

## CLASSIFICAÇÃO DE REFRIGERANTES ATRAVÉS DE ANÁLISE DE IMAGENS E ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

Mariana da S. Godinho, Raquel O. Pereira, Keysi de O. Ribeiro, Fernando Schmidt e Anselmo E. de Oliveira\*

Instituto de Química, Universidade Federal de Goiás, CP 131, 74001-970 Goiânia – GO, Brasil

Sérgio B. de Oliveira

Centro Federal de Educação Tecnológica de Goiás, Rua 75, 46, 74055-110 Goiânia – GO, Brasil

Recebido em 29/11/07; aceito em 7/3/08; publicado na web em 26/8/08

CARBONATED SOFT DRINK CLASSIFICATION BASED ON IMAGE ANALYSIS AND PCA. This paper describes an approach for the colour-based classification of RGB (red-green-blue) images, acquired using a common scanner, of commercial carbonated soft drinks. Mean histograms of image colour channels were evaluated for the PCA classification of 29 brands of Guaraná, Cola, and orange flavors. Loadings for principal component axes resulted in different patterns for sample grouping on score plots according to RGB histograms. pH, sorbic acid and sucrose measurements were also correlated to the analyzed brands through PCA score plots of the digitalized images.

Keywords: PCA; image analysis; carbonated soft drinks.

### INTRODUÇÃO

O Brasil é o terceiro maior produtor de refrigerantes do mundo, ficando atrás dos Estados Unidos e do México. Existem cerca de 750 fábricas e são produzidos, por ano, cerca de 11 bilhões de L pelas grandes indústrias, conhecidas internacionalmente, e também pelas fábricas regionais, que nos últimos anos tiveram um crescimento elevado em todo o País.<sup>1</sup> O estado de Goiás conta atualmente com 21 indústrias de refrigerantes que desenvolvem um comércio que vem crescendo sensivelmente e produziu, em 2002, cerca de 275 milhões de L.<sup>2</sup>

Vários fatores ambientais inter-relacionados determinam a susceptibilidade dos refrigerantes à contaminação e deterioração, daí a necessidade em se monitorar a qualidade desses produtos através de análises físico-químicas e pela sua cor. As análises físico-químicas são realizadas por técnicas analíticas convencionais, e a cor pode ser avaliada pela imagem digitalizada em um *scanner* e manipulada em um computador. As imagens formadas nas telas dos computadores, normalmente, utilizam o sistema RGB para a definição de cores. Nesse sistema, cada tom de cor é definido por três canais: R (vermelho), G (verde) e B (azul), que variam como índices inteiros entre 0 e 255, permitindo uma combinação de  $256^3$  (= 16.777.216) tonalidades em cada pixel. Ou seja, nesse formato, um tom de cor corresponde a um ponto em um espaço tridimensional formado pelos eixos R, G e B (Figura 1).

O uso de imagens, nos dias de hoje, está muito difundido, pois está associado ao desenvolvimento de diversos setores como aplicações industriais, comunicações e pesquisas.<sup>3</sup> Na Química, no entanto, esse uso ainda é pequeno, mas suas aplicações vêm crescendo nos últimos anos, como na determinação fotométrica de compostos em matrizes de papéis<sup>4</sup> e por cromatografia em camada delgada,<sup>5</sup> na determinação do teor de saponina em quinoa,<sup>6</sup> no estudo do fluxo de L-Glutamato em cérebro de ratos,<sup>7</sup> cromatografia gasosa bi-dimensional,<sup>8</sup> alimentos,<sup>9</sup> entre outras.

O presente trabalho teve por finalidade classificar os refrigerantes dos tipos Cola, Laranja e Guaraná, atualmente comercializados no estado de Goiás, utilizando análise de componentes principais (PCA)

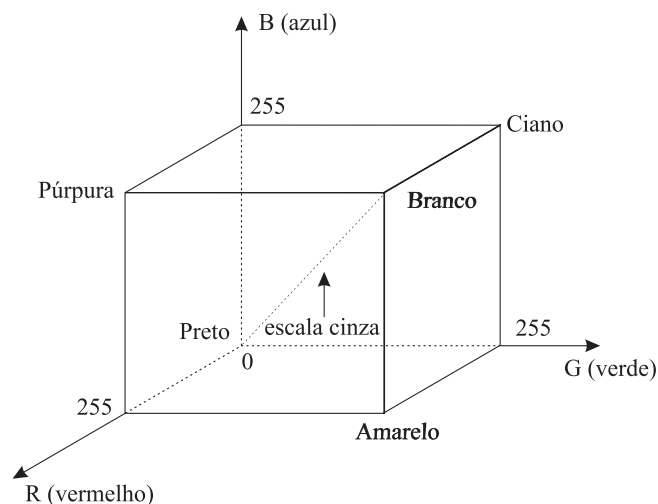


Figura 1. Cubo de cores RGB (R: vermelho; G: verde; B: azul). Cada eixo varia de 0 a 255 índices de cor. Nas extremidades do cubo estão as cores características

das imagens dos refrigerantes, obtidas em um *scanner* comum. Ao mesmo tempo, resultados de análises físico-químicas, para algumas marcas, são correlacionados com os resultados obtidos da PCA. A finalidade do presente trabalho decorre da proposição de um método simples de caracterização de amostras de refrigerantes com base na cor, como alternativa às falsificações dos refrigerantes.

### PARTE EXPERIMENTAL

Foram coletadas amostras de 29 marcas de refrigerantes comercializados em Goiás. Para cada marca foram adquiridas 5 amostras conforme as embalagens disponíveis, como latas de 350 mL e garrafas PET de 250, 500 mL e 2 L. As amostras de cada marca pertenciam ao mesmo lote, estando todas dentro do prazo de validade. Esses dados, além do local de fabricação, para todas as amostras analisadas constam da Tabela 1.

\*e-mail: elcana@quimica.ufg.br

As marcas numeradas de 1 a 5, para todas as três categorias, representam o mesmo nome comercial, diferindo apenas no sabor: cola, laranja ou guaraná. Para as demais, o nome comercial já é um indicativo do sabor. As amostras dos refrigerantes foram coletadas de forma aleatória em bares, distribuidoras e supermercados, levando-se em conta a variedade disponível, não sendo coletadas/analizadas amostras dos tipos *diet* ou *light*. Todas as amostras foram coletadas e suas imagens digitalizadas logo em seguida, não sendo submetidas a nenhum tipo de *stress*.

Resultados de análises físico-químicas<sup>10</sup> de pH, viscosidade relativa (VR), índice de refração relativo (IR) e os teores, em mg/100 mL, de ácido benzóico, ácido sórbico, sacarose e acidez total para algumas das marcas que tiveram suas imagens digitalizadas constam da Tabela 2. Essas análises físico-químicas foram feitas em amostras de lotes diferentes aos lotes que constam na Tabela 1, e a numeração das marcas utilizada na referência 10 é diferente da que é apresentada nas Tabelas 1 e 2.

### Análise das imagens

Foram analisadas 11 marcas de refrigerantes do tipo Cola, 8 do tipo Laranja e 10 do tipo Guaraná. Para cada marca, 5 amostras foram adquiridas, e 3 alíquotas para cada amostra foram coletadas, sendo que cada alíquota de 15,00 mL da amostra desgasificada,

em ultra-som, foi adicionada em uma placa de Petri. Essa imagem foi registrada por um *scanner* de mesa Genius (modelo ColorPage Vivid, 1200XE). O arquivo correspondente foi salvo no formato TIF e outras duas varreduras foram realizadas para a mesma amostra. Desse modo, 45 imagens foram digitalizadas para cada marca (3 varreduras no *scanner* x 3 alíquotas x 5 amostras). As imagens obtidas pelas varreduras foram lidas no programa Scilab,<sup>11</sup> de distribuição gratuita, com o *toolbox* SIP<sup>12</sup> (*Scilab Image Processing*), também distribuído gratuitamente,<sup>13</sup> sendo obtidos os respectivos histogramas para cada canal de cor (R, G e B). Esses histogramas foram decompostos em vetores linhas R, G e B, onde cada vetor tem 256 componentes (índices/tons de cor). Um novo vetor foi então obtido pela justaposição desses três vetores, gerando um perfil RGB para cada imagem, com 768 variáveis quantificadas, por imagem (Material Suplementar, Figura 1S), para 3 amostras do tipo Cola. Nesse caso, cada amostra consiste de uma média dos valores de 45 imagens.

A matriz de dados, **X**, para cada categoria/tipo de refrigerante tem a dimensão do produto do número de marcas analisadas pelo número de replicatas de imagens, por 768 colunas (índices de cor dos canais R, G e B). Desse modo, tem-se para as categorias Cola, Guaraná e Laranja matrizes 495 x 768, 450 x 768 e 360 x 768, respectivamente,

**Tabela 1.** Especificação das amostras dos refrigerantes adquiridos para as três categorias analisadas: Cola, Guaraná e Laranja. A coluna “Local” representa a cidade de fabricação, e a coluna “Embalagem” contém o tipo da embalagem, bem como seu volume

Categoria	Marca	Local	Validade	Lote	Embalagem
Guaraná	1	Goiânia-GO	15/09/2007	28	PET – 2 L
	2	Goiânia-GO	01/08/2007	07:09	PET – 2 L
	3	Trindade-GO	15/08/2007	16:25	PET – 250 mL
	4	Rio Verde-GO	22/07/2007	24	PET – 2 L
	5	Alexânia-GO	16/06/2007	421709	Lata – 350 mL
	6	Trindade-GO	02/09/2007	16:04	PET – 2 L
	7	Anápolis-GO	28/10/2007	14:13	Lata – 350 mL
	8	Trindade-GO	18/06/2007	P201206	Lata – 350 mL
	9	Leme-SP	25/08/2007	NAO47	Lata – 350 mL
	10	Trindade-GO	15/05/2007	P060307	PET – 2 L
Cola	1	Goiânia-GO	15/09/2007	27	PET – 2 L
	2	Goiânia-GO	10/07/2007	08:19	PET – 2 L
	3	Trindade-GO	03/08/2007	09:51	PET – 250 mL
	4	Rio Verde-GO	09/08/2007	42	PET – 2 L
	5	Alexânia-GO	19/06/2007	443120	PET – 250 mL
	6	Trindade-GO	14/09/2007	14:26	PET – 500 mL
	7	Trindade-GO	28/08/2007	P010307	Lata – 350 mL
	8	Trindade-GO	01/06/2007	05:23	PET – 2 L
	9	Anápolis-GO	30/10/2007	07:43	Lata – 350 mL
	10	Trindade-GO	18/07/2007	21:48	PET – 2 L
	11	Brasília-DF	04/05/2007	338	PET – 2 L
Laranja	1	Goiânia-GO	14/09/2007	24	PET – 2 L
	2	Goiânia-GO	31/08/2007	14:35	PET – 2 L
	3	Trindade-GO	05/08/2007	15:32	PET – 250 mL
	4	Rio Verde-GO	03/09/2007	70	PET – 2 L
	5	Alexânia-GO	20/06/2007	443121	PET – 250 mL
	6	Trindade-GO	10/09/2007	P140307	Lata – 350 mL
	7	Trindade-GO	15/08/2007	16:45	PET – 500 mL
	8	Anápolis-GO	17/11/2007	12:16	Lata – 350 mL

**Tabela 2.** Resultado das análises físico-químicas<sup>10</sup> para algumas marcas de refrigerantes

Categoria	Marca	pH	AB <sup>a</sup>	AS <sup>b</sup>	AT <sup>c</sup>	S <sup>d</sup>	VR <sup>e</sup>	IR <sup>f</sup>	
Guaraná	1	3,15	16,530	0,550	40,150	8,360	1,256	1,0099	
	2	3,84	39,125	0,150	75,560	6,745	1,160	1,0073	
	5	3,17	26,770	ND	106,880	12,670	1,266	1,0107	
	6	3,24	24,340	3,410	115,380	11,751	1,248	1,0086	
	7	3,28	39,310	3,920	99,840	8,230	1,241	1,0065	
	8	3,14	2,504	0,017	125,030	9,481	1,244	1,0079	
	10	3,05	9,780	ND	117,500	12,820	1,279	1,0113	
	Cola	1	2,98	26,780	0,030	71,178	8,414	1,300	1,0098
		2	2,07	15,150	ND <sup>g</sup>	70,820	9,280	1,178	1,0077
		5	1,95	11,920	ND	69,870	12,160	1,274	1,0109
6		2,74	15,261	ND	60,490	8,137	1,236	1,0103	
7		2,42	ND	ND	61,920	11,000	1,266	1,0114	
8		2,60	18,410	0,680	61,920	8,180	1,236	1,0054	
9		2,51	0,730	0,196	58,090	11,560	1,280	1,0118	
Laranja	1	3,11	35,310	0,660	111,940	9,700	1,361	1,0148	
	2	2,72	36,990	ND	88,380	9,970	1,195	1,0083	
	5	3,94	26,780	0,124	173,050	12,700	1,372	1,0128	
	6	3,44	28,500	0,014	167,790	10,680	1,382	1,0100	
	7	3,23	16,490	6,000	171,880	12,407	1,281	1,0118	

<sup>a</sup>ácido benzóico, <sup>b</sup>ácido sórbico, <sup>c</sup>acidez total, <sup>d</sup>sacarose, <sup>e</sup>viscosidade relativa, <sup>f</sup>índice de refração relativo, <sup>g</sup>ND = não detectado. Teores em mg/100 mL.

contendo os histogramas individuais de cada imagem digitalizada. O esquema geral para obtenção da matriz de dados pode ser visualizado na Figura 2S, Material Suplementar. Em seguida, alguns histogramas foram considerados discordantes, uma vez que o perfil de cor resultante para a marca apresentou índices de cor deslocados. Isso foi devido, provavelmente, aos erros de leitura do scanner. As imagens resultantes após a exclusão dos histogramas discordantes foram 498, 444 e 356 para as categorias Cola, Guaraná e Laranja, respectivamente. Esses histogramas discordantes representaram 1,2 %, em média, dos histogramas totais para cada categoria.

### Análise de componentes principais (PCA)

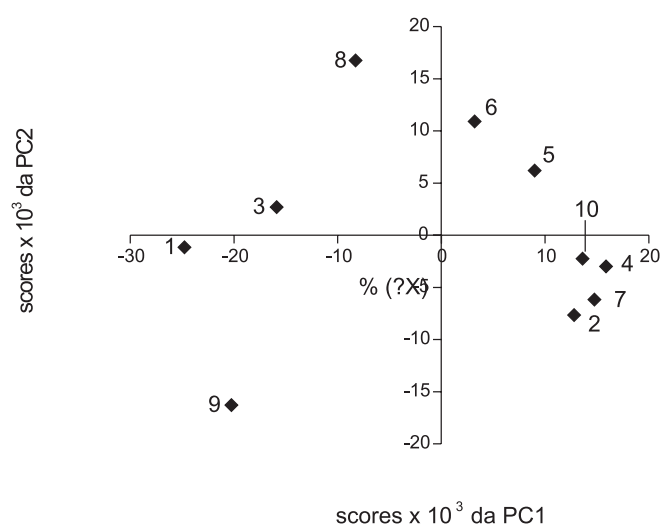
A ferramenta multivariada de análise de componentes principais (PCA) foi empregada para a classificação das imagens. PCA consiste em uma mudança de base dos dados multivariados, representados pela matriz  $\mathbf{X}$ , contendo os histogramas médios das imagens digitalizadas para cada categoria, encontrando o vetor  $\vec{v}$ , chamado autovetor, e um escalar,  $\tau$ , chamado autovalor, para o qual a igualdade  $\text{cov}(\mathbf{X})\vec{v} = \tau\vec{v}$  é satisfeita.  $\text{cov}(\mathbf{X})$  é o auto-sistema da covariância dos dados multivariados, de maneira que transforme um espaço multidimensional em um subespaço menor, como por exemplo, de dimensão um, gerado por um autovetor.<sup>14</sup> Ou seja,  $\mathbf{X}$  (multivariável) representada por muitos vetores, passa a ter as informações relevantes, representadas por um ou mais autovetores. Esse subsistema deve reter as informações relevantes e representar, de modo mais adequado, cada variável individualmente. Desse modo, PCA descreve, em termos geométricos das componentes principais, a covariância das variáveis (no presente caso, os canais RGB) usando o menor número de autovetores. O novo subsistema consiste, então, dos eixos das componentes principais, sendo as coordenadas das amostras nesses novos eixos denominados de *scores*. Como cada componente principal é uma combinação linear das variáveis originais, os coeficientes dessas variáveis (ou seja, os autovetores) são chamados de *loadings*.<sup>15</sup>

A matriz de dados dos histogramas médios para cada marca de refrigerante, dentro de cada categoria, foi centrada na média, e a vari-

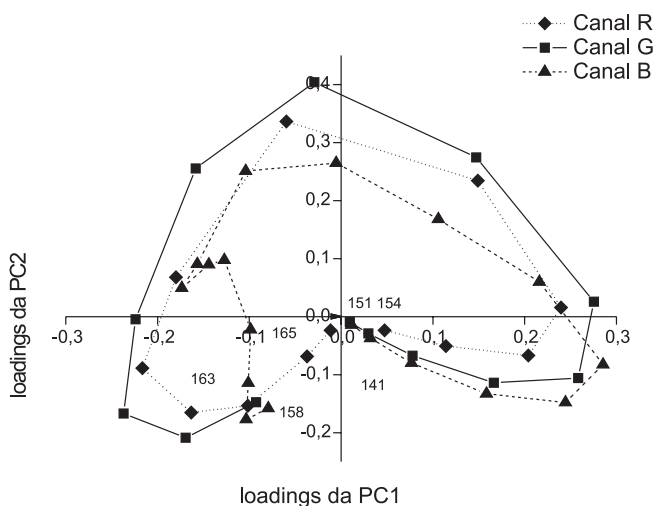
ância total explicada pelas duas primeiras componentes principais foi de 78,09% para Guaraná, 66,92% para Cola e 69,88% para Laranja. Os gráficos dos *scores* das duas primeiras componentes principais, para cada categoria, constam das Figuras 2, 4 e 5. O gráfico dos *loadings* das duas primeiras componentes principais para a categoria Guaraná consta da Figura 3, enquanto para as outras duas categorias esses gráficos constam das Figuras 3S e 4S, Material Suplementar.

### RESULTADOS E DISCUSSÃO

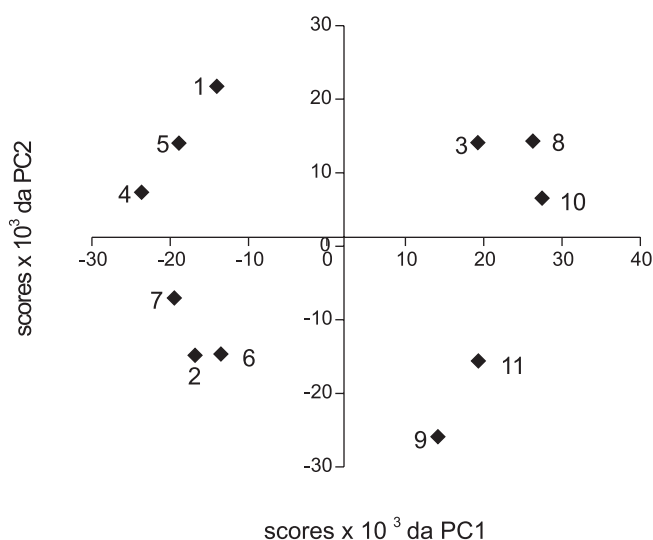
Para os refrigerantes da categoria Guaraná, o gráfico dos *scores* das duas primeiras componentes principais, Figura 2, explica 78,09%



**Figura 2.** Gráfico dos *scores* das duas primeiras componentes principais para as dez marcas analisadas dos refrigerantes da categoria Guaraná. PC1 explica 57,21% da variância total do dados, enquanto PC2 explica 20,88%



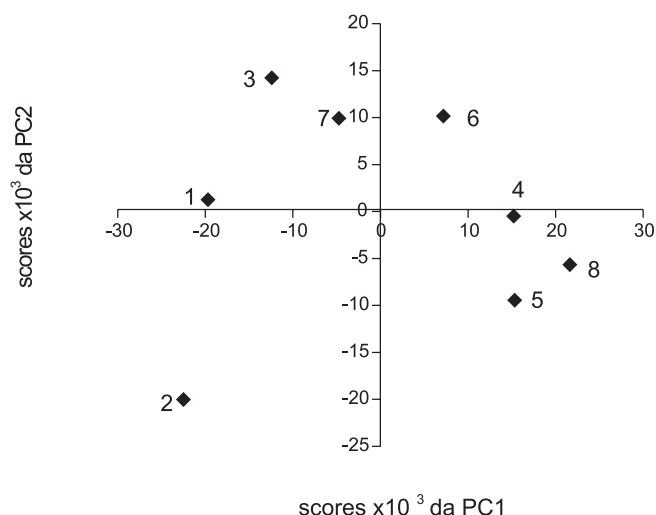
**Figura 3.** Gráfico dos loadings das duas primeiras componentes principais para as dez marcas analisadas dos refrigerantes da categoria Guaraná. As numerações, em destaque, referem-se aos valores dos índices de cor para os três canais, aumentando no sentido anti-horário: R de 154 a 165; G de 151 a 163; B de 141 a 158



**Figura 4.** Gráfico dos scores das duas primeiras componentes principais para as onze marcas analisadas dos refrigerantes da categoria Cola. PC1 explica 42,09% da variância total dos dados, enquanto PC2 explica 24,83%

da variância total dos dados, sendo que 57,21% é explicada por PC1 e 20,88% por PC2. Das 10 marcas analisadas, dois grupos formados pelas marcas 2, 4, 7 e 10, e 5 e 6 são separados em PC1. As outras marcas apresentam-se bem afastadas umas das outras, o que facilita a classificação por imagem em relação às demais. A distribuição dos pontos nessa figura está relacionada com os *loadings* dessas duas componentes principais, conforme a Figura 3. Nesse gráfico constam os *loadings* maiores que 0,01, sendo que para cada canal (R, G e B) estão representados os menores e os maiores índices de cor (os valores vão aumentando no sentido anti-horário para cada canal, e nessa figura estão representados apenas o menor e o maior índice de cor). Percebe-se, visualmente, que os *loadings* para todos os três canais de cor se distribuem de forma praticamente similar. Esse comportamento é reflexo da semelhança dos histogramas médios obtidos pelos canais de cor.

Os *loadings* positivos da PC1 encontram-se deslocados para tons



**Figura 5.** Gráfico dos scores das duas primeiras componentes principais para as oito marcas analisadas dos refrigerantes da categoria Laranja. PC1 explica 48,15% da variância total do dados, enquanto PC2 explica 21,73%

mais escuros (menores índices de cor; vide Figura 1) e os negativos, para tons mais claros. Já em PC2, os tons mais claros e os mais escuros apresentam valores positivos, enquanto que os valores negativos são característicos dos tons intermediários. Desse modo, a distribuição das marcas no gráfico dos *scores* de PC1 x PC2, Figura 2, vai da mais escura, 2, para a mais clara, 9, clareando no sentido anti-horário, na ordem 2-7-4-10-5-6-8-3-1-9. Com base nos *loadings* as marcas são classificadas, na Figura 2, conforme: as mais escuras são 2, 4, 7 e 10, e as mais claras 1 e 9. As marcas 3, 5, 6 e 8 apresentam tonalidades intermediárias, sendo que 5 e 6 são mais escuras que 3 e 8.

Análises físico-químicas realizadas para 7 das 10 marcas dessa categoria, cujas imagens foram digitalizadas, constam da Tabela 2. Dentre as variáveis nessa tabela, o teor de sacarose é o que guarda maior correlação com as imagens. No sentido horário, Figura 2, partindo da marca 1 até a 10 (seqüência 1-8-6-5-10) ocorre um aumento do teor de sacarose, o que indica as amostras mais escuras com índices maiores de sacarose. No entanto, essa tendência não é seguida para as amostras 7 e 2, com teores de 8,23 e 6,75 mg/100 mL, respectivamente.

Para os refrigerantes do tipo Cola, o gráfico dos *scores* das duas primeiras componentes principais explica 66,92% da variância dos dados, Figura 4, sendo que PC1 explica 42,09%, enquanto que PC2 explica 24,03%. Nessa figura cada quadrante, bem caracterizado, representa um grupo de classificação, com boa separação entre os grupos de marcas. Apesar da similaridade no comportamento obtido para os *loadings* dessa categoria (Material Suplementar, Figura 3S), suas tendências são contrárias às verificadas para a categoria Guaraná. Ou seja, na Figura 4, PC1 separa os tons mais escuros (lado esquerdo) dos tons mais claros (lado direito), enquanto que PC2 separa os tons intermediários (abaixo) dos tons mais e menos escuros (acima). As marcas mais escuras são 1, 4 e 5 e as menos escuras 3, 8 e 10. As demais apresentam tonalidades intermediárias, sendo que 2, 6 e 7 são mais escuras que 9 e 11. As marcas vão, então, se tornando menos escuras no sentido 1-5-4-7-2-6-9-11-10-8-3.

Análises físico-químicas realizadas para 7 das 11 marcas dessa categoria, cujas imagens foram digitalizadas, constam da Tabela 2. Dentre as variáveis nessa tabela, o teor de ácido sórbico é o que guarda maior correlação com as imagens. No lado esquerdo de PC1, Figura 4, estão as marcas nas quais o ácido ascórbico não foi detectado (marcas 2, 5, 6 e 7) e aquela que apresentou um valor muito baixo (marca 1 com 0,030 mg/100 mL), em contrapartida às amostras 8 (0,68 mg/100

mL) e 9 (0,196 mg/100 mL), situadas à direita em PC1.

Para a categoria Laranja, a Figura 5 representa os *scores* das duas primeiras componentes principais com 69,88% da variância explicada, sendo 48,15% por PC1 e 21,73% por PC2. Por se tratar de uma categoria cujas marcas apresentam cores mais claras, comparadas à categoria Cola, os índices de cor de maior ocorrência têm predominância em valores maiores para todos os três canais, conforme as variações obtidas nos *loadings*. Esses *loadings*, maiores que 0,01 (Material Suplementar, Figura 4S), ressaltam a correlação entre os canais de cor G e B, e as mudanças, no sentido horário, para valores menores (menos claros) para os valores maiores (mais claros). Logo, as marcas na Figura 5 estão distribuídas conforme a separação em PC1 entre tons mais claros (lado direito) e mais escuros/menos claros (lado esquerdo). Já PC2 separa as marcas conforme os tons mais claros e mais escuros (abaixo) dos tons intermediários (acima). As marcas, nessa figura, vão clareando no sentido 2-1-3-7-6-4-8-5.

Análises físico-químicas realizadas para 5 das 8 marcas dessa categoria, cujas imagens foram digitalizadas, constam da Tabela 2, sendo que o pH guarda uma maior correlação com as imagens. À medida que as imagens vão se tornando mais claras, no sentido horário, o pH vai aumentando. Seguindo a ordem das marcas 2, 1, 7, 6 e 5, Figura 5, o pH aumenta, respectivamente, na ordem 2,72, 3,11, 3,23, 3,44 e 3,94.

A separação das marcas analisadas para os refrigerantes do tipo Laranja pode ser percebida, de modo simplificado, pelos histogramas médios das imagens, por exemplo, com base nos índices de cor do canal B, Figura 6. Nessa figura, as amostras 1, 2 e 3 apresentam os histogramas mais à esquerda (tons mais escuros), enquanto que as amostras 5 e 8, mais à direita (tons mais claros).

Em suma, as diferenças apresentadas entre as marcas analisadas estão diretamente relacionadas às diferenças nas cores. O que faz uma determinada marca de um refrigerante da categoria Cola ser diferente de outra da mesma categoria é o fato de que, apesar de imperceptíveis, visualmente, as suas cores, em média, são diferentes. Naturalmente, isso é resultado da composição química do refrigerante. Assim sendo, a classificação pela cor é um modo alternativo, simples e barato em relação à classificação físico-química da amostra. Tal metodologia pode vir a facilitar a identificação de amostras de refrigerantes adulteradas e/ou falsificadas, com implicação direta no comércio desse produto.

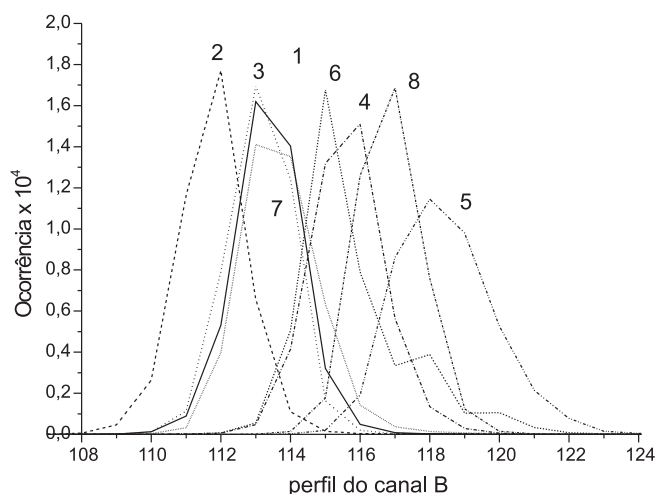


Figura 6. Perfil do canal de cor B dos histogramas médios das imagens das oito marcas dos refrigerantes da categoria Laranja

## CONCLUSÕES

Diferentemente da metodologia analítica convencional, imagens de refrigerantes podem ser empregadas para classificar diferentes marcas dessas bebidas, dentro de um mesmo tipo/sabor/categoria. Das 29 marcas de refrigerantes dos tipos Cola, Guaraná e Laranja analisadas, foi possível estabelecer padrões de similaridade dentro dos gráficos dos *scores* das componentes principais, com base nos valores médios dos histogramas dos canais de cor R, G e B, a partir das imagens geradas por um *scanner* de mesa. A mudança na cor das imagens é acompanhada no gráfico das componentes principais conforme o peso do índice de cor na componente principal, como resultado da tonalidade média de cada marca. Resultados de análises físico-químicas como o teor de sacarose, de ácido sórbico e o pH podem ser correlacionados com as imagens classificadas pela PCA. Em linhas gerais, marcas diferentes de refrigerantes podem ser classificadas pelas suas imagens.

## MATERIAL SUPLEMENTAR

A Figura 1S contém o perfil RGB dos histogramas médios para 3 marcas da categoria Cola; a Figura 2S contém um esquema gráfico da obtenção da matriz X; as Figuras 3S e 4S contêm os *loadings*, maiores que 0,01, das duas primeiras componentes principais para as categorias Cola e Laranja. Esse material encontra-se disponível gratuitamente em <http://www.quimicanova.s bq.org.br>, na forma de arquivo PDF.

## AGRADECIMENTOS

Ao CNPq e à FUNAPE pelo suporte financeiro.

## REFERÊNCIAS

1. <http://www.apexbrasil.com.br>, acessada em Setembro 2005.
2. *Jornal Diário da Manhã*, Caderno Universidade, Goiás, 13/10/2002.
3. Pratt, W. K.; *Digital Image Processing*, 2<sup>nd</sup> ed., John Wiley & Sons: New York, 1991.
4. Budantsev, A. Y.; *J. Anal. Chem.* **2004**, *59*, 703; Schmidt, F.; *Dissertação de Mestrado*, Universidade Estadual de Campinas, Brasil, 1997.
5. Hayakawa, T.; Hirai, M.; *Anal. Chem.* **2003**, *75*, 6728.
6. Souza, L. A. C.; Spehar, C. R.; Santos, R. L. B.; *Pesquisa Agropecuária Brasileira* **2004**, *39*, 397.
7. Hirano, A.; Moridera, N.; Akashi, M.; Saito, M.; Sugawara, M.; *Anal. Chem.* **2003**, *75*, 3775.
8. Reichenbach, S. E.; Ni, M.; Kottapalli, V.; Visvanathan, A.; *Chemom. Intell. Lab. Syst.* **2004**, *71*, 107.
9. Antonelli, A.; Cocchi, M.; Fava, P.; Foca, G.; Frachini, G. C.; Manzini, D.; Ulrici, A.; *Anal. Chim. Acta* **2004**, *515*, 3; Yu, H.; MacGregor, J. F.; *Chemom. Intell. Lab. Syst.* **2003**, *67*, 125.
10. de Oliveira, S. B.; Pereira, R. O.; Ribeiro, K. de O.; de Oliveira, A. E.; *Revista Processos Químicos* **2007**, *2*, 21.
11. <http://www.scilab.org>, acessada em Novembro 2006.
12. Fabri, R.; *Monografia*, Universidade de São Paulo, Brasil, 2002.
13. <http://siptoolbox.sourceforge.net>, acessada em Junho 2007.
14. Noble, B.; Daniel, J. W.; *Álgebra linear aplicada*, Prentice-Hall do Brasil: Rio de Janeiro, 1986.
15. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; *Quim. Nova* **1999**, *22*, 724.

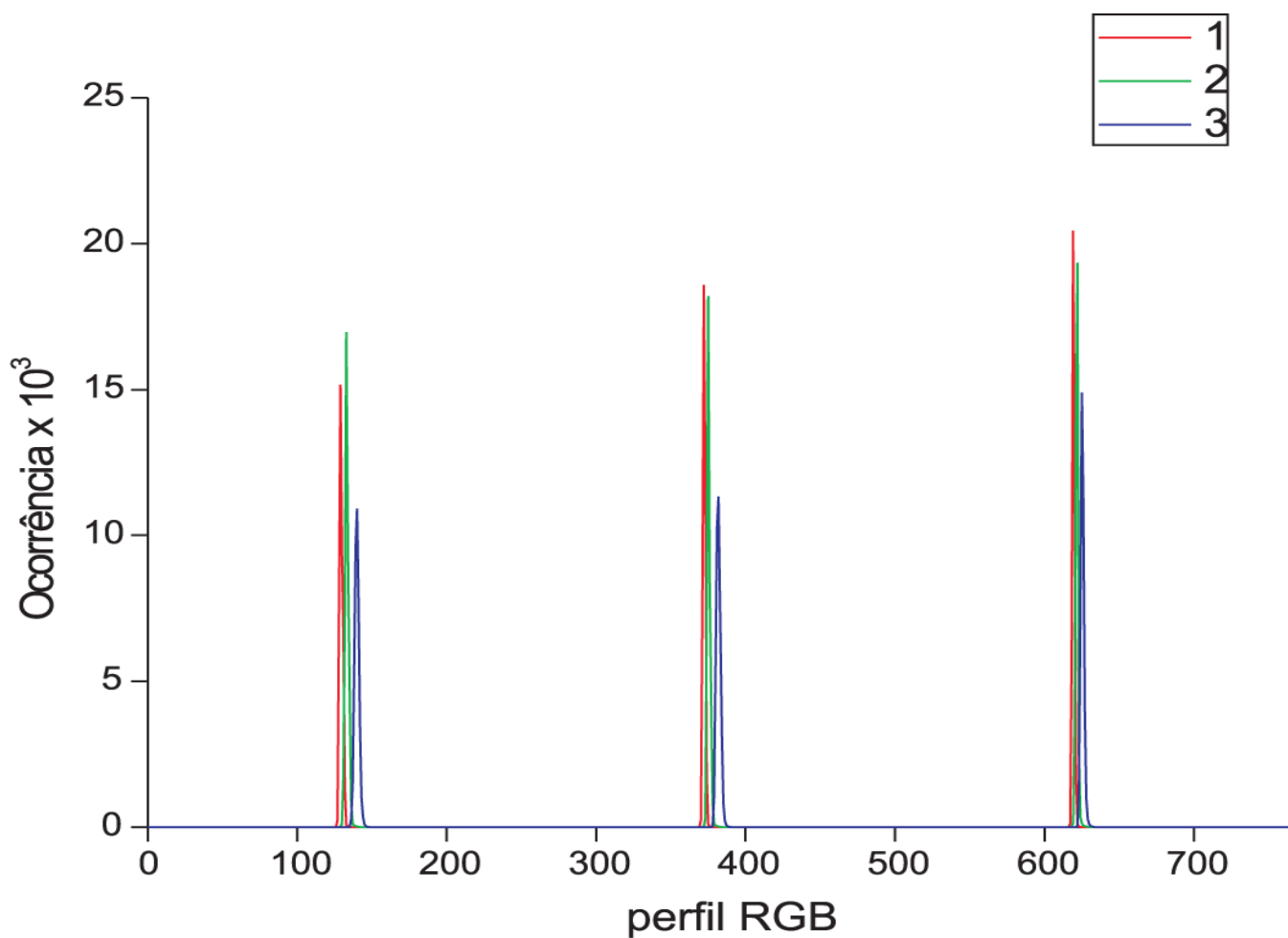
CLASSIFICAÇÃO DE REFRIGERANTES ATRAVÉS DE ANÁLISE DE IMAGENS E PCA

Mariana da S. Godinho, Raquel O. Pereira, Keysi de O. Ribeiro, Fernando Schmidt e Anselmo E. de Oliveira\*

Instituto de Química, Universidade Federal de Goiás, CP 131, 74001-970 Goiânia – GO, Brasil

Sérgio B. de Oliveira

Centro Federal de Educação Tecnológica de Goiás, Rua 75, 46, 74055-110 Goiânia – GO, Brasil



**Figura 1S.** Perfil RGB dos histogramas médios para três marcas da categoria Cola. O eixo “Ocorrência x 10<sup>3</sup>” representa o número de vezes que um determinado índice de cor (referente ao canal R, G ou B) está contido na área da imagem digitalizada. O eixo “perfil RGB” representa os índices das cores para os canais R (0 a 255), G (256 a 512) e B (513 a 768)



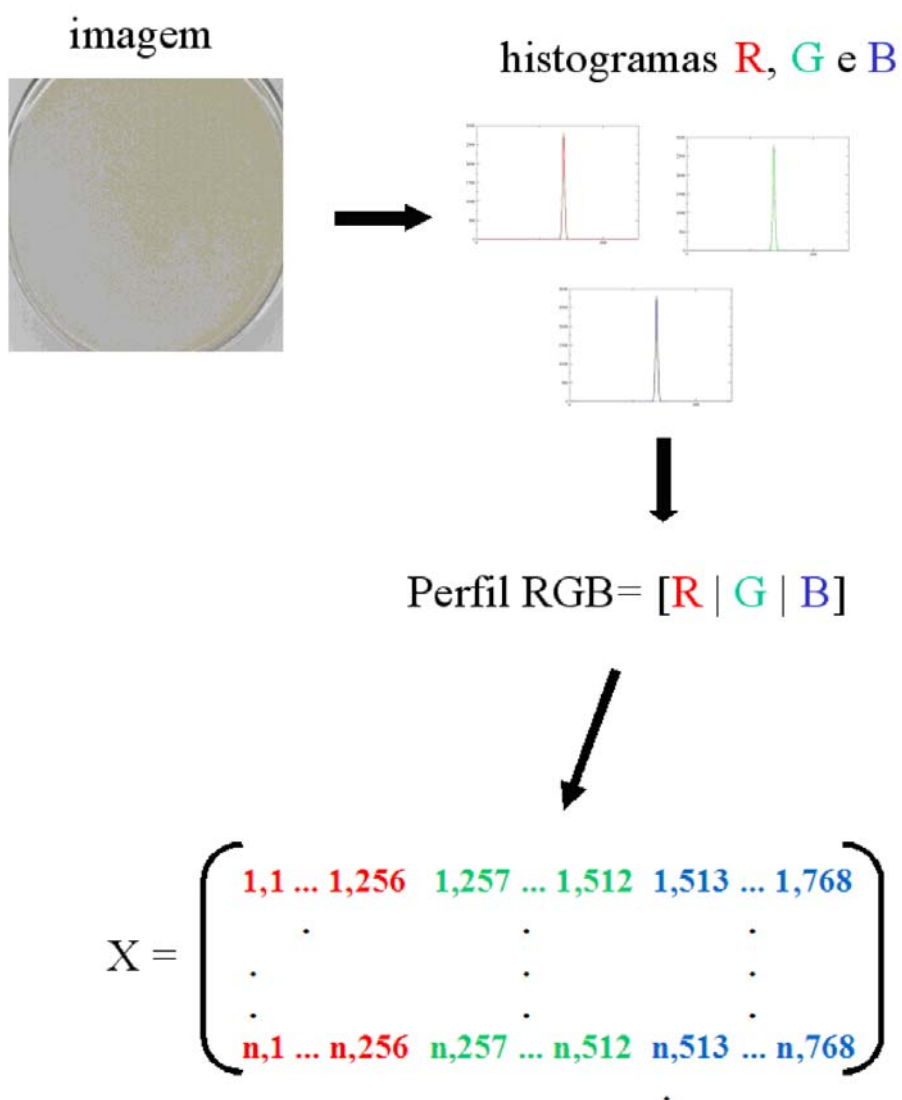
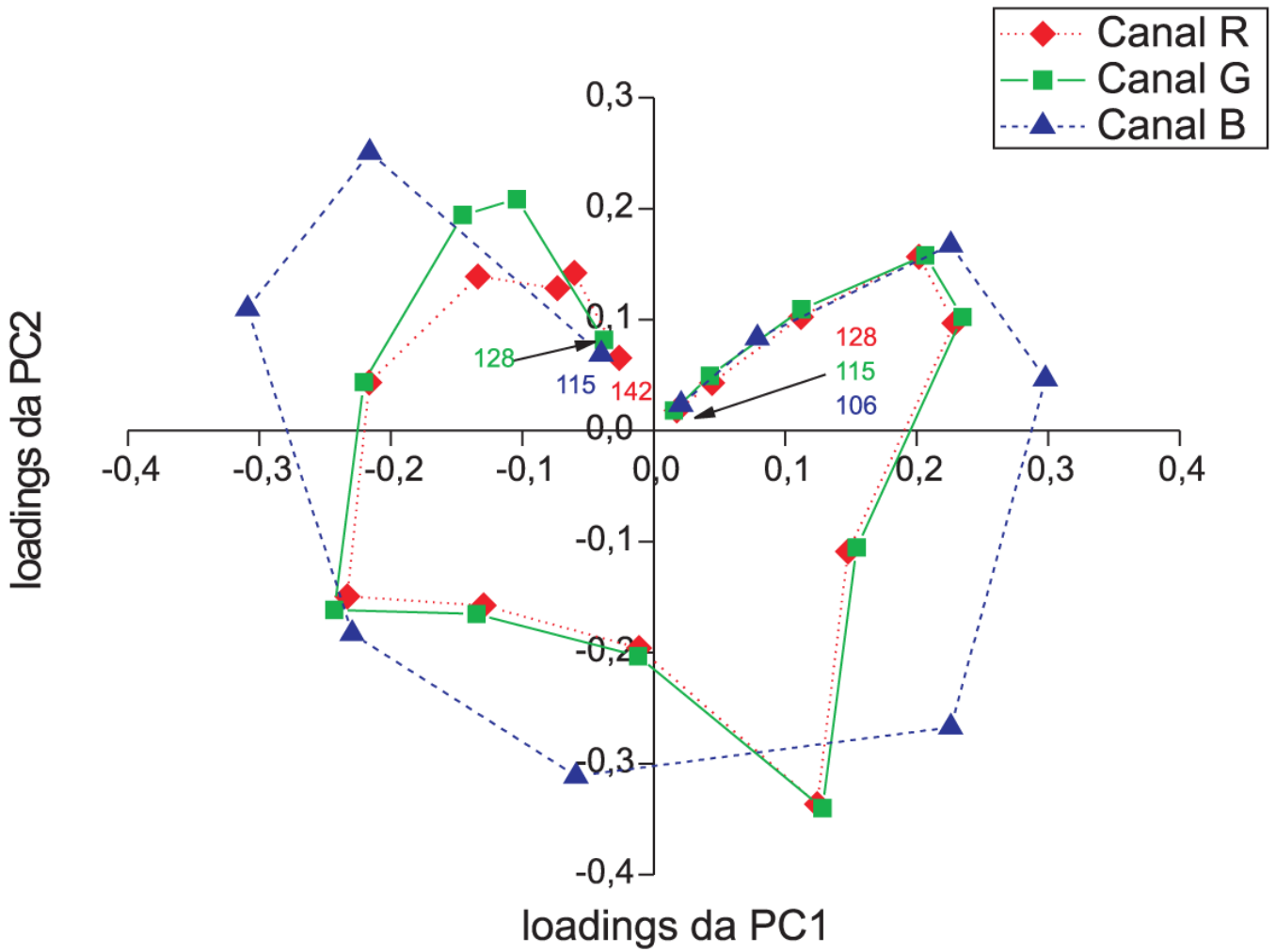
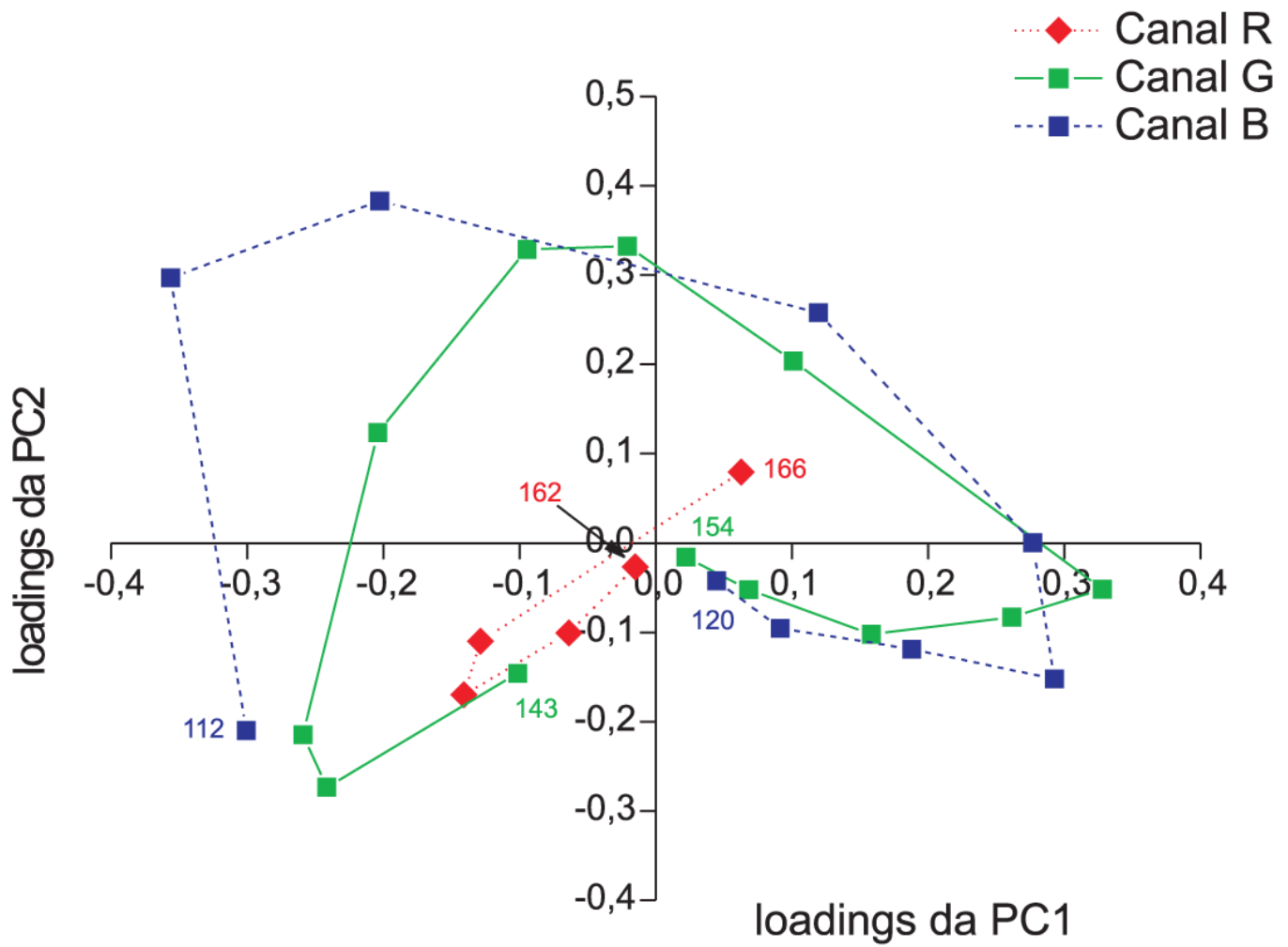


Figura 2S. Esquema de obtenção da matriz X



**Figura 3S.** Gráfico dos loadings das duas primeiras componentes principais para as onze marcas dos refrigerantes da categoria Cola. As numerações, em destaque, referem-se aos valores dos índices de cor para os três canais, aumentando no sentido anti-horário: R de 128 a 142; G de 115 a 128; B de 106 a 115





**Figura 4S.** Gráfico dos scores das duas primeiras componentes principais para as oito marcas analisadas dos refrigerantes da categoria Laranja. PC1 explica 48,15% da variância total do dados, enquanto PC2 explica 21,73%