

Análise de agrupamento da velocidade do vento no Nordeste do Brasil

Lêda V. R. Santana^{1†}, Antonio S. A. Silva²

¹Instituto Federal da Paraíba

²Universidade Federal Rural de Pernambuco. E-mail: samuelpmatematico@yahoo.com.br.

Resumo: *O presente trabalho objetiva identificar padrões no comportamento da velocidade do vento para a região Nordeste do Brasil. Foram utilizadas 94 séries históricas de velocidade do vento disponibilizadas pelo INMET com registros coletados as 00 e 12h (UTC), durante o período de 1961 a 2016. Nas análises utilizou-se o método hierárquico de agrupamento, por meio de diferentes medidas de distância (Euclidiana, Maximum, Manhattan, Canberra, Minkowski) e de agrupamento (Ward.D, Single, Complete, Average, Mcquitty, Median, Centroid) e, a partir do coeficiente de correlação cofenética foi escolhida a distância de canberra como medida de dissimilaridade e a média como medida de agrupamento. Os resultados apresentaram três grupos, o grupo 1 contendo 15 (15,96%) estações com menor médias da velocidade do vento, dispostas no Sul do Maranhão, Piauí e Oeste da Bahia. O grupo 2, correspondendo a 36 (38,3%) estações distribuídas por todos os estados exceto o RN. O grupo 3, apresentando 43 (45,74%) das 94 estações concentrando-as desde o estado do CE até AL e, com algumas estações na BA, ambas possuem maior média de velocidade do vento. O agrupamento permitiu a concentração de estações com desempenhos semelhantes em regiões distintas.*

Palavras-chave: padrões; INMET; agrupamento.

Abstract: *The present work aims to identify patterns in wind speed behavior for the Northeast region of Brazil. A total of 94 historical wind speed series were used by INMET, with records collected at 00 and 12h (UTC) during the period from 1961 to 2016. The hierarchical grouping method was used by means of different distance measures (Euclidiana, Maximum, Manhattan, Canberra, Minkowski) and clustering (Ward.D, Single, Complete, Average, Mcquitty, Median, Centroid) and, from the cophenetic correlation coefficient, measure of dissimilarity and the mean as a measure of grouping. The results presented three groups, group 1 containing 15 (15.96 %) stations with lower mean wind speeds, located in the South of Maranhão, Piauí and West of Bahia. Group 2, corresponding to 36 (38.3%) stations distributed by all states except the RN. Group 3, presenting 43 (45.74%) of the 94 stations concentrating them from the state of the CE to AL and, with some stations in BA, both have the highest average wind speed. The grouping allowed the concentration of stations with similar performances in different regions.*

Keywords: patterns; INMET; clustering.

[†]Autor correspondente: valeria_araua@hotmail.com.

Introdução

Diante da capacidade de geração de energia eólica do país pesquisadores tem explorado a região NE do Brasil com a finalidade de obter informações sobre o comportamento do vento local e assim poder contribuir em estudos que viabiliza a configuração e instalações de parques eólicos (LIRA et al., 2017; CAMELO et al., 2016; CARNEIRO; CARVALHO, 2015; DE ARAUJO et al., 2015; SANTANA et al., 2015; DE OLIVEIRA; STOSIC; STOSIC, 2012; OLIVEIRA; COSTA, 2011; SILVA et al., 2004; SILVA et al., 2002).

A região Nordeste (NE) detém o maior potencial eólico do país, além de, ser favorecida pela complementaridade sazonal do regime de ventos com os fluxos hidrológicos nos reservatórios hidrelétricos, com maior velocidade do vento no período de estiagem quando os níveis dos reservatórios baixam (ABEEÓLICA, 2019; GWEC, 2017).

Segundo o relatório do The Global Wind Energy Council (GWEC), a capacidade de energia eólica mundial instalada no final do ano de 2016 foi de 486.749 Megawatts (MW). A China é o país com maior capacidade mundial, 168.690 MW. O Brasil ocupa a nona posição no grupo dos países com capacidade instalada de 10 Gigawatts, fato que o fez líder do mercado latino americano em energia eólica, além de, ser o mais promissor até 2020 (GWEC, 2017).

Carneiro e Carvalho (2015) apresentam uma avaliação estatística do potencial eólico para os municípios de Maracanaú (CE), Petrolina (PE) e Parnaíba (PI). Foi observado que os padrões de vento com melhor potencial eólico ocorreram na Parnaíba (PI) e em Petrolina (PE), verificando também que ambas tem maior regularidade no comportamento dos ventos, com destaque para Petrolina (PE).

Camelo et al. (2016) em estudo da predição para médias mensais da velocidade do vento em municípios do Nordeste brasileiro, identificou as cidades de Fortaleza (CE), Parnaíba (PI) e São Luís (MA) com maior semelhança nos registros entre máximos e mínimos da velocidade do vento quando comparados os valores preditos e observados da velocidade do vento local.

A análise de agrupamento tem bastante aplicação na área climatológica, devido a importância em identificar grupos com propriedades similares. Weber e Kaufmann (1995) propõe um novo sistema de classificação para identificar padrões do fluxo do vento utilizando o a análise agrupamento hierárquico na Suíça. Os autores encontraram que grupos distintos contém diferentes padrões do fluxo.

Carreón-Sierra et al. (2015) utilizam dados de vento da rede meteorológica da Cidade do México para reconhecer e caracterizar padrões de circulação do vento através da análise de *clusters* hierárquicos, obtendo padrões semelhantes aos relatados por outros autores.

Clifton e Lundquist (2012) demonstram a utilidade da análise de agrupamento para identificar relação entre o vento na altura das turbinas e as variações climáticas, desenvolvendo um método adequado para prever o impacto das mudanças climáticas nos recursos eólicos.

Este trabalho tem como objetivo utilizar o método de agrupamento hierárquico para identificar padrões de velocidade do vento no Nordeste do Brasil através da média da velocidade do vento de estações do INMET.

Material e Métodos

Fonte de dados

Os dados utilizados são registros de 94 estações convencionais do Instituto Nacional de Meteorologia (INMET), 1, durante o período de 01/01/1961 a 31/12/2016. A base de dados pertence ao Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP) que apoia atividades de ensino e pesquisa com registros horários as 00 e 12 (UTC), a 10 m do solo, para a variável velocidade do vento na região NE, disponíveis em www.inmet.gov.br/portal/index.php?r=estacoes/estacoesconvencionais.

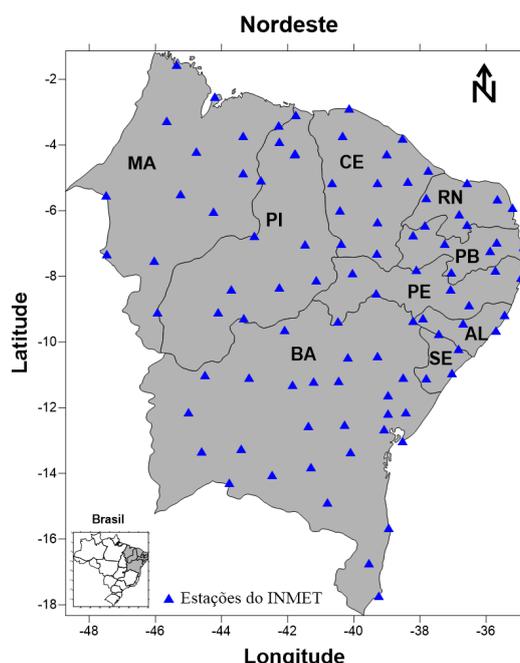


Figura 1: Estações do INMET sobre o NE do Brasil.

Análise de agrupamento (*Clustering*)

A análise de agrupamento (*Clustering*) é uma técnica estatística de classificação que visa avaliar a relação de similaridade entre as observações de um conjunto de dados formando grupos. Os grupos são formados considerando que elementos similares compõem um mesmo grupo e que grupos distintos apresentam pouca ou nenhuma similaridade entre si. Os algoritmos de agrupamentos podem ser divididos em agrupamento hierárquico, assim nomeado por apresentar grupos aninhados que formam uma árvore hierárquica ou, agrupamento particional sendo todos os *clusters* encontrados simultaneamente como uma partição dos dados. O agrupamento hierárquico ainda pode ser dividido em divisivo e aglomerativo, ou seja, dividindo composições ou juntando decomposições. Neste trabalho será abordado o método de agrupamento hierárquico aglomerativo, o qual tem várias versões e possuem diferentes critérios para construção dos grupos (KAUFMAN; ROUSSEEUW, 2005; ANDERBERG, 1973).

O agrupamento hierárquico aglomerativo tem como objetivo formar uma classificação hierárquica que usualmente é representada por um dendograma. O procedimento tem

dois passos básicos, primeiro calcula-se a matriz de similaridade para todos os pares do conjunto de dados, obtida através de uma medida de distância como as encontrada na 1. Em seguida, escolhe um método de agrupamento para determinar como os elementos serão agrupados a partir da matriz cofenética (KAUFMAN; ROUSSEEUW, 2005).

Tabela 1: Tipos de distâncias para obtenção da matriz de similaridade.

Distância	Fórmula
Euclidiana	$d_{i,j} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
Máxima	$d_{max} = \max_{dimensão\ i} x_i - y_i $
Manhattan	$d_{i,j} = \sum_{i=1}^n x_i - y_i $
Canberra	$d_{i,j} = \sum_{i=1}^n \frac{ x_i - y_i }{ x_i + y_i }$

Existem diferentes métodos de agrupamento hierárquico aglomerativo e, estes possuem diferentes critérios de decisão para construir os grupos. Alguns métodos amplamente utilizados serão apresentados a seguir (KAUFMAN; ROUSSEEUW, 2005).

Single linkage ou ligação simples, também chamado de *nearest neighbor* ou vizinho mais próximo, é o método mais antigo e mais simples, sendo definido como a mínima distância entre os pares de dados (x, y) mais próximos pertencentes a *clusters* diferentes 1.

$$d(R, Q) = \min d(x, y), \quad (1)$$

em que $x \in R, y \in Q$ e, R e Q são grupos distintos.

Complete linkage ou ligação completa define a dissimilaridade entre dois *clusters* como a maior distância entre elementos pertencentes a grupos diferentes. Portanto, este algoritmo é exatamente oposto ao *Single linkage* e frequentemente chamado de *furthest neighbor* ou vizinho mais longe.

$$d(R, Q) = \max d(x, y). \quad (2)$$

Average linkage ou ligação média, o método é baseado na média entre todos os elementos dos grupos e, os pares com menor distância são agrupados. Existem ainda, métodos baseados na média aritmética ponderada (*Weighted Pair-Group Method Using Arithmetic Averages* (WPGMA)) ou sem qualquer peso atribuído (*Unweighted Pair-Group Method Using Arithmetic Averages* (UPGMA)).

Centróide linkage ou centróide, o método é fundamentado no centróide com peso (*Weighted Pair-Group Method Using Arithmetic Averages* (WPGMC)) ou sem peso (*Unweighted Pair-Group Method Using Arithmetic Averages* (UPGMC)) atribuído. Cada grupo obtém um centróide através da média de cada grupo e, a similaridade entre os grupos é baseada na distância euclidiana 3.

$$d(R, Q) = d(\bar{R} - \bar{Q}), \quad (3)$$

em que \bar{R} e \bar{Q} são respectivamente os centróides dos agrupamentos R e Q e $d(R, Q)$ é a distância entre eles.

Ward esse tipo de agrupamento está embasado na variância dos grupos e procura minimizar a dissimilaridade associada a cada agrupamento, que é definida pela diferença da soma do erro quadrático (4) e a média da partição que está contido.

$$ESS_k = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2, \quad (4)$$

em que k é o agrupamento, n é o número de elementos do agrupamento k e x_i é o i -ésimo elemento do agrupamento k .

Para identificar qual o melhor método a ser utilizado, faz-se necessário verificar o quão diferentes são os grupos uns dos outros. Desta forma, utiliza-se o coeficiente de correlação cofenética para validar o agrupamento.

O coeficiente de correlação cofenética (CCC) foi proposto por Sokal e Rohlf (1962) com o objetivo de medir a consistência do agrupamento hierárquico através do coeficiente de correlação de Pearson entre a matriz de dissimilaridade e a matriz cofenética.

$$CCC = \frac{\hat{C}ov(F, C)}{\sqrt{\hat{V}(F) * \hat{V}(C)}}, \quad (5)$$

em que F é a matriz de dissimilaridade e, C é a matriz cofenética. Logo, para $ccc > 0.7$ o método é considerado adequado (SOKAL; ROHLF, 1962).

Com o propósito de identificar grupos com desempenho em comum da média da velocidade do vento nas diferentes coordenadas fornecidas pelo INMET, foi realizada a análise de agrupamento para as séries as 00 e 12h. A escolha do método para obter a matriz distância e cofenética foi determinado através do CCC, testando as diferentes medidas de dissimilaridade e agrupamento. Portanto, utilizou-se a medida de distância *canberra* para obter a matriz distância. Enquanto, a matriz cofenética para o agrupamento hierárquico foi obtido através do método de aglomeração média. Considerando este o agrupamento mais consistentes, por apresentar o maior coeficiente de correlação cofenética, $CCC = 0.78$, uma vez que para $CCC < 0.7$ o agrupamento é classificado impróprio (SOKAL; ROHLF, 1962).

Resultado

Na 2, observa-se o comportamento da média para a velocidade do vento as 00 (a) e 12h (b). Verifica-se que durante as 00h (2 (a)) a média da velocidade do vento é menor em relação a média das 12h, para todo o Nordeste, destacando o Oeste com média em torno de $1,5 \text{ m/s}^{-1}$. As 12h (2 (b)) as médias são maiores em relação as 00h, com locais atingindo média da velocidade do vento entre $4,5$ e $5,0 \text{ m/s}^{-1}$. Esse desempenho está relacionado as características local, estando a área Leste favorecida pela proximidade do oceano Atlântico e com poucas barreiras, as quais tem maior incidência quando os ventos avançam no sentido Oeste (AMARANTE et al., 2001). Observa-se, em uma pequena porção no Sul da Bahia que as médias também são mais intensas, essa região possui um dos maiores complexo eólico da América Latina, o Alto Sertão I (RENOVAENERGIA, 2017). Assim como, na divisa do Sul do Ceará com o Piauí, próximo a chapada do Araripe, região com condições geográficas favoráveis a exploração de energia eólica e, que detém o complexo Chapada do Piauí com 1414 MW (VENTOS, 2019).

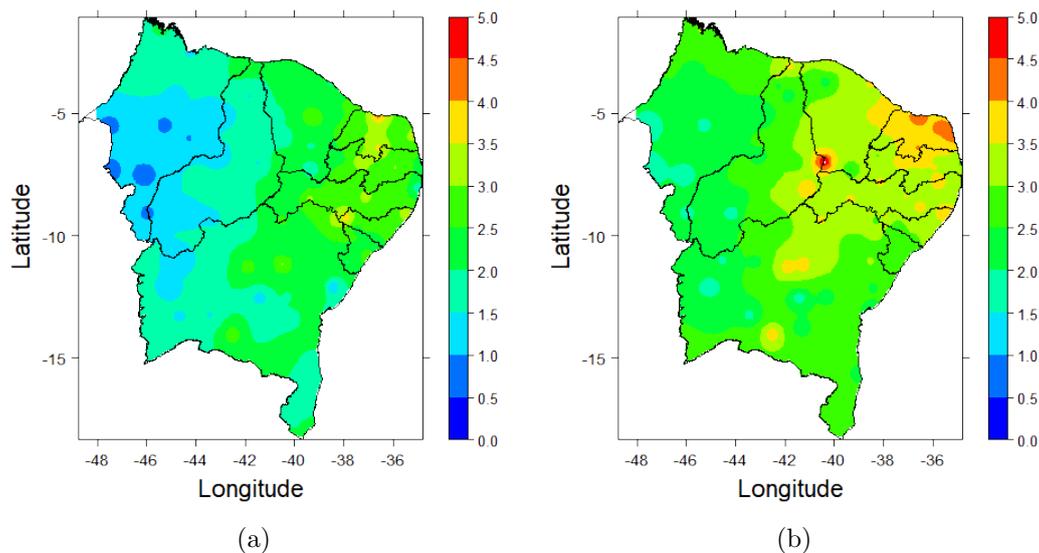


Figura 2: Média da velocidade do vento as 00h (a) e as 12h (b) para o NE brasileiro.

Na 3 (a), é mostrado o agrupamento das estações em três grupos por meio do dendograma. Verifica-se na 3 (b), a distribuição dos grupos sobre o NE. O grupo 3, tem o maior número de estações, concentrado nos estados do CE, RN, PB, PE, além de estações na BA e PI, tais localidades tem registros de maior média da velocidade com características geográficas que propiciam o aproveitamento do vento como fonte eólica. No entanto, o grupo 1 compreende as estações de baixas velocidade do vento sobre o Sul do MA e do PI e, no estado da Bahia. O grupo 2, abrange grande parte do NE exceto o estado do RN, agrupando estações em áreas que apresentam comportamento distinto e sofrem influência de diferentes sistemas como as ZCIT's e os ventos Alísios (AMARANTE et al., 2001).

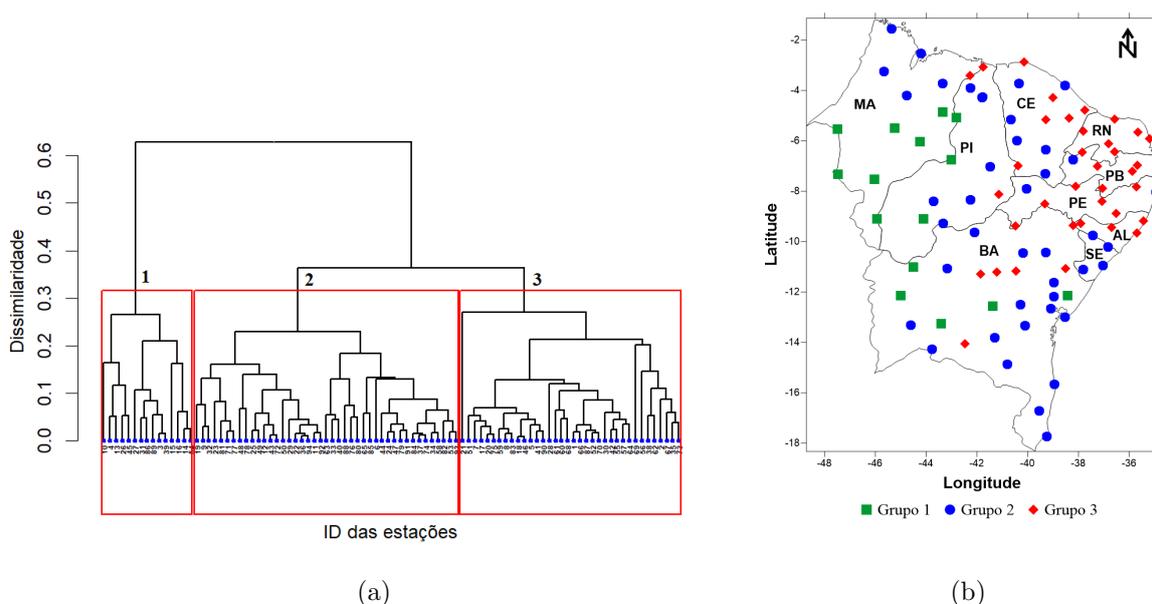


Figura 3: Agrupamento da média da velocidade do vento para as séries das 00 e 12h no NE brasileiro pertencentes ao INMET.

Na 2, estão expostos os grupos obtidos na análise de agrupamento afim de ter uma melhor visualização dos identificadores das estações apresentado na 3. O grupo 3, contém 43 estações, enquanto o grupo 2 abrange 36 estações e, o grupo 1 domina apenas 15 estações das 94 estações em estudo.

Tabela 2: Identificador das estações por grupo da análise de agrupamento

Grupos	Estações
1	3 4 10 13 14 15 16 26 27 31 39 45 54 86 89
2	6 9 11 12 19 22 23 24 25 29 32 33 34 36 37 40 43 44 47 48 49 50 53 58 63 65 71 72 74 76 77 78 79 80 81 82
3	1 2 5 7 8 17 18 20 21 28 30 35 38 41 42 46 51 52 55 56 57 59 60 61 62 64 66 67 68 69 70 73 75 83 87 90 84 85 88 91 92 93 94

Conclusão

A análise de agrupamento da velocidade do vento as 00 e 12h revela que o grupo 3 concentra estações com maior valor da velocidade do vento, as quais estão sobre regiões onde há grande exploração do vento como fonte de energia eólica. Além de mostra que o grupo 1, é responsável por concentrar as estações em locais com menor média da velocidade do vento.

Referências bibliográficas

ABEEÓLICA. Associação Brasileira de Energia Eólica. 2019. URL

<http://www.abeeolica.org.br/energia-eolica-o-setor/>.

AMARANTE, O. A.; BROWER, M.; ZACK, J.; SÁ, A. L. de. Atlas do potencial eólico brasileiro, 2001. URL

<http://www.cresesb.cepel.br/index.php?section=publicacoes&task=&cid=1>.

ANDERBERG, M. R. Cluster Analysis for Applications. First edition, first printing. Elsevier, Inc, Academic Press, 1973.

CAMELO, H. do N.; LUCIO, P. S.; GOMES, O. M.; LEAL JUNIOR, J. B. V. Predição de velocidade do vento em regiões do Nordeste brasileiro através de regressão linear e não linear para fins de geração eólica (wind speed prediction in areas of northeastern Brazil through linear regression and nonlinear for wind power). *Revista Brasileira de Geografia Física*, v. 9, n. 3, p. 927–939, 2016.

CARNEIRO, T. C.; CARVALHO, P. C. M. de. Caracterização de potencial eólico: estudo de caso para Maracanaú (CE), Petrolina (PE) e Parnaíba (PI). *Revista Brasileira de Energia Solar*, v. 6, n. 1, 2015.

CARREÓN-SIERRA, S; SALCIDO, A.; CASTRO, T.; CELADA-MURILLO, A. T. Cluster analysis of the wind events and seasonal wind circulation patterns in the mexico city region. *Atmosphere*, Multidisciplinary Digital Publishing Institute, v. 6, n. 8, p. 1006–1031, 2015.

Sigmae, Alfenas, v.8, n.2, p. 481-489, 2019.

64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).

18º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).

- CLIFTON, A.; LUNDQUIST, J. K. Data clustering reveals climate impacts on local wind phenomena. *Journal of Applied Meteorology and Climatology*, v. 51, n. 8, p. 1547–1557, 2012.
- DE ARAUJO, A. J.; STOSIC, T.; STOSIC, B.; DEZOTTI, C. H. Correlações de longo alcance em séries temporais de velocidade do vento no Nordeste. *Sigmae*, v. 2, n. 3, p. 81–84, 2015.
- DE OLIVEIRA, S. M.; STOSIC, T.; STOSIC, B. D. Long-term correlations in hourly wind speed records in Pernambuco, Brazil. *Physica A: Statistical Mechanics and its Applications*, Elsevier, v. 391, n. 4, p. 1546–1552, 2012.
- GWEC, G. W. P. C. Global Wind Energy Outlook. Brussels, 2017.
- KAUFMAN, L.; ROUSSEEUW, P. J. Finding groups in data: an introduction to cluster analysis. John Wiley & Sons, 2005.
- LIRA, M. A. T; MOITA NETO, J. M.; LOIOLA, J. V. L. de; SILVA, E. M. da; ALVES, J. M. B. Caracterização do Regime de Ventos no Piauí Para o Aproveitamento de Energia Eólica. *Revista Brasileira de Meteorologia*, scielo, v. 32, p. 77 – 88, 03, 2017. ISSN 0102-7786.
- OLIVEIRA, J.; COSTA, A. Estudo de variabilidade do vento em escala sazonal sobre o Nordeste brasileiro utilizando o rams: Os casos de 1973-1974 e 1982-1983. *Revista Brasileira de Meteorologia*, v. 26, n. 1, p. 53–66, 2011.
- RENOVAENERGIA. Companhia Brasileira de Geração de Energia Renovável. 2017. URL: <http://www.renovaenergia.com.br/pt-br/nossos-negocios/eolica/Paginas/default.aspx>.
- SANTANA, L. V. R; STOSIC, T.; DEZOTTI, C. H.; DE Albuquerque MOURA, G. B.; DE ARAÚJO, L. H. G. D.; DA SILVA, A. S. A. Spatial analyses of wind speed in the North-Brazil with data from ERA-40. *Revista Brasileira de Biometria*, v. 33, n. 3, p. 414–432, 2015.
- SILVA, B. B.; ALVES, J. J.; CAVALCANTI, E. P.; DANTAS, R. T. Potencial eólico na direção predominante do vento no Nordeste brasileiro. *Revista Brasileira de Engenharia Agrícola e Ambiental*, Departamento de Engenharia Agrícola-UFCG, 2002.
- SILVA, B. B. d.; ALVES, J. J. A.; CAVALCANTI, E. P.; Ventura, E. D. Variabilidade espacial e temporal do potencial eólico da direção predominante do vento no Nordeste do Brasil. *Revista Brasileira de Meteorologia*, v. 19, n. 2, p. 189–202, 2004.
- SOKAL, R. R.; ROHLF, F. J. The comparison of dendrograms by objective methods. *Taxon*, JSTOR, v. 11, n. 2, p. 33–40, 1962.
- VENTOS, C. dos. Casa dos Ventos. 2019. URL
- Sigmae**, Alfenas, v.8, n.2, p. 481-489, 2019.
- 64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).
18º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).

<http://casadosventos.com.br/pt/projetos/parques-eolicos>.

WEBER, R. O.; KAUFMANN, P. Automated classification scheme for wind fields. *Journal of Applied Meteorology*, v. 34, n. 5, p. 1133–1141, 1995.