Revista Brasileira de Geografia Física

Homepage: www.ufpe.br/rbgfe

Predição de velocidade do vento em municípios do Nordeste brasileiro através de regressão linear e não linear para fins de geração eólica

Henrique do Nascimento Camelo¹, Paulo Sérgio Lucio², Oseas Machado Gomes³, João Bosco Verçosa leal Junior⁴

¹Doutorando do Programa de Pós-Graduação em Ciências Climáticas, Universidade Federal do Rio Grande do Norte (PPGCC/UFRN), CEP 59.072-970, Natal – RN, Brasil. Prof. Msc. do Instituto Federal de Educação, Ciência e Tecnologia do Ceará – IFCE, campus Acaraú, CE, Brasil. Autor correspondente. Endereço eletrônico: henriquecamelo13@yahoo.com.br; ²Prof. Dr. (PPGCC/UFRN), e-mail: pslucio@ccet.ufrn.br; ³Doutorando (PPGCC/UFRN), e-mail: oseasmachado@gmail.com; ⁴Prof. Dr. da Universidade Estadual do Ceará, Depto. de Física (UECE), Fortaleza, CE, Brasil. e-mail: jblealjr@uece.br.

Artigo recebido em 07/03/2016 e aceito em 14/04/2016

RESUMO

Atualmente as mudanças climáticas em todo o planeta, constituem grandes preocupações para a humanidade no que se refere a qualidade de vida. Para atenuar poluentes atmosféricos que são os principais causadores dessas mudanças climáticas é necessário investir cada vez mais em pesquisas de fontes renováveis de energia como, por exemplo, a geração eólica. Esses trabalho trata sobre aplicação de modelos de Regressão Linear Simples (RLS) e Regressão Não Linear (RNL) para predição de médias mensais de velocidade do vento em municípios do nordeste brasileiro voltado a geração eólica. Os dados de vento são ao nível de 10 m de altura no período de 2010 à 2014. Para variável explicativa em ambos modelos foi utilizado o Deficit de Pressão de Vapor d'água no ar (DPV), o qual depende da temperatura do ar e umidade do ar. Na configuração dos modelos de RNL foram ajustados, modelo exponencial crescente (RNL1) e modelo quadrático (RNL2). Foram aplicados testes estatísticos como, Shapiro-Wilk, Durbin-Watson e Breusch-Pagan, todos sobre os resíduos dos modelos para averiguar a destreza dos mesmos. O menor valor da Média do Erro Absoluto Percentual (MAPE) foi encontrado em aproximadamente 12,94% para o modelo de RLS. Já para os modelo de RNL1 e RNL2, os valores encontrados foram 12,87% e 18,80%, respectivamente. O presente trabalho possui ainda como objetivo o incentivo de pesquisas de fontes de energias renováveis como é o caso do recurso eólico e consequentemente seu uso como alternativa energética para o avanço social e econômico do país.

Palavras-chaves: Velocidade do vento, Predição, Regressão linear, Regressão não-linear, Software R.

Wind speed prediction in northeastern Brazil municipalities through Linear Regression and Nonlinear for wind power generation purposes

ABSTRACT

Currently climate change around the world, are major concerns for humanity as regards the quality of life. To mitigate air pollutants that are the main causes of these climate change is necessary to invest more in research on renewable energy sources such as, for example, wind generation. These work deals with application of simple linear regression models (RLS) and Nonlinear Regression (NNI) for wind speed average monthly prediction of municipalities in northeastern Brazil facing wind generation. The wind data is the level of 10 m high in the period 2010 to 2014. For explanatory variable in both models was used the Vapor Pressure Deficit of water in the air (VPD), which depends on the air temperature and air humidity. In the configuration of RNL models were adjusted, increasing exponential model (RNL1) and quadratic model (RNL2). Statistical tests were applied as Shapiro-Wilk, Durbin-Watson and Breusch- Pagan, all the waste of models to ascertain the ability of the same. The lower value of the average absolute error percentage (MAPE) was found in approximately 12.94% to the model of RLS. As for the model RNL1 and RNL2, the values were 12.87% and 18.80%, respectively. This work also has the objective of encouraging renewable energy sources research such as the wind resource and therefore its use as an energy alternative for social and economic advancement of the country.

Keyword: Wind, Prediction, Linear regression, Nonlinear regression, Software R.

Introdução

A geração de energia eólica vem

crescendo significativamente ao redor do mundo. Isso 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 as fontes convencionais que usam em sua matéria-prima o petróleo para a combustão na geração de energia e também para atenuar mudanças climáticas.

Os benefícios da utilização de fontes renováveis de energia como é o caso da eólica são vários, ou seja, desde ambientais até mesmo de substituição energética. No entanto, a questão ambiental deve ser preponderante para o avanço em pesquisas e consequentemente na atividade prática. Na medida em que haja cada vez mais a utilização de recursos renováveis de energia ao redor do mundo, será possível alcançar o Desenvolvimento Sustentável.

De acordo com a Global Wind Energy Council (GWEC, 2014), na atualidade aproximadamente 50% de toda energia elétrica produzida proveniente de geração eólica no mundo concentra-se em três países (China, EUA, Espanha). Em termos de Brasil também de acordo com (GWEC, 2014) o percentual de geração eólica está em aproximadamente 2%. Nesse sentido o país necessita avançar muito para ter uma alternativa energética não poluidora.

No Brasil é importante destacar o grande potencial eólico de seu território que de acordo com o Atlas do Potencial Eólico Brasileiro, estima-se que a capacidade de geração de energia elétrica proveniente de eólica é cerca de 143 mil MW de potência eólica, sendo que 51% desta energia está sobre a região nordeste do país (Amarante, 2001). A velocidade do vento nessas regiões possui médias anuais, a 50 m de altura, da ordem de aproximadamente $8,5 \text{ m.s}^{-1}$ o que tem feito com que o haja bastante atenção de empresas do setor eólico. 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^{-1} ao mesmo nível de altura (Troen e Petersen, 1989).

Esse trabalho tem como objetivo utilizar os modelos de Regressão Linear e Não Linear, para realizar predição de médias mensais de velocidade do vento em região do nordeste brasileiro. Dessa forma, pretende-se identificar a confiabilidade dos modelos propostos no sentido de representar os dados observados, e a partir disso uma possível ferramenta que poderá ser utilizada para o estudo e avanço da geração eólica em diversas regiões do país.

Material e métodos

Na literatura existem diversos trabalhos que descrevem o potencial eólico de diversas regiões do país utilizando os mais variados métodos. No entanto, em termos da utilização de modelos de Regressão os trabalhos ainda são reduzidos em especial para a região nordeste. O presente trabalho poderá ter utilidade como uma ferramenta para o estudo e avanço da geração eólica em diversas regiões do país.

Os dados de velocidade do vento, utilizados nesse trabalho são de médias horárias para um período de cinco anos, ou seja, de janeiro de 2010 até dezembro de 2014 e possuem unidade de metro por segundo (m.s^{-1}). Para esse período foram utilizados dados de municípios de três estados Ceará (CE), Piauí (PI) e Maranhão (MA) da região nordeste do Brasil, relativos as torres meteorológicas instaladas em aeroportos, conforme (Tabela 1), abaixo. Vale ressaltar que as medidas de velocidade do vento são realizadas a uma altura de 10 m em relação ao solo.

Os dados de velocidade do 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 que os dados meteorológicos estão disponíveis gratuitamente no endereço eletrônico (<http://www.airmetar.com.br/>).

Tabela 1: Localização geográfica dos dados de velocidade do vento.

Localidade	Latitude	Longitude
Fortaleza – CE	-3,77°	-38,52°
Parnaíba – PI	-2,92°	-41,75°
São Luís – MA	-2,6°	-44,22°

O termo Regressão em estatística representa a expressão de uma variável como

função de uma ou mais variáveis (Cecon *et al.*, 2012). Avaliar a possível relação entre uma

variável dependente com uma ou mais variáveis independentes é tarefa muito comum em estatística. É possível atingir esse objetivo através dos modelos de Regressão, os quais podem se dividir em duas classes: Regressão Linear ou Regressão Não-Linear.

O modelo de Regressão Linear (RL) é comumente representado através da (Equação 1), abaixo:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

em que, n é o tamanho da amostra, Y_i é a observação da variável dependente para a i -ésima amostra (também chamada de variável preditora ou ainda a variável na qual se deseja realizar a predição), $X_i = (x_{i1}, x_{i2}, \dots, x_{ik})'$ é um vetor de observações das variáveis independentes para i -ésima amostra, $\beta = (\beta_0, \beta_1, \beta_2, \dots, \beta_k)'$ é um vetor de coeficientes de regressão (que podem ser encontrados através do método de máxima verossimilhança ou ainda pelo método dos mínimos quadrados) e ε_i é um componente do erro aleatório ou também chamado de resíduo. A forma como está ilustrado a (Equação 1) recebe o nome de Regressão Linear Múltipla (RLM), já que existem várias variáveis independentes para se obter uma variável dependente. Ou seja, a RLM é uma forma geral de RL, pois caso exista apenas uma variável dependente é sugerido falar em Regressão Linear Simples (RLS). Maiores detalhes sobre a RL utilizada nesse trabalho poderão ser obtidos através de (Gujarati, 2008; Larson e Farber, 2010).

Para o modelo apresentado na (Equação 1), foi utilizado nesse trabalho a variável resposta velocidade do vento (medida em m.s^{-1}) e a variável explicativa Deficit de Pressão de Vapor d'água no ar (DPV), ou seja, uma variável psicométrica que é representada pela (Equação 2), abaixo.

$$DPV = (e_s - e_a), \quad (2)$$

em que a unidade é dada em (kPa), sendo que e_s é a pressão de saturação de vapor d'água no ar em (kPa) e, e_a representa a pressão parcial de vapor d'água em (kPa). O DPV foi calculado a partir dos valores de temperatura do ar (T dada em $^{\circ}\text{C}$), utilizados para determinar a pressão de saturação de vapor d'água no ar, conforme proposto por Tetens (1930), visualizado na (Equação 3) abaixo.

$$e_s = 0,6108 \times 10^{\left[\frac{7,5T}{237,5+T} \right]} \quad (3)$$

Em seguida, a partir da umidade relativa (UR dada em %) e da pressão de saturação de vapor d'água no ar e_s , determinou-se a pressão parcial de vapor d'água no ar, mostrada na (Equação 4) abaixo.

$$e_a = \frac{UR \cdot e_s}{100} \quad (4)$$

Portanto, a diferença entre a pressão de saturação e a pressão parcial de vapor d'água no ar possibilitou o cálculo do DPV, que é a medida da quantidade de vapor d'água. Ainda maiores detalhes sobre as (Equações 2, 3, 4), poderão ser obtidos em (Alvarenga *et al.*, 2014).

Para identificar a confiabilidade dos modelos RL nesse trabalho foram utilizados testes estatísticos comumente empregados na literatura para análise dos resíduos. Através da (Equação 5), os resíduos podem ser matematicamente representados por:

$$\varepsilon_i = Y_i - \hat{Y}_i = \dots \dots = Y_i - (\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}), \quad (5)$$

na qual, $\hat{Y}_i$ representa a resposta ou predição fornecida pela RLM de acordo com Gujarati (2010).

De acordo com Souza (1998), modelos de Regressão Não Linear (RNL) podem ser escritos da seguinte maneira:

$$y_i(x_i) = f(x_i, \theta) + \varepsilon_i, \quad i = 1, \dots, n, \quad (6)$$

em que a função resposta $f(x_i, \theta)$ é não linear, y_i representa a observação da variável dependente, x_i representa um vetor de observações em k variáveis preditoras, $\theta = [\theta_1 \dots \theta_p]'$ é um vetor de parâmetros p dimensional desconhecidos e ε_i representa o erro aleatório.

Para a aplicação da RNL foram testados diversos modelos, porém será apresentado apenas os que forneceram melhores ajustes comparado aos dados observados. Em todos os modelos RNL a variável resposta da mesma forma que na RLS é a velocidade do vento (medida em m.s^{-1}) e a variável explicativa Deficit de Pressão de Vapor d'água no ar (DPV). A (Equação 7) representa um modelo exponencial que será simbolizado de RNL1.

$$y_i(x_i) = a \cdot e^{bx_i}, \quad (7)$$

em que a e b são coeficientes cuja unidade (m.s^{-1}) e (kPa^{-1}) respectivamente sendo esses valores encontrados através do ajuste aos dados observados. A (Equação 8) representa um modelo quadrático que será simbolizado de RNL2.

$$y_i(x_i) = A \cdot x_i^2, \quad (8)$$

na qual A é coeficiente dado em (m.s^{-1}).(kPa^{-1}) do modelo quadrático sendo esse valor encontrado através do ajuste aos dados observados.

Para que seja possível identificar se o ajuste aos dados observados pelos modelos de RLS e RNL é satisfatório é necessário averiguar se os seguintes pressupostos são atendidos:

a) Identificar quais combinações em variáveis explicativas fornecem o maior R^2 ajustado (esse resultado indica o quanto de

variação em $\hat{Y}_t$ pode ser explicada pelas variáveis explicativas), e também identificar se o p-valor dos coeficientes da (Equação 1) são significativos, ou seja, estão abaixo do nível de significância de 0,05.

b) Análise dos resíduos através de:

- Os resíduos ε_i são variáveis aleatórias de variância constante (σ^2) – hipótese de homoscedasticidade;
- Os resíduos $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, são independentes;
- Os resíduos ε_i seguem uma distribuição normal: $\varepsilon_i \sim N(0, \sigma^2)$.

Para identificação desses pressupostos

existem alguns testes na literatura que são possíveis de aplicação sobre os resíduos. Dentre eles destaca-se o teste de verificação de normalidade dos resíduos através do Teste de Shapiro-Wilk (TSW), maiores detalhes podem ser obtidos em (Royston, 1983), teste de independência dos resíduos através do Teste de Durbin-Watson (TDW), 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 (TBP), maiores detalhes em (Gujarati, 2010). Na (Tabela 2), tem-se a representação resumida dos testes utilizados, sendo que os mesmos serão aplicados ao nível de significância de $\alpha = 0,05$.

Tabela 2: Apresentação dos testes para identificar pressupostos dos modelos de RLS e RNL.

Teste de Shapiro-Wilk (TSW)	
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 α , ou seja, $p > 0,05$ (não rejeita H_0).	
Teste Durbin-Watson (TDW)	
H_0 : os resíduos são independentes	H_1 : os resíduos não são independentes.
Tomar decisão: se o p-valor for maior do α , ou seja, $p > 0,05$ (não rejeita H_0).	
Teste Breusch-Pagan (TBP)	
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).	

Outro método capaz de identificar se os modelos utilizados nesse trabalho são capazes de representar os dados reais são os métodos estatísticos de erros. O Erro Médio Absoluto, mede o valor médio de erro entre as séries observadas e preditas. Nessa análise essa variável será representada por *MAE* (do inglês *Mean Absolute Error*), cuja representação matemática é dada por,

$$MAE = \frac{1}{o} \sum_{i=1}^o |v_p - v_r|, \quad (9)$$

onde v_p representa valor individual da série predita e v_r 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, ou $m.s^{-1}$.

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 preditas, que pode ser definido matematicamente por,

$$RMSE = \sqrt{\frac{1}{o} \sum_{i=1}^o (v_p - v_r)^2} \quad (10)$$

Da mesma forma o *RMSE* na (Equação

10) possui unidade de $m.s^{-1}$. 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 preditas, e valores próximos de zero indicam uma predição quase perfeita. Maiores detalhes sobre a Raiz do Erro Quadrático Médio podem ser obtidos em (Pedro e Coimbra, 2012).

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,

$$MAPE = \frac{1}{o} \sum_{i=1}^o \left| \frac{v_p - v_r}{v_r} \right| \times 100 \quad (11)$$

Maiores detalhes sobre estatística de erros, em especial às (Equações 9, 10, 11) 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-PROJECT, 2014). 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).

Resultados e discussão

Na (Tabela 3) são apresentados os resultados dos testes sobre os resíduos dos modelos de RLS e RNL para averiguar se os pressupostos que garantem a qualidade da aderência dos ajustes são atendidos. Esses resultados são apresentados em termos dos p-valores encontrados. Os

pressupostos nas localidades de Fortaleza e São Luís para o modelo de RLS são atendidos em parte, ou seja, ao nível de significância de $\alpha = 0,05$ de acordo com os p-valores, possuem distribuição normal (TSW) e homoscedasticidade (TBP). No entanto, o critério de independência (TDW) dos resíduos não é atendido. Com respeito ao modelo exponencial de RNL1 da mesma forma os pressupostos são atendidos em parte, ou seja, ao nível de significância de $\alpha = 0,05$ para as localidades de Fortaleza e São Luís os resíduos possuem distribuição normal (TSW) e homoscedasticidade (TBP). Já para o teste de independência (TDW) dos resíduos não foi atendido. Por fim para o modelo quadrático os resultados na análise dos resíduos mostram que para as localidades de Fortaleza e Parnaíba também os pressupostos são atendidos em parte, sendo que nesse caso com os p-valores encontrados ao nível de significância de $\alpha = 0,05$ é possível afirmar que existe distribuição normal (TSW) e também homoscedasticidade (TBP). Mas para o critério de independência (TDW) dos resíduos não houve acerto.

Tabela 3: Averiguação dos pressupostos para identificar a aderência dos modelos de RLS e RNL. Para cada teste aplicado sobre os resíduos são apresentados os p-valores.

RLS	TSW	TDW	TBP
Fortaleza - CE	0,2139	$1,114 \times 10^{-6}$	0,3964
Parnaíba - PI	0,0191	$2,734 \times 10^{-10}$	0,0268
São Luís - MA	0,0726	$8,373 \times 10^{-10}$	0,2531
RNL1	TSW	TDW	TBP
Fortaleza - CE	0,2603	$1,121 \times 10^{-6}$	0,4417
Parnaíba - PI	0,0160	$4,176 \times 10^{-6}$	0,7608
São Luís - MA	0,0865	$4,846 \times 10^{-6}$	0,0651
RNL2	TSW	TDW	TBP
Fortaleza - CE	0,6981	$2,575 \times 10^{-12}$	0,4611
Parnaíba - PI	0,5128	$1,577 \times 10^{-11}$	0,2251
São Luís - MA	0,0437	$3,720 \times 10^{-13}$	0,1482

Muito embora os pressupostos para garantir a aderência dos modelos de RLS e RNL não são atendidos em sua totalidade, conforme mostrado, mas na estatística de erros é possível observar que a comparação direta entre os dados de velocidade de vento (observados) e (preditos), existe possibilidade de ser considerado os modelos propostos, haja vista que os valores dos erros são pequenos, conforme ilustrado na (Tabela 4). Exemplo disso, o menor valor de *MAPE*

encontrado para o modelo de RLS foi de 12,94% na localidade de Fortaleza. Já o maior valor de *MAPE* para o mesmo modelo foi encontrado de 19,11% na localidade de Parnaíba. No tocante ao modelo exponencial da RNL1, na (Tabela 4) é possível observar que o menor valor de *MAPE* foi de 12,87% para Fortaleza, já o maior valor foi de 18% para Parnaíba, ou seja, valores praticamente semelhantes aos encontrados para a RLS. Na análise do modelo quadrático da RNL2, na (Tabela

4) é possível observar que os erros para a variável *MAPE* são maiores quando comparado a RLS e RNL1, ou seja, o menor valor encontrado foi de 18,80% para Fortaleza, já o maior valor foi de 29,68% para São Luís. A conclusão que se pode chegar na observação da variável *MAPE* é que o

modelo RNL1 possui uma maior tendência para ajuste aos dados observados da velocidade do vento. Esse último resultado também é observado para a variável *MAE*, também mostrado na (Tabela 4).

Tabela 4: Estatística de erros – comparativo entre os modelos de RLS e RNL com os dados observados.

RLS	<i>MAE</i> (m/s)	<i>RMSE</i> (m/s)	<i>MAPE</i> (%)
Fortaleza - CE	0,62	0,74	12,94
Parnaíba - PI	0,98	1,11	19,11
São Luís - MA	0,56	0,64	15,51
RNL1	<i>MAE</i> (m/s)	<i>RMSE</i> (m/s)	<i>MAPE</i> (%)
Fortaleza - CE	0,62	0,75	12,87
Parnaíba - PI	0,97	1,10	18,65
São Luís - MA	0,55	0,63	15,00
RNL2	<i>MAE</i> (m/s)	<i>RMSE</i> (m/s)	<i>MAPE</i> (%)
Fortaleza - CE	0,89	1,09	18,80
Parnaíba - PI	1,30	1,60	26,80
São Luís - MA	0,98	1,18	29,68

Na (Figura 1) é ilustrado a dispersão entre as variáveis observadas DPV (kPa) e a velocidade do vento (m.s^{-1}), ambas representadas em (círculos vazios) e aplicado a RLS (representada em reta na cor azul) para cada localidade. Para Fortaleza e São Luís o ajuste do modelo aos dados observados é semelhante, conforme ilustrado nas (Figuras 1A e 1C), os valores dos R^2 ajustados foram respectivamente de aproximadamente 0,67 e 0,68. Já para Parnaíba o modelo apresentou R^2 ajustado inferior no valor de 0,51. Vale ressaltar que na construção do modelo para São Luís os coeficientes foram significativos ao nível de $\alpha = 0,05$ cujos os valores são $\beta_1 = 4,64 \times 10^{-16} (\text{m.s}^{-1}).(\text{kPa}^{-1})$ e $\beta_0 = 0,0463 \text{ m.s}^{-1}$.

A (Figura 2) ilustra a comparação entre os dados observados (círculo vazio) com dados preditos (círculo sólido) para o modelo de RLS em todo o período de estudo, ou seja, de janeiro de 2010 até dezembro de 2014, sendo que as (Figuras 2A, 2B e 2C), representam respectivamente Fortaleza, Parnaíba e São Luís. Nas localidades de

Fortaleza e São Luís é possível identificar que os valores preditos conseguem acompanhar o perfil dos valores observados com semelhanças de máximos e mínimos de velocidade do vento em todo o período de estudo, mostrando assim, a eficácia do modelo de RLS. Por exemplo, em Fortaleza para abril de 2014 o valor da velocidade do vento (observado) foi de aproximadamente 3,34 m/s, já para o predito o valor foi de aproximadamente 3,51 m/s, ou seja, de fato os dois valores muito próximos, representando uma diferença de 0,17 m/s. Já para São Luís em novembro de 2010, o valor da velocidade do vento (observado) foi de aproximadamente 5,17 m/s, já para o predito foi de aproximadamente 5,38 m/s, representando uma diferença de 0,21 m/s. Para Parnaíba o mesmo não é evidenciado, ou seja, em grande parte os valores preditos subestimam em muito os valores observados em especial a partir de maio de 2011.

Figura 1: Modelo de Regressão Linear Simples com a variável independente DPV (kPa) e a variável dependente velocidade do vento (m/s). (A) Fortaleza, (B) Parnaíba e (C) São Luís.

Figura 2: Comparação entre dados observados (círculo vazio) com dados preditos (círculo sólido) pelo modelo de RLS. (A) Fortaleza, (B) Parnaíba e (C) São Luís.

A (Figura 3) mostra a dispersão entre as variáveis observadas DPV (kPa) e a velocidade do vento (m.s^{-1}), ambas representadas em (círculos vazios), sendo que é aplicado o modelo exponencial da RNL1 (representado em linha curva na cor preta) para cada localidade. Para Fortaleza, Parnaíba e São Luís o ajuste do modelo aos dados

observados é semelhante, conforme ilustrado nas (Figuras 3A, 3B e 3C), os valores dos R^2 ajustados foram respectivamente de 0,66; 0,69 e 0,69. Esses resultados são melhores quando comparados aos encontrados no modelo de RLS.

Figura 3: Comparação do modelo de RNL1 (linha na cor preta) com os dados observados (círculo vazio). (A) Fortaleza, (B) Parnaíba e (C) São Luís.

A (Figura 4) ilustra a comparação entre os dados observados (círculo vazio) com dados preditos (círculo sólido) para o modelo de RNL1 em todo o período de estudo, ou seja, de janeiro de 2010 até dezembro de 2014, sendo que as (Figuras 4A, 4B e 4C), representam respectivamente Fortaleza, Parnaíba e São Luís. Nas três regiões de estudo para o modelo de RNL1 é possível identificar que os valores preditos conseguem acompanhar o perfil dos valores observados com semelhanças de máximos e mínimos de velocidade do vento em todo o período de estudo mostrando. Desta forma, pode-se afirmar a eficácia deste

modelo no sentido de representar os dados observados de velocidade de vento. Exemplo disso, em maio de 2014 em Fortaleza o valor da velocidade do vento (observado) foi de aproximadamente 4,61 m/s, já para o predito o valor foi de aproximadamente 4,49 m/s, ou seja, de fato os dois valores muito próximos, representando uma diferença de 0,12 m/s. Já para São Luís em abril de 2014, o valor da velocidade do vento (observado) foi de aproximadamente 2,85 m/s, já para o predito foi de aproximadamente 2,95 m/s, representando uma diferença de 0,10 m/s.

Figura 4: Comparação entre dados observados (círculo vazio) com dados preditos (círculo sólido) pelo modelo de RNL1. (A) Fortaleza, (B) Parnaíba e (C) São Luís.

A (Figura 5) mostra a dispersão entre as variáveis observadas DPV (kPa) e a velocidade do vento (m.s^{-1}), ambas representadas em (círculos vazios), sendo que é aplicado o modelo quadrático da RNL2 (representado em linha curva na cor preta) para cada localidade. Em termos de R^2 ajustado os resultados foram melhores quando

comparados ao modelo RNL1, ou seja, na (Figura 5A) a qual representa Fortaleza, o valor encontrado foi de 0,66. Agora na (Figura 5B) a qual representa Parnaíba, o valor encontrado foi de 0,70. E finalmente para São Luís, representada na (Figura 5C) o valor encontrado foi de 0,70.

Figura 5: Comparação do modelo de RNL2 (linha na cor preta) com os dados observados (círculo vazio). (A) Fortaleza, (B) Parnaíba e (C) São Luís.

A (Figura 6) ilustra a comparação direta entre os dados observados (círculo vazio) com dados preditos (círculo sólido) para o modelo quadrático de RNL2 em todo o período de estudo, ou seja, de janeiro de 2010 até dezembro de 2014, sendo que as (Figuras 6A, 6B e 6C), representam respectivamente Fortaleza, Parnaíba e São Luís. Muito embora os valores do R^2 ajustado sejam levemente maiores do que os encontrados nos modelos anteriores, apenas em Fortaleza é possível identificar que os valores preditos conseguem

acompanhar o perfil dos valores observados com semelhanças de máximos e mínimos de velocidade do vento em todo o período de estudo analisado. Exemplo disso, em outubro de 2012 o valor da média da velocidade do vento (observado) foi de 7,55 m/s. Já para o predito o valor encontrado foi de 7,63 m/s, representando para este mês uma diferença de 0,08 m/s. Para Parnaíba e São Luís o modelo de RNL2 não se mostrou eficiente havendo subestimação como também superestimação dos dados observados por parte dos preditos.

Figura 6: Comparação entre dados observados (círculo vazio) com dados preditos (círculo sólido) pelo modelo de RNL2. (A) Fortaleza, (B) Parnaíba e (C) São Luís.

Conclusão

Os resultados dos modelos de RLS e RNL mostram uma tendência de que o melhor ajuste aos dados observados de velocidade de vento em Fortaleza, Parnaíba e São Luís pode ser fornecido para o modelo exponencial de RNL1. Essa afirmativa é baseada em uma maior semelhança entre máximos e mínimos de velocidade do vento por parte dos valores preditos comparado aos valores observados. E aliado a esse resultado a variável *MAPE* é menor para esse modelo conforme mostrado. Muito embora os pressupostos para garantir a aderência dos modelos de RLS e RNL não são atendidos em sua totalidade nesse estudo, mas na estatística de erros é possível

observar que na comparação direta entre os dados de velocidade de vento (observados) e (preditos), existe possibilidade de ser considerado os modelos propostos tendo em vista que os valores dos erros são pequenos em especial com respeito a variável *MAPE*. Para próximos trabalhos poderão ser testado outros modelos de Regressão envolvendo outras regiões e consequentemente novos dados de velocidade do vento buscando desta forma a perfeição no ajuste. O presente trabalho poderá ter utilidade como uma ferramenta para o estudo e avanço da geração eólica em diversas regiões do país.

Agradecimentos

Os autores gostariam de agradecer ao Projeto Airmetar que foi desenvolvido pela kSoftwares pela doação dos dados, e também ao Programa de Pós-Graduação em Ciências Climáticas (PPGCC) da Universidade Federal do Rio Grande do Norte (UFRN). Finalmente, agradecimento à Agência Nacional de Petróleo (ANP) / PRH-51 pela bolsa de doutorado para o autor correspondente.

Referências

- Amarante, O. A. C. Brower, M., J. Zack e A.L. Sá, 2001. Atlas do Potencial Eólico Brasileiro. Ministério de Minas e Energia, Brasília.
- Alvarenga, C. B., Teixeira, M. M., Zolnier, S., Cecon, P. R., Siqueira, D. L., Rodrigues, D. E., Sasaki, R. S., Rinaldi, P. C. N., 2014. Efeito do deficit de pressão de vapor d'água no ar na pulverização hidropneumática em alvos artificiais. Bioscience Journal 30, 182-193.
- Cecon, P. R.; Silva, A. R.; Nascimento, M.; Ferreira, A., 2012. Métodos Estatísticos, 1 ed. Editora UFV, Viçosa.
- Doane, D. P.; Seward, L. E., 2014. Estatística aplicada à administração e economia. 4 ed. McGraw-Hill, São Paulo.
- GWEC. GLOBAL WIND ENERGY COUNCIL, 2014. Global Wind Report Annual Market Update. Disponível: http://www.gwec.net/wp-content/uploads/2015/03/GWEC_Global_Wind_2014_Report_LR.pdf. Acesso: 15 fev. 2016.
- Gujarati, D. N.; Porter, D. C., 2010. Econometria Básica, 5 ed. McGraw-Hill, São Paulo.
- Montgomery, D. C., Jennings, C. L., Kulahci, M., 2008. Introduction to Time Series Analysis and Forecasting. 6 ed. Wiley-Interscience, New York.
- Pedro, H.T.C., Coimbra, C.F.M., 2012. Assessment of forecasting techniques for solar power production with no exogenous inputs. Solar Energy 86, 2017-2028.
- Larson, R.; Farber, B., 2010. Estatística Aplicada. 4 ed. Pearson, São Paulo.
- R-PROJECT, 2014. The R Project for Statistical Computing. Disponível: <http://www.r-project.org/>. Acesso 10 jun. 2015.
- Royston, J. B., 1983. Some techniques for assessing multivariate based on the Shapiro-Wilk W. Applied Statistics, London 32, 121-133.
- Souza, G. S., 1998. Introdução aos modelos de regressão linear e não-linear. 1 ed. Embrapa, Brasília.
- Tetens, V. O., 1930. Über einige meteorologische Begriffe, Zeitschrift Geophysic 6, 297-309.
- Troen, I., Petersen, E.L., 1989. European Wind Atlas. Rise National Laboratory, Roskilde, Denmark.