UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA

DISSERTAÇÃO DE MESTRADO

Classificação de Cigarros Usando Espectrometria NIRR e Métodos Quimiométricos de Análise

Por

Edilene Dantas Teles Moreira

Orientador: Prof. Dr. Mário César Ugulino de Araújo

João Pessoa – março/2007

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA

Classificação de Cigarros Usando Espectrometria NIRR e Métodos Quimiométricos de Análise

Por

Edilene Dantas Teles Moreira

Dissertação submetida ao programa de pós-graduação em química, da Universidade Federal da Paraíba, como requisito parcial à obtenção do título de Mestre em Química, área de concentração "Química Analítica".

Orientador: Prof. Dr. Mário César Ugulino de Araújo.

João Pessoa – março/2007

A Deus, Pai Todo Poderoso e Misericordioso, por permitir e me tornar capaz de conquistar tantas coisas em minha vida. Por me amparar e me encorajar nos momentos de fraquezas. Por fechar janelas e abrir milhares de portas, no caminho que eu escolhi percorrer.

Ao meu amado esposo, Pablo Moreira, por acreditar na minha capacidade e por estar sempre me apoiando e incentivando. Por todo amor, carinho, companheirismo, lealdade e sinceridade. E acima de tudo, pela compreensão nos momentos em que ele precisou muito de mim e eu estava ausente.

Com muito AMOR e GRATIDÃO,
Eu dedico.

AGRADECIMENTOS

- Aos meus queridos pais, Regina Lúcia e Edivaldo Dantas, pelo amor, carinho e pela educação que me deram ao longo de todo o meu crescimento, permitindo a realização de mais esta vitória. Por todo o exemplo de vida que eles são para mim.

- Aos meus irmãos, Márcio Henrique, Alessandro Michel e Sandra Michele, pela família fiel e unida que somos.

- A minha querida sobrinha, Méline Ansermet, que me trouxe muita alegria e me proporcionou momentos inesquecíveis.

- As minhas eternas e verdadeiras amigas, Alessandra Félix e Glauciene Paula por estarem comigo em todos os momentos, rindo e chorando juntas, dividindo e compartilhando as nossas histórias.

- Ao Prof. Dr. Mário Ugulino pela oportunidade de trabalho, orientação, apoio, confiança e amizade.

- Ao Prof. Dr. Sérgio B. Santos pelos conselhos, orientações e pela amizade sempre sincera e alegre.

- Ao amigo Márcio C. Pontes pela ajuda, mais do que imprescindível, para a realização deste trabalho e também pela valiosa amizade.

- Aos demais professores do LAQA pelas contribuições acadêmico-científicas.

- À Amália Gama e Osmundo Dantas pelo companheirismo e amizade conquistada.

- A Dannielle Muniz, Janaine Geiza, Wellington Lira, Simone Simões e Ricardo Alexandre pela amizade.

- A Sara Regina e Elaine Cristina, pelo apoio na etapa inicial deste trabalho.

- A todos que fazem o LAQA pela boa convivência, brincadeiras e também pela troca de idéias, de conhecimentos e experiências.

- A CAPES pelo fornecimento da bolsa.

- Enfim, a todos que, de maneira direta ou indireta, estiveram contribuindo com essa realização profissional.

Muito Obrigada!

A	Amostra desconhecida
ABCF	Associação Brasileira de Combate à Falsificação
ANVISA	Agência Nacional de Vigilância Sanitária
CG-MS	Cromatografia e espectrometria de massa
$CHCl_3$	Clorofórmio
CO	Monóxido de Carbono
DI	Dimensionalidade Inerente
FAR	Infravermelho Distante
F_{cal}	Valor calculado para o teste F
F_{crit}	Valor crítico adotado para o teste F
FI	Injeção em Fluxo
FTIR	Infravermelho com Transformada de Fourier
G	Erro médio de uma classificação incorreta
HCA	Análise Hierárquica de Agrupamentos
H_i	Influência
I_o	Radiação de referência
I_R	Radiação refletida
KBr	Brometo de Potássio
K_{KM}	Constante de proporcionalidade que equivale ao produto da absortividade molar pelo caminho óptico.
LDA	Análise Discriminante Linear
MIR	Infravermelho Médio
MLR	Regressão Linear Múltipla
NIR	Infravermelho Próximo
NIRR	Infravermelho Próximo de Reflectância
NO	Óxido Nitroso
OMS	Organização Mundial de Saúde
PC's	Componentes Principais
PCA	Análise de Componentes Principais
PCR	Regressão em Componentes Principais
PLS	Mínimos Quadrados Parciais

R	Reflectância
R_{am}	Reflectância absoluta da amostra
R_{ef}	Reflectância efetiva da amostra
R_{ref}	Reflectância da substância de referência
RMSEP	Erro médio Quadrático de Previsão
SFA	Análise Fatorial Subjanelas
SG	Savitzky-Golay
S_i	Raiz quadrada da variância residual
SIMCA	Modelagem Independente Flexível por Analogia de Classe
SPA	Algoritmo das Projeções Sucessivas
S_{xy}	Similaridade entre duas amostras x e y quaisquer
x	Distância entre a projeção de A na direção da PC e a fronteira
y	Distância entre a amostra desconhecida A e o eixo da PC
z	Proximidade de A até a caixa

O cigarro é uma droga que traz muitos malefícios à saúde humana. Os falsificados e/ou contrabandeados além desses problemas geram enormes prejuízos aos governos na arrecadação de impostos. A inspeção visual é um método simples, barato e rápido, mas não tão eficaz em termos de fiscalização da qualidade. Este problema tem sido superado usando métodos cromatográficos, mas estes são caros, laboriosos e lentos. Para contornar esses inconvenientes, neste trabalho foi desenvolvida uma nova metodologia que usa a espectrometria de reflectância no infravermelho próximo (do inglês: *Near-Infrared Reflectance - NIRR*), o algoritmo das projeções sucessivas (do inglês: *Sucessive Projections Algorithm – SPA*) e a análise de discriminante linear (do inglês: *Linear Discriminant Analysis LDA*). No seu desenvolvimento foram usadas 210 amostras de cigarros de 4 marcas diferentes, adquiridas em estabelecimentos comerciais da grande João Pessoa - PB. O modelo elaborado com o método NIRR-SPA-LDA proposto selecionou 2 (5056 e 4903 cm^{-1}) dentre as 1168 variáveis da faixa espectral NIRR (de 5420 a 4252 cm^{-1}) medida e conseguiu classificar corretamente (100% de acerto) todos os cigarros nas 4 marcas usadas no estudo. Este resultado é muito animador, principalmente ao compará-lo com aquele obtido pelo SIMCA (do inglês: *Soft Independent Modeling of Class Analogy*), método de classificação tradicional, cujo índice de acerto obtido nas análises das mesmas amostras foi de 94%, com um nível de significância 5%. Portanto, o método NIRR-SPA-LDA proposto é uma boa alternativa para a fiscalização da qualidade de cigarros de forma não-invasiva, pouco laboriosa, rápida e precisa, sem produzir resíduos químicos. Além disso, deduz-se, a partir do método proposto que usa apenas 2 variáveis espectrais, que instrumentos de baixo custo, tais como os fotômetros NIRR, podem ser construídos usando diodos emissores de luz (do inglês: *Light Emitting Diodes-LEDs*) nos comprimentos de onda selecionados.

Palavras-Chave: fiscalização de cigarros; espectrometria NIRR; SPA, LDA, SIMCA.

Cigarette is a drug that brings a lot of harms to the human health. Besides this problem, falsified and/or smuggled cigarettes generate(s) enormous damages to the governments in the collection of taxes. The visual inspection is a simple, cheap and fast method, but it is not so effective in terms of fiscalization of the quality. This problem has been overcome using chromatographic methods, but these methods are expensive, onerous and slow. To avoid these drawbacks, in this work was developed a new methodology that uses the Near Infrared Reflectance Spectrometry (NIRR), the Successive Projections Algorithm (SPA) and the Linear Discriminant Analysis (LDA). In its development it was used 210 samples of cigarettes of 4 different marks, acquired in commercial establishments from João Pessoa city, State of Paraíba. The model elaborated with the NIRR-SPA-LDA proposed method selected 2 (5056 and 4903 cm^{-1}) among the measured 1168 variables in the NIRR spectral range (from 5420 to 4252 cm^{-1}) and it was able to classify correctly (100% of success) all of the cigarettes of the 4 marks used in the study. This result is very exciting, mainly when comparing it with that ones obtained by SIMCA (Soft Independent Modeling of Class Analogy), a traditional classification method, whose success index obtained in the analyses of the same samples was of 94%, with a level of significance of 5%. Therefore, the NIRR-SPA-LDA proposed method is a good alternative for the inspection of cigarettes quality in a not invasive, little laborious, fast and precise way without to produce chemical residues. Besides, it can be deduced, from the proposed method that uses only 2 spectral variables, that low cost instruments, such as NIRR photometers, can be built using LEDs (Light Emitting Diodes) with the selected wavelengths.

Keywords: fiscalization of cigarettes; NIRR spectrometry; SPA, LDA, SIMCA.

"Ninguém se engane a si mesmo: se alguém dentre vós se tem por sábio neste século, faça-se estulto para se tornar sábio. Porque a sabedoria deste mundo é loucura diante de Deus, porquanto está escrito: Ele apanha os sábios na própria astúcia deles".
1 Coríntios 3:18,21.

C A P Í T U L O 1

INTRODUÇÃO

"Ninguém se engane a si mesmo: se alguém dentre vós se tem por sábio neste século, faça-se estulto para se tornar sábio. Porque a sabedoria deste mundo é loucura diante de Deus, porquanto está escrito: Ele apanha os sábios na própria astúcia deles".

1 Coríntios 3:18,21.

1.1. O Tabaco e a Origem do Cigarro

Tabaco é o nome comum dado às plantas do gênero *Nicotiana*, pertencente à família das Solanáceas, em particular a espécie botânica *Nicotiana tabacum*. É desta planta que se extrai a substância chamada nicotina[1]. A denominação nicotina foi dada em homenagem ao médico e pesquisador francês Jean Nicot, que a usava como medicamento para curar as enxaquecas de uma rainha. Em 1737, recebeu a classificação científica feita pelo biólogo Linneu que a registrou com o nome *Nicotiana* em quatro variedades distintas: *rústica*, *glutinosa, penicilata e tabacum*[2].

A nicotina contida no tabaco foi originada da América Central e acompanhou as migrações dos índios até chegar ao território brasileiro. Muitas culturas indígenas faziam rituais mágico-religiosos e sagrados, utilizando o fumo do tabaco. Esta erva chamou a atenção de conquistadores por possuir propriedades consideradas curativas e também pelo seu sabor agradável[3]. O tabaco chegou ao mundo civilizado no século XVI e a partir de então passou a ser consumido sob várias formas: cachimbo, rapé, charuto, entre outros. Só então, no final do século XIX, o tabaco passou a ser consumido na forma de cigarro[2].

O cigarro é uma pequena porção de tabaco seco e picado, enrolado em papel fino ou em palha de milho. Pode, ou não, dispor de um filtro, geralmente de esponja ou papel. Inicialmente Paris foi invadida pelo cigarro em 1860 e logo depois ocorreu uma verdadeira explosão nos Estados Unidos, na década de 1880. O cigarro expandiu-se rapidamente por se tratar de um produto mais econômico, mais cômodo de carregar e de usar, comparado às outras formas de consumo[1,2].

1.2. Consumo e Tipos de Cigarros

No mundo, o consumo anual de nicotina é de 73 mil toneladas contidas em 7,3 trilhões de cigarros. Estes são fumados por aproximadamente 1,3 bilhões de tabagistas, sendo que 80% vivem nos países em desenvolvimento. No Brasil, cerca de 30% da população adulta é viciada em nicotina[2].

Mesmo parecendo iguais, os cigarros apresentam diferenças que vão desde variações de ingredientes até alterações estratégicas de *marketing* que atendam às preferências dos consumidores. Em geral, o produto final é constituído de três partes: fumo, papel e filtro. Além destes, existem aditivos que são adicionados, tais como: açúcares, extratos vegetais e agentes de sabor. Assim, os cigarros podem variar em função do fumo e dos aditivos utilizados na sua fabricação, bem como no *designer*, ou seja, seu desenho (comprimento, circunferência, tipo de filtro, tipo de papel, etc)[3]. Atualmente, estão registradas na Agência Nacional de Vigilância Sanitária (ANVISA), 15 empresas de cigarros, entre fabricantes e fornecedores, representados por 89 marcas. Com exceção dos exportados, cerca de 180 tipos de cigarros com teores, sabores e preços diversos são colocados à disposição dos consumidores de diversos níveis sociais[4,5].

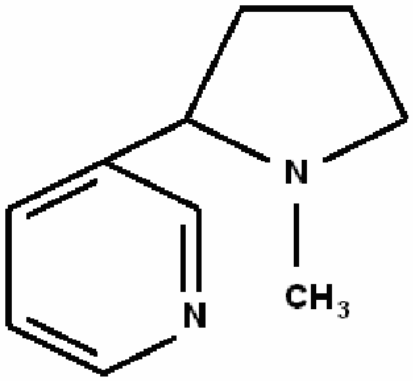
1.3. Composição Química dos Cigarros

A nicotina não ingressa de forma pura no organismo. Ela está num invólucro que é o tabaco contido no cigarro e, portanto, é administrada ao organismo com milhares de outras substâncias tóxicas. De um total de 6.700 substâncias, 4.720 já foram quimicamente identificadas[2,6]. Dentre estas a nicotina, o monóxido de carbono e o alcatrão são as mais prejudiciais[2].

1.3.1 Nicotina, Monóxido de carbono e Alcatrão

A nicotina é um alcalóide vegetal e sua principal fonte é a planta do tabaco. Embora a concentração varie em diferentes espécies de tabaco, as maiores concentrações estão em regiões altas e próximas ao talo da planta. Trata-se de um composto orgânico, uma amina terciária composta de anéis de piridina e pirrolidina, de forma molecular $C_{10}H_{14}N_2$ e estrutura mostrada na **Figura 1.1**. Ela é líquida à temperatura ambiente, incolor, inodora e oleosa e quando exposta à luz ou calor adquire uma coloração marrom e odor característico do tabaco. Em pequenas doses, a nicotina estimula especialmente o sistema nervoso vegetativo, favorecendo a

liberação de adrenalina e, em grandes doses, paralisa o sistema nervoso autônomo provocando convulsões que podem levar à morte[2,7-8].



NICOTINA

Figura 1.1. Representação da molécula da nicotina (S)-1-metil-2-(3-piridil) pirolidina.

O monóxido de carbono (CO) é um gás altamente tóxico. Este gás tem numerosos efeitos negativos sobre o organismo, sendo o mais importante dificultar o transporte de oxigênio para os tecidos do corpo. O CO tem mais afinidade com a hemoglobina do sangue do que o próprio oxigênio, deixando o corpo do fumante, ativo ou passivo, intoxicado[9].

Quanto ao alcatrão, este é uma mistura de centenas de substâncias químicas, apresentando-se de forma viscosa e cor negra. Entre as substâncias encontradas no alcatrão, incluem-se substâncias químicas orgânicas e inorgânicas e uma extensa variedade de produtos orgânicos voláteis e semivoláteis que são substâncias altamente carcinogênicas, ou seja, que atuam no desenvolvimento do câncer. O alcatrão também provoca a obstrução dos pulmões e causa perturbações respiratórias e, assim, à semelhança de outros componentes dos cigarros, é responsável pela sua toxicidade, provocando a dependência do tabaco e várias doenças associadas ao seu consumo[1,10].

1.4. Fiscalização da Qualidade de Cigarros

A Organização Mundial de Saúde (OMS) estima que cerca de 4 milhões de pessoas morrem, mundialmente, devido ao consumo do tabaco. No Brasil, a estimativa é que cerca de 200.000 mortes/ano sejam decorrentes do tabagismo. Este é

responsável por 90% dos cânceres de pulmão, 25% das mortes por doenças coronarianas, 85% das mortes por doença pulmonar, 25% das mortes por doença cérebro-vascular, impotência sexual, abortos e muitas outras doenças[6].

O Brasil não possui uma legislação específica para regulamentar os ingredientes adicionados ao cigarro, mas as indústrias de cigarros brasileiras atuam de acordo com legislações de outros países. A ANVISA desempenha um trabalho de fiscalização sobre a composição química dos derivados do tabaco, com o objetivo de controlar os níveis de exposição aos compostos tóxicos e cancerígenos do fumo e da fumaça, tanto para o consumidor como para o fumante passivo. Os níveis de nicotina, monóxido de carbono e alcatrão permitidos pela ANVISA em cada cigarro são os seguintes: 1,00 mg de nicotina; 2 mg de monóxido de carbono e 12 mg de alcatrão. Com isso, tem-se o estímulo da criação de tecnologias próprias para a análise de cigarros, auxiliando no estabelecimento de medidas sanitárias eficazes para fiscalização destes produtos que trazem tantos riscos a saúde humana[4,11-12].

1.5. Falsificação e Contrabando de Cigarros

Sabe-se que cigarro é muito prejudicial à saúde, e quando se trata de cigarro falsificado e/ou contrabandeado, estes conseguem ser ainda mais prejudiciais. Este tipo de cigarro é uma ameaça para saúde dos consumidores e também para a indústria e o comércio formal. O governo deixa de arrecadar impostos e perde bilhões de reais todo ano, devido ao contrabando e as vendas ilegais. A cada três cigarros fumados no Brasil, um é contrabandeado, e a venda destes não se restringe apenas ao comércio informal, podendo ser encontrados em padarias, bares, restaurantes e bancas. Assim, os consumidores acabam comprando cigarros falsificados e/ou contrabandeados sem notar qualquer diferença, comparado a um cigarro original[13].

Em estudo recente, feito pela Associação Brasileira de Combate à Falsificação (ABCF), foram analisados cigarros de 45 marcas comercializadas ilegalmente. Em todas elas foram encontrados materiais estranhos como pêlo de rato, grãos de areia, pedaços de barbante, asas de insetos, plásticos, penas de aves e até

mesmo inseticidas que têm a utilização proibida no Brasil. Também foram identificados teores de alcatrão, nicotina e monóxido de carbono acima dos níveis permitido. Raramente esses estudos chegam ao conhecimento dos consumidores e por se tratar de falsificações tão sofisticadas é que milhões de brasileiros tragam essas substâncias sem perceber qualquer alteração[12]. Geralmente, o trabalho de identificação de cigarros realizado pelos peritos dos órgãos de fiscalização é feito através de uma inspeção visual. Trata-se de um método simples, barato e rápido, mas não tão eficaz em termos da qualidade do cigarro.[3]

Diante dos fatos relatados acima, torna-se indispensável o desenvolvimento de metodologias analíticas que permitam análises rápidas e seguras do próprio fumo do cigarro. No contexto dos métodos baseados em análise multivariada, em especial os métodos de classificação, têm-se encontrado poucos trabalhos[14-20] que exploram medidas instrumentais, sobretudo as espectroscópicas, com o propósito de auxiliar a fiscalização de cigarros.

1.6. Análises e Classificação de Cigarros – Revisão Bibliográfica

Huang et al[14] empregaram a técnica de cromatografia gasosa acoplada a espectrometria de massas (do inglês: *Gas Chromatography - Mass Spectrometry* - GC–MS), combinada com métodos quimiométricos, para realizar uma análise comparativa de componentes voláteis de tabaco, obtidos de cinco lugares diferentes. As amostras foram submetidas a processos de destilação simultânea e a métodos de extração para obtenção de seus componentes voláteis. Com a análise fatorial subjanelas (do inglês: *Subwindow Factorial Analysis* - SFA) foram identificados e quantificados 102 componentes voláteis entre 138 picos distintos, correspondendo a 88,9% do conteúdo total. O método foi desenvolvido com intuito de ser usado na comparação de tabacos extraídos de diferentes locais e para o controle de qualidade na produção de cigarros.

Zimmermann et al[15] usaram as medidas de Py-SPI-TOFMS (*pyrolysis single-photon ionisation-time-of-flight mass spectrometry*) e métodos quimiométricos

de análise, PCA (*Principal Component Analysis*), LDA (*Linear Discriminant Analysis*) e PCR (*Principal Component Regression*) para discriminação de três tipos de tabaco (Burley, Virgínia e Oriental). Os resultados forneceram informação sobre a composição química e as características da fumaça derivadas de cada tipo de tabaco e permitiram conclusões sobre o cultivo das plantas citadas.

Guardia et al[16] propuseram um procedimento completamente automatizado que utiliza o infravermelho com transformada de Fourier (FTIR) e análise por injeção em fluxo (*Flow Injection Analysis - FIA*) para a determinação de nicotina em tabaco. O método baseou-se na extração *on-line* de nicotina com clorofórmio ($CHCl_3$). Uma alíquota de 400µl do extrato foi introduzida numa micro-cela de fluxo usando o $CHCl_3$ como carregador enquanto o espectro de infravermelho era registrado continuamente. A absorbância foi medida na faixa de comprimento de onda de 1334-1300 cm^{-1}. O método apresentou um limite de detecção de 0,1 $mg.ml^{-1}$ de nicotina, desvio padrão relativo abaixo de 2% e freqüência analítica de 6 análises por hora.

Wuang et al[17] utilizaram o método SIMCA (*Soft Independent Modelling of Class Analogy*) e PLS (*Partial Least Square Regression*) combinados à espectroscopia no infravermelho próximo com transformada de Fourier (FT-NIR) para avaliar a qualidade de papéis de cigarro. Os métodos de classificação foram estabelecidos para a discriminação de papéis de cigarro, enquanto os modelos de calibração foram estabelecidos para a determinação da gramatura, espessura, permeabilidade, umidade e cinzas do papel. Os coeficientes de correlação para os modelos ficaram em torno de 0,96.

A cromatografia gasosa com detecção espectrofotométrica com arranjo de diodo na região ultravioleta (167-330 nm) foi utilizado por Hatzinikolaou et al[18] para a análise da fumaça de cigarro. Foram identificados mais de 20 compostos voláteis e os principais compostos orgânicos voláteis foram determinados com boa precisão e reprodutibilidade, usando uma única corrida cromatográfica, executada em menos de 50 minutos.

Fatemi et al[19] utilizaram a espectrometria de absorção atômica com atomização eletrotérmica para a determinação de cádmio nos componentes do cigarro e da fumaça do cigarro. As faixas de concentração de cádmio encontradas no cigarro, na cinza do cigarro e fumaça do cigarro foram 1,3-3,1; 0,12-0,50 e 0,039-0,170 μg/g, respectivamente, para as diferentes amostras de cigarros utilizadas. Não foi detectado cádmio na corrente principal da fumaça do cigarro, indicando que, possivelmente, os fumantes passivos correm um maior risco de intoxicação por cádmio em relação aos fumantes ativos.

Borden et al[20] utilizaram a espectrometria no infravermelho e a técnica quimiométrica LDA com o objetivo de verificar a capacidade do uso destas técnicas em diferenciar amostras de soro sanguíneo de fumantes e não-fumantes. Os espectros foram registrados na faixa de 900-1800 cm^{-1} com uma resolução de 2 cm^{-1} e depois pré-processados utilizando o método de derivação de Savitzky-Golay. Eles conseguiram distinguir as amostras de soro de fumantes e não-fumantes com uma exatidão de 96,7 % para o conjunto de treinamento e de 82,8% para o conjunto de validação.

Não foram encontradas, até o momento, metodologias de classificação que possam ser utilizadas para fins de fiscalização rápida, simples e barata do fumo de cigarros. Além disso, os órgãos de fiscalização que realizam as análises desses produtos, com o intuito de verificar a autenticidade e integridade dos mesmos, utilizam métodos analíticos mais sofisticados, a exemplo dos cromatográficos[4-5,11]. Tratam-se de técnicas convencionais e, apesar de serem seguras e bem estabelecidas, apresentam desvantagens, pois necessitam de muita manipulação analítica, apresentam um alto custo de operação e de manutenção, são invasivas (destroem as amostras) e consomem reagentes químicos nocivos ao meio ambiente.

Tais inconvenientes podem ser superados através do desenvolvimento de métodos analíticos baseados no uso da espectrometria de reflectância no infravermelho próximo (NIRR). Entretanto, a técnica NIRR apresenta uma grande quantidade de dados gerados, uma alta sobreposição espectral e baixa intensidade dos

sinais e, por isso, para que uma metodologia baseada na espectrometria NIRR possa ser aplicada com eficácia, é necessária a utilização de técnicas quimiométricas.

A combinação da espectrometria NIRR com técnicas quimiométricas possibilita o desenvolvimento de metodologias bastante promissoras para a classificação de cigarros, entre outros motivos, por que as medidas NIRR podem ser associadas tanto às propriedades químicas (qualitativas e quantitativas), como às propriedades físicas das amostras (granulometria, densidade, etc), de forma não-invasiva, pouco laboriosa, rápida e precisa, sem produzir resíduos químicos.

1.7. Espectrometria na Região do Infravermelho Próximo (NIR)[21-28]

A Espectrometria NIR é um tipo de espectroscopia vibracional com energias de transição, que caracterizam um pequeno intervalo da região infravermelha do espectro eletromagnético que está compreendida entre 780 e 2500 nm (12820 e 4000 cm^{-1}).

No NIR, a ocorrência de transições eletrônicas é rara. Quando uma molécula absorve energia da região NIR observam-se, majoritariamente, bandas provenientes de sobretons ou bandas de combinação. Os sobretons provêm das transições que ocorrem entre o nível vibracional fundamental e o segundo ou terceiro nível mais energético. Estes possuem intensidades de absorção menores que as de combinações, as quais resultam das transições entre níveis vibracionais de ligações envolvendo átomos de massa relativamente baixa ou de ligações bastante energéticas como C-H, O-H e N-H. Ainda podem ocorrer acoplamentos ou ressonâncias entre diferentes vibrações em ligações do mesmo grupo funcional que geram bandas de absorção NIR.

As bandas que ocorrem na região NIR não são muito fortes. São 100 a 1000 vezes menos intensas do que as observadas no infravermelho médio (MIR) e infravermelho distante (FAR), além de ocorrer muitas sobreposições. No entanto, a dificuldade com a baixa sensibilidade devido às fracas transições pode ser superada

pelo uso de fontes de radiação mais intensas e detectores de alta eficiência que contribuem para o aumento da relação sinal/ruído.

Até 1950, espectroscopias vibracionais, como o Infravermelho Próximo (NIR), foram muito pouco utilizadas devido à complexidade dos espectros que, por apresentarem bandas largas resultantes de sobreposições de picos individuais, tornavam os espectros difíceis de serem interpretados através de métodos univariados.

O desenvolvimento das técnicas matemáticas e estatísticas (Quimiometria), a disponibilidade de softwares e o desenvolvimento de novas tecnologias instrumentais, que vêm ocorrendo desde as duas últimas décadas estão tornando a tecnologia NIR uma das técnicas mais promissoras no campo das análises precisas e confiáveis, compatíveis com as técnicas clássicas de referência. O grande número de metodologias analíticas desenvolvidas com base na técnica NIR firma-se na capacidade e na habilidade desta em realizar análises rápidas e não-invasivas em diversas áreas. Como exemplos, podem-se mencionar as indústrias petroquímica, têxtil, de carvão, farmacêutica, de polímeros, de tintas, etc.

Uma das vantagens do NIR com relação às outras técnicas que exploram o espectro eletromagnético vibracional, como o MIR, é que as medidas na região do infravermelho próximo exigem sistemas ópticos e detectores facilmente disponíveis, simplicidade relativa na instrumentação e a grande quantidade de compostos orgânicos possíveis de serem analisados.

O espectro na região NIR envolve uma radiação com os menores comprimento de onda da região infravervemelha e, consequentemente, de maior energia, favorecendo uma maior interação da radiação nas amostras. Diversos são os tipos de medidas para obtenção dos espectros, tais como transmitância/absorbância (amostras transparentes), transflectância, reflectância difusa (amostras sólidas), interactância e transmitância por dispersão média (amostras sólidas densas).

1.7.1. Espectrometria NIRR[29-34]

Dentre os mecanismos de medida na região do infravermelho próximo, a reflectância é o mais utilizado para análise de amostras sólidas. Os espectros de reflectância, embora semelhantes na aparência, não são idênticos aos espectros de absorção correspondentes, pois carregam tanto informação química quanto física, podendo ser utilizados tanto para análises qualitativas como quantitativas.

Medidas por reflectância são obtidas a partir da radiação refletida pela estrutura física da amostra. Esse mecanismo de medida é utilizado quando se deseja ou se necessita manter a integridade da amostra, ou mesmo quando esta apresenta pouca viabilidade aos mecanismos de absorção, transmitância e transflectância.

A reflectância é definida como sendo a razão entre as potências ou fluxos da radiação refletida (I_R) e a da radiação incidente (I_o) numa superfície:

$$R = \frac{I_R}{I_0} \qquad \textbf{(1.1)}$$

As medidas absolutas da reflectância dependem de condições experimentais, do ângulo de incidência e de reflexão, da espessura e estado físico da amostra, da temperatura, etc. Portanto, a reflectância é calculada comparando-se a quantidade medida na amostra com a intensidade medida numa referência, tal como a reflectância do Brometo de Potássio (KBr).

$$R_{ef} = R_{am} - R_{ref} \qquad \textbf{(1.2)}$$

onde R_{ef} é a reflectância efetiva da amostra, R_{am} é a reflectância absoluta da amostra e R_{ref} é a reflectância da substância de referência.

A reflectância, de acordo com o tipo de reflexão, pode ser: Regular ou Especular, Difusa e Total. Dentre estas, a que apresenta o maior número de aplicações, principalmente na região do NIR, é a técnica de reflectância difusa. Esta ocorre quando a radiação é refletida após penetrar e viajar dentro da estrutura física da amostra. O processo é mais bem reproduzido em superfícies rugosas que se apresentam na forma contínua ou de pó. Amostras com tais características facilitam o espalhamento da radiação incidente refletindo-a em todas as direções levando um

número maior de informações até atingir o detector. Estritamente, não se trata de um fenômeno de superfície, pois ela provém da interação entre a amostra e a radiação NIR incidente, de forma que, ocorre excitação com alteração dos modos vibracionais gerando absorção energética por parte da molécula em análise. Em uma medida instrumental, o feixe da radiação penetra nas primeiras camadas da superfície, interage com a matéria e retorna difusamente até o detector. (**Figura 1.2.**).

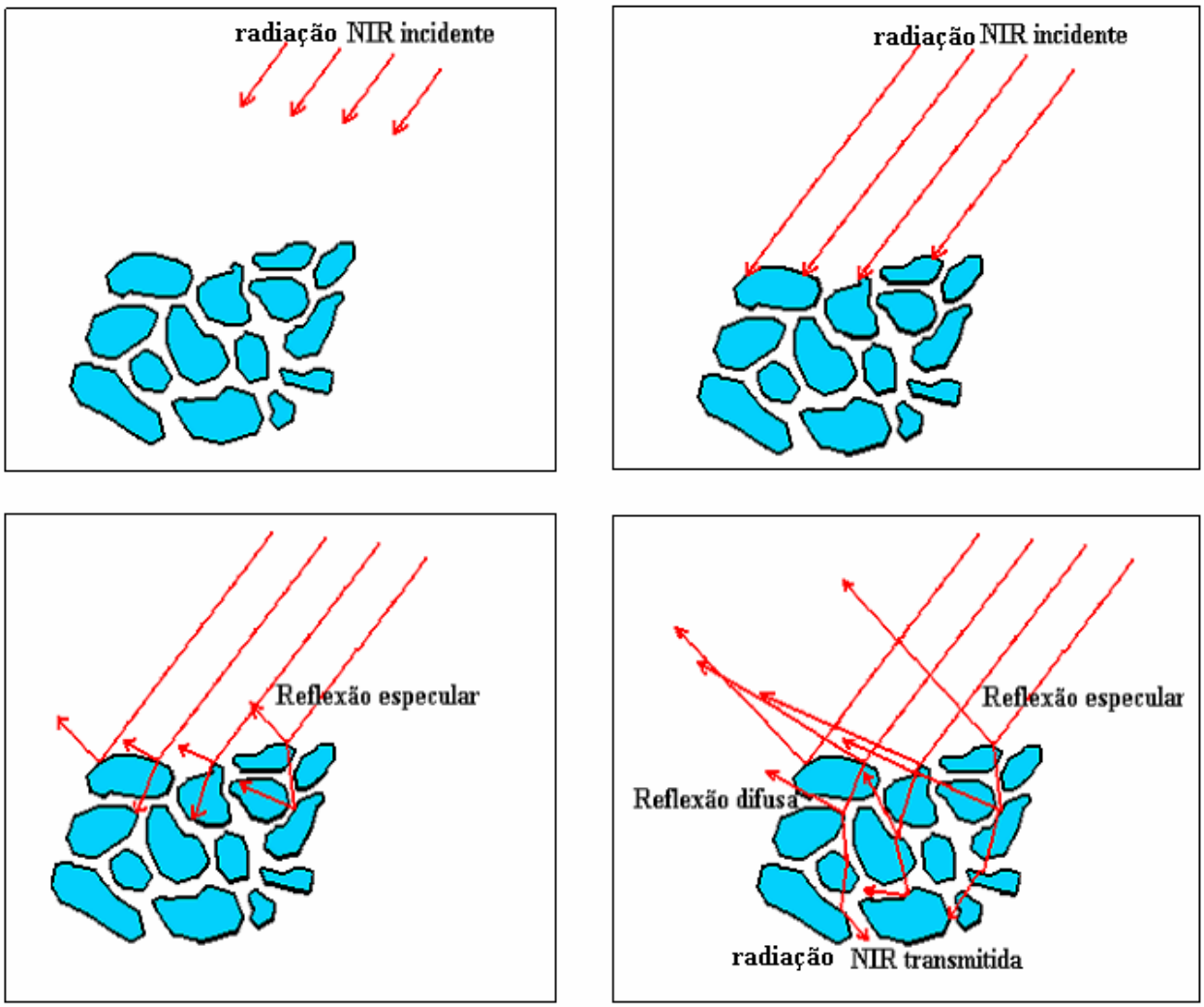


Figura 1.2. Demonstração da reflexão da REM NIRR numa amostra em pó.

Como descrito antes, a técnica de reflectância difusa é descrita pela penetração da radiação na amostra, percorrendo diferentes trajetórias até retornar ao detector (**Figura 1.2.**). Dessa forma o caminho percorrido pela radiação (caminho ótico) torna-se de difícil definição. Como resultado o espectro de reflectância de um dado composto não apresenta uma relação linear com a concentração. Um tratamento

teórico para este tipo de situação foi desenvolvido por Kubelka e Munk. Segundo esse modelo, é obtida a seguinte função entre reflectância medida e a concentração:

$$F(R)_\lambda = \frac{(1-R_\lambda)^2}{2R_\lambda} = K_{KM} \quad \textbf{(1.3)}$$

$$F(R)_{1+2+\ldots+n} = c_1F(R)_1 + c_2F(R)_2 + \ldots + c_nF(R)_n \quad \textbf{(1.4)}$$

onde *R* corresponde à reflectância medida na amostra em cada comprimento de onda (λ) sob condições em que a espessura da amostra pode ser considerada infinita, K_{KM} é o coeficiente de proporcionalidade que equivale ao produto da absortividade molar pelo caminho óptico, c é a concentração mássica e F(R) é a função de Kubelka-Munk

Como já foi dito, as medidas de reflectância difusa não requerem grande preparo da amostra, entretanto, para obter melhores resultados, algumas precauções devem ser consideradas: manter o tamanho da partícula de amostras sólidas (em pó) uniformizadas, o grau de compactação das amostras não deve sofrer variações significativas e a superfície das amostras deve estar plana.

A espectrometria de reflectância difusa no infravermelho próximo é um método capaz de fornecer informações quantitativas a respeito da composição da amostra e, portanto, torna-se uma ferramenta analítica muito útil para análise de diversos tipos de amostras sólidas. Entretanto, os sinais de baixa intensidade, a grande dificuldade em se elaborar uma satisfatória correlação entre os grupos de átomos envolvidos na molécula, bandas de difícil interpretação e de alta complexidade fazem com que o uso da espectrometria NIR para realizar análises precisas e confiáveis, dependa quase que totalmente da utilização de métodos quimiométricos para tratamento de dados multivariados.

1.7.2. Quimiometria

A evolução tecnológica e o avanço das diversas técnicas instrumentais de análises químicas têm impulsionado o crescente desenvolvimento de uma nova área da ciência química, conhecida como quimiometria, que a partir de então tem se

tornado uma ferramenta muito importante na área das ciências, tornando possível vencer grandes desafios.

A Quimiometria pode ser definida como o ramo da Química que utiliza métodos matemáticos e estatísticos em dados de origens distintas para obtenção de informação química[35]. A Quimiometria desempenha grande papel em análises que geram dados complexos e de pouca informação "a priori". De fato, o objetivo do tratamento de dados multivariados é extrair o máximo de informação presentes nos mesmos, com o mínimo de esforço, visando facilitar a sua interpretação[36,37].

A Quimiometria auxilia a espectrometria NIR nas mais variadas situações, como pré-processamento dos dados espectrais, planejamento e otimização de experimentos[38,39], processamento de sinais[40,41], seleção de variáveis e amostras[42-46], calibração multivariada[43-45], reconhecimento de padrões, classificação de amostras, transferência de calibração[46], etc.

As técnicas quimiométricas podem ser genericamente divididas em três classes distintas: análise exploratória de dados, construção de modelos quantitativos de calibração e construção de modelos qualitativos de classificação. A análise exploratória utiliza basicamente os métodos de Análises de Componentes Principais (PCA) e a Análise Hierárquica de Agrupamentos (do inglês: *Hierarquical Clustering Analysis* – HCA), enquanto que o modelo de classificação mais utilizado é o Modelagem Flexível Independente por Analogia de Classes (SIMCA)[35].

1.7.2.1. Pré-Processamento dos Dados

O tratamento inicial dos dados brutos, gerados numa análise, é feito na etapa de pré-processamento. Muitas vezes as análises instrumentais apresentam resultados com fortes ruídos instrumentais, diferenças em ordens de grandezas, variações sistemáticas na linha de base, etc. Nestes casos, antes da realização da modelagem multivariada propriamente dita, é necessário que se faça um pré-processamento para remover fontes de variações indesejáveis que as variáveis e/ou amostras possam apresentar ao conjunto de dados como um todo[35,47].

A apresentação de dados de natureza multivariada é feita, geralmente, em forma de tabelas. Nela os objetos (amostras) são distribuídos nas linhas e as variáveis (medidas de alguma propriedade das amostras) são colocadas nas colunas. Portanto, a organização deste conjunto de dados pode ser representada por uma matriz A de dimensão *m* x *n* (**Figura 1.3**). Essa Matriz (A) contém *m* amostras e *n* medidas experimentais. No caso de dados espectroscópicos, *n* podem ser os comprimentos ou números de onda.

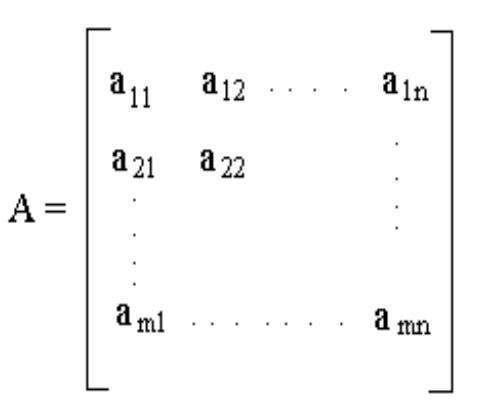


Figura 1.3. Representação de uma matriz de dados quaisquer.

O pré-processamento dos dados aplicado às variáveis pode ser realizado através de três maneiras: centralização dos dados na média, que consiste em subtrair o valor de cada elemento do vetor coluna (variável) pelo valor médio dos elementos dessa coluna; escalonamento, onde cada elemento de uma linha é dividido pelo desvio padrão de sua respectiva variável. Com isto, todos os eixos da coordenada são conduzidos a uma mesma escala e, conseqüentemente, cada variável fica com a mesma influência no modelo; e o auto-escalonamento que consiste em centralizar os dados na média e, em seguida, efetuar o escalonamento. Desta forma, as variáveis terão médias zero e desvios padrão igual a um. Tanto o escalonamento quanto o auto-escalonamento são utilizados quando se pretende atribuir os mesmos pesos às variáveis do sistema de investigação[35,36].

Além do pré-processamento dos dados aplicado às variáveis, também devem ser consideradas as variações existentes entre as amostras. Para contornar tais problemas os métodos mais utilizados são a derivação e suavização. Os métodos de

derivação são comumente aplicados quando, em um conjunto de dados, deseja-se corrigir problemas relacionados com a variação da linha de base. Além disso, permite uma melhor visualização de picos existentes nos sinais originais. A primeira derivada remove deslocamentos constantes (do inglês: *offsets*) da linha de base e a segunda derivada elimina uma variação linear da linha de base, normalmente devido a efeitos de espalhamento. Como o cálculo das derivadas é feito utilizando-se diferenças entre valores de pontos adjacentes do conjunto de dados, a relação sinal/ruído torna-se pior e, por isso, antes da derivação é comum aplicar-se aos dados algum tipo de suavização (do inglês: *smoothing*). Para isto, o algoritmo mais utilizado é o de Savitzky-Golay[35]. Este algoritmo ajusta um polinômio de baixa ordem aos pontos de uma janela pelo método dos mínimos quadrados. A escolha do número adequado de pontos é muito importante para que não haja perda de informação nem tão pouco a permanência de ruídos[35,36].

É importante frisar que não existe uma regra específica para a aplicação de pré-processamentos em amostras e/ou variáveis, pois estes dependem da natureza dos dados e, por isso, devem ser avaliados preliminarmente cabendo ao analista a decisão de tal aplicação, o qual deve ter todo o entendimento do processo químico ou físico do sistema em estudo.

1.7.2.2. Técnicas Quimiométricas

1.7.2.2.1. Análise Hierárquica de Agrupamentos (HCA)

A HCA é uma técnica de reconhecimento de padrões não-supervisionada que examina as similaridades ou diferenças existentes entre as amostras ou variáveis de um conjunto de dados multivariados[35].

O exame é feito utilizando distâncias interpontos, correspondentes às *m* amostras ou *n* variáveis no espaço n-dimensional de uma matriz A qualquer, representando-as em forma de um gráfico bidimensional chamado *dendrograma* (**Figura 1.4**). Trata-se de um gráfico facilmente compreensível, tornando possível reconhecer os agrupamentos formados em função da similaridade[48].

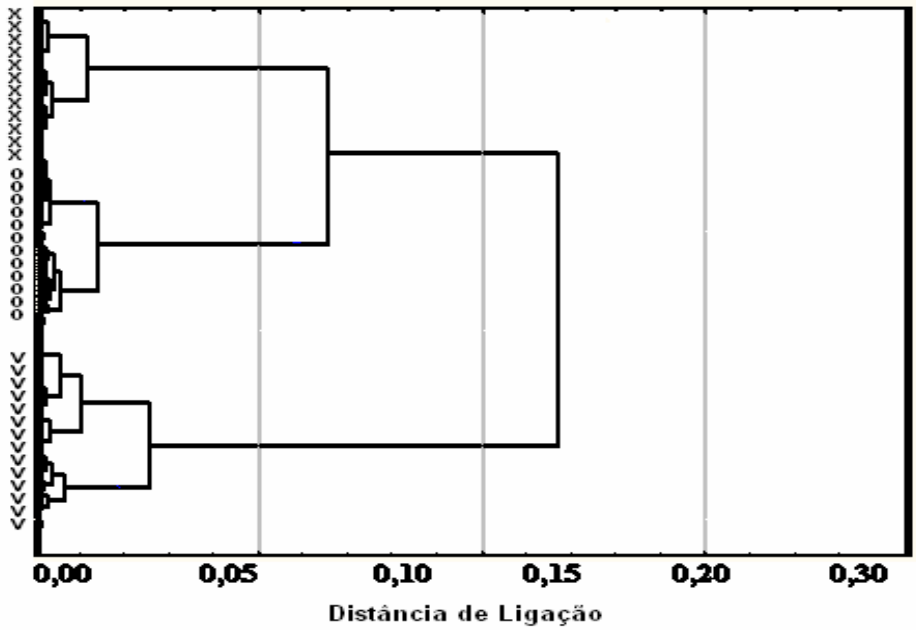


Figura 1.4. Dendrograma mostrando a separação de amostras em três agrupamentos, representados pelos símbolos x, o e v, usando uma distância interpontos quaisquer.

A similaridade, S_{xy}, entre duas amostras x e y quaisquer é calculada pela seguinte equação:

$$S_{xy} = 1 - \frac{d_{xy}}{d_{max}} \qquad \textbf{(1.5)}$$

onde d_{xy} é a distância interpontos entre as amostras x e y e d_{max} é a distância máxima entre todas as amostras consideradas. Daí as amostras são ditas similares se elas apresentarem valores de S_{xy} muito próximos de um[35].

Na construção de *dendrogramas* cada amostra ou variável é considerada, inicialmente, como um agrupamento. Usando um método de cálculo de distância interpontos apropriado, definem-se as amostras ou variáveis mais próximas, formando subagrupamentos. Os subagrupamentos formados são unidos através de uma técnica de conexão até que se tenha um único agrupamento com membros suficientemente similares entre si.

Há diversas maneiras de se calcular as distâncias interpontos, associados às amostras ou variáveis, como por exemplo: distância de Mahalanobis, distância Euclidiana, distância de Chebychev, coeficiente de correlação de Pearson, etc.

Dentre os métodos mais utilizados podemos destacar a distância de Chebychev[48]. Essa medida de distância pode ser apropriada em casos em que se necessita definir dois objetos como "diferentes", se eles são diferentes em qualquer uma das dimensões. A distância é calculada da seguinte forma:

$$D_{(chebychev)}(x,y) = \text{Máximo}|\mathbf{x_i} - \mathbf{y_i}| \quad \textbf{(1.6)}$$

em que x_i e y_i são as coordenadas das amostras x e y na i-ésima dimensão do espaço linha ("i" varia de 1 a *n* variáveis)[47].

Depois de definido o método para calcular a distância interpontos é necessário escolher o critério com que os subagrupamentos serão ligados, e para isso diversas técnicas de conexão são utilizadas: single-linkage (conexão simples), complete-linkage (conexão completa), centróide-link, método de Ward's, etc.

O método de Ward's revela-se bastante eficiente e se destaca por priorizar a formação de pequenos agrupamentos. Este método usa a análise da variância aproximada para avaliar as distâncias entre os subagrupamentos[47].

Sendo assim, numa análise por HCA é sempre recomendável que se testem os diferentes tipos de distância interpontos e técnicas de conexão, escolhendo aqueles que melhor representem a realidade informativa dos dados originais.

1.7.2.2.2. Análise de Componentes Principais (PCA)

A PCA é uma técnica quimiométrica de reconhecimento de padrões não-supervisionada que faz manipulações matemáticas num conjunto de dados de natureza multivariada, objetivando a representação das variações existentes em forma de fatores ou componentes principais (PC`s) (**Figura 1.5**). As PC's são as representações dos novos eixos formados a partir da combinação linear das variáveis originais e nelas estão contidas as informações mais relevantes dos dados. Elas são sempre ortogonais (perpendiculares) umas as outras, e sucessivas PC's descrevem quantidades decrescentes da variância explicada dos dados. Em geral, poucas PC's são necessárias para que toda a variância dos dados seja explicada[35,36].

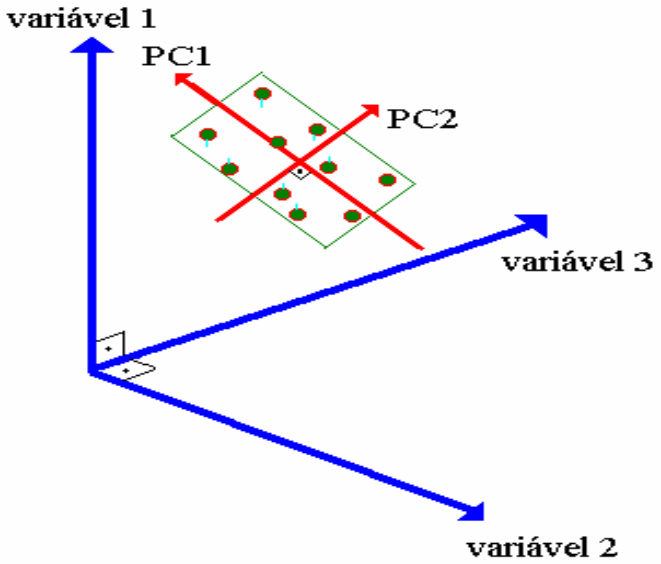


Figura 1.5. Primeira e segunda PC`s no espaço tridimensional das variáveis.

Da mesma forma que as amostras têm coordenadas definidas pelas variáveis originais, elas também possuem coordenadas relativas aos novos eixos PC's e são denominadas escores (do inglês*: scores*). A contribuição que cada variável original exerce sobre uma determinada PC é denominada peso (do inglês*: loading*) que, matematicamente, pode ser definida como sendo o cosseno do ângulo entre o eixo da variável e o eixo da PC. Se um determinado número de PC's não conseguir descrever toda a variância dos dados, então essa parte não descrita será representada em uma nova matriz chamada de resíduos[35,36].

O número de PC's necessários para descrever os dados é definido como sendo a dimensionalidade inerente (DI). A DI está relacionada a uma característica física e/ou química encontrada no conjunto de dados. A partir desta é possível definir o número adequado de PC's que fornecem as informações de interesse para o analista, para que não se inclua muito ruído nem tão pouco exclua informações pertinentes e assim, prejudique a interpretação dos dados. Para isso, recorre-se ao uso de ferramentas de diagnóstico, que investigam e exploram três aspectos do conjunto de dados: o modelo, as amostras e as variáveis. Dentre estas ferramentas, pode-se citar o gráfico dos resíduos, gráfico dos escores e o gráfico dos pesos. Outro ponto tão importante quanto à determinação do número de PC's é a detecção de amostras anômalas (*outliers*). Estas amostras apresentam características não esperadas pelas

ferramentas e por isso podem afetar nos resultados obtidos, podendo então ser descartadas do conjunto de dados[35,36].

A Análise de Componentes Principais (PCA) é uma técnica bastante utilizada e serve de base para a maioria das outras técnicas quimiométricas multivariadas, inclusive as técnicas de reconhecimento de padrões supervisionado como o SIMCA.

1.7.2.2.3. SIMCA

O SIMCA é uma técnica de reconhecimento de padrões supervisionado que utiliza a análise de componentes principais (PCA) para modelar a forma e a posição das classes no espaço, formados pelas amostras, para definir uma classe[35]. Uma embalagem ou caixa multidimensional é construída para cada classe e a classificação de amostras é feita determinando-se em qual das embalagens (classe) a amostra se encaixa ou pertence[49]. Esta técnica também é capaz de detectar se uma amostra desconhecida pertence ou não a uma das caixas modeladas. A forma como uma amostra desconhecida (A) é classificada usando o modelo SIMCA está ilustrada na **Figura 1.6**:

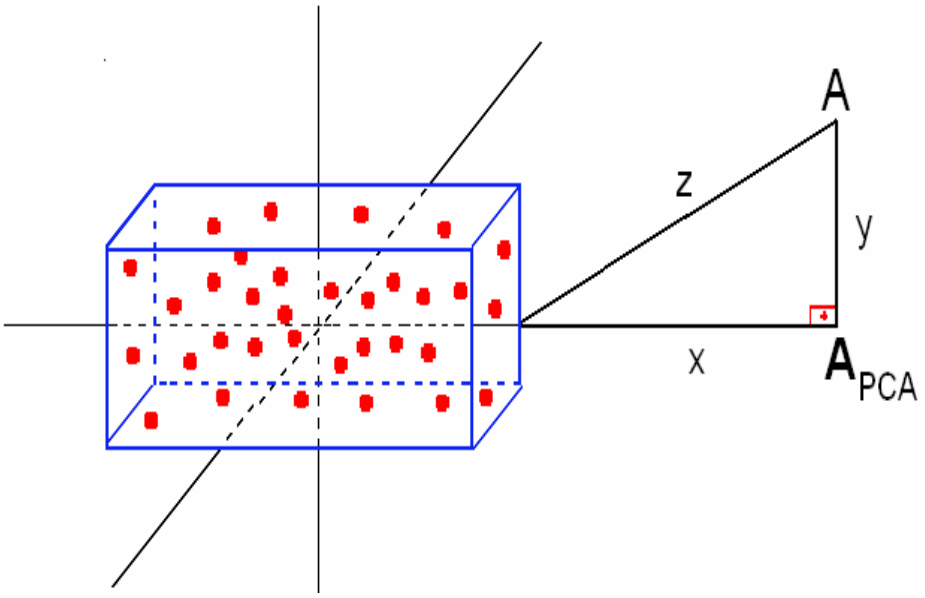


Figura 1.6. Esboço de um modelo SIMCA, onde "*y*" é a distância entre a amostra desconhecida A e o eixo da PC, "*x*" é a distância entre a projeção de A na direção da PC e a fronteira da classe e "*z*" é a distância de A até a caixa.

O valor de z^2 é calculado pela seguinte expressão:

$$z^2 = x^2 + y^2 \quad (1.7)$$

Ao dividir o valor de z^2 calculado pela variância da classe obtêm-se um valor chamado de F_{calc}. Esse valor de F_{calc} será comparado com outro valor chamado F_{crit} que é escolhido estatisticamente ou a partir de uma tabela. Esse procedimento é denominado de teste F e a amostra A será classificada como pertencente à classe das amostras se o $F_{calc} < F_{crit}$. Um dos métodos utilizados pelos pacotes quimiométricos[50] para representar a classificação feita pelo SIMCA é o gráfico bidimensional S_i x H_i (**Figura 1.7**).

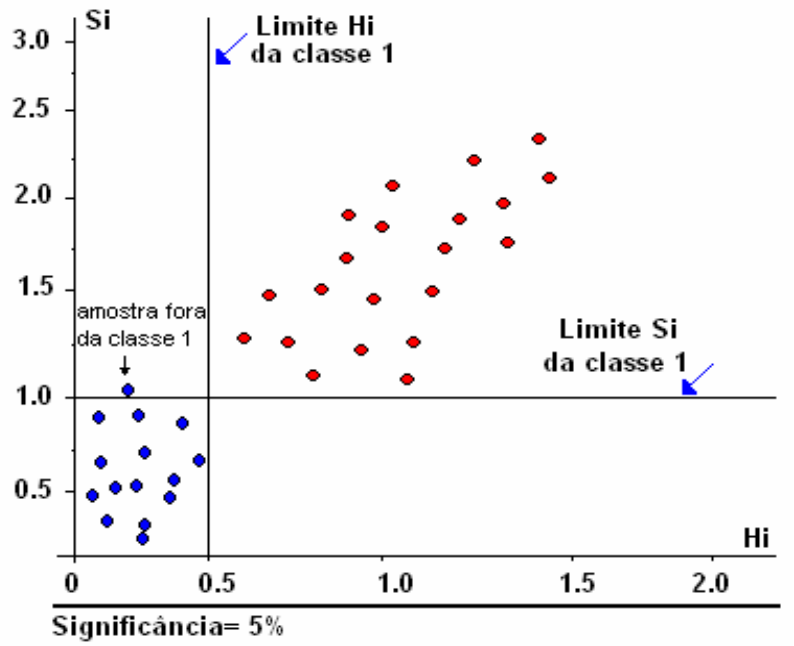


Figura 1.7. Gráfico de S_i x H_i, mostrando a classificação de amostras quaisquer pertencentes às classes 1 (azul) e 2 (vermelho), realizada pelo SIMCA.

Este gráfico mostra os dois limites usados para a classificação: a distância da "nova" amostra em relação ao modelo da classe (raiz quadrada da variância residual, S_i) e a influência (Leverage, H_i) que é a distância da amostra projetada ao centro do modelo. Então as amostras que se encontram dentro de ambos os limites para uma classe específica são ditas pertencentes àquela classe.

1.7.2.2.4. Análise Discriminante Linear (LDA)[51,52]

A análise Discriminante Linear (LDA) é uma técnica que consiste em estimar uma combinação linear de duas ou mais variáveis independentes, chamada de função discriminante, que pode distinguir os pontos relativos a dados de dois ou mais grupos diferentes. Se essa função de fato existe, então é possível dizer que os pontos pertencentes a esses grupos são linearmente separáveis. A discriminação é feita determinando-se o conjunto ótimo de pesos para as variáveis independentes de tal maneira que se maximize a variância entre os grupos em relação à variância dentro dos grupos. Uma das maneiras mais simples de se obter uma função linear discriminatória é ilustrada na **Figura 1.8**.

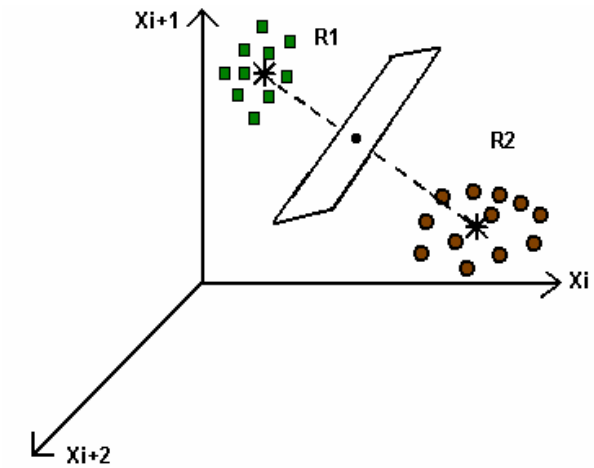


Figura 1.8. A técnica da análise discriminante linear. Os centros de gravidade dos grupos definidos por círculos e quadrados são determinados e em seguida traça-se o plano de decisão perpendicularmente à linha que une os centros de gravidade. O espaço fica dividido em duas regiões R1 e R2.

Se um ponto qualquer aparece na região R1, é classificada como pertencente à categoria 1 e 2 se aparece na região R2.

É importante citar que numa análise discriminante, as amostras são sempre classificadas como pertencente a uma ou outra categoria e, portanto não contempla a possibilidade da amostra em estudo pertencer a uma categoria ainda não definida.

Ao considerar as análises espectroscópicas em diversos tipos de aplicações qualitativas, constata-se que, atualmente, a grande maioria emprega métodos de reconhecimento de padrões conforme descritos acima. Contudo, uma etapa crítica para aumentar a capacidade preditiva dos modelos que utilizam tais métodos é a

seleção de variáveis. Esta etapa deve, idealmente, eliminar variáveis não-informativas e reter aquelas que resultem na exatidão máxima, principalmente quando ocorre alta sobreposição espectral.

1.7.2.2.5. Seleção de Variáveis

O aperfeiçoamento de instrumentos capazes de registrar um espectro em diversos tipos de regiões e em curtos intervalos de tempo veio levantar um aspecto importante em problemas de classificação: a conveniência de se utilizar modelos de classificação com todos os comprimentos de onda em que foram efetuadas as medidas. Embora tenham sido propostos modelos que processam a região espectral inteira[53-57], alguns trabalhos mostram que a capacidade preditiva dos modelos pode ser melhorada mediante uma seleção conveniente de variáveis. Vários algoritmos têm sido propostos para seleção de variáveis[58-61], dentre eles podemos destacar o algoritmo das projeções sucessivas - APS (do inglês: *Successive Projection algorithm* – SPA), desenvolvido por Araújo et al[42] para seleção de variáveis no contexto da calibração multivariada, em particular, quando aplicado a modelos de regressão linear múltipla (do inglês: *Multiple Linear Regression* –MLR).

1.7.2.2.5.1. Algoritmo das projeções Sucessivas (SPA)

O objetivo do SPA consiste em buscar um conjunto representativo pequeno de variáveis espectrais com ênfase na minimização de colinearidade. Desta maneira, torna-se possível usar modelos MLR's que, embora simples e relativamente de fácil interpretação, podem ser severamente afetados por problemas de colinearidade.[35]

Os fundamentos teóricos, bem como os detalhes das projeções do SPA têm sido extensivamente apresentados em outros trabalhos[41-45]. Apesar disto, faz-se necessário uma apresentação de um exemplo ilustrativo.

Considere um conjunto de j = 5 variáveis espectrais: X1=400 nm; X2= 450 nm; X3= 500 nm; X4= 550 nm e X5= 600 nm para a determinação de dois analitos

(A=2) usando 3 misturas de calibração (Mcal=3). Para a seleção das variáveis, a seqüência de operações do algoritmo é realizada na seguinte ordem:

a) Para o cálculo das projeções, toma-se como vetor (k(0)) de partida aquele que possuir a maior norma. Contudo, os demais vetores são posteriormente testados para a formação de cadeias de variáveis candidatas.

b) Cálculo das projeções dos demais vetores (comprimentos de onda) em um subespaço ortogonal ao do vetor inicial, conforme **Figura 1.9**.

c) Organização das cadeias de variáveis, onde cada cadeia é composta pelo vetor de partida (neste exemplo X_3) além das variáveis mais ortogonais (X_1 e X_5).

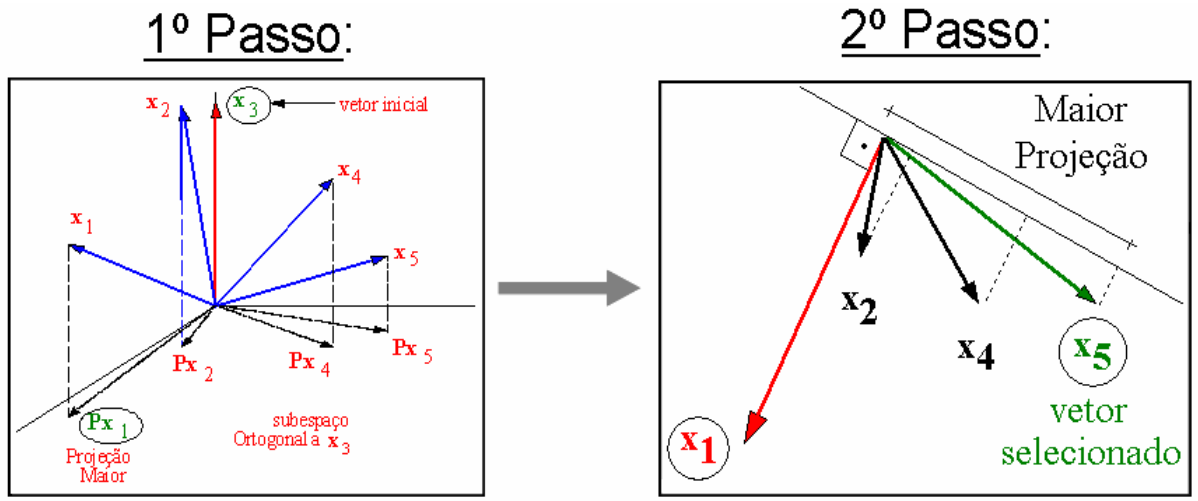


Figura 1.9. Exemplo de operações de projeções efetuadas com o uso do APS para selecionar comprimentos de onda a partir de 5 variáveis espectrais (J =5), 3 amostras de calibração (Mcal = 3), e comprimento de onda tomado como partida (k(0)) = 3, resultado da primeira iteração: k(1)=1.

De acordo com o problema decorrente de inversão de matrizes em modelos MLR's, o número de variáveis selecionadas em cada cadeia é limitado pelo número de misturas (amostras) do conjunto de calibração. Na demonstração deste exemplo, existem 3 amostras de calibração. Portanto, este será o número máximo de comprimentos de onda que serão selecionados para a cadeia de variáveis.

A **Figura 1.9** mostra que a primeira cadeia de variáveis é formada pelos comprimentos de onda X3, X1 e X5 que seriam (500, 400 e 600 nm).

Para se determinar qual a melhor cadeia de variáveis, faz-se uma regressão por MLR para cada uma das cadeias, empregando-se como matriz X os

comprimentos de onda da cadeia de variáveis e como matriz Y as concentrações dos dois metais avaliados. Realiza-se, através da equação 1.8, o cálculo do erro quadrático médio de previsão (RMSEP) de amostras de um conjunto de teste associado a cada um dos modelos MLR construídos.

$$RMSEP = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(y^m - \hat{y}^m)^2} \qquad \textbf{(1.8)}$$

onde y^m e $\hat{y}^m$ são os valores de referência e previsto pelo modelo, de um parâmetro de interesse. A cadeia de variáveis, cujo modelo MLR levar ao menor RMSEP será escolhida pelo algoritmo.

1.7.2.2.5.2. SPA para Classificação

Recentemente, Pontes et al[41] adaptaram o SPA para o contexto dos métodos de classificação. Neste caso, a informação da modelagem consiste nos dados das respostas instrumentais (matriz X) e um índice de classe para cada amostra. Como se trata de classificação, os valores de referência dos y parâmetros não podem ser disponíveis para processar a estatística do RMSEP (Função de custo adotada pelo SPA-MLR para a seleção das variáveis)[42]. Então, quando um subconjunto de variáveis sob avaliação é empregado, a função de custo adotada é o risco médio ***G*** de uma classificação incorreta, dada pela Análise Discriminante Linear.

Esta função é calculada no conjunto de validação como:

$$G = \frac{1}{Kv}\sum_{k=1}^{Kv} g_k \qquad \textbf{(1.9)}$$

onde $\mathbf{g_k}$ (risco de uma classificação incorreta de $\mathbf{K^{th}}$ validação da amostra $\mathbf{x_k}$) está definido como :

$$g_k = \frac{r^2(\mathbf{x}_k, \boldsymbol{\mu}_{Ik})}{\min_{Ij \neq Ik} r^2(\mathbf{x}_k, \boldsymbol{\mu}_{Ij})} \qquad \textbf{(1.10)}$$

Na equação anterior (1.10), o numerador $r^2(x_k,\mu_{Ik})$ é o quadrado da distância de Mahalanobis entre o objeto X_k (de índice de classe *Ik*) e a média da amostra μ_{Ik} de sua classe verdadeira (ambos os vetores da linha). Esta distância é calculada como:

$$r^2(\mathbf{x}_k, \boldsymbol{\mu}_{Ik}) = (\mathbf{x}_k - \boldsymbol{\mu}_{Ik})\boldsymbol{\Sigma}^{-1}(\mathbf{x}_k - \boldsymbol{\mu}_{Ik})^T \quad \textbf{(1.11)}$$

onde a média da amostra μ_{Ik} e a covariância Σ são calculados com os dados do conjunto de calibração.

1.8. Objetivo do Trabalho

O objetivo deste trabalho é desenvolver uma nova metodologia para classificação de diferentes tipos de cigarros, com intuito de auxiliar na fiscalização da qualidade dos mesmos. Para isso, propõe-se o uso combinado da espectrometria NIRR com o algoritmo das projeções sucessivas (SPA) e a análise discriminante linear (LDA).

C A P Í T U L O 2

EXPERIMENTAL

2.1. Aquisição das Amostras

As amostras de cigarros foram adquiridas em diversos estabelecimentos comerciais da cidade de João Pessoa – PB, no período de setembro a outubro de 2006. No total, foram coletadas 210 amostras, sendo:

- *45 amostras da marca A;*
- *57 amostras da marca B;*
- *57 amostras da marca C;*
- *51 amostras da marca D.*

Para garantir uma maior representatividade e variabilidade da composição das amostras de cigarros, a amostragem foi feita de forma aleatória e sem repetição nos lotes. Para tanto as amostras foram coletadas individualmente, ou seja, em cada estabelecimento comercial foi coletado apenas um único cigarro por maço.

2.2. Divisão do Conjunto de Amostras

Para a realização da modelagem SIMCA e NIRR-SPA-LDA, os dados foram divididos em três conjuntos: treinamento, teste e previsão utilizando o algoritmo Kennard-Stone (KS)[62] que é um método clássico para extrair um conjunto representativo de amostras em um determinado conjunto de dados. O algoritmo foi aplicado separadamente às amostras de cada uma das marcas de cigarros. O conjunto de treinamento foi inicialmente separado e as amostras remanescentes foram divididas, alternadamente, em teste e previsão. A **Tabela 2.1.** mostra o número de amostras selecionadas pelo KS para cada marca de cigarro.

Tabela 2.1. Número de amostras de treinamento, teste e previsão selecionadas pelo KS para as quatro marcas de cigarros.

Marca	Conjuntos		
	Treinamento	Teste	Previsão
A	27	12	12
B	27	15	15
C	25	10	10
D	27	15	15

2.3. Equipamentos

Para a obtenção dos espectros NIRR, foi utilizado um espectrômetro de Infravermelho com Transformada de Fourier, FTIR, da Perkin Elmer, modelo Spectrum, Série GX, apresentado na **Figura 2.1.**

(a) **(b)**

Figura 2.1. **(a)** Fotografia do Espectrômetro com Transformada de Fourier da Perkin Elmer. **(b)** *Zoom* do acessório de medida de reflectância difusa no espectrômetro.

Um acessório RD, acoplado ao equipamento, foi utilizado para o registro dos espectros. Este acessório é composto de uma base, uma plataforma e um recipiente metálico, denominado "panelinha", onde é colocada a amostra a ser analisada. (ver detalhes na **Figura 2.2.**).

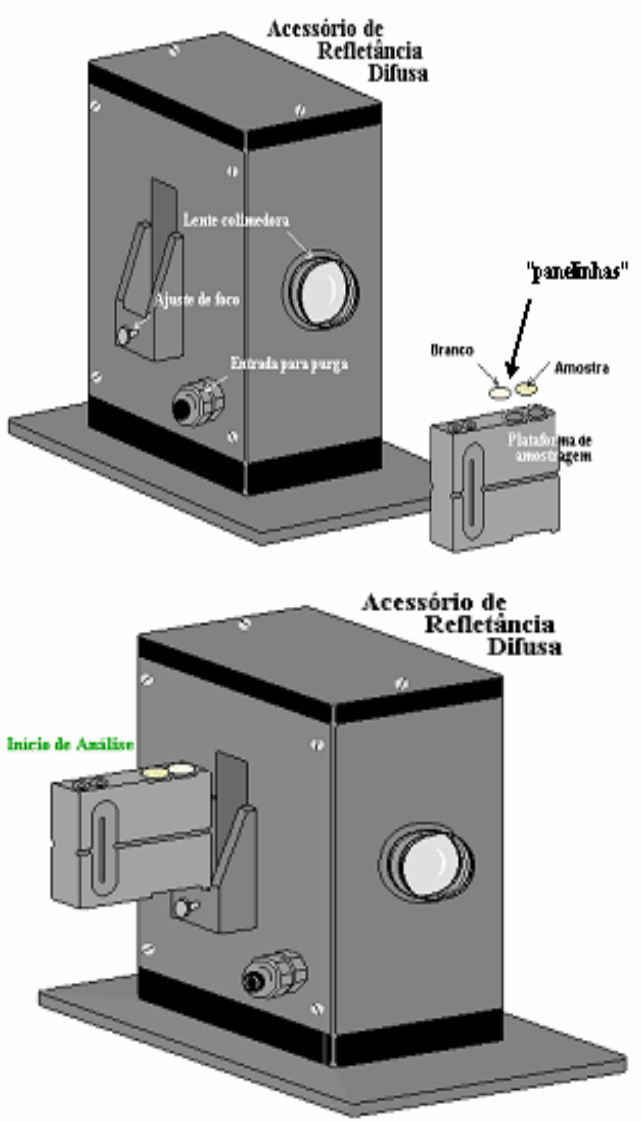


Figura 2.2. Ilustração do acessório de reflectância difusa do espectrômetro Perkin Elmer.

2.4. Procedimento Analítico

2.4.1. Preparação da Amostra

Antes de registrar os espectros NIRR das amostras de cigarro, estas foram submetidas a algumas etapas de processamento físico para melhor se adaptarem às condições de medidas do espectrômetro NIRR e para atenderem às precauções exigidas pela técnica de reflectância difusa, citadas na **Seção 1.8.1**. As etapas do processamento foram:

1. Em cada cigarro foi extraído todo o fumo e descartados o papel e o filtro;
2. O fumo foi transferido para um envelope de papel manteiga e secado na estufa a 60°C por um período de 24 horas (**Figura 2.3 a**);

3. Após a etapa 2, o material seco foi macerado e em seguida peneirado (**Figura 2.3 b**);
4. Por fim, as amostras foram guardadas no dessecador até o momento das medidas.

(a) **(b)**

Figura 2.3. **(a)** Fotografia do fumo seco e **(b)** do fumo seco, macerado e peneirado.

2.4.2. Registro dos Espectros NIRR

O registro do sinal do branco foi feito com brometo de potássio (KBr) macerado e peneirado da mesma forma que as amostras.

Cada uma das 210 amostras foi analisada em triplicata, ou seja, tomaram-se três porções, do fumo macerado e peneirado, de cada cigarro e registrou-se o espectro, gerando um total de 630 espectros. Em todo tratamento quimiométrico foi sempre utilizado, para cada cigarro, o espectro médio de sua triplicata. Todos esses espectros foram sempre registrados na região de 15.000 a 2.700 cm^{-1}, empregando uma resolução de 4 cm^{-1}, 32 varreduras (32 *scans*), a uma temperatura em torno de 24°C e umidade relativa do ar em torno de 48%. Esses espectros foram arquivados e posteriormente utilizados para a aplicação das ferramentas quimiométricas de análise, utilizadas para o desenvolvimento da nova metodologia proposta, para classificação de diferentes tipos de cigarros.

2.5. Software

O programa *Unscrambler®* 9.7 (CAMO S.A.) foi utilizado para o pré-processamento dos dados, PCA e SIMCA. A HCA e a aplicação dos algoritmos KS e SPA-LDA foram realizados utilizando os pacotes quimiométricos *Statística®* 6.0 e *Matlab®* 6.1, respectivamente.

"Todas as coisas me são lícitas, mas nem todas as coisas convêm".
Corintios 1:12.

C A P Í T U L O 3

RESULTADOS E DISCUSSÕES

2.5. Seleção da Região Espectral de Trabalho

Os espectros NIRR das 210 amostras de cigarros foram registrados em toda faixa operacional do espectrômetro de medida, 15000 a 2700cm^{-1} (660 – 3700 nm), conforme mostrado na **Figura 3.1**.

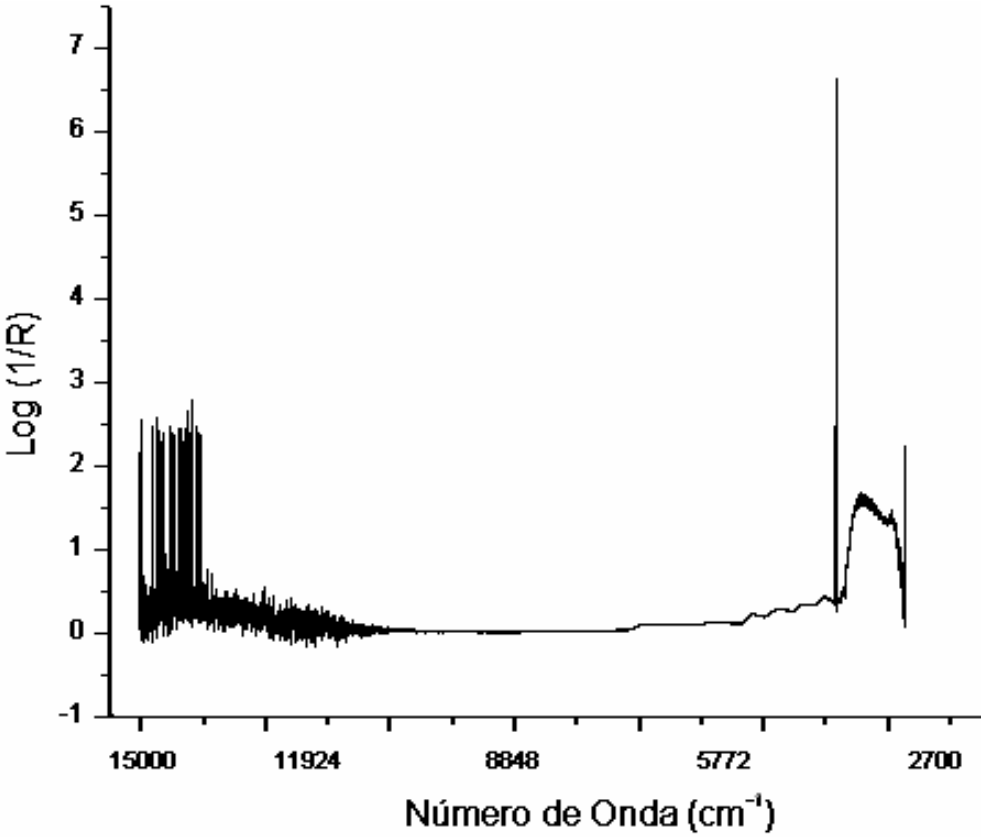


Figura 3.1. Espectro de uma amostra de cigarro na região de 2700 a 15000 cm^{-1}.

As regiões espectrais que compreendem os intervalos de 15000-10000 cm^{-1} e 4253-2700 cm^{-1} foram descartadas, pois apresentavam sinais muito ruidosos e/ou saturados. O intervalo de 10000 a 5419 cm^{-1} também foi descartado, pois neste caso os sinais de reflectância eram tão baixos que se confundiam com o sinal da linha de base (baixa relação sinal/ruído). Portanto, a sub-região compreendida na faixa de 5420 a 4252 cm^{-1} foi selecionada como a região de trabalho para a aplicação das ferramentas quimiométricas de análise e o desenvolvimento da nova metodologia para classificação de diferentes tipos de cigarros, com intuito de auxiliar na fiscalização da qualidade dos mesmos. Na **Figura 3.2** é mostrado um espectro de uma amostra na região de trabalho selecionada.

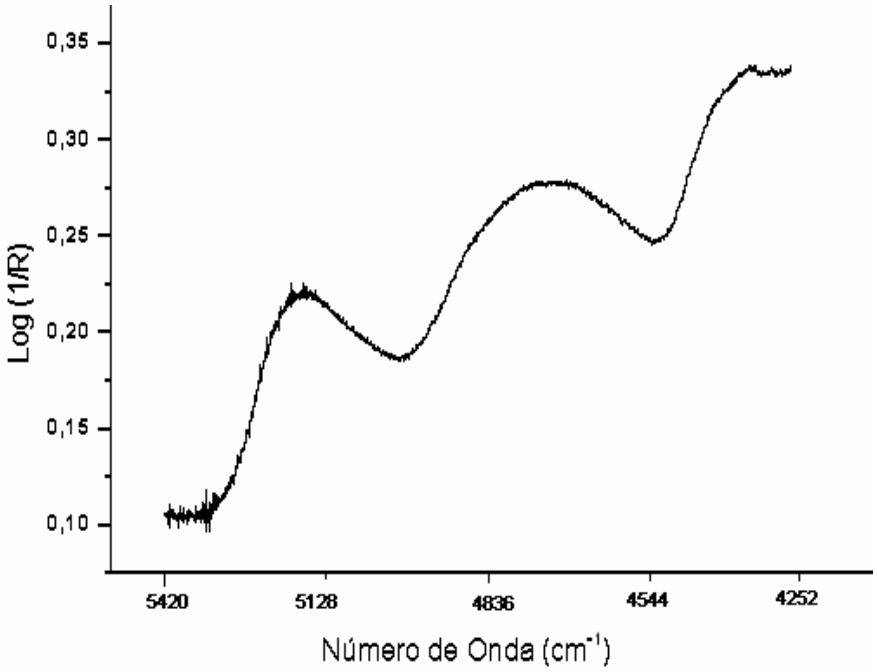


Figura 3.2. Espectro de uma amostra de cigarro na região espectral de trabalho selecionada.

3.2. Associações de Grupos Funcionais às Bandas dos Espectros NIRR[28,63,64]

O espectro da **Figura 3.2** mostra que a região selecionada das amostras de cigarros apresenta, essencialmente, três bandas distintas. Tratam-se de bandas largas e ruidosas. A primeira banda que aparece na região entre 5300 e 4970 cm^{-1} (1886 e 2012 nm) com o pico máximo em 5183 cm^{-1-} (1929 nm) está associada tanto ao segundo sobreton de transições vibracionais fundamentais relativas à ligações C=O de grupos carboxílicos e ésteres, como também a bandas de combinação relativas à ligação OH de água. A segunda banda, bastante larga e arredondada compreende o intervalo 4960-4500 cm^{-1} (2016 e 2222 nm) e pode ser atribuída tanto às bandas de combinação de transições fundamentais relativas às ligações OH de álcoois, mas principalmente, a bandas de combinação de ligação NH de compostos nitrogenados. Finalmente, tem-se uma terceira banda que vai de 4490 cm^{-1} até 4252 cm^{-1}. Este intervalo pode ser associado às bandas de combinação relativas às ligações CH e CC de compostos alifáticos e aromáticos. Como o cigarro contém milhares de compostos químicos, fica difícil a associação destas bandas a algum composto em específico, mas podemos destacar alguns importantes, como a nicotina, o metanol, a amônia,

(**Figura 3.3**), o monóxido de carbono e a água, todos eles presentes na composição química dos cigarros.

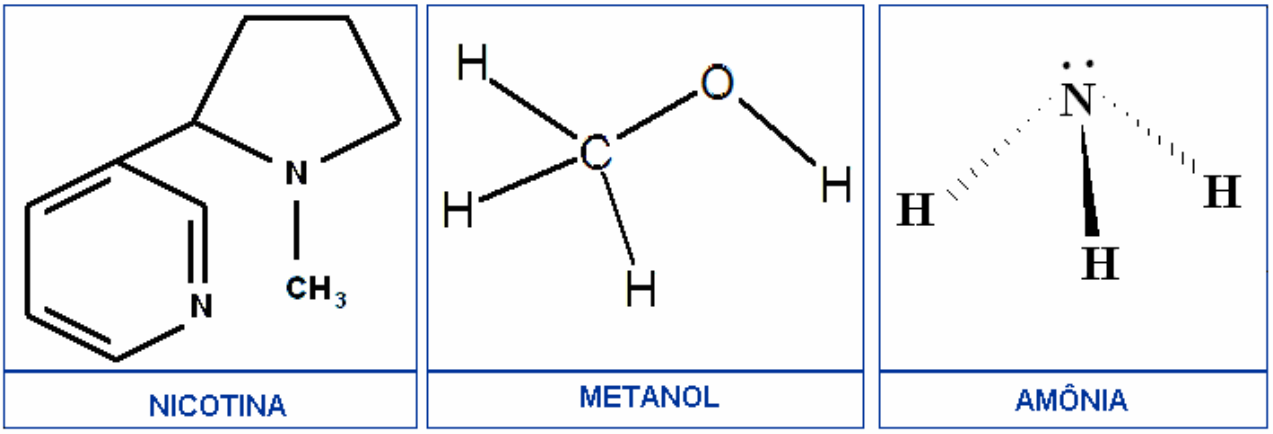


Figura 3.3. Representação das moléculas de Nicotina, Metanol e Amônia.

3.3. Pré-Processamento dos espectros NIRR

3.3.1. Pré-Processamento Aplicado às Amostras

Observou-se que os espectros das amostras de cigarros apresentaram problemas de variação sistemática da linha de base, bem como a presença de ruídos. Para correção desses problemas foram testados vários tipos de pré-processamentos aplicáveis às amostras. Melhores resultados foram alcançados empregando o método de derivação (1ª derivada) de Savitzky-Golay[35], usando um polinômio de segunda ordem e uma janela de 121 pontos. Os espectros derivativos médios das 210 amostras de cigarros, resultantes da aplicação desse pré-processamento, são apresentados na **Figura 3.4**.

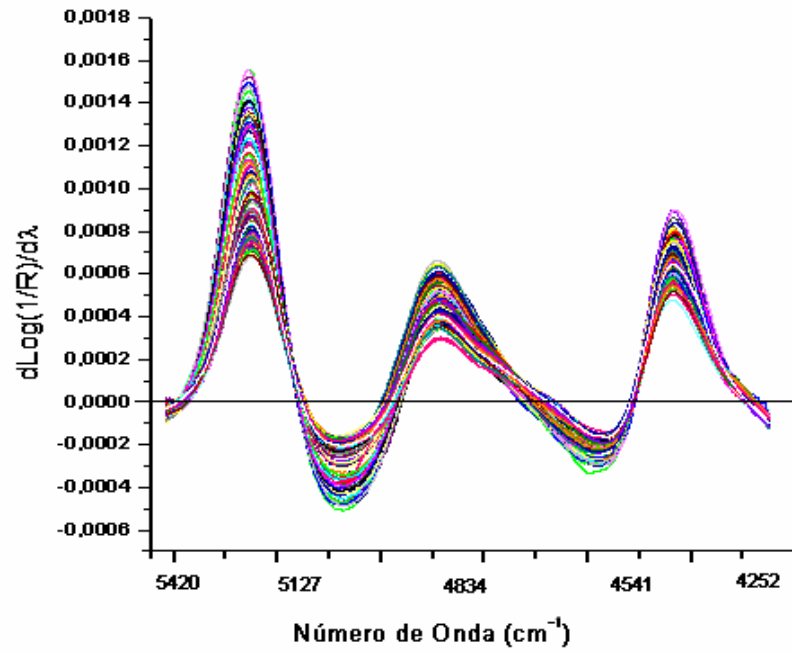


Figura 3.4. Espectros NIRR derivativos médios das 210 amostras de cigarros.

3.2.2. Pré-Processamento Aplicado às Variáveis

A eliminação de algumas variáveis não informativas (15000 a 5419 cm^{-1} e 4253 a 2700 cm^{-1}) e a escolha da região de trabalho (5420 a 4252 cm^{-1}), conforme descrito na **Seção 3.1**, foi o primeiro pré-processamento realizado sob as variáveis. Além deste, a centralização dos dados na média se fez necessário.

A **Figura 3.4** mostra que todas as amostras analisadas possuem perfis espectrais semelhantes e sobrepostos. Isto está relacionado à grande semelhança existente na composição química dos cigarros analisados, bem como a natureza complexa do sinal obtido pelo NIRR. Desta forma, torna-se impossível a distinção entre as diferentes classes de cigarros por meio de uma simples inspeção visual dos espetros. Portanto, o uso de ferramentas quimiométricas, conforme descrito na **Seção 1.9**, torna-se indispensável para caracterização e discriminação dos diferentes tipos de amostras de cigarros aqui utilizados.

3.4. Análise Exploratória dos Dados

Com o propósito de investigar se havia similaridades e/ou diferenças entre as amostras dos cigarros analisados, usando espectros NIRR, realizou-se uma análise exploratória no conjunto dos dados pré-processados, aplicando as técnicas de reconhecimentos de padrões não-supervisionados HCA e PCA.

3.4.1. Aplicação da HCA

Como técnica de reconhecimento de padrões não supervisionado, a HCA foi utilizada para identificar as possíveis formações de *clusters* ou agrupamentos nos dados pré-processados.

Foram testadas várias medidas de distâncias e técnicas de conexão de agrupamentos em todo o conjunto das 210 amostras analisadas. Dentre todos os dendrogramas construídos, o que apresentou os melhores resultados utilizou a medida de distância *Chebychev* e o método de Ward's (**Seção 1.9.2.1**), como técnica de conexão. O dendrograma resultante é mostrado na **Figura 3.5**.

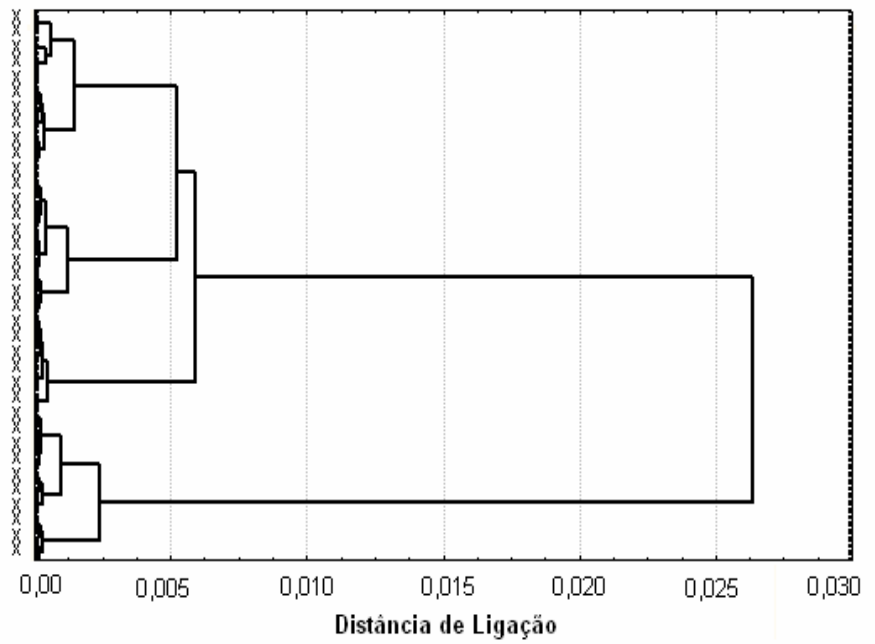


Figura 3.5. Dendrograma de uma HCA das 210 amostras de cigarros não rotulados, obtidos usando a distância de *Chebychev* e regra de ligação baseado no método de Ward´s.

Pode-se observar que a uma distância próxima a 0,0065 define-se a existência de dois agrupamentos de amostras. Já numa distância próxima a 0,005 constata-se a formação de quatro agrupamentos. Ao rotular as amostras, fica evidente que os quatro agrupamentos revelados referem-se aos 4 diferentes tipos de cigarros (A, B, C e D), conforme mostrado na **Figura 3.6**.

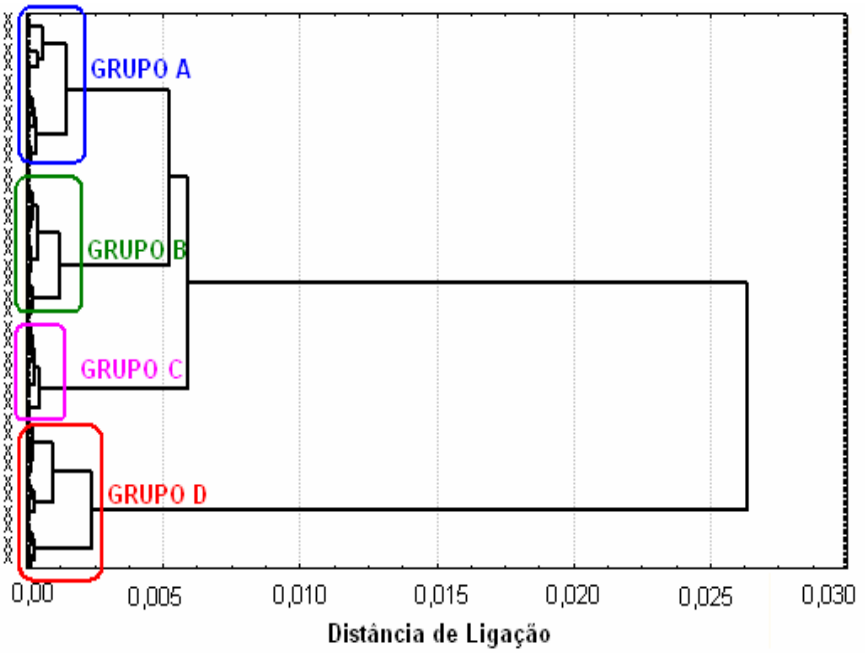


Figura 3.6. Dendrograma de uma HCA obtido pela distância de *Chebychev* e regra de ligação baseado no método de Ward´s, com as amostras rotuladas como A, B, C e D para os diferentes tipos (marcas) de cigarros.

A formação dos agrupamentos na distância próxima a 0,0065 separa a classe de cigarros do tipo D das outras três classes. Se os teores de nicotina e alcatrão forem comparados, de acordo com os informados pelos rótulos de seus respectivos fabricantes, observam-se dois aspectos interessantes:

1. Os cigarros do tipo D, inicialmente separados dos demais, são os que apresentam os menores teores de nicotina e alcatrão;
2. Na distribuição das classes apresentadas no dendrograma acima, percebe-se uma distribuição em ordem decrescente dos grupos formados, em relação aos teores de nicotina e alcatrão.

A **Tabela 3.1** contém as informações sobre os teores de nicotina e alcatrão indicados nos rótulos dos maços de cigarros analisados.

Tabela 3.1 Teores de nicotina de alcatrão das quatro marcas de cigarros.

Tipo de Cigarro	A	B	C	D
Teor de Alcatrão (mg)	13	10	8	6
Teor de Nicotina (mg)	1,10	0,8	0,7	0,6

Isso pode ser uma evidência de que a separação ocorre com base nos teores de nicotina e/ou alcatrão destes cigarros.

3.4.2. Aplicação da PCA

Inicialmente foi realizada uma PCA, em todo o conjunto das 210 amostras de cigarros, para avaliar possíveis sobreposições entre as classes. Essa análise mostrou que, apesar das amostras estarem muito próximas umas das outras, não houve sobreposição de classes, ou seja, as amostras encontraram-se bem distribuídas ao longo de PC1 e PC2. O gráfico dos escores apresentado na **Figura 3.7** mostra a formação de quatro grupos de amostras que correspondem às quatro marcas de cigarros.

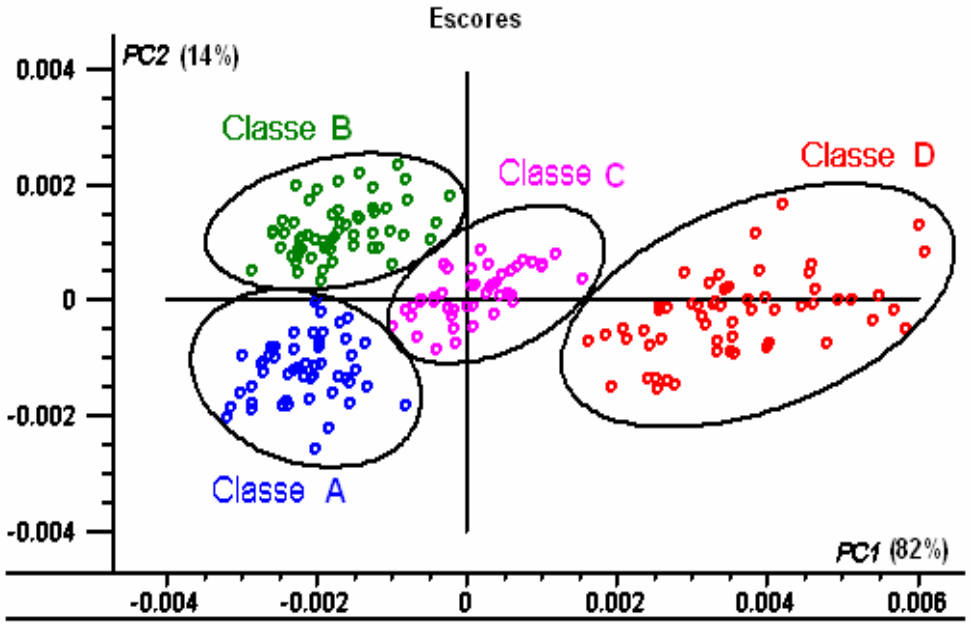


Figura 3.7. Gráfico dos escores de PC1 versus PC2 das 210 amostras de cigarros pertencentes as quatro classes de cigarros: A (azul), B(verde), C(rosa) e D(vermelho).

A classe de cigarros do tipo D está separada das classes A e B pelos valores de escores apresentados em PC1 (82% da variância explicada), ou seja, em escores com valores positivos, encontram-se as amostras da classe D, e em negativos, A e B. A classe do tipo C está no ponto intermediário desta separação, apresentando tanto escores negativos como positivos. Esta distribuição das amostras ao longo de PC1 está de acordo com os resultados apresentados pela HCA. Assim, a classe de

cigarros do tipo D, que apresenta os menores teores de nicotina e alcatrão, encontra-se separada das demais classes ao longo da PC1.

A PC2, com 14% da variância explicada, é responsável pela separação entre as classes A e B, sendo que a classe A está em escores negativos e as amostras da classe B em escores positivos. Todas as informações relevantes necessárias para a formação e/ou discriminação dos agrupamentos foram descritas por apenas duas PC's, totalizando 96% da variância explicada dos dados. Esta informação é confirmada pelo gráfico da variância explicada *versus* o número de PC's **Figura 3.8**.

O gráfico da variância explicada *versus* o número de componentes principais (PC's) (**Figura 3.8**) mostra que a partir da segunda PC a variância dos dados permanece quase que constante, indicando que duas PC's já são suficientes para distinguirem os quatro tipos de cigarros.

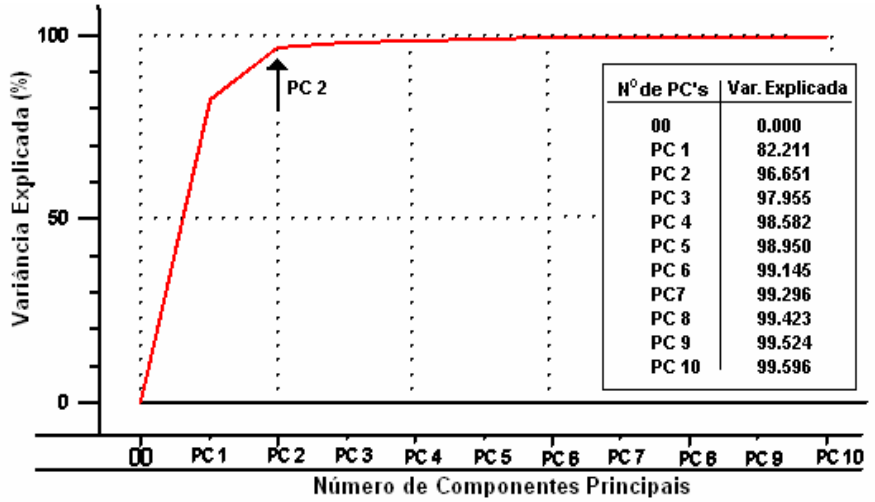


Figura 3.8. Gráfico da variância explicada *versus* Nº de PC's para todas as classes de cigarros.

Ao analisar o gráfico dos pesos (**Figura 3.9**)**,** observa-se que o maior valor de peso para a PC1 corresponde à região próxima de 5268 cm^{-1}. Como já visto na **Seção 3.2**, esta região está associada ao segundo sobreton de transições vibracionais de ligações C=O de grupos carboxílicos e ésteres. Além destas, bandas de combinação de transições fundamentais relativas à ligação OH da água e álcoois também estão presentes. Já em PC2, o maior valor dos pesos está em torno de 4903 cm^{-1}, que corresponde à região de combinações de transições fundamentais relativas à ligação NH de compostos nitrogenados.

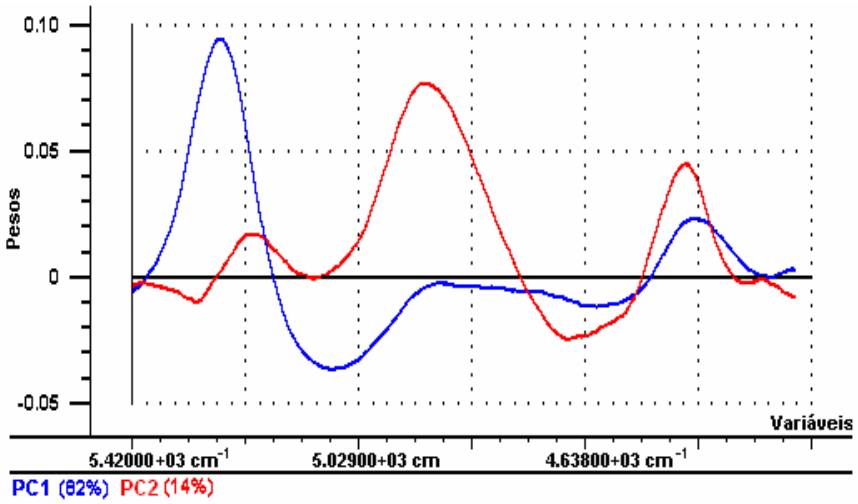


Figura 3.9. Gráfico dos pesos *versus* variáveis de PC1 (azul) e PC2 (vermelho) de todas as classes.

Os resultados da técnica PCA corroboram aqueles alcançados na aplicação da HCA. Isto pode ser evidenciado comparando os dendrogramas das **Figuras 3.5** e **3.6** com o gráfico dos escores, mostrado na **Figura 3.7**.

3.5. Modelagem e Classificação SIMCA

Por se tratar de uma técnica de reconhecimentos de padrões supervisionados, no SIMCA é necessária a construção de modelos individuais (para cada classe de amostras) que possam ser usados futuramente para a previsão de amostras desconhecidas. Desta forma, modelos de classificação para cada tipo de cigarro foram construídos com as amostras do conjunto de treinamento e validados com as amostras do conjunto teste, previamente selecionadas pelo algoritmo KS (como discutido na **Seção 2.2**).

3.5.1 Construção dos modelos SIMCA das classes A, B, C e D

Duas ferramentas de diagnósticos foram utilizadas para a validação destes modelos: Gráfico da variância explicada *versus* o número de PC's e o gráfico dos escores. Como a determinação do número de variáveis latentes é um fator preponderante na avaliação da capacidade preditiva dos modelos, conhecer um número ótimo de PC's para uma determinada classe modelada torna-se indispensável. O gráfico da variância explicada *versus* o número de componentes principais pode ser

útil na escolha adequada destas PC's. Os gráficos para as 4 classes de cigarros são mostrados nas **Figuras 3.10** a **3.13**.

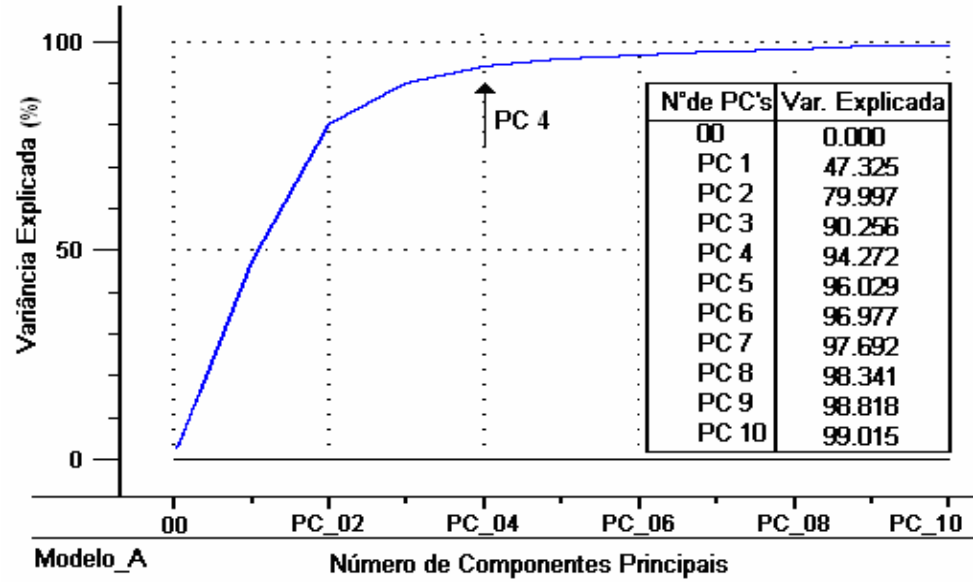


Figura 3.10. Gráfico da Variância explicada *versus* o Nº de PC's para a classe de cigarros do tipo A.

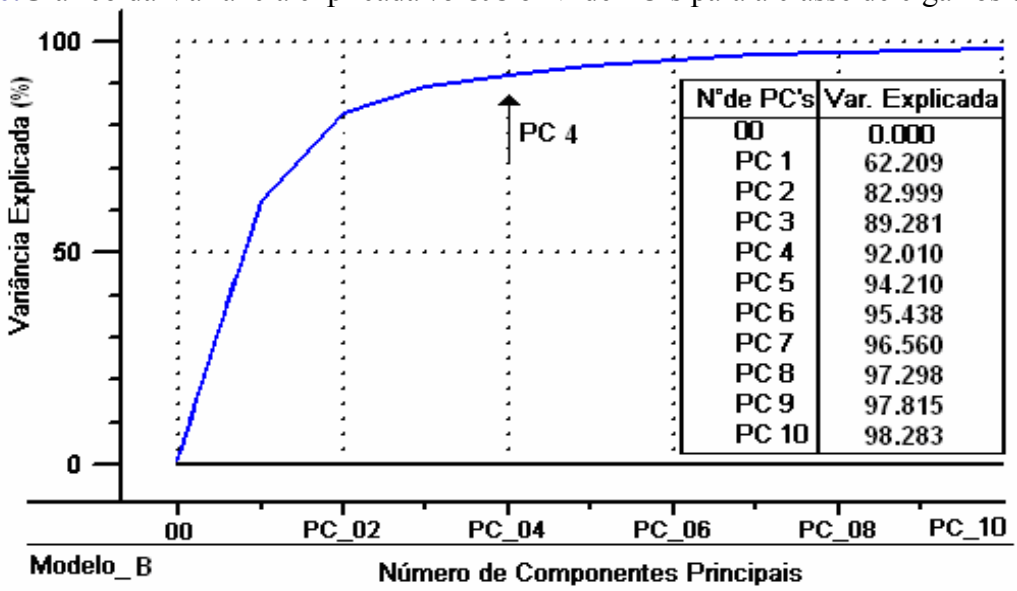


Figura 3.11. Gráfico da Variância explicada *versus* o Nº de PC's para a classe de cigarros do tipo B.

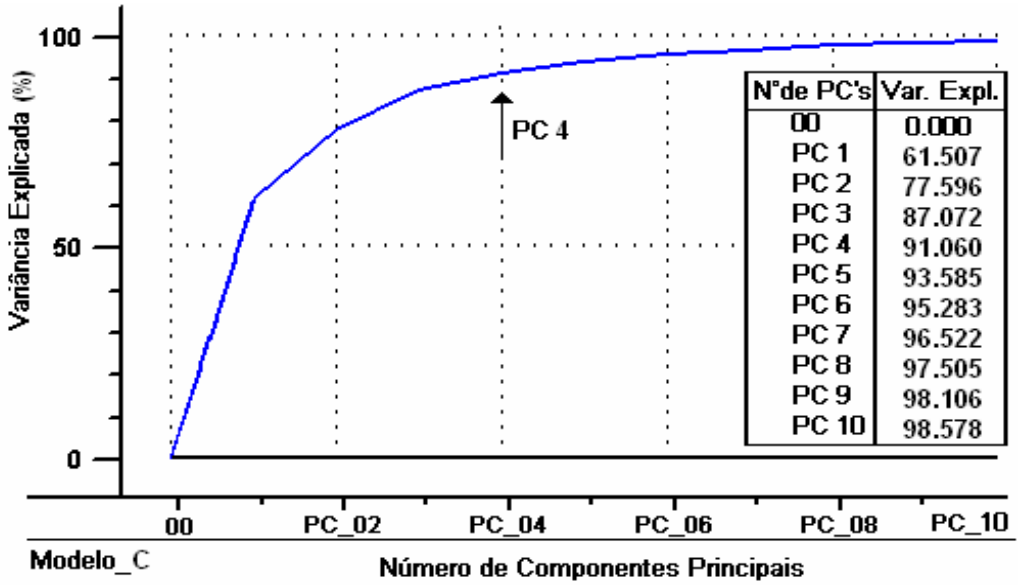


Figura 3.12. Gráfico da Variância explicada *versus* o Nº de PC's para a classe de cigarros do tipo C.

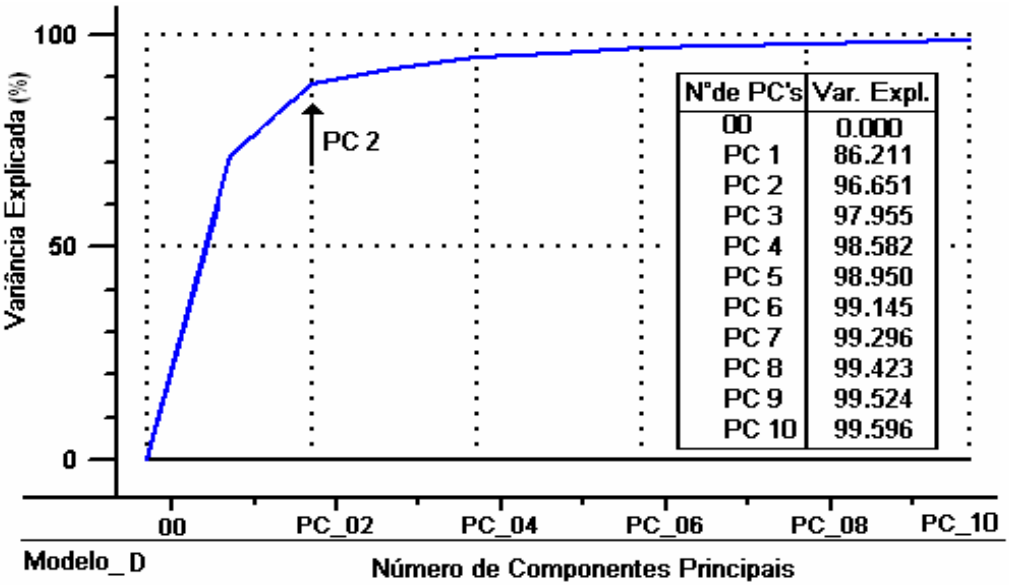


Figura 3.13. Gráfico da Variância explicada *versus* o Nº de PC's para a classe de cigarros do tipo D.

Os gráficos apresentados nas **Figuras 3.10** a **3.12** mostram que para as classes de cigarros A, B e C, o número ideal de PC's é quatro, totalizando 94,3%, 92,1% e 91,06% da variância explicada dos dados, respectivamente. Observa-se nestes gráficos que, depois da PC4, a variância se manteve quase que constante. Para a classe D, o número ótimo de PC's é dois (**Figura 3.13**), correspondendo a 96,7% da variância explicada dos dados. Vale salientar, que estes também foram os números de PC's de cada classe de cigarro, escolhidos pelo *Unscrambler*, segundo o critério adotado e estabelecido como *Default* do programa, o qual não é acessível aos seus usuários.

Por fornecer uma visão privilegiada das amostras ao longo das PC's, a análise dos gráficos dos escores é também considerada uma ferramenta de diagnóstico de grande importância na validação dos modelos construídos. Com ela, torna-se possível examinar com maior detalhe as semelhanças, diferenças, localização e anomalias das diferentes amostras usadas no processo de modelagem.

Os gráficos dos escores das 4 classes de cigarros são apresentados nas **Figuras 3.14** a **3.17**.

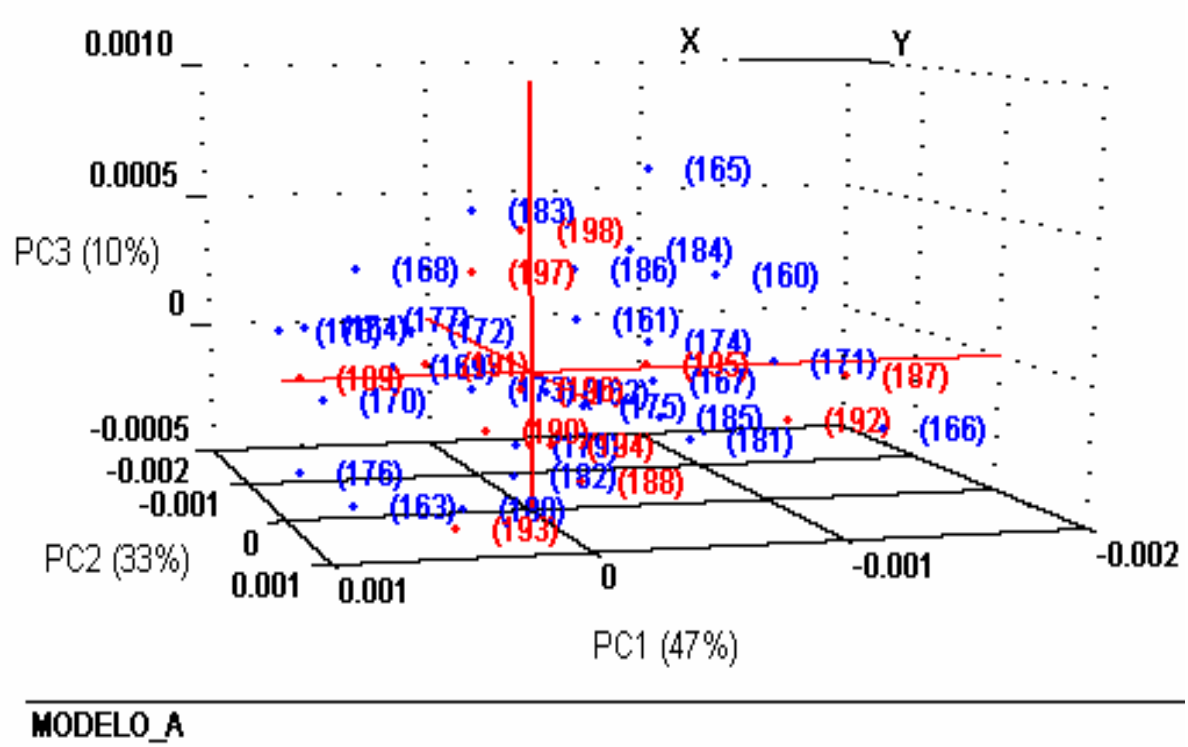


Figura 3.14. Gráfico dos escores de PC1 *versus* PC2 *versus* PC3 para a classe de cigarros do tipo A. As amostras dos conjuntos de treinamento estão pintadas em azul e dos de teste em vermelho.

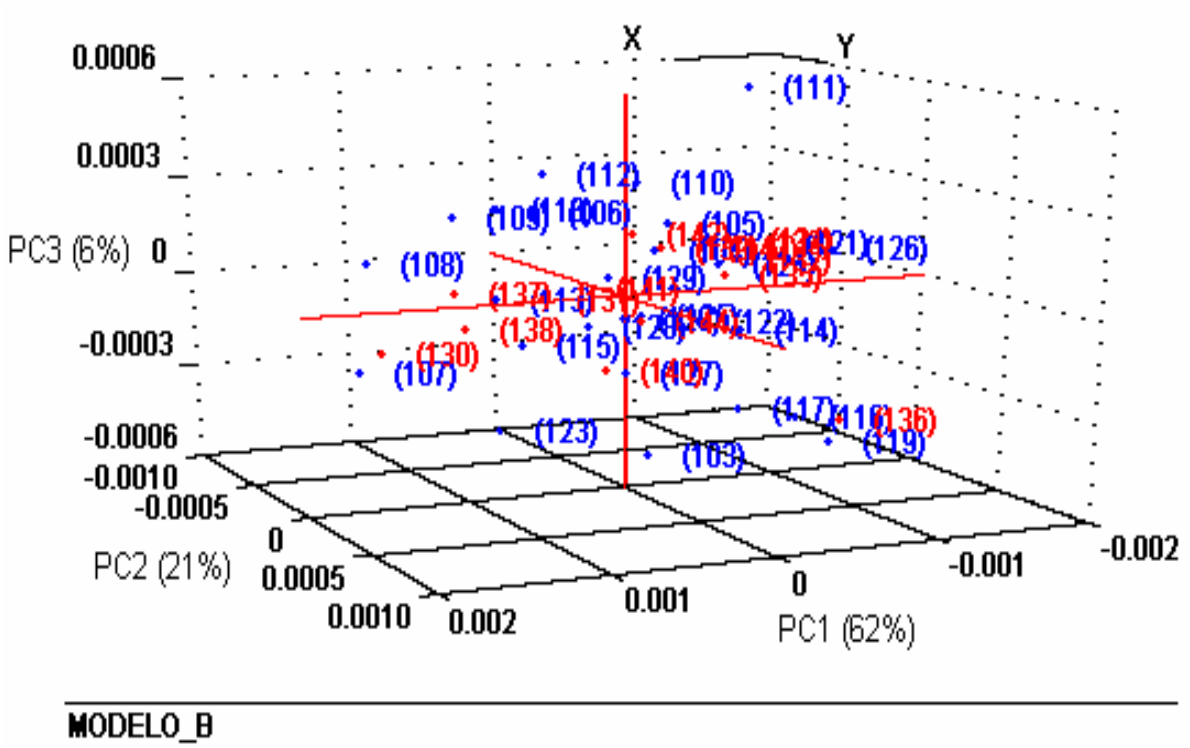


Figura 3.15. Gráfico dos escores de PC1 *versus* PC2 *versus* PC3 para a classe de cigarros do tipo B. As amostras dos conjuntos de treinamento estão pintadas em azul e dos de teste em vermelho.

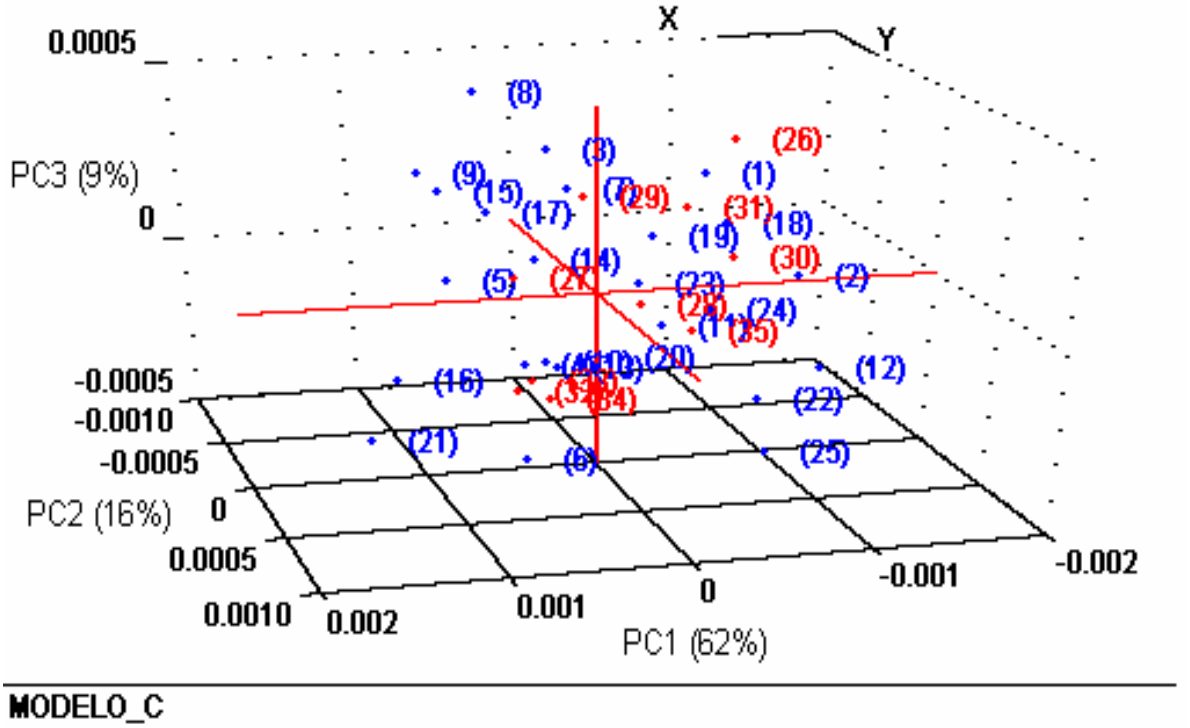


Figura 3.16. Gráfico dos escores de PC1 *versus* PC2 *versus* PC3 para a classe de cigarros do tipo C. As amostras dos conjuntos de treinamento estão pintadas em azul e dos de teste em vermelho.

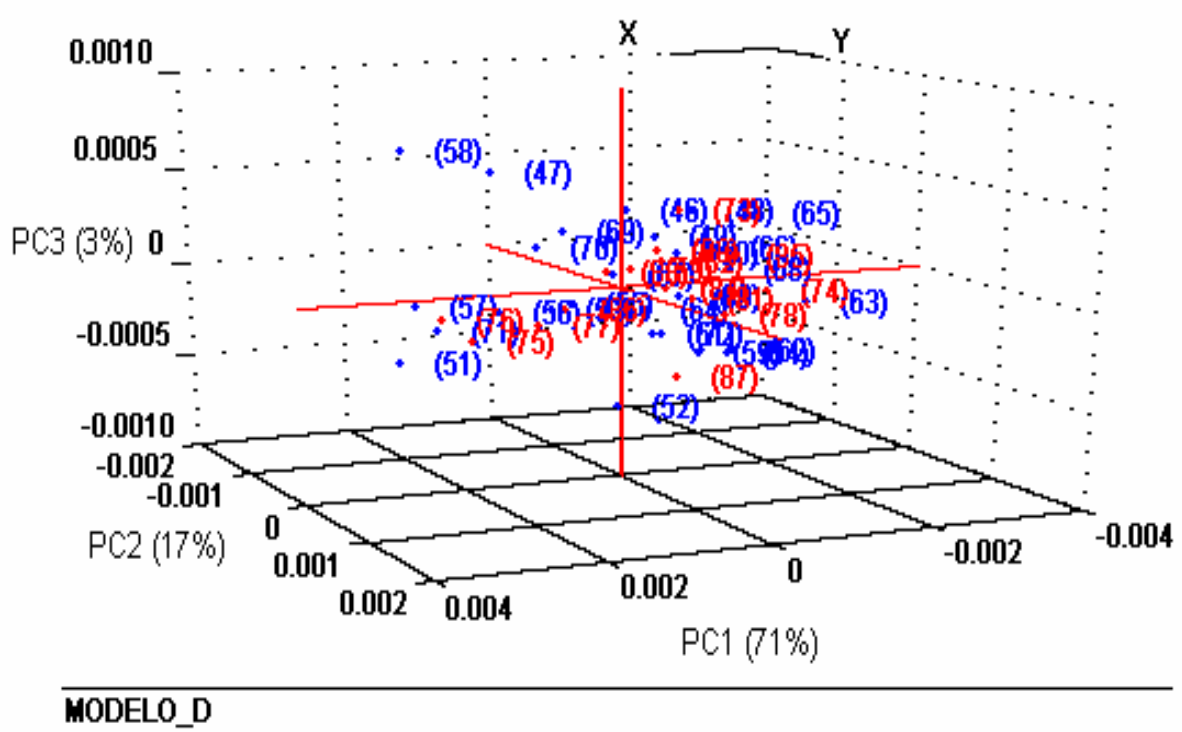


Figura 3.17. Gráfico dos escores de PC1 *versus* PC2 *versus* PC3 para a classe de cigarros do tipo D. As amostras dos conjuntos de treinamento estão pintadas em azul e dos de teste em vermelho.

Os quatros gráficos de escores apresentados nas **Figuras 3.14** a **3.17** mostram que todas as amostras do conjunto de treinamento, escolhidas separadamente para cada classe pelo algoritmo KS, encontram-se bem distribuídas ao longo das 3 PC's, cobrindo as fronteiras das amostras do conjunto teste. Isto mostra a eficiência do KS no tocante à escolha de tais conjuntos.

3.5.2 Validação dos modelos SIMCA das classes A, B, C e D

Os modelos SIMCA das classes A, B, C, e D construídos são agora aplicados na previsão das amostras do conjunto de treinamento e também as de teste, com intuito de se fazer uma avaliação da sua capacidade de previsão, ou seja, fazer uma validação dos modelos construídos. Para isso foram construídos gráficos S_i x H_i (**Seção 1.9.2.3**), empregando um nível de confiança de 95%. A avaliação do modelo SIMCA de cada classe são apresentadas nas seções a seguir.

Classificação de todos os Cigarros usando o modelo SIMCA da classe A

No gráfico S_i x H_i da classe A (**Figura 3.18**) observa-se que nenhuma amostra de cigarro tipo A foi classificado como não pertencente a sua classe. Entretanto, uma amostra do tipo C foi classificada como sendo pertencente à classe A.

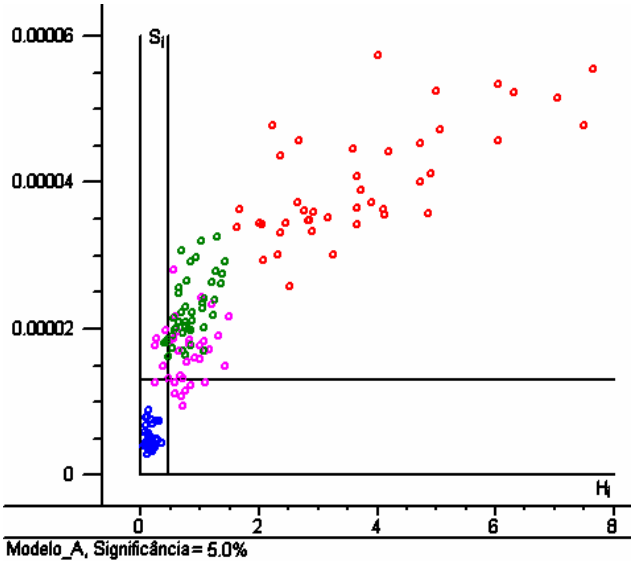


Figura 3.18. Gráfico de S_i x H_i (nível de significância de 5%) do modelo da classe A. Os símbolos nas cores azul, verde, rosa e vermelha indicam as classes A, B, C e D respectivamente.

Classificação de todos os Cigarros usando o modelo SIMCA da classe B

Verifica-se no gráfico S_i x H_i da classe B (**Figura 3.19**) que duas amostras de cigarro do tipo A e três do tipo C foram classificadas como pertencentes à classe B**.** Contudo, todas as amostras da classe B foram classificadas corretamente.

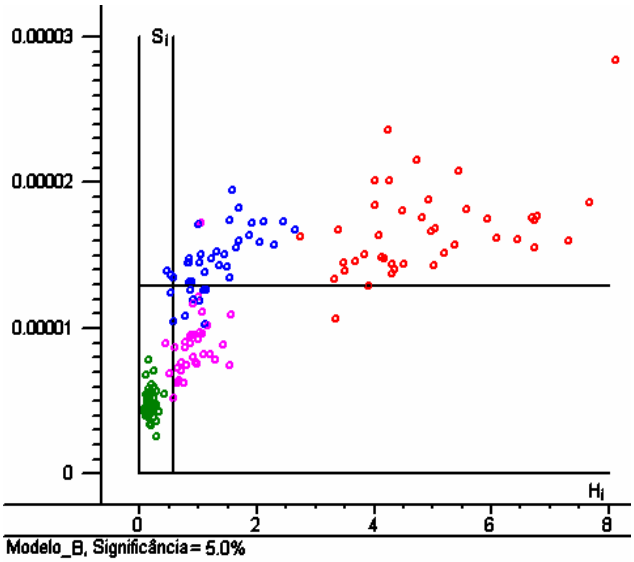


Figura 3.19. Gráfico de S_i x H_i (nível de significância de 5%) do modelo da classe B. Os símbolos nas cores azul, verde, rosa e vermelha indicam as classes A, B, C e D respectivamente.

Classificação de todos os Cigarros usando o modelo SIMCA da classe C

O gráfico S_i x H_i da classe C (**Figura 3.20**) mostra que nenhuma amostra de cigarro tipo C foi classificada incorretamente. Entretanto, quatro amostras da classe A e uma da classe B foram classificadas como sendo pertencentes à classe C.

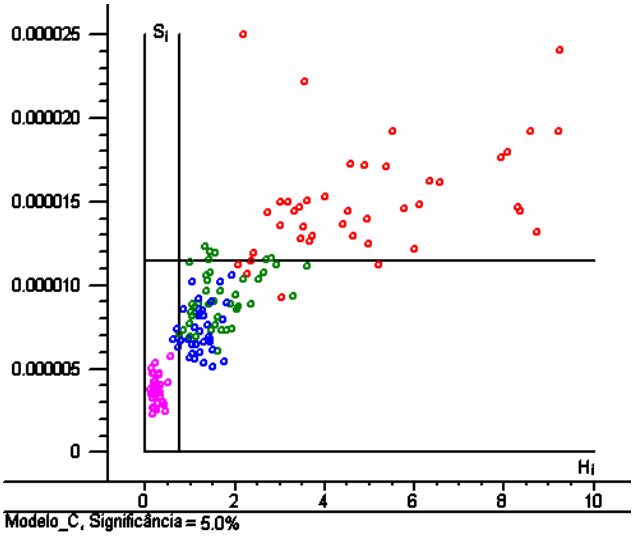


Figura 3.20. Gráfico de S_i x H_i (nível de significância de 5%) do modelo da classe C. Os símbolos nas cores azul, verde, rosa e vermelha indicam as classes A, B, C e D respectivamente.

Classificação de todos os Cigarros usando o modelo SIMCA da classe D

Para a classe D, o gráfico S_i x H_i (**Figura 3.21**) indica que cinco amostras de cigarros do tipo C foram classificadas como sendo do tipo D. Além disso, diferentemente dos 3 modelos SIMCA das classes A, B e C, o modelo SIMCA da classe D apresentou uma amostra classificada incorretamente.

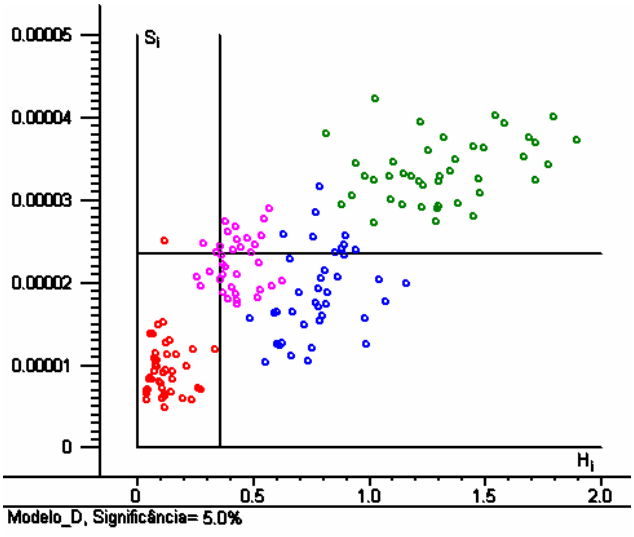


Figura 3.21. Gráfico de S_i x H_i (nível de significância de 5%) do modelo da classe D. Os símbolos nas cores azul, verde, rosa e vermelha indicam as classes A, B, C e D respectivamente.

Devido aos erros de classificação apresentados nesta avaliação, pode-se afirmar que os modelos SIMCA construídos para as marcas de cigarros A, B, C e D não estão eficientemente validados, embora várias outras tentativas de melhorá-los tenham sido testadas e não foi obtido sucesso. Isto pode ter ocorrido devido a:

1. Limitações da própria modelagem SIMCA;
2. Espectrometria NIRR não ser discriminatória o suficiente para fornecer classificações com 100% de acerto;
3. Necessidade de um maior número de amostras representativas no conjunto de treinamento.

Apesar do exposto acima, os modelos SIMCA construídos foram também usados na classificação de amostras de um conjunto externo, denominado de conjunto de previsão.

3.5.3. Uso dos Modelos SIMCA das classes A, B, C e D no conjunto de Previsão

Os modelos SIMCA das classes A, B, C e D foram usados na classificação das amostras do conjunto de previsão. Os resultados são apresentados na **Tabela 3.2**. Vale salientar que os erros apresentados numa classificação podem ser de dois tipos: I, quando a amostra avaliada não é classificada na sua classe verdadeira e II, quando a amostra é classificada em sua classe e também em uma classe errada. Pode também acontecer simultaneamente os dois tipos de erros, ou seja, quando uma amostra não é classificada na sua classe verdadeira (erro tipo I), mas é classificada em uma classe errada (erro tipo II).

Verifica-se na **Tabela 3.2** que três amostras da classe C (C-36, C-38 e C-41) apresentaram erro do tipo II e que todas as demais amostras foram classificadas corretamente, a um nível de confiança de 95%. Apesar dos bons resultados dos modelos SIMCA, buscou-se uma alternativa, de modo a se obter modelos sem erros do tipo I e II, ou seja, com 100% de acerto. Para isso, uma nova metodologia foi

desenvolvida usando as medidas dos espectros NIRR dos cigarros, o algoritmo das projeções sucessivas e a análise de discriminante linear (NIRR-SPA-LDA).

Tabela 3.2 Classificação SIMCA das amostras de previsão das quatro classes de cigarros. A indicação com asterisco (*) indica que a amostra foi incluída na determinada classe.

Amostra Testada	Classe A	Classe B	Classe C	Classe D
A-198	*			
A-199	*			
A-200	*			
A-201	*			
A-202	*			
A-203	*			
A-204	*			
A-205	*			
A-206	*			
A-207	*			
A-208	*			
A-210	*			
B-145		*		
B-146		*		
B-147		*		
B-148		*		
B-149		*		
B-150		*		
B-151		*		
B-152		*		
B-153		*		
B-154		*		
B-155		*		
B-156		*		
B-157		*		
B-158		*		
B-159		*		
C-36		*	*	
C-37			*	
C-38			*	*
C-39			*	
C-40			*	
C-41			*	*
C-42			*	
C-43			*	
C-44			*	
C-45			*	
D-88				*
D-89				*
D-90				*
D-91				*
D-92				*
D-93				*
D-94				*
D-95				*
D-96				*
D-97				*
D-98				*
D-99				*
D-100				*

D-101				*
D-102				*

3.6. Modelagem e Classificação NIRR-SPA-LDA

3.6.1 Construção do Modelo NIRR-SPA-LDA

Os espectros NIRR das amostras de cigarros do conjunto de treinamento e teste foram, inicialmente, usados pela modelagem SPA-LDA de modo a selecionar as variáveis espectrais na mesma faixa espectral de trabalho utilizadas na construção dos modelos HCA, PCA e SIMCA.

O gráfico apresentado na **Figura 3.22** mostra o número de variáveis selecionadas pelo SPA *versus* os valores obtidos com a função de custo, que fornece o risco médio G de uma classificação incorreta pela LDA, conforme discutido em detalhes na **Seção 1.9.2.4** do **Capítulo 1**.

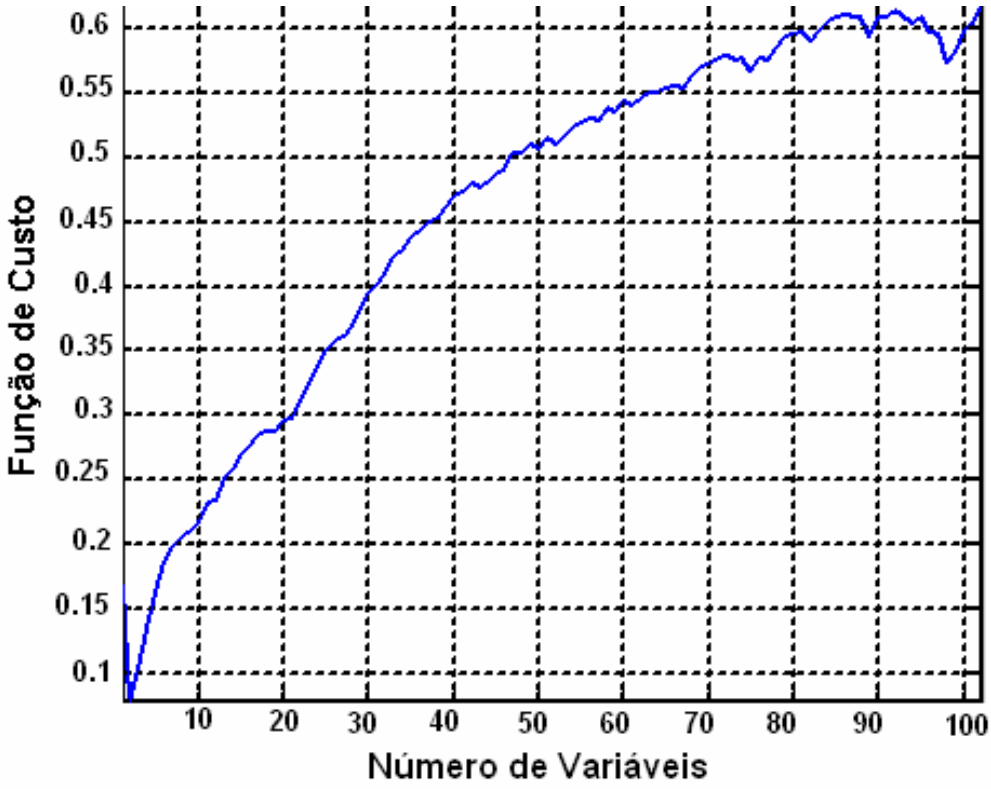


Figura 3.22. Número de variáveis selecionadas pelo SPA *versus* os valores da função de custo.

A **Figura 3.22** indica que o modelo NIRR-SPA-LDA que apresenta menor valor de risco médio G é aquele que usa apenas duas variáveis. Essas variáveis correspondem aos números de onda 4903 cm^{-1} e 5056 cm^{-1}.

A **Figura 3.23** mostra o espectro NIRR derivativo, de uma amostra de cigarro, com as variáveis selecionadas pelo SPA destacadas (4903 e 5056 cm^{-1}). Elas estão em regiões em que ocorrem tanto o segundo sobreton de C=O de grupos carboxílicos e ésteres, bandas de combinação relativas à ligação OH de água e álcoois (por volta de 5056 cm^{-1}) como também, bandas de combinação de transições fundamentais relativas à ligação NH de compostos nitrogenados, tais como aminas e amidas (na região em torno de 4903 cm^{-1}).[28,63,64]

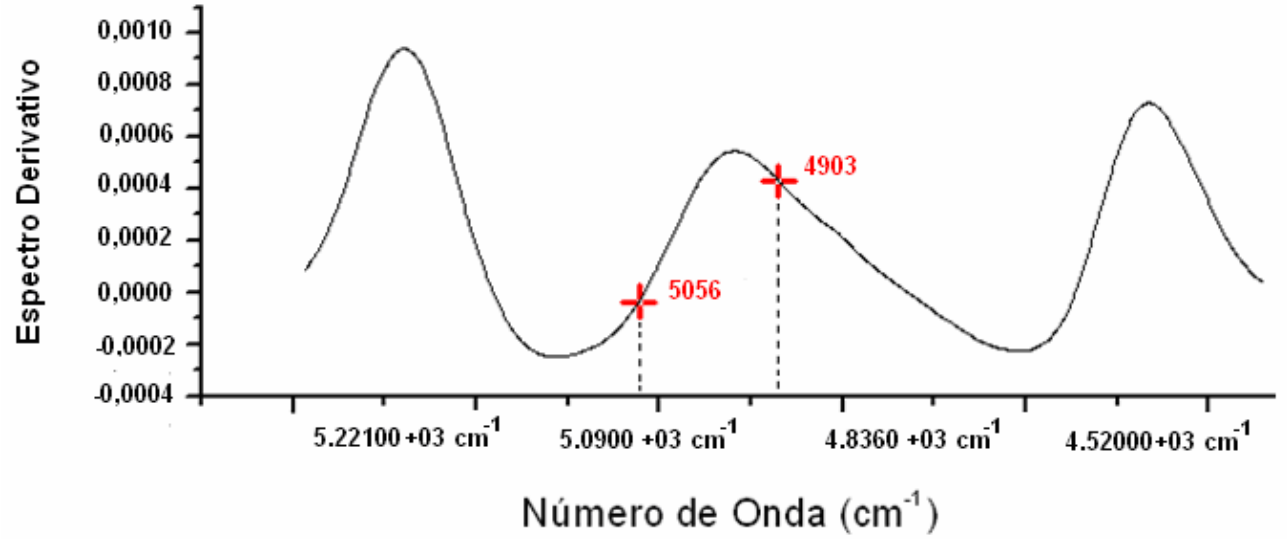


Figura 3.23 Espectro NIRR derivativo de uma amostra de cigarro na região espectral de trabalho, com as variáveis selecionadas pelo NIR-SPA-LDA em destaque.

3.6.2 O Uso do Modelo NIRR-SPA-LDA na Previsão das Amostras de Cigarros

O modelo NIRR-SPA-LDA construído com as duas variáveis (5056 e 4903 cm^{-1}) foi então aplicado na classificação das amostras do conjunto de previsão. Nenhum erro foi obtido (100% de acerto) na classificação de todas as marcas de cigarros deste conjunto.

3.6.3 O Uso do Modelo NIRR-SPA-LDA na Previsão de Todas as Amostras

A **Figura 3.24** mostra o gráfico de dispersão das amostras obtido usando o modelo NIRR-SPA-LDA construído com as duas variáveis (5056 e 4903 cm^{-1}), quando este foi aplicado as 210 amostras de cigarros analisadas. Neste gráfico observa-se a formação bem definida dos quatro grupos de amostras, correspondentes às quatro marcas de cigarros analisadas. As classes esperadas para esse conjunto analisado estão bem distribuídas e não há qualquer sobreposição entre elas.

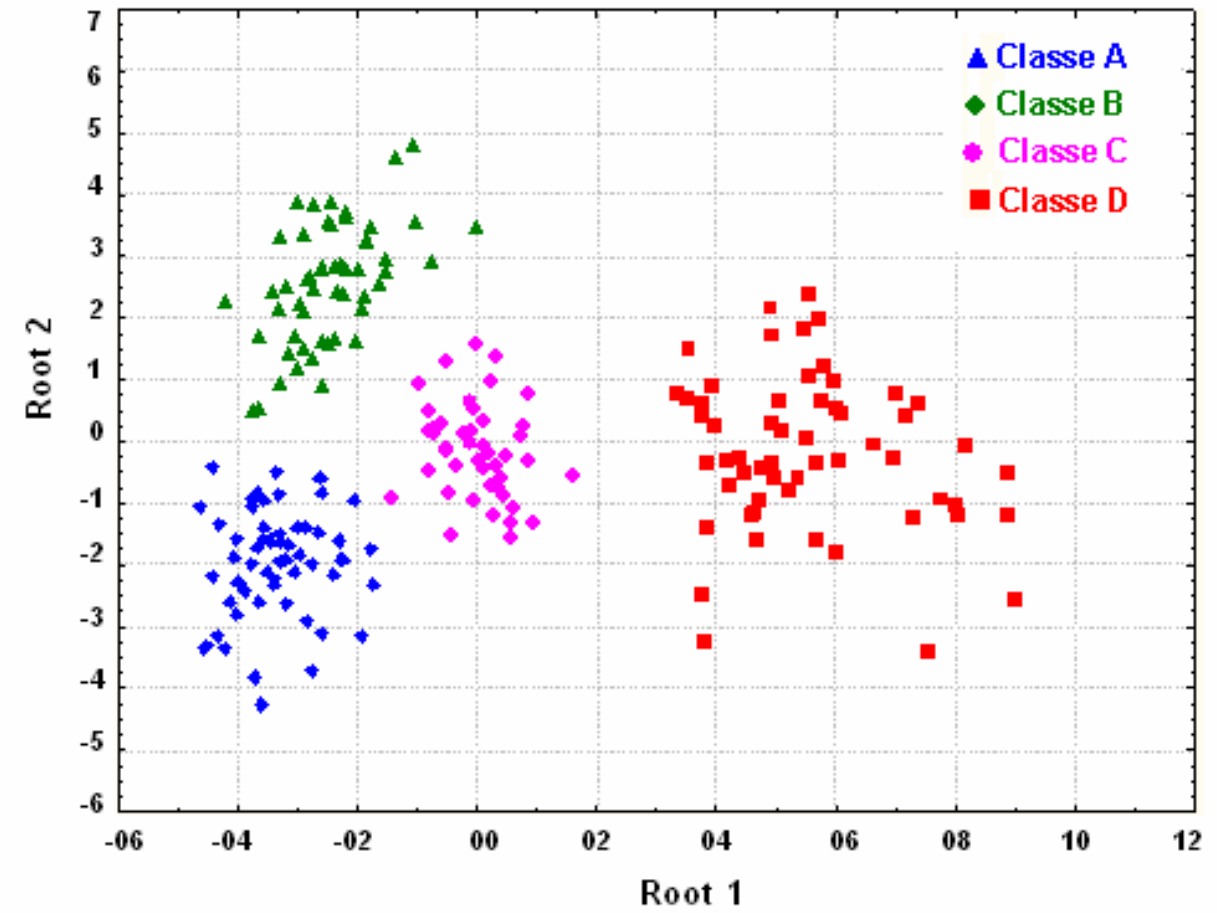


Figura 3.24. Gráfico de dispersão das 210 amostras cigarros A, B, C e D obtidas com o modelo NIRR-SPA-LDA. Root 1 e 2 são as funções discriminantes 1 e 2, respectivamente.

3.7. Comparação entre os modelos NIRR-SPA-LDA e SIMCA

Modelos SIMCA, em 4 níveis de significância do teste *F* (1, 5, 10 e 25%), foram construídos com o intuito de comparar os modelos resultantes com aquele obtido pelo NIRR-SPA-LDA. Estes modelos foram aplicados na classificação das amostras do conjunto de previsão e os resultados são apresentados na **Tabela 3.3**.

Tabela 3.3 Erros de classificação dos modelos NIRR-SPA-LDA e SIMCA (em 4 níveis de significância do teste *F*) para as amostras de cigarros do conjunto de previsão.

	NIRR-SPA-LDA	SIMCA (1%)	SIMCA (5%)	SIMCA (10%)	SIMCA (25%)
Erro Tipo I	**0**	**0**	**0**	**0**	**1**
Erro Tipo II	**0**	**7**	**3**	**2**	**1**
Total de Erros	**0**	**7**	**3**	**2**	**2**

Como pode ser observado na **Tabela 3.3**, os erros do tipo II dos modelos SIMCA diminuem e do tipo I aumentam à medida que se eleva o nível de significância do teste *F*. Isso acontece porque as fronteiras dos modelos SIMCA aumentam e assim há uma maior probabilidade de outras amostras entrarem na embalagem de uma classe diferente da sua. Por outro lado, o NIRR-SPA-LDA não apresentou nenhum tipo de erro, então o número total de erros apresentados pelo SIMCA é maior do que os erros do SPA-LDA. Diante destes resultados, comprova-se que o método NIRR-SPA-LDA proposto é uma excelente alternativa para classificar cigarros com intuito de ser usado na fiscalização da qualidade de cigarros.

CAPÍTULO 4

CONCLUSÕES

4. Conclusões

O presente trabalho propôs o desenvolvimento de uma nova metodologia que usa a espectrometria *NIRR*, o algoritmo das projeções sucessivas (*SPA*) e a análise discriminante linear (*LDA*) para classificação de diferentes tipos de cigarros com intuito de auxiliar na fiscalização da qualidade dos mesmos.

Inicialmente, foi realizada uma análise exploratória com os espectros NIRR das 210 amostras de 4 marcas de cigarros, empregando HCA e PCA. Tanto o dendrograma do HCA, construído usando a regra de ligação Chebychev e a técnica de conexão baseado no método de Ward´s, como o gráfico dos escores de PCA, revelou que os espetros NIRR são capazes de discriminar os 4 agrupamentos esperados de acordo com as 4 marcas de cigarros que compõem as 210 amostras de cigarros analisadas.

Com base no estudo acima, partiu-se, então, para a construção de modelos SIMCA que permitissem classificar, baseados em medidas de espectros NIRR, as amostras das 4 marcas de cigarros analisadas. Apesar da validação dos 4 métodos SIMCA construídos, estes apresentaram erros de classificação do tipo I (I, quando a amostra avaliada não é classificada na sua classe verdadeira) e do tipo II (quando a mesma é classificada em sua classe e também em uma classe errada), com um nível de significância de 5%.

Buscou-se, portanto, desenvolver uma nova metodologia, usando as medidas dos espectros NIRR dos cigarros, o algoritmo das projeções sucessivas e a análise discriminante linear (NIRR-SPA-LDA), de modo a se obter modelos sem erros de classificação do tipo I e II, ou seja, com 100% de acerto. O método NIRR-SPA-LDA proposto conseguiu classificar corretamente (100% de acerto) todos os cigarros, nas quatro marcas usadas no estudo, usando apenas duas variáveis espectrais selecionadas (5056 e 4903 cm^{-1}), dentre as 1168 da faixa espectral NIRR (de 5420 a 4252 cm^{-1}) das medidas. Deduz-se a partir de tal fato que instrumentos de baixo custo, tais como os fotômetros NIRR, podem ser construídos usando diodos

emissores de luz (*LEDs*) nos comprimentos de onda selecionados (1978 e 2040 nm) e empregados nas medidas para a metodologia NIRR-SPA-LDA proposta.

Diante dos resultados apresentados, conclui-se finalmente que a metodologia NIRR-SPA-LDA desenvolvida é uma boa alternativa para a classificação de cigarros, que poderá ser utilizada na fiscalização da qualidade dos mesmos, de forma simples, rápida e eficiente, além de utilizar uma técnica não-invasiva e que não produz resíduos químicos e, conseqüentemente, não contamina o meio ambiente.

4.1. Propostas Para Continuidade do Trabalho

Pretende-se, como continuidade deste trabalho:

1. *Fazer estudos de aplicação de Tranformadas Wavelet para a filtragem de sinais, ou seja, maior eficiência na remoção dos ruídos apresentados nos espectros.*

2. *Estudar a possibilidade de aumentar o poder de classificação do modelo NIRR-SPA-LDA, ampliando suas fronteiras com amostras de outras marcas de cigarros e amostras oriundas de apreensões feitas por órgãos fiscalizadores de contrabando e adulterações de cigarros, comercializados no Brasil.*

3. *Tendo em vista que o modelo NIRR-SPA-LDA utilizado na classificação de cigarros alcançou resultados bastante satisfatórios a partir da seleção de dois comprimentos de onda (1978 e 2040 nm), propõe-se a construção de um fotômetro-NIRR que, baseado na metodologia NIRR-SPA-LDA desenvolvida, identifique amostras adulteradas e não adulteradas, simplificando e minimizando o custo das análises.*

REFERÊNCIAS

1. http://pt.wikipedia.org/wiki/tabaco (acessado no dia 30/11/2006).
2. ROSEMBERG, J.; ARRUDA, A.M.; MORAES, M.A. Nicotina Droga Universal, São Paulo, 2005.
3. http://www.sousacruz.com.br/oneweb/sites/sou_5RRP92.nsf/vwPagesweblive (acessado no dia 01/12/2006).
4. http://www.anvisa.gov.br/tabaco/lista_marcas.pdf (acessado em 10/10/2006)
5. http://www.receita.fazenda.gov.br/DestinacaoMercadorias/ProgramaNacCombCigarroIlegal/MarcasProdFabricantes.htm (acessado em 29/08/2006).
6. http://www.ocarreteiro.com.br/content.php (acessado no dia 01/09/2006).
7. MADURO, R. M., Equilíbrio líquido-líquido em sistemas nicotina+água +extratantes, *Dissertação de Mestrado,* UNICAMP, Campinas, 2005.
8. http://www.qmc.ufsc.br/qmcweb/artigos/nicotina.html (acessado em 01/09/2006).
9. http://www.naofumantes.com.br/cigarrocompo.htm (acessado em 01/09/2006).
10. PANKOW, J.F.; HENNINGFIELD, J.E.; GARRETT B.E. Ammonia and other chemical base tobacco additives and cigarette nicotine delivery: Issues and research needs, *Nicotine & Tobacco Research,* v.6 (2), p.199-205, 2004.
11. http://e-legis.bvs.br/leisref/public (acessado no dia 06/12/2006).
12. http://www.idec.org.br/noticia (acessado no dia 01/09/2006).
13. http://www.terra.com.br/istoe/1765/economia/1765_mafia_dos_cigarros_02 (acessado em 05/09/2006).
14. HUANG L.F.; ZHONG K.J.; SUN X.J.; WU M.H.; HUANG K.L.; LIANG Y.Z.; GUO F.Q.; LI Y.W. Comparative analysis of the volatile components in cut tobacco from different locations with gas chromatography-mass spectrometry (GC-MS) and combined chemometric methods, *Analytica Chimica Acta*, v. 575(2), p. 236-245, 2006.
15. ADAM T.; FERGE T.; MITSCHKE S.; STREIBEL T.; BAKER R.R.; ZIMMERMANN R. Discrimination of three tobacco types (Burley, Virginia and

Oriental) by pyrolysis single-photon ionisation-time-of-flight mass spectrometry and advanced statistical methods, *analytical and bioanalytical chemistry,* v. 381 (2), p. 487-499, 2005.

16. GARRIGUES, J.M.; PEREZ-PONCE, A.; GARRIGUES, S.; DE LA GUARDIA, M. Flow injection Fourier transform infrared determination of nicotine in tobacco, *Analyst,* v. 124 (5), p. 783-786, 1999.
17. WANG J.J.; WANG F., MA L. The quality assessment of cigarette paper by SIMCA and PLS combined with near infrared spectrum, *spectroscopy and spectral analysis,* v. 26 (10), p. 1858-1862, 2006.
18. HATZINIKOLAOU, D.G.; LAGESSON, V.; STAVRIDOU, A.J.; POULI, A.E.; LAGESSON, A. L.; STAVRIDES, J.C. Analysis of the gas phase of cigarette smoke by gas chromatography coupled with UV-diode array detection, *analytical chemistry*, v. 78 (13), p. 4509-4516, 2006.
19. FATEMI, S.J.; SHEIKH A.; NOROOZIAN E.; AFZALI D. GF-AAS determination of cadmium in different cigarettes and cigarette smoke, *Asian Journal Of Chemistry,* v. 18 (2), p. 1285-1289, 2006.
20. BORDEN, J.T.; MAN, A.; SCOTT, D.A.; LIU, K.Z. Tobacco-induced alterations to the Fourier-transform infrared spectrum of serum, *Journal Molecular Medicine*, v. 81, p. 788–794, 2003.
21. SKOOG, D.A.; LEARY, J.J. Principles of Instrumental Analysis, Fourth Edition, Saunders College, 1992.
22. PASQUINI, C. Fundamentals, practical aspects and analytical applications. *J. Brazilian Chemical Society,* 14 (2) (2003), 198-219.
23. FIDÊNCIO, P.H. Análises de solos por espectroscopia no infravermelho próximo e aplicação de métodos quimiométricos. *Tese de Doutorado,* UNICAMP, Campinas. 2001.
24. FRANKLIN E. BARTON, Theory and principles of near infrared spectroscopy, *Spectroscopy Europe*, v.14, p.12-18, 2002

25. BRAGA, J.W.B. Avaliação de figuras de mérito em calibração multivariada, aplicada na determinação de carbamazepina por espectroscopia no infravermelho próximo e médio. *Dissertação de mestrado,* UNICAMP, Campinas. 2004.
26. MCCLURE, W. F., Near-Infrared Spectroscopy – The Giant is Running Strong *Anal. Chem*, 66, 43A-53A, 1994.
27. BOKOBZA, L. "Near Infrared Spectroscopy", *J. Near Infrared Spectrosc*., v. 6, p. 3-17, 1998.
28. WORKMAN Jr., J.J. Interpretative Spectroscopy for Near Infrared. *Appl. Spectrosc*. Rev. 31 (3), p. 251-320, 1996.
29. WETZEL, D.L. Near-Infrared Reflectance analysis. Sleeper among spectroscopic techniques, *Anal. Chem*., v.55, p.1165A-1176A, 1983
30. MALLEY, D.F.; WILLIAMS, P.C.; STAINTON, M.P. Rapid measurement of suspended C, N, and P precambriam shield lakes using near-infrared reflectance spectroscopy. *Water Res.,* v. 30, p. 1325-1322, 1996.
31. NILSSON, M.B. et.al. Quantifying relationships between near-infrared reflectance spectra of lake sediments and water chemistry. *Environ Sci Technol.,* v. 30, p. 2586-2590, 1996.
32. CONCEIÇÃO, P. R. N., COMIM, K. O., PETER, C. O., Determinação do espectro de reflectância de misturas de caulim através da Função de Kubelka-Munk, *Rev. Esc. Minas*, v.54 (4), 2001.
33. HANA, M.; MCCLURE, W.F.; WHITAKER, T.B. *J. Infrared Spectrosc*., 3, 133 (1995).
34. W.W.M. Wendlant; H. G. Hecht, *Reflectance Spectroscopy*, Jonh Wiley e Sons, New York, 1996.
35. FERREIRA, M.M.C.; et. al. Quimiometria I: calibração multivariada, um tutorial. *Química Nova,* v. 22(5), p. 724-731, 1999.
36. BEEBE, K.R.; PELL, R.J; SEASHOLTZ, M.B.; Chemometrics A Practical Guide. *John Wiley & Sons,* New York, 1998.

37. PIMENTEL, M. F.; SALDANHA, T. C. B.; ARAÚJO, M. C. U. Effects of Experimental Design on Calibration Curve Precision in Routine Analysis, *Journal of Automatic Chemistry*, v. 20(1), p. 9-15, 1998.
38. HONORATO, F. A.; HONORATO, R. S.; PIMENTEL, M. F.; ARAUJO, M. C. U. Analytical curve or standard addition method: how to Elect and Design – A strategy applied to copper determination in sugarcane spirits using AAS. *Analyst*, v. 127, n. 11, p.1520-1525, 2002.
39. GALVÃO, R. K. H.; JOSÉ, G. E.; DANTAS FILHO, H. A.; ARAUJO, M. C. U.; SILVA, E. C.; PAIVA, H. M.; SALDANHA, T. C. B.. Optimal Wavelet Filter Construction Using X and Y Data. *Chemometrics And Intelligent Laboratory Systems*, v. 70, n. 1, p. 1-10, 2004.
40. GALVÃO, R. K. H.; ARAUJO, M. C. U.; SALDANHA, T. C. B.; VISANI, V.; PIMENTEL, M. F. Estudo Comparativo sobre Filtragem de Sinais Instrumentais Usando Transformada de Fourier e de Wavelet. *Química Nova*, São Paulo-Brasil, v. 24(6), p. 874-884, 2001.
41. PONTES, M. J. C.; GALVÃO, R. K. H.; ARAUJO, M. C. U.; MOREIRA, P. N. T.; PESSOA NETO, O. D.; JOSÉ, G. E.; SALDANHA, T. C. B. The Successive Projections Algorithm for Spectral Variable Selection in Classification Problems. *Chemometrics And Intelligent Laboratory Systems*, v78 p.11-18, 2005.
42. DANTAS FILHO, H. A.; GALVÃO, R. K. H.; ARAUJO, M. C. U.; SILVA, E. C.; SALDANHA, T. C. B.; JOSÉ, G. E.; PASQUINI, C.; RAIMUNDO JR., I. M.; ROHWEDDER, J. J. R. A Strategy for Selecting Calibration Samples for Multivariate Modelling. *Chemometrics And Intelligent Laboratory Systems* v. 72, p. 83-91, 2004.
43. ARAUJO, M. C. U.; SALDANHA, T. C. B.; GALVÃO, R. K. H.; YONEYAMA, T.; CHAME, H. C.; VISANI, V. The Successive Projections Algorithm for Variable Selection in Spectroscopy Multicomponent Analysis. *Chemometrics And Intelligent Laboratory Systems*. v.57, n. 2, p. 65-73, 2001.

44. GALVÃO, R. K. H.; PIMENTEL, M. F.; ARAÚJO, M. C. U.; YONEYAMA, T., VISANI, V. Aspects of the Successive Projections Algorithm for Variable Selection in Multivariate Calibration Applied to Plasma Emission. *Analytica Chimica Acta*, v. 443 (1), p. 107-115, 2001.
45. HONORATO, F. A.; GALVÃO, R. K. H.; PIMENTEL, M. F.; BARROS NETO, B.; ARAUJO, M. C. U.; CARVALHO, F. R.. Robust Modeling for Multivariate Calibration Transfer by the Successive Projections Algorithm. *Chemometrics And Intelligent Laboratory Systems*, IN PRESS, 2005.
46. KOWALSKI, B.R.; SEASHOLTZ, M.B.; *J. Chemom.*, v. 5, p. 129, 1991.
47. MASSART, D. L.; VANDEGINSTE, B. G. M.; BUYDENS, S. J.; LEWI, P. J.; SMEYERS-VERBEKE, J., Handbook of Chemometrics and Qualimetrics: Parte B, Elsevier, Amsterdam, 1997.
48. Stasoft Brasil, Manual do Usuário, STATÍSTICA, versão 6.0, São Caetano do Sul-Brasil, 2001.
49. FLATEN, G. R.; GRUNG, B.; KVALHEIM, O. M. A method for validation of reference sets in SIMCA modeling, *Chemometrics And Intelligent Laboratory Systems*, v.72, p.101-109, 2004.
50. CAMO S.A. Manual do Usuário. UNSCRAMBLER, versão 7.5. Noruega, 1998.
51. BRUNS, Roy E.; FAIGLE, J.F.G., Quimiometria, *Química Nova,* p84-99, 1985.
52. COSTA, N. A. A., A reciclagem do resíduo de construção e demolição: uma aplicação da análise multivariada, Universidade Federal de Santa Catarina, *Tese de Doutorado*, Florianópolis, 2003.
53. L.R Schimleck, A.J. Michell, P Vinden. Appita Journal 49 (5) (1996) 319-324.
54. D. A. Rusak, L. M. Brown, S. D. Martin. Journal of Chemical Education 80 (5) (2003) 541-543.
55. D. Bellido-Milla, M. M. Moreno-Perez, M. P. Hernández-Artiga. Spectrochimica Acta Part B.55 (2000) 855 864.
56. M. Palma, C. G. Barroso. 58 (2002) 265-271.

57. D. Cozzolino, H. E. Smyth, M. Gishen, J. Agric. Food Chem., 51 (2003) 7703-7708.
58. M. Gestal, M. P. Gomez-Carracedo, J. M. Andrade, J. Dorado, E. Fernandez, D. Prada, A. Pazos. Analytica Chimica Acta 524 (1-2) (2004) 225-234.
59. S. Spycher, M. Nendza. Qsar & Combinatorial Science 23 (9) (2004) 779-791.
60. W. Wu, D. L. Massart .Analytica Chimica Acta 349 (1-3) (1997) 253-261.
61. S. J. Steel, N. Louw, N. J. Le Roux, Journal of Statistical Computation And Simulation 65 (2) (2000) 157-172.
62. R.W. Kennard, L. A. Stone, Technometrics 11 (1969) 137-148
63. HIBBARD, R.R.; CLEAVES, A.P. Carbon-Hidrogen groups in hydrocarbons-caracterization by 1.10 to 1.25 micron infrared absorptions. *Anal. Chem.*, v.21 (4), p. 496-492, 1949.
64. EVANS, A.; HIBBARD, R.R.; POWEL, A.S. Determination of carbon-hidrogen groups in high molecular weight hydrocarbons by near infrared absorptions. *Anal. Chem.*, v.23 (11), p.1604-1610, 1951.