

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

Francisco dos Santos Panero^{a*}, Maria de Fátima Pereira Vieira^b, Ângela Maria Fagundes da Cruz^b, Maria de Fátima Vitória Moura^c

^aDepartamento de Química – Universidade Federal de Roraima, UFRR, CEP 69304-000 Boa Vista – RR, Brasil.

^bDepartamento de Química – Universidade Federal do Rio Grande do Norte, UFRN, CEP: 59072-970 Lagoa Seca – RN, Brasil.

^cDepartamento de Farmácia – Universidade Federal do Rio Grande do Norte, UFRN, CEP: 59010-180 Petrópolis – RN, Brasil.

Recebido em 25/07/2009; revisado em 12/09/2009; aceito em 03/10/2009. Publicado na internet em 07/11/2009

RESUMO: Os teores de Cu, Zn, Na, Fe, K, Ca, Mn, Mg, PO_4^{3-} , Cl⁻ e SO_4^{2-} foram determinados nas 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. O objetivo deste trabalho é a aplicação de dois métodos de análise exploratória de dados: Análise de Componentes principais - PCA e Análise de Agrupamentos Hierárquicos - HCA na discriminação geográfica do quiabo proveniente dos estados do Rio Grande do Norte e Pernambuco. Os resultados mostraram que Cl⁻ e Na foram os principais elementos para a diferenciação das amostras do Estado do Rio Grande do Norte e as amostras de Pernambuco apresentaram os maiores teores de Fe, Cu, Mn, Mg, Ca, Zn, K, PO_4^{3-} e SO_4^{2-} . Os dois métodos de análise exploratória de dados: PCA e HCA são eficientes para discriminação geográfica do quiabo proveniente do Rio Grande do Norte e Pernambuco.

Palavras-chave: análise exploratória de dados, quimiometria, PCA, HCA, quiabo

* Autor de correspondência. E-mail: fspaneroit@yahoo.com.br

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 [3]. 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 ser um alimento de grande consumo, é necessário estudar a composição do fruto do quiabeiro cultivado em solo potiguar e pernambucano, 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 sofisticções 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]. O reconhecimento de padrões não supervisionado foi introduzido há algumas décadas no tratamento de dados químicos, dando origem a uma nova disciplina denominada quimiometria, uma importante ferramenta que permite a extração de informação útil e relevante dos dados químicos [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-alcoólicas, empregando métodos de reconhecimento de padrões não supervisionados 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.

Reconhecimento de Padrões Não Supervisionado

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 do reconhecimento de padrões não supervisionado 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 reconhecimento de padrões não supervisionado 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, tornam-se perfeitamente aplicáveis em experimentos de rastreabilidade [12].

A análise de componentes principais (PCA) é um método de reconhecimento de padrão não supervisionado utilizado para projetar dados n-dimensionais em um espaço de baixa dimensão, ou seja, as informações contidas no espaço de dimensão n são comprimidas através de 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 através de 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 às 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 através dos gráficos dos "scores" e dos "loadings" [16-17]. O estudo do conjunto de "scores" e "loadings" permite ainda estimar a influência de cada variável em cada amostra [10].

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, na qual agrupa amostras ou variáveis em função da similaridade [10, 17]. Neste diagrama, a escala varia de zero (amostras sem similaridades) a um (amostras similares) [14].

Morgano e colaboradores realizaram a diferenciação de diversos tipos de vegetais (feijão, lentilha, soja, batata, cenoura, aipo, brócolis, couve-flor, chicória, alface, grão de bico, repolho, beterraba, batata doce e inhame) através de PCA e HCA utilizando a composição de macro e micronutrientes e minerais (Na, K, Ca, Mg, P e Fe) [20].

Neste trabalho as técnicas de reconhecimento de padrões não supervisionados (HCA e PCA) foram aplicadas ao conjunto de dados formados por variáveis medidas por amostras de minerais feitas no quiabo, visando avaliar a constituição nutricional do quiabo, e a discriminação geográfica dos quiabos 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), VSA-PE (município de Vitória de Santo Antão), ambas no Estado de Pernambuco; no Estado do Rio Grande do Norte, teve-se, C-RN (Coqueiro - município de Ceará-Mirim), M-RN (município de Macaíba) e E-RN (Estivas - município de Extremoz).

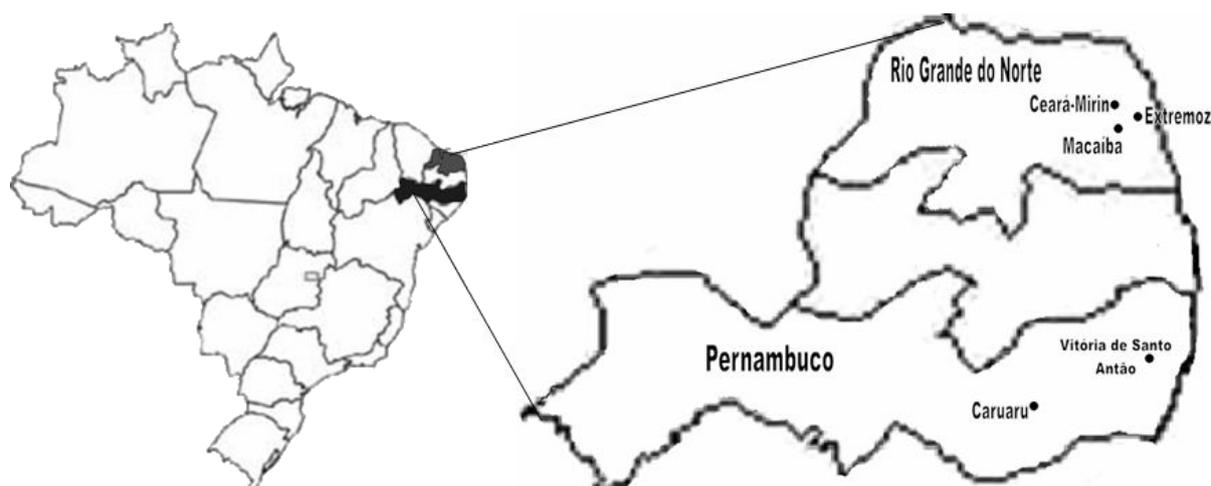


Figura 1. Localização das amostras, municípios dos Estados de Pernambuco e Rio Grande do Norte.

Preparação das amostras

As amostras (frutos verdes de quiabo verde, cerca de 2,0 kg) foram previamente limpas com papel toalha para retirada de resíduos não desejáveis (areia, restos de folhas etc.). Em seguida, as amostras foram cortadas em rodela e distribuídas para a determinação de umidade, cinzas e minerais, sendo desidratadas a 105 °C e depois de frias maceradas e acondicionadas na forma de pó, em frascos de vidro previamente limpos, para a execução das determinações de macro e micronutrientes.

Análise de metais e espécies aniônicas

A quantificação dos minerais foi obtida a partir das cinzas do quiabo, a qual foi dissolvida em HCl 10% (v/v). As determinações dos metais (Cu, Zn, Na, Fe, K, Ca, Mn, Mg) foram feitas usando-se um 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 Molecular através de um equipamento Varian, modelo Cary 1E (*), utilizando-se o método do molibdovanadato. O cloreto e o

sulfato foram analisados 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 – Departamento 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 [18]. As determinações químicas foram realizadas em triplicata para cada amostra. Utilizou-se estas triplicatas para a construção da matriz de dados (15 x 11), 15 amostras e onze variáveis foram utilizadas na análise exploratória de dados (PCA e HCA).

Os dados foram previamente auto-escalados antes de serem submetidos à 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 [19, 20]. Para obtenção do dendrograma do HCA foram utilizadas a distância euclidiana e o método de conexão incremental.

Resultados e Discussão

A determinação de minerais do quiabo no resíduo seco (RS), proveniente dos estados de Pernambuco e Rio Grande do Norte é apresentada na Tabela 1.

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

Amostra	Cu	Zn	Na	Fe	K	Ca	Mn	Mg	PO ₄	SO ₄	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

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.

Análise de componentes principais (PCA)

Realizando a análise de componentes principais (PCA) foi verificado que as componentes PC1 e PC2 descrevem 83,27% da variação total dos dados e fornecem informações discriminatórias das amostras. Sendo que a primeira componente principal (PC1) descreve 65,42% da variação total, 17,85% da variação total dos dados é descrita pela segunda componente principal (PC2). Através do gráfico de "scores" (Figura 2) observa-se a formação de dois agrupamentos.

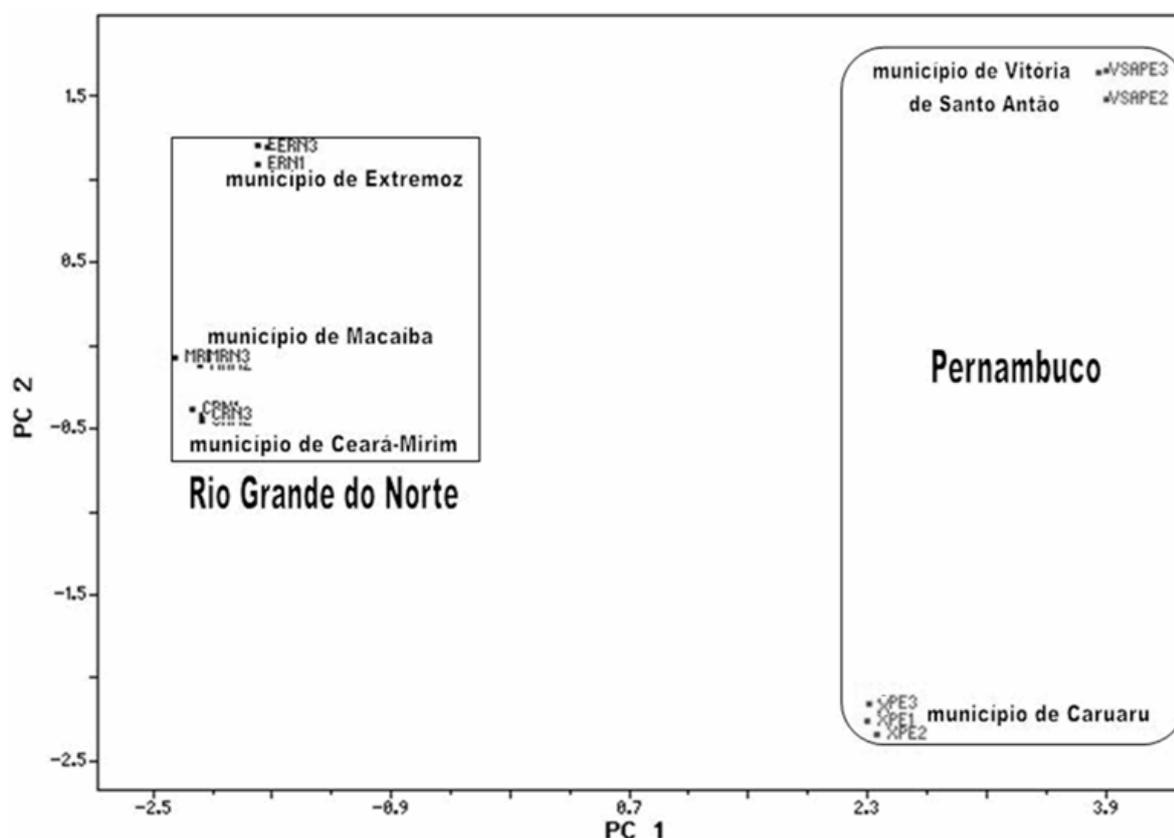


Figura 2. Gráfico dos "scores", PC2 vs PC1, autoescalado, var. tot. 83,27%, discriminação geográfica dos quiabos provenientes dos estados de Pernambuco e Rio Grande do Norte.

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). Através do gráfico dos

"loadings" (Figura 3) é possível observar a influência das variáveis sobre as amostras, gráfico dos "scores". 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, PO₄ e SO₄.

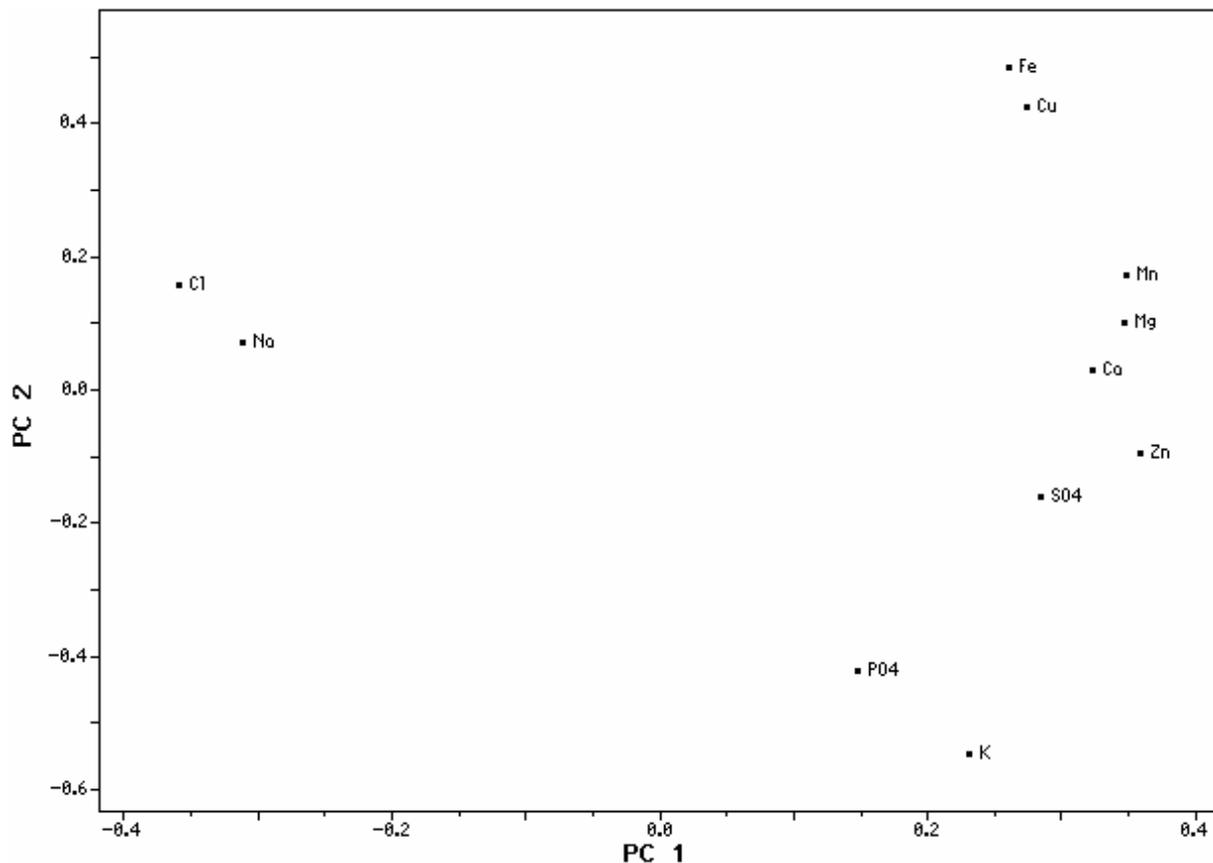


Figura 3. Gráfico dos "loadings", PC2 vs PC1, autoescalamamento, var. tot. 83,27%, influência das variáveis na discriminação geográfica dos quiabos provenientes dos estados de Pernambuco e Rio Grande do Norte.

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 do Rio Grande do Norte. Dentro do grupo formado pelos municípios de Pernambuco também pode ser observado um deslocamento entre as amostras: XPE e VSAPE. Isso ocorreu, pois as amostras de quiabo provenientes do XPE possuem maiores teores de PO₄ e K, e as amostras de quiabo de VSAPE possuem maiores teores de Fe e Cu do estado de PE.

Análise de Agrupamentos Hierárquicos (HCA)

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

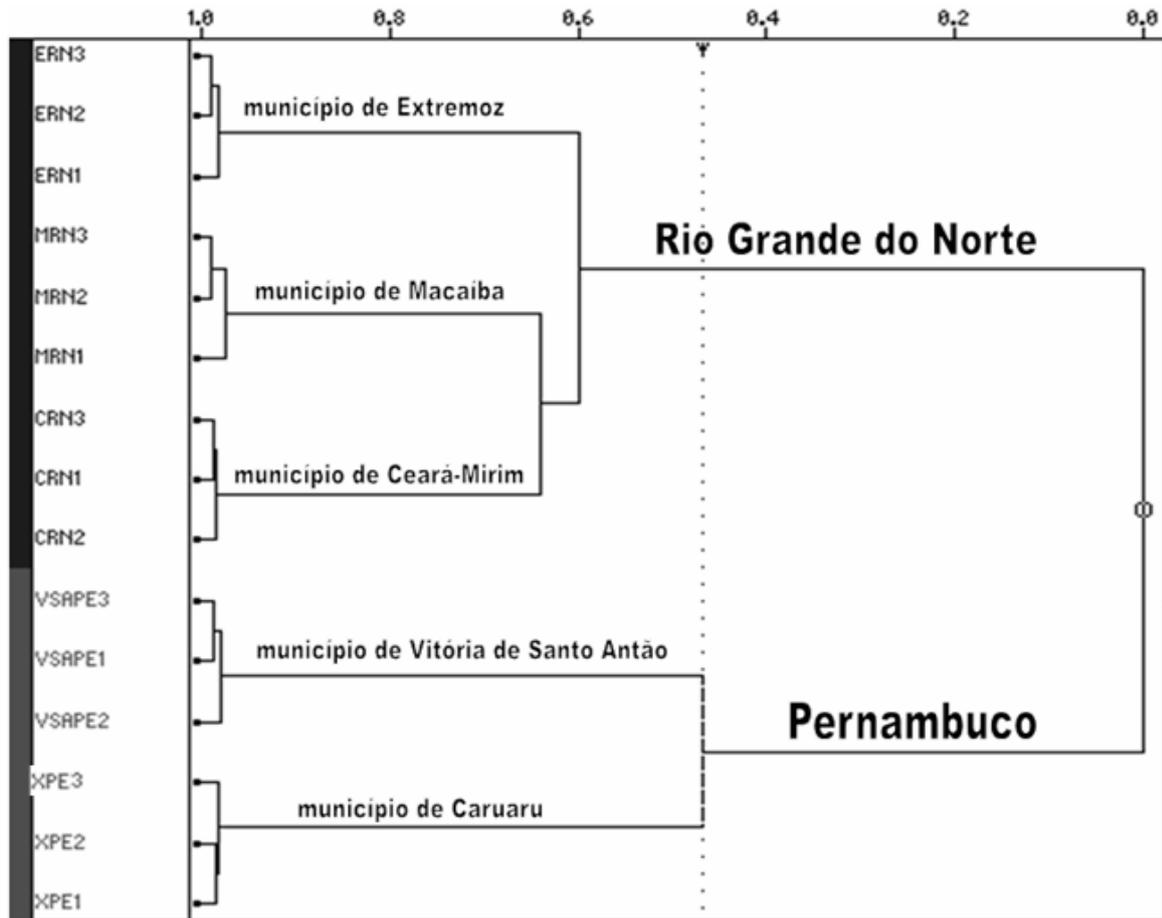


Figura 4 – Dendrograma das amostras (HCA), autoescalamento, distância Euclidiana, técnica de conexão incremental, discriminação geográfica dos queijos provenientes dos estados de Pernambuco e Rio Grande do Norte.

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 VSAP (município de Vitória de Santo Antão) localizam-se no dendrograma separadas das amostras do Rio Grande do Norte, formando outro agrupamento.

Conclusões

As amostras foram agrupadas de acordo com o esperado, mostrando que a

análise exploratória de dados permite que se obtenham informações rápidas e eficientes sobre a similaridade entre as amostras através de visualização gráfica. A PCA identificou as amostras provenientes de Pernambuco e Rio Grande do Norte, e a HCA confirmou os resultados produzidos pela PCA, discriminando geograficamente as amostras de quiabo. Com isso foi possível demonstrar a aplicação da análise exploratória de dados (HCA e PCA) na discriminação geográfica de alimentos. Desse modo, como o avanço tecnológico dos programas computacionais nesta área tornou-os disponíveis e a utilização deles é relativamente simples, é possível que os analistas de alimentos utilizem os métodos multivariados corretamente, para a construção de modelos de previsão ou classificação, com o objetivo de prever a origem de amostras alimentares.

ABSTRACT

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

The contents of Cu, Zn, Na, Fe, K, Ca, Mn, Mg, PO₄³⁻, Cl⁻ and SO₄²⁻ were determined in samples of okra of the municipal districts of Caruaru and Vitória de Santo Antão, in Pernambuco, as well as in the municipal districts of Ceará-Mirim, Macaíba and Extremoz in the state of Rio Grande do Norte. The objective of this work is the application of two methods of exploratory analysis of data: Principal Component Analysis - PCA and Hierarquical Cluster Analysis - HCA in the geographical discrimination of okra originating in the states of Rio Grande do Norte and Pernambuco. The results showed that Cl⁻ and Na were the main elements for the differentiation of the samples of Rio Grande do Norte and, the samples of Pernambuco presented the largest amount of Fe, Cu, Mn, Mg, Ca, Zn, K, PO₄³⁻, and SO₄²⁻. Boths the methods of exploratory analysis of data investigated are efficient for geographical discrimination of okra originating in Rio Grande do Norte and Pernambuco.

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

Referências Bibliográficas

- [1] Balbach, A.; Boarim, D. S. F. *As Hortaliças na Medicina Natural*, 1ª ed. São Paulo: Missionária, 1992.

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