas matemáticas, quando necessário. O *R* está disponível como software livre, nos termos da GNU na forma de código aberto. Pode ser compilado e “roda” em um grande número de plataformas UNIX e sistemas semelhantes (incluindo *FreeBSD* e Linux), *Windows* e *MacOS*.

### 3 Resultados e discussão

A seleção do melhor modelo ARIMA para cada localidade foi resultado do menor valor para o critério Akaike. Conforme a Tabela 4, mostra o modelo selecionado e a análise dos resíduos realizados através dos testes citados. Os resultados mostram que para a maioria das localidades o modelo predominante foi o SARIMA e os pressupostos comentados para os resíduos são atendidos, ou seja, são homocedásticos; independentes; e possuem distribuição normal. Vale ressaltar que esses resultados são indicativos de que existe viabilidade para a utilização do modelo selecionado.

Tabela 4 – Apresentação do modelo para cada localidade e também testes aplicados aos resíduos

Localidade (Modelo)	Shapiro-Wilk	Durbin-Watson	Breusch-Pagan
Aracaju SARIMA(1,0,0)(1,0,0)	p-valor: 0,533. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,422. Decisão: os resíduos são independentes.	p-valor: 0,813. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
Fortaleza SARIMA(1,1,0)(1,0,0)	p-valor: 0,209. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,363. Decisão: os resíduos são independentes.	p-valor: 0,767. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
João Pessoa SARIMA(1,1,0)(1,0,0)	p-valor: 0,008. Decisão: os resíduos não possuem distribuição normal.	p-valor: 0,463. Decisão: os resíduos são independentes.	p-valor: 0,334. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
Maceió ARIMA(1,1,0)	p-valor: 0,769. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,667. Decisão: os resíduos são independentes.	p-valor: 0,052. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
Natal SARIMA(1,0,0)(1,0,0)	p-valor: 0,276. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,408. Decisão: os resíduos são independentes.	p-valor: 0,265. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
Parnaíba ARIMA(2,1,2)	p-valor: 0,084. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,396. Decisão: os resíduos são independentes.	p-valor: 0,310. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
Recife SARIMA(1,0,0)(1,0,0)	p-valor: 0,093. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,252. Decisão: os resíduos são independentes.	p-valor: 0,284. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
Salvador SARIMA(1,0,0)(1,0,0)	p-valor: 0,361. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,343. Decisão: os resíduos são independentes.	p-valor: 0,906. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
São Luís SARIMA(2,0,0)(1,1,0)	p-valor: 0,492. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,731. Decisão: os resíduos são independentes.	p-valor: 0,411. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).

Conforme a Tabela 5, mostra a estatística de erros para identificar a acurácia da modelagem Box – Jenkins em cada localidade de estudo. Como é possível perceber no tocante aos erros de MAE e RMSE seus valores na maioria são inferiores a 1 m/s. Para o erro de MAPE é possível destacar que o menor valor encontrado foi de 4,66% para Aracaju, sendo que esse valor pode ser um indicativo de que nessa região o modelo ARIMA se ajustou melhor comparada as demais regiões. No entanto, o maior valor encontrado foi de aproximadamente 41% em Maceió valor este que pode indicar que o modelo ARIMA não se ajusta aos dados observados nesta localidade.

Tabela 5 – Estatística de erros para identificar acurácia da modelagem Box – Jenkins no período de previsão de janeiro de 2014 até dezembro de 2014

Erros	Fortaleza	Natal	Parnaíba	São Luís
MAE	0,79	0,61	1,49	0,74
RMSE	1,01	0,74	1,78	0,91
MAPE	17,41	10,77	31,68	22,53
Erros	Recife	Aracaju	João Pessoa	Salvador
MAE	0,47	0,22	0,36	0,26
RMSE	0,51	0,29	0,45	0,32
MAPE	11,93	4,66	10,67	6,21

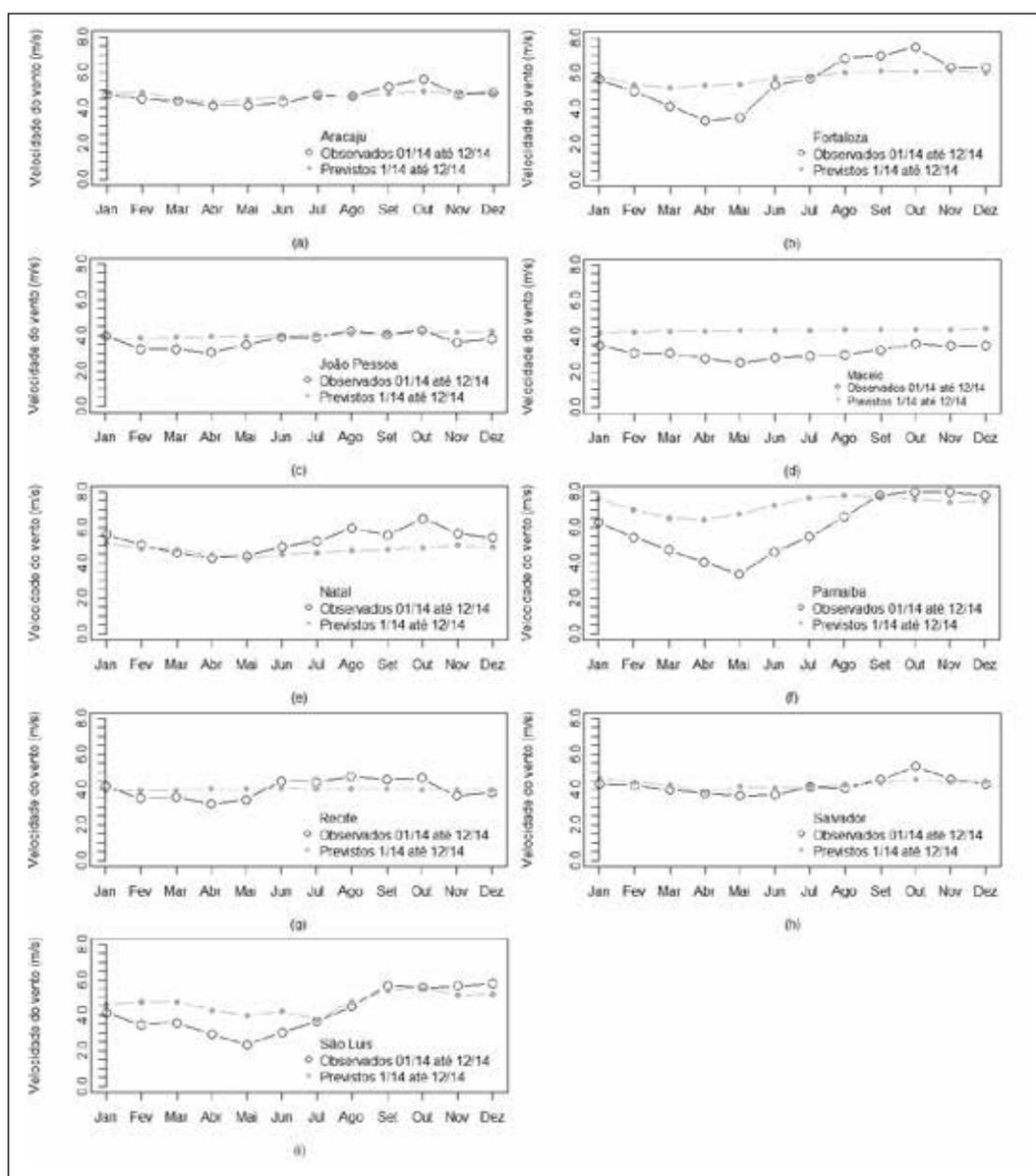


Figura 1 – Comparação entre as séries observadas (cor preta, linha e círculo vazio) e previstas (cor cinza, linha e círculo cheio). (a) Aracaju, (b) Fortaleza, (c) João Pessoa, (d) Macció, (e) Natal, (f) Parnaíba, (g) Recife, (h) Salvador, e (i) São Luís

Conforme a Figura 1, mostra os resultados para a comparação entre a previsão dos modelos ARIMA e a série temporal observada no período de janeiro de 2014 até dezembro de 2014 para cada localidade mencionada. De acordo com a análise de erros conforme a Tabela 5, as regiões onde houve melhores ajustes para todo o ano em questão foram em Aracaju conforme a Figura 1-a, João Pessoa conforme a Figura 1-c, e Salvador conforme a Figura 1-h. Nessas regiões existe uma forte tendência de que as séries temporais previstas (representadas na cor cinza, sendo linha com círculo sólido) possam acompanhar o perfil das séries temporais observadas (representadas na cor preta, sendo linha com círculo vazio). É importante destacar que em algumas regiões o ajuste é parcial, ou seja, para alguns meses do ano como, por exemplo, em Natal conforme a Figura 1-e somente é possível falar que a série temporal prevista consegue acompanhar o perfil da série temporal observada no primeiro semestre de 2014. Da mesma forma, em São Luís, há tendência de que a previsão se ajuste aos dados observados para o segundo semestre de 2014.

Conforme as Figuras 2-a, 2-b, 2-c, 2-d, 2-e, 2-f, 2-g, 2-h, e 2-i que representam respectivamente as localidades de Aracaju, Fortaleza, João Pessoa, Maceió, Natal, Parnaíba, Recife, Salvador, e São Luís, sendo que a série temporal observada em todo período analisado (linha na cor preta), ou seja, janeiro de 2010 até dezembro de 2013, e em seguida a previsão para o período de janeiro de 2014 até dezembro de 2014 (linha cor azul). E ainda a região na cor cinza escuro é referente ao intervalo de confiança de 95%. Nas regiões de Aracaju e Salvador é possível identificar que a série prevista tende a possuir comportamento semelhante a série observada no tocante a semelhança de máximos e mínimos de média mensal de velocidade de vento. Vale destacar que essas localidades, mostraram menores valores para o intervalo de confiança, isso pode ser um indicativo de melhor desempenho na previsão das citadas regiões.

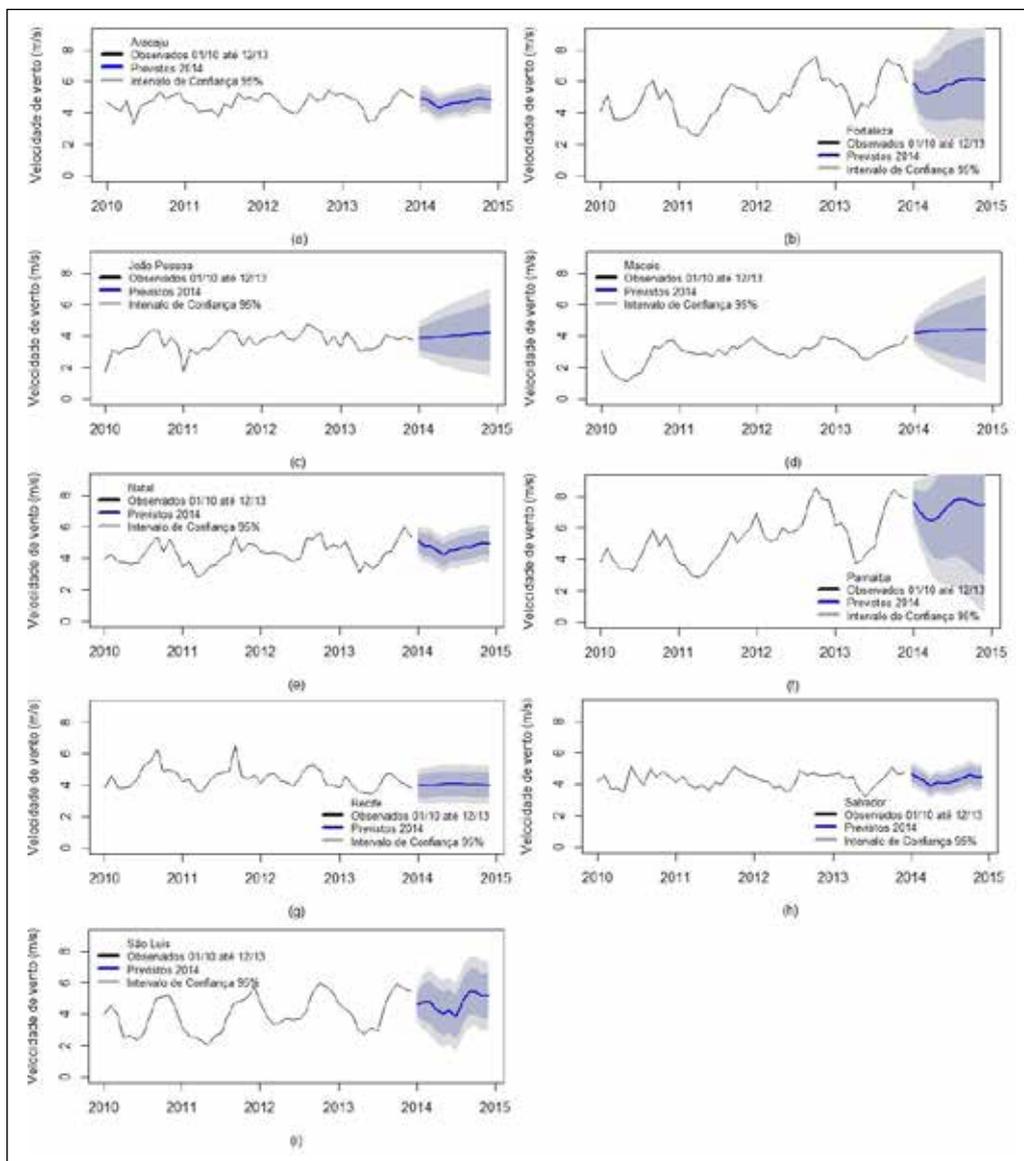


Figura 2 - Série temporal observada em todo período (linha cor preta), e série prevista para o ano de 2014 na (linha cor azul). A região na cor cinza escuro é referente ao intervalo de confiança de 95%. (a) Aracaju, (b) Fortaleza, (c) João Pessoa, (d) Maceió, (e) Natal, (f) Parnaíba, (g) Recife, (h) Salvador, e (i) São Luís

## 4 Conclusões e Perspectivas Futuras

A utilização da metodologia Box – Jenkins para a previsão de média mensal de velocidade do vento em regiões do nordeste brasileiro, mostrou uma predominância para a aplicação do modelo SARIMA. Com esse resultado é possível afirmar que o modelo de previsão ideal nas regiões de estudo, necessariamente deve ser introduzido a componente sazonal do modelo ARIMA, para minimizar os erros quando comparado as séries temporais observadas e previstas. As regiões onde houve melhores ajustes para todo o período de previsão, ou seja, o ano de 2014 foram em Aracaju, João Pessoa, e Salvador. Nessas regiões existe uma forte tendência de que as séries temporais previstas possam acompanhar o perfil das séries temporais observadas. É importante destacar que em algumas regiões o ajuste é parcial, ou seja, para alguns meses como, por exemplo, em Natal somente é possível falar que a série temporal prevista consegue acompanhar o perfil da série temporal observada no primeiro semestre de 2014. Já em São Luís, é o oposto de Natal, ou seja, há tendência de que a previsão se ajuste aos dados observados para o segundo semestre de 2014.

Para trabalhos futuros poderá ser testado, a partir da disponibilidade de dados de velocidade do vento, nessas regiões de estudo, uma previsão com uma maior série temporal e, portanto, identificar se com isso haverá uma redução ainda maior na análise de erros. Também poderá ser testado outros modelos de previsão para comparação como, por exemplo, Regressão, Holt-Winters, RNAs e dentre outros. Esse trabalho poderá ser utilizado como uma ferramenta de previsão de velocidade do vento, visando o estudo e avanço da potência eólica em diversas regiões do país, proporcionando aos tomadores de decisão a instalação de novos parques eólicos, visto que haverá possibilidade de conhecer o regime local de intensidade do vento futuro.

## Referências

- AKAIKE H. Information theory and an extension of the maximum likelihood principle. In: Selected Papers of Hirotugu Akaike. New York: Springer; 1998.
- AQUILA G, DE OLIVEIRA PAMPLONA E, DE QUEIROZ AR, JUNIOR PR, FONSECA MN. An overview of incentive policies for the expansion of renewable energy generation in electricity power systems and the Brazilian experience. *Renew. Sustain. Energy Rev.* 2017;70:1090-1098.
- BOX GPE, JENKINS GM. *Time Series Analysis: Forecasting and Control*. 1st ed. São Francisco: Holden Day; 1978.
- CADENA E, RIVERA W. Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA - ANN model. *Renew. Energy* 2010;35(12):2732-2738.
- COCHRAN WG. *Sampling techniques*. 1 st ed. New York: John Wiley & Sons; 1977.
- DOANE DP, SEWARD LE. *Estatística aplicada à administração e economia*. 4 st ed. São Paulo: McGraw-Hill; 2014.
- FAVA VL. Metodologia de Box-Jenkins para modelos univariados. In: Vasconcellos MAS, Alves D. *Manual de econometria: nível intermediário*. 1st ed. São Paulo: Atlas; 2000.
- GOMES TLO. Utilização do Software R para Previsão de Preço do Dia Seguinte de uma Série Temporal. *UNOPAR Cient. Exatas Tecnologia* 2014;13(1):61-68.
- GUJARATI DN. *Econometria Básica*. 5 st ed. São Paulo: McGraw-Hill; 2010.
- KAVESSERI RG, SEETHARAMAN K. Day-ahead wind speed forecasting using f-ARIMA models. *Renew. Energy* 2009;34:1388-1393.
- MARCHEZAN A, SOUZA AM. Previsão do preço dos principais grãos produzidos no Rio Grande do Sul. *Ciência Rural* 2010;40(11):2368-2374.
- MORETTIN PA, TOLOI CMC. *Análise de Séries Temporais*. 2 st ed. São Paulo: Edgard Blücher; 2006.

MONTGOMERY DC, JENNINGS L, KULAHCI M. Introduction to Time Series Analysis and Forecasting. 6 st ed. New York: Wiley-Interscience; 2008.

PINDYCK RS, RUBINFELD DL. Microeconomia. 5 st ed. São Paulo: Prentice Hall; 2004.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>, 2017.

ROYSTON JB. Some techniques for assessing multivariate based on the Shapiro-Wilk W. Appl. Stat. 1983;32(2):121-133.

TROEN I, PETERSEN EL. European Wind Atlas. 1989 [cited 2017 oct 31]. In: Rise National Laboratory. Roskilde, Denmark. Available from: [http://www.orbit.dtu.dk/files/112135732/European\\_Wind\\_Atlas.pdf](http://www.orbit.dtu.dk/files/112135732/European_Wind_Atlas.pdf).

WERNER L, RIBEIRO JLD. Previsão de demanda: Uma aplicação dos modelos Box-Jenkins na área de assistência técnica de computadores pessoais. Gestão & Produção 2003;10(1):47-67.

VARGAS AS, PESSANHA JF. Previsão probabilística de curto prazo da geração de energia eólica. XLVI Simpósio Brasileiro de Pesquisa Operacional; 2014 setembro; Salvador;BA: 2014.

#### Henrique do Nascimento Camelo

Universidade Federal do Rio Grande do Norte, Brasil

Doutorando em Ciências Climáticas

Email: [henriquecamelo13@yahoo.com.br](mailto:henriquecamelo13@yahoo.com.br)

Participação do autor: confecção do modelo Box-Jenkins dentro do software R, e desta foram sendo possível fornecer previsão de velocidade do vento, e também na parte escrita do artigo

#### Paulo Sérgio Lucio

Universidade Federal do Rio Grande do Norte, Brasil

Departamento de Ciências Atmosféricas e Climáticas

Email: [pslucio@ccet.ufrn.br](mailto:pslucio@ccet.ufrn.br)

Participação do autor: execução da pesquisa repassando informações necessárias para o andamento da pesquisa

#### João Bosco Verçosa Leal Junior

Universidade Estadual do Ceará, Brasil

Doutorado em Física

Email: [jblealjr@uece.br](mailto:jblealjr@uece.br)

Participação do autor: contribuiu na revisão da parte escrita do artigo

#### Paulo Cesar Marques de Carvalho

Universidade Estadual do Ceará, Brasil

Departamento de Engenharia Elétrica

Email: [carvalho@dee.ufc.br](mailto:carvalho@dee.ufc.br)

Participação do autor: contribuiu com sua larga experiência em geração eólica, e desta maneira realizando revisões na parte escrita do artigo as quais abordaram este tema