

Artigo

## Métodos de Previsão de Séries Temporais e Modelagem Híbrida ambos Aplicados em Médias Mensais de Velocidade do Vento para Regiões do Nordeste do Brasil

Henrique do Nascimento Camelo<sup>1</sup>, Paulo Sérgio Lucio<sup>1</sup>, João Bosco Verçosa Leal Junior<sup>2</sup>,  
Paulo Cesar Marques de Carvalho<sup>3</sup>

<sup>1</sup>*Programa de Pós-Graduação em Ciências Climáticas, Universidade Federal do Rio Grande do Norte, Campus Lagoa Nova, Natal, RN, Brasil.*

<sup>2</sup>*Mestrado Acadêmico em Ciências Física Aplicadas, Universidade Estadual do Ceará, Campus Itaperi, Fortaleza, CE, Brasil.*

<sup>3</sup>*Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Ceará, Campus do Pici, Fortaleza, CE, Brasil.*

Recebido em 19 de Setembro de 2016 – Aceito em 16 de Março de 2017

### Resumo

Esse trabalho tem como objetivo realizar previsões de séries temporais da velocidade do vento em termos de médias mensais no nordeste brasileiro. Foram testados os seguintes modelos de séries temporais, Auto-Regressivo Integrado de Médias Móveis (ARIMA) e Holt-Winters (HW), e também inteligência artificial computacional com o uso de Redes Neurais Artificiais (RNA). E ainda foram testados dois modelos híbridos, o primeiro com combinação dos modelos ARIMA e RNA, o qual existem trabalhos na literatura, e o segundo trata-se de uma tentativa de combinação dos modelos HW e RNA. As séries ajustadas obtidas pelos modelos híbridos, mostram-se eficientes em acompanhar o perfil das séries observadas das regiões de estudo, com semelhanças aos dados observados em termos de máximos e mínimos, indicando assim, a capacidade dos modelos em representar sazonalidades. O cálculo da estatística de erros envolvendo o modelo híbrido (HW e RNA) obteve os menores valores em Fortaleza, São Luís e Natal, por exemplo, com valores de erros percentuais de 3,80%, 4,91% e 2,85%, respectivamente. A redução das variáveis da estatística de erros por parte dos modelos híbridos quando comparado separadamente a utilização dos modelos (ARIMA, HW e RNA) poderá influenciar nas precisões das velocidades dos ventos previstas.

**Palavras-chaves:** energias renováveis, ARIMA, RNA, Holt-Winters.

## Time Series Forecasting Methods and Hybrid Modeling both Applied on Monthly Average Wind Speed for Regions of Northeastern Brazil.

### Abstract

The objective of this work is to perform time series forecasts of wind speed in terms of monthly averages in the Brazilian Northeast. The following time-series models, Auto-Regressive Integrated Moving Average (ARIMA) and Holt-Winters (HW) were tested, as well as computational artificial intelligence with the use of Artificial Neural Networks (ANN). In addition, two hybrid models were tested, the first with a combination of the ARIMA and ANN models, which work in the literature, and the second is an attempt to combine HW and ANN models. The adjusted series obtained by the hybrid models, are efficient to follow the profile of the observed series of the study regions, with similarities to the data observed in terms of maxima and minima, thus indicating the capacity of the models to represent seasonalities. The calculation of error statistics involving the hybrid (HW and ANN) model obtained the lowest values in Fortaleza, São Luís and Natal, for example, with percent error values of 3.80%, 4.91% and 2.85%, respectively. The reduction of the statistical variables of errors by the hybrid models when comparing the use of the models (ARIMA, HW and ANN) separately may influence the predictions of expected wind velocities.

**Keywords:** renewable energy, ARIMA, ANN, Holt-Winters.

## 1. Introdução

As energias renováveis estão se tornando em algumas nações a principal estratégia no sentido de atenuar mudanças climáticas, em decorrência de poluentes atmosféricos oriundos de atividades industriais (Akella *et al.*, 2009). Uma das fontes renováveis promissoras para geração de energia elétrica trata-se da geração eólica, o crescimento desse recurso certamente é motivado por pesquisas que garantem sua viabilidade em termos técnicos e ambientais. O impulso desse crescimento pode ter ocorrido a partir de constantes debates em diferentes meios acadêmicos, principalmente alertando para a necessidade de buscar alternativas energéticas em relação às fontes convencionais que usam em sua matéria-prima o petróleo para combustão na geração de energia e também para atenuar mudanças climáticas.

Nos últimos anos uma técnica de pesquisa sobre geração eólica que vem sendo estudada, trata-se da previsão de séries temporais da velocidade do vento. Por exemplo, Kavasseri e Seetharaman (2009) utilizam o modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA), o qual também é conhecido por modelagem Box-Jenkins (Box e Jenkins, 1976), para 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 da velocidade do vento é a utilização de Redes Neurais Artificiais (RNA), por exemplo, Cadenas e Rivera (2009) realizaram previsão a curto prazo horária na região de La Venta, Oaxaca, México. Também outro modelo que vem sendo estudado para previsão da 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.* (2010) os autores mostram que a SES fornece menores erros quando comparado aos modelos ARIMA e RNA aplicados em região litorânea do México. Também é possível encontrar trabalhos com a combinação dos modelos ARIMA e RNA para previsão de séries temporais, ou seja, os chamados modelos híbridos. Liu *et al.* (2012) analisam a utilização desse tipo de modelagem na previsão da velocidade do vento em fazendas eólicas da China.

Este trabalho tem como objetivo utilizar modelos de previsão de séries temporais para aplicação em médias mensais da velocidade do vento em região do nordeste brasileiro. Além disso, será também proposto um modelo híbrido para as regiões de estudo. Pretende-se identificar a confiabilidade dos modelos no sentido de representar os dados observados, e assim uma ferramenta que poderá auxiliar no estudo e avanço da geração eólica em diversas regiões do país, configurando-se em uma tentativa a mais de garantias aos tomadores de decisão na exploração do potencial eólico de uma dada região, tendo em vista a

possibilidade de realizar previsões da intensidade do vento local, ou seja, conhecer o regime de ventos no futuro.

## 2. Material e Métodos

### 2.1 Região de estudo e dados coletados

Os dados observados da velocidade do vento representam médias mensais entre o período de janeiro de 2010 até dezembro de 2014 ao nível de 10 m de altura em relação ao solo. As regiões de estudo são localizadas no Nordeste do Brasil: São Luís (2,58° S e 44,23° W); Parnaíba (2,89° S e 41,73° W); Fortaleza (3,77° S e 38,53° W); e Natal (5,75° S e 35,35° W). Os dados foram coletados de torres meteorológicas instaladas nos aeroportos das respectivas localidades.

### 2.2. Modelos de séries temporais

Os modelos de previsão de séries temporais utilizados foram:

(I) A metodologia de Box-Jenkins (Box e Jenkins, 1976), também conhecida como modelo Auto-Regressivo Integrado de Médias Móveis (ARIMA). Este modelo vem sendo muito utilizado para previsão de séries temporais; ver por exemplo, Kavasseri e Seetharaman, (2009); Cadenas e Rivera (2009) e Liu *et al.* (2012). O modelo ARIMA possui características flexíveis, permitindo ajustes e adaptações em seus parâmetros, contribuindo para que diferentes tipos de séries temporais sejam representadas, assumindo assim, uma estrutura de correlação linear entre os valores das séries temporais (Makridakis *et al.*, 1998).

A modelagem ARIMA é comumente aplicada em processos lineares não estacionários, quando as séries temporais não são estacionárias (por exemplo, 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 (Makridakis *et al.*, 1998). Para estes processos o modelo é comumente representado por  $ARIMA(p, d, q)$ .

De acordo com Makridakis *et al.* (1998) o número de vezes que a série original tem que ser diferenciada antes de resultar em uma série estacionária é denominada “ordem de homogeneidade”. Estes processos são descritos pelos modelos Auto-Regressivos, Integrados (número de diferenciações), e Médias Móveis de ordens  $p$ ,  $d$ , e  $q$ , respectivamente, sendo representado por  $ARIMA(p, d, q)$  que podem ainda, ter incluído um operador sazonal, representado então por  $SARIMA(p, d, q)(P, D, Q)$  (Makridakis *et al.*, 1998). Para estimação dos parâmetros do modelo ARIMA, uma vez determinados os valores de  $p$ , e  $q$ , pode-se estimar os  $p$  parâmetros  $\phi$  (coeficientes relativos ao filtro de médias móveis), os  $q$  parâmetros  $\theta$  (coeficientes relativos ao filtro autorregressivo estacionário), e os erros do modelo  $\epsilon_t$ , também chamados de resíduos, conforme a Eq. (1):

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \dots \quad (1)$$

$$\dots + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Maiores detalhes sobre a Eq. (1) e sua forma sazonal, podem ser encontrados em Makridakis *et al.* (1998). A seleção do modelo ARIMA mais adequado para se ajustar aos dados observados em cada localidade de estudo foi realizado através da utilização do critério de Akaike (cuja abreviatura é AIC, do inglês *Akaike Information Criterion*) (Akaike, 1973). Esse critério é baseado na identificação do menor valor para AIC, o qual teoricamente resulta no melhor modelo para ajustar-se aos dados observados.

Para avaliar se os modelos ARIMA possuem viabilidade para a previsão de médias mensais da velocidade do vento foi realizado uma análise de resíduos. Conforme em Gujarati e Porter (2009) para que o modelo ARIMA representado na Eq. (1) seja viável no tocante ao ajuste dos dados observados, o termo de erro  $\varepsilon_t$  deste modelo deve se comportar como um ruído branco, ou seja, possuir média zero, variância constante (são homocedásticos) e ser independentes. Além disso, o termo  $\varepsilon_t$  deve seguir uma distribuição normal. Para identificação destes 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, para maiores detalhes, veja Royston (1983), teste de independência dos resíduos a partir do Teste de Durbin-Watson, (Doane e Seward, 2014), e o teste que verifica a igualdade da variância dos resíduos, ou seja, hipótese de homoscedasticidade a partir do Teste de Breusch-Pagan, (Gujarati e Porter, 2009). 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 para utilização 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 que $\alpha$ , 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 que $\alpha$ , 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 que $\alpha$ , ou seja, $p > 0,05$ , (não rejeita $H_0$ ).

(II) Holt-Winters (HW) (Makridakis *et al.*, 1998), no ano de 1957 Holt expandiu o modelo de suavização exponencial simples para lidar com dados que apresentavam tendência linear, e assim realizar previsões que fossem mais precisas que as realizadas com suavização exponencial simples, por ajustá-las a esta condição. Em 1960, Winters estendeu o modelo de Holt, incluindo uma nova equação que possibilitasse ser acrescentada nas previsões o comportamento da componente sazonal dos dados que se estivessem trabalhando, gerando assim o método de HW (Makridakis *et al.*, 1998).

A equação referente ao método de suavização exponencial com sazonalidade e tendência linear, ou seja, com a componente sazonal sendo tratada de forma aditiva em (Makridakis *et al.*, 1998), é representada da seguinte forma:

$$Y_{t+n} = a_t + b_t \cdot h + s_{t+p+n} \quad (2)$$

em que:  $a_t$  (representa o nível da série, cuja unidade é dada em m/s, e está relacionada de que forma a série temporal prevista evolui com o tempo, sendo que é identificado se a mesma varia lentamente no tempo ou, excepcionalmente, sofre variações bruscas. Por exemplo, a série prevista pode ter movimento de crescimento lento de forma linear);  $b_t$  (tendência, cuja unidade é  $m/s^2$ , e está relacionada com o fato da série temporal prevista possuir movimentos de crescimento ou decrescimento que pode ocorrer em intervalos de tempos distintos. Por exemplo, uma série temporal prevista hipotética possui uma tendência de crescimento linear no seu primeiro ano, já no segundo ano a mesma foi identificada com um decrescimento linear);  $s_t$  (componente sazonal, cuja unidade é o m/s que está relacionada com o fato da série temporal prevista existir padrões cíclicos de variação que se repetem em intervalos relativamente constantes de tempo. Por exemplo, uma série temporal prevista hipotética durante um ano apresenta comportamento ondulatório a cada três meses);  $Y_{t+n}$  (previsão para o período  $n$  adiante, e possui unidade de m/s);  $p$  (período sazonal);  $n = 1, 2, 3, \dots, h$  (horizonte da previsão).

Para a realização da previsão com os modelos propostos através dos dados citados em cada localidade de estudo foram seguidas as seguintes etapas: (i) aplicação dos modelos propostos aos dados da velocidade do vento no período de janeiro de 2010 até dezembro de 2014. Essa etapa consiste em fazer com que os modelos propostos possam captar características das séries temporais tais como, tendência, nível e sazonalidade; (ii) cálculo de medidas de erros para identificar acurácia dos modelos e (iii) a partir da descoberta do modelo que fornece os melhores ajustes aos dados observados, realizar previsões da velocidade do vento para os meses do ano seguinte aos dados observados, como uma forma de ilustrar a capacidade do mesmo em representar características das séries observadas.

### 2.3. Redes Neurais Artificiais (RNA)

O modelo de Pitts e McCulloch era constituído de apenas uma saída e uma função de entrada, calculada através da soma ponderada de diversos valores. Esse modelo assemelha-se com o neurônio humano, onde sinais são recebidos pelo mesmo com intensidades diferentes, provocando uma determinada reação por parte da célula nervosa. Embora biologicamente inspiradas no neurônio humano, as RNAs encontraram aplicações em diferentes áreas científicas.

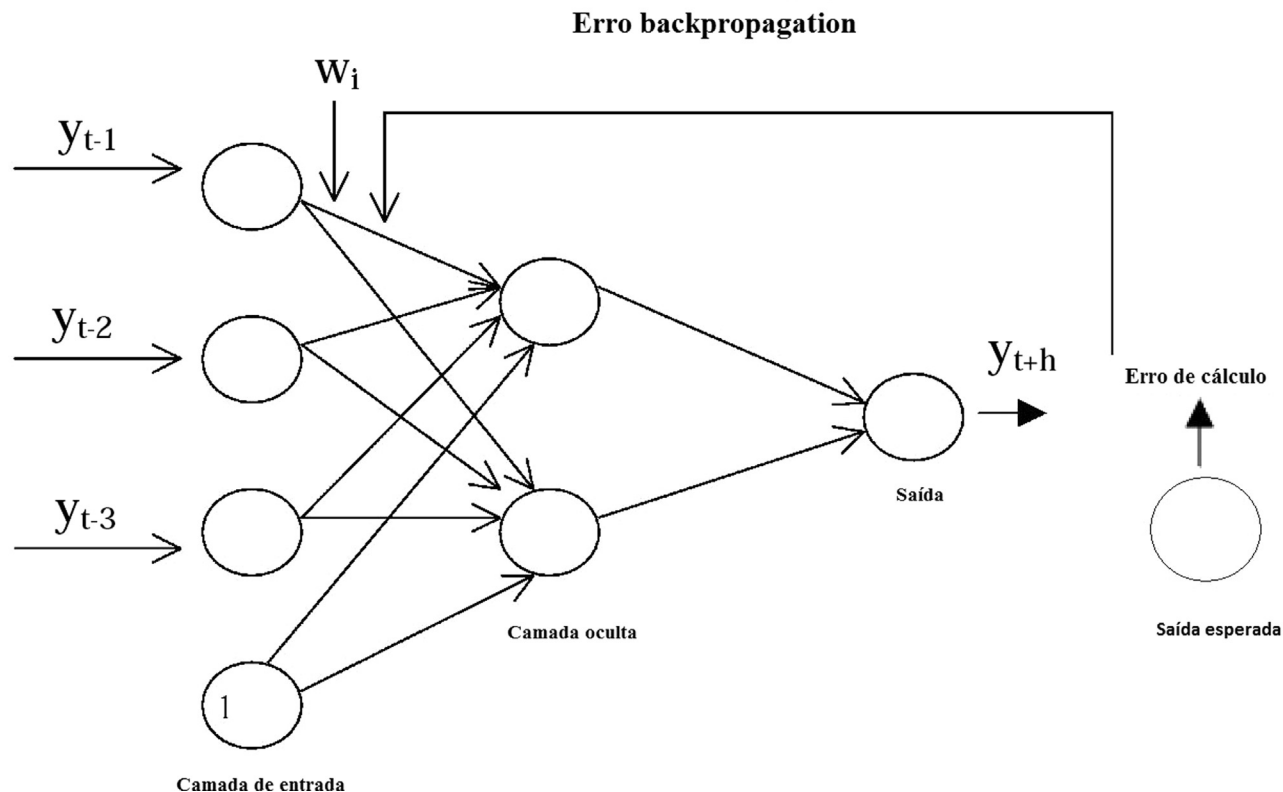
Conforme Zhang *et al.* (1998), em anos recentes uma das técnicas de inteligência computacional comumente usadas na tentativa de prever séries temporais é o treinamento de RNA. Estas são baseadas na arquitetura e aprendizagem do cérebro humano. As RNAs funcionam conceitualmente de forma similar ao cérebro humano, ou seja, tentando reconhecer regularidades e padrões de dados sendo capazes de aprender com a experiência e fazer generalizações baseadas no seu conhecimento previamente acumulado. As RNAs em sua estrutura podem possuir ambos modelos não-linear e linear e assim obter melhores resultados quando comparado com outros modelo de previsão (Zhang, 2003).

Uma RNA pode ser pensada como uma rede de “neurônios” organizados em camadas. Os preditores (ou entradas  $y_{t-i}$ ) formam a camada inferior, e as previsões (ou saídas  $y_{t+h}$ ) formam a camada superior. Pode haver camadas inter-

mediárias que contenham “neurônios ocultos”. A Fig. 1 mostra um exemplo para a estrutura de uma RNA com 4 entradas e com 1 camada oculta. Os coeficientes ligados aos preditores são chamados de “pesos” e comumente representados por  $w_i$ , os pesos são selecionados através de um “algoritmo de aprendizagem”, como é o caso do *backpropagation* (que surge do fato de que o algoritmo se baseia na retropropagação dos erros para realizar os ajustes de pesos das camadas intermediárias), que minimiza o erro entre as séries temporais previstas e observadas (Cadenas e Rivera, 2009).

Previsão de séries temporais é o processo de prever valores futuros de uma série temporal a partir do conhecimento de seus valores passados. Dentre as técnicas utilizadas para previsão de séries temporais estão aquelas baseadas em diferentes arquiteturas de RNA, do tipo *feedforward* como a Multi-Layer Perceptron (MLP). As redes MLP foram originalmente concebidas para executar tarefas de natureza estática, não foram, portanto, idealizadas para tratar problemas temporais. O método de janela de tempo foi a primeira adaptação da rede MLP treinada com o algoritmo *backpropagation* para processamento dinâmico (Cadenas e Rivera, 2009).

Na estrutura da RNA, por exemplo, mostrada na Fig. 1, as saídas de nós em uma camada são entradas para a próxima camada. O resultado é, em seguida, modificado por uma função não linear antes da saída. Por exemplo, as



**Figura 1** - Exemplo de estrutura RNA – MLP com algoritmo backpropagation com 3 entradas, 1 camada oculta e 1 saída.



entradas em neurônios  $j$  em camadas ocultas podem ser linearmente combinadas a partir da expressão:

$$Z_j = b_j + \sum_{i=1}^3 w_{i,j} y_i \quad (3)$$

Na camada oculta, a Eq. (3) é modificada utilizando uma função não linear, tal como uma sigmóide dada pela seguinte expressão:

$$s(z) = \frac{1}{1 + e^{-z}} \quad (4)$$

A Eq. (4), por sua vez, serve de entrada para a camada seguinte, sendo que essa estratégia tende a reduzir o efeito de valores extremos de entrada, proporcionando um melhor funcionamento da RNA. Os parâmetros  $b_1, b_2, b_3$ , e  $w_{1,1}, w_{3,2}$  são descobertos na etapa de “aprendizagem” a partir dos dados observados da série temporal em questão. Vale ressaltar que os pesos inicialmente podem assumir valores aleatórios, por sua vez, serão atualizados usando os dados observados. Maiores detalhes sobre a utilização de RNA para previsão de séries temporais podem ser obtidos em Zhang *et al.* (1998). Na literatura é possível destacar trabalhos sobre a utilização de RNA para previsão de série temporal da velocidade do vento. Flores *et al.* (2005) realizaram a previsão da velocidade dos ventos em uma fazenda eólica usando RNA, sendo baseada no algoritmo de aprendizagem *backpropagation*, cuja avaliação foi realizada com medição de dados reais de duas diferentes localizações.

Ressalta-se que a utilização de RNA pelo programa de computador  $R$  é proporcionada pelo pacote *forecast* apresentado por Hyndman e Khandakar (2008) a partir da função *nnetar*, por sua vez, é simbolizada por  $NNAR(p, P, k)(12)$ , em que,  $p$  (representa quantidade de termos auto-regressivo),  $P$  (é referente a dados sazonais), e  $k$  (representa o número de sub-camadas), e por fim, o número 12 é referente a quantidade de entradas  $y_i(t)$  da primeira camada da RNA.

#### 2.4. Modelagem híbrida

Neste artigo foram testados duas formas de modelagem híbrida para previsão da média mensal da velocidade do vento das regiões de estudo. Trata-se da combinação dos modelos (ARIMA + RNA) e (HW + RNA), os quais serão descritos a seguir:

##### I) Modelo híbrido (ARIMA + RNA)

Em Zhang (2003) o autor propõe a combinação dos modelos ARIMA e RNA, afirmando que este modelo pode capturar padrões diferentes nos dados. Os modelos estatísticos do tipo ARIMA são capazes de capturar os padrões lineares. Entretanto, por causa da sua flexível capacidade de modelagem não-linear, tem-se empregado frequentemente RNA para previsões de séries temporais (Aladag, 2009). O modelo híbrido combinado de previsão

para séries temporais sazonais ARIMA e RNA, é representado da seguinte forma:

$$hibrido(1) = ARIMA_{previs\tilde{a}o} + RNA_{previs\tilde{a}o(1)} \quad (5)$$

em que *hibrido(1)* representa a previsão do modelo híbrido, a qual é resultado da adição da componente linear prevista pelo modelo  $ARIMA_{previs\tilde{a}o}$ , com a componente não linear prevista pelo modelo  $RNA_{previs\tilde{a}o(1)}$ , sendo que esta última componente é modelada a partir dos resíduos do modelo ARIMA, os quais são comprovados que não são lineares pelo Teste de Breusch-Pagan, conforme mostrado na Tabela 1. É importante destacar que essa metodologia já foi empregada em diversos trabalhos de previsão de séries temporais da velocidade do vento, por exemplo, Cadenas *et al.* (2010) propuseram um modelo híbrido baseado no ARIMA e RNA para a previsão da velocidade dos ventos em três locais diferentes do México.

##### II) Modelo híbrido (HW + RNA)

A utilização deste modelo híbrido é semelhante ao descrito anteriormente, ou seja, é composto de uma componente linear e de outra componente não-linear, sendo representado pela seguinte expressão:

$$hibrido(2) = HW_{previs\tilde{a}o} + RNA_{previs\tilde{a}o(1)} \quad (6)$$

em que *hibrido(2)* representa a previsão do modelo híbrido, a qual é resultado da adição da componente linear prevista pelo modelo  $HW_{previs\tilde{a}o}$ , com a componente não linear prevista pelo modelo  $RNA_{previs\tilde{a}o(2)}$ , sendo que esta última componente é modelada a partir dos resíduos do modelo HW.

#### 2.5. Medidas de acurácia

Para que seja possível identificar a acurácia dos modelos de previsão propostos é necessário averiguar os seguintes métodos de estatística de erros. O Erro Médio Absoluto, mede o valor médio de erro entre as séries observadas e ajustadas. Nesta análise esta variável será representada por MAE (do inglês *Mean Absolute Error*), cuja representação matemática é dada por,

$$MAE = \frac{1}{o} \sum_{i=1}^o |v_{aju} - v_{obs}| \quad (7)$$

em que  $v_{aju}$  representa valor individual da série prevista e  $v_{obs}$  representa valor individual da série observada, e  $o$  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 RMSE, sendo esta abreviatura do inglês *Root Mean Squared Error*) representa as diferenças individuais quadráticas entre as séries temporais observadas e ajustadas:

$$RMSE = \sqrt{\frac{1}{o} \sum_{i=1}^o (v_{aju} - v_{obs})^2} \quad (8)$$

Da mesma forma o RMSE na Eq. (8) 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 ajustadas, e valores próximos de zero indicam um ajuste quase perfeito. Maiores detalhes sobre RMSE 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 esta expressão está na sua representação em termos percentuais (%) que fornece um rápido entendimento, e 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 esta variável é representada por

$$MAPE = \frac{1}{o} \sum_{i=1}^o \left| \frac{v_{aju} - v_{obs}}{v_{obs}} \right| \times 100. \quad (9)$$

Maiores detalhes sobre estatística de erros, em especial as Eqs. (7), (8), e (9) que serão usadas podem ser obtidos em (Cochran, 1977; Montgomery *et al.*, 2008).

Outra medida para identificar a qualidade dos ajustes, trata-se do coeficiente de eficiência de Nash-Sutcliffe (NS) (Nash e Sutcliffe, 1970). O coeficiente de eficiência pode variar entre  $-\infty$  a 1, sendo o valor 1 indicativo de um perfeito ajuste. Conforme Silva *et al.* (2008), quando o valor do NS resultar maior que (0,75) o desempenho do modelo é considerado bom. Para valores de NS entre (0,36 e 0,75) o desempenho é considerado aceitável, enquanto valores de NS inferiores a (0,36) fazem com que o modelo seja julgado como inaceitável. A expressão utilizada para o cálculo de NS é dada por

$$NS = 1 - \frac{\sum_{i=1}^o (v_{obs} - v_{aju})^2}{\sum_{i=1}^o (v_{obs} - \bar{v}_{obs})^2} \quad (10)$$

em que  $\bar{v}_{obs}$  representa a média da série temporal observada.

## 4. Resultados

A seleção do melhor modelo ARIMA para cada localidade é resultado do menor valor para o critério Akaike. A Tabela 2 mostram os modelos selecionados e as análises dos resíduos realizadas através dos testes citados. Os resultados mostram que para todas as regiões de estudo os pressupostos para os melhores modelos ARIMA são atendidos, satisfazendo a condição de ruído branco, ou seja, possuem homocedasticidade (Teste Breusch-Pagan); independência (Teste Durbin-Watson); e distribuição normal (Teste Shapiro-Wilk). Com isso ressalta-se que estes resultados são indicativos de que existe viabilidade para a utilização dos modelos ARIMA selecionados para as regiões de estudo. Em todas as localidades de estudo o melhor ajuste na utilização da modelagem Box e Jenkins foi fornecido com o uso do modelo SARIMA, ou seja, nestas localidades é necessário considerar a componente sazonal do modelo ARIMA. Uma vez que os pressupostos foram atendidos para os modelos ARIMA, a partir disso foi possível a utilização da modelagem híbrida (ARIMA+RNA), ou seja, esta é a condição necessária para modelar os resíduos através de RNA considerando-os referentes à parte não linear, o que será mostrado adiante.

A Tabela 3 mostra a estatística de erros na comparação entre as duas séries temporais, ou seja, observada e ajustada pelos modelos propostos entre o período de janeiro de 2010 até dezembro de 2014. Nas localidades de Fortaleza, São Luís e Natal os resultados mostram que o modelo híbrido (HW+RNA) apresenta menores valores da estatística de erro das variáveis MAE, RMSE e MAPE. Já para a localidade de Parnaíba o modelo híbrido (ARIMA+RNA) obteve o melhor desempenho em termos de menores valores para a estatística de erros.

Ainda a respeito da estatística de erros na maioria das análises foi identificado que a utilização das modelagens híbridas proporcionam melhores desempenhos para os ajustes aos dados observados em comparação aos outros modelos testados nesta pesquisa, por exemplo, na localidade de Parnaíba o uso do modelo híbrido (ARIMA+RNA)

**Tabela 2** - Apresentação dos modelos ARIMA para cada localidade e também testes aplicados aos resíduos.

Localidade (Modelo)	Shapiro – Wilk	Durbin – Watson	Breusch – Pagan
Fortaleza SARIMA (0,1,0)(1,0,0)	p-valor: 0,128. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,725. Decisão: os resíduos são independentes.	p-valor: 0,518. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
Parnaíba SARIMA (0,1,0)(1,0,0)	P-valor: 0,167. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,190. Decisão: os resíduos são independentes.	p-valor: 0,590. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
São Luís SARIMA (1,0,1)(1,1,1)	p-valor: 0,078. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,532. Decisão: os resíduos são independentes.	p-valor: 0,980. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).
Natal SARIMA (1,1,0)(1,0,1)	p-valor: 0,922. Decisão: os resíduos possuem distribuição normal.	p-valor: 0,426. Decisão: os resíduos são independentes.	p-valor: 0,460. Decisão: as variâncias dos resíduos são iguais (homoscedasticidade).

**Tabela 3** - Estatística de erros na comparação da série temporal observada com a obtida pelos modelos de previsão propostos.

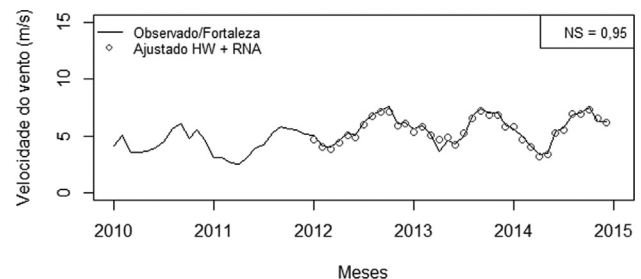
Fortaleza	ARIMA	HW	RNA	ARIMA+RNA	HW+RNA
MAE (m/s)	0,56	0,48	0,43	0,54	0,20
RMSE (m/s)	0,72	0,62	0,52	0,69	0,25
MAPE (%)	11,41	9,86	8,74	10,86	3,80
Parnaíba	ARIMA	HW	RNA	ARIMA+RNA	HW+RNA
MAE (m/s)	1,41	0,51	0,52	0,14	0,22
RMSE (m/s)	1,66	0,72	0,64	0,25	0,28
MAPE (%)	25,02	9,02	9,85	2,81	3,85
São Luís	ARIMA	HW	RNA	ARIMA+RNA	HW+RNA
MAE (m/s)	0,26	0,31	0,38	0,37	0,20
RMSE (m/s)	0,35	0,43	0,45	0,47	0,26
MAPE (%)	7,46	7,78	10,50	9,89	4,91
Natal	ARIMA	HW	RNA	ARIMA+RNA	HW+RNA
MAE (m/s)	0,39	0,43	0,28	0,25	0,13
RMSE (m/s)	0,50	0,55	0,37	0,57	0,26
MAPE (%)	8,68	9,71	6,52	2,98	2,85

possui um desempenho melhor do que somente a utilização do ARIMA. Essa afirmativa é baseada no fato de que este modelo híbrido proporciona uma melhoria na redução da variável MAPE de aproximadamente 2,81%, valor este muito menor do que somente a utilização do modelo ARIMA cujo o valor do MAPE é de aproximadamente 25,02%. Já em relação ao modelo híbrido (HW+RNA), para Parnaíba o ajuste proporciona também uma melhora significativa com redução da variável MAPE, em aproximadamente 3,81%, valor este quase três vezes menor em comparação a utilização somente do modelo HW que possui valor de aproximadamente 9,02%.

Para efeito de comparação com trabalhos semelhantes na literatura é possível destacar Cadenas *et al.* (2010) estudo em que os autores utilizam o modelo híbrido (ARIMA+RNA) para previsão de séries temporais de médias horárias da velocidade do vento em três regiões do México. Os resultados deste estudo mostraram que o modelo híbrido produz maior precisão das previsões da velocidade do vento apresentando menores valores da estatística de erros do que utilizando os modelos ARIMA e RNA separadamente. Como exemplo disso, o menor valor encontrado no modelo híbrido para a variável MAE é da ordem de 0,5 m/s. No trabalho de Rossi (2013), desenvolveu-se metodologia híbrida utilizando RNA do tipo *perceptron* multicamadas e os modelos estatísticos ARIMA com a abordagem de Box e Jenkins, com o objetivo de prever séries temporais de velocidade dos ventos na região da Ilha de Solteira do estado de São Paulo no sudeste do Brasil. Este trabalho revelou valor de MAPE da ordem de 22% considerando-se a previsão no horizonte de 24 h. Outros estudos comparativos podem ser encontrados, em Hu *et al.* (2013) os autores utilizaram o modelo ARIMA

para previsão de média mensal da velocidade do vento entre o período de janeiro de 2001 até dezembro de 2006. Neste estudo é possível observar que o menor valor de MAPE encontrado na comparação entre dados observados e ajustados pelo modelo para região montanhosa da China é de aproximadamente 13,70%. Já em termos de trabalhos de previsão da velocidade do vento com o uso do modelo HW é possível destacar previsão de médias horárias como em Malgorzata e Piotr (2014), os autores encontraram para as variáveis MAE e RMSE valores de 2,22 m/s e 2,93 m/s respectivamente, com horizonte de previsão de 6 h em região da Polônia. De acordo com os comentários realizados acima é possível afirmar que os resultados para a previsão de média mensal da velocidade do vento em regiões do Nordeste do Brasil, a partir dos modelos propostos estão em conformidade com os encontrados na literatura.

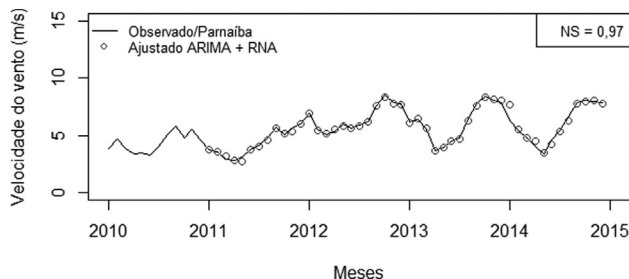
A Fig. 2 mostra o melhor modelo para ajuste aos dados observados em Fortaleza, ou seja, o modelo híbrido (HW+RNA), baseado nos resultados da estatística de erros apresentados na Tabela 3. A série temporal ajustada (representada em formato de círculos vazios) consegue acom-

**Figura 2** - Apresentação do modelo híbrido (HW + RNA) para previsão de média mensal da velocidade do vento em Fortaleza.

panhar o perfil da série temporal observada (representada na linha contínua cor preta), com semelhanças entre máximos e mínimos da velocidade do vento em alguns meses, por exemplo, no ano de 2014 o mês de abril representa o período de mínimo nas duas séries, ajustada e observada, com valores próximos de 4,0 m/s. No mês de outubro também de 2014, os valores de máximos são coincidentes nas duas séries em aproximadamente 7,0 m/s. O NS encontrado foi de 0,95 o que mostra o bom desempenho do modelo na eficiência do ajuste aos dados observados.

De acordo com estudo realizado a respeito da normal climatológica em todo território brasileiro pelo Instituto Nacional de Meteorologia (INMET), o qual pode ser consultado em (INMET, 2016), 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.* (2008) a média mensal da intensidade da velocidade do vento em Fortaleza é menor quando comparado ao período de baixa precipitação ou também chamado de período seco entre os meses de outubro até dezembro. Ainda na Fig. 2, em Fortaleza é possível identificar que o modelo híbrido (HW+RNA) consegue captar menores intensidades da média mensal da velocidade do vento no período chuvoso e também maiores intensidades para o período seco como é o caso do mês de outubro.

A Fig. 3 mostra o melhor modelo de previsão da média mensal da velocidade do vento em Parnaíba, trata-se do modelo híbrido (ARIMA+RNA) de acordo com os resultados da estatística de erros mostrados na Tabela 3. O NS encontrado na avaliação do modelo foi de 0,93 o qual demonstra a eficiência do ajuste (representado em formato de círculos vazios) aos dados observados (representado na linha contínua cor preta) em Parnaíba. A série temporal ajustada pelo modelo híbrido consegue acompanhar o perfil da série temporal observada em todo o período de comparação. Ocorrem equivalências de máximos e mínimos da velocidade do vento entre ambas as séries, por exemplo, em março de 2012 representa o mês de ocorrência do valor mínimo para as duas séries temporais que, por sua vez, os valores são coincidentes em aproximadamente 5,0 m/s. Para o mesmo ano, o mês de outubro foi o período que coincidiu com os valores máximos encontrados nas duas

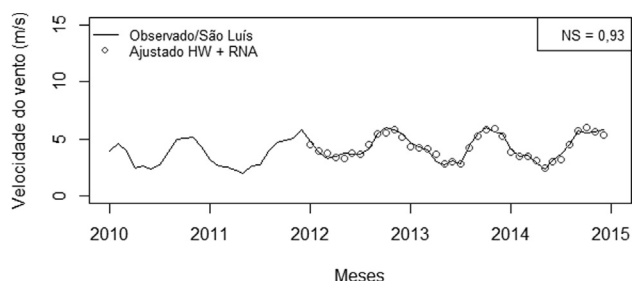


**Figura 3** - Apresentação do modelo híbrido (ARIMA + RNA) para previsão de média mensal da velocidade do vento em Parnaíba.

séries temporais, possuindo valores semelhantes em aproximadamente 8,0 m/s. É possível afirmar que o modelo híbrido (ARIMA+RNA) proposto para Parnaíba, consegue representar características sazonais da série temporal observada nessa localidade, ou seja, diferenciando períodos de menores e maiores intensidades da velocidade do vento. Essas diferenças assim como ocorreram em Fortaleza, podem estar associadas também aos períodos chuvoso e seco em Parnaíba, uma vez que, nessa cidade de acordo com sua normal climatológica os meses de maiores intensidades de precipitações estão entre fevereiro e maio, e os meses de menores índices de precipitações estão entre agosto e novembro (INMET, 2016).

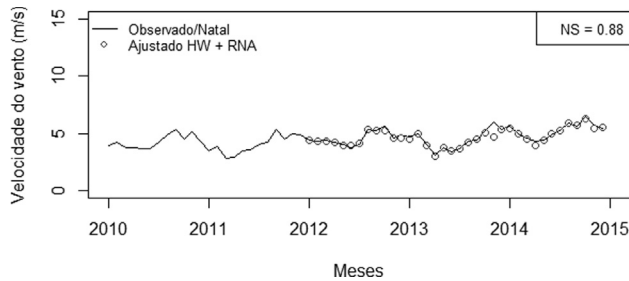
A Fig. 4 mostra os resultados do modelo de previsão híbrido (HW+RNA) para São Luís, o qual apresentou os menores valores da estatística de erros conforme a Tabela 3. Como é possível identificar a série temporal ajustada (representada em formato de círculos vazios) se assemelha por demais a série temporal observada (linha contínua cor preta) em termos de equivalências de valores máximos e mínimos da velocidade do vento. O NS encontrado mostra a eficiência do modelo no valor de 0,97. Diferentemente de Fortaleza e Parnaíba, para São Luís a série ajustada possui maior semelhança com a série observada, provavelmente esse resultado deve estar associado ao fato de que em São Luís as observações não possuem comportamentos desiguais em todo o período, ou seja, a sazonalidade é bem definida apresentando períodos de menores (entre fevereiro e maio) e maiores (entre agosto e novembro) intensidades da velocidade do vento. De acordo com a normal climatológica em São Luís, com respeito a precipitação, esses períodos mencionados também são referentes aos chuvosos e secos (INMET, 2016). De forma que as relações das intensidades das velocidades dos ventos, com os distintos períodos chuvosos e secos podem ser indicativos de que os ventos em São Luís sofrem influência direta da precipitação local.

A Fig. 5 mostra os resultados do modelo híbrido (HW+RNA) em Natal, o qual apresentou os menores valores para a estatística de erros conforme mostrado na Tabela 3, e com valor de NS de aproximadamente 0,88 que mostra a eficiência do ajuste (representado em formato de círculos vazios) aos dados observados (representados na linha conti-



**Figura 4** - Apresentação do modelo híbrido (HW + RNA) para previsão de média mensal da velocidade do vento em São Luís.





**Figura 5** - Apresentação do modelo híbrido (HW + RNA) para previsão de média mensal da velocidade do vento em Natal.

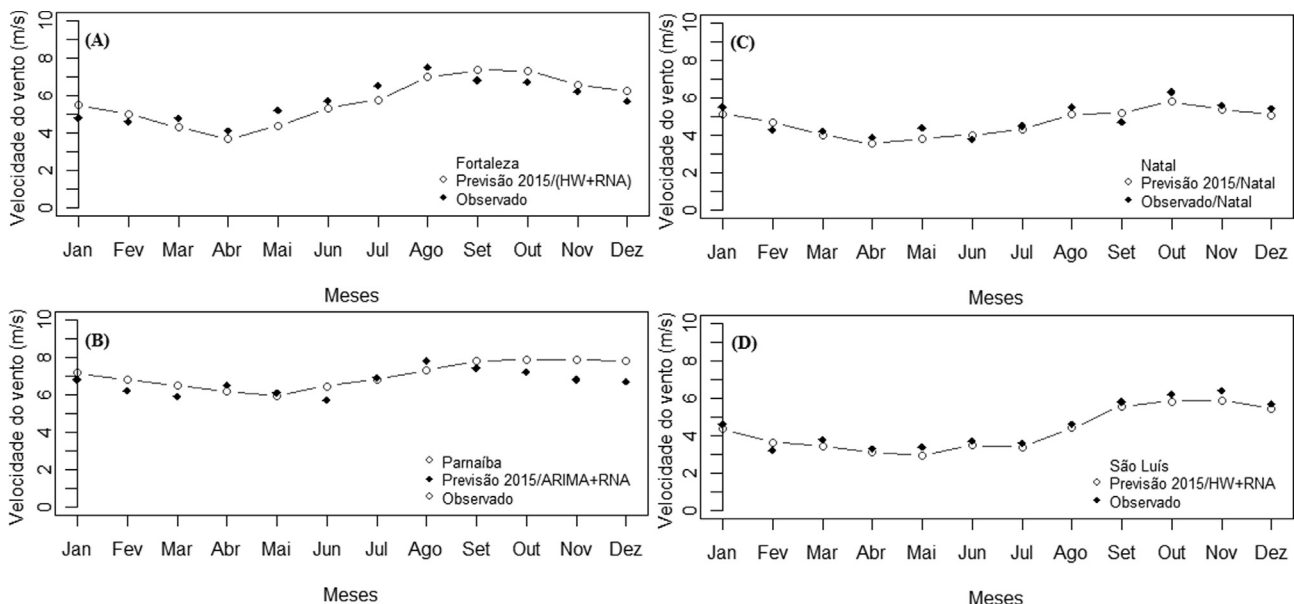
nua cor preta). Na maioria dos meses as séries temporais ajustadas e observada são similares, ou seja, possuindo valores da intensidade da velocidade do vento muito próximos, por exemplo, em abril de 2013 representa o valor mínimo em ambas as séries com valores muito próximo de 3,0 m/s. Em Natal os períodos de maiores precipitações estão entre (março e julho), já os de menores estão entre (setembro e dezembro), de acordo com informações obtidas no (INMET, 2016). Os dados observados da velocidade do vento em Natal apresentam intensidades distintas nos períodos chuvosos e secos mencionados, este resultado pode indicar a existência de sazonalidades para esta variável, sendo que o modelo híbrido proposto consegue representar estas sazonalidades, reforçando sua qualidade para aderência aos dados observados.

Com a finalidade de ilustrar ainda mais a qualidade dos modelos híbridos propostos (ou seja, os que apresentaram melhores acurácias, mostrados nas Figs. 2, 3, 4 e 5) na tentativa de realizar previsões da velocidade do vento, a Fig. 6 mostra as previsões (em formato de círculos vazio) de cada modelo para 2015, ou seja, ano seguinte das observações (2010 até 2014), as quais serviram para os ajustes

dos modelos. A Fig. 6A representa previsão para (Fortaleza), (Parnaíba) Fig. 6B, (Natal) Fig. 6C e (São Luís) Fig. 6D. Essas previsões são comparadas com suas respectivas séries temporais observadas (em círculos preenchidos cor preta) no mesmo ano de 2015. Os resultados dessas comparações mostram a eficácia dos modelos híbridos nas previsões das projeções de médias mensais da velocidade do vento das regiões de estudo, muito embora, já era esperado por ocasião da influência recebida dos perfeitos ajustes obtidos pelos modelos em questão. Outro detalhe identificado a cerca das previsões, refere-se ao fato de que as mesmas também conseguem representar as sazonalidades existentes das suas respectivas séries temporais ajustadas, ou seja, menores e maiores intensidades de acordo com os diferentes períodos chuvosos e secos.

## 5. Conclusões

As séries temporais ajustadas de médias mensais da velocidade do vento obtidas a partir de modelagens híbridas compostas de modelos de séries temporais e inteligência artificial, mostram-se eficientes em acompanhar o perfil das séries temporais observadas de médias mensais da velocidade do vento em regiões do nordeste brasileiro (Fortaleza, Parnaíba, Natal e São Luís). Além disso, é possível identificar que os modelos híbridos fornecem dados ajustados que possuem semelhanças de valores máximos e mínimos com os dados observados, indicando assim, a capacidade dos modelos em representar sazonalidades da velocidade do vento. O modelo híbrido (HW+RNA) proposto nesse artigo obteve os menores valores da estatística de erros nas regiões de Fortaleza, São Luís e Natal com valores de MAPE da ordem de 3,80%, 4,91% e 2,85%, respectivamente. E para Parnaíba o modelo híbrido



**Figura 6** - Comparações entre séries temporais previstas e observadas no período de 2015. (A) Fortaleza, (B) Parnaíba, (C) Natal e (D) São Luís.

(ARIMA+RNA) da literatura obteve os menores valores da estatística de erros, por exemplo, com valor de MAPE de aproximadamente 2,81%. A redução das variáveis de erros por parte dos modelos híbridos quando comparado separadamente a utilização dos modelos (ARIMA, HW e RNA) poderá influenciar nas precisões das velocidades dos ventos previstas. Esse trabalho poderá auxiliar em diversas áreas de interesse em previsão de velocidade do vento, por exemplo, no setor eólico, configurando-se em uma tentativa a mais de garantias aos tomadores de decisão na exploração do potencial eólico de uma da região, tendo em vista, a possibilidade de realizar previsões da intensidade do vento local, ou seja, conhecer o regime de ventos no futuro e assim otimizar a demanda de geração de energia elétrica proveniente de geração eólica.

### Agradecimentos

Os autores gostariam de agradecer ao Programa de Pós-Graduação em Ciências Climáticas (PPGCC) da Universidade Federal do Rio Grande do Norte (UFRN). Agradecimento por parte do autor correspondente ao apoio financeiro fornecido pela Agência Nacional de Petróleo (ANP)/PRH-51 a partir de bolsa de doutorado. Agradecimento por parte de co-autor ao apoio financeiro dado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa PQ2 - Proc. 301553/2016-5, e do Projeto de Pesquisa do CNPq - Proc. 400053/2014-4.

### Referências

- AKAIKE, H. Information theory and an extension of the maximum likelihood principle. Proc. 2<sup>a</sup> Inter. Symposium on Information Theory, 267-281, Budapest, 1973.
- AKELLA, A.K.; SAINI, R.P.; SHARMA, M.P. Social, economical and environmental impacts of renewable energy systems. **Renew. Energy**, v. 32, n. 2, p. 390-396, 2009.
- ALADAG, C.H.; EGRIOGLU, E.; KADILAR, C. Forecasting nonlinear time series with a hybrid methodology. **Appl. Math. Lett.**, v. 22, n. 9, p. 1467-1470, 2009.
- BOX, G.E.P.; JENKINS, G.M. Time Series Analysis: Forecasting and Control. Fifth Edition. **Wiley**, 712 p, 1976.
- CADENAS, E.; RIVERA, W. Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks. **Renew. Energy**, v. 34, n. 1, p. 274-278, 2009.
- CADENAS, E.; JARAMILLO, O.A.; RIVERA, W. Analysis and forecasting of wind velocity in chetumal, quintana roo, using the single exponential smoothing method. **Renew. Energy**, v. 35, n. 5, p. 925-930, 2010.
- 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.
- COCHRAN, W.G. Sampling techniques. Third Edition. **John Wiley & Sons**, 428 p, 1977.
- DINCER, F. The analysis on wind energy electricity generation status, potential and policies in the world. **Renew. Sustain. Energy Rev.**, v. 15, n. 9, p. 5135-5142, 2011.
- DOANE, D.P.; SEWARD, L.E. Estatística aplicada à administração e economia. Fourth Edition. **McGraw-Hill**, 840 p, 2014.
- EPE – **Empresa de Pesquisa Energética (2014)**: Balanço Energético Nacional, séries completas. 2015.
- FLORES, P.; TAPIA, A.; TAPIA, G. Application of a control algorithm for wind speed prediction and active power generation. **Renew. Energy**, v. 30, n. 4, p. 523-56, 2005.
- GUJARATI, D.N.; PORTER, D.C. Basic Econometrics. Fourth Edition. **McGraw-Hill**, 922 p, 2009.
- HU, J.; WANG, J.; ZENG, G. A hybrid forecasting approach applied to wind speed time series. **Renew. Energy**, v. 60, n. 1, p. 185-194, 2013.
- HYNDMAN, R.J.; KHANDAKAR, Y. Automatic Time Series Forecasting: The forecast Package for R. **J. Stat. Softw.**, v. 27, n. 3, p. 1-22, 2008.
- INMET – **Instituto Nacional de Meteorologia**. (2016). Página eletrônica: <http://www.inmet.gov.br/portal/index.php?r=clima/normaisclimatologicas>. Consultado em 11/05/2016.
- KAVASSERI, R.; SEETHARAMAN, K. Day-ahead wind speed forecasting using f-ARIMA models. **Renew. Energy**, v. 34, n. 5, p. 13388-1393, 2009.
- LIU, H.; TIAN, H.; LI, Y. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction. **Appl. Energy**, v. 98, p. 415-424, 2012.
- LUND, H. Renewable energy strategies for sustainable development. **Energy**, v. 32, n. 6, p. 912-919, 2007.
- MAKRIDAKIS, S.G.; WHEELWRIGHT, S.C.; HYNDMAN, R.J. Forecasting: methods and applications. Third Edition. **John Wiley and Sons**, 656 p, 1998.
- MALGORZATA, J.; PIOTR, O. Comparative analysis of wind speed short term forecasts for wind farms. **Agr. Eng.**, v. 152, n. 4, p. 81-90, 2014.
- McCULLOCH, W.; PITTS, W.A. A Logical Calculus of Ideas Immanent in Nervous Activity. **Bull. Math. Biophysics**, v. 5, p. 115-133, 1943.
- MONTGOMERY, D.C.; JENNINGS, C.L.; KULAHCI, M. Introduction to Time Series Analysis and Forecasting. Sixth Edition. **Wiley-Interscience**, 472 p, 2008.
- NASH, J.E.; SUTCLIFFE, J.V. River flow forecasting through conceptual models: a discussion of principles. **J. Hydrol.**, v. 10, p. 282-290, 1970.
- PANWAR, N.L.; KAUSHIK, S.C.; KOTHARI, S. Role of renewable energy sources in environmental protection: A review. **Renew. Sustain. Energy Rev.**, v. 15, n. 3, p. 1513-1524, 2011.
- ROSSI, D.J. Previsão da velocidade dos ventos por redes neurais artificiais e ARIMA de Box & Jenkins. **Dissertação de Mestrado – Universidade Estadual Paulista (UNESP)**, São Paulo, SP. 76 p., 2013.
- ROYSTON, J.B. Some techniques for assessing multivariate based on the Shapiro-Wilk W. **Appl. Statistics**, v. 32, n. 2, p.121-133, 1983.
- SILVA, P.M.O.; MELLO, C.R.; SILVA, A.M.; COELHO, G. Modelagem da hidrógrafa de cheia em uma bacia hidrográfica da região Alto Rio Grande. **Rev. Bras. Eng. Agríc. Ambient.** v. 12, n. 3, p. 258-265, 2008.
- ZHANG, G.; PATUWO, B.E.; HU, Y.M. Forecasting with artificial neural networks: The state of the art. **Int. J. Forecasting**, v. 4, n. 1, p. 35-62, 1998.
- ZHANG, G. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, v. 50, p. 159-175, 2003.