



ISSN: 2525-815X

Journal of Environmental Analysis and Progress

Journal homepage: www.jeap.ufrpe.br/[10.24221/jeap.3.1.2018.1652.024-036](https://doi.org/10.24221/jeap.3.1.2018.1652.024-036)

Avaliação da distribuição estatística e elaboração de modelo de regressão múltipla linear da velocidade média do vento

Statistical distribution evaluation and modeling of multiple linear regression of average wind speed

Cícero de Deus Rosa Filho^a, Suellem Attila Parisi^a, Willames de Albuquerque Soares^a

^a Universidade de Pernambuco-UPE, Escola Politécnica de Pernambuco-POLI, Rua Benfica, n. 455, Madalena, Recife-PE. CEP: 50720-001. E-mail: cicerofilho@outlook.com, suellem.parisi@gmail.com, was@poli.br.

ARTICLE INFO

Recebido 19 Out 2017

Aceito 11 Jan 2018

Publicado 22 Jan 2018

ABSTRACT

The shortage of fossil fuels coupled with the need to reduce pollution has led man to seek alternative sources of energy, such as wind. In addition, climate changes can directly influence the speed of the winds, changing their behavior. Thus, this study aimed to verify the adhesion of the average velocities of the winds of the Curado region in Recife to the different probabilistic distributions, as well as to determine a multiple linear regression model for this variable. It was obtained data of the average wind speed and other 13 variables of the National Institute of Meteorology (INMET) for the Curado region of Recife. Subsequently, the Kolmogorov-Smirnov test (K-S test) was applied to the Normal, Log-Normal, Weibull and Gamma probabilistic distributions. After analyzing the sample by the K-S method, it was verified that only the Log-Normal distribution did not adhere to the test, since its result was above the critical value, while the Weibull distribution had the best result. In the regression study between the thirteen variables and mean wind velocity, correlation, covariance, variance inflation factor and determination coefficient tests were performed to verify possible relationships between these variables. The multiple linear regression results showed that the variables average maximum wind speed, tar evaporation, total precipitation, number of days with precipitation and total insolation together can be used as predictors of the average wind speed due to the good coefficient of determination.

Keywords: Weibull, adhesion test, probabilistic distribution.

RESUMO

A escassez de combustíveis fósseis aliada à necessidade de reduzir a poluição levou o homem a procurar fontes alternativas de energia, como a eólica. Ademais, as mudanças climáticas podem influenciar diretamente na velocidade dos ventos, alterando seu comportamento. Dessa forma, este estudo objetivou verificar a aderência das velocidades médias dos ventos da região do Curado no Recife às diferentes distribuições probabilísticas, assim como determinar um modelo de regressão múltipla linear para essa variável. Foram obtidos dados da velocidade média do vento e de outras 13 variáveis do Instituto Nacional de Meteorologia (INMET) para a região do Curado no Recife. Posteriormente, foi aplicado o teste de Kolmogorov-Smirnov (teste K-S) às distribuições probabilísticas Normal, Log-Normal, Weibull e Gama. Após analisar a amostra pelo método K-S, verificou-se que apenas a distribuição Log-Normal não aderiu ao teste, visto que seu resultado ficou acima do valor crítico, enquanto a distribuição Weibull apresentou o melhor resultado. No estudo da regressão entre as treze variáveis e a velocidade média do vento, foram realizados testes de correlação, covariância, fator de inflação de variância e coeficiente de determinação para verificar possíveis relações entre essas variáveis. Os resultados de regressão linear múltipla mostraram que as variáveis velocidade máxima média do vento, evaporação do piche, precipitação total, número de dias com precipitação e insolação total, juntas, podem ser utilizadas como

variáveis preditoras da velocidade média do vento devido ao bom coeficiente de determinação encontrado.

Palavras-Chave: Weibull, teste de aderência, distribuição probabilística.

Introdução

O uso de combustíveis fósseis pela população está gradativamente diminuindo as reservas naturais desses materiais. Além disso, os impactos do uso desenfreado de certas fontes de energia acabam sendo muito prejudiciais ao meio ambiente. As denominadas fontes “limpas” ou renováveis são uma alternativa para substituir os combustíveis poluentes, capazes de suportar o crescimento econômico e tecnológico dos países. Um exemplo dessas fontes renováveis é a energia eólica, em que se utiliza o vento como gerador de energia (Pes et al., 2005).

O estudo da velocidade média dos ventos é de grande importância para a sociedade principalmente na produção de energia sustentável, além de sua utilização em projetos estruturais de engenharia. Segundo Beruski et al. (2009), é a partir desses dados que se torna possível verificar se uma determinada região possui potencial eólico satisfatório, viabilizando tecnicamente a instalação de torres de captação de energia. A energia eólica vem ganhando espaço dentro da política de expansão energética brasileira nos últimos quatro anos. O Brasil possui o potencial de vento adequado para exploração eólica e o governo brasileiro tem apoiado o desenvolvimento do setor com o programa PROINFA (Programa de Incentivo às Fontes Alternativas de Energia Elétrica). Além da capacidade de geração de eletricidade, a energia eólica tem participação fundamental na questão de segurança energética brasileira (Pes, 2015).

Segundo Amarante et al. (2001), grande parte da capacidade de produção energética pela força do vento se concentra em apenas 0,8% do território nacional. A maior parte dessa produção se concentra na região Nordeste, seguida do Sudeste, Sul, Norte e Centro-Oeste (Silva, Rosa & Araujo, 2005).

Um fator de grande importância relacionado à inserção da geração eólica no Brasil é a complementaridade com a geração hidrelétrica, que remonta ao conceito de segurança energética discutido amplamente após a crise energética de 2001. O maior potencial eólico brasileiro ocorre na região nordeste e tem a melhor ocorrência de ventos no período de menor potencial hídrico do ano naquela região (ANEEL, 2008).

A energia eólica foi a fonte mais impulsionada pelo PROINFA, tanto que entre os anos de 2005 e 2013, o país passou de 27 MW de energia eólica instalada, para 2.207 MW (Ben,

2014 *apud* Pes, 2015). Ademais, o Brasil conseguiu desenvolver uma indústria nacional de aerogeradores com capacidade de produção de, aproximadamente, 750 MW por ano e com índice de 70% de nacionalização (EPE, 2007).

O estudo realizado por Breslow & Sailor (2002) aponta uma diminuição de 1% a 4,5% na velocidade do vento decorrente das mudanças climáticas globais para um período de 100 anos na região continental dos Estados Unidos. Além disso, segundo Pryor, Schoof & Barthelmie (2005), existe uma projeção de queda na velocidade média e na densidade do vento para as projeções climáticas entre 2080-2100.

Sailor et al. (2000) realizaram um estudo sobre os efeitos das mudanças climáticas no vento em superfície para o Nordeste dos Estados Unidos em que foi projetada uma diminuição de 40% no recurso eólico para ambos os cenários nos períodos de primavera e verão. Considerando o Brasil, o litoral do Nordeste, segundo Pes (2015), terá ocorrências de ventos médios acima de $8,5 \text{ m}\cdot\text{s}^{-1}$, aumentando ao longo das projeções, enquanto que em outras regiões do país, a velocidade do vento diminuirá significativamente. O balanço dessas variações para as diferentes regiões brasileiras mostra que o potencial eólico de 2100 em relação a 2001 terá um decréscimo de 31% a 60%. A análise desses dados e seu comportamento pode ser realizada através das funções de distribuição da estatística (Kiss & János, 2008).

Este estudo objetivou avaliar a aderência das velocidades médias dos ventos da região do Curado no Recife às diferentes distribuições probabilísticas e identificar um modelo de regressão múltipla linear para essa variável.

Material e Métodos

A análise da aderência foi realizada com dados do Instituto Nacional de Meteorologia (INMET) para o cálculo da velocidade média dos ventos da região do Curado no Recife durante o período de Jan/1986 a Dez/2016, totalizando 347 dados. As variáveis analisadas foram: velocidade média dos ventos (VVM), velocidade máxima média do vento (VMMM), evaporação do piche (EP), evapotranspiração potencial BH (EPBH), evapotranspiração real BH (ERBH), insolação total (IT), nebulosidade média (NM), número de dias com precipitação (NDP), precipitação total (PT), pressão atmosférica média (PAM), temperatura máxima média (TMaxM), temperatura

compensada média (TCM), temperatura mínima média (TMinM) e umidade relativa média (URM).

O teste K-S foi escolhido por ser um dos testes de aderência mais utilizado na literatura, assim como ocorreu com a escolha das distribuições probabilísticas Gama (Beruski et al., 2009), Log-Normal (Santana et al., 2013) e Weibull (Sansigolo, 2005; Silva, 1999; Pes et al., 2005). Para esse teste, quanto mais aderida a distribuição está, mais próximo de zero será o resultado do teste.

No caso da distribuição de Weibull, a literatura utilizou dados da velocidade dos ventos.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad \text{Eq.(1)}$$

$$f(x; \mu, \sigma) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[\frac{-(\log(x) - \mu)^2}{2\sigma^2}\right], & \text{se } x > 0 \\ 0, & \text{caso contrário} \end{cases} \quad \text{Eq.(2)}$$

onde $\alpha > 0$ e $\beta > 0$ são os parâmetros de forma e escala:

$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right], x > 0 \quad \text{Eq.(3)}$$

$$f(x) = \begin{cases} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}, & \text{se } x \geq 0 \\ 0, & \text{caso contrário} \end{cases} \quad \text{Eq.(4)}$$

$$D_n = \sup_{-\infty < x < +\infty} |F_n(x) - F_0(x)| \quad \text{Eq.(5)}$$

Para a análise da regressão múltipla da velocidade média do vento com as outras variáveis foram realizados testes de correlação e covariância entre essas variáveis, sendo, posteriormente, selecionadas aquelas que obtiveram melhores resultados para integrar a regressão. Esses valores passaram por uma análise do fator de inflação da

Neste teste considera-se a estatística como uma medida da discrepância entre a função de distribuição da amostra (F_n) e a função de distribuição proposta (F_0). Abaixo estão as distribuições Normal (Equação 1), Log-Normal (Equação 2), Weibull (Equação 3) e Gama (Equação 4), calculadas a partir de suas respectivas funções de densidade. Dessa diferença é obtido o valor máximo (D_n) para todos os valores da amostra, segundo a Equação 5.

Considerando μ a média e σ o desvio padrão:

variância (VIF), segundo a Equação 6, onde as variáveis com valores menores que 10 eram aceitas. Ao realizar a regressão foram verificados o coeficiente de determinação (R_{quad}) e os índices de p-valores. O resultado do p-valor de cada variável da equação da regressão deveria ser menor que 0,10 para ser aceito.

$$VIF = \frac{1}{1 - R_{quad}} \quad \text{Eq.(6)}$$

A regressão linear múltipla foi baseada no modelo da Equação 7, onde β_0 é o intercepto,

β_1 a β_n são os coeficientes angulares e X_1 a X_n são as variáveis e ε é o erro.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad \text{Eq.(7)}$$

Resultados

Inicialmente foi executado um histograma dos dados analisados para verificar as faixas de

valores com maior repetição na amostra. A Figura 1 mostra a distribuição de frequência dos valores da amostra da velocidade média dos ventos.

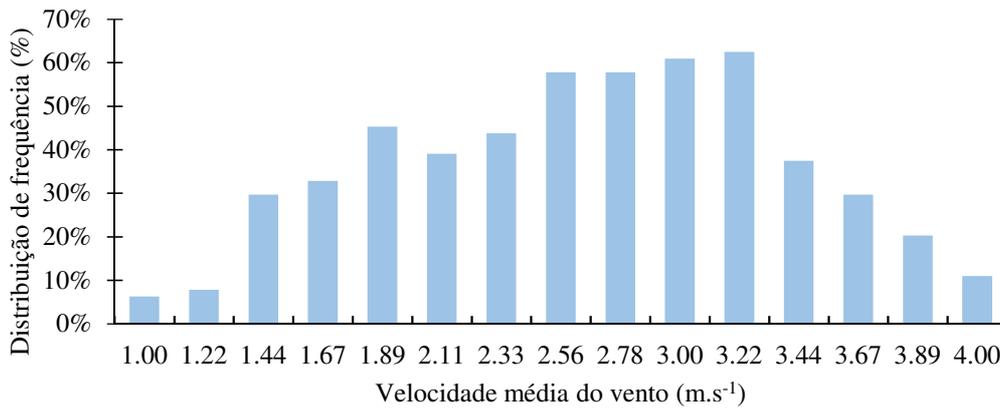


Figura 1. Distribuição de frequência dos valores da amostra da velocidade média dos ventos na região do Curado (Recife/PE) entre janeiro/1986 e dezembro/2016.

Em sequência foi possível verificar que grande parte dos valores se concentram entre as velocidades de $2,5 \text{ m.s}^{-1}$ a $3,22 \text{ m.s}^{-1}$.

Após a escolha das distribuições a serem analisadas, foi aplicado o teste de aderência K-S para cada uma delas.

Segundo o teste K-S, existe um valor crítico onde apenas as distribuições que estão abaixo dele são consideradas aderidas. Como a

amostra desse estudo está constituída por 347 dados, o valor crítico é de 0,073009.

Para a distribuição Normal foi calculada a média e o desvio padrão da amostra, com valores de 2,53 e 0,61, respectivamente. Posteriormente, calculou-se a distribuição normal dos dados da amostra, obtendo como resultado a Figura 2.

Aplicando-se o teste K-S para esta distribuição foi obtido o valor de 0,0451.

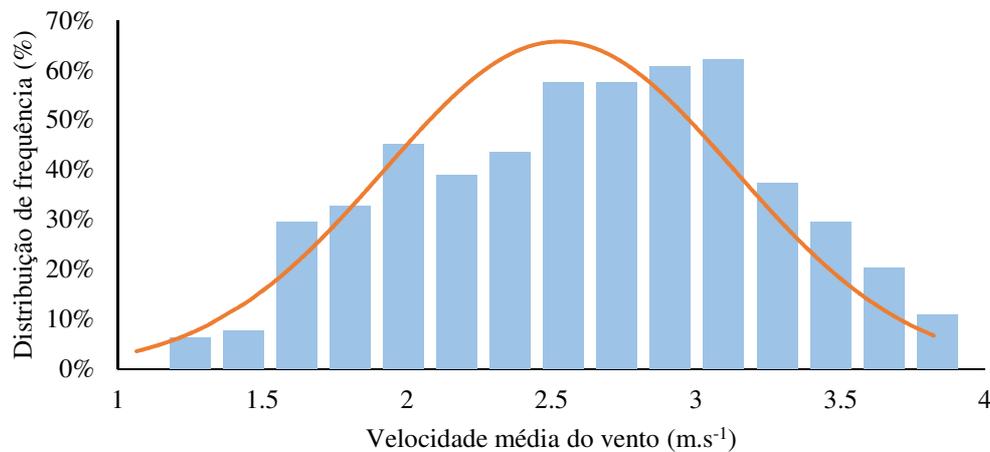


Figura 2. Distribuição Normal dos dados da amostra da velocidade média dos ventos na região do Curado (Recife/PE) entre janeiro/1986 e dezembro/2016.

Para a distribuição Weibull foram calculados os parâmetros alfa e beta, conforme Naghettini & Pinto (2007), encontrando-se valores de 4,76 e 2,76, respectivamente. Posteriormente,

calculou-se a distribuição dos dados da amostra, obtendo como resultado a Figura 3.

Aplicando-se o teste K-S para esta distribuição, foi obtido como resultado o valor de 0,0418.

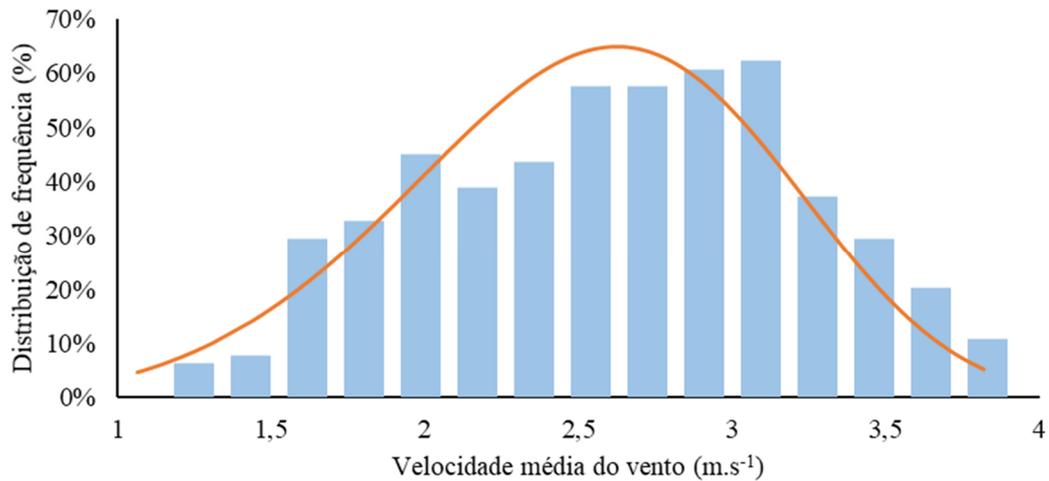


Figura 3. Distribuição Weibull dos dados da amostra da velocidade média dos ventos na região do Curado (Recife/PE) entre janeiro/1986 e dezembro/2016.

Para a distribuição Gama foram calculados os parâmetros alfa e beta pelo Método dos Momentos, conforme Naghettini & Pinto (2007), encontrando-se valores de 17,36 e 0,14, respectivamente. Posteriormente, calculou-se a

distribuição dos dados da amostra, obtendo como resultado a Figura 4.

Aplicando-se o teste K-S para esta distribuição foi obtido como resultado o valor 0,0725.

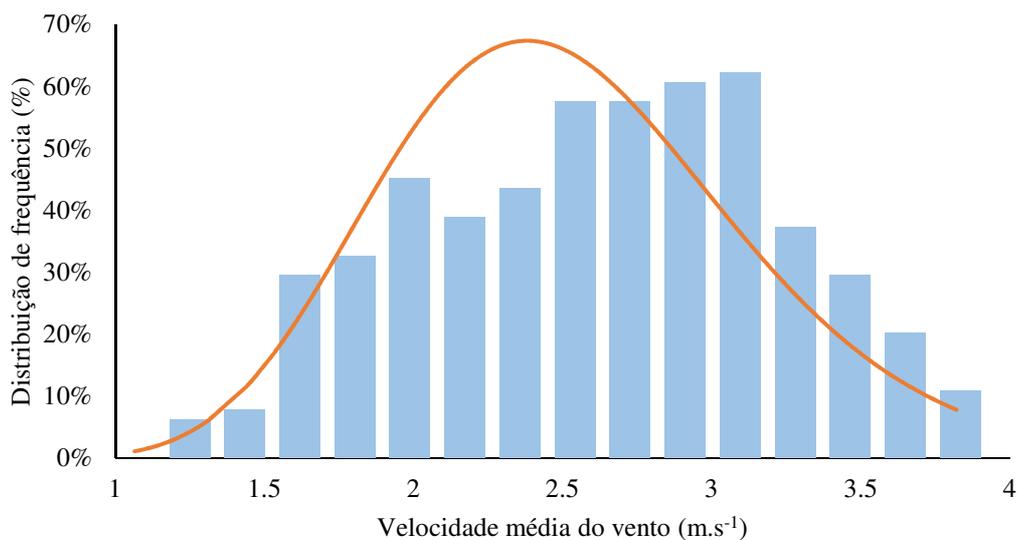


Figura 4. Distribuição Gama dos dados da amostra da velocidade média dos ventos na região do Curado (Recife/PE) entre janeiro/1986 e dezembro/2016.

Para a distribuição Log-Normal foram calculados os valores da média e do desvio padrão para a função $Y = \ln(x)$, sendo "x" os dados da amostra, cujos valores foram 0,89 e 0,26. Posteriormente, calculou-se a distribuição dos

dados da amostra, obtendo como resultado a Figura 5.

Aplicando-se o teste K-S para esta distribuição foi obtido como resultado o valor de 0,0818.

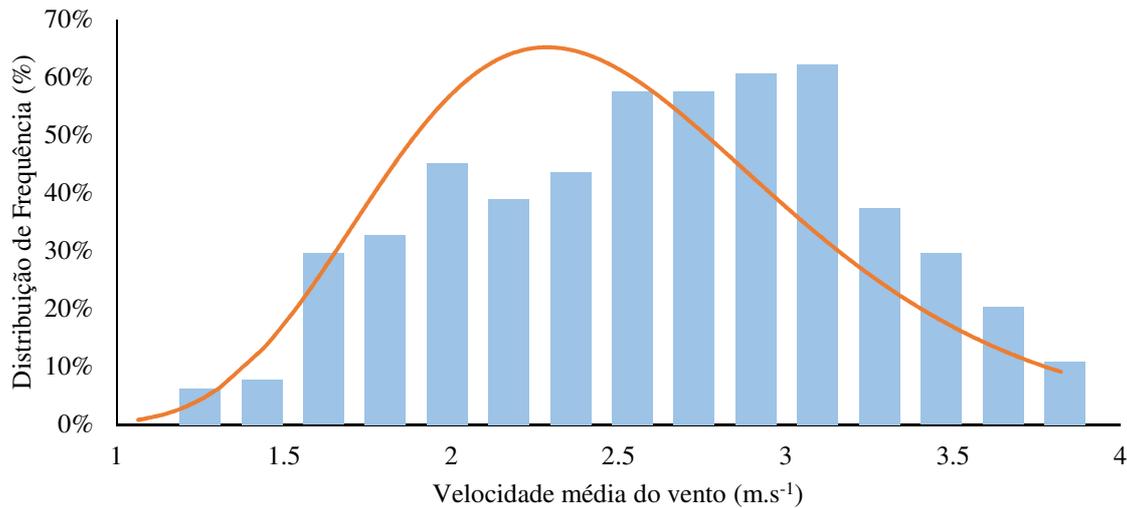


Figura 5. Distribuição Log-Normal dos dados da amostra da velocidade média dos ventos na região do Curado (Recife/PE) entre janeiro/1986 e dezembro/2016.

Após analisar a amostra pelo método K-S, verificou-se que apenas a distribuição Log-Normal não aderiu ao teste, visto que seu resultado ficou

acima do valor crítico. Todas as outras aderiram, tendo a distribuição Weibull o melhor resultado, conforme a Tabela 1.

Tabela 1. Comparação dos valores de cada distribuição e valor crítico dos dados da amostra da velocidade média dos ventos no bairro do Curado entre janeiro/1986 e dezembro/2016.

Distribuições	Weibull	Normal	Gama	Log-Normal
Teste K-S	0,0418	0,0451	0,0725	0,0818
Valor crítico	0,0730	0,0730	0,0730	0,0730

Comparando os gráficos das distribuições investigadas, consegue-se verificar que aquele que possui melhor aderência visual aos dados é o da distribuição de Weibull (Figura 6). Também se

pode perceber que a distribuição Log-Normal está mais afastada da curva que representa a frequência dos dados. Dessa forma, pode-se ver que o teste K-S descreve em números (Figura 6).

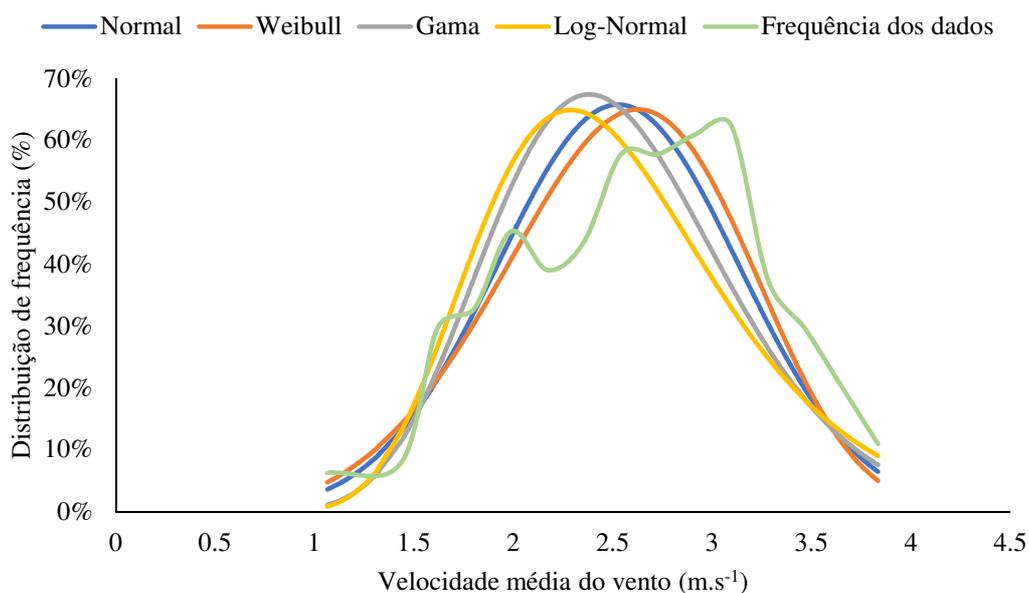


Figura 6. Gráfico comparativo das distribuições analisadas da velocidade média dos ventos na região do Curado (Recife/PE) entre janeiro/1986 e dezembro/2016.

O resultado encontrado por este estudo também foi encontrado por outros três que referenciaram este estudo (Sansigolo, 2005; Silva, 1999; Pes et al., 2005). Todos encontraram melhor aderência para o K-S na distribuição Weibull.

Para a análise da regressão foram dispostas todas as 14 variáveis com seus respectivos dados. Verificou-se que as variáveis pressão atmosférica média, evapotranspiração potencial BH e evapotranspiração real BH possuíam poucos dados comparados com as outras variáveis, com respectivamente 48, 134 e 134 dados. Devido a esse fato, decidiu-se eliminá-las da comparação, pois o intuito de encontrar a equação de regressão era justamente para preencher os dados vazios da velocidade média do vento, e, com essas variáveis, isso não seria possível.

Portanto, restaram 11 variáveis que formavam uma matriz com 300 dados cada, em que as análises de correlação e covariância eram possíveis.

O primeiro passo foi fazer a análise de covariância das variáveis para verificar, em primeira instância, quais as variáveis mais dependentes com a velocidade média do vento. A Tabela 2 mostra os resultados da covariância entre as 10 variáveis com a velocidade média do vento. Nesse caso, quanto mais próximo de zero o resultado estiver, mais independentes são as variáveis entre si; por outro lado, quanto mais afastadas de zero (em módulo) mais dependentes elas são.

Tabela 2. Resultados da covariância entre a variável velocidade média dos ventos e demais variáveis selecionadas: VVM = velocidade média dos ventos, VVMM = velocidade máxima média do vento, EP = evaporação do piche, IT = insolação total, NM = nebulosidade média, NDP = número de dias com precipitação, PT = precipitação total, TMaxM = temperatura máxima média, TCM = temperatura compensada média, TMinM = temperatura mínima média, URM = umidade relativa média.

	VVM	
VVM	-	
VVMM	0,36856	
EP	17,8972	
IT	8,40735	
NM	-0,03537	
NDP	-1,16814	
PT	-43,7226	
TMaxM	-0,0175	
TCM	0,06465	
TMinM	0,03204	
URM	-1,28331	

Resultado bom
 Resultado ruim
 Resultado aceitável

As variáveis com melhor covariância, ou seja, mais dependentes entre si são: EP, IT e PT (Tabela 2). As variáveis NM, TMaxM, TCM e TMinM, obtiveram valores muito próximos de zero. As variáveis VVMM, NDP e URM obtiveram valores que poderiam ser aceitos ou não, seriam necessários testes adicionais.

Em seguida foi verificado o coeficiente de correlação entre as variáveis. Para essa relação, quanto mais próximo de 1, mais dependentes são as variáveis, e, quanto mais próximas de zero, mais independentes elas são. A Tabela 3 mostra os valores obtidos com o coeficiente de correlação entre a VVM e as outras 10 variáveis.

Tabela 3. Valores obtidos com o coeficiente de correlação entre a variável velocidade média dos ventos e demais variáveis selecionadas: VVM = velocidade média dos ventos, VVMM = velocidade máxima média do vento, EP = evaporação do piche, IT = insolação total, NM = nebulosidade média, NDP = número de dias com precipitação, PT = precipitação total, TMaxM = temperatura máxima média, TCM = temperatura compensada média, TMinM = temperatura mínima média, URM = umidade relativa média.

	VVM	
VVM	-	
VVMM	0,59072	
EP	0,72816	
IT	0,37587	
NM	-0,10938	

Resultado bom
 Resultado ruim

NDP	-0,31257	■ Resultado aceitável
PT	-0,48191	
TMaxM	-0,02478	
TCM	0,09831	
TMinM	0,05999	
URM	-0,42779	

Analisando esses valores, pode-se verificar o que tinha sido constatado na Tabela 2. Os piores valores de correlação (NM, TMaxM, TCM e TMinM) foram os das mesmas variáveis da covariância. Ressalta-se, porém, que a VVMM a qual obteve um valor não tão bom na covariância, teve um valor bom na correlação. A variável EP foi a que obteve o melhor valor.

Após essas análises, decidiu-se eliminar as variáveis NM, TMaxM, TCM e TMinM por apresentarem valores muito baixos para a

covariância ($<0,07$), evidenciando sua independência com a velocidade média do vento.

Posteriormente à verificação da correlação e covariância, verificou-se o VIF entre as variáveis. Os valores ideais estariam abaixo de 10, aqueles acima de 10 deveriam ser eliminados. Assim, primeiramente, foi preciso calcular o coeficiente de determinação (r_{quad}), que é o coeficiente de correlação elevado ao quadrado. Depois, esses valores foram aplicados na fórmula do VIF. A Tabela 4 mostra os valores obtidos com o cálculo do VIF para todas as variáveis entre si.

Tabela 4. Valores obtidos com o cálculo do VIF entre as variáveis analisadas: VVM = velocidade média dos ventos, VVMM = velocidade máxima média do vento, EP = evaporação do piche, IT = insolação total, NM = nebulosidade média, NDP = número de dias com precipitação, PT = precipitação total, TMaxM = temperatura máxima média, TCM = temperatura compensada média, TMinM = temperatura mínima média, URM = umidade relativa média.

	VVM	VVMM	EP	IT	NM	NDP	PT	TMaxM	TCM	TMinM	URM
VVM	-	1,54	2,13	1,16	1,01	1,11	1,30	1,00	1,01	1,00	1,22
VVMM	1,54	-	1,06	1,00	1,05	1,04	1,00	1,10	1,06	1,03	1,01
EP	2,13	1,06	-	1,91	1,26	1,86	1,98	1,21	1,33	1,14	2,27
IT	1,16	1,00	1,91	-	2,88	2,25	2,48	1,49	1,47	1,11	2,57
NM	1,01	1,05	1,26	2,88	-	1,94	1,79	1,45	1,32	1,05	1,88
NDP	1,11	1,04	1,86	2,25	1,94	-	2,46	1,76	1,73	1,19	3,85
PT	1,30	1,00	1,98	2,48	1,79	2,46	-	1,28	1,30	1,06	3,44
TMM	1,00	1,10	1,21	1,49	1,45	1,76	1,28	-	15,86	2,65	1,84
TCM	1,01	1,06	1,33	1,47	1,32	1,73	1,30	15,86	-	4,47	1,97
TMM	1,00	1,03	1,14	1,11	1,05	1,19	1,06	2,65	4,47	-	1,26
URM	1,22	1,01	2,27	2,57	1,88	3,85	3,44	1,84	1,97	1,26	-

Apenas os valores para TMaxM e TCM ficaram acima de 10 (Tabela 2). Porém, essas variáveis tinham sido eliminadas pelo teste da covariância e correlação; todas as outras variáveis foram aprovadas no teste do VIF.

Antes de aplicar a regressão nas variáveis selecionadas, decidiu-se verificar os diagramas de dispersão dessas variáveis *versus* a velocidade média do vento para visualizar uma possível relação. Esse tipo de gráfico mostra o

comportamento de duas variáveis entre si, onde é possível verificar possíveis tendências, conforme mostrado nas Figuras 7-11.

Mesmo não apresentando um bom coeficiente de covariância, era esperado que a velocidade máxima média do vento mostrasse uma boa relação com a VVM. É possível verificar este fato a partir da análise da Figura 7, onde os pontos estão bem próximos da reta.

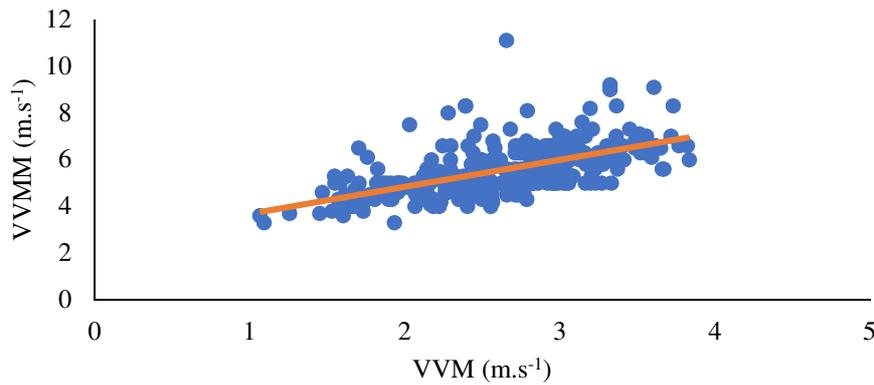


Figura 7. Dispersão entre as variáveis velocidade média dos ventos (VVM) e velocidade máxima média dos ventos (VVMM).

A evaporação do piche tinha apresentado bons coeficientes de covariância e correlação com a VVM. A dispersão dessas duas variáveis

mostradas na Figura 8 ratifica os valores calculados.

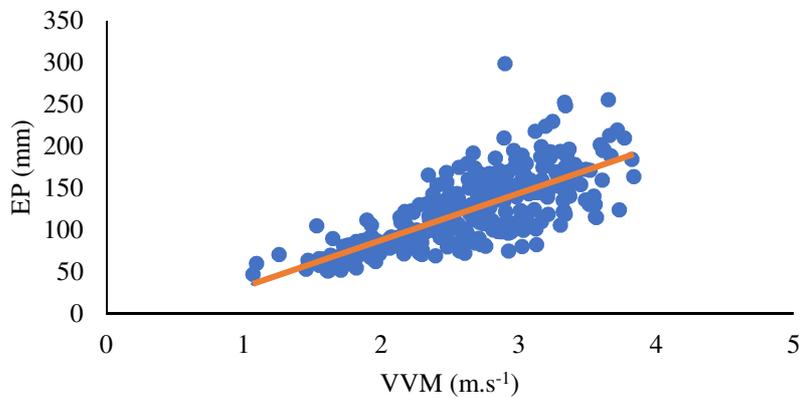


Figura 8. Dispersão entre as variáveis velocidade média dos ventos (VVM) e evaporação do piche (EP).

A insolação total possuiu um bom coeficiente de covariância, porém um aceitável coeficiente de correlação. A dispersão entre essas

variáveis, segundo a Figura 9 não é tão boa quanto a das duas variáveis anteriores, porém é possível verificar uma dependência entre elas.

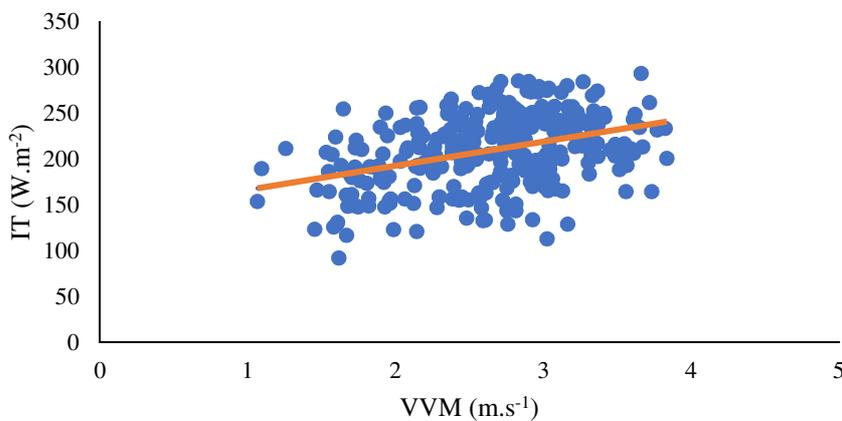


Figura 9. Dispersão entre as variáveis velocidade média dos ventos (VVM) e insolação total (IT).

A Figura 10 não apresentou boa dispersão das variáveis VVM e NDP. Os valores de covariância e correlação dessas duas variáveis

também não tinham sido tão bons, mas nada que inviabilizasse sua utilização e dependência.

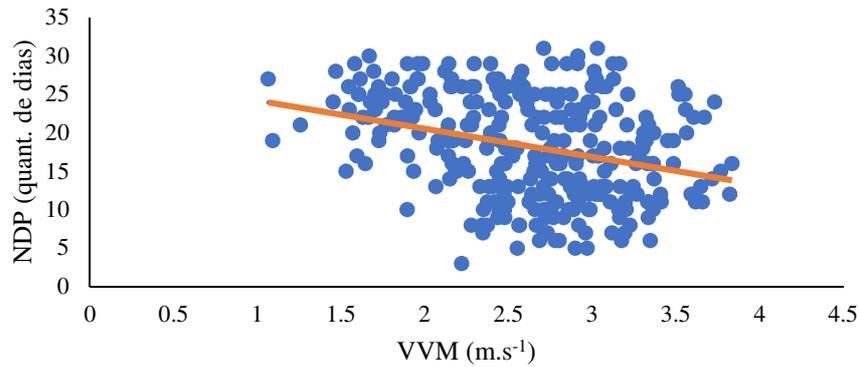


Figura 10. Dispersão entre as variáveis velocidade média dos ventos (VVM) e número de dias com precipitação (NDP).

O mesmo acontece com a Figura 11, onde os pontos se apresentam bastante afastados da reta. Entretanto, a precipitação total atingiu bom valor

de covariância com a VVM, verificando sua possível dependência.

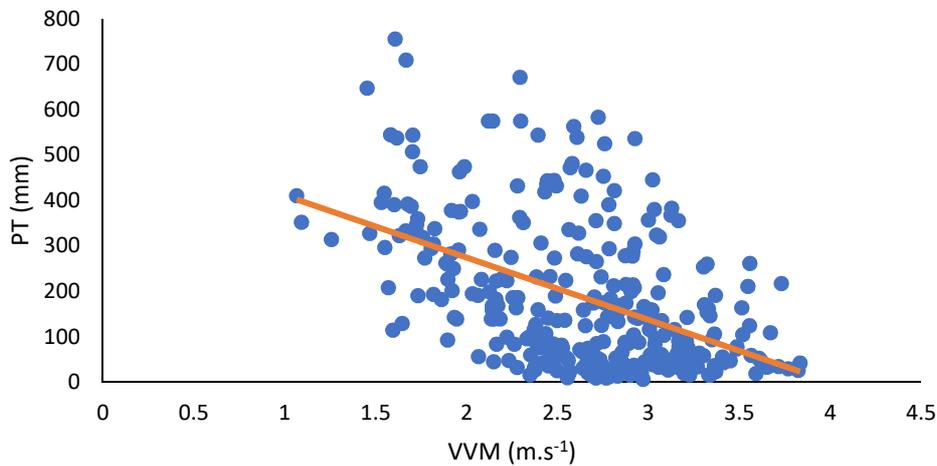


Figura 11. Dispersão entre as variáveis velocidade média dos ventos (VVM) e precipitação total (PT).

Esta etapa objetivou verificar, visualmente, o comportamento das variáveis com a VVM. Dessa forma, nenhuma das variáveis foram eliminadas nessa fase.

Após a análise das etapas anteriores, decidiu-se, inicialmente, fazer a regressão da VVM com as variáveis VVMM, EP, IT, NDP, PT e URM. As variáveis deveriam possuir o p-valor inferior a 0,10 e, caso não possuíssem, eram retiradas da equação e a regressão era realizada novamente até que todas elas estivessem abaixo do valor desejado.

Na primeira combinação, com as 6 variáveis citadas, a URM obteve um p-valor de 0,53 que estava acima do limite máximo de 0,10. Dessa forma, retirou-se essa variável para fazer uma nova regressão com as outras cinco. Na segunda combinação, todas as variáveis alcançaram valores abaixo de 0,10. Além disso, a segunda combinação obteve um melhor coeficiente de determinação e um menor erro que a primeira, conforme mostra a Tabela 5.

Tabela 5. Resultado da regressão com diferentes combinações entre as variáveis. VVMM = velocidade máxima média do vento, EP = evaporação do piche, IT = insolação total, NDP = número de dias com precipitação, PT = precipitação total, URM = umidade relativa média.

	Variáveis comparadas com a VVM					Dados da regressão		
	VVMM	EP	IT	NDP	PT	URM	Rquad ajustado	Erro
Combinação 1								
p-valor	9,0E-27	1,9E-25	2,1E-03	1,9E-02	1,4E-05	0,537	0,7367	0,291

Combinação 2									
p-valor	5,0E-27	4,0E-28	2,5E-03	2,0E-02	6,5E-07	-		0,7373	0,290

Portanto, a combinação 2 é a que melhor representa a previsão dos pontos da VVM, visto que, aproximadamente 74%, dos pontos podem ser

representados por ela. A Equação 8 representa a equação final da combinação 2:

$$VVM = 0,8729 + 0,2155 \times VVMM + 0,0082 \times EP - 0,0022 \times IT + 0,0111 \times NDP - 0,00097 \times PT \quad \text{Eq. (8)}$$

Com a equação dos valores esperados para VVM, fez-se uma comparação entre os dados coletados (obtidos pelo INMET) e os dados

calculados (obtidos a partir da equação de regressão). Tal resultado pode ser verificado na Figura 12.

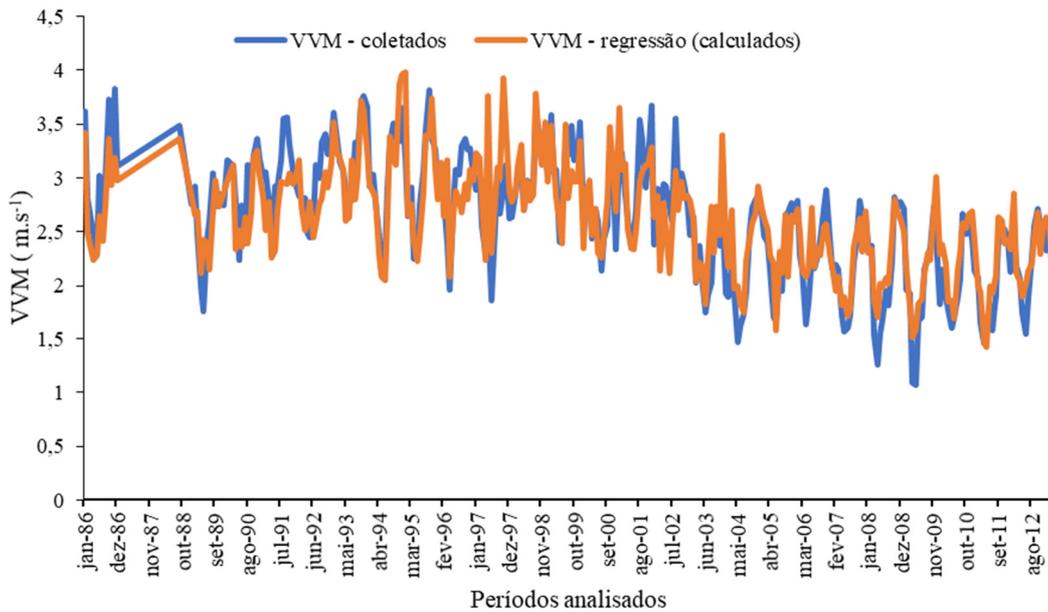


Figura 12. Comparação dos dados da velocidade média dos ventos para o bairro do Curado entre janeiro/1986 e agosto/2012: coletados x calculados.

É possível perceber a proximidade dos dados originais com os dados calculados pela equação, ratificando o fato da combinação ser representativa. Outra forma de demonstrar a boa

relação entre os dados é através do diagrama de dispersão delas (Figura 13). A boa proximidade dos pontos na reta também ratifica a equação encontrada.

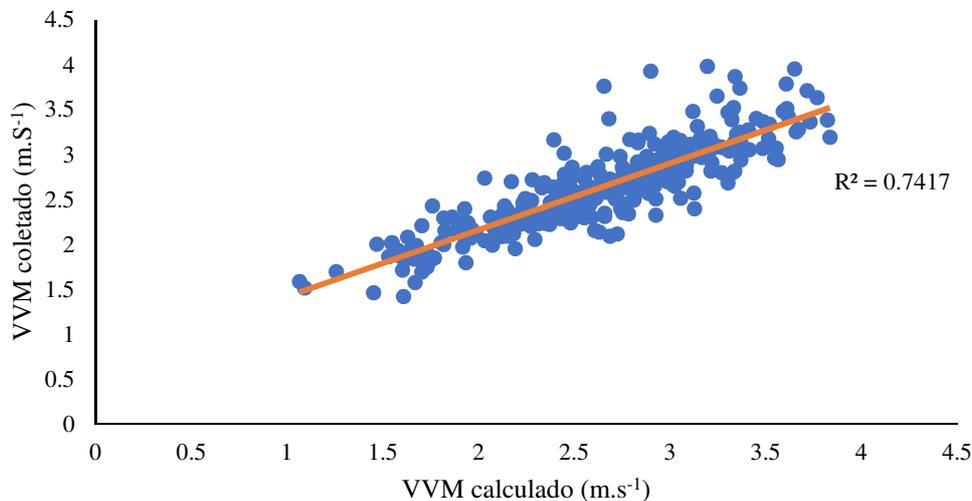


Figura 13. Dispersão entre os dados de velocidade média dos ventos calculados e coletados.

Considerando que, das cinco variáveis dessa equação, apenas duas possuíam dados durante todo o período estudado e as outras três não tinham dados exatamente no mesmo período da VVM, não foi possível preencher os dados inexistentes para a VVM. As outras variáveis que tinham sido eliminadas durante o processo seletivo para a regressão possuíam valores muito baixos de coeficiente de determinação, que não garantiriam validade para o estudo.

Discussão

Dentre os estudos encontrados, verificou-se a aderência das velocidades médias dos ventos às distribuições Gama (Beruski et al., 2009), Log-Normal (Santana et al., 2013) e Weibull (Sansigolo, 2005; Silva, 1999; Pes et al., 2005). Entretanto, dentre as distribuições analisadas, encontrou-se uma melhor aderência da velocidade média dos ventos à distribuição Weibull, sendo também possível aceitar os dados das distribuições Normal e Gama. A distribuição Log-Normal, mesmo tendo melhor aderência no estudo de Santana et al. (2013), não alcançou o valor mínimo para ser considerada aderida neste estudo.

Apesar da distribuição Weibull apresentar uma boa aproximação, é importante compreender, conforme Rajapaksha & Perera (2016), que esse modelo não é tão preciso em locais fortemente influenciados pelas monções (mudanças de ventos sazonais nas estações de seca e de chuva) por apresentarem comportamentos bastante divergentes. Para o presente estudo, tal fato não ocorre.

Considerando a análise da regressão linear baseada em outras variáveis meteorológicas, não foram encontrados resultados que pudessem ser comparados a este estudo, devido a grande diferença das variáveis analisadas. Entretanto, foram encontrados alguns estudos que relacionaram a evaporação e a evapotranspiração com a velocidade do vento, a umidade relativa, a insolação total e a precipitação total. Borges Júnior et al. (2012) utilizaram a regressão linear para relacionar as variáveis citadas para a microrregião de Garanhuns/PE e encontraram coeficientes de determinação com valores de 0,7105 nos semestres de primavera-verão e 0,8150 nos semestres de outono-inverno, resultados próximos ao 0,7417 encontrado neste estudo. Silva et al. (2005) encontraram valores máximos para o coeficiente de determinação de 0,66 ao relacionar as variáveis citadas na cidade de Petrolina/PE. Dessa forma, verificou-se que o valor de 0,7417 aqui encontrado pode ser considerado bom, se comparado aos outros dois estudos citados acima.

Conclusão

O estudo confirmou que a distribuição Weibull é bastante assertiva para a análise de dados da velocidade média do vento da amostra estudada. Além disso, também apresentou outros modelos de distribuição que podem ser aceitos para a região em questão, porém com uma menor precisão, como as distribuições Normal e Gama.

A análise de regressão dos dados coletados permitiu calcular uma equação capaz de identificar valores da velocidade média do vento utilizando amostras de cinco outras variáveis (VVMM, EP, IT, NDP e PT). A equação calculada se mostrou eficiente e viável para prever valores da velocidade média dos ventos.

A partir da análise conjunta da distribuição estatística e da equação de regressão é possível verificar a viabilidade de instalação de equipamentos para a geração de energia renovável baseada na correta previsão da velocidade média do vento.

Agradecimentos

Os autores agradecem o apoio da Escola Politécnica da Universidade de Pernambuco (UPE-POLI).

Referências

- AMARANTE, O. A. C.; BROWER, M.; ZACK, J.; SA, A. L. 2001. Atlas do Potencial Eólico Brasileiro. CRESESB, Brasília.
- ANEEL. 2008. Agência Nacional de Energia Elétrica. Ministério de Minas e Energia. Atlas de Energia Elétrica do Brasil. 3. ed. Brasília, 236p.
- BERUSKI, G. C.; LEITE, M. L.; VIRGENS FILHO, J. S.; ADACHESKI, P. A.; OLIVEIRA, A. V. 2009. Análise probabilística da velocidade média e caracterização da direção predominante do vento no município da Lapa/PR. R. RA'EGA, n. 17, p. 65-75.
- BORGES JÚNIOR, J. C. F.; ANJOS, R. J.; SILVA, T. A. J.; LIMA, J. R. S.; ANDRADE, C. L. T. 2012. Métodos de estimativa da evapotranspiração de referência diária para a microrregião de Garanhuns, PE. Revista Brasileira de Engenharia Agrícola e Ambiental, v. 16, n. 4, p. 380-390.
- BRESLOW, P. B.; SAILOR, D. J. 2002. Vulnerability of wind power resources to climate change in the continental United States. Renewable Energy Journal, v. 27, n. 4, p. 585-598.
- KISS, P.; JÁNOSI, I. M. 2008. Comprehensive empirical analysis of ERA-40 surface wind speed

distribution over Europe. *Energy Conversion and Management*, v. 49, i. 8, p. 2142-2151.

EPE. 2007. Empresa de Pesquisa Energética. Ministério de Minas e Energia. Plano nacional de energia 2030. Disponível em: http://www.epe.gov.br/PNE/20070626_1.pdf. Acesso em: junho 2017.

NAGHETTINI, M.; PINTO, E. J. A. 2007. *Hidrologia Estatística*. Belo Horizonte, CPRM-Serviço Geológico do Brasil, Superintendência Regional de Belo Horizonte. 552p.

PES, M. P. 2015. Impactos das variações e mudanças climáticas sobre os ventos extremos e seus efeitos no setor elétrico brasileiro. Tese de Doutorado do Curso de Pós-Graduação em Ciência do Sistema Terrestre, Instituto Nacional de Pesquisas Espaciais – INPE, São José dos Campos, São Paulo, Brasil. 174p.

PES, M. P.; GUEDES, M.; ROSETTO, J. G.; BIAZI, L. A.; GUARNIERI, R. A.; MARTINS, F. R.; PEREIRA, E. B.; SCHUCH, N. J. 2005. Distribuição estatística dos ventos na região de São Martinho da Serra por meio da função de Weibull: dados de agosto de 2004 a julho de 2005. *Ciência e Natura*, v. 27, p. 229-232.

PRYOR, S. C.; SCHOOF, J. T.; BARTHELMIE, R. J. 2005. Empirical downscaling of wind speed probability distributions. *Journal of Geophysical Research: Atmospheres*, v. 110, D. 19109, p. 1-12.

RAJAPAKSHA, K. W. G. D. H.; PERERA, K. 2016. Wind speed analysis and energy calculation based on mixture distributions in Narakkalliya, Sri

Lanka. *Journal of the National Science Foundation of Sri Lanka*, v. 44, i. 4, p. 409-416.

SANSIGOLO, C. A. 2005. Distribuições de probabilidade de velocidade e potência do vento. *Revista Brasileira de Meteorologia*, v. 20, n. 2, p. 207-214.

SAILOR, D. J.; HU, T.; LI, X.; ROSEN, J. N. 2000. A neural network approach to local downscaling of GCM output for assessing wind power implications of climate change. *Renewable Energy*, v. 19, p. 359-378.

SANTANA, L. V. R.; DEZOTTI, C. H.; STOSIC, T.; CHAGAS, E. S.; SILVA, H. S. 2013. Análise da distribuição da velocidade dos ventos em Piracicaba/SP. Anais do 58ª Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria e 15º Simpósio de Estatística Aplicada à Experimentação Agrônômica, Campina Grande, PB, SEAGRO.

SILVA, N. F.; ROSA, L. P.; ARAÚJO, M. R. 2005. The utilization of wind energy in the Brazilian electric sector's expansion. *Renewable and Sustainable Energy Reviews*, v. 9, p. 289-309.

SILVA, P. C.; GUEDES, V. G.; ARAÚJO, M. R. P.; HIRATA, M. H. 1999. Otimização dos parâmetros da distribuição de Weibull. Anais do XV Congresso Brasileiro de Engenharia Mecânica, Águas de Lindóia, SP, ABCM.

SILVA, V. P. R.; BELO FILHO, A. F.; SILVA, B. B.; CAMPOS, J. H. B. C. 2005. Desenvolvimento de um sistema de estimativa da evapotranspiração de referência. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v. 9, p. 547-553.