

## Aplicação da análise exploratória de dados na discriminação geográfica do quiabo do Rio Grande do Norte e Pernambuco

F. S. Panero<sup>3\*</sup>, M. F. P. Vieira<sup>1</sup>, Â. M. F. Cruz<sup>1</sup>, M. F. V. Moura<sup>2</sup> e H. E. B. da Silva<sup>3</sup>

<sup>1</sup>Departamento de Química – Universidade Federal do Rio Grande do Norte, UFRN, CEP: 59072-970 Lagoa Seca – RN.

<sup>2</sup>Departamento de Farmácia – Universidade Federal do Rio Grande do Norte, UFRN, CEP: 59010-180 Petrópolis – RN.

<sup>3</sup>Departamento de Química – Universidade Federal de Roraima, UFRR, CEP 69304000 Boa Vista – RR.

\*fspaneroit@yahoo.com.br

**Resumo:** Foram analisadas amostras de quiabo dos municípios de Caruaru e Vitória de Santo Antão, em Pernambuco, assim como nos municípios de Ceará-Mirim, Macaíba e Extremoz no estado do Rio Grande do Norte. A aplicação de dois métodos de análise exploratória de dados: Análise de Componentes principais - PCA e Análise de Agrupamentos Hierárquicos - HCA permitiu a discriminação geográfica do quiabo proveniente dos dois estados.

**Palavras-chave:** Análise exploratória de dados; Quimiometria; PCA; HCA; quiabo.

### INTRODUÇÃO

O quiabo é um fruto proveniente do quiabeiro (*Hibiscus esculentus*) que é uma planta hortícola de origem africana pertencente à família das Malváceas[1]. Possui alto valor nutritivo, medicinal e comercial, tem um ciclo vegetativo rápido, de fácil cultivo e de alta rentabilidade<sup>3</sup>. É bastante encontrado no nordeste brasileiro, em especial no Rio Grande do Norte que tem se destacado como um dos grandes produtores. Além de ser usado na alimentação como um ingrediente importante em diversos pratos da cozinha brasileira, também é utilizado na medicina, atuando como laxante (refrescando o intestino), nos casos de pneumonia, bronquites e tuberculose pulmonar.

Por se tratar de um alimento de grande consumo, faz-se necessário estudar a composição do fruto do quiabeiro cultivado em solo potiguar e pernambucano (fontes para o comércio na CEASA e supermercados de Natal), para compor tabelas de macro e micronutrientes, de modo a auxiliar na

aplicação de dietas adequadas, principalmente nos casos de deficiência nutricional[2-4].

Os crescentes avanços e sofisticações das técnicas instrumentais de análise química, proporcionam a obtenção de um maior volume de resultados, havendo assim a necessidade de utilização de técnicas de tratamentos de dados mais complexas do ponto de vista matemático[5]. A análise multivariada foi introduzida há algumas décadas no tratamento de dados químicos, dando origem a uma nova disciplina denominada quimiometria, uma importante ferramenta permitindo a extração de uma quantidade de informação que dificilmente seria gerada empregando métodos univariados[6].

Procedimentos para controle de qualidade ou discriminação de origem geográfica de uvas, amostras de mel, vinho, água mineral, água superficial, água subterrânea e bebidas não-alcólicas, empregando métodos estatísticos multivariados têm sido propostos na literatura ao longo da última década[7-10]. Diferentes estratégias podem

ser utilizadas para se obter a informação desejada em experimentos dessa natureza.

### Análise multivariada

Os métodos estatísticos multivariados consideram as amostras e as variáveis em seu conjunto, permitindo extrair informações complementares que a análise univariada não consegue evidenciar. Um dos objetivos na utilização da análise multivariada é reduzir a representação dimensional dos dados, organizando-os em uma estrutura que facilita a visualização de todo o conjunto de dados[11].

As duas técnicas de análise multivariada mais conhecidas são as análises de agrupamento hierárquico (HCA) e a análise de componentes principais (PCA). A PCA e a HCA são metodologias exploratórias que visam evidenciar similaridades ou diferenças entre amostras em um determinado conjunto de dados. Desse modo, torna-se perfeitamente aplicáveis em experimentos de rastreabilidade[12].

A análise de componentes principais (PCA) é um método de análise multivariada utilizado para projetar dados n-dimensionais em um espaço de menor dimensão, ou seja as informações contidas no espaço de dimensão n são comprimidas por combinações lineares das variáveis originais a um espaço geralmente de ordem 2 ou 3 [13]. A PCA é um método exploratório porque auxilia na elaboração de hipóteses gerais a partir dos dados coletados. É também capaz de separar a informação importante da redundante e aleatória. Em uma análise de componentes principais, o agrupamento das amostras define a estrutura dos dados pela construção gráficos de *scores* e *loadings*, cujos eixos são componentes principais (PCs) nos quais os dados são projetados[14-15]. Os *scores* fornecem a composição das PCs em relação as amostras, enquanto os *loadings* fornecem essa mesma composição em relação às variáveis. Como as PCs são ortogonais, é possível examinar as relações entre amostras e variáveis com os gráficos dos *scores* e dos *loadings*[16-17]. O estudo conjunto de *scores* e *loadings* permite ainda estimar a influência de cada variável em cada amostra[18].

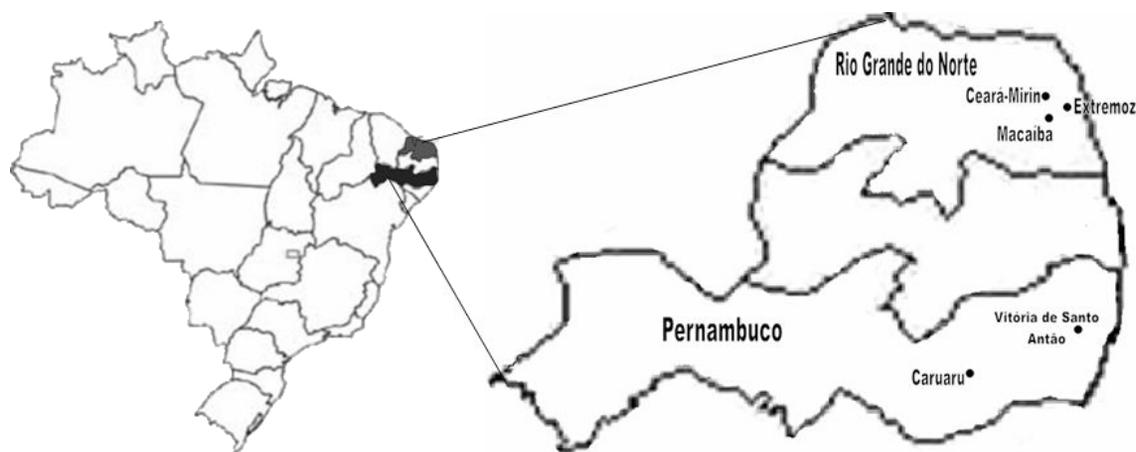
A análise de agrupamento hierárquico (HCA) é um processo hierárquico no qual, em cada passo a matriz de dados é diminuída em uma dimensão, pela reunião de pares semelhantes, até a reunião de todos os pontos em um único grupo[15,16]. O objetivo da HCA é exibir os dados em um espaço bidimensional de maneira a enfatizar os seus agrupamentos e padrões naturais. A distância entre os pontos (amostras ou variáveis) reflete a similaridade de suas propriedades, sendo útil para determinar a semelhança entre amostras. O método relaciona as amostras de forma que as mais semelhantes são agrupadas entre si. Os resultados são apresentados na forma de um dendrograma que agrupa amostras ou variáveis em função da similaridade[17,18]. Neste diagrama, a escala varia de zero (amostras sem similaridades) a um (amostras similares)[14].

Neste trabalho as técnicas de análises multivariadas (HCA e PCA) foram aplicadas ao conjunto de dados formados por variáveis medidas por análises de conteúdo de minerais no quiabo, visando avaliar a constituição nutricional deste fruto, e a discriminação geográfica do quiabo provenientes dos Estados de Pernambuco e Rio Grande do Norte.

## MATERIAL E MÉTODOS

### Amostras

Foram coletadas amostras de quiabo oriundas de cinco diferentes municípios dos Estados de Pernambuco e Rio Grande do Norte (**Figura 1**): X-PE (Chã Grande - município de Caruaru) e VSA-PE (município de Vitória de Santo Antão) no Estado de Pernambuco; C-RN (Coqueiro - município de Ceará-Mirim), M-RN (município de Macaíba) e E-RN (Estivas - município de Extremoz) no estado do Rio Grande do Norte.



**Figura 1** – Localização das amostras, municípios dos Estados de Pernambuco e Rio Grande do Norte.

### Preparação das amostras

As amostras (frutos verdes de quiabo verde, cerca de 2,0 kg) foram previamente limpas, com papel toalha, cortadas em rodelas e distribuídas para a determinação de umidade, cinzas e minerais. Foram desidratadas a 105 °C e após resfriamento, maceradas e acondicionadas na forma de pó em frascos de vidro.

### Análise de espécies metálicas

A quantificação dos minerais foi obtida a partir das cinzas do quiabo dissolvidas em HCl 10% (v/v). A determinação dos metais (Cu, Zn, Na, Fe, K, Ca, Mn, Mg) foi realizada em Espectrofotômetro de Absorção Atômica Varian, modelo SpectrAA 110 (\*), com lâmpada de cátodo oco, calibrado em condições específicas de comprimento de onda, fenda e mistura de gases para cada elemento. O fosfato foi determinado por Espectroscopia de Absorção através com equipamento Varian, modelo Cary 1E (\*), utilizando-se o método do molibdovanadato. As concentrações de cloreto e sulfato foram determinadas pelos Métodos de Möhr e Turbidimétrico, respectivamente. Para a quantificação de cada mineral foram construídas curvas de calibração, com exceção do cloreto. Os padrões e reagentes utilizados foram da marca

Merck com certificado de análise rastreado com padrões NIST.

(\*) Central Analítica – Depto de Química – UFRN – RN

### Análise estatística

Para efetuar a análise de componentes principais (PCA) e a análise de agrupamento hierárquico (HCA) foi utilizado o software EINSIGHT<sup>19</sup>. As análises químicas foram realizadas em triplicata, para a construção da matriz de dados 15x11 (15 amostras e onze variáveis).

Os dados foram previamente auto-escalados, antes de serem submetidos a análise de componentes principais e a análise de agrupamento hierárquico uma vez que há uma grande variação de respostas das diversas variáveis, ou seja, diferem em ordem de grandeza, atribuindo-se assim um mesmo peso para todas as variáveis[20-21]. Para a obtenção do dendrograma do HCA foram utilizadas a distância euclidiana e o método de conexão incremental.

## RESULTADOS E DISCUSSÃO

A **Tabela 1** apresenta o conjunto de dados experimentais utilizado para realizar as análises

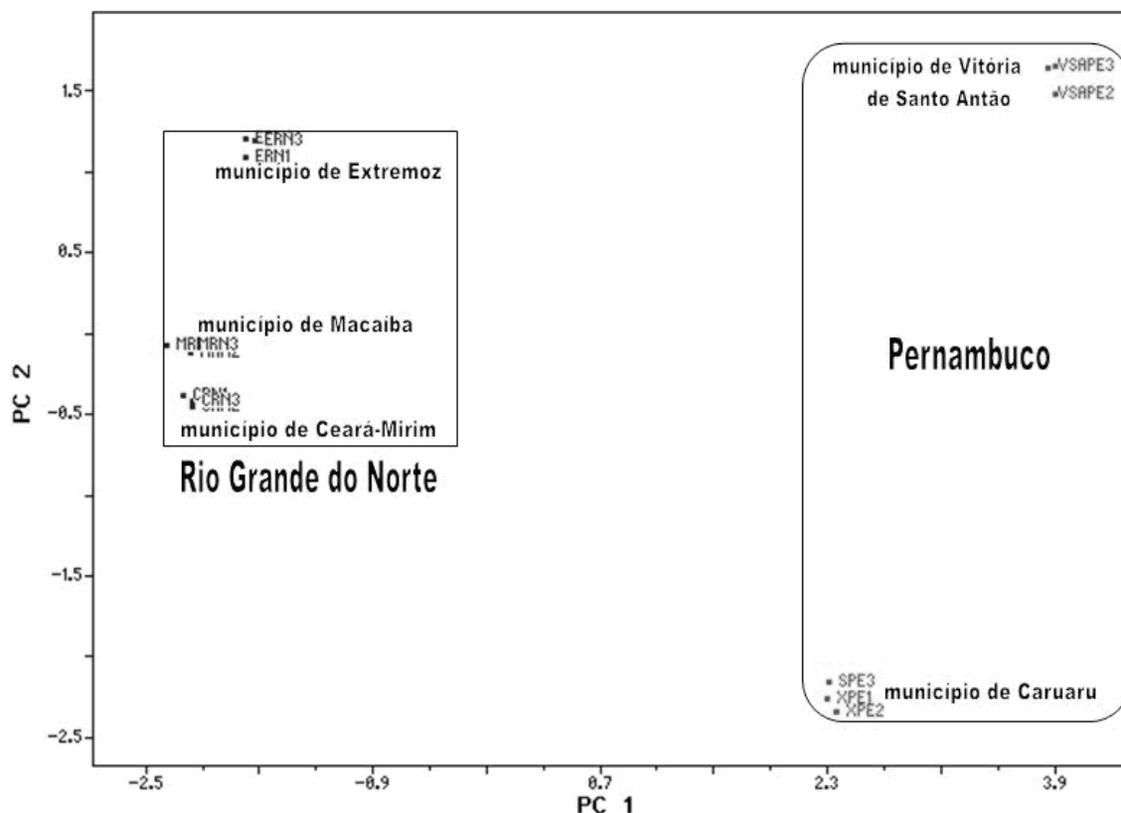
exploratórias PCA e HCA. As técnicas de análise multivariada foram utilizadas visando resumir em poucas e importantes dimensões a maior parte da variabilidade da matriz para a obtenção de relações entre os dados obtidos, extraindo desta maneira a informação química mais relevante.

**Tabela 1-** Determinação de espécies metálicas do quiabo no resíduo seco (RS), (mg de espécie metálica/100g de amostra)

Sample	Cu	Zn	Na	Fe	K	Ca	Mn	Mg	PO4	SO4	Cl
XPE1	1,02	3,69	5,98	3,50	2888,17	977,49	5,48	676,59	114,07	489,38	12,80
XPE2	1,01	3,71	5,86	3,54	2927,02	970,27	5,46	677,36	115,17	492,13	12,60
XPE3	1,03	3,68	6,10	3,52	2812,28	973,52	5,48	676,82	114,72	494,33	12,85
VSAPÉ1	1,35	3,72	1,50	4,34	2053,79	1041,03	6,80	935,33	92,46	497,88	12,04
VSAPÉ2	1,33	3,78	1,41	4,35	2078,32	1043,19	6,81	938,57	94,44	503,38	12,12
VSAPÉ3	1,37	3,69	1,58	4,34	2041,27	1038,02	6,83	937,73	93,56	501,78	12,09
CRN1	0,98	2,48	11,81	3,23	1813,61	620,67	4,10	300,13	93,50	471,11	37,63
CRN2	0,97	2,50	12,23	3,26	1826,43	624,86	4,12	302,17	94,55	477,61	37,75
CRN3	0,99	2,52	11,95	3,21	1802,29	621,66	4,14	303,37	94,12	476,96	37,69
MRN1	0,84	2,18	9,87	3,37	1935,17	843,71	3,31	506,76	85,49	384,51	34,41
MRN2	0,86	2,15	9,02	3,40	1942,36	852,55	3,32	507,19	87,76	391,16	34,57
MRN3	0,87	2,20	9,36	3,38	1938,85	848,47	3,34	506,57	87,12	388,11	34,48
ERN1	1,10	2,65	18,01	3,84	1466,31	778,35	4,67	364,81	102,46	404,86	39,90
ERN2	1,12	2,63	17,82	3,84	1470,29	785,90	4,66	366,22	100,83	403,84	40,12
ERN3	1,12	2,66	17,23	3,85	1468,78	780,94	4,69	365,72	101,23	404,23	40,07

### Análise de componentes principais (PCA)

Verifica-se que as componentes PC1 e PC2 descrevem 83,27% da variação total dos dados e fornecem informações discriminatórias das amostras. A primeira componente principal (PC1) descreve 65,42% da variação total e a segunda (PC2) 17,85%. A figura 2 apresenta o gráfico de *scores* e observam-se dois agrupamentos.

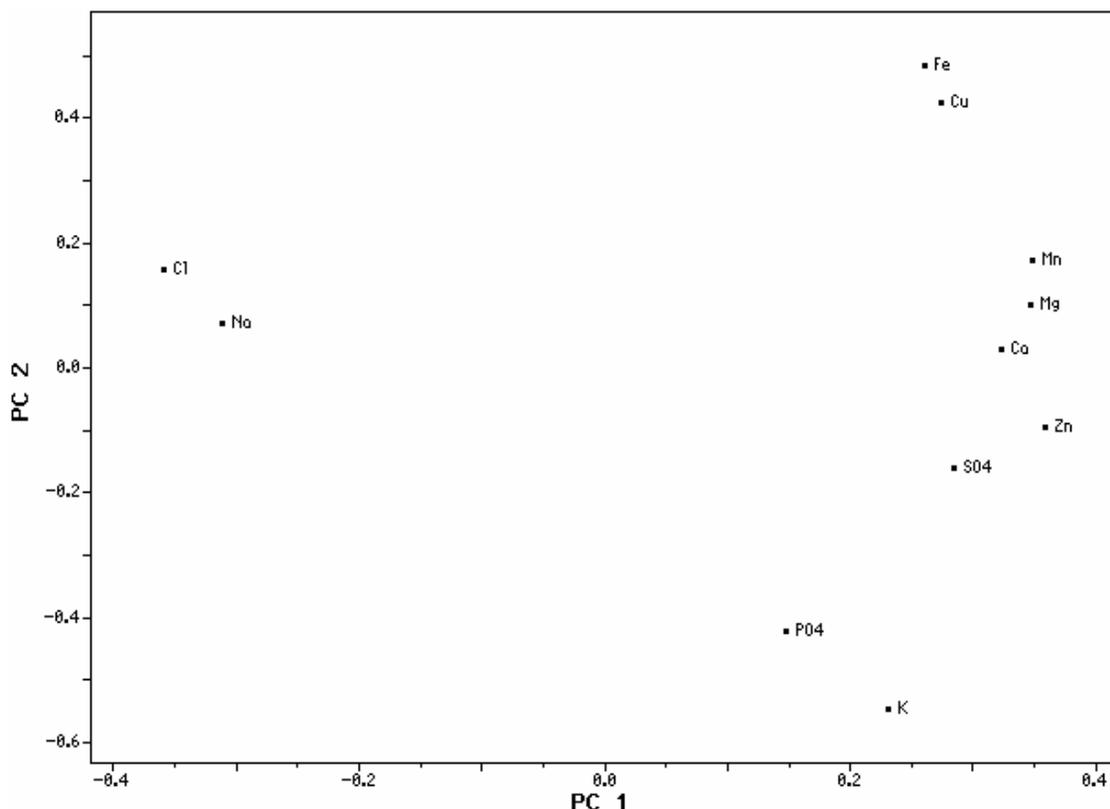


**Figura 2** – Gráfico dos scores, PC2 vs PC1, autoescalamento, var. tot. 83,27%, discriminação geográfica dos quiabos provenientes dos estados de Pernambuco e Rio Grande do Norte.

Analisando a PC1 é possível observar a discriminação de dois grandes grupos, ou seja, as amostras provenientes do Estado de Pernambuco: XPE (Chã Grande - município de Caruaru) e VSAPE (município de Vitória de Santo Antão) as amostras provenientes do Estado do Rio Grande do Norte: CRN (Coqueiro - município de Ceará-Mirim), MRN (município de Macaíba) e ERN (Estivas - município de Extremoz).

O gráfico dos *loadings* na **Figura 3** mostra a influência das variáveis sobre as amostras. O comportamento distinto apresentado pelas amostras CRN, MRN e ERN, ou seja, a discriminação geográfica destas amostras em relação as amostras do Estado de Pernambuco, foi causada pelos teores de Cl(cloreto), Na(sódio), pois as amostras de quiabo provenientes do Estado do Rio Grande do Norte apresentam maiores teores de Cl e Na em relação as demais amostras. Já as amostras

de quiabo provenientes do Estado de Pernambuco XPE e VSAPE apresentam os maiores teores de Fe, Cu, Mn, Mg, Ca, Zn, K,  $PO_4^{3-}$  e  $SO_4^{2-}$ .

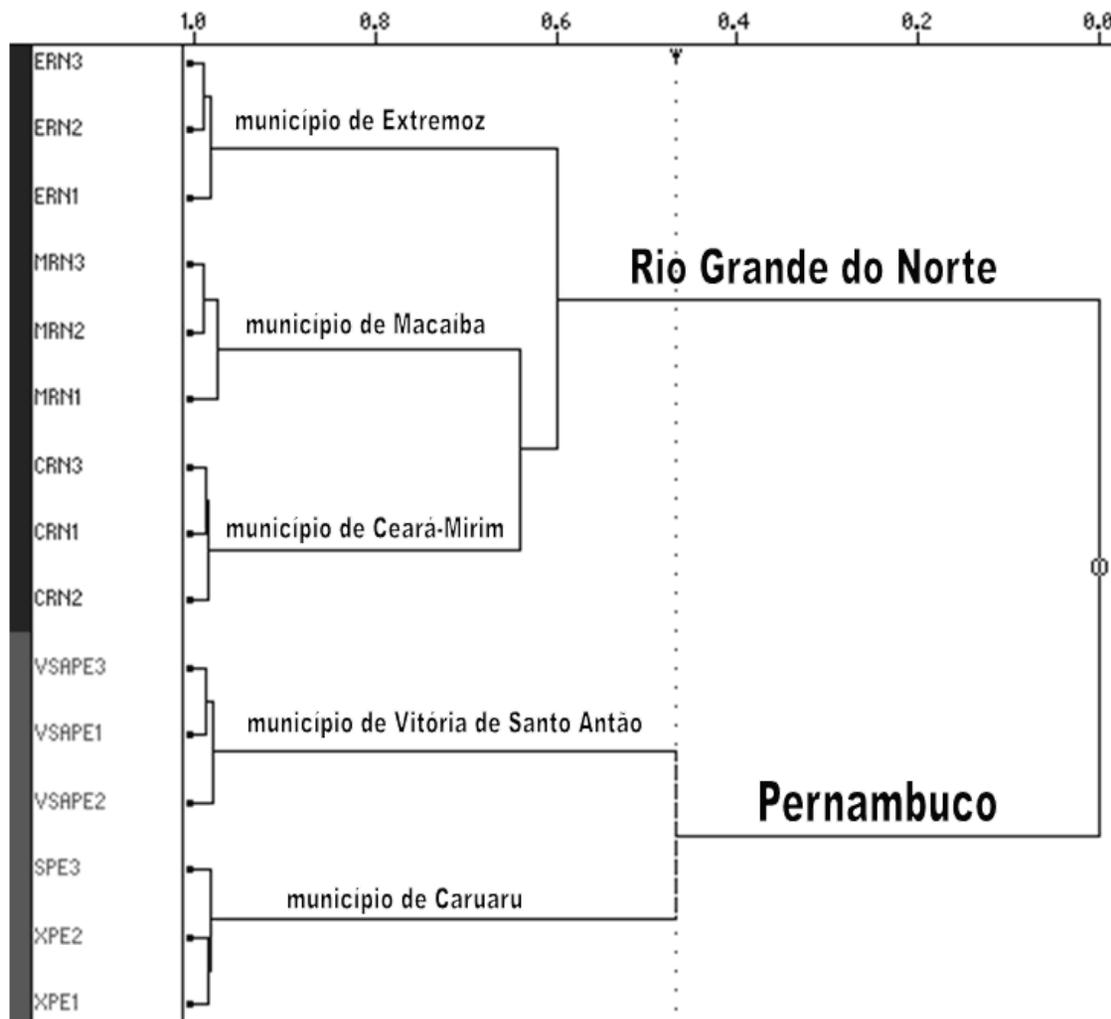


**Figura 3** – Gráfico dos loadings, PC2 vs PC1, autoescalamento, var. tot. 83,27%, influência das variáveis na discriminação geográfica dos quiabos provenientes dos estados de Pernambuco e Rio Grande do Norte.

Analisando a PC2 na figura 2, é possível observar um ligeiro deslocamento da amostra ERN das demais amostras provenientes do Rio Grande do Norte. Isso ocorreu devido aos maiores teores de Fe e Cu em relação aos municípios do estado RN. Dentro do grupo formado pelos municípios de Pernambuco também pode ser observado um deslocamento entre as amostras: XPE e VSAPÉ. Isso ocorreu pois as amostras de quiabo provenientes do XPE possuem maiores teores de  $PO_4^{3-}$  e K, e as amostras de quiabo de VSAPÉ possuem os maiores teores de Fe e Cu do estado de PE.

#### Análise de Agrupamentos Hierárquicos (HCA)

As tendências observadas através da análise de componente principais (PCA) foram confirmadas através do dendrograma obtido pela HCA (**Figura 4**), ou seja, é possível observar a formação de dois grandes agrupamentos.



**Figura 4** – Dendrograma das amostras (HCA), autoescalamto, distância Euclidiana, técnica de conexão incremental, discriminação geográfica dos quiabos provenientes dos estados de Pernambuco e Rio Grande do Norte.

As amostras provenientes do estado do Rio Grande do Norte: CRN (Coqueiro - município de Ceará-Mirim), MRN (município de Macaíba) e ERN (Estivas - município de Extremoz), formaram um agrupamento. Já as amostras provenientes do Estado de Pernambuco: XPE (Chã Grande - município de Caruaru) e VSAPE (município de Vitória de Santo Antão), localizam-se no dendrograma separadas das amostras do RN, formando outro agrupamento.

## CONCLUSÕES

As amostras foram agrupadas de acordo com o esperado, mostrando que a análise exploratória de dados permite a obtenção de informações rápidas e eficientes sobre a similaridade entre as amostras pela visualização gráfica. A PCA identificou as amostras provenientes de PE e RN, e a HCA confirmou os resultados produzidos pela PCA, discriminando geograficamente as amostras de quiabo. A aplicação da análise exploratória de dados (HCA e PCA) na discriminação geográfica

de alimentos é hoje relativamente simples, e torna possível que analistas de alimentos utilizem estes métodos multivariados corretamente, para a construção de modelos de previsão ou classificação, com o objetivo de prever a origem de amostras alimentares.

Application of the exploratory analysis of data in the geographical discrimination of okra from Rio Grande do Norte and Pernambuco

Received November 03 2008

Accepted 04 2008

**Abstract:** Samples of okra from Caruaru and Vitória of Santo Antão, in the State of Pernambuco, and Ceará-Mirim, Macaíba and Extremoz in the State of Rio Grande do Norte have been analysed. Two different methods were applied in the data treatment allowing to geographically discriminate samples from different origins: Principal Component Analysis - PCA and Hierarchical Cluster Analysis - HCA.

**Keywords:** exploratory analysis; chemometrics; PCA; HCA; okra

## REFERÊNCIAS

- [1] A. Balbach, D. S. F. Boarim, *As Hortaliças na Medicina Natural*, Missionária, São Paulo, 1ª ed., 1992.
- [2] F. S. Almeida, *Guia de Herbicidas: recomendações para o uso adequado em plantio direto e convencional*, IAPAR, Londrina, 1985.
- [3] L. S. Camargo, *As Hortaliças e seu Cultivo*. revista e atualizada, Fundação Cargill, Campinas, 3ª ed., 1992.
- [4] E. Malavolta, *Manual de Química Agrícola*. Agronômica Ceres, São Paulo, 1976.
- [5] R. E. Bruns, J. F.G. Faigle, *Quim. Nova* 8 (1985) 84.
- [6] K. R. Beebe, B. R. Kowalski, *Anal. Chem.* 59 (1987) 1007A.
- [7] F. V. Silva, M. Y. Kamogawa, M. M. C. Ferreira, J. A. Nóbrega, A. R. A. Nogueira, *Eclat. Quím.*, 27 (2002) 91.
- [8] F. S. Panero, H. E. B. Silva, 1th International Meeting on Chemometrics and Multivariate Analysis Applied to Cultural Heritage and Environment, Nemi-Rome, Italy, 2006.
- [9] H. E. B. da Silva, F. S. Panero, M. A. S. Israel, 10th International Conference on Chemometrics in Analytical Chemistry, Águas de Lindóia, Brazil, 2006.
- [10] F. S. Panero, H. E. B. Silva, J. S. Panero, 7th International Symposium on Environmental Geochemistry, Beijing, China, 2006.
- [11] M. C. S. Moura, A. N. C. Lopes, G. C. Moita, J. M. M. Neto, *Quim. Nova* 29(3) (2006) 429.
- [12] R. A. de Sousa, W. B. Neto, R. J. Poppi, N. Baccam, S. Cadore, *Quim. Nova*, 29(4) (2006) 654.
- [13] E. C. Ferreira, S. H. B. G. Rodrigues, M. M. C. Ferreira, J. A. Nóbrega, A. R. A. Nogueira, *Eclat. Quím.* 27 (2002) 77.
- [14] F. S. Panero, H. E. B. da Silva, *Microchemical Journal* 88(2) (2008) 194.
- [15] P. R. M. Correia, M. M. C. Ferreira, *Quim. Nova* 30(2) (2007) 481.
- [16] K. R. Beebe, R. J. Pell, M. B. Seasholtz, *Chemometrics: a practical guide*, John Wiley & Sons: New York, 1997.
- [17] M. A. Sharaf, D. L. Illman, B. R. Kowalski, *Chemometrics*, John Wiley & Sons, New York, 1986.
- [18] F. S. Panero, J. S. Panero, H. E. B. Silva, *Chinese J. of Geochemistry*, 25(Suppl.) (2006) 165.
- [19] EIN\*SIGHT 3.0, Infometrix, Seattle, USA, 1991
- [20] A. P. de Oliveira, J. A. G. Neto, M. M. C. Ferreira, *Eclat. Quím.* 31(1) (2006) 7.
- [21] M. A. Morgano, S. C. N. Queiroz, M. M. C. Ferreira, *Braz. J. Food Technol.*, 2 (1999) 73.