

**UTILIZAÇÃO DA ESPECTROSCOPIA NA  
REGIÃO DO INFRAVERMELHO (FTIR) E  
QUIMIOMETRIA NA IDENTIFICAÇÃO DO  
CAFÉ TORRADO E MOÍDO ADULTERADO**

**MARIANA CRISTINA SOUZA SANTOS**

**2005**

**MARIANA CRISTINA SOUZA SANTOS**

**UTILIZAÇÃO DA ESPECTROSCOPIA NA REGIÃO DO  
INFRAVERMELHO (FTIR) E QUIMIOMETRIA NA IDENTIFICAÇÃO  
DO CAFÉ TORRADO E MOÍDO ADULTERADO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Agroquímica e Agrobioquímica, para a obtenção do título de “Mestre”.

Orientador

Prof. Dr. Walclée de Carvalho Melo

LAVRAS  
MINAS GERAIS – BRASIL  
2005

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Santos, Mariana Cristina Souza

Utilização da espectroscopia na região do infravermelho (FTIR) e quimiometria na identificação do café torrado e moído adulterado / Mariana Cristina Souza Santos. -- Lavras : UFLA, 2005.

65 p. : il.

Orientador: Walclée de Carvalho Melo.

Dissertação (Mestrado) – UFLA.

Bibliografia.

*1. Espectroscopia. 2. FTIR. 3. Quimiometria. 4. Café. I. Universidade Federal de Lavras. II. Título.*

CDD-543.0858

-630.24

-633.73

**MARIANA CRISTINA SOUZA SANTOS**

**UTILIZAÇÃO DA ESPECTROSCOPIA NA REGIÃO DO  
INFRAVERMELHO (FTIR) E QUIMIOMETRIA NA IDENTIFICAÇÃO  
DO CAFÉ TORRADO E MOÍDO ADULTERADO**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Agronomia, área de concentração em Agroquímica e Agrobioquímica, para a obtenção do título de “Mestre”.

APROVADA em 4 de março de 2005

Profa. Celeste Maria Patto de Abreu

UFLA

Profa. Luciana Maria Vieira L. Mendonça

EAFMUZ

Prof. Dr. Walclée de Carvalho Melo  
(UFLA)  
Orientador

LAVRAS  
MINAS GERAIS – BRASIL

## AGRADECIMENTOS

Primeiramente a Deus, por tudo que sou, por tudo que tenho, por tudo que vivo e por tudo que conquistei...

À minha família, meu tudo. A meu pai, por todo seu esforço; à minha mãe, por seu carinho e amor incondicionais e aos meus queridos irmãos, companheiros pra toda a vida.

Em especial ao Prof. Dr. Mário Guerreiro, um amigo, um mestre e um grande exemplo de perseverança, humildade, inteligência e paciência.

Ao Prof. Dr. Walclée, pela orientação deste trabalho e por sua amizade.

À Prof. Dra. Malu, pela amizade e convivência.

Aos professores do Departamento de Química, Ciências Exatas e Ciências dos Alimentos, pelos ensinamentos em suas disciplinas.

À Dra. Celeste e à Dra. Luciana, pela participação na banca de defesa desta dissertação.

Aos colegas de curso, pela amizade, convivência e grandiosas trocas de experiência.

Às amigas Cris, Maraísa, Fabiana, Andressa e Ester, pelo apoio e carinho.

À secretária Miriam, por todo apoio e atenção.

À Xulita, por toda sua disposição e ajuda imprescindíveis.

Ao pessoal do Departamento de Ciências Exatas, pela paciência, apoio, convivência e amizade.

A todo o pessoal do CAPQ, pela convivência diária.

À Universidade Federal de Lavras, por ter me acolhido por mais dois anos, dando-me mais esta oportunidade.

Ao CNPq, pelo financiamento do trabalho por meio de bolsa de estudos.

Ao NECAF, pela doação das amostras de café utilizadas neste trabalho.

Aos meus familiares: tios, tias, vó Geralda, vó Carlinho que agora olha por mim lá de cima junto com vó Walfrides e vó Mariana e aos meus queridos primos e primas.

Aos meus grandes amigos Má, Wal, Ju, Bruninho, Su, Ivo, Paula, Flavinha, Carol Souza, Vivi, Dri, Rick e Breno. Vocês são especiais. Obrigada de coração!

Enfim, a todos aqueles que participaram e torceram (de longe ou de perto) por esta conquista. Muito obrigada!

## SUMÁRIO

	<b>Página</b>
<b>RESUMO</b> .....	i
<b>ABSTRACT</b> .....	ii
<b>1 INTRODUÇÃO</b> .....	1
<b>2 REFERENCIAL TEÓRICO</b> .....	3
2.1 O café.....	3
2.1.1 O café hoje.....	5
2.1.2 O café e sua qualidade.....	5
2.1.3 O café e sua composição química.....	6
2.2 Fraudes.....	7
2.3 Métodos de detecção de fraudes.....	9
2.4 Espectroscopia.....	12
2.4.1 Transmitância.....	12
2.4.2 Reflectância difusa (DRIFT).....	13
2.4.3 Reflectância total atenuada (ATR).....	16
2.5 Quimiometria.....	17
2.5.1 Quimiometria aplicada à química de alimentos.....	18
2.6 Aplicativo R.....	20
2.7 Métodos qualitativos de análise.....	21
2.7.1 Análise de componentes principais (PCA).....	21
2.7.2 Análise de agrupamentos hierárquicos (HCA).....	24
2.8 Métodos quantitativos de análise.....	26

2.8.1 Calibração multivariada.....	26
2.8.2 Regressão por mínimos quadrados parciais (PLS).....	28
<b>3 MATERIAL E MÉTODOS.....</b>	<b>30</b>
3.1 Amostras de café.....	30
3.2 Adulteração das amostras.....	30
3.3 Liofilização das amostras.....	30
3.4 Leituras no instrumento FTS 3000X.....	31
3.5 Análises espectrais.....	31
3.5.1 Análises de transmitância.....	31
3.5.2 Análises de reflectância difusa (DRIFT).....	31
3.5.3 Análises de reflectância total atenuada (ATR).....	32
3.6 Organização dos dados.....	32
3.7 Processamento dos dados espectrais.....	32
3.8 Análise estatística.....	33
<b>4 RESULTADOS E DISCUSSÃO.....</b>	<b>34</b>
4.1 Transmitância.....	34
4.1.1 Análise de componentes principais (PCA).....	35
4.1.2 Análise de agrupamentos hierárquicos (HCA).....	39
4.1.3 Regressão por mínimos quadrados parciais (PLS).....	41
4.2 Reflectância difusa (DRIFT).....	42
4.2.1 Análise de componentes principais (PCA).....	43
4.2.2 Análise de agrupamentos hierárquicos (HCA).....	47
4.2.3 Regressão por mínimos quadrados parciais (PLS).....	49
4.3 Reflectância total atenuada (ATR).....	50
4.3.1 Análise de componentes principais (PCA).....	51

4.3.2 Análise de agrupamentos hierárquicos (HCA).....	55
4.3.3 Regressão por mínimos quadrados parciais (PLS).....	57
<b>5 CONCLUSÕES</b> .....	59
5.1 Perspectivas futuras.....	60
<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	61

## RESUMO

SANTOS, Mariana Cristina Souza. **Utilização da espectroscopia na região do infravermelho (FTIR) e quimiometria na identificação do café torrado e moído adulterado.** Lavras: UFLA, 2005. 65p. (Dissertação - Mestrado em Agronomia)\*

Neste trabalho são combinadas as potencialidades das técnicas quimiométricas e as facilidades na obtenção de espectros apresentadas pelas técnicas de espectroscopia na região do infravermelho com transformada de Fourier (FTIR), através de reflectância total atenuada (ATR), reflectância difusa (DRIFT) e transmitância, visando à identificação de cevada (substância usada como adulterante) em café torrado e moído. A região espectral analisada foi a região do infravermelho médio. Foram utilizadas amostras de café arábica de bebida dura. Os grãos de café foram torrados com torra de padrão comercial e moídos em seguida. Ao café torrado e moído foi adicionada cevada nas proporções de 0%, 0,5%, 3%, 5%, 7%, 10%, 15% e 20%. Foram feitas três repetições de cada, totalizando 24 amostras. As amostras adulteradas foram submetidas à liofilização para se evitar a interferência das bandas de água no espectro e deixadas em dessecador com sílica gel por 24h. Em seguida, foram realizadas as leituras dos espectros. Os dados foram organizados em matrizes e centrados na média. Os métodos quimiométricos empregados foram: análise de componentes principais (PCA), análise de agrupamentos hierárquicos (HCA) utilizando distância euclidiana para cálculo da matriz de distâncias e o método do vizinho mais distante, e regressão por mínimos quadrados parciais (PLS), por meio de validação cruzada. Utilizou-se o software R para o tratamento dos dados com as funções adequadas. Os métodos empregados permitiram separar o café com cevada do café puro por meio da reflectância difusa (DRIFT) e da reflectância total atenuada (ATR), com exceção das amostras contendo 0,5% de cevada. Os métodos quimiométricos possibilitaram, também, distinguir amostras com 20% de cevada pelas técnicas de reflectância total atenuada (ATR) e reflectância difusa (DRIFT). A técnica de transmitância não foi eficaz na identificação de cevada em café torrado e moído. Na análise de regressão por mínimos quadrados parciais foram obtidos resultados satisfatórios somente para a análise de ATR, na qual se estabeleceu um bom ajuste entre os valores verdadeiros e os valores estimados de cevada. Um coeficiente de correlação de 98% foi encontrado.

---

\* Comitê Orientador: Dr. Walclée de Carvalho Melo – UFLA (Orientador), Dr. Mário César Guerreiro – UFLA

## ABSTRACT

SANTOS, Mariana Cristina Souza. **Utilization of spectroscopy in the infrared region (FTIR) and chemometry in the identification of adulterated toasted and ground coffee.** Lavras: UFLA, 2005. 65p. (Dissertation – Master in Agronomy)\*

In this work the potentialities of the chemometric techniques and the eases in the obtaining of spectra presented by the spectroscopy techniques in the infra-red region with Fourier transform (FTIR) through the attenuated total reflectance (ATR), drift reflectance (DRIFT) and transmittance, aiming at the identification of barley (substance used as an adulterant) in a ground and toasted coffee are combined. The analyzed spectral region was the medium infrared region. Arabic coffee samples of hard beverage were utilized. The coffee beans were toasted with standard commercial toasting and ground next. To the toasted and ground coffee was added barley at the ratios of 0%, 0,5%, 3%, 5%, 7%, 10%, 15% and 20%. Three replicates of each one amounting to 24 samples were done. The adulterated samples were submitted to freeze-drying to avoid the interference of the water bands into the spectrum and left in a desiccator with silica gel for 24 h. Next, the spectra were obtained. The data were organized into matrices and centered on the mean. The chemometric methods employed were: principal component analysis (PCA), hierarchical clustering analysis (HCA), by utilizing euclidean distance for calculation of the distance and the complete linkage method and partial least squares regression (PLS) with cross validation. The R software was utilized for the treatment of the data with adequate functions. The employed methods allowed to separate the coffee with barely from the pure coffee using the diffuse reflectance (DRIFT) and attenuated total reflectance (ATR) with the exception of the samples containing 0.5% of barley. The chemometric methods also allowed to distinguish samples with 20% of barley through the techniques of attenuated total reflectance (ATR) and diffuse reflectance (DRIFT). The transmittance technique was not effective in identifying barley in toasted and ground coffee. In the partial least squares regression analysis, satisfactory results were obtained only for the ATR technique, finding a good adjust between the true values and the estimated values of barley. A correlation coefficient of 98% was found.

---

\* Guidance Committee: Dr. Walclée de Carvalho Melo – UFLA (Adviser), Dr. Mario César Guerreiro – UFLA

## 1 INTRODUÇÃO

O café é reconhecido mundialmente pela grande receptividade que tem sua bebida, sendo considerado um dos produtos de maior comercialização no mercado internacional. Este grão possui particular importância para o Brasil, pois seu cultivo e industrialização geram um grande número de empregos e divisas para o país. Ainda hoje, o Brasil é reconhecido como um dos maiores produtores e exportadores de grãos de café e de café processado. Esses fatores deixam claro que estudos relacionados à avaliação da qualidade do café, como composição do café e detecção de fraudes são indiscutivelmente importantes.

A detecção de impurezas e misturas no pó de café torrado e moído é uma preocupação constante no que diz respeito, principalmente, à garantia da qualidade do produto.

Em todas as partes do mundo onde o café é consumido, as autoridades sanitárias sempre se preocuparam com as fraudes no produto. O material usado para adulterar o café é o mais barato, disponível em grande quantidade e apresenta perfeita semelhança com o café ao ser torrado e moído.

Por fraude, considera-se a mistura, intencional ou não, de materiais estranhos ao produto, normalmente de baixo custo, que alteram a sua qualidade e causam danos ao consumidor, especialmente, os de ordem econômica.

No Brasil, as impurezas ou misturas encontradas, com maior frequência, no café torrado e moído são cascas e paus, milho, cevada, trigoilho, açúcar mascavo, soja e café esgotado, entre outras. Um dos principais problemas encontrados na identificação dessas substâncias refere-se aos métodos analíticos empregados, pois são demorados, subjetivos e apresentam resultados muitas vezes, discordantes.

Com relação ao café torrado e moído, por exemplo, o processo convencionalmente utilizado consiste na preparação de lâminas microscópicas

para análise visual. Por tratar-se de análise microscópica, este processo constitui-se em um método subjetivo e, conseqüentemente, a confiabilidade dos resultados é pequena, já que depende da experiência do analista e, por isso, está sujeita a erros humanos, além de ser demorado e oneroso.

Uma técnica instrumental de análise química, amplamente utilizada nos dias de hoje, é a técnica de espectroscopia de absorção na região do infravermelho, que permite as análises qualitativa e quantitativa e a identificação de uma substância pelos grupos funcionais presentes na molécula em consideração.

Uma das características mais interessantes desses modernos instrumentos é o número das variáveis que podem ser medidas em uma única amostra. Um exemplo notável é a intensidade de absorção em mil ou mais comprimentos de onda, que é rotineiramente registrada em único espectro.

A aquisição de dados, principalmente na área de química analítica, atingiu um ponto bastante sofisticado com o interfaceamento de instrumentos aos computadores, produzindo uma enorme quantidade de informação, muitas vezes complexa e variada.

De posse de tal quantidade de dados, aumentou necessidade de ferramentas novas e mais sofisticadas para tratá-los e extrair informações relevantes, dando origem à quimiometria, que é uma área especificamente destinada à análise de dados químicos de natureza multivariada.

Dessa forma, o presente trabalho objetivou verificar a viabilidade do uso da espectroscopia na região do infravermelho para a identificação de cevada adicionada ao café torrado e moído, como adulterante, associada às potencialidades da quimiometria, utilizando-se um software de fonte aberta para o tratamento dos dados.

## **2 REFERENCIAL TEÓRICO**

### **2.1 O café**

A planta de café é originária da Etiópia, centro da África, onde ainda faz parte da vegetação natural. Foi a Arábia, o país responsável pela propagação da cultura do café. A palavra café não é originário da Kaffa, local de origem da planta e sim da palavra árabe qahwa, que significa vinho. Por esse motivo, o café era conhecido como "vinho da Arábia", quando chegou à Europa no século XIV (ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ, 2005).

Os manuscritos mais antigos que mencionam a cultura do café datam de 575 no Yêmen, onde, consumido como fruto in natura, passou a ser cultivado. Somente no século XVI, na Pérsia, os primeiros grãos de café foram torrados para se transformar na bebida que hoje conhecemos (ABIC, 2005).

A partir de 1615, o café começou a ser saboreado no Continente Europeu, trazido por viajantes em suas freqüentes viagens ao oriente. Até o século XVII, somente os árabes produziam café. Alemães, franceses e italianos procuravam desesperadamente uma maneira de desenvolver o plantio em suas colônias. Mas foram os holandeses que conseguiram as primeiras mudas e as cultivaram nas estufas do jardim botânico de Amsterdã, fato que tornou a bebida do café uma das mais consumidas no velho continente, passando a fazer parte definitiva dos hábitos dos europeus (ABIC, 2005).

Com as experiências holandesa e francesa, o cultivo de café foi levado para outras colônias européias. O crescente mercado consumidor europeu propiciou a expansão do plantio de café em países africanos e a sua chegada ao Novo Mundo. Pelas mãos dos colonizadores europeus, o café chegou ao Suriname, a São Domingos, a Cuba, a Porto Rico e às Guianas. Foi por meio das Guianas que chegou ao norte do Brasil. Desta maneira, o segredo dos árabes se espalhou por todos os cantos do mundo (ABIC, 2005).

O café chegou ao norte do Brasil, mais precisamente em Belém, em 1727, trazido da Guiana Francesa pelo Sargento-Mor Francisco de Mello Palheta a pedido do governador do Maranhão e Grão Pará, que o enviara às Guianas com essa missão. Já naquela época o café possuía grande valor comercial (ABIC, 2005).

Devido às nossas condições climáticas, o cultivo de café se espalhou rapidamente, com a produção voltada para o mercado interno. Em sua trajetória pelo Brasil, passou pelo Maranhão, Bahia, Rio de Janeiro, São Paulo, Paraná e Minas Gerais. Num espaço de tempo relativamente curto, passou de uma posição relativamente secundária para a de produto-base da economia brasileira. Desenvolveu-se com total independência, ou seja, apenas com recursos nacionais, sendo, afinal, a primeira realização exclusivamente brasileira que visou a produção de riquezas (ABIC, 2005).

Por quase um século, o café foi a grande riqueza brasileira e as divisas geradas pela economia cafeeira aceleraram o desenvolvimento do Brasil e o inseriram nas relações internacionais de comércio. A cultura do café ocupou vales e montanhas, possibilitando o surgimento de cidades e dinamização de importantes centros urbanos por todo o interior do estado de São Paulo, sul de Minas Gerais e norte do Paraná. Ferrovias foram construídas para permitir o escoamento da produção, substituindo o transporte animal e impulsionando o comércio inter-regional de outras importantes mercadorias. O café trouxe grande contingente de imigrantes, consolidou a expansão da classe média, a diversificação de investimentos e até mesmo intensificou movimentos culturais. A partir de então, o café e o povo brasileiro passaram a ser indissociáveis (ABIC, 2005).

Implantado com o mínimo de conhecimento da cultura, em regiões que mais tarde se tornaram inadequadas para seu cultivo, a cafeicultura no centro-sul do Brasil começou a ter problemas em 1870, quando uma grande geada atingiu

as plantações do oeste paulista, provocando prejuízos incalculáveis (ABIC, 2005).

Depois de uma longa crise, a cafeicultura nacional se reorganizou e os produtores, industriais e exportadores voltaram a alimentar esperanças de um futuro melhor. A busca pela região ideal para a cultura do café se estendeu por todo o país firmando-se em regiões do estado de São Paulo, Minas Gerais, Paraná, Espírito Santo, Bahia e Rondônia. O café continua, hoje, a ser um dos produtos mais importantes para o Brasil e é, sem dúvida, o mais brasileiro de todos. Hoje, o país é o primeiro produtor e o segundo consumidor mundial do produto (ABIC, 2005).

### **2.1.1 O café hoje**

Atualmente, os maiores produtores mundiais de café são o Brasil, a Colômbia, o México e a Guatemala, onde predomina o cultivo de *Coffea arabica* L., e Indonésia, Vietnã, Costa do Marfim, Índia, Uganda e Etiópia, onde se cultiva a espécie *Coffea canephora* Pierre. No final da década de 1970, a área cultivada com café no mundo equivalia a 10 milhões de hectares, evoluindo para 13,5 milhões de hectares no final dos anos 1980, quando começou a reduzir. Hoje, situa-se em torno de 11,5 milhões de hectares. O Brasil é o maior produtor, com 28% da produção mundial, seguido de Colômbia com 14%, da Indonésia com 7%, e do México e Vietnã com 5%. A produção mundial de café nos últimos dez anos situou-se, em média, em 97 milhões de sacas de 60kg/ano (O café..., 2005).

### **2.1.2 O café e sua qualidade**

No setor produtivo, a qualidade do café está relacionada à obtenção de produtos que conquistem excelentes preços no mercado. No campo da difusão de tecnologia, entende-se por qualidade a execução de todas as etapas de

produção, com eficiência, economia e produtividade, com base no planejamento preciso de cada fase (Mendonça, 2004).

Para o setor agroindustrial, que agrega valor ao produto, as propriedades físicas, químicas, sensoriais e higiênico-sanitárias devem estar de acordo com padrões estabelecidos para que o café seja considerado de qualidade (Mendonça, 2004).

Para regulamentar a produção de cafés com qualidade mínima para o consumo, no Brasil, existem leis definidas para o café beneficiado e o torrado e moído.

A Portaria nº 377, de 26 de abril de 1999 (Brasil, 1999), foi criada com o objetivo de fixar a identidade e as características mínimas de qualidade a que devem obedecer o café torrado em grão e o café torrado e moído. A legislação especifica os requisitos obrigatórios em relação à composição, características sensoriais e físico-químicas e estabelece normas para o uso de aditivos e para a presença de contaminantes. O teor máximo de impurezas tolerado em café torrado é de 1%.

Segundo a Portaria nº 377, o café torrado não deve ser consumido quando estiver alterado ou adulterado por qualquer forma ou meio, inclusive pela adição de corantes ou outros produtos que modifiquem a sua especificação, cujo emprego é vedado (Brasil, 1999).

### **2.1.3 O café e sua composição química**

Os diversos constituintes físicos e físico-químicos presentes no café são os responsáveis pela aparência do grão torrado, pelo sabor e aroma característicos das bebidas. Entre esses constituintes destacam-se os compostos voláteis, fenólicos (ácido clorogênico), ácidos graxos, proteínas, açúcares, acidez, índice de coloração e degradação de parede celular dos grãos, com conseqüentes alterações em seus constituintes e algumas enzimas, cuja presença,

teores e atividade conferem ao café um sabor e aroma peculiares (Pimenta, 2003).

O odor característico do café é proporcionado pela presença de compostos voláteis, principalmente na forma de aldeídos, cetonas e ésteres metílicos, que são formados durante a torração e ficam retidos na estrutura celular dos grãos torrados (Pimenta, 2003).

Com relação à composição química dos grãos de café, Rinantonio (1987) verificou para cafés arábica, teores entre 55% a 65,5% de carboidratos; ácidos, 8% a 11%, correspondendo a 0,1% de ácido acético, propiônico, butírico e valérico; 0,7% a 1,4% ao ácido cítrico; 0,3% a 0,7% ao málico; 0,3% a 0,5% ao quínico; 6,7% a 9,2% ao ácido clorogênico e 1% a 3% à lignina; lipídeos, 15% a 18%; compostos nitrogenados, 11% a 15%, sendo 0,2% a 0,8% aminoácidos; 8,5% a 12% de proteína; 0,8% a 1,4% de cafeína e 0,6% a 1,2% de trigonelina e minerais, 3% a 5,4%.

## **2.2 Fraudes**

Dos artifícios desonestos utilizados no comércio, a fraude é o mais comum e é tão velho quanto a própria humanidade. Em todo setor especulativo onde esteja em jogo o interesse econômico, a ambição sem limites de aproveitadores sem escrúpulos faz sentir logo a sua indesejável presença (Menezes Jr., 1952).

A concorrência desleal desta casta de exploradores da economia popular se torna mais criminosa e desprezível quando se insurge no terreno da saúde pública, modificando e alterando a composição de produtos alimentícios, prejudicando e pondo em perigo a saúde e a vida de uma população (Menezes Jr., 1952).

Por fraude, considera-se a mistura, intencional ou não, de materiais estranhos ao produto, normalmente de baixo custo, que alteram a sua qualidade e

causam danos ao consumidor, especialmente os de ordem econômica (Assad et al., 2002).

Anular este desonesto expediente foi sempre a preocupação máxima das autoridades sanitárias, em todos os quadrantes da Terra e em todas as épocas (Menezes Jr., 1952).

Entre os gêneros alimentícios, o café é um produto que está sujeito a um grande número de fraudes, por ser muito extensa a lista de substâncias que se prestam a esse fim. É o produto em pó o que melhor se presta à prática de fraudes. Seu aspecto exterior granuloso, sua textura oleosa e aderente e sua cor, variando do castanho-avermelhado ao pardo-escuro, contribuem grandemente a se tornarem imperceptíveis, a olho nu, substâncias estranhas, as mais diversas, a ele adicionadas (Menezes Jr., 1952).

Desde que estejam com o mesmo grau de torração do café, estas substâncias são mascaradas pela absorção do óleo e aderência das partículas mais finas do pó de café à sua superfície, tornando-se difícil seu reconhecimento sem o auxílio de aparelhos e métodos analíticos especiais (Menezes Jr., 1952).

Devido ao preço elevado que o café alcança nos países importadores a sofisticação do produto em pó toma proporções alarmantes. São numerosas as substâncias utilizadas para esse fim e, entre elas, umas são consideradas como “sucedâneos do café” e outras servem, unicamente, para aumentar-lhe o peso e o volume (Menezes Jr., 1952).

Uma substância é tanto mais propícia a uma fraude quanto mais semelhança tiver com o produto visado, maior for a quantidade existente e menor o seu preço aquisitivo (Menezes Jr., 1952).

Chama-se de impureza toda substância que não é café, resultante do mal beneficiamento do produto, como cascas, paus, areia, etc., enquanto misturas são os produtos adicionados ao café para fraudá-lo, como milho, cevada, caramelo, etc. (Instituto..., 1997).

No Brasil, as impurezas ou misturas encontradas com maior frequência, no café torrado e moído, são cascas e paus, milho, cevada, trigoilho, açúcar mascavo e soja, entre outras (Assad et al., 2002).

### **2.3 Métodos de detecção de fraudes em café**

A detecção de impurezas e misturas no pó de café torrado e moído é uma preocupação constante no que diz respeito, principalmente, à garantia da qualidade do produto (Assad et al., 2002).

Um dos principais problemas encontrados na identificação dessas substâncias refere-se aos métodos analíticos empregados, pois são demorados, subjetivos e apresentam resultados muitas vezes discordantes (Assad et al., 2002).

Com relação ao café torrado e moído, por exemplo, o processo convencionalmente utilizado consiste na preparação de lâminas microscópicas e na sua análise visual (Assad et al., 2002).

Por tratar-se de análise microscópica, este processo constitui-se em um método subjetivo e, conseqüentemente, a confiabilidade dos resultados é pequena, já que depende da experiência do analista e, por isso, está sujeita a erros humanos, além de ser demorado e oneroso (Assad et al., 2002).

Assad et al. (2002) utilizaram princípios físicos de sensoriamento remoto empregando-se comportamento espectral e análises de imagens digitais para quantificar impurezas e misturas adicionadas ao pó de café puro. Embora tenham obtido bons resultados, o tempo gasto na preparação das amostras foi muito grande.

A espectroscopia no infravermelho, tal como NIR (infravermelho próximo) e MIR (infravermelho médio), tem sido usada em pesquisas sobre a autenticidade de muitos alimentos (Downey, 1996 e Wilson, 1990). Os métodos usados em espectroscopia no infravermelho são simples, de custo-benefício

favorável, rápidos e não destrutivos e poderiam servir como uma ferramenta potencial para a detecção de adulteração econômica ou para análises de rotina de calibração adequada e procedimentos de validação com protocolos de aquisição de dados estabilizados (Paradkar et al., 2003).

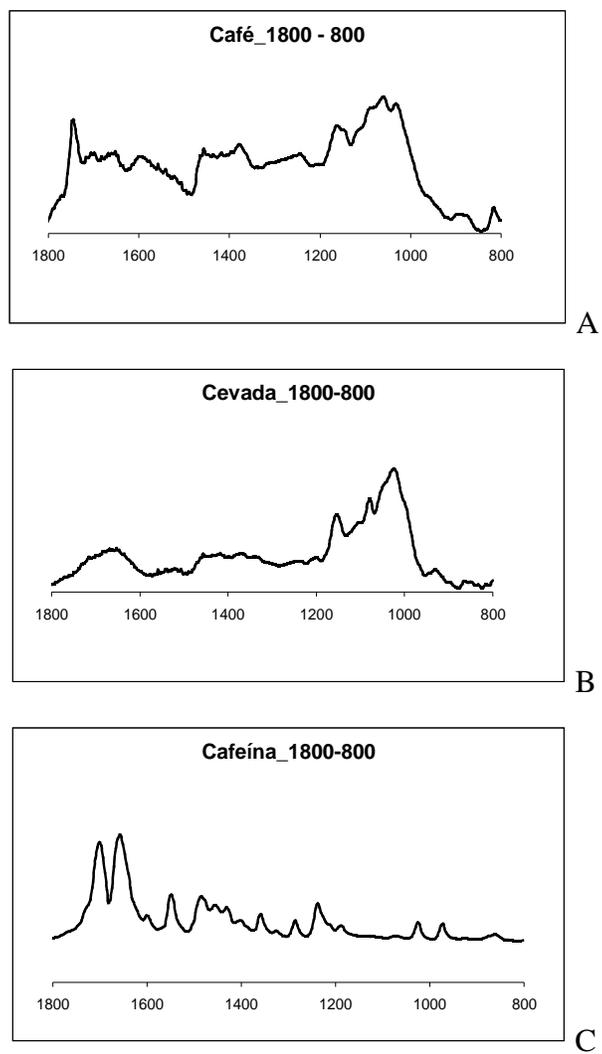
Briandet et al (1996a) desenvolveram um método para a detecção de materiais desconhecidos em café solúvel por meio da espectroscopia no infravermelho, equipado com sistemas: DRIFT (Reflectância Difusa) e ATR (Reflectância Total Atenuada). O processamento dos dados foi realizado usando diferentes métodos de análise multivariada, como: LDA (*Linear discriminant analysis*), PLS (*Partial least squares*) e ANN (*Artificial neural network*).

O café é um alimento de grande complexidade química. Os compostos mais importantes encontrados em café liofilizado são carboidratos, minerais, cafeína, ácidos clorogênicos, proteínas e lipídeos. Uma análise do espectro infravermelho desta matriz é muito difícil. Porém, sabe-se que a cafeína apresenta bandas marcantes na região entre  $1.750-1.550\text{ cm}^{-1}$  (Briandet et al., 1996b). Como a cevada não possui cafeína, é possível observar algumas diferenças entre os espectros do café e cevada nessa região.

Na Figura 1 estão representados os espectros do café (1 – A), da cevada (1 – B) e da cafeína (1 – C), com o intuito de se destacar as regiões com maiores diferenças entre esses produtos. Nota-se uma diferença na região de  $1.600$  a  $1.800\text{ cm}^{-1}$  que pode ser atribuída à ausência de cafeína na cevada.<sup>1</sup>

---

<sup>1</sup> Os espectros da Figura 1 foram obtidos pelo autor deste trabalho por meio da técnica de reflectância total atenuada (ATR), usando um espectrofotômetro baseado em transformada de Fourier, com uma resolução de  $8\text{ cm}^{-1}$  e detector DTGS.



**FIGURA 1.** Espectros de infravermelho na região espectral de  $1.800\text{-}800\text{ cm}^{-1}$  do café (A), da cevada (B) e da cafeína (C)

## 2.4 Espectroscopia

### 2.4.1 Transmitância

A espectroscopia estuda a interação da matéria com a radiação eletromagnética. Dependendo da energia de radiação incidente, um ou mais dos seguintes processos podem ocorrer: reflexão, espalhamento, fluorescência/fosforescência, reação fotoquímica ou absorção (Amaya, 1999).

A absorção da energia da radiação eletromagnética pode ocorrer devido a transições eletrônicas entre orbitais atômicos ou moleculares ou a mudanças de estados rotacionais ou vibracionais das moléculas. Na região do infravermelho ocorrem absorções devido a estas duas últimas transições e, para o estado condensado, as rotacionais podem ser negligenciadas por apresentarem intensidades muito baixas (Amaya, 1999).

A região do infravermelho pode ser ainda dividida em três sub-regiões distintas: o infravermelho distante, o infravermelho próximo (NIR, *Near Infrared*), que compreende a radiação com números de onda entre 4.000 e 15.000  $\text{cm}^{-1}$  e o infravermelho médio (MIR, *Mid Infrared*), entre 200 e 4.000  $\text{cm}^{-1}$ . No infravermelho médio, ocorrem as transições fundamentais, assim chamadas porque a molécula passa do estado fundamental (ou de maior energia) para o estado excitado imediatamente superior. Esta característica faz da espectroscopia no infravermelho médio uma técnica excelente na caracterização de compostos orgânicos, pois cada ligação característica de um grupo funcional apresenta uma banda de vibração em uma frequência específica. Entretanto, quantificações com esta região do espectro só foram possíveis com a utilização de técnicas quimiométricas (Amaya, 1999).

#### **2.4.2 Reflectância difusa (DRIFT)**

Inicialmente, o uso da reflectância difusa foi restrito às regiões do ultravioleta, visível e infravermelho próximo (Chalmers & Mackenzie, 1985). Com o aperfeiçoamento dos equipamentos (maior razão sinal ruído, aumento na sensibilidade dos detectores, maior sensibilidade e estabilidade da fonte) e acessórios de reflectância difusa (com habilidade de capturar o máximo da reflexão difusa) a técnica tem sido utilizada como prática de rotina na região do infravermelho médio, oferecendo inúmeras vantagens (Coleman, 1993).

O rápido aumento na popularidade da espectroscopia no infravermelho por reflectância difusa surgiu devido ao fornecimento dos espectros de pós, sólidos e espécies adsorvidas em sólidos. Também, novos métodos de amostragem têm introduzido abrasivos para coletar espectros de amostras que não são pós, aumentando a utilidade da técnica sem a necessidade de extensiva preparação da amostra. Além de reduzir o tempo de preparação da amostra, obtêm-se informações de materiais na forma mais próxima da original (Mackenzie, 1988).

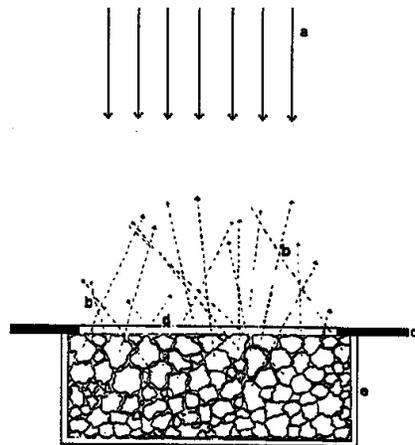
O potencial da técnica foi reconhecido como um novo método de análise para aplicações quantitativas. A aplicação da técnica requer um entendimento teórico de suas limitações para a obtenção de espectros de boa qualidade. Resultados quantitativos podem ser obtidos se for feita uma preparação adequada da amostra (Coleman, 1993).

Vários acessórios comerciais são empregados para obter os espectros nos quais a radiação espalhada é coletada em montagem óptica própria e é dirigida para o detector (Griffiths, 1975 e Teverucht & Griffiths, 1989). Na reflectância, a referência tomada é um corpo branco que reflete praticamente toda a radiação incidente. No entanto, o caminho percorrido pela luz não pode ser controlado e ele varia de acordo com o tamanho da partícula e o índice de refração (Fidêncio, 2001).

Então, reflectância pode ser definida como sendo a razão das potências ou fluxos (luminosos) da radiação refletida ( $I_R$ ) e da radiação de referência ( $I_0$ ) numa superfície (Fidêncio, 2001).

$$R=I_R/I_0 \quad (1)$$

A reflectância difusa ocorre quando a radiação penetra na camada superficial das partículas (Figura 2), excitando de modo vibracional a molécula a ser analisada e, então, tem-se um espalhamento em todas as direções (Fidêncio, 2001).



**FIGURA 2** Reflectância difusa: **a** = radiação incidente, **b** = radiação refletida difusamente, **c** = capa da cela de plástico, **d** = janela e **e** = cela de armazenamento da amostra

Na ausência de absorção, com espessura infinita e com múltiplos espalhamentos, um máximo de radiação retorna à superfície. Quando ocorre absorção, a radiação é transmitida e espalhada nas interfaces e no interior da amostra antes de retornar à superfície (Messerschmidt, 1999).

As medidas absolutas da reflectância dependem das condições experimentais, do ângulo de incidência e de reflexão, da espessura da amostra, da temperatura, do estado da superfície da amostra, etc. A reflectância é, portanto, obtida comparando-se a quantidade medida com uma referência, tal como brometo de potássio (KBr), sulfato de bário ( $\text{BaSO}_4$ ), óxido de magnésio, etc. (Fidêncio, 2001).

O tamanho, a forma das partículas, o espaço livre entre elas e o grau de compactação afetam a quantidade (concentração) do material por meio do qual a radiação é transmitida. Alguns desses fatores também afetam o caminho entre as interfaces individuais. A morfologia da superfície da amostra é uma consideração importante na magnitude da componente especular (Wendlant & Hecht, 1966, Wetzal, 1983 e Olinger & Griffiths, 1993).

Chalmers e Mackenzie (1985) chamam a atenção para o fato de que a razão entre as intensidades das bandas não permanece constante com a alteração do tamanho das partículas.

Sendo assim, a reflectância, de acordo com o tipo de reflexão pode ser:

- a) regular ou especular, em que a reflexão segue as leis da óptica de reflexão e a razão  $R$  é dada pela radiação regularmente refletida em relação ao fluxo incidente;
- b) difusa;
- c) total: quando a reflexão é do tipo especular e difusa (Fidêncio, 2001).

### **2.4.3 Reflectância total atenuada (ATR)**

Reflectância total atenuada é uma técnica de análise que normalmente requer uma mínima preparação da amostra para análise rápida e fácil para obtenção de espectros no infravermelho de materiais que são espessos e fortemente absorventes para serem analisados por espectroscopia de transmitância. É útil para dar informações sobre as propriedades da superfície de um material, incluindo identificação, modificação e adsorção em superfícies (Coleman, 1993).

Existem múltiplos elementos de ATR e uma variedade de acessórios é desenvolvida para obterem-se espectros, em geral, de materiais líquidos, sólidos e viscosos para larga faixa de tipos de amostras, incluindo filmes, resíduos, papel, revestimentos sobre papel, pós, tintas, tecidos, espumas, minerais, vidros, etc. (Mirabela, 1985).

O fenômeno da reflexão interna é observado sob certas condições. Quando a radiação entra em um prisma feito com um índice de refração alto em relação ao meio externo (cristal de ATR), a radiação será refletida totalmente em seu interior. Esta reflexão interna cria uma onda evanescente, a qual se estende na amostra em contato com o cristal. Na região do infravermelho, onde a amostra absorve energia, a onda evanescente será atenuada (Messerschmidt, 1999).

Não existe nenhum material universal para ATR. Um dos materiais mais comumente usados é o ZnSe, cujo índice de refração é 2,4. Este material é insolúvel em água (Coleman, 1993).

Dois parâmetros controláveis são o ângulo de incidência e a escolha do elemento de ATR. A espessura efetiva e a profundidade de penetração aumentam quando o ângulo crítico é aproximado, podendo-se variar a profundidade de penetração ajustando o ângulo de incidência. Contudo, o

número de reflexões também varia com o ângulo de incidência (Coleman, 1993 e Mirabela, 1985).

Um problema encontrado é a dificuldade de se obter boa reprodutibilidade no contato das amostras com o elemento de ATR. Esse efeito é observado na variação da intensidade das bandas com a pressão aplicada; aumentando a pressão, a eficiência de contato é aumentada e, conseqüentemente, as intensidades das bandas aumentam (Coleman, 1993 e Mirabela, 1985).

A área de contato entre o cristal e a amostra é um fator que também influencia a intensidade das bandas. Toda a área do cristal deve estar em contato com a amostra, pois irregularidades na superfície da amostra dificultam um bom contato (Coleman, 1993 e Mirabela, 1985).

## **2.5 Quimiometria**

A moderna instrumentação de análises químicas é capaz de gerar uma quantidade considerável de dados, sobre uma única amostra, em um curto espaço de tempo: um espectrômetro pode registrar sinais provenientes de mais de mil comprimentos de onda ou um único cromatograma pode apresentar mais de cem picos (Robinson, 2001 e Voet et al., 2000). Assim, para que informação útil seja obtida deste grande volume de dados, é necessário que se utilizem técnicas matemáticas adequadas, sendo a quimiometria um dos campos de estudo da química que fornece tais ferramentas (Robinson, 2001 e Lajolo & Nutti, 2003).

A quimiometria pode ser definida como a aplicação de métodos matemáticos e estatísticos a dados de origens distintas para a obtenção de informação química. Consiste de um conjunto de técnicas de cálculo com o objetivo de promover a obtenção de informação útil de um conjunto complexo de dados, incluindo conceitos de planejamento experimental, pré-processamento de dados, estatística e análise multivariada (Robinson, 2001 e Gould, 1992).

As técnicas quimiométricas podem ser genericamente divididas em três classes distintas: a análise exploratória de dados, a construção de modelos quantitativos de calibração e a construção de modelos qualitativos de classificação. A análise exploratória utiliza basicamente os métodos de análise de componentes principais (PCA) e a análise de agrupamentos hierárquicos (HCA), enquanto que os modelos de classificação quantitativos mais utilizados são os de regressão por componentes principais (PCR) e regressão por mínimos quadrados parciais (PLS) (Robinson, 2001 e Voet et al., 2000).

Dentre as técnicas quimiométricas, as de calibração multivariada destacam-se porque possibilitam a realização de determinações quantitativas de compostos presentes na ordem de porcentagem ou até de micro-constituintes, em matrizes complexas. Outras vantagens de sua aplicação são a redução considerável da necessidade de preparo das amostras, bem como da utilização de reagentes químicos por vezes nocivos ao homem ou ao meio ambiente, e a possibilidade da utilização de métodos não-destrutivos que podem ser aplicados *on line* em processos fabris. Além disso, a aplicação de métodos quimiométricos reduz a necessidade de seletividade instrumental pela utilização de um número relativamente grande de sinais, abrangendo uma gama maior de informação do modelo matemático (Robinson, 2001 e Voet et al., 2000).

Devido à sua natureza inerentemente multivariada, as técnicas quimiométricas utilizam-se extensivamente da manipulação de vetores e matrizes, objeto de estudo do ramo da matemática denominado álgebra linear (Vogel, 1992 e Sadler et al., 1990).

### **2.5.1 Quimiometria aplicada à química de alimentos**

Métodos quimiométricos são largamente empregados em química de alimentos, não apenas na forma de modelos de calibração, mas também para a

realização de análises exploratórias e construção de modelos de classificação (Pedro, 2004).

Análises exploratórias são comumente utilizadas para determinar possíveis relações entre propriedades físicas e constituintes químicos em matrizes alimentícias, correlações entre as variáveis instrumentais ou subjetivas ou na definição do comportamento dos alimentos em função de variações em sua composição ou na forma de seu processamento (Robinson, 2001 e Voet et al., 2000).

Métodos de calibração multivariada encontraram terreno fértil para o desenvolvimento de aplicações em alimentos devido, principalmente, à complexidade e variedade de matrizes que esses produtos apresentam, tornando demorados e caros os métodos clássicos de análise química. Mesmo os instrumentos mais sofisticados dependem de complicados procedimentos de tratamento de amostras. Assim, uma das principais vantagens da aplicação da calibração multivariada à “química de alimentos” é a redução ou até eliminação do tratamento de amostras. Essas etapas são normalmente substituídas por técnicas matemáticas de tratamento do sinal instrumental (Robinson, 2001 e Voet et al., 2000).

A quimiometria também começou a se popularizar entre os químicos de alimentos, graças aos vários instrumentos que já são fornecidos com softwares de tratamento de dados. Também existem programas independentes que apresentam interfaces gráficas extremamente amigáveis e de fácil utilização (Pedro, 2004).

Devido à complexidade das matrizes alimentícias, os métodos de análise são, em sua grande maioria, destrutivos, impedindo sua aplicação para controle *on line* de processo. Desse modo, a utilização de métodos espectroscópicos, aliados à calibração multivariada, é atualmente extensivamente estudada em

química de alimentos, visando, principalmente, à redução do tempo e da extensiva manipulação das amostras (Robinson, 2001 e Voet et al., 2000).

Existem, na literatura, relatos de inúmeros trabalhos científicos nesta área. Como exemplo, pode-se citar o trabalho desenvolvido por Ferrão et al. (2004), que utilizaram as técnicas de espectroscopia por reflexão no infravermelho próximo (NIRRS) e por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS), empregando-se o método de regressão multivariada por mínimos quadrados parciais (PLS) para a determinação simultânea de proteína e cinza em amostras de farinha de trigo.

Um outro exemplo, é o trabalho desenvolvido por Morgano et al. (2003) que utilizaram espectroscopia no infravermelho com transformada de Fourier (FTIR) com acessório de reflectância total atenuada (ATR) para a quantificação de sacarose em açúcar cristal.

## **2.6 Aplicativo R**

O aplicativo R é uma linguagem e um ambiente para a computação estatística e gráficos. É um projeto GNU (do inglês, *general public license*) similar à linguagem e ambiente S, desenvolvido no *Bell Laboratories*, por John Chambers e colegas (Introdução..., 2005).

O R disponibiliza uma grande variedade de métodos estatísticos (modelagem linear e não linear, testes estatísticos clássicos, séries temporais, classificação, métodos multivariados, etc.) e técnicas gráficas. Um dos pontos fortes do R é a facilidade com que gráficos bem delineados e de alta qualidade para impressão podem ser produzidos com possibilidade de inclusão de fórmulas e símbolos matemáticos quando necessário (Introdução..., 2005).

O R é disponibilizado, sob os termos da GNU da *Free Software Foundation*, na forma de código aberto. O R pode ser compilado para um grande número de plataformas UNIX e similares (incluindo FreeBSD e Linux).

Também pode ser compilado para os sistemas operacionais Windows 9x/NT/2000 e MacOS (Introdução..., 2005).

## 2.7 Métodos qualitativos de análise

### 2.7.1 Análise de componentes principais (PCA)

A análise de componentes principais (PCA) (Geladi & Kowalski, 1996) é uma ferramenta quimiométrica utilizada na redução de dados e visualização de amostras. Comumente é empregada para identificação de grupos distintos e na seleção de amostras na construção de modelos para calibração multivariada. O objetivo de PCA é reduzir o número de variáveis envolvidas no modelamento, por meio de combinações lineares estabelecidas entre os dados (Fidêncio, 2001).

A PCA é um método para decompor uma matriz de dados  $\mathbf{X}$  de posto  $r$  ou “rank”  $r$ , como uma soma de matrizes de posto igual a 1, como pode ser visto na equação 2

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \dots + \mathbf{M}_h \quad (2)$$

sendo o posto um número que expressa a verdadeira dimensão de uma matriz (Fidêncio, 2001).

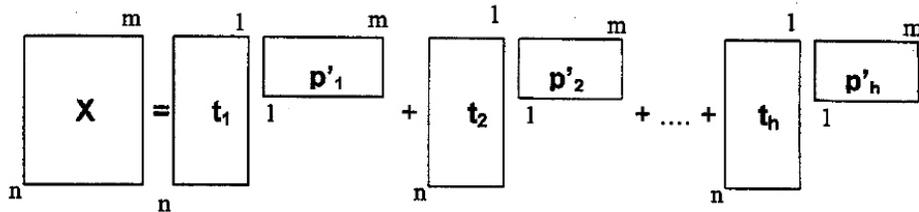
Assim, essas novas matrizes de posto igual a 1 são produtos de vetores chamados escores  $t_h$  e pesos ou “loadings”  $p_h$ , como mostrado na equação 3. Os escores e os pesos podem ser calculados par a par, por um processo iterativo (Fidêncio, 2001).

$$\mathbf{X} = t_1p_1 + t_2p_2 + \dots + t_hp_h \quad (3)$$

Esta equação pode ser representada, de maneira geral, para expressar a somatória dos vetores em matrizes, como mostra a equação 4 (Fidêncio, 2001).

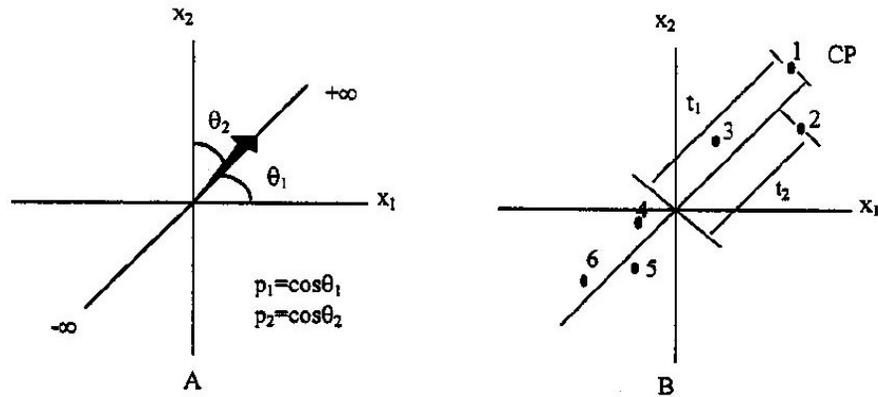
$$X = TP' \quad (4)$$

De uma forma mais clara, esta soma de vetores pode ser vista por uma representação gráfica de matrizes dos vetores, como mostra a Figura 3 (Fidêncio, 2001).



**FIGURA 3.** Representação matricial da somatória de produtos dos vetores escores e pesos

Uma melhor exemplificação da PCA pode ser vista na Figura 4, no qual os vetores  $t_h$  e  $p'_h$  estão representados no plano de duas variáveis  $x_1$  e  $x_2$ . A Figura 4A mostra uma componente principal, que é a reta que aponta para a direção de maior variabilidade das amostras da Figura 4B, aqui numeradas de 1 a 6. Na Figura 4A, estão representados os pesos ( $p_1$  e  $p_2$ ) que são os co-senos dos ângulos do vetor de direção; na Figura 4B estão representados os escores ( $t_1$  e  $t_2$ ) que são as projeções das amostras na direção da componente principal (Fidêncio, 2001).



**FIGURA 4.** Representação de uma componente principal para as variáveis  $x_1$  e  $x_2$

A análise das componentes principais é aplicada em casos nos quais as variáveis  $x$  são colineares. Nesta análise, redundâncias e pequena variabilidade de ruídos podem ser removidas (Martens & Naes, 1996). Os dados de FTIR são registrados em muitas frequências e, geralmente, existem muitas informações correlacionadas ou redundantes. Neste tipo de medida, essas informações podem ser comprimidas de maneira que possam reter as informações essenciais e mostrar mais facilmente cada variável individualmente. As informações essenciais não consistem em qualquer variável individual, mas em como as variáveis mudam umas em relação às outras, ex: covariância (Wise & Gallagher, 1998).

A análise de componentes principais é um meio útil de simplificar a série de dados. Pelo emprego dos fatores, podem-se utilizar gráficos com duas ou três dimensões para distinguir grupos de amostras com dados de alta

dimensionalidade. Desse modo, pode-se avaliar as tendências naturais de agrupamentos da série de dados pelo uso do PCA e gráfico dos escores e “loadings” dos fatores (Fidêncio, 2001).

### **2.7.2 Análise de agrupamentos hierárquicos (HCA)**

As técnicas de agrupamento são utilizadas com o objetivo de investigar as relações existentes dentro de um conjunto multivariado em que, *a priori*, nenhuma caracterização é conhecida. Ela pode ser de dois tipos: aglomerativa ou divisiva. No primeiro, cada amostra é considerada inicialmente um grupo e, de acordo com suas semelhanças, elas vão sendo agrupadas em subgrupos, até que todas elas formem um único grupo. Na técnica divisiva ocorre o contrário. Todas as amostras constituem um único grupo que será dividido em subgrupos, também de acordo com as similaridades entre as mesmas, até que cada amostra forme um único grupo (Verbeke et al., 1984).

A análise de agrupamentos hierárquicos (HCA) é uma técnica aglomerativa não supervisionada que examina as distâncias interpontuais entre todas as amostras do conjunto de dados e representa essa informação na forma de um gráfico bidimensional chamado dendograma. Por meio do dendograma podem-se visualizar os agrupamentos e similaridade entre as amostras e ou variáveis (Lauritsena & Bürgi, 1985).

A construção dos dendogramas é feita com base na proximidade existente entre as amostras no espaço. Isto é feito calculando-se a distância entre todas as amostras (agrupamentos) do conjunto, em pares, e então definindo uma matriz de similaridade cujos elementos são os chamados índices de similaridade. Um índice alto indica uma distância pequena entre dois agrupamentos e, portanto, uma alta similaridade (equação 5) (Beebe et al., 1998). A cada passo, os dois grupos mais similares vão se juntando e o processo vai se repetindo até que forme um único agrupamento (Lauritsena & Bürgi, 1985).

$$s_{ij} = d_{ij}/d_{max} \quad (5)$$

em que:  $s_{ij}$  é a similaridade entre duas amostras (ou agrupamentos)

$d_{ij}$  é a distância euclidiana entre as mesmas;

$d_{max}$  é a maior distância encontrada entre todas as amostras do conjunto (Lauritsena & Bürgi, 1985).

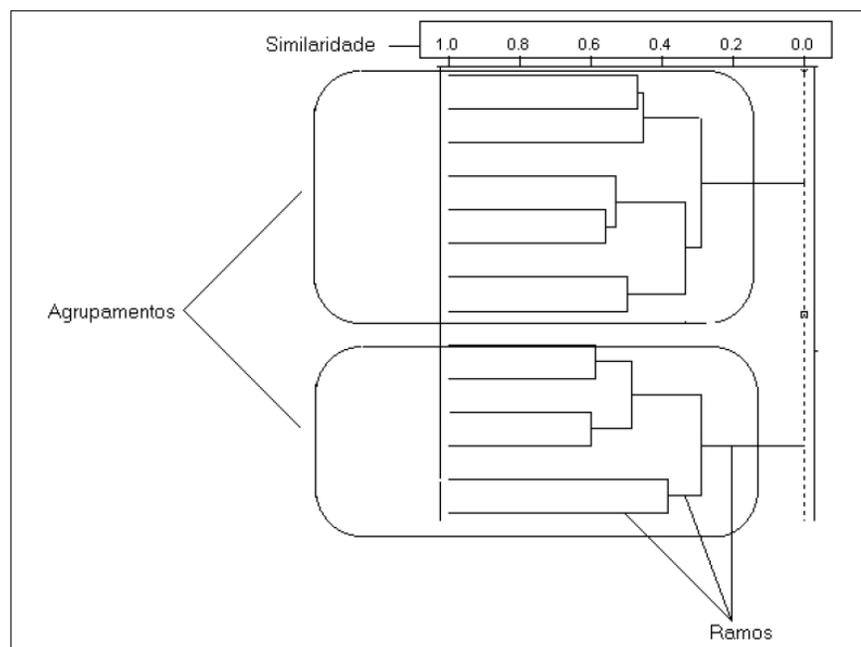
A escolha da proximidade entre dois agrupamentos pode ser feita basicamente, por três métodos: do vizinho mais próximo, do vizinho mais distante ou pela média (que pode ser calculada de várias maneiras) e a mais simples medida de similaridade entre pontos num conjunto de dados é sua distância Euclidiana (Lauritsena & Bürgi, 1985).

A conexão pelo vizinho mais próximo é feita buscando-se inicialmente, a maior similaridade (ou menor distância) entre dois grupos. A partir daí, a matriz de similaridade vai sendo continuamente atualizada, sempre procurando as menores distâncias e aproximando os agrupamentos mais similares, até que um único agrupamento seja formado (Beebe et al., 1998).

Pelo método do vizinho mais distante, busca-se sempre a maior distância entre as amostras e, dentre estas, o par de maior similaridade é agrupado (Bratchell, 1989).

Os agrupamentos encontrados podem ser então visualizados por um dendograma.

No dendograma exemplificado na Figura 5, as amostras são listadas do lado esquerdo do gráfico e os ramos indicam quais amostras estão em dado agrupamento. O eixo horizontal é uma medida da distância interagrupamentos e a posição das linhas verticais indica as distâncias entre dois destes pontos (ou similaridade) (Beebe et al., 1998).



**FIGURA 5.** Exemplo de dendograma

Este método pode ser aplicado na identificação de grupos de um conjunto de dados para testar hipóteses de agrupamentos, na identificação de membros de um dado grupo ou na formação mais conveniente de agrupamentos com características um tanto quanto diversas (Bratchell, 1989).

## **2.8 Métodos quantitativos de análise**

### **2.8.1 Calibração multivariada**

A calibração multivariada tem como idéia básica relacionar dois grupos de dados ( $\mathbf{X}$  e  $\mathbf{Y}$ ), em que  $\mathbf{X}$  ( $n, m$ ) é o bloco dos dados experimentais contendo as variáveis independentes, representado em suas linhas pelas amostras do

conjunto e em suas colunas pelas variáveis medidas e o outro bloco ( $Y$ ), formado pela(s) variável(eis) dependente(s) (Ferreira et al., 1999). Dessa forma, na calibração multivariada, ao contrário do que ocorre na univariada, utilizam-se múltiplas respostas para relacionar com a(s) propriedade(s) medida(s) das amostras (Beebe et al., 1998).

Este tipo de calibração tem como principais vantagens o fato de permitir determinações simultâneas de mais de um analito de interesse, permitir determinações, mesmo na presença de interferentes e apresentar uma diminuição do erro estimado no modelo devido ao fato de ser um método que utiliza múltiplas variáveis.

A calibração, de maneira geral, pode ser dividida em duas etapas consecutivas:

- modelagem: estabelece-se uma relação entre o sinal medido e a propriedade que se deseja quantificar da amostra;
- validação: assegura que o modelo reflete o comportamento do analito.

A construção do modelo de calibração pode ser feita por vários métodos e a escolha entre cada um deles está relacionada ao perfil do conjunto de dados (Ribeiro, 2001).

Uma vez construído o modelo, este deve ser validado, ou seja, testado para garantir que os valores obtidos para a variável dependente sejam iguais ou bastante próximos dos experimentais (Parreira, 2003).

A validação pode ser de dois tipos: utilizando um conjunto externo ou por meio da validação cruzada (Parreira, 2003).

No primeiro caso, o conjunto de dados é dividido em dois subconjuntos de calibração e validação. O modelo é construído usando as amostras do conjunto de calibração e depois é validado utilizando o conjunto com as amostras restantes (Ribeiro, 2001). Os resultados obtidos para a variável dependente nessa etapa são comparados aos valores experimentais, sendo os

resíduos calculados. A eficácia do modelo construído será maior quanto menores forem os resíduos encontrados (Ribeiro, 2001).

O método de validação cruzada funciona da seguinte forma: a matriz de dados é dividida em pequenos grupos. Um determinado grupo é então removido da matriz original e esta, agora reduzida, é decomposta normalmente em escores e pesos, e o modelo de calibração é estabelecido sobre esse novo conjunto. A partir deste, a propriedade das amostras removidas será prevista e os resíduos obtidos entre os valores reais e os estimados são comparados. Isso é repetido para todos os pequenos grupos do conjunto de dados (Ribeiro, 2001).

### **2.8.2 Regressão por mínimos quadrados parciais (PLS)**

O método dos mínimos quadrados parciais (PLS) (Martens & Naes, 1989) vem sendo utilizado no modelamento de dados. O PLS pode ser usado na modelagem de sistemas que possuem comportamento linear (PLS-linear) ou não (PLS\_polinomial). O método PLS pode ser dividido em duas etapas básicas. A primeira consiste em montar as matrizes de dados  $\mathbf{X}$  a partir dos dados espectrais (variáveis independentes) e  $\mathbf{Y}$  a partir das propriedades de interesse (variáveis dependentes). A segunda etapa consiste em obter-se a relação entre as matrizes  $\mathbf{X}$  e a matriz  $\mathbf{Y}$ , que possui a propriedade de interesse (Fidêncio, 2001).

Os dados mais comumente tratados por este método de calibração são os obtidos instrumentalmente. O PLS é muito empregado em dados químicos por ser um método de resposta rápida e pela facilidade de entrada dos dados, para que sejam processados e a resposta desejada obtida (Fidêncio, 2001).

No PLS, a informação da matriz e ou vetor  $\mathbf{Y}$  (variáveis dependentes) é usada no cálculo das chamadas variáveis latentes (equivalentes às componentes principais) (Geladi & Kowalski, 1996).

Nesse processo, o ideal é encontrar o número ideal de variáveis latentes, que irá fornecer uma melhor previsão das novas amostras e que forneça um menor erro de previsão (Fidêncio, 2001).

### 3 MATERIAL E MÉTODOS

#### 3.1 Amostras de café

Foram usados grãos crus de café arábica (*Coffea arabica* L.), bebida dura, originados do Centro de Ensino Pesquisa e Extensão do Agronegócio Café (CEPECAFE) da Universidade Federal de Lavras (UFLA).

Os grãos foram torrados em um torrador Pinhalense tipo TP – 2, a uma temperatura de mais ou menos 200°C por, aproximadamente, 8 minutos. O grau de torração foi baseado em torra de padrão comercial.

Os grãos torrados foram moídos em moinho *Probat Emmerick*, com moagem nº 2 e embalados em frascos de vidro.

#### 3.2 Adulteração das amostras

Ao café torrado e moído foi adicionada cevada comercial *Superbom*® (torrada e moída) em diferentes proporções.

Os percentuais de cevada adicionados ao café torrado e moído foram: 0%, 0,5%, 3%, 5%, 7%, 10%, 15% e 20%.

Para cada amostra foram obtidas três repetições em oito tratamentos (% de cevada comercial), totalizando 24 amostras. A amostra contendo 3% de cevada para a análise de transmitância foi perdida, portanto, foram feitas três repetições de cada um dos sete teores de cevada, totalizando 21 amostras.

#### 3.3 Liofilização das amostras

As amostras foram submetidas à liofilização que foi realizada para se evitar a interferência das bandas de água nos espectros. Antes da leitura dos espectros, as amostras foram colocadas em um dessecador contendo sílica gel, por 24 horas.

### **3.4 Leituras no instrumento FTS 3000 MX**

Foram obtidas leituras no espectrofotômetro baseado em transformada de Fourier, modelo FTS 3000 MX, utilizando o detector DTGS, com uma resolução de  $4\text{ cm}^{-1}$ . Os espectros foram obtidos na região de 400 a  $4.000\text{ cm}^{-1}$  para as análises de transmitância e reflectância difusa (DRIFT) e na região de 600 a  $4.000\text{ cm}^{-1}$  para a análise de reflectância total atenuada (ATR). O branco utilizado neste equipamento foi o brometo de potássio (KBr).

Antes de iniciar a leitura espectral das amostras e do branco, estes foram triturados com um gral de ágata, para que as partículas ficassem mais homogêneas.

### **3.5 Análises espectrais**

As análises espectroscópicas foram realizadas na região do infravermelho médio.

#### **3.5.1 Análise de transmitância**

As amostras foram colocadas em um recipiente próprio para a formulação de pastilhas. O recipiente com a amostra foi submetido a uma pressão de  $3000\text{kgF/cm}^2$  por meio de uma prensa hidráulica. As amostras contendo 3% de cevada foram perdidas durante o processo de preparação. Foram produzidas três pastilhas de cada amostra, totalizando 21 amostras. As amostras foram diluídas em KBr, na proporção de 1% e foram pesadas para que ficassem mais homogêneas.

#### **3.5.2 Análise de reflectância difusa (DRIFT)**

As amostras foram colocadas em acessório apropriado e, em seguida, procedeu-se a leitura dos espectros na região espectral de 400 a  $4.000\text{ cm}^{-1}$ . Foram feitas três repetições de cada, totalizando 24 amostras.

### 3.5.3 Análise de reflectância total atenuada (ATR)

Foram construídas pastilhas com as amostras, para que as mesmas tivessem um maior contato com a superfície do cristal de seleneto de zinco (ZnSe). Um ângulo de incidência de 45° foi utilizado para a leitura dos espectros. Foram feitas três repetições de cada, totalizando 24 amostras. A leitura dos espectros foi realizada na região espectral de 600 a 4000  $\text{cm}^{-1}$ .

### 3.6 Organização dos dados

A análise multivariada requer a organização do conjunto de dados em questão numa matriz ( $n \times p$ ), em que as linhas desta matriz representam o conjunto das amostras e as colunas o das variáveis medidas, ou seja, os comprimentos de onda.

Uma matriz de dados  $\mathbf{X}$ , contendo  $p$  medidas experimentais (variáveis) obtidas para  $n$  amostras, pode ser graficamente representada por  $n$  pontos num espaço  $p$ -dimensional.

Para conjuntos com muitas variáveis, a alta dimensionalidade apresentada pode dificultar o tratamento dos dados e uma ferramenta matemática que possibilite uma melhor visualização espacial dos mesmos (como por exemplo, a análise de componentes principais e análise de agrupamentos hierárquicos) torna-se de grande valia.

### 3.7 Processamento dos dados espectrais

As regiões onde ocorreram as principais diferenças entre os espectros foram de 450 a 1.900  $\text{cm}^{-1}$  e 2.450 a 3.800  $\text{cm}^{-1}$  para as análises de transmitância e DRIFT; e de 600 a 1.900  $\text{cm}^{-1}$  e 2.450 a 3.800  $\text{cm}^{-1}$  para a análise de ATR. Portanto, os dados utilizados nos cálculos foram relativos aos destas regiões.

A transformação dos dados mais comum é centrá-los na média. Esta operação foi feita na matriz de dados das medidas espectrais ( $\mathbf{X}$ ). Isso

corresponde a subtrair o valor de cada elemento da coluna pelo valor médio dos elementos dessa coluna, resultando numa matriz, em que todas as colunas têm a média zero. Esta operação, geralmente, reduz a dimensão do modelo (Thomas, 1994).

### 3.8 Análise estatística

As análises realizadas foram:

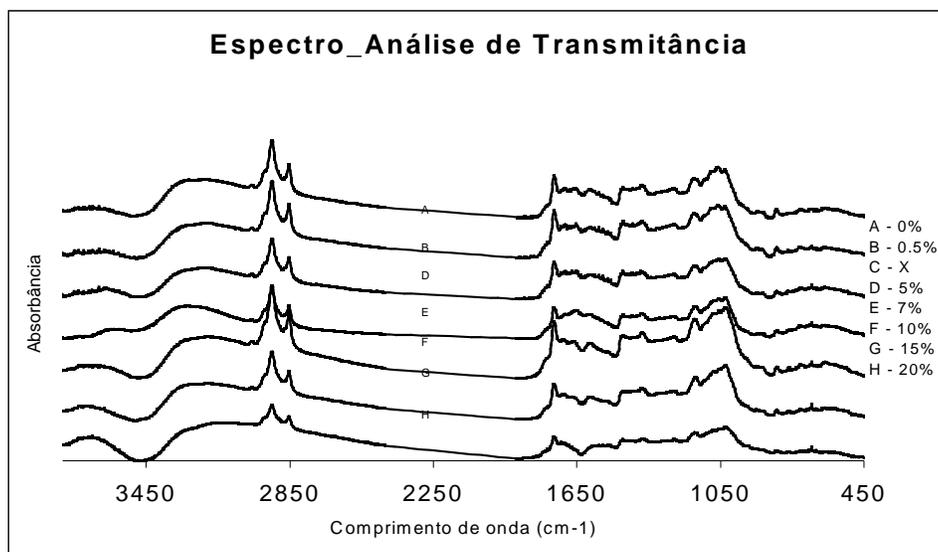
- análise de componentes principais (PCA);
- análise agrupamentos hierárquicos (HCA);
- regressão por mínimos quadrados parciais (PLS).

O software utilizado para o tratamento dos dados foi o R. Para a análise de PCA, a função utilizada foi a *prcomp* (R Development Core Team, 2004). Os dados foram centrados na média antes da realização dos cálculos. Para a análise de HCA, utilizou-se a função *hclust* (R Development Core Team, 2004). A distância utilizada foi a euclidiana e o método utilizado foi o método do vizinho mais distante (*complete linkage*). Já para a análise regressão, a função utilizada foi a *mvr*, por meio do pacote *pls.pcr* (Ron Wehrens, 2004) e o método utilizado foi o SIMPLS, utilizando validação cruzada como validação do modelo.

## 4 RESULTADOS E DISCUSSÃO

### 4.1 Análise de transmitância

Na Figura 6 estão representados os espectros do café e das amostras com cevada, em diferentes proporções, obtidos por transmitância. Pequenas diferenças podem ser observadas nas regiões de 450 a 1.900  $\text{cm}^{-1}$  e de 2.450 a 3.800  $\text{cm}^{-1}$ . No entanto, essas diferenças são tão pequenas e os espectros tão complexos, que estatísticas multivariadas são necessárias para obter uma diferenciação entre as amostras.

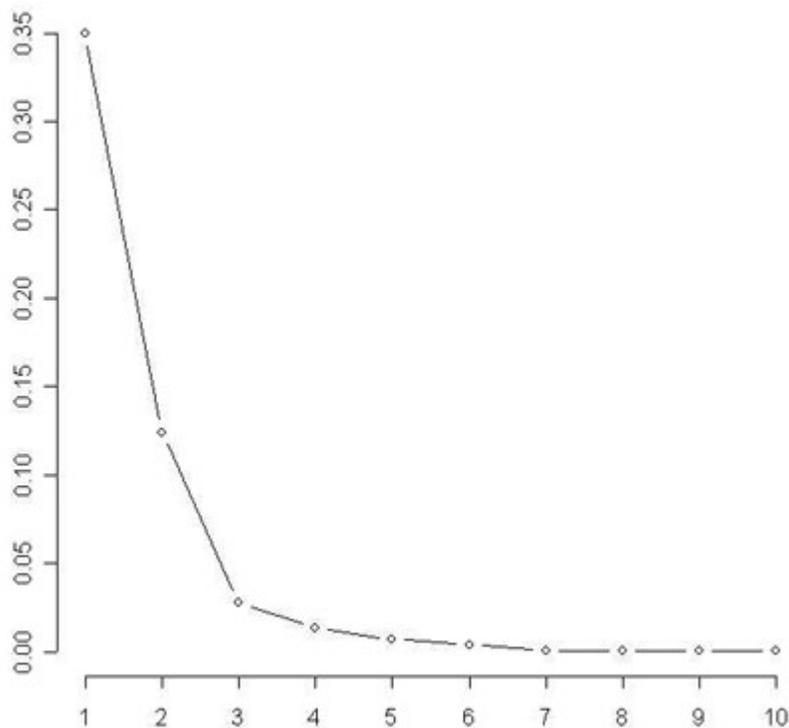


**FIGURA 6.** Espectros do café e das amostras com cevada, em diferentes proporções, obtidos por FTIR por meio da análise de transmitância

#### 4.1.1 Análise de componentes principais (PCA)

A análise de componentes principais (PCA) foi aplicada aos dados obtidos na análise de transmitância.

Por meio do gráfico de cotovelo (Figura 7), é possível determinar o número de componentes principais necessárias para resumir a variação amostral total. Este gráfico é um processo de determinação do número apropriado de componentes a ser retido.



**FIGURA 7.** Gráfico de cotovelo para as variâncias das dez primeiras componentes principais dos dados obtidos por FTIR (nas regiões espectrais de 450-1.900 e 2.450-3.800  $\text{cm}^{-1}$ ), por meio da análise de transmitância

Pela análise da Figura 7 observa-se que um cotovelo é formado na posição sete. Isso significa que as componentes acima de 7 possuem a mesma magnitude e são relativamente pequenas. Sendo assim, as sete primeiras componentes são suficientes para resumir a variação amostral total.

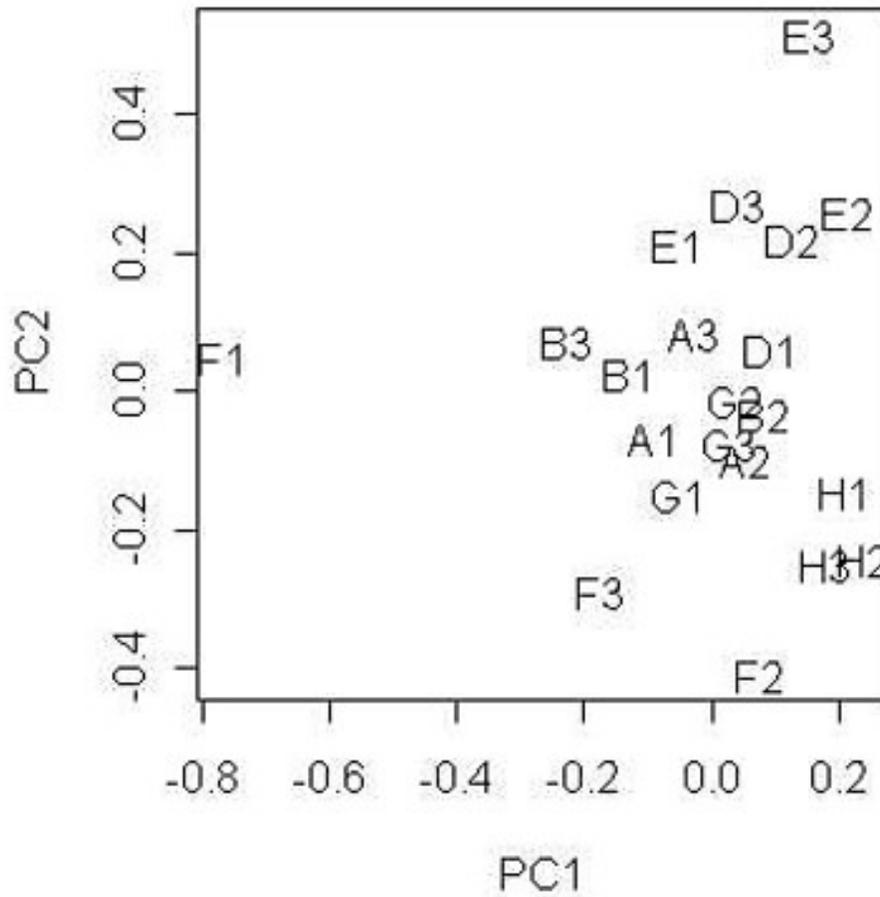
A porcentagem de variância das dez primeiras componentes principais está apresentada na Tabela 1.

**TABELA 1.** Porcentagem de variância e porcentagem de variância acumulada para a análise de PCA dos dados obtidos por FTIR (nas regiões espectrais de 450-1.900 e 2.450-3.800  $\text{cm}^{-1}$ ), por meio da análise de transmitância em amostras de café puro e café com cevada.

PC escores	% de variância	% de variância acumulada
1	66,30	66,30
2	23,40	89,70
3	5,19	94,89
4	2,50	97,39
5	1,35	98,74
6	0,77	99,51
7	0,15	99,66
8	0,09	99,75
9	0,07	99,82
10	0,06	99,88

Para a análise de transmitância, 7 componentes principais explicam 99,66% da variabilidade dos dados analisados, dos quais 66,30% são explicados pela componente principal 1 (PC1), 23,40% pela componente principal 2 (PC2) e 5,19% pela componente principal 3 (PC3).

A dispersão das amostras é mostrada por meio do gráfico de escores da primeira (PC1) contra a segunda componente principal (PC2) (Figura 8).



A	B	C	D	E	F	G	H
0%	0,5%	X	5%	7%	10%	15%	20%

**FIGURA 8.** Gráfico de dispersão da primeira (PC1) e segunda (PC2) componentes principais. (As letras correspondem às diferentes proporções de cevada e os índices 1, 2 e 3 referem-se à repetição)

O gráfico de dispersão não apresenta uma separação especial das diferentes amostras. Apenas o café, com 10% de cevada (F1, F2 e F3) está separado das demais e apresenta dispersão bem irregular. As amostras de café com 20% de cevada (H1, H2 e H3) estão bem agrupadas e separadas das demais. Já para as outras amostras não é possível verificar agrupamento, não permitindo uma separação conveniente das mesmas. Sendo assim, pode-se dizer que a técnica de transmitância não foi capaz de separar o café puro do café adulterado; porém, é possível detectar amostras com teores de cevada acima de 20%.

#### **4.1.2 Análise de agrupamentos hierárquicos (HCA)**

Após a PCA, foi realizada a análise de agrupamentos hierárquicos (HCA) utilizando-se, assim, um número reduzido de variáveis. A matriz de similaridade foi obtida pelo cálculo da distância euclidiana e o método utilizado foi o do vizinho mais distante (*complete linkage*). Os grupos formados estão representados na Figura 9, em um dendograma completo.



A	B	C	D	E	F	G	H
0%	0,5%	X	5%	7%	10%	15%	20%

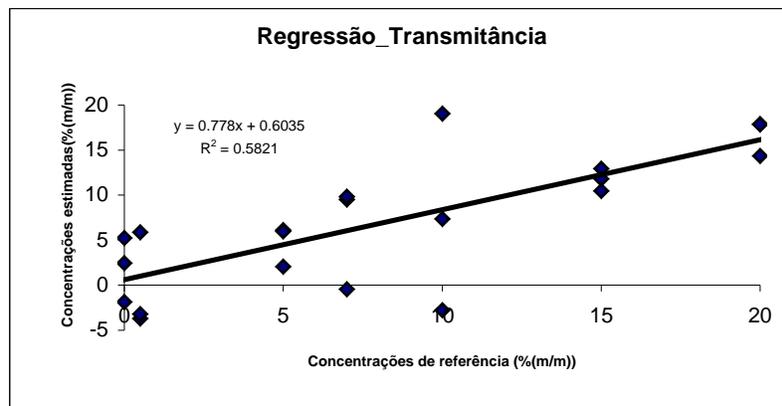
**FIGURA 9.** Dendrograma completo de 21 amostras de café puro e café com cevada, obtido por distância euclidiana e pelo método do vizinho mais distante

Como já havia sido observado na análise de componentes principais, as amostras com 10% de cevada (F1, F2 e F3) distribuem-se de forma irregular, não permitindo agrupá-las. Porém, observa-se um grupo formado pelas amostras com 20% de cevada (H1, H2 e H3), o que confirma o resultado observado na dispersão das amostras no gráfico de escores. Não foi possível separar o café

puro do café adulterado por meio da técnica de HCA, confirmando os resultados obtidos na PCA.

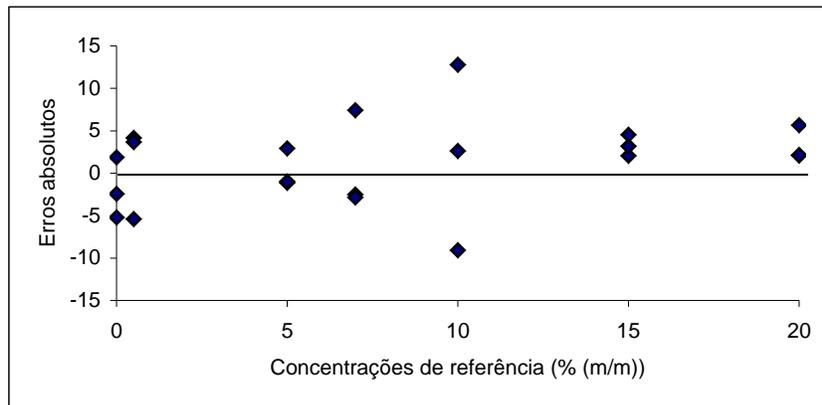
#### 4.1.3 Regressão por mínimos quadrados parciais (PLS)

A regressão por mínimos quadrados parciais (PLS) foi realizada para as amostras de café puro e de café com cevada. O modelo PLS construído utilizou as variáveis originais e o método utilizado foi o *SIMPLS*, usando validação cruzada. O modelo de regressão ótimo foi obtido com 5 PLS escores (variáveis latentes). O modelo PLS aplicado não apresentou uma boa correlação entre as concentrações de cevada estimadas e as concentrações reais de cevada. O  $r^2$  observado foi de apenas 58%. Os valores de concentração de cevada estão plotados contra os valores estimados no modelo na Figura 10.



**FIGURA 10.** Valores de referência versus valores estimados pelo modelo PLS com cinco variáveis latentes (Método *SIMPLS* e validação cruzada)

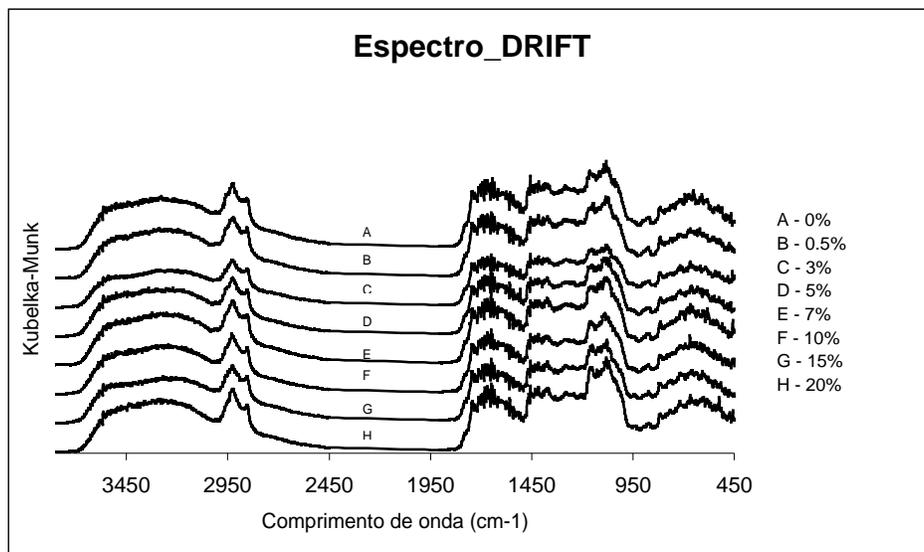
Na Figura 11, estão representados os erros absolutos do modelo PLS aplicado. O erro absoluto é a diferença entre o valor verdadeiro ( $y$ ) e o valor estimado ( $\hat{y}$ ) (erro absoluto =  $y - \hat{y}$ ). Observa-se que os valores estimados são, na maioria, inferiores aos valores verdadeiros. Os maiores problemas são verificados nas concentrações mais baixas. Observa-se que, para concentrações de cevada menores que 15%, o erro é muito significativo, superando até mesmo o valor verdadeiro.



**FIGURA 11.** Gráfico de erros absolutos contra as concentrações de referência

#### 4.2 Reflectância difusa (DRIFT)

Os espectros obtidos por reflectância difusa (DRIFT) para o café e para as amostras com cevada em diferentes proporções são mostrados na Figura 12. Algumas diferenças podem ser observadas nas regiões espectrais de 450-1.900 e 2.450-3.800  $\text{cm}^{-1}$ . Porém, essas diferenças são tão pequenas e os espectros tão complexos, que estatísticas multivariadas são necessárias para se obter uma melhor diferenciação entre as amostras.

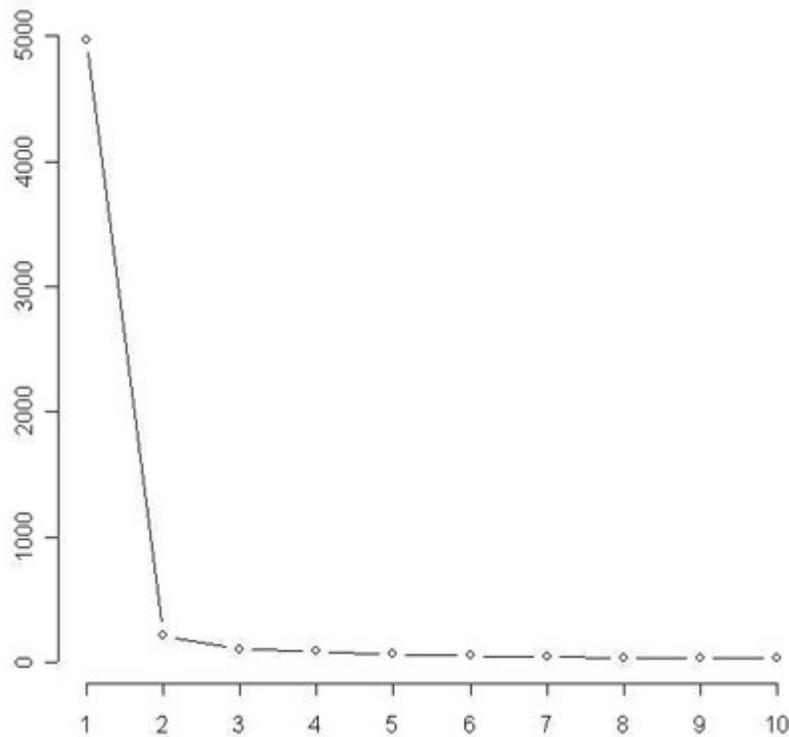


**FIGURA 12.** Espectros de infravermelho obtidos por reflectância difusa (DRIFT) para café puro e café com cevada, em diferentes proporções

#### 4.2.1 Análise de componentes principais (PCA)

A análise de componentes principais (PCA) foi aplicada a todos os dados obtidos por DRIFT nas regiões de 450-1.900 e 2.450-3.800 cm<sup>-1</sup>.

O gráfico de cotovelo (Figura 13) é um processo de determinação do número apropriado de componentes a ser retido. Por meio deste, é possível determinar o número de componentes principais necessárias para resumir a variação amostral total.



**FIGURA 13.** Gráfico de cotovelo para as variâncias das dez primeiras componentes principais dos dados obtidos por reflectância difusa (DRIFT), nas regiões espectrais de 450-1.900 e 2.450-3.800  $\text{cm}^{-1}$

Pela análise da Figura 13 observa-se que um cotovelo é formado aproximadamente na posição quatro. Isso significa que as componentes acima de quatro possuem a mesma magnitude e são relativamente pequenas. Conclui-se então, que apenas quatro componentes principais são necessárias para se explicar o comportamento das variâncias das variáveis estudadas.

A porcentagem das dez primeiras componentes principais está apresentada na Tabela 2.

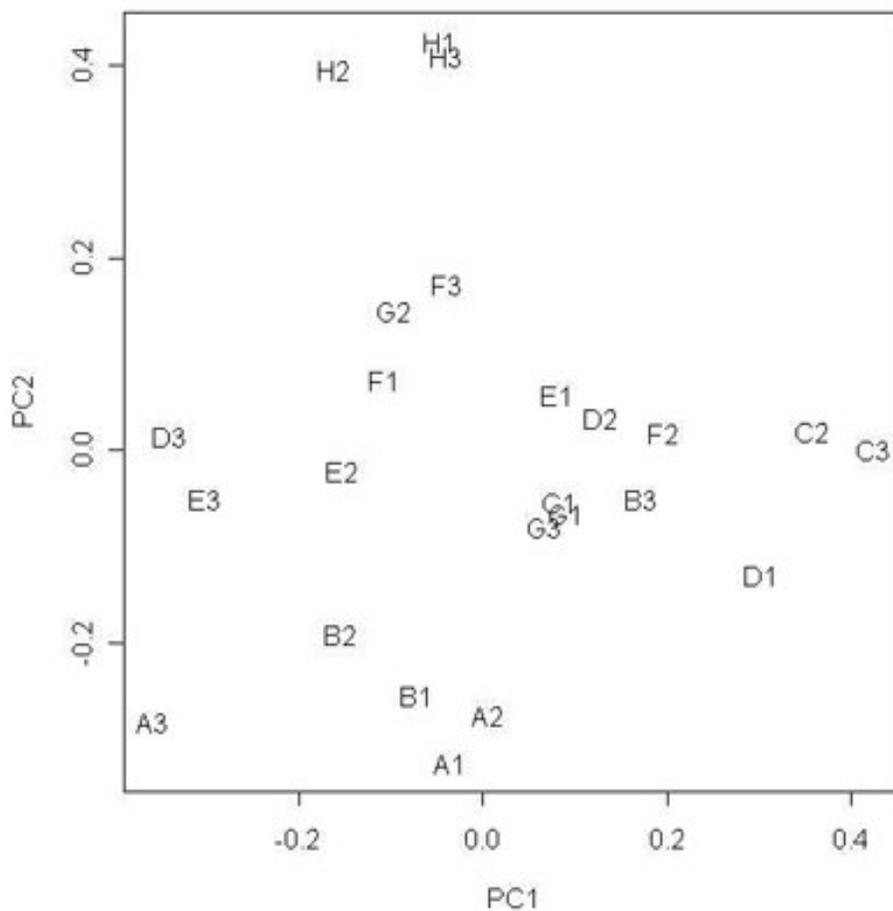
**TABELA 2.** Porcentagem de variância e porcentagem de variância acumulada para a análise de PCA, para os dados obtidos por DRIFT nas regiões espectrais de 450-1.900 e 2.450-3.800  $\text{cm}^{-1}$ .

PC escores	% de variância	% de variância acumulada
1	85,00	85,00
2	3,60	88,60
3	1,79	90,39
4	1,44	91,83
5	1,08	92,91
6	0,85	93,76
7	0,77	94,53
8	0,59	95,12
9	0,54	95,66
10	0,51	96,17

Em termos de percentual acumulado de variância, verifica-se que cerca de 92% da variância são explicados por quatro componentes principais.

No gráfico de escores da primeira (PC1) contra a segunda componente principal (PC2), está representada a dispersão das amostras analisadas por

reflectância difusa (DRIFT) nas regiões espectrais de 450-1.900 e 2.450-3.800 $\text{cm}^{-1}$  (Figura 14).



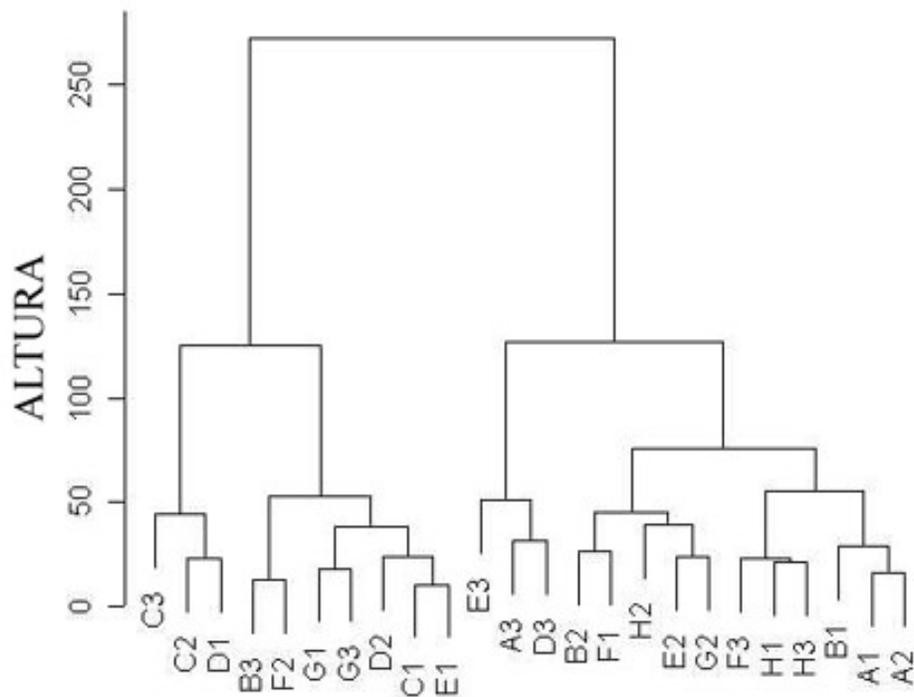
A	B	C	D	E	F	G	H
0%	0,5%	3%	5%	7%	10%	15%	20%

**FIGURA 14.** Gráfico de dispersão da primeira (PC1) e segunda (PC2) componentes principais (as letras correspondem às diferentes proporções de cevada e os itens 1,2 e 3 referem-se a repetição)

Na análise do gráfico de dispersão das amostras, observa-se que estas se encontram distribuídas de forma irregular. Não houve separação completa, porém, as amostras com 20% de cevada (H1, H2 e H3) e as amostras de café puro (A1, A2 e A3) encontram-se separadas das demais. Porém, estão agrupadas com duas amostras com 0,5% de cevada (B1 e B2). Portanto, foi possível separar o café puro de quase todas as amostras adulteradas, com exceção daquelas com teores de 0,5% de cevada. O modelo permitiu, ainda, a distinção das amostras com teor de 20% de cevada.

#### **4.2.2 Análise de agrupamentos hierárquicos (HCA)**

A Análise de Agrupamentos Hierárquicos foi realizada com base na PCA, utilizando-se, assim, um número reduzido de variáveis. A matriz de similaridade foi obtida por meio do cálculo da distância euclidiana e o método utilizado foi o do vizinho mais distante (*complete linkage*). A matriz utilizada para o cálculo da matriz de similaridade foi a matriz de escores. Os grupos formados estão representados em um dendograma completo (Figura 15).



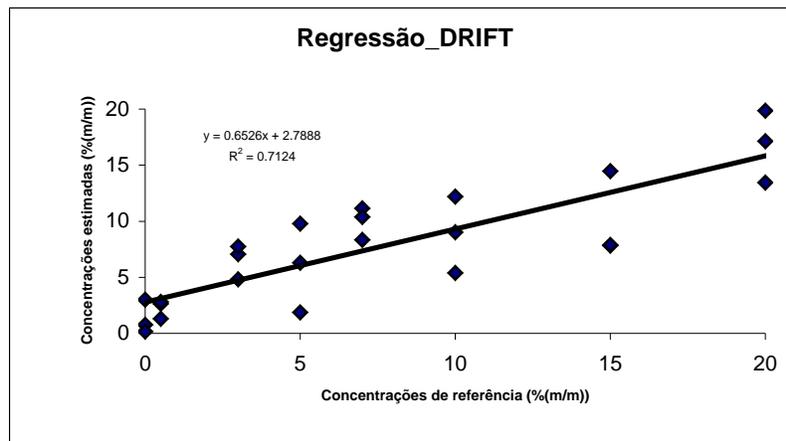
A	B	C	D	E	F	G	H
0%	0,5%	3%	5%	7%	10%	15%	20%

**FIGURA 15.** Dendrograma completo de 24 amostras, obtido por distância euclidiana e pelo método do vizinho mais distante (as letras correspondem às diferentes proporções de cevada e os itens 1, 2 e 3 referem-se à repetição)

Apesar da aparente distinção das amostras com 20% de cevada observada no gráfico de dispersão, não foi possível obter uma boa separação das amostras com a técnica de HCA.

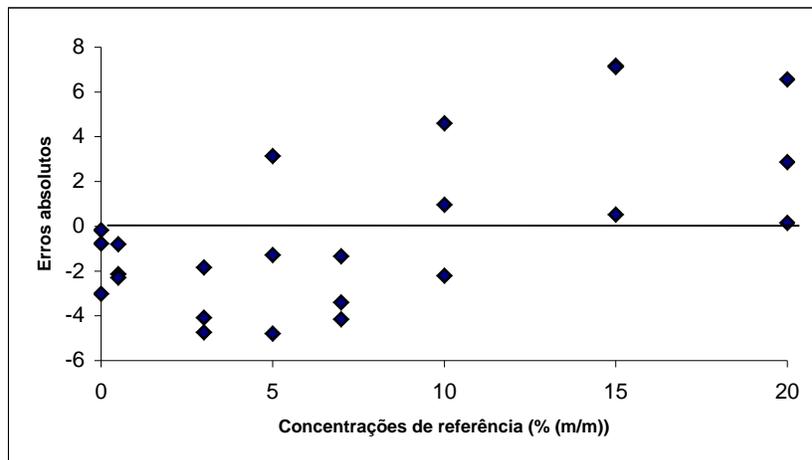
#### 4.2.3 Regressão por mínimos quadrados parciais (PLS)

A regressão por mínimos quadrados parciais (PLS) foi realizada para as amostras de café puro e para as amostras com cevada em diferentes proporções. Foram utilizados os dados originais na construção do modelo PLS e o método utilizado foi o *SIMPLS*. Usando validação cruzada, o modelo de regressão ótimo foi obtido com 4 PLS escores (variáveis latentes). O modelo aplicado não apresenta uma boa correlação entre as concentrações reais de cevada e as concentrações de cevada estimadas. O coeficiente de correlação ( $r^2$ ) observado foi de 71% apenas. Os valores de referência estão plotados contra os valores estimados no modelo na Figura 16.



**FIGURA 16.** Valores de referência versus valores estimados pelo modelo PLS com quatro variáveis latentes (método *SIMPLS* e validação cruzada)

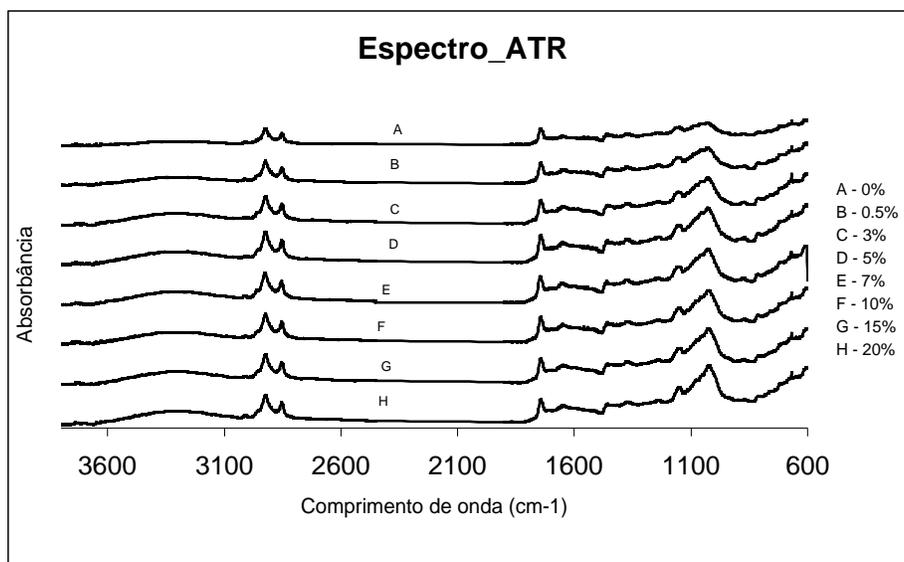
Os erros absolutos do modelo PLS aplicado estão representados na Figura 17. O erro absoluto é a diferença entre o valor verdadeiro ( $y$ ) e o valor estimado ( $\hat{y}$ ) (erro =  $y - \hat{y}$ ). Observa-se uma tendência dos valores mais baixos de concentração de cevada serem superestimados enquanto que os valores mais altos de concentração de cevada são subestimados.



**FIGURA 17.** Gráfico de erros absolutos contra as concentrações de referência

### 4.3 Reflectância total atenuada (ATR)

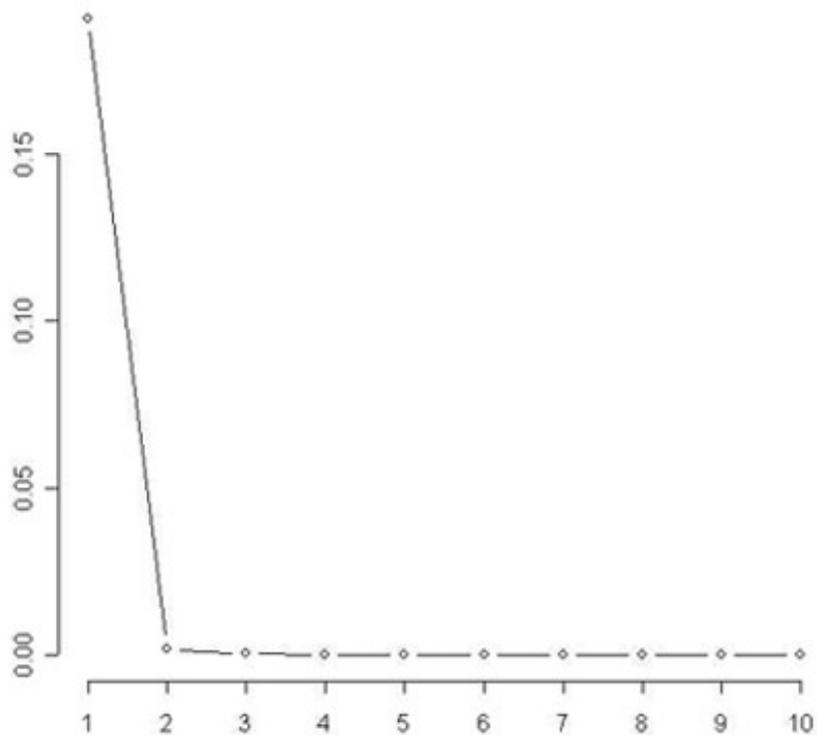
Na Figura 18 estão representados os espectros obtidos por reflectância total atenuada (ATR) para o café puro e para as amostras com cevada em diferentes proporções. Observam-se pequenas diferenças nas regiões de 600-1.900 e 2.450-3.800  $\text{cm}^{-1}$ . No entanto, essas diferenças são tão pequenas e os espectros tão complexos, que estatísticas multivariadas são necessárias para se obter uma melhor diferenciação entre as amostras.



**FIGURA 18.** Espectros de infravermelho obtidos por reflectância total atenuada (ATR) para café puro e café com cevada, em diferentes proporções

#### 4.3.1 Análise de componentes principais (PCA)

A análise de componentes principais (PCA) foi aplicada aos dados obtidos por ATR nas regiões de 600-1.900 e 2.450-3.800  $\text{cm}^{-1}$ . É possível determinar o número de componentes principais necessárias para resumir a variação amostral total por meio da análise do gráfico de cotovelo (Figura 19). Este gráfico é um processo de determinação do número apropriado de componentes a ser retido.



**FIGURA 19.** Gráfico de cotovelo para os dados de infravermelho (ATR) das variâncias representadas para as dez primeiras componentes principais

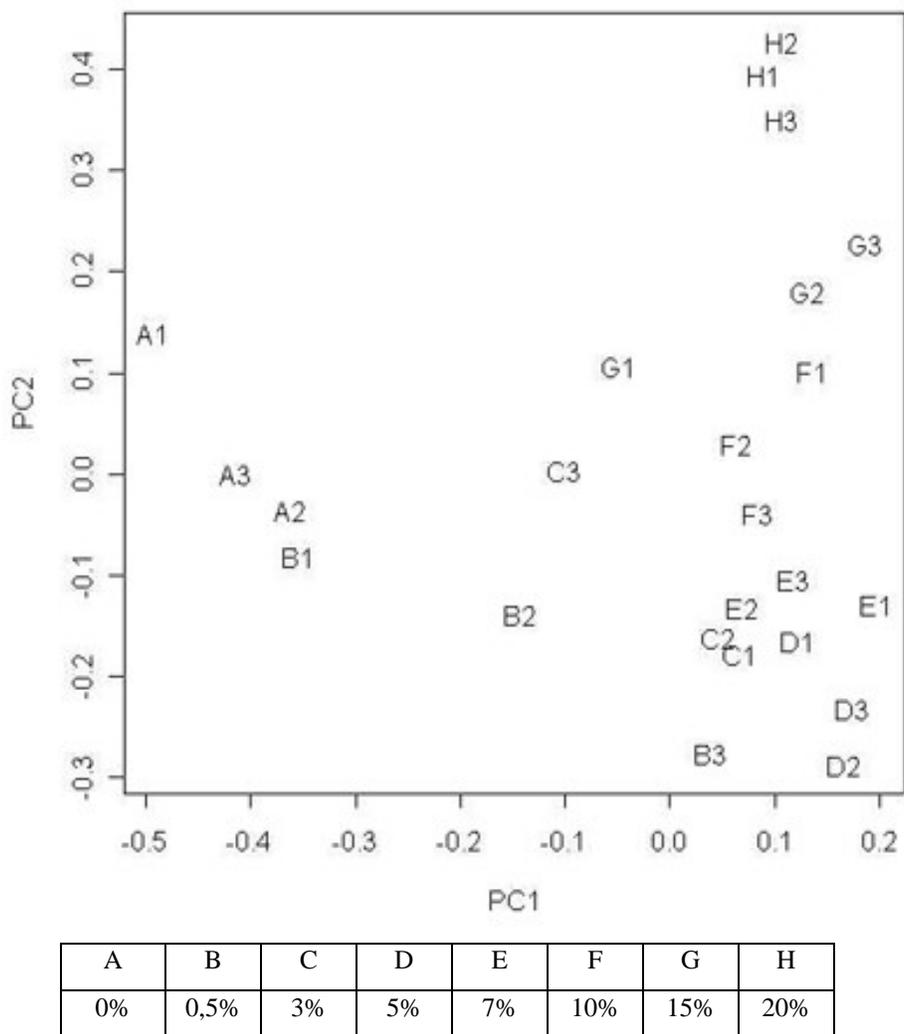
Pelo gráfico da Figura 19, observa-se que um cotovelo é formado exatamente na posição dois. Isso significa que as componentes acima de dois possuem aproximadamente a mesma magnitude e são relativamente pequenas, indicando que as duas primeiras componentes são suficientes para explicar a variância amostral total. A porcentagem das dez primeiras componentes principais está apresentada na Tabela 3.

**TABELA 3.** Porcentagem de variância e porcentagem de variância acumulada para a análise de PCA, para os dados obtidos por reflectância total atenuada (ATR), nas regiões espectrais de 600-1.900 e 2.450-3.800  $\text{cm}^{-1}$ .

PC escores	% de variância	% de variância acumulada
1	98,500	98,500
2	0,926	99,426
3	0,197	99,623
4	0,145	99,768
5	0,129	99,897
6	0,062	99,959
7	0,030	99,989
8	0,008	99,997
9	0,005	100,005
10	0,003	100,008

Para a análise de ATR, duas componentes principais explicam 99,43% da variabilidade dos dados analisados, sendo 98,5% explicados pela componente principal 1 (PC1) e 0,93% pela componente principal 2 (PC2).

A dispersão das amostras é mostrada por meio do gráfico de escores da primeira (PC1) contra a segunda componente principal (PC2) (Figura 20).

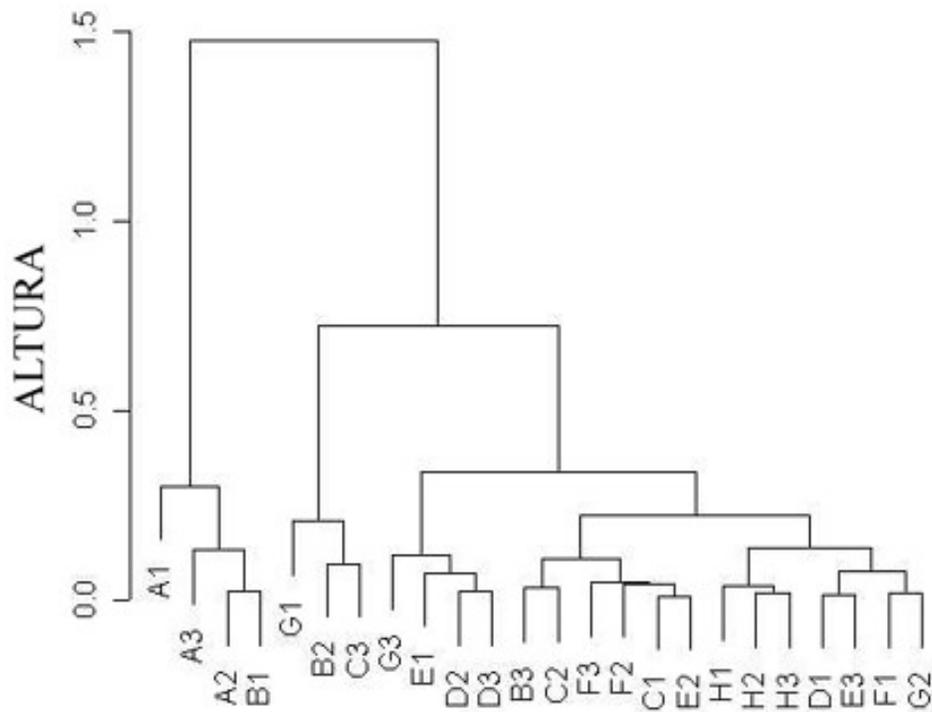


**FIGURA 20.** Gráfico de dispersão da primeira (PC1) e segunda (PC2) componentes principais (as letras correspondem às diferentes proporções de cevada e os itens 1, 2 e 3 referem-se à repetição)

Por meio do gráfico de escores, observa-se que as amostras encontram-se separadas de acordo com o teor de cevada que apresentam, sendo evidenciada a separação do café puro das demais amostras, com exceção de uma amostra com 0,5% de cevada. As amostras que continham 20% de cevada ficaram mais separadas. Sendo assim, pode-se dizer que a técnica de espectroscopia por ATR é capaz de distinguir café puro de café adulterado, com exceção de amostras com 0,5% e permite detectar amostras de café com 20% de cevada.

#### **4.3.2 Análise de agrupamentos hierárquicos (HCA)**

Após a PCA, foi realizada a análise de agrupamentos hierárquicos utilizando-se assim um número reduzido de variáveis. O método utilizado foi o do vizinho mais distante (*complete linkage*) e a matriz de similaridade foi obtida pelo cálculo da distância euclidiana. A matriz utilizada para o cálculo da matriz de similaridade foi a matriz de escores. Os grupos formados estão representados em um dendograma completo (Figura 21).



A	B	C	D	E	F	G	H
0%	0,5%	3%	5%	7%	10%	15%	20%

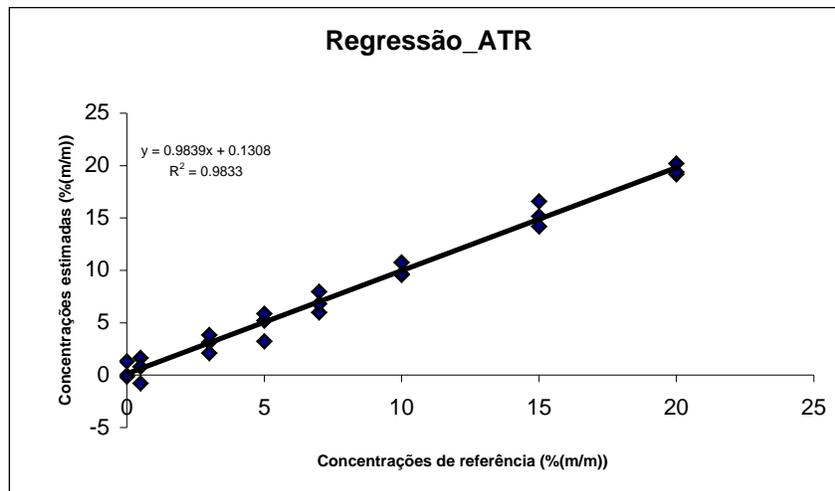
**FIGURA 21.** Dendograma completo de 24 amostras, obtido por distância euclidiana e pelo método do vizinho mais distante (as letras correspondem às diferentes proporções de cevada e os itens 1, 2 e 3 referem-se à repetição)

No dendograma, observa-se a separação de dois grupos principais, sendo um grupo formado pelo café puro e o outro pelas amostras com cevada. Além disso, observa-se que há uma tendência de agrupamento entre as amostras com

teores de cevada aproximados. No grupo formado pelas amostras com cevada observa-se um subgrupo formado pelas três amostras, com teor de 20% de cevada. Dessa forma, confirma-se o resultado obtido na observação do gráfico de escores, em que houve separação do café puro, com exceção de uma amostra contendo 0,5% de cevada e a distinção das amostras com 20%.

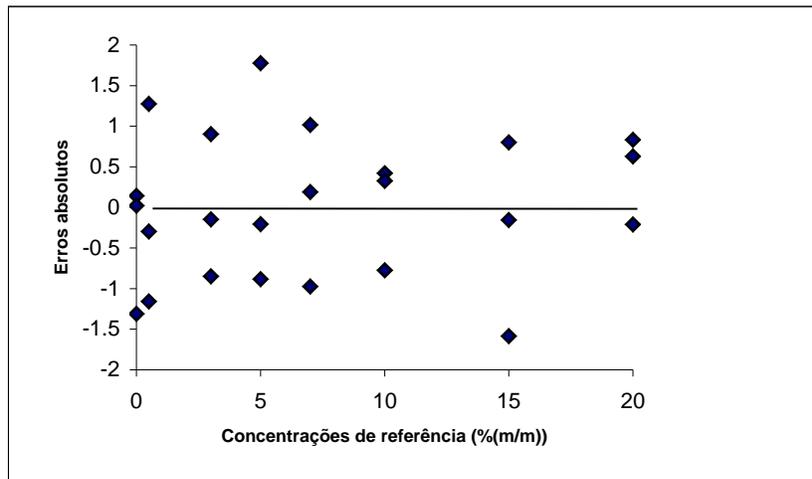
#### 4.3.3 Regressão por mínimos quadrados parciais (PLS)

Foi realizada a regressão por mínimos quadrados parciais (PLS) para as amostras de café puro e de café com cevada. O modelo PLS construído utilizou as variáveis originais e o método PLS aplicado foi o *SIMPLS*. Usando validação cruzada, o modelo de regressão ótimo foi obtido com 3 PLS escores (variáveis latentes). Na Figura 22 estão plotados os valores de referência contra os valores estimados no modelo.



**FIGURA 22.** Valores de referência versus valores estimados no modelo PLS com três variáveis latentes (método *SIMPLS* e validação cruzada)

No gráfico de dispersão dos erros absolutos observa-se uma boa distribuição dos erros em torno dos valores verdadeiros, mostrando que no modelo utilizado, não há efeitos de tendências. Apesar dos erros absolutos serem elevados para as concentrações baixas de cevada, foi observada uma boa correlação entre os valores verdadeiros e os valores estimados. O coeficiente de correlação ( $r^2$ ) encontrado foi de 98%. Na Figura 23, estão representados os erros absolutos do modelo.



**FIGURA 23.** Gráfico de erros absolutos contra as concentrações de referência

## 5 CONCLUSÕES

Os resultados obtidos neste trabalho permitem concluir que:

- as técnicas de espectroscopia na região do infravermelho médio mostraram-se eficazes na avaliação de amostras de café com cevada, tendo que a técnica de reflectância total atenuada (ATR) apresentou uma melhor eficiência;
- pelos métodos quimiométricos de análise foi possível detectar a presença de cevada (substância usada como adulterante) no café torrado e moído de maneira simples e com bons resultados;
- em se tratando de um conjunto de dados tão grande, a utilização de métodos quimiométricos é de grande valia, uma vez que sem estes fica bastante difícil a análise conjunta de todas as variáveis;
- o uso de aplicativos de fonte aberta permite a obtenção de resultados de qualidade, ampliando assim, a inclusão tecnológica de grupos de pesquisa com menor acesso aos recursos disponíveis;
- a análise de PCA foi muito eficiente na redução da dimensionalidade dos dados. A dispersão das amostras foi representada em apenas duas componentes principais, com uma porcentagem de explicação da variância maior que 90%. Foi possível separar o café puro do café com cevada nas análises de ATR e DRIFT, com exceção de amostras contendo 0,5% de cevada;

- a análise HCA permitiu separar o café puro do café com cevada em dois grupos distintos, com a técnica de ATR, além de apresentar uma tendência de agrupamento das amostras semelhantes;
- a regressão por mínimos quadrados parciais (PLS) mostrou-se de grande eficiência para a técnica de ATR. Estabeleceu-se uma boa correlação entre os valores verdadeiros e os valores estimados de cevada com coeficiente de correlação ( $r^2$ ) de 98%.

### **5.1 Perspectivas futuras**

Vale ressaltar que os parâmetros discutidos neste trabalho não devem ser tomados isoladamente como parâmetros para apontar a presença de adulterantes em café e sim apenas como indicativo quantitativo. Sendo assim, como perspectivas futuras, tem-se a possibilidade de realização da determinação de outros tipos de adulterantes do café.

Estas determinações, possivelmente, serão ser feitas usando a técnica de espectroscopia na região do infravermelho médio, com aplicação de métodos quimiométricos e calibração multivariada.

## REFERÊNCIAS BIBLIOGRÁFICAS

AMAYA, D.B.R. **A guide to carotenoid analysis in foods**. Washington: ILSI, 1999.

ASSAD, E.D. et al. Identificação de impurezas e misturas em pó de café por meio de comportamento espectral e análise de imagens digitais. **Pesquisa Agropecuária Brasileira**, Brasília, v.37, n.2, p.211-216, fev. 2002.

ASSOCIAÇÃO BRASILEIRA DA INDÚSTRIA DE CAFÉ. **A história do café: origem e trajetória**. Disponível em: <<http://www.abic.com.br/scafe.historia.html>>. Acesso em 10 de jan. 2005.

BEEBE, K.R.; PELL, R.J.; SEASHOLTZ, M.B. **Chemometrics: a practical guide**. Nova York: J. Wiley & Sons, 1998.

BRASIL. Ministério da Saúde. Portaria nº377 de 26 de abril de 1999. **Diário Oficial da República Federativa do Brasil**, Brasília, n.80, p.22-29, abr. 1999. Seção 1.

BRATCHELL, N. **Cluster analysis**. Chemometrics and Intelligent Laboratory Systems, v.6, p.105-125, 1989.

BRIANDET, R.; KEMSLEY, E.K.; WILSON, R.H. Approaches to adulteration detection in instant coffees using infrared spectroscopy and chemometrics. **Journal of the Science of Food and Agricultural**, v.71, p.359-366, 1996a.

BRIANDET, R.; KEMSLEY, E.K.; WILSON, R.H. Discrimination of arabica and robusta in instant coffee by FTIR and chemometrics. **Journal Science Food Chemical**, v.44, p.170-174, 1996b.

CHALMERS, J.M.; MACKENZIE, M.W. Some industrial applications of FT-IR diffuse reflectance spectroscopy. **Applied Spectroscopy**, v.34, p.533-539, 1985.

COLEMAN, P.B. **Practical sampling techniques for infrared analysis**. London: CRC, 1993.

DOWNEY, G. Authentication of food and food ingredients by near infrared spectroscopy. **Journal of Near Infrared Spectroscopy**, v.4, p.47-61, 1996.

FERRÃO, M.F. et al. Determinação simultânea dos teores de cinza e proteína em farinha de trigo empregando NIR-PLS e DRIFT-PLS. **Ciência e Tecnologia dos Alimentos**, v24, n.30, p.333-340, jul./set. 2004.

FERREIRA, M.M.C. Quimiometria I: calibração multivariada, um tutorial, **Química Nova**, v.22, p.724-731, 1999.

FIDÊNCIO, P.H. **Análise de solos por espectroscopia no infravermelho próximo e aplicação de métodos quimiométricos**. 2001. Tese (Doutorado em Ciências-Química Analítica)-Universidade de Campinas, Campinas, SP.

GELADI, P.; KOWALSKI, B.R. Partial least squares regression: a tutorial, **Analytica Chimica Acta**, v.185, p.1-17, 1996.

GOULD, W.A. **Tomato production, processing & technology**. 3.ed. Baltimore: CTI, 1992.

GRIFFITHS, P.R. **Chemical infrared Fourier transform spectroscopy**. New York: J. Wiley & Sons, 1975.

INSTITUTO Adolfo Lutz: metodologia incontestável. **Jornal do Café**. São Paulo, v.6, n.66, p.10-12, fev. 1997.

INTRODUÇÃO ao R: tutoriais do programa R. Disponível em: <<http://www.est.ufpr.br/Rtutorial/>>. Acesso em: 4 jan. 2005.

LAJOLO, F.M.; NUTTI, M.R. **Transgênicos: bases científicas da sua segurança**, São Paulo: SBAN, 2003.

LAURITSENA, L.N.; BÜRGI, H.B. Cluster analysis of periodic distributions; application to conformational analysis. **Journal of Computational Chemistry**, v.6, p.216-228, 1985.

MACKENZIE, M.W. **Advances in applied Fourier transform infrared spectroscopy**. Chichester: J. Wiley & Sons, 1988.

MARTENS, H.; NAES, T. **Multivariate calibration**, New York USA: J. Wiley & Sons, 1989.

MARTENS, H.; NAES, T. **Multivariate calibration**. New York: J. Wiley & Sons, 1996.

MENDONÇA, L.M.V.L. **Características físico-químicas e sensoriais de cultivares de *Coffea arabica* L.** 2004. 153p. Tese (Doutorado em Ciências dos Alimentos)-Universidade Federal de Lavras, Lavras, MG.

MENEZES Jr., J.B.F. Fraudes do café. **Revista do Instituto Adolfo Lutz**, São Paulo, v.12, n. único, p.11-144, 1952.

MESSERSCHMIDT, I. **Análise quantitativa por espectroscopia no infravermelho médio empregando técnicas de reflectância e calibração multivariada.** 1999. Tese (Doutorado em Química)-Universidade de Campinas, Campinas-SP.

MIRABELA, F.M. International reflection spectroscopy. **Applied Spectroscopy**, v.21, p.45-178, 1985.

MORGANO, M.A.; MORIYA, C.; FERREIRA, M.M.C. Determinação quantitativa o teor de sacarose em açúcar cristal por espectroscopia FTIR/ATR e regressão multivariada, **Btaz. Journal of Food Technology**, v.6, p.77-83, 2003.

O CAFÉ no mundo, produção e consumo. Londrina, Disponível em: <<http://www.pr.gov.br/iapar/café/m&prodcons.html>>. Acesso em: 10 jan. 2005

OLINGER, J.M.; GRIFFITHS, P.R. Effects of sample dilution and particle-size morphology on diffuse reflection spectra of carbohydrate systems in the near-infrared and mid-infrared. 1. Single Analytes, **Applied Spectroscopy**, v.47, p. 687-694, 1993.

PARADKAR, M.M.; SIVAKESAVA, S.; IRUDAYARAJ, J. Discrimination and classification of adulterants in maple syrup with the use of infrared spectroscopic techniques. **Journal of the Science of Food and Agricultural**, v.83, p.714-721, 2003.

PARREIRA, T.F. **Utilização de métodos quimiométricos em dados de natureza multivariada.** 2003. Dissertação (Mestrado em Química)-Universidade de Campinas, Campinas, SP.

PEDRO, A.M.K. **Determinação simultânea e não-destrutiva de sólidos totais e solúveis, licopeno e beta-caroteno em produtos de tomate por espectroscopia no infravermelho próximo utilizando calibração multivariada,** 2004. Dissertação (Mestrado em Físico-Química)-Universidade de Campinas, Campinas, SP.

PIMENTA, C.J. **Qualidade de café**. Lavras: UFLA. 304p, 2003

R DEVELOPMENT CORE TEAM. **R - a language and environment for statistical computing**. Vienna, Austria: R Foundation for Statistical Computing, 2004. ISBN 3 – 900051 – 07 – 0, URL Disponível em: <<http://www.R-project.org>>. Acesso em: 10 set. 2004

RIBEIRO, F.A.L. **Aplicação de métodos de análise multivariada no estudo de hidrocarbonetos policíclicos aromáticos**. 2001. Dissertação (Mestrado em Química)-Universidade de Campinas, Campinas, SP.

RINANTONIO, V. Coffee. In: \_\_\_\_\_. **Ullman's encyclopedia of industrial chemistry**. New York. v.A7, p.315-338, 1987.

ROBINSON, C. **Genetic modification technology and food**. Brussels: ILSI Europe, 2001.

RON WEHRENS. **pls.pcr: PLS and PCR functions**. R package version 0.2.6, 2004.

SADLER, G.; DAVIS, J.; DEZMAN, D. Rapid extraction of lycopene and beta-carotene from reconstituted tomato paste and pink grapefruit homogenates. **Journal of Food Science**, v.55, p.1460-1461, 1990.

TEVRRUCHT, M.L.E.; GRIFFITHS, P.R. A simple sample packing accessory for diffuse reflectance infrared spectrometry. **Applied Spectroscopy**, v.43, p.1492-1494, 1989.

THOMAS, E. V. A primer on multivariate calibration, **Analytical Chemistry**, v.66, p.795A-804A, 1994.

VERBEKE, J. S. et al. Clustering applied to an organic air pollutants data set. **Analysis**, v.12, p.486-489, 1984.

VOET, D.; VOET, J.; PRATT, C.W. **Fundamentos de bioquímica**. Porto Alegre: Artmed, 2000.

VOGEL, A.I. **Vogel** – análise química quantitativa. 2.ed. Rio de Janeiro: Guanabara Koogan, 1992.

WENDLANT, W.W.M.; HECHT, H.G. **Reflectance spectroscopy**. New York: Interscience, 1966.

WETZEL, D. L. Near-infrared reflectance analysis-Sleeper among spectroscopic techniques, **Analytical Chemistry**, v.55, p.1165-1168, 1983.

WILSON, R. H. Fourier transform mid-infrared spectroscopy for food analysis. **Trends in Analytical Chemistry**, v.9, p.127-131, 1990.

WISE, B.M.; GALLAGHER, N.B. **PLS-Toolbox for use with MATLAB**, version2.0. Eigenvector Research, 1998.

This document was created with Win2PDF available at <http://www.daneprairie.com>.  
The unregistered version of Win2PDF is for evaluation or non-commercial use only.