

Análise multivariada para classificação da velocidade do vento no estado do Ceará

Diego A. Gomes^{1†}, Fábio S. Santos², Kerolly K. F. Nascimento³, Juliana Maria C. Araújo⁴, Denise S. A. Ferreira⁵, Kenikywayne K. F. Nascimento⁶, Guilherme R. Moreira⁷

¹ Universidade Federal Rural de Pernambuco - Deinfo/UFRPE.

² Universidade Federal Rural de Pernambuco - Deinfo/UFRPE. E-mail: fabio.sandropb@gmail.com

³ Universidade Federal Rural de Pernambuco - Deinfo/UFRPE. E-mail: kerollyfn@gmail.com

⁴ Centro Universitário - UNIFACISA. E-mail: julianamariacosta25@gmail.com

⁵ Universidade Federal Rural de Pernambuco - Deinfo/UFRPE. E-mail: denialmeida@gmail.com

⁶ Instituto Federal de Educação Ciência e Tecnologia da Paraíba - IFPB.

E-mail: keni_wane@hotmail.com

⁷ Universidade Federal Rural de Pernambuco - Deinfo/UFRPE. e-mail: guirocham@gmail.com

Resumo: *Objetivando verificar a existência de zonas com velocidades dos ventos distintas, foi utilizado a técnica multivariada de análise de agrupamento no estado do Ceará, foram utilizadas 15 estações, uma estação por município. As séries são oriundas do Instituto Nacional de Meteorologia - INMET, com dimensão de 04/11/2017 até 03/11/2018. Para verificar o melhor agrupamento foi utilizado 3 distâncias; Euclidiana, Chebyshev e Mahalanobis. Posteriormente aplicou-se os métodos aglomerativos de ligação; completa, simples e média. Gerando 9 dendrogramas, aplicando o coeficiente de correlação cofenética para determinar a melhor distância com o método aglomerativo, cujo o valor 0,82 indicou a distância de Mahalanobis com o método de ligação média. A análise de agrupamento indicou grupos com velocidade do vento distintas intra grupo e semelhantes entre grupos, consequentemente a quantidade de grupos é a quantidade de zonas com velocidades distintas. Baseando-se no índice de Ratkowsky do dendrograma da distância de Mahalanobis com o método de ligação média nessas estações existe 5 zonas no estado do Ceará.*

Palavras-chave: Energia eólica; Análise de agrupamento; Zonas.

Abstract: *Objectivating to verify the existence of a zone with suction cup variation, a multivariate cluster analysis technique was used in the state of Ceará, 15 stations were used, one station per municipality. The series come from the National Institute of Meteorology (INMET), from 04/11/2017 until 03/11/2018. To find the best grouping, 3 distances were used; Euclidian, Chebyshev and Mahalanobis. Subsequently, the binding agglomeration methods were applied; complete, simple and average. Generating 9 dendrograms, applying the coefficient of cofenética key to determine the best distance with the agglomerative method, whose value is 0,82 indicated a test distance with the method of average connection. The cluster analysis indicated groups with speed of separation of groups and groups between groups, consequently an amount of groups is an amount of zones with distinct speeds. Based on the Ratkowsky index of Mahalanobis dendrogram dendrogram with the own temperature control line of the 5 stations in the state of Ceará.*

Keywords: Wind Energy; Analysis of clustering; Areas.

[†]Autor correspondente: diegoalvesestatistica@gmail.com.

Introdução

A forma como o ser humano tem se relacionado com a natureza vem mudando nos últimos anos, e uma dessas mudanças é a percepção da necessidade de se obter novas fontes de energias, que sejam menos nocivas ao meio ambiente. É preciso garantir que o uso da energia atual não afete as gerações futuras (DO NASCIMENTO CAMELO et al., 2008).

A energia eólica vem demonstrando potencial para geração de energia sustentável e limpa, garantindo um desenvolvimento sustentável. O Brasil se destaca no mundo pelos seus recursos naturais e o Nordeste brasileiro possui características para geração de energia eólica com ventos regulares (Lopes et al., 2017; Santos et al., 2019). O estado do Ceará é um dos estados com reconhecido potencial eólico no nordeste brasileiro (Sá, 2001). Por ser um estado com capacidade de produção eólica, é preciso dispor de várias ferramentas que possam auxiliar no desenvolvimento e manutenção de pesquisas científicas. Segundo ABEólica, (2019) a produção eólica do Nordeste brasileiro é de mais de 10GWs.

A análise multivariada através da análise de agrupamento, permite identificar elementos com características semelhantes intra grupos e distintos entre grupos (JOHNSON; WICHERN, 2014). Utilizando a análise de agrupamentos em dados de velocidade dos ventos é possível verificar se existem zonas com velocidade dos ventos distintas. Após aplicação das técnicas de agrupamentos; distâncias e métodos de ligação. Os valores médios, máximos e desvios padrão da velocidade dos ventos podem ser organizada em grupos. Em que cada grupo seria uma zona com estações de características de ventos semelhantes intra grupo (SOMBRA, 2008).

O presente estudo descreve a análise de agrupamento como meio de identificar zonas com velocidade dos ventos distintas, em 15 estações dos estado do Ceará.

Material e Métodos

Os dados de velocidade do vento (m/s) utilizados neste trabalho foram obtidos junto ao Instituto Nacional de Meteorologia (INMET) no período de 04-11-2017 a 03-11-2018 em quinze estações meteorológicas automáticas situadas no estado do Ceará (Figura 1). As estações coletavam informações a cada hora no padrão brasileiro. Todas as análises foram feitas utilizando o software R (R CORE TEAM, 2018).

As variáveis utilizadas para aplicar os agrupamentos foram calculadas usando as 15 séries de observações das estações aqui estudadas; médias, desvios padrão e valor máximo como descrito na Tabela 1. As distâncias aplicadas nas três variáveis foram distâncias; Euclidiana, Chebychev e Mahalanobis. Diversas são as distâncias que podem ser utilizadas em análise de agrupamento (JOHNSON; WICHERN, 2014).

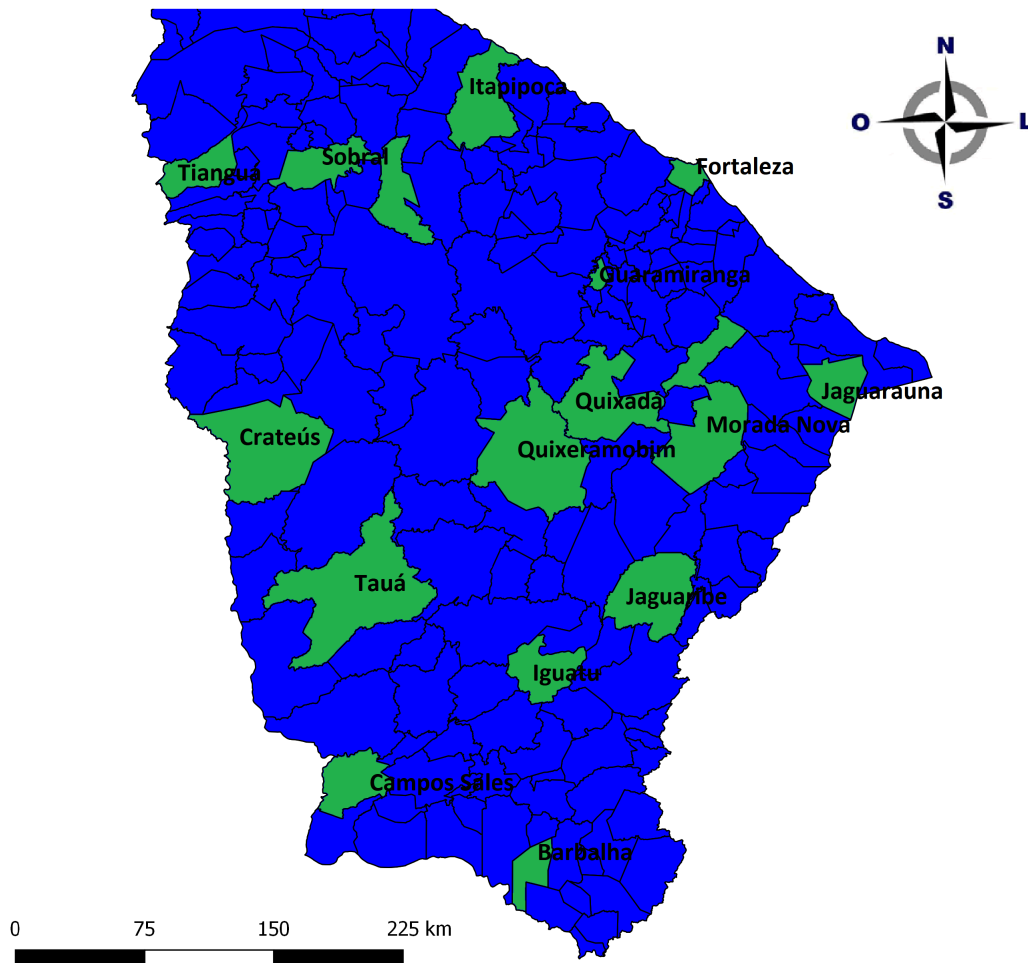
Distância Euclidiana é a distância entre dois elementos x_i e y_j é a raiz quadrada do somatório dos quadrados das diferenças entre valores de i e j para todas as variáveis ($r = 1, 2, \dots, p$), é definida por:

$$d(x_i, y_j) = \sqrt{\sum_{r=1}^p (x_{ir} - y_{jr})^2}.$$

Distância de Chebyshev é a distância entre dois elementos a_i e b_j , em que é o valor

máximo para todas as variáveis, das diferenças entre dois elementos, é definida por:

$$d(a_i, b_j) = \max |a_{iv} - b_{jv}|.$$



Fonte: Elaborada pelo Autor

Figura 1: Localização geográfica das 15 estações situada no estado do Ceará utilizada neste estudo

Distância de Mahalanobis é a distância entre dois elementos p_i e q_j , $i \neq j$, é definida por:

$$d(x_i, y_j) = (x_i - y_j)' S^{-1} (x_i - y_j).$$

Após aplicar as distâncias é organizado uma matriz com as dissimilaridades entre as estações. Na matriz com as dissimilaridades é aplicado o método aglomerativo hierárquico de ligação: completa, simples e média para obter uma estrutura com grupos denominada Dendrograma. Os métodos de ligação são definidos de acordo com Charra, (2014).

Método de ligação completa é a distância D_{ij} entre dois grupos C_i e C_j é a distância máxima entre dois pontos x e y , com $x \in C_i$ e $y \in C_j$:

$$D_{ij} = \max_{x \in C_i, y \in C_j} d(x, y).$$

Sigmae, Alfenas, v.8, n,2, p. 90-97, 2019.

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

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

Método de ligação simples é a distância D_{ij} entre dois grupos C_i e C_j é a distância mínima entre dois pontos x e y , com $x \in C_i$ e $y \in C_j$:

$$D_{ij} = \min_{x \in C_i, y \in C_j} d(x, y).$$

Método de ligação média: a distância D_{ij} entre dois grupos C_i e C_j é a média das distâncias entre o par de pontos x e y , onde $x \in C_i$ e $y \in C_j$:

$$D_{ij} = \sum_{x \in C_i, y \in C_j} \frac{d(x, y)}{n_i \times n_j}.$$

A avaliação do agrupamento com o método de ligação que melhor descrever os agrupamentos, é realizada com base no coeficiente de correlação cofenética. Para a organização dos grupos ser boa o r precisar ser maior que 0,7. O coeficiente é definido por:

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

O índice de Ratkowsky realiza um corte no dendrograma que demonstrar a quantidade de grupos ideal, este índice é definido pela seguinte expressão:

$$\text{Ratkowsky} = \frac{\bar{S}}{q^{1/2}}$$

em que

$$\begin{aligned} \bar{S}^2 &= \frac{1}{p} \sum_{j=1}^p \frac{BGSS_j}{TSS_j} \\ BGSS_j &= \sum_{k=1}^q n_k (c_{kj} - \bar{x}_j) \\ TSS_j &= \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2. \end{aligned}$$

Resultados e Discussão

É possível observa os valores médios, desvios padrão e valores máximos da velocidade do vento obtidos nas séries das 15 estações dos municípios do estado do Ceará, na Tabela 1. A maior média é 3,747 m/s observada no município de Itapipoca - CE e a menor média 0,093 m/s no município de Campos Sales - CE. O maior valor máximo 10,2 m/s foi observado no município de Jaguaribe - CE e o menor valor 4,5 m/s no município de Quixadá- CE, demonstrando variação nos níveis de velocidades dos ventos nas estações analisadas.

Para definir a melhor formação dos grupos e conseqüentemente verificar se existe zonas com variação na velocidade do vento. Foram testas as seguintes distâncias: distância Euclidiana, distância de Chebyshev e distância de Mahalanobis. Em que, foram aplicados os métodos aglomerativos hierárquicos de ligação completa, ligação simples e ligação

média. Os dendrogramas gerados das distâncias e dos métodos de ligação, estão descritos respectivamente da esquerda para direita nas Figuras 1, 2 e 3.

A aplicação das técnicas de agrupamentos em estudos de velocidade do vento vem trazendo bons resultados em outros países, na Turquia (UNAL; KINDAP; KARACA, 2003) e em Bangladesh (SIRAJ-UD-DOULAH; ISLAM, 2019).

Tabela 1: Valores médios, desvios-padrão (DP) e valor máximo da velocidade do vento nas 15 estações nos municípios do estado do Ceará

Municípios	Estações	Médias	DP.	Máx
1 Barbalha	A315	1,747	0,940	6,3
2 Campos Sales	A347	0,093	0,565	6,3
3 Crateus	A342	2,250	1,088	7,3
4 Fortaleza	A305	2,350	1,337	6,5
5 Guaramiranga	A314	3,039	1,155	7,7
6 Iguatu	A319	3,114	1,332	8,4
7 Itapipoca	A359	3,747	1,694	9,5
8 Jaguaribe	A358	2,658	1,402	10,2
9 Jaguarauna	A339	1,142	1,780	8,3
10 Morada Nova	A332	2,592	1,762	8,6
11 Quixadá	A369	1,281	0,874	4,5
12 Quixeramobim	A325	3,036	1,505	8,5
13 Sobral	A306	1,756	1,761	9,2
14 Tauá	A324	1,587	1,773	8
15 Tianguá	A368	3,018	1,439	7,6

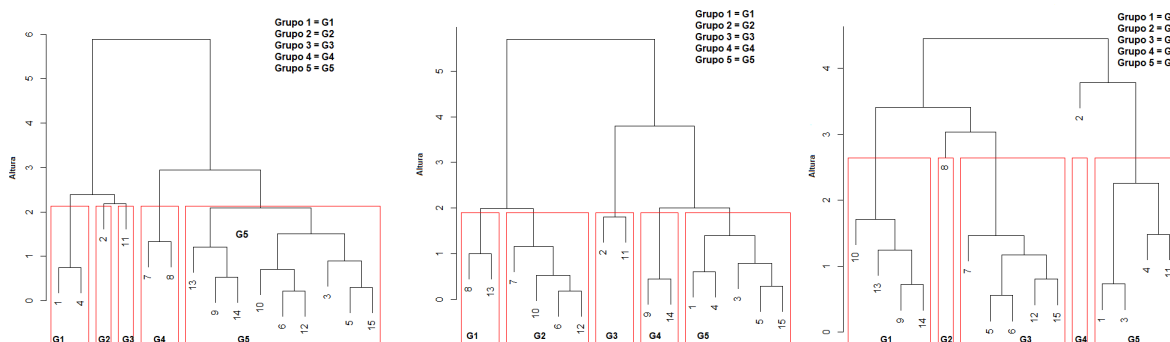


Figura 2: Dendrogramas com as distâncias Euclidiana, Chebyshev e Mahalanobis, utilizando o método de ligação completa respectivamente da esquerda para a direita

Todos os dendrogramas das Figuras 1, 2 e 3 com o índice de Ratkowsky particionam as estações em cinco grupos G1, G2, G3, G4 e G5. Indicando a existência de 5 zonas com velocidades dos ventos distintas. A análise multivariada consegue avaliar bem as diferenças entre zonas de ventos (SIRAJ-UD-DOULAH; ISLAM, 2019).

Cada grupo é formado por estações com velocidades dos ventos semelhantes intra grupos e diferentes entre grupos. A indicação dos grupos foi realizada de acordo com o

índice de Ratkowsky, o corte que o índice realiza em cada altura nos dendrogramas pode ser visto em Charra, (2014).

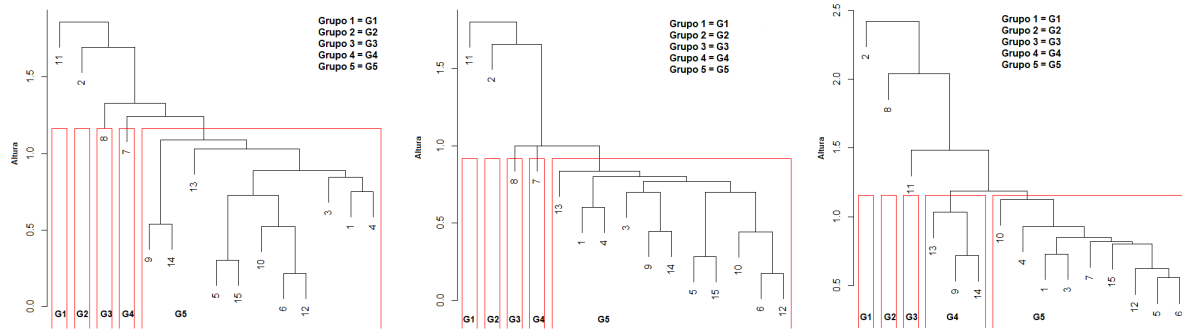


Figura 3: Dendrogramas com as distâncias Euclidiana, Chebyshev e Mahalanobis, utilizando o método de ligação simples respectivamente da esquerda para a direita

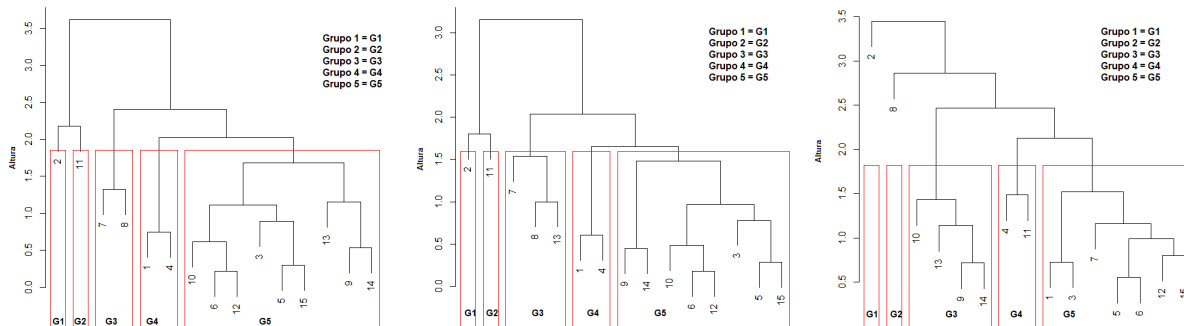


Figura 4: Dendrogramas com as distâncias Euclidiana, Chebyshev e Mahalanobis, utilizando o método de ligação média respectivamente da esquerda para a direita

Na Tabela 2 o maior coeficiente de correlação cofenética é 0,82 para a distância de Mahalanobis com o método de ligação média, indicando que a distância de Mahalanobis com o método de ligação média organizou os grupos com maior precisão. O uso do coeficiente de correlação cofenética como critério de ajuste dos grupos pode ser visto em De Oliveira, (2019).

Tabela 2: Resultados do coeficiente de correlação cofenética para os métodos de ligações completa, simples e média aplicados nas distâncias hierárquicas nos dados de 15 estações do estado do Ceará

Métodos de Ligação	Euclidiana	Chebychev	Mahalanobis
Completa	0,708	0,512	0,587
Simple	0,770	0,734	0,753
Média	0,799	0,754	0,815

Fonte: Elaborada pelo Autor

Conclusão

As distâncias multivariadas Euclidiana, Chebyshev e Mahalanobis combinadas com os métodos de ligação completa, simples e média apresentam divisões em grupos para as séries de velocidade do vento no estado do Ceará. Indicando zonas com rajadas distintas, porém utilizando o coeficiente de correlação cofenética e o índice de Ratkowsky como critério de seleção para a melhor formação dos grupos, a distância de Mahalanobis com o método de ligação média, é a combinação mais adequada. Dividindo o estado do Ceará em 5 zonas (grupos) com velocidades dos ventos distintas. A análise multivariada demonstrou ser uma eficaz técnica em estudos exploratórios de velocidade do vento no estado do Ceará.

Agradecimentos

Os autores agradecem ao Instituto Nacional de Meteorologia (INMET) pela concessão dos dados. A Universidade Federal Rural de Pernambuco (UFRPE). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Referencias Bibliográficas

- ABEEOLICA. ASSOCIAÇÃO BRASILEIRA DE ENERGIA EÓLICA. *Boletim de Dados* de Março de 2019. 2019. Disponível em: <http://abeeolica.org.br/noticias/geracao-de-energia-eolica-cresce-15-em-2018-2/i>. Acesso em: acessado em 28/03/2019.
- CHARRAD, M.; GHAZZALI, N.; BOITEAU, V.; NIKNAFS, A. Package nbclust. *Journal of statistical software*, v. 61, p. 1-36, 2014.
- DE OLIVEIRA, A. C. R.; CECON, P. R.; NASCIMENTO, M.; FINGER, F. L.; PEREIRA, G. M.; PUIATTI, G. A. Genetic divergence between pepper accessions based on quantitative fruit traits. *Científica*, v. 47, n. 1, p. 83-90, 2019.
- DO NASCIMENTO CAMELO, H.; CARVALHO, P. C.; JUNIOR, J. B. V. L.; ACCIOLY FILHO, J. B. P. Análise estatística da velocidade de vento do estado do Ceará. *Revista Tecnologia*, v. 29, n. 2, 2008.
- DOS SANTOS, F. S.; DO NASCIMENTO, K. K. F.; DE JESUS, E. S.; DA SILVA JALE, J.; STOSIC, T.; FERREIRA, T. A. E. Análise estatística da velocidade do vento em Petrolina-PE utilizando as distribuições Weibull e a Burr. *Journal of Environmental Analysis and Progress*, v. 4, n. 1, p. 057-064, 2019.
- JOHNSON, Richard; WICHERN, Dean. *Applied Multivariate Statistical Analysis*. 6. ed. Harlow: Pearson Education Limited, 2014. 770 p.
- LOPES, I.; DE CERQUEIRA JÚNIOR, E. P.; DE MELO, J. M. M.; LEAL, B. G. 2017. Potencial de geração de energia eólica no Submédio São Francisco. *Journal of Sigmae*, Alfenas, v.8, n.2, p. 90-97, 2019.
- 64ª Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS). 18º Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).

Environmental Analysis and Progress, v. 2, n. 3, p. 330-340.

SÁ, A. L. Atlas do Potencial Eólico Brasileiro. *DEWI Magazine*, no. 19, p. 63-73, 2001.

SIRAJ-UD-DOULAH, M.; ISLAM, M. N. Defining Homogenous Climate zones of Bangladesh using Cluster Analysis. *International Journal*, v. 6, n. 1, p. 119-129, 2019.

SOMBRA, S. S. Modelagem numérica em alta resolução para previsão de geração de energia eólica no Ceará. *Revista Brasileira de Meteorologia*, v. 23, n. 4, p. 477-489, 2008.

UNAL, Y.; KINDAP, T.; KARACA, M. Redefining the climate zones of Turkey using cluster analysis. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, v. 23, n. 9, p. 1045-1055, 2003.

R CORE TEAM. *R: A language and environment for statistical computing*. *R Foundation for Statistical Computing*, Vienna, Austria. 2012. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Sigmae, Alfenas, v.8, n.2, p. 90-97, 2019.

64^a Reunião da Região Brasileira da Sociedade Internacional de Biometria (RBRAS).
18^o Simpósio de Estatística Aplicada à Experimentação Agronômica (SEAGRO).