JOÃO FERNANDO VILLARRUBIA LOPES MUNHOZ

# Método Alternativo para Detecção de Não Conformidade em Gasolina Comercial Brasileira do Tipo C: Uso da Cromatografia Ultrarrápida Aliada a Ferramentas Quimiométricas

Dissertação apresentada ao Instituto de Química, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Mestre em Química

Orientador: Prof. Dr. José Eduardo de Oliveira

Co-orientador: Prof. Dr. Danilo Luiz Flumignan

Araraquara

2013

**JOÃO FERNANDO VILLARRUBIA LOPES MUNHOZ**

Dissertação apresentada ao Instituto de Química, Universidade Estadual Paulista, como parte dos requisitos para obtenção do título de Mestre em Química.

Araraquara, 09 de abril de 2013.

**BANCA EXAMINADORA**

Prof. Dr. José Eduardo de Oliveira (Orientador)
Instituto de Química – UNESP, Araraquara - SP

Prof. Dr. Aristeu Gomes Tininis
Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP, Matão-SP

Dr. Fabio da Silva Vinhado
Centro de Pesquisas e Análises Tecnológicas - CPT – ANP, Brasília - DF

DADOS CURRICULARES

Nome do Programa: QUÍMICA
Curso: MESTRADO
Nome: JOÃO FERNANDO VILLARRUBIA LOPES MUNHOZ Matrícula nº: 85666-1

| Disciplinas | Ano | Período Letivo | Créditos | Carga Horária | Freq | Conceito |
|---|---|---|---|---|---|---|
| Química Orgânica Avançada | 2011 | 1º Semestre | 12 | 180 | 87 | C |
| Quimiometria | 2011 | 1 º Semestre | 12 | 180 | 100 | B |
| Teoria e Métodos de Separação, Isolamento e Purificação de Compostos Orgânicos | 2011 | 2 º Semestre | 12 | 180 | 87 | B |
| Tópicos Especiais | | | | | | |
| Confiabilidade Analítica | 2011 | 2 º Semestre | 6 | 90 | 100 | B |
| Técnicas Analíticas em Minerais | 2011 | 2 º Semestre | 2 | 30 | 100 | A |
| Empreendedorismo | 2012 | 1 º Semestre | 8 | 120 | 87 | A |
| Total de créditos em disciplinas | | | 52 | 780 | | |
| TOTAL GERAL DE CRÉDITOS | | | 52 | 780 | | |

Dedico esse trabalho a minha Amada pelo apoio, ajuda e, principalmente, pelo amor que tem dado a mim.

A minha família pela ajuda, pelo apoio e incentivo.

Aos meus amigos, pela convivência e pela atenção.

E a você, Leitor, que deseja aventurar-se neste mundo intrigante chamado Química.

# AGRADECIMENTOS

Primeiramente, quero agradecer a Deus por mais este trabalho realizado. Foram anos de pesquisa e desenvolvimento dois quais não esquecerei.

Quero agradecer a minha namorada, Laisa, por alegrar a minha vida. Tenho muita sorte em ter uma mulher carinhosa, amorosa, caridosa, inteligente, como você em minha vida. Obrigado por estar ao meu lado.

Agradeço ao meu pai, João Munhoz, a minha mãe, Alzira, a minha irmã, Fernanda, as minhas tias, Irma e Ivana, ao meu tio Nelson e a minha tia Penha, as minhas primas Lívia e Miriam, ao meu primo Pedro e a todos da minha família, cujo nome não foi citado. Obrigado pelo apoio em mais essa conquista em minha vida.

Agradeço ao meu sogro, Aristeu, a minha sogra, Maria, ao meu cunhado, Erick, a minha cunhada e seu esposo, Thaís e Ricardo. Obrigado pelo apoio e por me aceitarem em suas vidas.

Agradeço também aos meus amigos que, embora eu não tivesse tido muito tempo para conversar, sempre estiveram ali para me ouvir. Agradeço ao Ronan, que finalmente concluiu o curso de Química, ao Bruno, à Lais, ao Jonatas, ao Guilherme, ao Raul, aos Rodrigos (Snorlax e Porpeta) e a todos os outros cujo nome não me lembro agora enquanto escrevo.

Quero agradecer também a todos na biblioteca que tanto me ajudaram e me aturaram neste período. Agradeço à Cris, à Rita, à Bel, à Valéria e a todos os que trabalham ou trabalharam na biblioteca.

Muitas pessoas me ajudaram durante o desenvolvimento deste trabalho. Agradeço ao Prof. Dr. José Eduardo de Oliveira e ao Prof. Dr. Danilo Luiz Flumignan pela orientação durante o meu Mestrado, ao Prof. Luiz Antônio D'Ávila por ceder os solventes utilizados neste trabalho, ao Dr. Fábio da Silva Vinhado e ao Prof. Dr. Aristeu Gomes Tininis que tiveram a paciência de avaliar este trabalho na minha Defesa.

Enfim, são muitas as pessoas que me ajudaram neste período. Se o seu nome não apareceu peço desculpas e agradeço desde já.

Poderia terminar com alguma frase. Ao invés disso, quero que o leitor a termine por mim. Afinal, o fim de um trabalho representa o começo do próximo.

"Se ao colocar a mão no bolso, retirar apenas seu punho cerrado lembre-se do que você sempre carrega consigo: ..."

João Fernando

“A paciência é uma árvore de raíz amarga, mas de frutos doces.” (Aforismo Persa)

# RESUMO

O Governo brasileiro intervia diretamente na comercialização de combustíveis até a década de 90. Com a abertura do mercado, pequenos revendedores sem contrato exclusivo com qualquer distribuidora, chamados de 'bandeiras brancas', surgiram no ambiente concorrencial. Com o propósito de assegurar a qualidade, a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, ANP, criou portarias especificando as características físico-químicas da gasolina comercial brasileira, tal como a Portaria nº 309. Com a crescente quantidade de informações, a diminuição do tempo de análise é uma necessidade. Uma técnica promissora é a cromatografia gasosa ultrarrápida, capaz de realizar uma corrida em poucos minutos, ou até mesmo em alguns segundos. Para isso, o comprimento da coluna capilar e seu diâmetro interno são reduzidos e seu aquecimento é feito através de uma resistência encapsulada. Assim, é possível obter os mesmos resultados em tempo diminuído e com a mesma qualidade. O aumento do volume de informações obtido experimentalmente exige ferramentas matemáticas mais avançadas para seu tratamento. Desde a década de 70, a quimiometria tem sido aplicada na obtenção, representação e transformação destes dados. O objetivo deste trabalho é criar um método de cromatografia gasosa ultrarrápida para análise de gasolinas comerciais brasileiras e, com ferramentas quimiométricas, classificar quanto à qualidade segundo a ANP. No desenvolvimento, cerca de 580 amostras de gasolinas comerciais foram coletadas mensalmente na região centro-oeste do estado de São Paulo. Os ensaios físico-químicos obtidos revelaram que grande parte não estava em conformidade com as portarias estabelecidas, possuindo apenas um parâmetro em discordância: a quantidade de etanol em sua composição. Deste modo, a adição de adulterantes a algumas amostras foi necessária para aumentar a variabilidade dos dados. Mensalmente, 50 das 580 amostras foram selecionadas através da análise hierárquica de agrupamento, de tal modo que representassem a população coletada no mesmo período. A coleta ocorreu por nove meses, de julho de 2011 a abril de 2012, o número de amostras passou de 5220 para 426 (59,6% conformes e 40,3 % não conformes). Somando com as adulteradas, a quantidade de amostras utilizadas neste trabalho foi de 643. Com o uso da técnica da cromatografia ultrarrápida, o tempo de análise de cada corrida foi de 171 segundos, com boa resolução dos picos. Os cromatogramas obtidos foram alinhados pelo método de *Correlation Optimized Warping*, para a realização do tratamento quimiométrico. Por fim, após encontrar o pré-tratamento ótimo, a modelagem independente suave de analogia de classes conseguiu separar as 643 amostras em dois grupos: gasolinas de boa qualidade e adulteradas. Com base no que foi obtido, pode-se afirmar que a cromatografia gasosa ultrarrápida é uma técnica eficaz na análise de gasolina comercial e com uso de métodos quimiométricos foi possível distinguir uma gasolina de boa qualidade de uma adulterada.

**Palavras-chave**: gasolina, quimiometria, cromatografia gasosa ultrarrápida, cromatografia gasosa, petróleo.

# ABSTRACT

The Brazilian government intervened directly in the commercialization of fuel until de 90s. By opening of the marketing, small dealers without exclusive contract with any distributor, called 'white flags', appeared in the competitive environment. In order to ensure quality, the National Agency of Petroleum, Natural Gas and Biofuels, ANP, has created ordinances which specify physico-chemicals characteristics of Brazilian commercial gasoline, such as the Ordinance No 309. By increasing the amount of information, reducing the analysis time is a necessity. A promising technique is the ultra-fast gas chromatography, able to perform a chromatographic run in just few minutes, or even in seconds. Thereunto, the capillary column length and its internal diameter are reduced and its heating is done via an encapsulated resistor. It is possible to obtain the same results in reduced time and with the same quality. The increasing volume of information obtained experimentally requires more advanced mathematical tools for their treatment. Since the 70s, Chemometrics has been applied in acquisition, representation and processing these data. The objective of this work is to create an ultra-fast gas chromatography method for analisys of the Brazilian commercial gasoline and, with chemometric tools, sort by their quality according to ANP. In development, about 580 commercial gasoline samples were collected monthly in the central-western state of São Paulo. The physico-chemical assays reveled that a great fraction of which was not in accordance with the ordinances established, having only one parameter in disagreement: the amount of ethanol in its composition. The addition of adulterants to some samples was necessary to increase the variability of the data. Monthly, 50 of 580 samples were selected by the hierarchical cluster analysis which represented the population collected in the same period. The gathering occurred for nine months, from July 2011 to April 2012, the number of samples decreased from 5220 to 426 (59,6% in accordance e 40,3% in non-accordance). With the adulterated, the amount of samples used in this study was 643. Using the ultrafast chromatography technique, the analysis time for each run was 171 seconds. The chromatograms obtained were aligned by the Correlation Optimized Warping method to perform the chemometric treatment. Finally, after finding the optimal pretreatment, the soft independent modeling of class analogy managed to separate the 643 samples into two groups: good quality gasoline and adulterated. Based in which has been obtained, it can be affirmed that ultra-fast gas chromatography is an effective technique for analysis of commercial gasoline by using chemometrics methods. It has successfully distinguished a good quality gasoline from an adulterated.

**Key words**: gasoline, chemometrics, ultra-fast gas chromatography, gas chromatography, petroleum.

# LISTA DE FIGURAS

# LISTA DE TABELAS

# LISTA DE QUADROS

# LISTA DE ABREVIATURAS E SIGLAS

**AEAC** – Álcool Anidro Etílico Combustível
**ANN** – Redes Neurais Artificiais
**ANP** – Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
**Aro** - Aromático
**ASTM** – *American Society for Testing and Materials*
**Bz** - Benzeno
**CEMPEQC** – Centro de Monitoramento e Pesquisa da Qualidade de Combustíveis, Biocombustíveis, Petróleo e Derivados
**CFC** – Craqueamento por Fluido Catalítico
**CG** – Cromatografia Gasosa
**CG-C** – Cromatografia Gasosa Convencional
**CGL** – Cromatografia gás-líquido
**CG-MR** – Cromatografia Gasosa Muito Rápida
**CG-R** – Cromatografia Gasosa Rápida
**CGS** – Cromatografia gás-sólido
**CG-UR** – Cromatografia Gasosa Ultrarrápida
**CLS** – Quadrados Mínimos Clássicos
**COW** – Correlação de Deformação Otimizada
**DOU** – Diário Oficial da União
**DP** – Programação Dinâmica
**DTW** – Deformação Dinâmica do Tempo
**ECD** – *Eléctron Capture Detector*
**FID** – Detector de Chama Ionizante
**GA** – Algoritmo Genético
**HCA** – Análise Hierárquica de Agrupamentos
**IAD** – Índice Antidetonante
**ILS** – Quadrados Mínimos Inversos
**KNN** – Regra dos Vizinhos mais Próximos
**MAPA** – Ministério da Agricultura, Pecuária e Abastecimento
**MEG** – Mistura Metanol-Etanol-Gasolina
**MEsp** – Massa Específica
**MLR** – Regressão Linear Múltipla
**MME** – Ministério de Minas e Energia
**MON** – *Motor Octane Number*
**MS** – Espectroscopia de Massa
**MSC** – Correção Multiplicativa de Sinal
**NIPALS** – Quadrados Mínimos Parciais Não Interativos
**N-PLS** – Mínimos Quadrados Parciais Não-lineares
**Ole** – Olefínicos
**PC** – Componentes Principais
**PCA** – Análise de Componentes Principais
**PCR** – Regressão por Componentes Principais
**PFE** – Ponto Final de Ebulição
**PLS** – Quadrados Mínimos Parciais
**PMC** – Produtos de Marcação Compulsória
**PRESS** – Predição do Erro Residual da Soma de Quadrados
**PVR** – Pressão de Vapor REID
**Res** – Resíduo de Destilação
**RMN** – Ressonância Magnética Nuclear
**RON** – *Research Octane Number*
**RP** – Reconhecimento de Padrões
**Sat** – Hidrocarbonetos Saturados
**SECV** – Erro Padrão da Validação Cruzada
**SEF** – *Speed Enhancement Factor*
**SEP** – Predição do Erro Padrão
**SIMCA** – Modelagem Independente Suave de Analogia de Classes

**SNV** – Método da Variação de Padrão Normal
**T10%** - Temperatura após 10% destilado
**T50%** - Temperatura após 50% destilado
**T90%** - Temperatura após 90% destilado
**TCD** – *Thermal Conductivity Detector*
**TMP** – Poder de Modelamento Total
**UFM** – Módulo Ultrarrápido
**Unesp** – Universidade Estadual Paulista Júlio de Mesquita Filho

# SUMÁRIO

## *1 INTRODUÇÃO*

Energia. O mundo atual é muito dinâmico. Tudo está em constante movimento. A velocidade de aquisição e processo de informações cresce a cada dia e para que tudo continue em ação, continue infalível, precisa-se de energia.

Uma das principais fontes de energia é o combustível fóssil. Sua origem provém da decomposição de matéria orgânica e da deposição de sedimentos através de milhares de anos. Posto que sua composição é rica em hidrocarbonetos, sua combustão gera uma grande quantidade de energia. Sendo de suma importância para o mundo atual, seu uso é dos mais variados; pode ser empregado em atividades domésticas, na forma de gás de cozinha, ou até mesmo na movimentação de veículos de grande porte. Todos os dias, de alguma forma, faz-se uso deste tipo de combustível (CONAWAY, 1999).

Um das suas principais aplicações é na forma de gasolina. Sendo esta comercializável por várias distribuidoras, para consumo em veículos automotivos. No Brasil, a gasolina comercial é vendida, principalmente, com Álcool Etílico Anidro Combustível (AEAC) misturado em sua composição. Esta mistura é denominada de gasolina comercial brasileira do tipo C (PETROBRAS, 2000).

Até a década de 90, o Governo brasileiro intervia no mercado de combustíveis. Após a desregulamentação em 01 de Janeiro de 2002, formou-se um ambiente concorrencial no setor de derivados de petróleo. Em 1993, o Ministério de Minas e Energia (MME) autorizou a participação de revendedoras sem contrato exclusivo com qualquer distribuidora, os chamados ‘bandeiras brancas’.

Com o surgimento dos ‘bandeiras brancas’ e de pequenas distribuidoras, algumas práticas irregulares foram facilitadas, para assegurarem sua sobrevivência neste novo espaço competitivo. Algumas delas são: sonegação de impostos, descumprimento de contratos de exclusividade, contrabando e adulteração de gasolina.

A fim de fiscalizar sua qualidade, a Agencia Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) fixou vários parâmetros físico-químicos para a gasolina comercial brasileira. Porém, apenas alguns laboratórios especializados conseguem realizar todos os ensaios necessários para averiguação da qualidade da

gasolina comercial brasileira. O número de amostras coletadas mensalmente é elevado. Um laboratório de analise de combustíveis pode receber acima de 500 amostras de gasolina por mês a serem examinadas para um grande número de parâmetros, cada um determinado por um ensaio específico. Isto somado a outros combustíveis tais como, etanol, diesel e biodiesel, resulta em um grande esforço mensal do laboratório.

Trabalhos anteriores mostraram que é possível através de técnicas espectrométricas ou dos perfis cromatográficos aliados a ferramentas quimiométricas, realizar uma estimativa de parâmetros físico-químicos de gasolinas brasileiras, diminuindo significativamente os tempos de avaliação das amostras (FLUMIGNAN, 2006; FLUMIGNAN, 2007).

Várias técnicas de análise são exploradas e desenvolvidas, com o objetivo de diminuir os tempos de análise. Uma destas é a cromatografia gasosa ultrarrápida (CG-UR) capaz de realizar corridas com tempos inferiores a 1 minuto. Sua diferença para com a cromatografia gasosa convencional (CG-C) está no menor comprimento da coluna capilar, de 2 a 10m, diâmetro interno reduzido, entre 0,05 e 0,10mm, aquecimento resistivo da coluna, ao invés do uso do forno convencional e taxa elevada de aquecimento da coluna, superior a 60ºC $min^{-1}$(SEQUINEL, 2010). Além disso, a CG-UR utiliza os mesmos tipos de injetores e detectores que a CG-C, não havendo a necessidade de um equipamento distinto para ambos, bastando apenas acoplar, ao CG convencional, o módulo ultrarrápido (também chamado “gaiola”), que consiste em um sistema dedicado de aquecimento da coluna capilar. Em outras palavras é possível ter as duas técnicas em um mesmo equipamento.

Outro contratempo a se considerar é o grande volume de informação gerado destes ensaios. Ferramentas matemáticas mais avançadas são necessárias para seu tratamento. Assim, a rápida evolução dos computadores, bem como o avanço da estatística, facilitou o emprego da quimiometria na obtenção, representação e transformações dos dados obtidos através de medições experimentais.

Os cálculos quimiométricos podem ser usados para distinguir e classificar as amostras de gasolina comercial brasileira quanto à qualidade determinada pela resolução expedida pela ANP e assim, assegurar a qualidade do derivado de petróleo para o consumidor final.

As coletas mensais das amostras deste trabalho foram feitas em diversos postos distribuidores de combustíveis na região centro-oeste do estado de São Paulo. Os ensaios físico-químicos foram realizados no Centro de Monitoramento e Pesquisa da Qualidade de Combustíveis, Biocombustíveis, Petróleo e Derivados (CEMPEQC) localizado no Instituto de Química da Universidade Estadual Paulista Júlio de Mesquita Filho, Unesp, de Araraquara/SP.

O período de coleta foi de julho de 2011 a abril de 2012. Em 1º de Outubro de 2011, o Diário Oficial da União (DOU) publicou uma Portaria que alterava a quantidade de AEAC na gasolina comercial brasileira de 25±1% para 20±1%. Assim, houve a necessidade de separar o banco de dados totais em: Banco I, que corresponde às amostras com 25%±1% de AEAC, conforme determinado pela Portaria nº 143 do Ministério de Agricultura, Pecuária e Abastecimento, MAPA de 27/06/2007, e Banco II, com 20±1% de AEAC, de acordo com a Portaria nº 678 do MAPA de 31/11/2011.

Como o número de amostras era muito alto, realizou-se a seleção de amostras representativas, do grupo completo, através da análise hierárquica de agrupamentos, HCA, de forma que o resultado final seja o mesmo que o obtido com todas as amostras disponíveis.

As amostras separadas foram armazenadas e, então, submetidas a análises cromatográficas ultrarrápidas. Por fim, os cromatogramas obtidos dos Bancos I e II, foram alinhados através do método de *Correlation Optimized Warping* (COW) e, em seguida, analisados pela Modelagem Independente Suave de Analogia de Classes (SIMCA). Sendo que, este último classificava as amostras quanto à conformidade com a ANP.

O objetivo deste trabalho é a busca de alternativas mais rápidas e eficientes de análise de gasolina comercial brasileira. Aliando-se a CG-UR com métodos quimiométricos, é possível ter uma grande diminuição no tempo de análise de um grande número de amostras, não somente na área de combustíveis, mas em qualquer outra que se deseja explorar.

## *2 OBJETIVOS*

Desenvolver um método de cromatografia gasosa ultrarrápida para análise de gasolinas comerciais brasileiras. Aplicar a metodologia envolvida em um conjunto de amostras coletadas de diferentes fornecedores, correlacionando os perfis cromatográficos com as especificações descritas na Resolução nº. 57 de 2011 da Agência Nacional de Petróleo, Gás Natural e Biocombustíveis. Classificar as amostras quanto à sua qualidade e atendimento às especificações legais, através da aplicação de métodos quimiométricos de reconhecimento de padrões multivariados.

# *3 LEVANTAMENTO BIBLIOGRÁFICO*

## *3.1 Gasolina Comercial Brasileira*

### 3.1.1 Extração e Obtenção

Originalmente, a gasolina era um sub-produto indesejado na indústria de refinamento de petróleo, cujo interesse estava na obtenção de querosene para iluminação. No final do século XIX, com o advento dos motores à combustão, surgiu o interesse no uso da gasolina como combustível, devido à sua alta energia de combustão, alta volatilidade e compressibilidade (FLUMIGNAN, 2006).

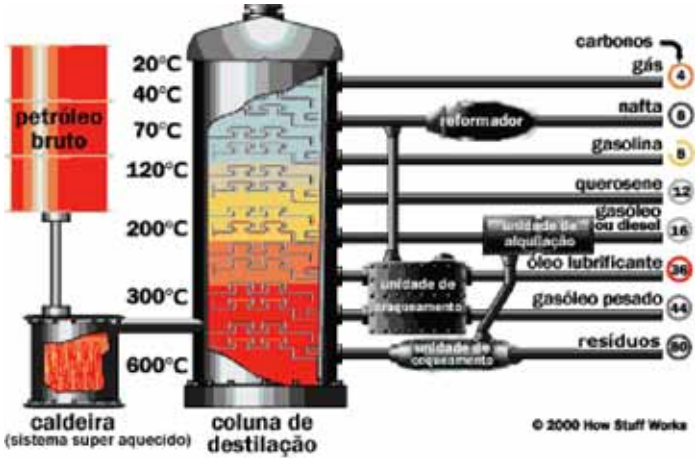
Hoje em dia, a gasolina é um dos combustíveis fósseis mais consumidos no mundo. É constituída por uma mistura de hidrocarbonetos parafínicos, olefínicos naftênicos e aromáticos e, em menor quantidade, de produtos oxigenados e sulfurados. Os hidrocarbonetos apresentam estruturas variadas em diferentes proporções, com cadeias carbônicas entre 4 e 12 átomos de carbono por moléculas e pontos de ebulição entre 30 e 220ºC (FLUMIGNAN, 2006).

A gasolina comercial é obtida através do processo de destilação fracionada do petróleo bruto, conforme indicado na figura 1. A fim de se obter um maior rendimento, novas técnicas surgiram. Uma delas é denominada *craqueamento*, que consiste em quebrar (do inglês *to crack*) moléculas maiores em outras menores, dentro da faixa pertencente à fração da gasolina. Este processo é ilustrado na figura 2.

Até 1937, essa técnica era realizada em altas temperaturas, recebendo o nome de craqueamento térmico. Com o passar do tempo, foi introduzido o craqueamento catalítico, atualmente em uso, que consiste na quebra de moléculas maiores em altas temperaturas, com o auxílio de um catalisador (podendo ser zeólitas, hidrossilicato de alumínio, bauxita, ou alumino-silicatos), para aumentar a velocidade da reação. Este método é mais econômico e eficaz que seu predecessor. No presente, o

processo mais empregado é o craqueamento por fluido catalítico (CFC) (FLUMIGNAN, 2006).

**Figura 1** - Diagrama de um processo de destilação fracionada do petróleo bruto.

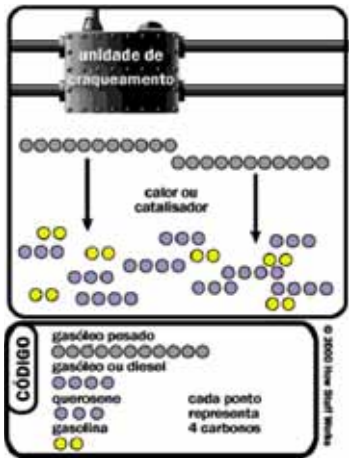



**Fonte**: HOW STUFF WORKS, 2012.

As vantagens apresentadas pelo craqueamento catalítico em relação ao craqueamento térmico são: isomerização de maior número de olefinas, maior controle da saturação das duplas ligações, menor produção de olefinas líquidas na faixa da gasolina e maior produção de hidrocarbonetos aromáticos. Isto faz com que a gasolina obtida por esta técnica apresente um alto índice de octano e baixo teor de goma lavada (FLUMIGNAN, 2006).

Há ainda outras técnicas de obtenção de gasolina como a polimerização, que consiste na conversão de olefinas gasosas e de cadeia pequena (propeno e buteno) em moléculas maiores, a alquilação, processo que combina uma olefina e uma molécula de isobutano, e isomerização, conversão de hidrocarbonetos de cadeia normal em hidrocarbonetos de cadeia ramificada (FLUMIGNAN, 2006).

**Figura 2** - Diagrama de um processo de craqueamento do petróleo bruto.




**Fonte**: HOW STUFF WORKS, 2012.

### 3.1.2 Bandeira Branca

A estrutura do mercado de combustíveis no Brasil foi marcada por excessiva intervenção governamental até a década de 90. A partir daí, iniciou-se o processo de desregulamentação, até a total abertura do mercado, que ocorreu em 01 de Janeiro de 2002. O estabelecimento de um ambiente concorrencial no setor de derivados de petróleo tem por objetivo a proteção aos interesses do consumidor quanto a preço, qualidade e oferta dos produtos e a promoção da livre concorrência, de acordo com a Lei nº 9.478 de 06 de agosto de 1997 art. 1º, inciso III e IX (PINTO, 2004).

Foi autorizada em 1993, com a Portaria do MME nº 362, a participação no mercado de revendedoras sem contrato exclusivo com qualquer distribuidora. Tal agente foi chamado de ‘bandeira branca’. O mercado brasileiro tinha sua indústria de combustíveis caracterizada por contratos exclusivos entre distribuidoras e revendedoras, ou seja, cada revendedora, obrigatoriamente, estabelecia um contrato de exclusividade com uma distribuidora e apenas adquiria combustível desta. A mudança desta estrutura foi, sem dúvida, um fator importante para o estabelecimento de uma nova dinâmica de formação de preços, governada por forças de mercado (PINTO, 2004).

Com o surgimento do revendedor de bandeira branca e de pequenas distribuidoras no mercado, várias práticas irregulares foram facilitadas, como por

exemplo, o descumprimento de contratos de exclusividade, sonegação de impostos, contrabando de gasolina e adulteração de combustíveis. Tais práticas causam distorções no funcionamento do mercado, inviabilizam a competição, lesam o consumidor e o contribuinte, reduzem a arrecadação dos estados e da união, estimulam a corrupção e o crime organizado (PINTO, 2004).

### 3.1.3 Especificações da Gasolina Comercial Brasileira

As especificações e o controle da qualidade das gasolinas automotivas brasileiras seguem uma variedade de parâmetros físico-químicos estabelecidos pela Resolução nº 57 da ANP expedida em 20 de outubro de 2011. Especificamente, o teor de AEAC foi estabelecido pela Portaria nº 143 do MAPA publicada no DOU em 10 de Outubro de 2001, cujo foi fixado em 25±1% (v/v). (CAMARGO, 2009; AGÊNCIA..., 2013a; BRASIL, 2013a)

Com a alta no preço do álcool etílico no mercado brasileiro, em 1º de Outubro de 2011, a Portaria MAPA nº 678 fixou em 20±1% (v/v), o percentual obrigatório de adição de AEAC à gasolina comercial brasileira. Alguns dos parâmetros estabelecidos pela ANP e pelo MAPA estão descritos na tabela 1 (CAMARGO, 2009; BRASIL, 2013b).

**Tabela 1** - Especificações vigentes da gasolina, segundo a Resolução nº 57 da ANP de 2011 e as Portarias nº 143 e 678 do MAPA.

| **Características** | **Unidade** | **Especificação Gasolina Comum** | **Método** | |
|---|---|---|---|---|
| | | | **ABNT NBR** | **ASTM** |
| AEAC[1] | % vol. | 25±1 (Portaria MAPA nº 143)<br>20±1 (Portaria MAPA nº678) | Cromatografia/13992 | - |
| Teor de Metanol, máx. | % vol. | 0,5 | Cromatografia | |
| Cor | - | Incolor a amarelada | 14954 | D4176 |
| Aspecto | - | Límpido e isento de impurezas | | |
| Massa Específica a 20 ºC | $Kg/m^3$ | Anotar | 7148/14065 | D1298/D4052 |
| Destilação | | | | |
| 10% evaporado, máx.<br>50% evaporado, máx.<br>90% evaporado, máx.<br>PFE, máx.<br>Residuo, máx. | ºC<br><br><br><br>%vol. | 65,0<br>80,0<br>190,0<br>220,0<br>2,0 | 9619 | D86 |
| Nº de Octano Motor – MON, min. | - | 82,0 | - | D2700 |
| Índice Antidetonante – IAD, min. | - | 87,0 | - | D2699/D2700 |
| Benzeno – máx. | % vol. | 1,0 | - | D3606/D5443/D6277 |
| Hidrocarbonetos | | | | |
| Aromáticos – máx.<br>Olefínicos – máx. | % vol. | 45<br>30 | 14932 | D1319 |

[1] Valor fixado após 1º de Outubro de 2011. Antes desta data, a percentagem de AEAC era de 25±1%.

**Fonte**: AGÊNCIA..., 2013a; BRASIL, 2013a; BRASIL, 2013b.

### 3.1.4 Tipos de Gasolina

Definidos pela mesma Resolução nº 57 da ANP, há dois tipos de gasolinas: a gasolina A e a gasolina C. Estas podem ser descritas como:

- Gasolina A Comum: combustível produzido por processo de refino de petróleo ou formulado por meio da mistura de correntes provenientes do refino de petróleo e processamento de gás natural, destinado aos veículos automotivos dotados de motores ciclo Otto, isento de componentes oxigenados (AGÊNCIA..., 2013a). Este produto é a base da gasolina disponível nos postos revendedores (PETROBRAS, 2000);

- Gasolina A Premium: isenta de compostos oxigenados e formulada a partir de uma mistura de naftas de elevada octanagem, sejam craqueadas, sejam alcoiladas, sejam reformadas, as quais fornecem maior resistência à detonação que a gasolina do tipo A Comum. Constitui a base da gasolina PREMIUM, destinada aos consumidores finais (PETROBRAS, 2000).

- Gasolina C Comum: combustível obtido da mistura de gasolina A e AEAC, nas proporções definidas pela legislação em vigor (AGÊNCIA..., 2013a). Produto final disponível no mercado, para atender automóveis, motos, embarcações aquáticas etc. (PETROBRAS, 2000);

- Gasolina C Premium: obtida pela mistura de gasolina A Premium e AEAC, nas proporções e especificações definidas pela legislação em vigor e que atenda ao Regulamento Técnico (AGÊNCIA..., 2013a). Esta gasolina foi desenvolvida com o intuito de atender aos veículos, nacionais ou importados, de altas taxas de compressão e alto desempenho e que tenham recomendação de utilizar combustíveis de elevada resistência à detonação (PETROBRAS, 2000).

Além destes quatro tipos, existem outras formulações de gasolinas com maior especificidade.

- Gasolina Aditivada: formulada através da adição, à gasolina C, de aditivos, os quais, dentre diversas características, são detergentes e dispersantes, tendo a finalidade de melhorar o desempenho do produto, minimizando a formação de depósitos no carburador e bicos injetores, bem como no coletor e hastes das válvulas de admissão. Além disso, a gasolina aditivada recebe um corante de identificação para fácil distinção visual (PETROBRAS, 2000);

- Gasolina Padrão: produzida especialmente para uso na indústria automobilística. Comumente, utilizadas nos ensaios de avaliação do consumo e das emissões de poluentes dos veículos recém-produzidos pelas montadoras. É produzida sob encomenda (PETROBRAS, 2000);

- Gasolina de Aviação: apresenta alto índice de desempenho e alta octanagem, além de outras características especiais. Uso exclusivo em aeronaves que possuem motores de ignição por centelhas (PETROBRAS, 2000);

- Gasolinas Especiais: formulada apenas sob encomenda, para aplicações específicas, possuindo composições diferentes das demais apresentadas. É o caso da gasolina produzida pela PETROBRAS para a Fórmula 1 e da gasolina produzida para a Fórmula 3 (PETROBRAS, 2000).

### 3.1.5 Características da Gasolina Comercial

Os parâmetros físico-químicos, descritos na tabela 1, serão abordados mais detalhadamente abaixo.

#### *3.1.5.1 Aspecto*

Indicação visual da qualidade e de possível contaminação do combustível. A gasolina deve apresentar-se límpida e isenta de materiais em suspensão. Se presentes, estes podem reduzir a vida útil dos filtros de combustíveis dos veículos e prejudicar o funcionamento dos motores. O teste é realizado observando-se contra a luz uma amostra de 0,9L do produto contida em um recipiente vítreo transparente, com capacidade total de 1L (PETROBRAS, 2000).

#### *3.1.5.2 Cor*

Indicação da tonalidade característica. No caso das gasolinas A e C, sem aditivos, a cor pode variar de incolor a amarelo. Já as aditivadas, recebem corantes específicos, para a diferenciação das demais. Pode apresentar quaisquer cores, exceto a azul, destinada para a gasolina de aviação e rosa, reservada para a mistura Metanol-Etanol-Gasolina, MEG. Alteração na coloração pode ocorrer devido à presença de contaminantes e impurezas, ou mesmo à oxidação de compostos instáveis, como olefinas e compostos nitrogenados (PETROBRAS, 2000).

*3.1.5.3 Teor de Enxofre*

O enxofre é um componente indesejável em qualquer combustível, devido à ação corrosiva de seus compostos e à formação de gases tóxicos como o $SO_2$ e $SO_3$, que ocorre durante a combustão do produto. Em veículos dotados de catalisadores, quando a carga de material catalítico não é adequada, ou quando não está devidamente dimensionada, o enxofre pode levar à formação de $H_2S$, o qual é tóxico e apresenta odor desagradável (AMERICAN..., 2007).

A análise é realizada incidindo raios X em uma célula contendo amostra do produto. Neste teste, os átomos de S absorvem energia de um comprimento de onda específico numa quantidade proporcional à sua concentração na gasolina. Outra forma de análise aplicável é o "método da lâmpada", que se consiste na queima de uma amostra com uma lamparina especial em atmosfera especial (30% de $O_2$ e 70% de $CO_2$). Os óxidos gerados são absorvidos em uma solução de $H_2O_2$, produzindo ácido sulfúrico. Por fim, através de cálculos do ácido, determina-se a concentração de enxofre (AMERICAN..., 2007).

*3.1.5.4 Destilação*

A destilação é um dos testes que tem como objetivo avaliar as características de volatilidade da gasolina. O teste é feito tomando-se 100mL da amostra do produto, que são colocados em um balão de vidro especial que, a seguir, é submetido a aquecimento para destilação em condições controladas. Após esta operação, as temperaturas anotadas são corrigidas levando-se em conta as perdas, que ocorrem por evaporação de pequena parte do produto e a pressão barométrica. Este teste, além de ser usado no controle da produção da gasolina, pode ser utilizado para identificar a ocorrência de contaminação por derivados mais pesados como o óleo diesel, óleo lubrificante, querosene etc. A seguir, há uma breve explicação acerca dos pontos especificados na destilação da gasolina e que visam atender às diferentes condições de operação dos motores (AMERICAN..., 2009).

*3.1.5.5 10% Evaporado*

Ponto da curva de destilação da gasolina que indica a temperatura, na qual 10% do produto são destilados. O controle deste ponto da curva tem por objetivo garantir que o combustível possua uma quantidade mínima de frações leves que se vaporizem e queimem, com facilidade, na temperatura de partida a frio do motor, facilitando o início de funcionamento do veículo. Ao mesmo tempo em que facilita a partida a frio, uma concentração muito alta de frações leves pode dificultar a partida a quente e prejudicar a dirigibilidade do veículo, devido à geração de bolhas na linha de combustível (AMERICAN..., 2009).

*3.1.5.6 50% Evaporado*

Ponto da curva de destilação da gasolina que indica a temperatura, na qual 50% do produto são destilados. O controle desta característica também visa possibilitar uma partida fácil do motor, porém sua principal influência se faz sentir no tempo necessário ao seu aquecimento. Isto reflete à concentração de componentes de temperatura de destilação intermediária que apresentam queima relativamente fácil e fornecem uma quantidade de energia superior àquela fornecida pelas frações mais leves representadas pelos 10% inicialmente evaporados. Assim, é uma fração cujas características contribuem, diretamente, para que o motor entre em regime de operação uniforme (AMERICAN..., 2009).

*3.1.5.7 90% Evaporado*

Ponto da curva de destilação da gasolina que indica a temperatura, na qual 90% do produto são destilados. A fixação de um valor máximo para a temperatura de destilação de 90% da gasolina visa minimizar a formação de depósitos na câmara de combustão e nas velas de ignição, o que ocorre se a temperatura correspondente for

muito elevada. Quanto mais alta for a temperatura necessária para vaporizar as frações mais pesadas de uma gasolina, maior quantidade dela sobrará na câmara de combustão sem queimar. Essa parte do combustível, que não se queima, tende a vazar para o cárter do motor, entrando em contato com o cilindro e contaminando o óleo lubrificante. Esta ocorrência prejudica a lubrificação e acelera o desgaste das camisas, devido à remoção da película protetora formada pelo óleo lubrificante. No que diz respeito à emissão de poluentes, uma gasolina que exige alta temperatura para a evaporação de suas frações finais contribui mais na emissão de hidrocarbonetos pelos veículos, do que uma gasolina que é evaporada em baixas temperaturas. A Resolução da ANP nº 57 define um valor máximo, a ser atendido pela gasolina, para esta e as demais características. O objetivo é possibilitar a identificação de contaminantes que possam a vir estar presente no produto (AMERICAN..., 2009).

*3.1.5.8 Ponto Final de Ebulição*

O Ponto Final de Ebulição (PFE) é a mais alta temperatura verificada durante a destilação da gasolina. O PFE pode ou não coincidir com a vaporização total do produto, tendo em vista que a parte final da gasolina pode sofrer decomposição térmica, craqueamento, fazendo com que se formem frações mais leves, que reduzem a temperatura na fase final do teste. Neste caso, a maior temperatura lida durante o ensaio é registrada como ponto final de ebulição. Não havendo decomposição da fração final da gasolina, o PFE corresponderá à temperatura em que se verifica a destilação de todo o volume da gasolina tomado para o teste (AMERICAN..., 2009).

O controle do PFE da gasolina visa minimizar a formação de depósitos na câmara de combustão e nas velas de ignição. De um modo geral, um alto valor para o PFE apresenta influências para a destilação de 90% do volume do produto. Valores do PFE fora da especificação podem ser indicativos de contaminação por óleo diesel, querosene, ou lubrificante (AMERICAN..., 2009).

*3.1.5.9 Resíduo de Destilação*

Parte da gasolina que sobra após ter-se alcançado o ponto final de ebulição. A ocorrência de alta percentagem de resíduos pode estar relacionada com a instabilidade térmica das frações finais remanescentes após a destilação contínua do combustível em questão. Um alto teor de resíduo, assim como um elevado ponto final de ebulição, indica um alto teor de frações pesadas advindas do processo de produção ou de contaminação da gasolina por produtos mais pesados, como óleo diesel. O uso de gasolina com alto teor de resíduo contribui para maior formação de depósitos no motor e pode provocar carbonização das velas de ignição, o que leva o motor a funcionar precariamente (AMERICAN..., 2009).

*3.1.5.10 Pressão de Vapor REID*

A Pressão de Vapor REID (PVR) tem como objetivo avaliar a tendência da gasolina a evaporar-se de modo que, quanto maior é a pressão de vapor, mais facilmente a gasolina evapora, o que é influenciado pela quantidade de frações leves presentes no produto. Este ensaio é utilizado para indicar as exigências que devem ser satisfeitas para o transporte, armazenamento e uso do produto, de modo a evitar acidentes, minimizar as perdas por evaporação e garantir a partida fácil do veículo (AMERICAN..., 2006).

Uma pressão de vapor muito alta pode levar à ocorrência de tamponamento do fluxo de combustível. Este efeito é provocado pelos vapores da gasolina que bloqueiam a tubulação impedindo que o combustível seja bombeado até o carburador ou bicos injetores. Isto dificulta a partida dos veículos em dias quentes, quando a evaporação é acelerada pelo calor ambiente, inclusive calor do asfalto quente, que é transmitido para a parte inferior do tanque dos veículos. Uma alta pressão de vapor também contribui para um aumento da poluição devido à ocorrência, em maiores níveis, de emissões evaporativas de gasolina através do suspiro do tanque de combustível (AMERICAN..., 2006).

O teste é feito colocando-se uma amostra da gasolina resfriada, entre 0 a 1ºC, na câmara de amostra da bomba para pressão de vapor, que também é provida de um manômetro. A seguir, esta aparelhagem é imersa em um banho a 37,8ºC. Procede-se à agitação da bomba de pressão de vapor, prosseguindo-se com o teste até que se obtenha uma pressão constante. A leitura do manômetro é reportada como PVR, a 37,8ºC (AMERICAN..., 2006).

### *3.1.5.11 Número de Octano*

Embora haja diferentes tipos de gasolina sua característica fundamental de qualidade é a octanagem e seu valor está diretamente relacionado ao bom funcionamento de motores de combustão interna. Para sua determinação, podem ser utilizados os métodos:

- *Motor Octane Number* (MON): avalia a resistência da gasolina à detonação, na situação em que o motor está em plena carga e em alta rotação (AMERICAN..., 2012c);
- *Research Octane Number* (RON): avalia a resistência da gasolina à detonação, na situação em que o motor está em carregado e em baixa rotação (até 3000 rpm) (PETROBRAS, 2011);
- Índice Antidetonante (IAD): obtido através da média aritmética entre MON e RON, (MON + RON) /2 (PETROBRAS, 2011).

### *3.1.5.12 Goma Atual Lavada*

Representa a quantidade de produto de aspecto resinoso presente na gasolina e que não é facilmente evaporável e nem se queima com facilidade. A goma, ou "verniz", constitui-se de moléculas de cadeias carbônicas grandes que tendem a se precipitar, separando do produto. Sua formação ocorre quando compostos olefínicos

presentes em gasolina sofrem oxidação, pelo oxigênio do ar, ou quando reagem entre si ou com outros hidrocarbonetos, na presença de luz ou calor (AMERICAN..., 2012b).

Logo após a produção da gasolina, a percentagem de goma presente é relativamente baixa, cerca de 1mg/100mL na gasolina tipo A. Este teor pode vir a aumentar durante a estocagem, em que o combustível tem contato com o ar atmosférico. O uso de gasolina com alta concentração de goma pode provocar a formação de grandes quantidades de depósitos nas hastes das válvulas de admissão e no carburador do veículo, podendo estrangular o "gigleur" e trancar a borboleta do segundo grupo do carburador. Nos bicos injetores, estes depósitos podem prejudicar a vazão e a forma do jato de combustível, prejudicando o desempenho do motor (AMERICAN..., 2012b).

Para minimizar a formação da goma até um teor aceitável, regulado pela ANP, a PETROBRAS acrescenta à gasolina, durante a fase da produção, um composto antioxidante que retarda o início das reações de degradação do produto, resultando na obtenção de uma gasolina de boa estabilidade. Ainda assim, quando a gasolina é colocada em contato com cobre e suas ligas, muitas vezes presentes em válvulas, conexões e filtros das instalações de armazenamentos e postos de revenda, a reação que leva à formação de goma é catalisada, provocando a degradação do produto. Além disso, o álcool etílico, presente na gasolina tipo C, costuma apresentar cobre como contaminante (PETROBRAS, 2000).

O teste para determinar o teor de goma atual lavada é realizado tomando 50mL do produto, após sua filtração em funil de vidro sintetizado, e aquecê-lo em banho a uma temperatura de 160 a 165ºC, com uma corrente de ar dirigida para o interior do poço do banho à razão de 1000±150mL $s^{-1}$. Passados 30min, remove-se e resfria-se por 2h em dessecador e, em seguida pesa-se a amostra. Depois, adiciona-se 25mL de n-heptano, a fim de remover aditivos e óleos não voláteis, se presentes. Após a lavagem, o recipiente com o resíduo remanescente é novamente colocado no banho evaporador para secar por 5min. Após este tempo e resfriar por 2h, faz-se a pesagem final e calcula-se o teor de goma lavada em mg/100mL (AMERICAN..., 2012b).

*3.1.5.13 Período de Indução*

Teste realizado, a fim de avaliar a estabilidade da gasolina quanto à estocagem. Esta característica é extremamente importante para a amostra, visto que a mesma deve apresentar um mínimo de resistência à oxidação, a qual leva a formação da goma durante a armazenagem (PETROBRAS, 2000).

O período de indução é definido como o período de tempo decorrido, em minutos, até que uma amostra de gasolina apresente uma queda de pressão de 2 PSI, submetida a um teste de oxidação em atmosfera de oxigênio a 100 ± 2 PSI e a 100 ºC, num intervalo de 15 minutos, acompanhada de outra queda de pressão, não inferior a 2 PSI, nos 15 minutos seguintes. Tal fato ocorre devido ao consumo de oxigênio na reação de oxidação do produto (PETROBRAS, 2000).

*3.1.5.14 Percentagem de Álcool Etílico Anidro*

Indicação do teor de álcool etílico presente na gasolina automotiva. A avaliação desta característica é de grande importância, tendo em vista que quando este produto é adicionado em excesso ou em teor menor do que o especificado, podendo comprometer o bom funcionamento dos veículos que, no Brasil, devem ser regulados para consumir uma gasolina C com 20 a 25% em volume de etanol. Esta adição de álcool contribui para uma pequena elevação da octanagem da gasolina, o que também pode ser obtido com pequenas alterações na formulação da gasolina utilizando-se puramente os produtos derivados do petróleo (ASSOCIAÇÃO..., 2003).

O consumo de uma gasolina com o teor de álcool etílico inferior ao mínimo especificado poderá levar o motor do veículo a apresentar problemas como detonação (batida de pino), formação de depósitos generalizados de fuligem e carbonização das velas de ignição, o que pode causar falhas no seu funcionamento. Tais ocorrências se devem à insuficiência de ar para queima completa do combustível, tendo em vista que, com a redução do teor de álcool, menos oxigênio estará disponível para participar da queima da gasolina. Além destes problemas, poderá ocorrer um aumento significativo

do teor de CO, proveniente da queima incompleta do combustível, enquanto poluentes como aldeídos e óxidos de nitrogênio são reduzidos (ASSOCIAÇÃO..., 2003).

Por outro lado, o uso de gasolina com o teor de álcool superior ao especificado e para o qual o motor foi regulado, levará o veículo a apresentar perda de potência acompanhada de um aumento no consumo de combustível o que se pode creditar ao menor poder energético do álcool e ao desequilíbrio da relação ar/combustível (ASSOCIAÇÃO..., 2003).

### *3.1.5.15 Teor de Metanol*

Identificação do teor de metanol presente na gasolina comercial brasileira. A determinação quantitativa de metanol é de importância, pois além da sua alta toxicidade, o imposto atribuído a esta substância é bem menor do que ao AEAC. Facilitando, assim, a sonegação e outras práticas irregulares. Além disso, dependendo da quantidade presente, pode causar desgaste, corrosão e deterioração do motor e do sistema de combustível. A análise de metanol presente em gasolina comercial (ou em AEAC) é descrita pela ASTM NBR 16041.

A amostra é injetada em um sistema de cromatografia a gás, sendo os componentes separados em coluna capilar de camada porosa e detectados por ionização de chama de hidrogênio. A identificação do metanol e etanol na amostra é realizada por comparação dos respectivos tempos de retenção com os de padrões analisados sob as mesmas condições cromatográficas e a quantificação por padronização externa (ASSOCIAÇÃO..., 2012).

### *3.1.5.16 Composição da Gasolina*

A composição da gasolina é que define suas características químicas e físico-químicas e faz com que seja um produto de elevada complexidade. A composição desse combustível tem uma grande importância no seu desempenho, tanto no que diz

respeito ao desempenho no motor, como na emissão de poluentes. A gasolina deve se apresentar adequada para atender aos requisitos ambientais definidos pela legislação e aos requisitos dos motores dos mais diversos tipos de veículos, isto sem agredir as peças e os componentes com que mantém contato. Quando produtos estranhos, com características solventes ou similares, são adicionados à gasolina, esta possui composição e propriedades alteradas. A mistura final se apresenta inadequada para consumo nos veículos, pois não atende aos requisitos técnicos mínimos exigidos para um bom funcionamento dos motores (PETROBRAS, 2000).

### 3.1.5.16.1 Teor de Aromáticos

Expressa o teor de compostos aromáticos presentes na gasolina. Tais compostos conferem à gasolina uma boa resistência à detonação apresentando, isoladamente, um maior número de octanagem MON e RON do que os demais componentes do combustível. Entretanto, deve-se atentar à toxicidade destes compostos. Entre as substâncias aromáticas presentes na gasolina, encontram-se benzeno, carcinogênico, tolueno e o xileno, ambos também tóxicos (PETROBRAS, 2000).

Os aromáticos possuem a tendência de gerar mais fumaça e depósitos de carbono, durante a queima do motor, do que o verificado para os compostos saturados e olefínicos. Também possuem capacidade de promover ataque aos componentes confeccionados de material plástico e borracha dos veículos, o que exige que estes materiais sejam especificados para que suportem o contato com a gasolina (PETROBRAS, 2000).

### 3.1.5.16.2 Teor de Olefinas

Indica a concentração de hidrocarbonetos insaturados (C=C). As olefinas são responsáveis pela instabilidade química da gasolina, pois apresentam a tendência de

reagir entre si e com outros hidrocarbonetos, na presença de oxigênio, luz ou calor. Estas reações geram polímeros, goma ou verniz, e alteram a cor do produto. A goma assim formada, conforme visto anteriormente, prejudica a qualidade da gasolina e pode provocar sérios danos, devido à formação de incrustações no sistema de combustível e na câmara de combustão dos motores. Além disso, altas concentrações de olefinas contribuem para maior nível de emissão de $NO_x$, poluentes indesejáveis devido aos problemas respiratórios que provocam nos seres humanos (PETROBRAS, 2000).

3.1.5.16.3 Teor de Hidrocarbonetos Saturados

Expressa o teor de hidrocarbonetos saturados, seja cadeia carbônica cíclica (parafinas e naftênicos), seja linear, presentes no produto. Estes compostos representam cerca de 40% da gasolina automotiva e fornece à mesma, dentre outras características, uma boa estabilidade química, fundamental para que não se deteriore com o tempo (PETROBRAS, 2000).

3.1.5.16.4 Teor de Chumbo

O chumbo tetaetila-CTE, foi, durante muitos anos, além de outros compostos deste elemento, incorporado à gasolina de vários países com o fim de melhorar a sua octanagem. Com o crescimento da preocupação com o meio ambiente, compostos de chumbo foram suprimidos da formulação da gasolina automotiva, principalmente por serem tóxicos para o ser humano, mas também por inviabilizar a adoção de catalisadores nos veículos, impedindo assim que este importante componente antipoluente tenha a eficiência necessária (AMERICAN..., 2012d).

O teor de chumbo reflete a concentração de substâncias, contendo este elemento, na gasolina. O teste deve ser feito quando houver suspeita da ocorrência de contaminação da gasolina por compostos de chumbo. A análise pode ser realizada por espectrofotometria de absorção atômica (AMERICAN..., 2012d).

*3.1.5.17 Densidade a 20/4ºC*

A densidade é uma característica da gasolina que pode ser relacionada ao seu potencial energético total, pois, quanto maior seu valor, maior a massa do combustível que estará sendo injetada no motor, para um mesmo volume considerado. Grandes variações na densidade levam a uma significativa variação na massa de combustível injetada, impossibilitando a obtenção de uma mistura ar/combustível balanceada (PETROBRAS, 2000).

Ao contrário da gasolina Padrão, os tipos A e C não possuem limites de densidade fixados por regulamentação. Apesar disso, o teste é normalmente realizado pelas refinarias para todas as gasolinas liberadas para venda. O valor da densidade é utilizado para converter o volume do produto, medido a uma temperatura qualquer, para o volume a 20ºC e que é considerado para efeito de faturamento (PETROBRAS, 2000).

O teste é feito imergindo-se um densímetro de vidro de 1000mL contendo amostra do produto, segundo ASTM D1298. Neste caso, o resultado é expresso como densidade 20/4ºC (PETROBRAS, 2000).

*3.1.5.18 Corrosividade ao Cobre*

Este teste dá uma indicação do potencial de corrosividade da gasolina às peças metálicas em geral, inclusive àquelas confeccionadas em cobre e suas ligas. O caráter corrosivo da gasolina é normalmente associado à presença de enxofre elementar e gás sulfídrico. O teste é feito imergindo uma lâmina de cobre devidamente preparada numa amostra do produto mantida a 50ºC por 3h. Decorrido este tempo, a lâmina é retirada, lavada e sua coloração é comparada com lâminas padrão da ASTM e, assim, definindo o grau de corrosividade da gasolina (AMERICAN..., 2012a).

### 3.1.6 Classificação Quanto a Qualidade

Por fim, é preciso ressaltar que a gasolina comercial brasileira é classificada quanto à conformidade e a adulteração. A conformidade se refere ao enquadramento com os parâmetros especificados pela Resolução nº 57 da ANP em 20/10/2011 e pelas Portarias nº 143 e 678 do MAPA em 27/06/2007 e 31/11/2011 respectivamente. Já a adulteração se remete a adição de solventes e de metanol na gasolina. Prática realizada com o propósito de obtenção de lucro, através da evasão fiscal e da concorrência desleal.

Desta forma, pode-se separar a gasolina em quatro grupos:

- *Conforme* e *não adulterada*: gasolina de boa qualidade para consumo, cujos parâmetros físico-químicos estão em acordo com o estabelecido pelas Portarias;
- *Não conforme* e *não adulterada*: suas características físico-químicas estão em desacordo com o especificado. Pode causar danos ao veículo e ao meio ambiente;
- *Conforme* e *adulterada*: possui os parâmetros físico-químicos dentro dos especificados, porém com adição irregular de solventes em sua composição;
- *Não conforme* e *adulterada*: possui solventes adicionados irregularmente, que modificam suas características originais. Pode causar danos ao veículo e ao meio ambiente.

Como forma de monitoramento da qualidade, a ANP instituiu o Programa de Marcação Compulsória em todo o território nacional. Este Programa estabelece a obrigatoriedade da adição de marcadores em Produtos de Marcação Compulsória (PMC) (AGÊNCIA..., 2013b).

PMC são solventes e eventuais derivados do petróleo indicados pela ANP. Já os marcadores são substâncias identificáveis, qualitativa e quantitativamente e, que depois de adicionadas aos PMC, resulte em concentração máxima de 1ppm e não interfira nas características físico-químicas e no grau de segurança para manuseio e uso dos PMC (AGÊNCIA..., 2013b).

Contudo, a tecnologia de análise de detecção do marcador é confidencial e de conhecimento exclusivo do Fornecedor de Marcador, da ANP e das instituições contratadas pela ANP.

## *3.2 Quimiometria*

A Química está enfrentando grandes mudanças. A evolução rápida dos computadores, tanto nos *softwares* como nos *hardwares*, trouxe novas possibilidades para os pesquisadores. Novos dados, parâmetros e análises, que antes não se podiam encontrar, hoje são facilmente obtidos. Esse novo volume de informações exige ferramentas matemáticas mais avançadas para tratá-lo (HASWELL, 1992).

O termo *quimiometria* foi introduzido por Svante Wold e Bruce R. Kowalski no começo dos anos 70 (WOLD, 1987). Desde então, tem sido desenvolvida e aplicada na Química, especialmente na química analítica. A Quimiometria é definida formalmente como uma disciplina química que emprega métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e para fornecer o máximo de informação química com a análise dos dados obtidos (FERREIRA, 1999).

### 3.2.1 Áreas de Atuação

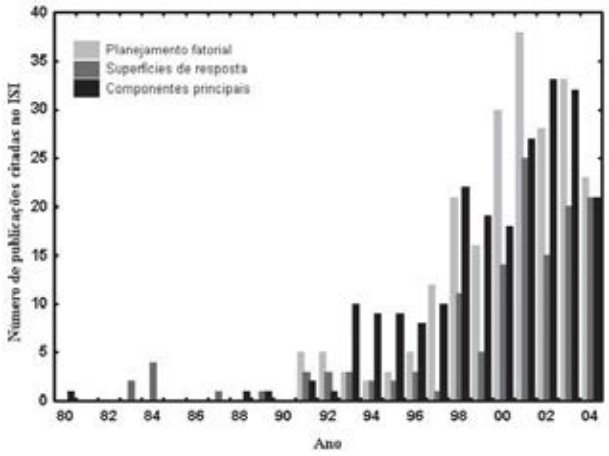
A Quimiometria tem diversas áreas de atuação. As três principais são: *planejamento de experimentos*, *reconhecimento de padrões* e *calibração multivariada*. Estas serão detalhadas a seguir.

### *3.2.1.1 Planejamento de Experimentos*

O uso de experimentos estatisticamente planejados é largamente utilizado na Química, Engenharia e Biotecnologia. Além disso, várias indústrias utilizam de métodos quimiométricos de planejamento, tais como Oxiteno, Pirelli, Braskem, Clariant, Nitroquímica, 3M do Brasil, Unilever, Petrobrás, Petroflex e Masterfoods do Brasil (NETO, 2006).

As técnicas de planejamento e otimização de experimentos têm sido usadas cada vez mais nos dias de hoje. A figura 3 mostra a evolução do número de trabalhos brasileiros publicados que utilizam desta ferramenta.

**Figura 3** - Publicações em três áreas da Quimiometria, entre 1980 e 2004, com a participação de brasileiros.




**Fonte**: NETO, B. B.; SCARMINIO, I. S.; BRUNS, R. E., 2006.

### *3.2.1.2 Reconhecimento de Padrões*

Os métodos de Reconhecimento de Padrões (RP) podem ser aplicados com diferentes finalidades, entre elas a análise exploratória de dados, a classificação de amostras e a resolução de curvas (FOO-TIM, 2004) (NETO, B. B., 2006).

Uma matriz de dados químicos possui certo número de objetos, cada um descrito por um número n de variáveis. Os objetos são amostras, compostos, cromatogramas, espectros, etc. as variáveis são propriedades dos objetos, associadas à composição destes. Em outras palavras, são as características que os descrevem, como concentração, absorbância, intensidade do pico cromatográfico, etc. A análise exploratória é utilizada para tentar detectar padrões de associação no conjunto de dados, a partir dos quais se podem estabelecer relações entre objetos e variáveis, agrupar objetos ou descobrir objetos anômalos, *outliers*. Os dois métodos mais empregados com tal fim é a HCA, e a Análise de Componentes Principais, PCA (FOO-TIM, 2004) (NETO, B. B., 2006). Ambos serão oportunamente apresentados.

Outro estudo típico do RP é a classificação de amostras de acordo com uma propriedade específica, baseada em medidas indiretamente relacionadas com a propriedade. Essa regra de classificação é desenvolvida para os objetos cujas variáveis são conhecidas. Os métodos de classificação que usam esse processo são conhecidos como métodos supervisionados de RP. Os mais utilizados em química são a Regra dos Vizinhos mais Próximos (KNN) e a SIMCA (NETO, B. B., 2006). Ambos serão oportunamente apresentados.

### *3.2.1.3 Calibração Multivariada*

A concentração dos analitos de interesse é comumente estimada através de valores físico-químicos do sistema obtidos experimentalmente. Desta forma, faz-se necessária a construção modelos de calibração, para determinar a relação entre propriedades medidas e concentrações das espécies estudadas. A calibração multivariada é uma das mais bem sucedidas combinações de métodos estatísticos com dados químicos, seja na química analítica, seja na química teórica. Os métodos mais usados incluem Regressões por Componentes Principais (PCR) (do acrônimo em inglês *Principal Component Regression*) e por Quadrados Mínimos Parciais (PLS) o Algoritmo Genético (GA) as Redes Neurais Artificiais (ANN) e os Quadrados Mínimos Parciais Não Lineares (N-PLS) (NETO, B. B., 2006).

### 3.2.2 Pré-tratamento

Antes de se realizar qualquer técnica quimiométrica, os dados precisam passar por uma etapa de pré-tratamento. Este pode ser dividido em *transformações*, aplicadas às amostras (linhas) de uma matriz de dados, representada por $\boldsymbol{X}$, e em *pré-processamentos*, aplicados às variáveis (colunas) da matriz de dados $\boldsymbol{X}$.

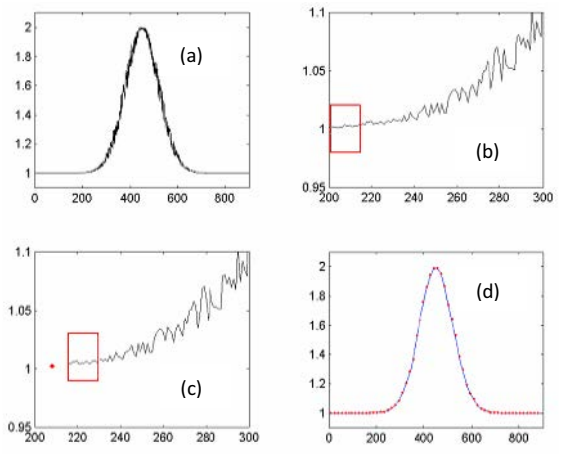
#### *3.2.2.1 Transformadas Matemáticas*

As transformações, ou transformadas, possuem como objetivo remover matematicamente fontes de variação indesejáveis que poderão gerar erros durante a análise dos dados. Tais variações podem ser *aleatórias*, como no caso de ruído experimental, ou *sistemáticas*, como no caso de desvios da linha base.

A remoção de variações aleatórias é feita por meio de técnicas de alisamento (*smoothing*), fazendo com que a relação sinal/ruído aumente. Dentre as técnicas, as mais utilizadas são *alisamento pela média* e *alisamento pela média móvel* (FERREIRA, 2013).

O alisamento pela média é empregado, quando se deseja diminuir o número de variáveis $j$. Primeiramente, selecionam-se uma janela de abertura com $n+1$ variáveis, sendo $n$ um número inteiro e par. Calcula-se a média das respostas, que passará a ser a primeira variável da curva alisada, com valor de abscissa igual ao centro da janela, ou seja $(n/2)+1$. O processo é repetido para $n+2, n+3, \ldots 2n+1$. O resultado final é uma curva com um número de variáveis igual ao inteiro mais próximo de $J/(n+1)$. A figura 4 ilustra o processo de alisamento pela média.

**Figura 4** - Exemplo de alisamento pela média de um espectro simples: (a) espectro bruto, (b) início do alisamento em $n + 1$, (c) alisamento em $n + 2$, (d) espectro alisado.

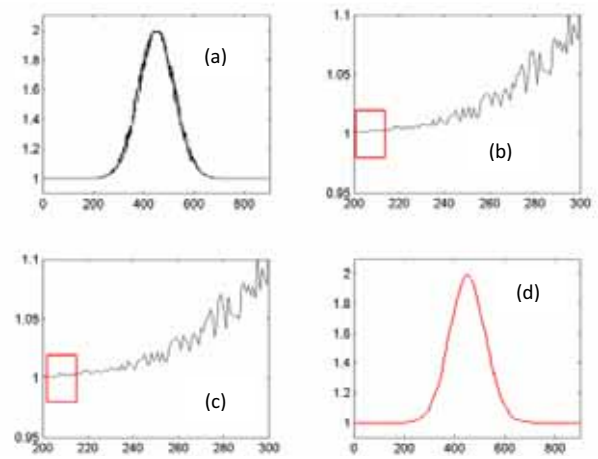



**Fonte**: FERREIRA, 2013.

O alisamento pela média móvel, figura 5, é semelhante ao alisamento pela média, exceto que a janela move de elemento em elemento ao invés de janela em janela. O espectro alisado contém o mesmo número de variáveis que o original.

Já a remoção de variações sistemáticas é feita por correções da linha base. Dentre as técnicas, as mais usadas são *primeira derivada*, *segunda derivada*, *logaritmo na base 10*, *correção multiplicativa de sinal* e *variação de padrão normal* (FERREIRA, 2013).

**Figura 5** - Exemplo de alisamento pela média de um espectro simples: (a) espectro bruto, (b) início do alisamento com janela móvel, (c) primeira janela alisada, (d) espectro alisado.




**Fonte**: FERREIRA, 2013.

A primeira derivada é utilizada é utilizada para correção do sinal obtido, se este estiver deslocado de uma quantidade constante (FERREIRA, 2013). O algoritmo mais utilizado para tal fim é o de Savitsky-Golay:

$$2\delta \frac{dy}{dx}(x_j) \cong \Delta y(x_j) = y_{x_j+\delta} - y_{x_j-\delta} \quad (01)$$

Em que $y$ é um representa a intensidade do sinal, em função de $x_j$, e $\delta$ é a quantidade constante em que $y$ está deslocado.

Já a segunda derivada é empregada em casos que apresentam problemas de inclinação de linha de base (FERREIRA, 2013). Assim, o algoritmo é representado por:

$$(2\delta)^2 \frac{d^2y}{dx^2}(x_j) \cong \Delta^2 y(x_j) = y_{x_j+\delta} + y_{x_j-\delta} - 2y_{x_j} \quad (02)$$

O logaritmo na base 10 é aplicado com objetivo de reduzir picos de alta intensidade, a fim de enfatizar os de intensidades baixas, importantes para análise. Também é usado na linearização dos dados, sem afetar sua interpretação. Espectros de transmitância, ou refletância, por exemplo, não são lineares com a concentração, havendo a necessidade de empregar o logaritmo:

$$T_\lambda = 10^{-\alpha_\lambda lc} \quad (03)$$

$$A_\lambda = -\log T_\lambda = -\log I/I_\mathrm{o} \quad (04)$$

Em que $T_\lambda$ é a transmitância observada para certo analito em um dado comprimento de onda $\lambda$, $\alpha_\lambda$ é a absortividade característica do analito em determinado $\lambda$, $l$ é o caminho óptico, $c$ é a concentração de analito, é a absorção observada para o analito em $\lambda$, $I_\mathrm{o}$ e $I$ são as intensidades de radiação incidente transmitida respectivamente.

A Correção Multiplicativa de Sinal (MSC) (do acrônimo em inglês *Multiplicative Scatter Correction*), é um método utilizado quando ocorrem variações multiplicativas nos sinais obtidos. É largamente empregado em leituras de reflectância difusa, pois corrige o efeito da dispersão da luz, causada pela falta de homogeneidade das amostras. Calcula-se um espectro médio referente a todos os espectros presentes no conjunto de calibração (FERREIRA, 2013). Posteriormente, cada espectro é recalculado por:

$$X_{ij(MSC)} = h + gX_{ij} \quad (05)$$

Em que $X_i$ é o espectro da amostra $i$, $X_i'$ é o espectro transformado ,$h$ e $g$ são os coeficientes escolhidos de modo que a diferença entre $X_i'$ e o espectro médio seja mínima (ALMEIDA, 2009).

Por último, um dos algoritmos de normalização mais utilizados é o Método da Variação de Padrão Normal (SNV) (do acrônimo em inglês *Standard Normal Variate Method*), que é empregado para minimizar interferências causadas pelo tamanho de partículas e diferenças de densidades das amostras.

Este método tem como princípio centrar cada sinal em torno de zero, por subtração da média (ajuste aditivo) e divisão pelo desvio padrão (ajuste multiplicativo) em todos os pontos do espectro. É bastante semelhante ao MSC, embora a determinação dos ajustes seja realizada de forma diferente (ALMEIDA, 2009).

Assim sendo, a equação 06 representa o cálculo do SNV.

$$X_{ij(SNV)} = \frac{X_i - m_i}{s_i} \quad (06)$$

Em que $X_{ij(SNV)}$ é o espectro transformado da amostra $i$, $X_i$ é o espectro da amostra $i$, $m_i$ é a média das leituras espectrais para a amostra $i$ e $s_i$ é o desvio padrão destas leituras.

*3.2.2.2 Pré-processamentos*

Os pré-processamentos servem para adequar as amostras do conjunto, visando maximizar ou minimizar o efeito de certas variáveis, ou mesmo tratá-las quando possuem informações correlacionadas. Podem ser feitos de quatro maneiras principais: *centrando-os na média*, *escalando-os pela variância*, *auto-escalando-os*, ou seja, centralizando-os na média com posterior escalação pela variância, ou *escalando-os pela amplitude* (PARREIRA, T. F., 2003).

A centralização pela média é convenientemente usada em casos nos quais todas as variáveis forem medidas numa mesma unidade, com uma mesma magnitude.

Permitindo assim, que a presença de ruídos não afete negativamente a análise. Neste pré-processamento, o centroide da matriz de dados $\boldsymbol{X}$ é levado à origem pela subtração de cada elemento de cada coluna pela média da respectiva coluna. A relação é dada pela equação 8.

$$x_{ij\,(CM)} = x_{ij} - \bar{x}_j \tag{07}$$

Em que $x_{ij\,(CM)}$ é o valor centrado na média, $x_{ij}$ é o valor da variável na linha $i$ e coluna $j$ e $\bar{x}_j$ é a média dos valores das amostras na coluna $j$. Este tipo de pré-tratamento é muito utilizado em análises espectroscópicas.

Já no escalamento pela variância, cada elemento de uma dada variável é dividido pelo desvio padrão dessa variável. Isto conduz todos os eixos da coordenada a terem o mesmo comprimento, dando a cada variável a mesma influência no modelo. Assim:

$$x_{ij(Var)} = \frac{x_{ij}}{s_j} \tag{08}$$

em que $x_{ij(Var)}$ é o valor escalado pela variância e $s_j$ é o desvio padrão dos valores da variável $j$.

No escalamento pela amplitude, ou escalamento pelo *range*, de cada elemento é subtraído o elemento de menor valor na matriz de dados e, em seguida, divide-se pela variação dos valores do mesmo banco de dados. Isto faz com que todos os elementos possuam valores entre 0 e 1. Matematicamente:

$$x_{ij(Ran)} = \frac{x_{ij} - x_{j(\min)}}{x_{j(\max)} - x_{j(\min)}} \tag{09}$$

em que $x_{ij(Ran)}$ é o valor escalado pela amplitude, $x_{j(\min)} = \min_{1<i<I}\left|x_{ij}\right|$, ou seja o elemento de menor valor para a variável $j$ e $x_{j(\max)} = \max_{1<i<I}\left|x_{ij}\right|$, ou seja o elemento de menor valor para a variável $j$.

Por fim, o auto-escalamento é feito pela centralização dos dados na média com posterior escalamento pela variância. As variáveis terão média zero e desvio padrão igual a 1. A relação é descrita pela equação 10.

$$x_{ij(AS)} = \frac{\bar{x}_j}{s_j} \tag{10}$$

Em que $x_{ij(AS)}$ é o valor auto-escalado da variável $j$ para a amostra $i$.

Tanto o modelo de escalamento pela variância, quanto o auto-escalamento são utilizados com a finalidade de dar a mesma importância (ou peso) a todas as variáveis medidas, posto que a PCA, por ser um método de quadrados mínimos, faz com que variáveis com alta variância possuam altos pesos.

Assim, com o pré-processamento escolhido, pode-se iniciar a análise exploratória de interesse.

### 3.2.3 Análise Hierárquica de Agrupamentos

As técnicas de agrupamento são utilizadas com a finalidade de relacionar os objetos de um conjunto de dados, no qual não se conhece nenhuma caracterização a primeira vista. O agrupamento pode ser feito de duas formas: *aglomerativa* ou *divisiva*. Na primeira, cada amostra é considerada, no princípio, um grupo e, por suas semelhanças, vão sendo agrupadas em subgrupos, até que todas formem um único grupo. Já na técnica divisiva, todas as amostras constituem um único grupo, o qual será dividido em subgrupos, também por suas semelhanças, até que cada amostra forme um único grupo (PARREIRA, 2003).

A Análise Hierárquica de Agrupamentos (HCA) (do acrônimo em inglês *Hierarchical Cluster Analysis*) é uma técnica aglomerativa não supervisionada que examina as distâncias interpontuais entre todas as amostras do conjunto de dados e representa essa informação na forma de um gráfico bidimensional, um *dendograma*. Por meio deste, é possível visualizar a similaridade e os agrupamentos entre as amostras ou variáveis (PARREIRA, 2003).

A construção dos dendogramas é realizada com base na proximidade entre as amostras no espaço. Ou seja, o agrupamento é feito através do cálculo das distâncias entre todas as amostras, gerando, assim, uma matriz de similaridade cujos elementos, também chamados de índices de similaridade, que variam entre 0 e 1. Um valor alto mostra uma pequena distância entre dois agrupamentos e, por consequência, uma alta similaridade. Conforme o processo continua, grupos similares vão sendo unidos até que

se forme um único agrupamento (PARREIRA, 2003). A similaridade é descrita na equação 11.

$$s_{ij} = 1 - \frac{d_{ij}}{d_{max}} \quad (11)$$

Em que $s_{ij}$ é a similaridade entre duas amostras, $d_{ij}$ é a distância entre as mesmas e $d_{max}$ é a maior distância encontrada entre todas as amostras do conjunto.

A escolha da proximidade entre dois agrupamentos pode ser feita por três métodos: pelo *vizinho mais próximo*, pelo *vizinho mais distante* ou *centróide* (DONI, 2004).

A conexão pelo vizinho mais próximo é feita inicialmente encontrando a menor distância entre dois indivíduos, ou um indivíduo e um grupo, e conectando-os. Em seguida, procura-se a segunda menor distância, unindo-os e, assim por diante até que um único agrupamento seja formado (DONI, 2004).

Já na conexão centróide, as amostras ligam-se aos agrupamentos cujos centros estiverem mais próximos. Existem vários métodos de conexão centróide, sendo que a diferença encontra-se na maneira como é calculado (DONI, 2004). A conexão incremental, por exemplo, é uma conexão centroide, cujo cálculo das distâncias entre grupos é feito por uma soma de quadrados.

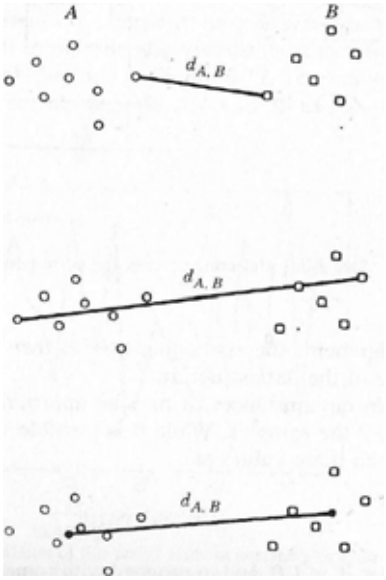
Por fim, pelo método do vizinho mais distante, busca-se sempre a maior distância entre as amostras e, dentre estas, o par de similaridade é agrupado. A figura 6 ilustra esses diferentes métodos de conexão.

**Figura 6** - Esquema dos diferentes métodos de conexão entre grupos de amostras.

Vizinho mais próximo

Vizinho mais distante

Conexão centróide




**Fonte**: SHARAF, M. A.; ILLMAN, D. L.; KOWALSKI, B. R., 1986.

As distâncias podem são classificadas em: *distância euclidiana*, *distância de* Manhattan, *distância de* Chebychev.

A distância euclidiana, equação 12, é a distância geométrica no espaço multidimensional. A distância euclidiana entre dois vetores $A = [A_1, A_2, \dots, A_n]$ e $B = [B_1, B_2, \dots, B_n]$ é definida como:

$$d_{AB(DE)} = \sqrt{(A_1 - B_1)^2 + (A_2 - B_2)^2 + \cdots + (A_n - B_n)^2} = \sqrt{\sum_{i=1}^{n}(A_i - B_i)^2} \quad (12)$$

Em que $d_{AB(DE)}$ é a distância euclidiana entre os elementos $A$ e $B$.

Quanto à distância *Manhattan*, equação 13, seus resultados, em muitos casos, são similares ao da distância euclidiana. Contudo, a grande diferença consiste na minimização das diferenças entre os elementos, uma vez que os elementos não são elevados ao quadrado.

$$d_{AB(DM)} = |A_1 - B_1| + |A_2 - B_2| + \cdots + |A_n - B_n| = \sum_{i=1}^{n}|A_i - B_i| \quad (13)$$

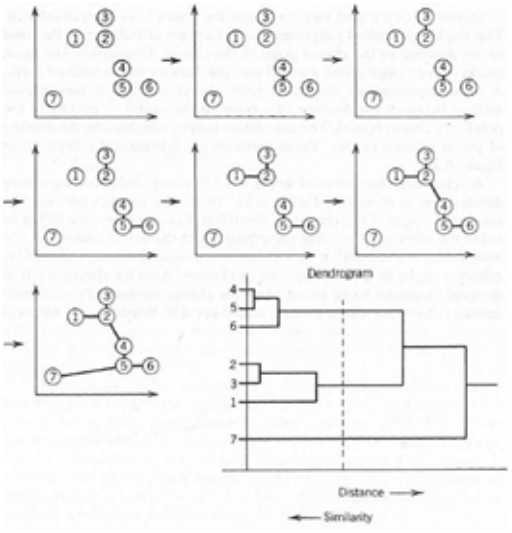
Em que $d_{AB(DM)}$ é a distância *Manhattan* entre os elementos $A$ e $B$.

Por último, a distância de *Chebychev*, equação 14, é apropriada em casos nos quais se deseja definir dois elementos como diferentes, se apenas umas das dimensões é diferente. É definida por:

$$d_{AB(DC)} = máximo(|A_1 - B_1| + |A_2 - B_2| + \cdots + |A_n - B_n|) = máximo_i|A_i - B_i| \quad (14)$$

Assim, com o método de conexão e o cálculo de medida definidos, pode-se construir um dendograma para um determinado conjunto de amostras. Uma das grandes vantagens da HCA é o aspecto visual fornecido pelo método, possibilitando a fácil percepção dos agrupamentos de amostras distinção, bem como o grau de similaridade entre si (DONI, 2004).

**Figura 7** - Esquema da construção de um dendograma simples.




**Fonte**: SHARAF, M. A.; ILLMAN, D. L.; KOWALSKI, B. R., 1986.

### 3.2.4 Análise de Componentes Principais

A PCA tem como princípio a redução da dimensionalidade original da matriz de dados $\boldsymbol{X}$, através da construção de novas variáveis, denominadas componentes principais, as quais descrevem o mesmo conjunto.

As componentes principais são calculadas, a fim de reterem a maior quantidade de variância presente nas variáveis originais. Podendo, assim, descrever os dados iniciais através de um menor número de variáveis.

*3.2.4.1 Tratamento Matemático*

Antes de se realizar o tratamento matemático da PCA, faz-se necessária a revisão de alguns conceitos. Em estatística, algumas das análises que podem ser feitas em um conjunto de dados são a *média aritmética*, $\bar{a}$ (equação 15), a *variância*, $s^2$ (equação 16), e o *desvio padrão*, $s$. Para um conjunto A de amostras tem-se:

$$\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{15}$$

$$s_a{}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(a_i - \bar{a})^2 \tag{16}$$

Sendo que $a_i$ indica a amostra na posição $i$ no conjunto de dados A e $n$ representa o número total de amostras. Contudo, ao se introduzir um novo conjunto de dados B, que se relaciona com o A através do tempo, faz-se necessária a introdução do conceito de *covariância*.

Para as séries A e B, a covariância fornece uma medida não padronizada do grau de relação e é estimada tomando o produto dos desvios da média para cada variável em cada período. A covariância está representada na equação 17.

$$s_{ab} = \sum_{i=1}^{n}(a_i - \bar{a})\left(b_i - \bar{b}\right) \tag{17}$$

Se houver mais de duas séries, é necessária a construção de uma *matriz de covariância*. Assim, para as séries A, B e C:

$$\boldsymbol{S_{ABC}} = \begin{pmatrix} s_{aa} & s_{ab} & s_{ac} \\ s_{ba} & s_{bb} & s_{bc} \\ s_{ca} & s_{cb} & s_{cc} \end{pmatrix} \tag{18}$$

em que a diagonal principal da matriz $\boldsymbol{S_{ABC}}$ contém as variâncias e as demais posições contém a correlação entre as direções.

Para amostras de vetores, o *vetor médio* é expresso por:

$$\bar{\mathbf{a}} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{a_i} \tag{19}$$

Semelhante à média aritmética $\bar{a}$. A matriz de covariância para $n$ amostras de vetores, com vetor médio $\bar{\mathbf{a}}$, pode ser calculado por:

$$\boldsymbol{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{a}\mathbf{a}^t - \bar{\mathbf{a}}\bar{\mathbf{a}}^t) \quad (20)$$

Tal matriz $\boldsymbol{S}$ será de suma importância no cálculo da PCA.

Com estes conceitos consolidados, é possível definir *autovetores* e *autovalores*. Define-se um autovetor υ de uma matriz quadrada **M** se a multiplicação **M**.υ resulta num multiplo λ de υ. Ou seja:

$$\boldsymbol{M}.\upsilon = \lambda\upsilon \quad (21)$$

Assim, λ é denominado, então, autovalor de **M** associado ao autovetor υ. Uma das propriedades dos autovetores é a sua ortogonalidade entre si. Isto é importante, pois torna possível expressar os dados em termos dos autovetores, ao invés dos eixos x, y e z.

Para matrizes quadradas de ordem 2, ou 3, os autovalores são representados por:

$$\det(\boldsymbol{M} - \lambda.\mathbf{I}) = 0 \quad (22)$$

Em que **I** é a matriz identidade, **M** a matriz de dados e os escalares não nulos, λ, que a solucionam serão os autovalores. No caso de dimensões maiores, aplica-se um algorítimo numérico interativo.

Da mesma forma, os autovetores associados aos autovalores serão vetores não-nulos no espaço solução da equação:

$$(\lambda.\mathbf{I} - \mathbf{M})\upsilon = 0 \quad (23)$$

Esse espaço é denominado *auto-espaço* de **M** associado a λ. As bases para cada um destes autovetores são chamadas *bases de auto-espaço*.

Considerando uma matriz quadrada $n \times n$, com $n$ autovalores linearmente independentes, então tal matriz é diagonalizável, ou seja, possui $n$ autovalores linearmente independentes, os quais serão os elementos da diagonal principal.

Para diagonalizar uma matriz **M**, deve-se encontrar seus autovalores linearmente independentes, $\upsilon_1, \upsilon_2, \ldots, \upsilon_n$, formar uma matriz $\Upsilon$ com estes vetores como

coluna. O produto $\Upsilon^{-1}\boldsymbol{M}\Upsilon$ será uma matriz diagonal, com elementos iguais aos autovalores na diagonal principal.

O último conceito a ser introduzido, necessário para o calculo da PCA, é a *transformada de Hotelling*. Lembrando que a matriz de covariância **S** é real e simétrica, é sempre possível encontrar um conjunto de valores de $n$ autovetores ortonormais, os quais são arranjados de modo decrescente, de acordo com os valores dos $n$ autovalores. Em outras palavras, o auto vetor $\boldsymbol{e}_1$ corresponde ao maior autovalor $\lambda_1$, o auto vetor $\boldsymbol{e}_2$ corresponde ao segundo maior autovalor $\lambda_2$ e assim por diante até o autovetor de menor valor $\boldsymbol{e}_{n-1}$ e seu correspondente autovalor $\lambda_{n-1}$.

Tomando uma matriz $\boldsymbol{A}$, cujas colunas são os autovetores de **S** ordenados como descrito anteriormente. A transformação definida por esta matriz será:

$$\boldsymbol{y}_H = \boldsymbol{A}(x - m_x) \tag{24}$$

Esta equação é denominada Transformada de Hotelling e irá mapear os valores $x$, em valores $y$, cuja média será zero, isto é $m_x = 0$, e cuja matriz de covariância dos valores $y$ pode ser obtida por:

$$\boldsymbol{S}_y = \boldsymbol{A}\boldsymbol{S}_x\boldsymbol{A}' \tag{25}$$

A matriz $\boldsymbol{S}_y$, equação 26, é diagonal e possui elementos ao longo da diagonal principal que são os autovalores de $\boldsymbol{S}_x$.

$$\boldsymbol{S}_y = \begin{bmatrix} \lambda_1 & 0 & 0 & & \cdots & \\ 0 & \lambda_2 & 0 & & \cdots & \\ 0 & 0 & \lambda_3 & & \cdots & \\ & & & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & 0 & \lambda_{n-2} & 0 \\ & & & 0 & 0 & \lambda_{n-1} \end{bmatrix} \tag{26}$$

Como os elementos fora da diagonal principal de $\boldsymbol{S}_y$ são nulos, os elementos dos vetores $\boldsymbol{y}$ são descorrelacionados. Já que os elementos da diagonal de uma matriz diagonal são seus autovalores, $\boldsymbol{S}_x$ e $\boldsymbol{S}_y$ possuem os mesmos autovalores e autovetores.

O efeito do uso da Transformada de Hotelling é o estabelecimento de um novo sistema de coordenadas, cuja origem é o centroide do conjunto de pontos, ou seja o vetor de média, e cujos eixos estarão na direção dos autovetores de $\mathbf{S}_x$. Assim, obtém-

se um alinhamento de autovetores pela rotação do sistema de eixos. Isto é o mecanismo que descorrelaciona os dados da matriz $\boldsymbol{X}$ a ser analisada pela PCA. Por sua vez, cada autovalor indica a variância do componente $\boldsymbol{y}_i$ ao longo do autovetor $\upsilon_i$.

A PCA consiste em promover uma transformação linear nos dados, de modo que os dados resultantes tenham suas componentes mais relevantes nas primeiras dimensões, em eixos denominados Componentes Principais (PC) (WOLD, 1987).

Por fim, a representação matemática da PCA dá-se da seguinte forma:

$$\boldsymbol{X} = \boldsymbol{T\Theta V'} = \boldsymbol{TP'} = \sum_{i=1}^{A} \boldsymbol{t}_i \boldsymbol{p}_i' \tag{27}$$

Em que $T\Theta V'$ representam os valores únicos de decomposição da matriz de dados $X$ e os termos $t_i$ e $v_i$ (para i=1, 2, ..., A) são vetores denominados *score* e *loading*, respectivamente, e são ortogonais entre si. A matriz diagonal $\Theta$ contém os autovalores da matriz de covariância $\boldsymbol{X'X}$, distribuídos em cada eixo ortogonal da componente principal (WOLD, 1987).

Durante qualquer análise, é comum a obtenção de dados com medidas de ruído, os quais não são analisados pela análise de fatores. Assim, podem-se expressar dados com ruídos na forma:

$$\boldsymbol{X} + \boldsymbol{E} = \boldsymbol{CS'} + \boldsymbol{E} = \boldsymbol{T\Theta V'} + \boldsymbol{E'} = \boldsymbol{TP'} + \boldsymbol{E'} \tag{28}$$

Ou pela forma matricial:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & x_{MN} \end{bmatrix} + \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1N} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & e_{MN} \end{bmatrix} =$$

$$\begin{bmatrix} t_{11} & \cdots & t_{1A} \\ t_{21} & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ t_{M1} & \cdots & t_{MA} \end{bmatrix} . \begin{bmatrix} p_{11} & \cdots & \cdots & p_{1N} \\ \cdots & \cdots & \cdots & \cdots \\ p_{M1} & \cdots & \cdots & p_{MN} \end{bmatrix} +$$

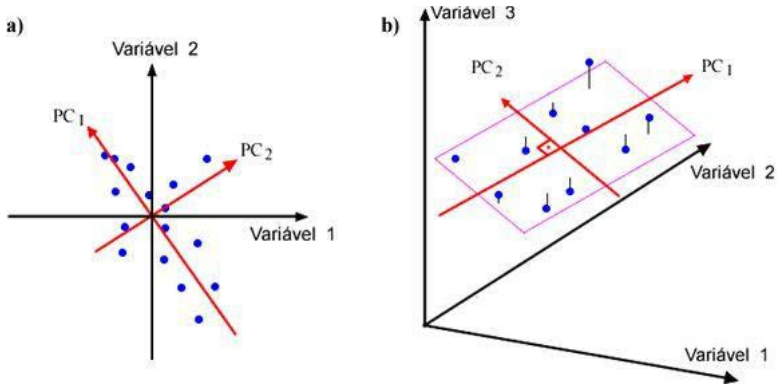
$$\begin{bmatrix} t_{1,A+1} & \cdots & t_{1A} \\ t_{2,A+1} & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ t_{M,A+1} & \cdots & t_{MN} \end{bmatrix} . \begin{bmatrix} p_{A+1,1} & \cdots & \cdots & p_{A+1,N} \\ \cdots & \cdots & \cdots & \cdots \\ p_{M1} & \cdots & \cdots & p_{MN} \end{bmatrix} =$$

$$\begin{bmatrix} t_{11} & \cdots & t_{1A} \\ t_{21} & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ t_{M1} & \cdots & t_{MA} \end{bmatrix} . \begin{bmatrix} p_{11} & \cdots & \cdots & p_{1N} \\ \cdots & \cdots & \cdots & \cdots \\ p_{M1} & \cdots & \cdots & p_{MN} \end{bmatrix} + \begin{bmatrix} e'_{11} & \cdots & \cdots & e'_{1N} \\ e'_{21} & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ e'_{M1} & \cdots & \cdots & e'_{MN} \end{bmatrix} \tag{29}$$

Em que a matriz $\boldsymbol{E}$ é a medida de ruídos e possui o mesmo tamanho da matriz de dados $\boldsymbol{X}$ e $\boldsymbol{E'}$ é a matriz de erros. Juntos são chamados de *modelo residual*. Se a medida de ruído for ruído branco, ou seja possui variância $s^2$ constante, então $E$ e $E'$ são aproximadamente iguais e ortogonais entre si, entre os valores $c_i$ e $s_i$ e $u_i$ e $p_i$ (FOO-TIM, 2004).

As matrizes $\boldsymbol{T}$ e $\boldsymbol{P'}$, ortogonais entre si, são chamadas matrizes de ***scores*** e ***loadings*** respectivamente. A matriz de *scores* contém os dados originais num sistema de coordenadas rotacionado e a matriz de *loadings* comporta os pesos de cada variável original, após o cálculo das PC. As matrizes de *scores* e *loadings* podem ser representadas graficamente em um *score plot* e um *loading plot* respectivamente (WOLD, 1987).

**Figura 8** - Representação esquemática de um conjunto de dados (a) definidos por duas componentes principais no plano e (b) definidos por três componentes principais no espaço.




**Fonte**: FSPANERO, 2013.

### 3.2.5 Calibração Multivariada

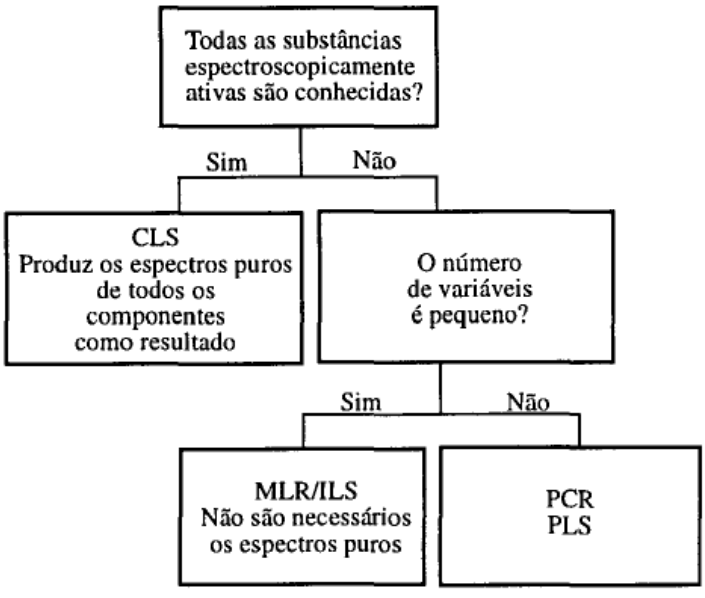
Calibração é o procedimento para encontrar um algoritmo matemático que produza propriedades de interesse a partir dos resultados instrumentais registrados. Uma vez encontrado, este algoritmo poderá ser usado para prever a concentração do componente químico de interesse em amostras de composição desconhecida, usando a resposta instrumental da mesma (FERREIRA, 1999).

As medidas instrumentais de cada amostra são organizadas em uma matriz de variáveis independentes $\boldsymbol{X}$, em que cada linha representa uma amostra e contém as respostas medidas para a mesma. O outro conjunto de dados é constituído das variáveis dependentes e organizado na matriz $\boldsymbol{Y}$, caso haja mais de uma variável (mais de um analito de interesse), ou pelo vetor $\boldsymbol{y}$, no caso de uma única variável. O total de elementos deste vetor é igual ao número de amostras e corresponde a propriedade de interesse que se espera prever no futuro (FERREIRA, 1999).

A calibração pode ser dividida em duas etapas: *modelagem*, em que se estabelece uma relação matemática entre $\boldsymbol{X}$ e $\boldsymbol{Y}$ no conjunto de calibração; *validação*, que otimiza a relação no sentido de uma melhor descrição do analito de interesse (FERREIRA, 1999).

Os métodos de regressão multivariada são classificados em dois tipos principais: direto, Quadrados Mínimos Clássicos (CLS) (do acrônimo em inglês *Classical Least Squares*) e inverso e Quadrados Mínimos Inversos (ILS) (do acrônimo em inglês *Inverse least Squares*). A figura a seguir mostra um esquema para a escolha do método de calibração a ser utilizado em determinada situação.

**Figura 9** - Fluxograma para escolha do método de calibração apropriado.




**Fonte**: FERREIRA, 1999

Construído o modelo, este deve ser validado, ou seja, testado para garantir que os valores obtidos para a variável dependente sejam iguais ou próximos dos

experimentais. A validação pode ser feita de duas formas: utilizando um *conjunto externo* ou por meio da *Validação Cruzada*.

No primeiro caso, o conjunto de dados é dividido em dois subconjuntos: um de calibração, usado para a construção do modelo, e outro de validação. Os resultados obtidos para a variável dependente nesta etapa são comparados com os valores experimentais, calculando os resíduos. Quanto menos resíduos encontrados, maior a eficácia do modelo.

Já na Validação Cruzada, a matriz de dados é dividida em pequenos grupos, sendo que um deles é removido da matriz original. Após isso, a matriz reduzida é decomposta, novamente, em *scores* e *loadings*, e o modelo de calibração é estabelecido sobre esse novo conjunto. A partir disso, a propriedade das amostras removidas será prevista e os resultados obtidos entre os valores serão estimados e calculados. Este processo é feito para todos os grupos pequenos do conjunto de dados.

A avaliação da eficácia dos modelos construídos é feita pela Predição do Erro Residual da Soma de Quadrados (PRESS) (do acrônimo inglês *Prediction Residual Error Sum of Squares*), equação 30, ou pelo Erro Padrão da Validação Cruzada/ Predição do Erro Padrão (SECV/SEP) (do acrônimo inglês *Standard Error of Cross Validation*/*Prediction*), equações 31 e 32. A diferença entre ambos consiste no fato do SECV levar em conta o número de componentes principais utilizados.

$$PRESS = \sum_i (\hat{y}_i - y_i)^2 \tag{30}$$

$$SECV = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{(n-k-1)}} \tag{31}$$

$$SEP = \sqrt{\frac{\sum_i (\hat{y}_i - y_i)^2}{n}} \tag{32}$$

Em que $\hat{y}_i$ é o valor previsto para a amostra $i$ utilizando o modelo; $y_i$ é o valor médido para a amostra $i$; $k$ é o número de componentes principais utilizadas; $n$ é o número de amostras do conjunto de calibração.

Após a validação, já é possível utilizar o modelo construído na previsão da propriedade modelada em amostras em que este valor é desconhecido (PARREIRA, 2003).

*3.2.5.1 Calibração por Quadrados Mínimos*

Este tipo de calibração é utilizado em casos, nos quais os componentes são conhecidos e suas respectivas curvas de respostas podem ser obtidas separadamente. A grande restrição desse método é essa exigência do conhecimento da concentração de cada espécie presente no sistema, o que normalmente é inviável para um grande número de amostras, ou uma matriz complexa (PARREIRA, 2003).

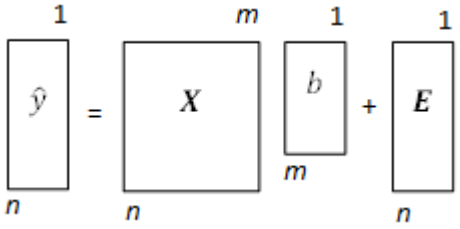
*3.2.5.2 Calibração por Quadrados Mínimos Inversos*

Os métodos de calibração inversa podem ser divididos em três grupos: Regressão Linear Múltipla (MLR) (do acrônimo em inglês *Multiple Linear Regression*), PCR, e Regressão por PLS (do acrônimo em inglês *Partial Least Squares*) (PARREIRA, 2003).

Na MLR, são medidas para $m$ variáveis $x_j$, para $j = 1 - m$ e para $y$, com a finalidade de encontrar uma relação linear, ou de primeira ordem, entre si. Isso pode ser representado matematicamente por:

$$\hat{\boldsymbol{y}} = \boldsymbol{Xb} + \boldsymbol{E} \tag{33}$$

Ou graficamente por:




(34)

Na equação 33, $\hat{\boldsymbol{y}}$ é o vetor das concentrações, $\boldsymbol{X}$ é a matriz das variáveis medidas, com $n$ amostras e $m$ variáveis, e $\boldsymbol{b}$ é o vetor que contém os coeficientes do modelo. Esta equação é usada para modelar a relação entre múltiplos analitos de interesse e a mesma matriz de resposta usando diferentes coeficientes do modelo. Com isso, é possível distinguir entre três casos:

- $m > n$. Há mais variáveis do que amostras. Neste caso, há infinitas soluções para $\boldsymbol{b}$, o que não é interessante.
- $m = n$. Esta situação pode não ser encontrada comumente. Contudo, dá uma solução única para $\boldsymbol{b}$, permitindo a equação:

$$e = \hat{\boldsymbol{y}} - \boldsymbol{Xb} = 0 \quad (35)$$

Em que $e$ é chamado de vetor residual. Neste caso, um vetor de zeros: 0.

- $m < n$. Há mais amostras que variáveis, não permitindo uma solução exata para $\widehat{\boldsymbol{b}}$. Porém, pode-se obter uma solução minimizando o tamanho do vetor residual $e$ na equação:

$$e = \hat{\boldsymbol{y}} - \boldsymbol{Xb} \quad (36)$$

O método mais popular para a determinação de $\boldsymbol{b}$ é feita pelo Método dos Quadrados Mínimos:

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{y}} \quad (37)$$

$(\boldsymbol{X}'\boldsymbol{X})$ é uma matriz quadrada, com número de linhas e colunas iguais ao número de variáveis medidas. A princípio, um número de amostras no conjunto de calibração maior ou igual ao número de variáveis é necessário para essa inversão. Contudo, em sistemas analíticos, o número de variáveis supera o número de amostras, ou surge um grande número de variáveis altamente correlacionadas, impossibilitando a inversão da matriz (PARREIRA, 2003).

*3.2.5.3 Regressão Linear Múltipla*

Ao se utilizar MLR, uma possível solução seria a redução do número de variáveis por meio de um método de seleção, tornando a matriz não singular (matriz que possa ser invertida). Entretanto, se o conjunto for muito grande, será necessário grande esforço. Recomenda-se que seja empregado apenas quando o número de variáveis for pequeno, ou quando apenas um subconjunto das variáveis medidas é desejado.

Ao contrário deste, os métodos PCR e PLS não requerem a seleção de variáveis, posto que estas são transformadas em componentes principais. Ainda assim, a vantagem do MLR está em sua simplicidade, exceto no caso de uma seleção errônea de variáveis (PARREIRA, 2003).

*3.2.5.4 Regressão por Componentes Principais*

O fundamento por trás destes métodos consiste em encontrar combinações lineares relevantes a partir dos valores das variáveis originais e, então, usá-las na equação de regressão. Assim, toda informação irrelevante será descartada da modelagem, resolvendo, desta forma, o problema da colinearidade e obtendo uma equação de regressão mais objetiva. (PARREIRA, 2003)

Nestes modelos, esta redundância é eliminada pela construção de uma nova matriz $\boldsymbol{P}$, com colunas formadas por combinações lineares das originais em $\boldsymbol{X}$. Essa nova matriz torna $\boldsymbol{P'P}$ inversível, possibilitando encontrar o vetor $\widehat{\boldsymbol{b}}$ (PARREIRA, 2003).

A diferença entre o PCR e o PLS está na forma de construção da matriz $\boldsymbol{P}$. Na PCR, utiliza-se, exclusivamente, a matriz $\boldsymbol{X}$ na determinação da combinação linear das variáveis e as concentrações são usadas apenas quando os coeficientes de regressão são estimados. Essa mudança é feita a partir de uma PCA e os resultados são relacionados com as variáveis independentes $y$ (PARREIRA, 2003).

Na Análise de Componentes Principais, tem-se que $\boldsymbol{X} = \boldsymbol{TP'}$. Multiplicando por $\boldsymbol{P}$:

$$\boldsymbol{XP} = \boldsymbol{TP'P} \tag{38}$$

Como $\boldsymbol{P}$ é ortonormal, $\boldsymbol{P'P} = \boldsymbol{I}$, portanto:

$$\boldsymbol{T} = \boldsymbol{XP} \tag{39}$$

A etapa de correlação pode ser descrita como:

$$\boldsymbol{y} = \boldsymbol{Tb} + \boldsymbol{E} \tag{40}$$

$$\boldsymbol{b} = (\boldsymbol{T}'\boldsymbol{T})^{-1}\boldsymbol{T}'\boldsymbol{y} \quad (41)$$

Devido à ortogonalidade entre as linhas da matriz dos *scores*, a inversão de $\boldsymbol{T}'\boldsymbol{T}$ não é mais problemática.

PCR resolve o problema da colinearidade, através da obtenção de uma matriz inversível no cálculo de $\boldsymbol{b}$, e possui a habilidade de eliminar componentes principais permintindo a redução de erros aleatórios, ruídos. Contudo, PCR é um método de dois passos e, assim, corre-se o risco de perder informações úteis, ou predições, presentes em componentes principais descartadas. Além disso, alguns ruídos permanecerão nas componentes usadas na regressão.

*3.2.5.5 Regressão por Quadrados Mínimos Parciais*

O método PLS contorna essa característica do PCR usando a informação das concentrações na obtenção dos fatores, o que só é justificável se tais concentrações tiverem valores confiáveis. Neste método, a covariância nas medidas com as concentrações é utilizada em conjunto com a variância em $\boldsymbol{X}$ para gerar $\boldsymbol{T}$, tirando vantagem da correlação existente entre os dados da matriz original e a variável dependente. No PLS, o termo *variável latente* designa as componentes principais. Isto se deve ao fato de sua construção ser feita a partir de informações contidas no vetor das variáveis dependentes. O número de variáveis latentes que será utilizado é determinado durante o processo de validação cruzada.

Uma grande vantagem do método PLS é sua robustez, ou seja, os parâmetros do modelo não se alteram de maneira significativa quando novas amostras são acrescentadas ou retiradas do conjunto de calibração. Essa robustez permite que seja possível trabalhar com sistemas industriais cujas características nem sempre são mantidas rigorosamente da mesma maneira durante todo o processo, ou seja, é possível acrescentar novas amostras, conforme apareçam, sem alterar os parâmetros do modelo criado inicialmente (PARREIRA, 2003).

O método PLS é construído através do algoritmo Quadrados Mínimos Parciais Não Interativos (NIPALS) (do acrônimo inglês *Non-Interative Partial Least Squares*). O algoritmo é descrito a seguir:

(1) Selecionar um vetor $\boldsymbol{x}_j$ de $\boldsymbol{X}$ e chama-lo de $\boldsymbol{u}_h$: $\boldsymbol{x}_j = \boldsymbol{u}_h$

(2) Calcular $\boldsymbol{p}_h^t$: $\boldsymbol{p}'_h = \boldsymbol{t}'_h\boldsymbol{X}/\boldsymbol{t}_h^t\boldsymbol{t}_h$

(3) Normalizar $\boldsymbol{p}'_h$ para o comprimento 1: $\boldsymbol{p}'_{h\ novo} = \boldsymbol{p}'_{h\ antigo}/\left\|\boldsymbol{p}'_{h\ antigo}\right\|$

(4) Calcular $\boldsymbol{t}_h$: $\boldsymbol{t}_h = \boldsymbol{X}\boldsymbol{p}_h/\boldsymbol{p}'_h\boldsymbol{p}_h$

(5) Comparar o $\boldsymbol{t}_h$ usado no passo 2 com o obtido no passo 4. Se forem iguais, parar (a interação convergiu). Se ainda diferirem, ir ao passo 2.

Semelhante ao PCR, a matriz de *scores* pode representar a matriz de dados. Um modelo simplificado consiste na regressão os scores das matrizes $\boldsymbol{X}$ e $\boldsymbol{Y}$. O modelo PLS consiste nas relações extrínsecas ($\boldsymbol{X}$ e $\boldsymbol{Y}$ individualmente) e nas relações intrínsecas (unindo ambos).

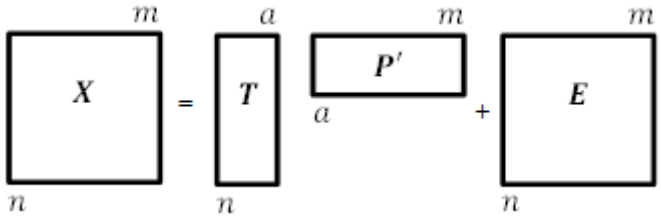
A relação exterior para $\boldsymbol{X}$ é:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}' + \boldsymbol{E} = \sum_{i=1}^{n} \boldsymbol{t}_i\boldsymbol{p}'_i + \boldsymbol{E} \qquad (42)$$

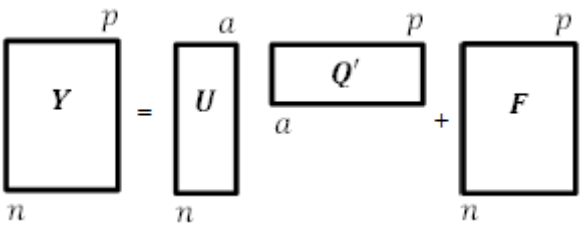
Pode-se construir para $\boldsymbol{Y}$ de forma semelhante:

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}' + \boldsymbol{F} = \sum_{i=1}^{a} \boldsymbol{u}_i\boldsymbol{q}'_i + \boldsymbol{F} \qquad (43)$$

Graficamente:




(44)




(45)

Podem-se descrever todas as componentes e fazer $\boldsymbol{E} = \boldsymbol{F} = 0$ ou não. A finalidade é descrever o máximo de $\boldsymbol{Y}$ e, por isso, fazer $\|\boldsymbol{F}\|$ ser o menor possível e, ao mesmo tempo, conseguir uma relação útil entre $\boldsymbol{X}$ e $\boldsymbol{Y}$. A relação intrínseca pode ser realizada observando um gráfico do *score* de $\boldsymbol{Y}$, $\boldsymbol{u}$, em função do *score* de $\boldsymbol{X}$, $\boldsymbol{t}$, para cada componente. O modelo mais simples é a relação linear:

$$\widehat{\boldsymbol{u}}_h = \boldsymbol{b}_h \boldsymbol{t}_h \tag{46}$$

Em que $\boldsymbol{b}_h = \boldsymbol{u}_h' \boldsymbol{t}_h / \boldsymbol{t}_h' \boldsymbol{t}_h$. O $\boldsymbol{b}_h$ funciona como coeficiente de regressão $\boldsymbol{b}$, nos modelos MLR e PCR. Contudo, esse modelo não é o melhor. Uma vez que as componentes principais são calculadas para cada matriz separadamente, relacionando fracamente entre si.

Um novo modelo, aplicando PCA duas vezes, pode ser escrito na forma de algoritmo, através de NIPALS. Para a matriz $\boldsymbol{X}$:

(1) Selecionar qualquer $\boldsymbol{x}_j$ e chama-lo de $\boldsymbol{t}_{início}$: $\boldsymbol{t}_{início} = \boldsymbol{x}_j$

(2) Calcular $\boldsymbol{p}'$: $\boldsymbol{p}' = \boldsymbol{t}'\boldsymbol{X}/\boldsymbol{t}'\boldsymbol{t}\ (= \boldsymbol{u}'\boldsymbol{X}/\boldsymbol{u}'\boldsymbol{u})$

(3) $\boldsymbol{p}'_{novo} = \boldsymbol{p}'_{antigo} / \|\boldsymbol{p}'_{antigo}\|$

(4) $\boldsymbol{t} = \boldsymbol{X}\boldsymbol{p}/\boldsymbol{p}'\boldsymbol{p}$

(5) Comparar $\boldsymbol{t}$ nos passos 2 e 4. Se forem iguais, parar. Se ainda diferirem, ir para 2.

Para a matriz $\boldsymbol{Y}$:

(1) Selecionar qualquer $\boldsymbol{y}_j$ e chama-lo de $\boldsymbol{u}_{início}$: $\boldsymbol{u}_{início} = \boldsymbol{y}_j$

(2) Calcular $\boldsymbol{q}'$: $\boldsymbol{q}' = \boldsymbol{u}'\boldsymbol{Y}/\boldsymbol{u}'\boldsymbol{u}\ (= \boldsymbol{t}'\boldsymbol{Y}/\boldsymbol{t}'\boldsymbol{t})$

(3) $\boldsymbol{q}'_{novo} = \boldsymbol{q}'_{antigo} / \|\boldsymbol{q}'_{antigo}\|$

(4) $\boldsymbol{u} = \boldsymbol{Y}\boldsymbol{q}/\boldsymbol{q}'\boldsymbol{q}$

(5) Comparar $\boldsymbol{t}$ nos passos 2 e 4. Se forem iguais, parar. Se ainda diferirem, ir para 2 (GELADI, 1986).

As relações acima são escritas como diferentes algoritmos. A troca de informação ocorre no passo 2, no qual $\boldsymbol{t}$ e $\boldsymbol{u}$ trocam de lugar. Deste modo, os algoritmos podem ser escritos sequencialmente:

(1) Selecionar qualquer $y_j$ e chama-lo de $\boldsymbol{u}_{início}$: $\boldsymbol{u}_{início} = \boldsymbol{y}_j$

(2) Calcular $\boldsymbol{p}'$: $\boldsymbol{p}' = \boldsymbol{t}'\boldsymbol{X}/\boldsymbol{t}'\boldsymbol{t}$ $(\boldsymbol{w}' = \boldsymbol{u}'\boldsymbol{X}/\boldsymbol{u}'\boldsymbol{u})$

(3) $\boldsymbol{p}'_{novo} = \boldsymbol{p}'_{antigo}/\|\boldsymbol{p}'_{antigo}\|$ $(\boldsymbol{w}'_{novo} = \boldsymbol{w}'_{antigo}/\|\boldsymbol{w}'_{antigo}\|)$

(4) $\boldsymbol{t} = \boldsymbol{X}\boldsymbol{p}/\boldsymbol{p}'\boldsymbol{p}$ $(\boldsymbol{t} = \boldsymbol{X}\boldsymbol{w}/\boldsymbol{w}'\boldsymbol{w})$

(5) $\boldsymbol{q}' = \boldsymbol{t}'\boldsymbol{Y}/\boldsymbol{t}'\boldsymbol{t}$

(6) $\boldsymbol{q}'_{novo} = \boldsymbol{q}'_{antigo}/\|\boldsymbol{q}'_{antigo}\|$

(7) $\boldsymbol{u} = \boldsymbol{Y}\boldsymbol{q}/\boldsymbol{q}'\boldsymbol{q}$

(8) Comparar $\boldsymbol{t}$ nos passos 4 e na repetição. Se forem iguais, com certo erro arredondado, parar. Se diferirem, ir ao passo 2 (Em casos para $\boldsymbol{Y}$ com uma variável, os passos 5-8 podem ser omitidos igualando $\boldsymbol{q} = 1$) (GELADI, 1986).

Apesar dos métodos PCR e PLS usarem componentes principais, muitas vezes faz-se necessária a seleção de variáveis, mas não devido às limitações, como descrito no método MLR, e sim porque o conjunto de dados pode conter variáveis de alta colinearidade, como no caso de espectros, ou mesmo variáveis que não contribuem fortemente para a construção de um modelo de regressão. Em tais casos, costuma-se recorrer a algum método de seleção com o objetivo de minimizar estes problemas e otimizar a construção dos modelos propriamente ditos.

### 3.2.6 Modelagem Independente Suave de Analogia de Classes

A SIMCA (do acrônimo em inglês *Soft Independent Modeling of Class Analogy*), é um método para classificação que considera informações da distribuição da população, estima um grau de confiança da classificação e pode prever novas amostras como pertencentes a uma ou mais classes ou nenhuma classe.

Na construção do modelo de classificação, calcula-se, para cada classe separadamente, o desvio padrão dos resíduos. Para o espaço descrito pelas PC, as variâncias das amostras são computadas em cada eixo. Por fim, esses dois parâmetros são usados na classificação de novas amostras.

O objetivo da SIMCA é criar um espaço limitado para cada classe. Isto pode ser melhor compreendido para uma classe descrita por duas componentes principais. Em termos geométricos, os resíduos desta classe correspondem às distâncias das amostras ao plano das componentes principais. Desta forma, o cálculo do desvio padrão dos resíduos dá origem a dois planos paralelos ao destes componentes, isto é, um acima e outro abaixo.

A classificação de uma nova amostra é feita através de sua projeção nas PC de cada classe. Em cada uma são calculadas as respectivas variâncias e resíduos. Em seguida, faz-se a comparação por um teste F, determinando em qual classe a amostra pertencerá. Dependendo do resultado, a amostra pode não pertencer a nenhuma classe (FLUMIGNAN, 2006).

*3.2.6.1 Tratamento Matemático*

Em SIMCA, cada conjunto de treinamento é descrito por suas próprias PC. Quando uma previsão é feita, novas amostras insuficientemente próximas ao espaço das componentes principais são consideradas *não membros*. Além disso, cada amostra de treinamento precisa ser atribuída a uma das $Q$ diferentes categorias, sendo que $Q > 1$. Assim, SIMCA pode ser considerada uma PCA com $Q$ classes, em que $Q > 0$ (PIROUETTE, 2008).

Os modelos de PC para duas classes são incorporados por duas matrizes divididas de loadings, $\boldsymbol{V}_1$ e $\boldsymbol{V}_2$. As amostras da classe 1 podem ser encaixadas no modelo da classe 2, ao projetar seu bloco pré-processado e sua transformada, $\boldsymbol{X}_1$, no espaço definido pelos *loadings* da classe 2 (PIROUETTE, 2008). Esta relação está descrita na equação 47.

$$\widehat{\boldsymbol{X}}_1 = \boldsymbol{X}_1\boldsymbol{V}_1\boldsymbol{V}_2' \qquad (47)$$

Os resíduos são definidos por:

$$\boldsymbol{E} = \widehat{\boldsymbol{X}}_1 - \boldsymbol{X}_1 \tag{48}$$

Em que $\boldsymbol{E}$ é a matriz contendo o vetor linha $\boldsymbol{e}_i$ para cada amostra da classe 1. O resíduo entre classes para cada amostra na classe 1 é definido por:

$$s_{12} = \left[\frac{1}{(m-k_2)n_1}\sum_i^{n_1} e_i e_i'\right]^{\frac{1}{2}} \tag{49}$$

Em que $k_2$ é o número de componentes principais no modelo da classe 2 e $n_1$ é o número de amostras na classe 1. O cálculo da equação 3 é repetido para todos os pares de combinações de cada $Q$ classes para produzir uma matriz $Q$ por $Q$. Esta matriz não é simétrica, ou seja $s_{12} \neq s_{21}$; encaixar as amostras da classe 1 no modelo de classe 2 não é equivalente a encaixar as amostras da classe 2 no modelo da classe 1. Para classes que estão bem encaixadas entre si e são bem separadas das outras, os elementos da matriz diagonal, que são os resíduos da classe encaixada em si mesma, deve ser consideravelmente menor que os valores fora da diagonal.

Informações sobre a separabilidade das classes podem ser apresentadas de outra forma. Para cada par de classes e entre os resíduos das classes, definidos anteriormente, a distância entre classes é dada por:

$$D_{12} = \left(\frac{s_{12}^2+s_{21}^2}{s_{11}^2+s_{22}^2}\right)^{\frac{1}{2}} - 1 = D_{21} \tag{50}$$

O cálculo da equação 50 pode ser repetido para todos os pares de combinação de $Q$ classes, gerando uma matriz $Q$ por $Q$. Essa matriz é simétrica e todos os elementos da diagonal são iguais a zero, pois $D_{11} = D_{22} = 0$. Grandes distâncias entre classes implicam em classes bem separadas. Por regra, as classes são consideradas separáveis se a distância entre si for maior que 3 (PIROUETTE, 2008).

O Poder de Modelagem Total (TMP) (do acrônimo em inglês *Total Modeling Power*) é uma medida simples da força de modelagem através de todas as $Q$ classes. Para cada classe $q$, a variância residual é computada para a $j$-ésima coluna de $E$:

$$\hat{s}_{jq}^2 = \frac{e_j' e_j}{n_q - k_q - 1} \tag{51}$$

Para cada classe, a variância total é dada por:

$$s_{0jq}^2 = \frac{1}{n_q - 1} \sum_i^{n_q} \left(x_{ij} - \bar{x}_j\right)^2 \quad (52)$$

A variância residual é corrigida para o número de fatores em cada classe do modelo e somada em todas as classes. Assim:

$$s_{jT}^2 = \sum_q s_{jq}^2 \frac{m}{m - k_q} \quad (53)$$

A variância total também é somada para todas as classes:

$$s_{0jT}^2 = \sum_q s_{0jq}^2 \quad (54)$$

Assim, o TMP é dado por:

$$TMP = 1 - \frac{s_{jT}}{s_{0jT}} \quad (55)$$

O TMP é útil para a determinação de variáveis que possuem pouca ou nenhuma importância para qualquer classe no conjunto de treinamento. Varia de 0 a 1, embora possa ter valores negativos (PIROUETTE, 2008).

Pode ser instrutivo também para se conhecerem quais variáveis são melhores para discriminação entre classes no conjunto de treinamento. Para cada variável, comparando a variância de resíduo médio de cada classe em relação a todas as outras e a variância residual de todas as classes, tem-se uma indicação do quanto uma variável discrimina entre uma classificação correta e incorreta. O Poder de Discriminação é definido por:

$$DP = \frac{1}{Q - 1} \frac{\sum_q^Q \sum_r^Q \left({}_q\hat{\boldsymbol{e}}_{jr}\right)' \left({}_q\hat{\boldsymbol{e}}_{jr}\right)}{\sum_r^Q \left({}_r\hat{\boldsymbol{e}}_{jr}\right) \left({}_r\hat{\boldsymbol{e}}_{jr}\right)} \quad (56)$$

O vetor residual ${}_q\hat{\boldsymbol{e}}_{jr}$ denota a coluna j-ésima da matriz residual, após o encaixe do conjunto de treinamento na classe $r$ para o modelo para classe $q$. O dobro da soma no numerador é avaliado para casos em que $q \neq r$.

Uma amostra desconhecida (uma amostra que não foi previamente atribuída a nenhuma classe) pode ser ajustada em ambas as classes 1 e 2 do modelo. Quando encaixada no modelo 1, produz um vetor residual:

$$\boldsymbol{e}_u = \widehat{\boldsymbol{x}}_u - \boldsymbol{x}_u = \boldsymbol{x}_u \boldsymbol{V}_1 \boldsymbol{V}_1' - \boldsymbol{x}_u \quad (57)$$

Em que o $u$ subscrito indica a amostra desconhecida (do inglês, *unknown*). A variância residual pode ser calculada através de $\boldsymbol{e}_u$:

$$s_{u1}^2 = \frac{\boldsymbol{e}_u \boldsymbol{e}_u'}{m-k} \quad (58)$$

A raiz quadrada da variância residual pode ser comparada com o valor crítico, como descrito na equação 59.

$$s_{crit1} = s_{01}(F_{crit})^{\frac{1}{2}} \quad (59)$$

Em que o valor crítico $F_{crit}$ é baseado nos graus de liberdade 1 e $n - k_1$ e $s_{01}$ é a raíz quadrada da variância da classe 1, definida por:

$$s_{01}^2 = \frac{1}{n_1 - k_1} \sum_i^{n_1} s_i^2 \quad (60)$$

Se $s_{u1}$ for menor que $s_{crit1}$, a amostra desconhecida é designada para a classe 1.

O mesmo processo é repetido para a classe 2. Se $s_{u2}$ for menor que $s_{crit2}$, a amostra desconhecida é designada para a classe 2. Se a amostra se qualificar para ambas as classes, a que possuir menor valor de amostra residual é considerada a melhor e a outra é denominada próxima melhor classe. Se a amostra exceder os dois valores críticos, não será designada para nenhuma classe (PIROUETTE, 2008).

### *3.3 Cromatografia Gasosa*

A Cromatografia Gasosa (CG) é um procedimento físico utilizado para separar uma amostra em seus componentes individuais. A base para esta separação é a distribuição da amostra entre duas fases, uma fase estacionária e uma fase gasosa móvel.

A fase móvel é denominada gás de arraste ou gás portador, já que se trata de um gás inerte cuja finalidade é transportar, através da coluna, as moléculas a serem

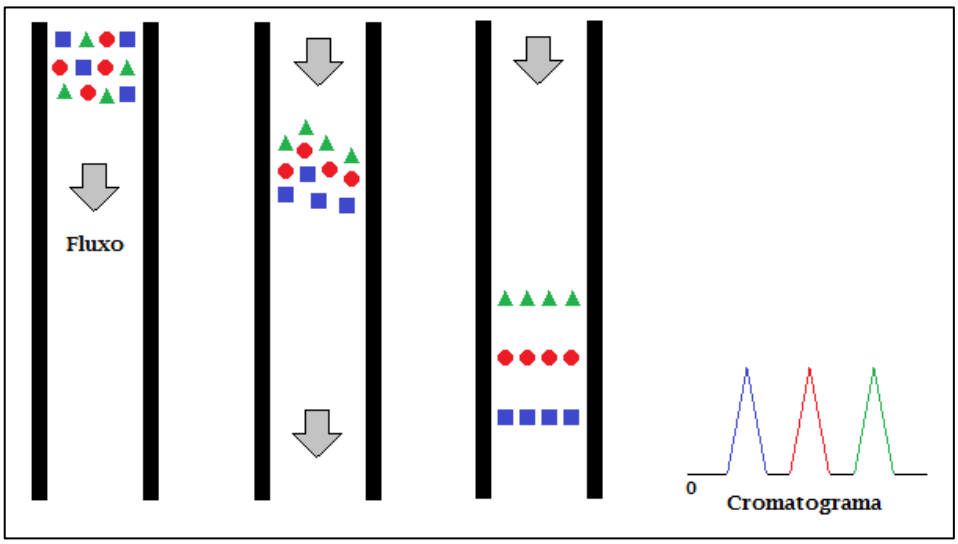
separadas. A figura 10 mostra um esquema da separação de três componentes de uma mesma matriz e a construção de um cromatograma simples.

De acordo com a natureza da fase estacionária, pode-se diferir a cromatografia gasosa em cromatografia gás-líquido (CGL) e cromatografia gás-sólido (CGS).

Na CGS, a fase estacionária é um sólido de grande área superficial, usualmente um adsorvente como carvão vegetal, sílica-gel, ou peneira molecular (zeólita sintética). A adsorção diferencial dos componentes da mistura a ser separada sobre a superfície sólida é a base para a separação cromatográfica na CGS, a qual é, geralmente, empregada na separação de gases como o nitrogênio, oxigênio, monóxido de carbono e outros (LANÇAS, 1993).

Na CGL, a fase estacionária é uma película delgada líquida, a qual recobre um sólido inerte, denominado suporte. A base para a separação cromatográfica, neste caso, é a distribuição da amostra dentro e fora desta película, ou seja, a partição da amostra entre a fase móvel e a fase líquida estacionária.

**Figura 10** - Esquema de uma separação cromatográfica.




**Fonte**: Elaborado pelo autor (2013).

Em um cromatógrafo a gás, a amostra é injetada em uma coluna capilar e, juntamente com o gás de arraste, eluída até um detector. Este é necessário para medir a quantidade de um componente da amostra no efluente da coluna. Bons detectores possuem alta sensibilidade, baixo ruído, uma ampla faixa de resposta, para todos os

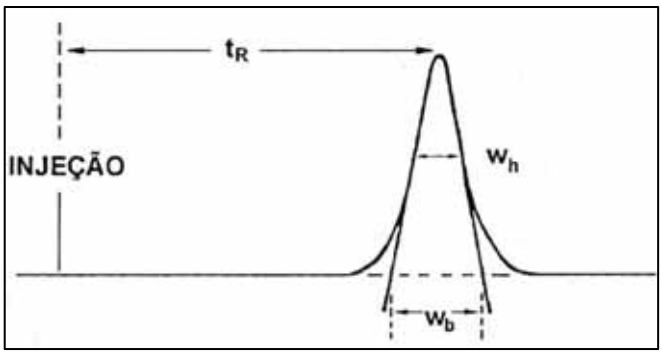
tipos de compostos. Uma baixa sensibilidade ao fluxo e a flutuações de temperatura é desejável, mas nem sempre possível (HADDEN, 1971).

### 3.3.1 Teoria Básica

Um cromatograma é o registro gráfico da análise, em que se indicam os compostos e o grau de concentração em que se encontravam presentes em um determinado tempo. Quando apenas o gás de arraste deixa a coluna, uma linha reta, paralela ao eixo do tempo é traçada. Esta linha é denominada linha de base. Ao se eluírem compostos presentes na amostra, mostrar-se-ão os perfis de suas concentrações, o que permite a obtenção de dois parâmetros fundamentais: o *tempo de retenção* e a *área do pico* (LANÇAS, 1993). A representação de um pico comum em cromatografia é feita na figura 11.

Através da largura do pico, $w_b$, ou da largura da meia altura, $w_h$; é possível determinar a concentração do respectivo componente. Assim, a área do pico é proporcional à concentração do constituinte da amostra (LANÇAS, 1993).

**Figura 11** - Representação de um pico cromatográfico comum.




**Fonte**: LANÇAS, F. M., 1993.

O tempo de retenção, $t_R$, é o tempo transcorrido desde o momento da injeção da amostra até que se tenha obtido o máximo do pico. Este tempo é característico do soluto, da fase líquida do fluxo do gás de arraste e da temperatura da coluna. O mesmo pico, na mesma coluna, em idênticas condições terá sempre o mesmo

tempo de retenção. Este fato permite a identificação dos picos, posto que são reprodutíveis (LANÇAS, 1993).

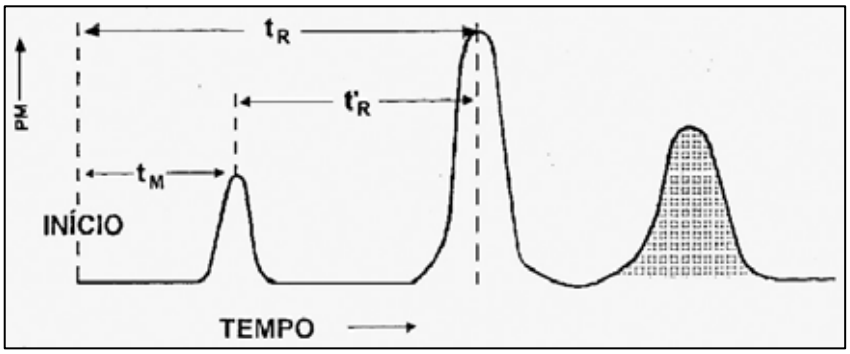
A posição do pico é determinada pela velocidade do fluxo de gás de arraste e pelo fator de retenção $k$, denominado relação de partição ou relação de distribuição.

$$k = \frac{quantidade\ de\ amostra\ na\ fase\ líquida}{quantidade\ de\ amostra\ na\ fase\ sólida} \qquad (61)$$

Esta relação entre a quantidade da amostra na fase líquida e na fase gasosa também está relacionada com o tempo de retenção:

$$t_R = t_M + t_M k = t_M + t'_R \qquad (62)$$

**Figura 12** - Representação gráfica de $t_R$, $t'_R$ e $t_M$.




**Fonte**: LANÇAS, F. M., 1993.

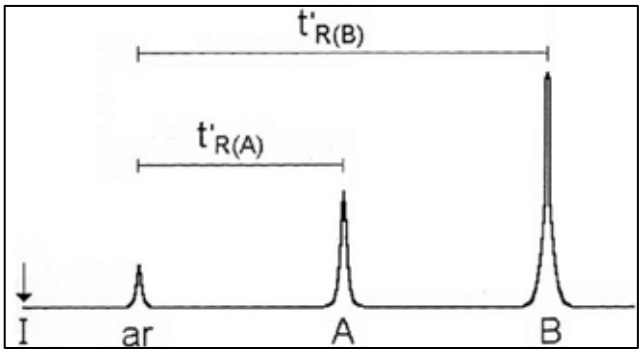
Em que $t_R$ é o tempo que transcorre desde o ponto da injeção até o máximo do pico e $t_M$ é o tempo transcorrido desde o ponto de injeção até obter-se um pico inerte, ar por exemplo. É uma medida do tempo que cada componente da amostra permanece na fase gasosa.

Pela figura 12, nota-se que o tempo de retenção é a soma dos tempos que a amostra permanece nas duas fases: o tempo na fase móvel, $t_M$, e o tempo na fase estácionária, $t'_R$.

O fator de separação, $\alpha$, é a relação existente entre o tempo que dois picos permanecem na fase móvel. Este fator é representado pela figura 13 e pela equação 63.

$$\alpha = \frac{t'_R(B)}{t'_R(A)} \qquad (63)$$

**Figura 13** - Representação gráfica do fator de separação $\alpha$.




**Fonte**: LANÇAS, F. M., 1993.

A principal utilidade de $\alpha$ é fornecer uma medida da seletividade da fase móvel com relação a dois componentes. Para $\alpha = 1$, os dois componentes terão a mesma solubilidade na fase estacionária e não poderão ser separados na mesma. Quanto maior o valor de $\alpha$, mais seletiva será a fase móvel e, consequentemente, melhor a separação de $A$ e $B$ (LANÇAS, 1993).

A eficiência de uma coluna é determinada pelo número de pratos teóricos, $N$, gerados. Quanto maior o valor de $N$, melhor a eficiência e, portanto, melhor a separação. A equação 64 mostra a relação entre $N$, $t_R$ e $w_b$.

$$N = 16\left(\frac{t_R}{w_b}\right)^2 \quad (64)$$

Em casos de picos mal resolvidos, nos quais $w_b$ é difícil de ser medida, ou havendo severa variação na linha de base, o número de pratos teóricos pode ser calculado através da meia altura do pico, $w_h$:

$$N = 5{,}545\left(\frac{t_R}{w_h}\right)^2 \quad (65)$$

Diversos fatores afetam o valor de $N$, tais como: tempo de retenção, comprimento da coluna, temperatura da coluna, soluto, fluxo de gás, tamanho da amostra, técnica de injeção e assim por diante. O número de pratos pode ser mudado desde que se variem estas condições, o que torna difícil uma comparação entre o número de pratos de diferentes colunas ou instrumentos (LANÇAS, 1993).

Os parâmetros operacionais de uma coluna podem ser avaliados por seu efeito sobre a altura equivalente a um prato teórico, $H$, que corresponde à razão entre o comprimento da coluna, $L$, e o número de pratos teóricos, $N$ (LANÇAS, 1993). Assim:

$$H = \frac{L}{N} \tag{66}$$

Em que $L$ é medido em centímetros ou milímetros. Portanto, $H$ é o comprimento necessário de uma coluna para gerar um prato teórico. Assim, quanto maior $N$, menor será $H$ e mais eficiente será a coluna (LANÇAS, 1993).

Os fatores que determinam a largura dos picos podem ser expressos na equação de Van Deemter:

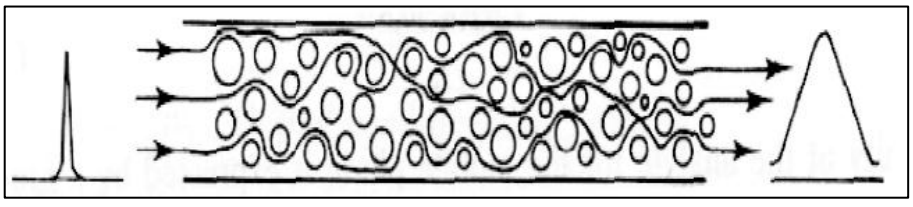
$$H = A + \frac{B}{\bar{\mu}} + C + \bar{\mu} \tag{67}$$

Em que $A$ é a Difusão de Eddy, $B$ é a Difusão longtudinal, $C$ são os Efeitos de Transferência de Massa e $\bar{\mu}$ é a velocidade linear média do gás de arraste dada por:

$$\bar{\mu} = \frac{L}{t_M} \tag{68}$$

A Difusão de Eddy, representada na figura 14, é calculada através dos diversos caminhos que um analito pode percorrer na fase estacionária. Uma vez que os picos dependem da quantidade de constituinte da amostra que sai da coluna, ao tomarem diferentes percursos, o pico é alargado. Contudo, em uma coluna capilar, $A = 0$, posto que não há partículas, considerando a coluna como um tubo simples (HAWKES, 1983).

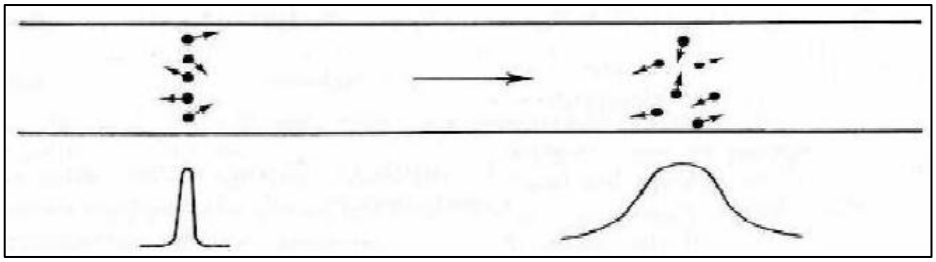
**Figura 14** - Representação esquemática da Difusão de Eddy.



**Fonte**: Elaborado pelo autor (2013).

Já o termo $B$, Difusão Longitudinal, representado na figura 15, descreve como as moléculas atravessam a coluna sob a influência da fase móvel. Devido à difusão molecular no interior da coluna, quanto maior o fluxo de gás, menor a dispersão do analito e menor será o valor de $H$, estreitando o pico (HAWKES, 1983).

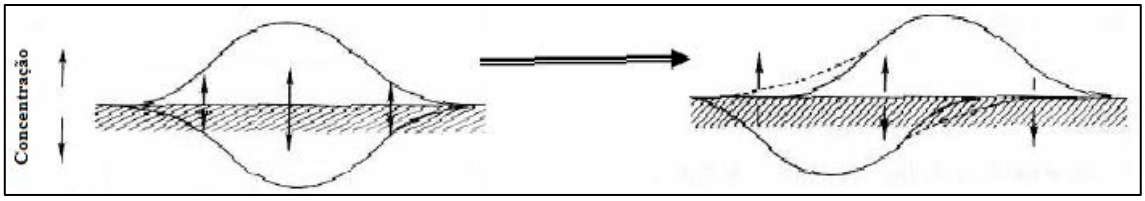
**Figura 15** - Representação esquemática da Difusão Longitudinal.



**Fonte**: Elaborado pelo autor (2013).

Por fim, o termo $C$, que remete aos Efeitos de Transferência de Massas, representado na figura 16, aumenta devido ao desequilíbrio, que não ocorre instantaneamente, entre a fase estacionária e a fase móvel. Consequentemente, parte do componente na fase móvel tende a ser eluído mais rapidamente do que a parte que está na fase estacionária. Isto gera um alargamento do pico (HAWKES, 1983).

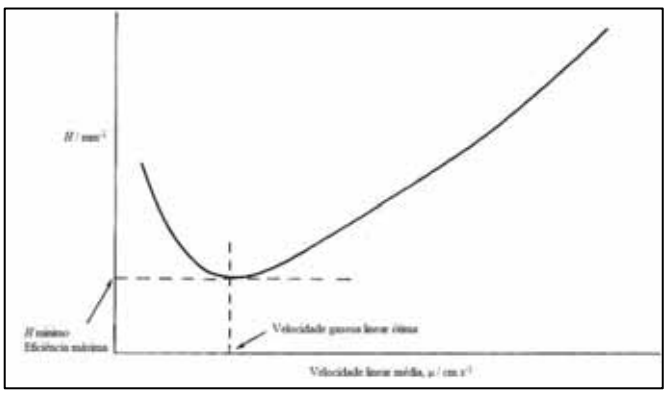
**Figura 16** - Representação esquemática dos Efeitos de Transferência de Massas.




**Fonte**: Elaborado pelo autor (2013).

A representação gráfica da equação de van Deemter é descrita na figura 17. O mínimo da curva representa a menor altura do prato teórico e o máximo de eficiência da coluna cromatográfica.

**Figura 17** - Representação gráfica da equação de van Deemter.



**Fonte**: LANÇAS, F. M., 1993.

Um último conceito a ser introduzido é a resolução, $R_S$. Trata-se de uma medida quantitativa da separação de dois picos consecutivos, sendo descrita como:

$$R_S = \frac{t_{R2} - t_{R1}}{\frac{1}{2}(w_{b1} + w_{b2})} = \frac{2\Delta t_R}{w_{b1} + w_{b2}} \quad (69)$$

O valor $\Delta t_R$ é uma medida da separação dos máximos dos picos adjacentes. Este termo poderá ser aumentado reduzindo-se a temperatura, ou escolhendo-se uma fase mais seletiva, maior $\alpha$. Já $w_b$ é uma medida de eficiência da coluna e está relacionada com a velocidade de alargamento da banda, podendo ser medido por $N$ ou por $H$ (LANÇAS, 1993).

Além do número de pratos teóricos e da seletividade, a resolução depende da posição relativa dos picos no cromatograma. Assim, pode-se reescrever a equação anterior na forma:

$$R = \left(\frac{\sqrt{N}}{4}\right)\left(\frac{k}{k+1}\right)\left(\frac{\alpha - 1}{\alpha}\right) \quad (70)$$

Esta equação é denominada equação mestra da resolução (LANÇAS, F. M., 1993).

### 3.3.2 Instrumentação

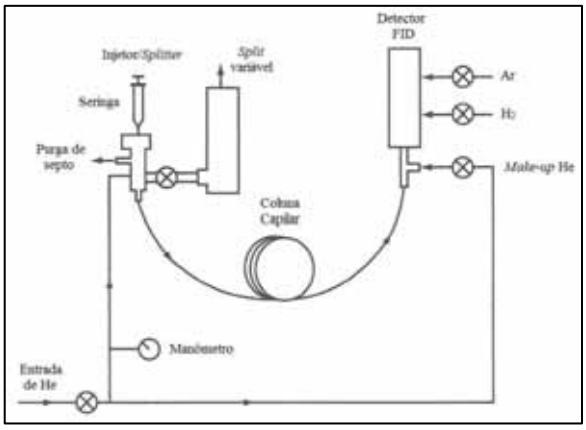
A figura 18 representa, esquematicamente, o sistema de um cromatógrafo a gás básico.

Um gás de arraste inerte, provindo de um grande cilindro de gás, flui continuamente através da porta de injeção, da coluna e do detector. A taxa de fluxo do gás é devidamente controlada para se assegurar a reprodutibilidade dos tempos de retenção e para reduzir os ruídos e desvios gerados pelo detector. A amostra é, então, injetada, com uma microsseringa (caso a amostra seja líquida), na porta de injeção previamente aquecida. Realizada a injeção, a amostra é vaporizada e carregada através da coluna, comumente capilar de 15 a 30m de comprimento, com um filme fino interno de 0,2 μm de espessura (fase estacionária). Os componentes da amostra são, assim,

separados pelas fases móvel e estacionária, baseados na solubilidade relativa e pressão de vapor relativos (McNAIR, 2009).

Em seguida, o gás de arraste e a amostra passam por um detector, o qual mede a quantidade de amostra e gera um sinal elétrico. Este sinal é processado, a fim de gerar um cromatograma correspondente. Na maior parte dos casos, o sistema de tratamento de dados, comumente um computador com o *software* próprio do cromatógrafo, integra a área dos picos, realiza os cálculos e imprime um relatório com os resultados quantitativos e tempos de retenção (McNAIR, 2009).

**Figura 18** - Esquema de um típico cromatógrafo a gás.




**Fonte**: McNAIR, H. M.; MILLER, J. M., 2009.

### *3.3.2.1 Gás de Arraste*

A principal funcionalidade do gás de arraste é, como o próprio nome diz, "arrastar" a amostra através da coluna, ou seja, agir como a fase móvel. Deve ser inerte e não pode interagir quimicamente com a amostra. Sua segunda função é prover uma matriz adequada para o detector diferenciar e os componentes da amostra. O quadro 1 mostra quais gases de arraste são ideais para cada tipo de detector (McNAIR, 2009).

**Quadro 1** - Gases de arraste ideais para cada detector.

| Detector | Gás de arraste |
|---|---|
| Condutividade Térmica | Hélio |
| Chama Ionizante | Hélio, Nitrogênio ou Hidrogênio |
| Captura de Elétrons | Nitrogênio muito seco |

**Fonte**: McNAIR, H. M.; MILLER, J. M., 2009.

### *3.3.2.2 Injetores*

Colunas capilares possuem alguns requisitos para injeções de amostras: a injeção precisa ser realizada rapidamente e a quantidade de amostra deve ser muito pequena, geralmente, menor que 1µg. Em uma coluna capilar típica de 25m, há cerca de 10mg de fase líquida. Isso mostra que se forem injetadas grandes quantidades de amostra, a coluna pode ser entupida, ou ocorrer vazamentos (McNAIR, 2009).

Picos provenientes de colunas capilares são muito estreitos, havendo larguras de poucos segundos, por isso, injeções rápidas são necessárias, a fim de minimizar o alargamento de banda produzido por uma injeção lenta (McNAIR, 2009).

Os injetores podem ser classificados em *split*, *splitless*, *on-column*, *cold on-column*. Os mais comuns são os dois primeiros e estão descritos a seguir.

### *3.3.2.3 Injeção* Split

A figura 19 mostra uma secção transversal de um injetor *split* comum.

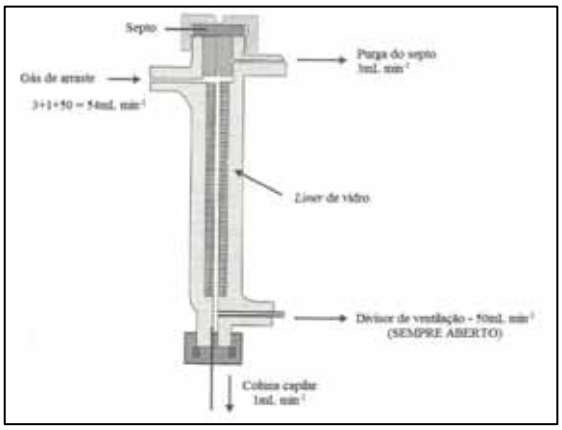
A injeção *split* é a técnica mais simples e a mais fácil a ser usada. O procedimento envolve, normalmente, a injeção de 1µL de amostra com uma seringa padrão através de uma porta de injeção aquecida contendo um *liner* de vidro desativado.

A amostra é rapidamente vaporizada e apenas uma fração, geralmente 1-2%, do vapor entra na coluna. O restante da amostra vaporizada e um grande fluxo de gás de arraste passam pela purga do septo.

A técnica é simples, pois o operador precisa apenas controlar a taxa de *split* abrindo ou fechando a purga do septo. A quantidade de amostra introduzida na coluna é muito pequena e a taxa de fluxo até o ponto de *split* é rápida (soma da taxa de fluxo da coluna e da ventilação). Isto resulta em separações de alta resolução. Outra vantagem é a introdução de amostras concentradas sem a necessidade de diluí-las, uma vez que se use uma taxa de *split* alta.

Uma desvantagem dessa técnica de injeção é a limitação da análise de traços, posto que apenas uma fração é inserida na coluna. Uma segunda desvantagem é que esse modo de injeção, algumas vezes, discrimina solutos com altos pesos moleculares na amostra, tornando a amostra que entra na coluna não representativa da amostra injetada (McNAIR, 2009).

**Figura 19** - Secção transversal de um injetor *split*.




**Fonte**: McNAIR, H. M.; MILLER, J. M., 2009.

### *3.3.2.4 Injeção* Splitless

A figura 20 mostra uma secção transversal de um injetor *splitless* comum.

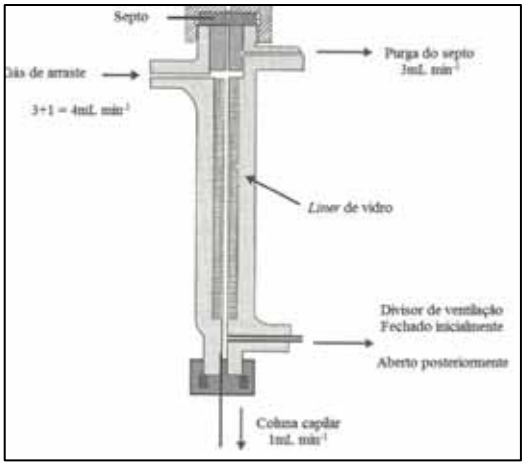
A injeção *splitless* utiliza o mesmo equipamento da injeção *split*, porém com a válvula de *split* inicialmente fechada. A amostra é diluída em um solvente volátil, como hexano ou metanol, e de 1 a 5µL é injetado na porta de injeção aquecida. A amostra é vaporizada e lentamente (com taxa de fluxo por volta de 1mL/min) carregada até uma coluna fria, na qual amostra e solvente são condensados. Após cerca de 60s, dependendo da taxa de fluxo da coluna, a válvula de separação é aberta e qualquer vapor residual deixado na porta de injeção é rapidamente varrido para fora do sistema. Purga do septo é essencial em injeções *splitless*.

A coluna possui temperatura programada e, no início, apenas os solventes voláteis são vaporizados e levados através da coluna. Enquanto isso, os analitos são concentrados em uma banda estreita do solvente residual. Após algum tempo, esses compostos são vaporizados pela coluna aquecida e separados cromatograficamente. Observa-se uma alta resolução dos constituintes com ponto de ebulição elevados.

A grande vantagem da injeção *splitless* é a melhora na sensibilidade em relação ao *split* e, como consequência, melhora-se análise de traços para amostras farmacêuticas, biomédicas e ambientais.

Por outro lado a injeção *Splitless* tem algumas desvantagens. Consome muito tempo, já que se inicia com uma coluna fria, tornando necessária a programação da temperatura. Deve-se diluir a amostra com solvente volátil. Por fim, não é uma injeção indicada para compostos voláteis. Para se obter um bom cromatograma, os primeiros picos de interesse devem ter pontos de ebulição com 30ºC acima do solvente (McNAIR, 2009).

**Figura 20** - Secção transversal de um injetor *splitless*.




**Fonte**: McNAIR, H. M.; MILLER, J. M., 2009.

### *3.3.2.5 Detectores*

Um detector capta os efluentes de uma coluna e grava uma resposta na forma de um cromatograma. Os sinais do detector são proporcionais à quantidade de cada analito, tornando possível uma análise quantitativa.

O mais comum é o Detector de Ionização em Chama (FID) (do acrônimo em inglês *Flame Ionization Detector*). Possui alta sensibilidade, linearidade e, além disso, é relativamente simples e com baixo custo. Outros detectores populares são o TCD (do acrônimo em inglês *Thermal Conductivity Detector*) e o ECD (do acrônimo em inglês *Electron Capture Detector*) (McNAIR, 2009).

#### 3.3.2.5.1 Detector de Ionização em Chama

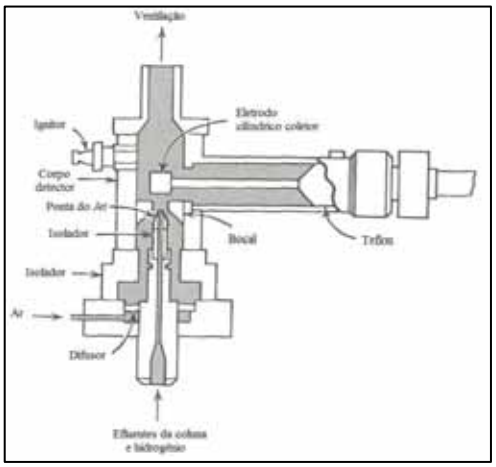
A figura 21 mostra a secção transversal de um típico detector FID.

Como fora dito, o FID é o mais largamente usado em CG. Os efluentes da coluna são queimados em uma pequena chama de ar sintético-hidrogênio, produzindo alguns íons no processo. Estes são coletados e formam uma pequena corrente que é

transformada em um sinal. Quando não há amostras na chama, ocorre pouca ionização e uma pequena corrente, cerca de $10^{-14}$A, proveniente de impurezas do hidrogênio e do ar sintético.

Conforme mostra a figura 18, os efluentes da coluna são misturados com hidrogênio, ou são previamente misturados se $H_2$ for usado como gás de arraste, e levados até a ponta de um pequeno queimador, o qual está cercado por alto fluxo de ar sintético, que auxilia a combustão. A ignição da chama é feita automaticamente. O eletrodo coletor é polarizado em cerca de +300V em relação à ponta da chama e a corrente coletada é amplificada por um circuito de alta impedância. Posto que há formação de água no processo de combustão, o detector precisa ser aquecido até pelo menos 125ºC, a fim de evitar a condensação da água e de amostras com alto ponto de ebulição. Grande parte dos FID funciona em 250ºC ou acima (McNAIR, 2009).

**Figura 21** - Secção transversal de um típico detector FID.




**Fonte**: McNAIR, H. M.; MILLER, J. M., 2009.

O exato mecanismo da ionização em chama ainda não é conhecido. O FID responde a todos os compostos orgânicos que queimam em chama de ar sintético-hidrogênio. O sinal é aproximadamente proporcional ao conteúdo de carbono, dando origem à regra *igual por carbono*, um fator de resposta constante obtido através da conversão de todos os átomos de carbono em um soluto orgânico em metano no processo de combustão no detector (McNAIR, 2009).

Assim, todos os hidrocarbonetos devem mostrar uma mesma resposta por átomo de carbono. Se houverem heteroátomos presentes, como oxigênio ou nitrogênio,

o fator de resposta diminui. Valores relativos são tabulados como Número Efetivo de Carbonos (NEC) e alguns estão descritos na tabela 2.

Para uma operação eficiente, os gases (hidrogênio e ar sintético) devem ser puros e livres de materiais orgânicos, para elevar a ionização de fundo. A taxa de fluxo precisa ser otimizada para cada diferente modelo de detector. A taxa de fluxo de hidrogênio tem um máximo de sensibilidade para cada taxa de fluxo de gás de arraste. Colunas tubulares abertas possuem fluxos por volta de 1mL $min^{-1}$, assim, um gás auxiliar (*make-up*) é adicionado ao gás de arraste para elevar o total para 30mL $min^{-1}$ (McNAIR, 2009).

**Tabela 2** - Número Efetivo de Carbonos (relativos ao heptano).

| Composto | NEC | NEC Teórico |
|---|---|---|
| Acetileno | 1,95 | 2 |
| Hexeno | 5,82 | 6 |
| Etanol | 2,52 | 2,5 |
| Butanal | 3,12 | 3 |
| Ácido Acético | 1.01 | 1 |
| Acetato de Metila | 1,04 | 2 |
| Acetona | 2,00 | 2 |

**Fonte**: McNAIR, H. M.; MILLER, J. M., 2009.

Gás hidrogênio pode ser usado como gás de arraste, desde que avaliados o fluxo de gás e o modelo do detector. Além de medidas de precaução, como uma fonte separada (em local afastado) de hidrogênio.

A taxa de fluxo de ar possui valores entre 300-400mL $min^{-1}$, o suficiente para a maioria dos detectores (McNAIR, 2009).

Compostos que não contém carbono orgânico não queimam e não são detectados. Os mais importantes estão listados no quadro 2. O mais significante é a água, já que produz picos largos e assimétricos. A ausência de picos de água permite a

utilização do FID em análises de amostras contendo $H_2O$, desde que não interfira no cromatograma. Aplicações típicas incluem contaminantes orgânicos em água, vinho e outras bebidas alcoólicas e produtos alimentícios (McNAIR, H. M.; MILLER, J. M., 2009).

**Quadro 2** - Compostos que dão baixa ou nenhuma resposta no FID.

| | | |
|---|---|---|
| He | $CS_2$ | $NH_3$ |
| Ar | COS | CO |
| Kr | $H_2S$ | $CO_2$ |
| Ne | $SO_2$ | $H_2O$ |
| Xe | NO | $SiCl_4$ |
| $O_2$ | $N_2O$ | $SiHCl_3$ |
| $N_2$ | $NO_2$ | $SiF_4$ |

**Fonte**: McNAIR, H. M.; MILLER, J. M., 2009.

Em suma, as vantagens do uso do FID são sua boa sensibilidade, grande linearidade, simplicidade, robustez e adaptabilidade a todos os tamanhos de colunas.

### 3.3.3 Cromatografia Gasosa Ultrarrápida

Com o passar dos anos, as técnicas cromatográficas evoluíram. O primeiro aspecto a ser melhorado foi a coluna cromatográfica, com o surgimento de novas fases estacionárias e de colunas capilares. Em seguida, a preparação da amostra a ser injetada no cromatógrafo. Só então, a atenção voltou-se para o tempo de análise. Assim, surgiram as técnicas de Cromatografia Gasosa Rápida, Cromatografia Gasosa Muito Rápida e CG-UR. A evolução será abordada a seguir.

#### *3.3.3.1 Evolução*

Por volta de 1960, vem surgindo teorias relacionadas à redução de tempo de análise na CG. Nesta época, ainda se utilizavam colunas empacotadas longas, com

diâmetro interno reduzido, e capazes de atingir desempenho de 10 pratos efetivos por segundo.

Em 1957, Golay introduziu as chamadas colunas capilares, que suprimiram algumas das limitações relacionadas as colunas empacotadas, tais como aumento de resolução e possibilidade de separação de matrizes mais complexas. Estudos posteriores mostraram a capacidade destas colunas de realizarem separações mais rápidas, conforme o diâmetro interno fosse reduzido (GOLAY, 1958).

Com o surgimento e aprimoramento de novas fases estacionárias e colunas capilares, a resolução cromatográfica em CG desenvolveu-se cada vez mais. Contudo, até 1980, a problemática relacionada ao aumento de velocidade não estava atrelada somente a limitações das colunas. Outras variáveis precisavam ser consideradas como limitações do equipamento de época e o tempo necessário para preparação de uma amostra e posterior tratamento dos dados obtidos do cromatograma, o que excedia e muito ao tempo da separação cromatográfica (SEQUINEL et al, 2010).

Conforme os cromatógrafos foram evoluindo, juntamente com o surgimento de novas fases estacionárias e do desenvolvimento de *softwares* avançados para tratamento de dados cromatográficos, o estudo do aumento da velocidade em uma separação cromatográfica voltou a ser foco de muitos pesquisadores (SEQUINEL et al, 2010).

Os usuários de cromatografia passaram a questionar algumas das condições experimentais convencionalmente utilizadas e chegaram à conclusão que a resolução cromatográfica poderia ser reduzida para um valor limite que não causasse prejuízos para a identificação dos compostos presentes em uma amostra. Isto reduziria o tempo de análise em CG. Assim, constatou-se que os três fatores mais importantes para se realizar uma separação rápida são:

1. Minimização da resolução cromatográfica para um valor apenas suficiente, de modo a reduzir o número de pratos teóricos, N, requeridos para uma dada separação. Pode ser feito reduzindo o comprimento da coluna e empregando altas velocidades do gás de arraste (conforme visto nas equações 59 e 61) (SEQUINEL et al, 2010);

2. Maximização da seletividade do sistema cromatográfico. Em separações rápidas, as propriedades das colunas são essenciais para melhorar a distinção entre pares críticos de componentes de uma amostra (com tempos de retenção próximos) (SEQUINEL et al, 2010);
3. Proposição de alternativas que reduzam o tempo de análise, mantendo a resolução constante: determinados parâmetros podem ser explorados caso o tempo de análise em uma resolução mínima aceitável ainda exceda o tempo desejável. Exemplo: redução do diâmetro interno da coluna, para compensar a inevitável perda da resolução, quando se diminui o comprimento da coluna (SEQUINEL et al, 2010).

*3.3.3.2 Definição*

Os conceitos que fundamentam o aumento de velocidade em CG são estabelecidos através de diversos parâmetros, ao invés do tempo de análise somente. Dagan e Amirav introduziram um fator para quantificar o aumento da velocidade: SEF (do acrônimo em inglês *Speed Enhancement Factor*). A partir deste, a CG foi subdividida em três categorias: *Cromatografia Gasosa Rápida* (CG-R), *Cromatografia Gasosa Muito Rápida* (CG-MR), e CG-UR, além da CG-C.

O SEF não necessariamente reflete a exata redução do tempo de análise, porém normaliza as separações em função da CG-C, que é caracterizada pela utilização de coluna de 30m com diâmetro interno reduzido, menor que 0,25mm, de tal forma que o fluxo de gás de arraste na coluna seja de 1mL $min^{-1}$ (correspondente a He com velocidade linear de 34cm $s^{-1}$). Este fator é definido pelo produto do fator de redução do comprimento da coluna e do fator que relaciona o aumento da velocidade do gás de arraste em relação a CG-C. Como exemplo, a CG-UR apresenta um fator de aumento de velocidade na faixa de 400 a 4000 vezes em relação à CG-C.

Vários autores apresentaram diferentes definições, baseadas nas larguras dos picos cromatográficos, $w_b$ e $w_h$, e no tempo total de analise, para a classificação dos diferentes tipos de CG. Magni *et al*. apresentaram uma definição mais detalhada, divergindo das apresentadas até o momento. Segundo esta nova definição, fatores como

dimensões da coluna e a taxa de aquecimento também foram considerados. Como apresentado na tabela 3, separações ultrarrápidas somente seriam obtidas com colunas com dimensões muito reduzidas e altas taxas de aquecimento, resultando em picos com largura entre 0,05 e 0,2s (SEQUINEL et al, 2010).

**Tabela 3** - Principais parâmetros para a caracterização das diversas modalidades de cromatografia gasosa.

| **Classificação segundo Magni *et al.*** | **CG-UR** | **CG-R** | **CG-CC**[2] | **CG-C** |
|---|---|---|---|---|
| Comprimento da coluna / m | 2 a 10 | 5 a 15 | 5 | 25 a 30 |
| Diâmetro interno / mm | 0,05 a 0,10 | 0,10 a 0,25 | 0,25 | 0,25 a 0,32 |
| Tempo de análise / min | $\leq$ 1 | $\leq$ 10 | $\leq$ 3 a 15 | $\leq$ 10 a 60 |
| Taxa de aquecimento / $^{o}C\ min^{-1}$ | $\geq$ 60 | 15 a 60 | 5 a 40 | 1 a 10 |
| Largura do pico / s | 0,05 a 0,2 | 0,5 a 2 | 1 a 5 | 1 a 10 |

[2] Cromatografia Gasosa com Coluna Curta.

**Fonte**: SEQUINEL et al, 2010

Embora os termos apresentados, CG-UR, CG-R dentre outros, sejam comumente encontrados na literatura, não existe uma posição acadêmica consolidada a respeito destas classificações. Permitindo, assim, uma classificação mais subjetiva por parte dos usuários de cromatografia, principalmente em relação à CG-UR (SEQUINEL et al, 2010).

### *3.3.3.3 Instrumentação da CG-UR*

O aumento da velocidade não necessariamente dependerá da substituição ou mesmo da adaptação dos equipamentos convencionais da CG, já que nem sempre os equipamentos convencionais são compatíveis com a velocidade requerida. Para uma redução de duas a quatro vezes no tempo total de uma análise cromatográfica, a utilização de colunas de dimensões reduzidas acompanhada por simples alterações feitas em alguns parâmetros cromatográficos, tais como aumento na taxa de aquecimento e otimização da velocidade do gás de arraste, podem ser suficientes. Contudo, para separações ultrarrápidas, redução de dez vezes ou mais no tempo de análise, pode

requerer uma adaptação significante dos cromatógrafos convencionais. Para isso, são obrigatórios amostradores automáticos de alta velocidade, sistemas para aquecimento rápido da coluna e detectores com altas taxas de aquisição (SEQUINEL et al, 2010).

#### 3.3.3.3.1 Amostradores e Injetores

São preferíveis os amostradores automáticos de alta velocidade, a fim de obter uma máxima diminuição do tempo necessário para se completar um ciclo analítico. A injeção de nanovolumes pode ser uma ferramenta muito útil em sistemas ultrarrápidos, uma vez que utilizam colunas com menor capacidade de processamento de amostra. Neste caso, seriam necessárias pequenas adaptações nos amostradores automáticos, como a utilização de seringas com êmbolo na agulha (*plunger-in-neddle syringe*) (SEQUINEL et al, 2010).

Dos injetores convencionais, todos poderiam ser aplicados em CG-UR em princípio: *split*/*splitless*, *cold on-column* etc. (SEQUINEL et al, 2010).

#### 3.3.3.3.2 Coluna Cromatográfica

As colunas capilares utilizadas em CG-UR possuem dimensões reduzidas, conforme mostrado na tabela 3. Tais colunas são de fundamental importância, pois diminuem o tempo de análise e aumentam a resolução dos cromatogramas. A variação do diâmetro interno em uma coluna cromatográfica promove um balanço entre dois fatores: a eficiência e a capacidade da coluna. A otimização de um destes requer a minimização do outro.

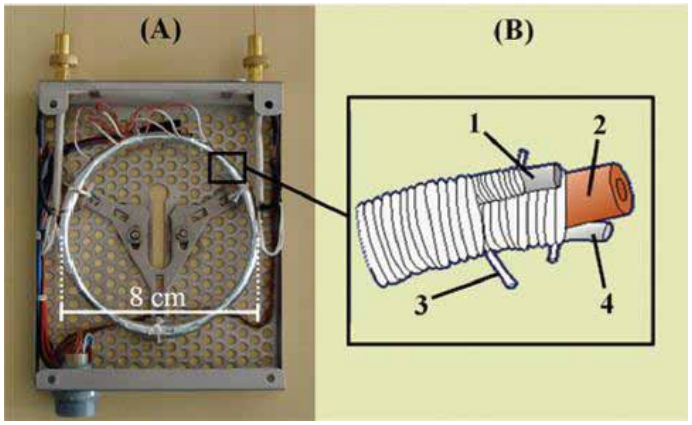
Em CG-UR, a utilização de colunas com diâmetro interno reduzido, as chamadas *narrow bore columns*, proporciona um aumento no valor de $N$, gerando picos mais estreitos, e surge como alternativa a perda de resolução em decorrência da diminuição do comprimento da coluna (relação vista na equação 66 da seção 3.3.1).

As separações ultrarrápidas também são facilitadas com o uso de colunas com depósito fino de fase estacionária, em que a resistência oferecida no processo de transferência de massa nesta fase é praticamente insignificante quando comparada à fase gasosa (diminuindo, assim o valor de $C$ na equação 67 da seção 3.3.1).

3.3.3.3.3 Sistema de Aquecimento da Coluna

Com relação às altas taxas de programação de temperatura, descrito na tabela 3, uma das opções é a utilização de sistemas dedicados ao aquecimento, em que a coluna narrow bore é colocada no interior de um módulo ultrarrápido (UFM) (do acrônimo em inglês *Ultra Fast Module*), conhecido usualmente como "gaiola". Neste sistema, a coluna recebe calor por meio de um aquecimento resistivo. Tanto o sensor de temperatura, quanto a resistência elétrica são revestidos por uma camada de teflon, ou fibra cerâmica, constituindo um conjunto de pouca massa térmica (SEQUINEL et al, 2010). A figura 22 contém um exemplo do módulo ultrarrápido.

**Figura 22** - Módulo ultrarrápido ("gaiola") usado em CG-UR. (A) Sistema dedicado de aquecimento resistivo da coluna (*Thermo Fisher Scientific Inc.*®). (B) Detalhe do revestimento da coluna capilar: 1-elemento aquecedor, 2-coluna capilar, 3-fibra cerâmica, 4-sensor de temperatura.




**Fonte**: SEQUINEL et al, 2010.

Assim, nenhum aquecimento do forno é exigido e são utilizadas as interfaces convencionais de injetor e detector do cromatógrafo convencional. O sistema modular para aquecimento permite alcançar maiores velocidades de aquecimento, em relação ao forno da CG-C. Em se tratando de CG-UR, este módulo é imprescindível

para a obtenção de separações cromatográficas no limite de alguns segundos (SEQUINEL et al, 2010).

### 3.3.4 Alinhamento dos Cromatogramas

Antes de se aplicarem técnicas quimiométricas aos cromatogramas, estes devem passar por um processo de alinhamento dos picos. Posto que, mesmo havendo reprodutibilidade das análises cromatográficas, sempre haverá um pequeno desvio entre os picos (tempos de retenção) da mesma substância em diferentes cromatogramas. Isto pode gerar erros durante a análise com a PCA, com a SIMCA ou qualquer outra ferramenta.

Assim, faz-se necessária uma técnica de alinhamento que, não só adeque os cromatogramas, quanto o faça com a menor distorção possível. A seguir, será detalhada a ferramenta chamada *Correlação de Deformação Otimizada*, comumente utilizada em alinhamentos cromatográficos, e fundamentada nas teorias de *Programação Dinâmica* e *Deformação Dinâmica do Tempo*.

As técnicas de alinhamento descritas nesta seção podem ser utilizadas para diferentes tipos de dados. Contudo, certos tipos de sinais, como espectros de Ressonância Magnética Nuclear (RMN) e Espectroscopia de Massa (MS) são mais restritos, devido aos seus picos agudos. Este tipo de dados espectrais é melhor alinhados por *bucketing* ou por técnicas que utilizem escolha de picos. (JELLEMA, 2009)

#### *3.3.4.1 Programação Dinâmica*

A Programação Dinâmica (DP) (do acrônimo em inglês *Dynamic Programming*), em técnicas de alinhamento, é usada para encontrar o menor caminho cumulativo através de um mapa de busca cumulativa $D$. Para a construção do mapa $D$, um mapa de distâncias $d$ é construído, contendo as distâncias entre todos os possíveis

pares de intensidades medidos em um sinal $x$ e um sinal $y$. Essa distância $d$ pode ser calculada através da distância Euclidiana, equação 12, para todos os pares de sinais.

Como primeiro passo na construção da matriz $D$, a distância matricial mínima local cumulativa é calculada no primeiro ponto $d(1,1)$. Para cada ponto na matriz $D$, que possui o mesmo tamanho de $d$, três distâncias são calculadas e o menor valor é inserido na matriz $D$. As três distâncias estimadas são calculadas conforme a equação 71 (JELLEMA, 2009).

$$D(i,j) = min\begin{Bmatrix} D(i-1,j) + d(i,j) \\ D(i-1,j-1) + d(i,j) \\ D(i,j-1) + d(i,j) \end{Bmatrix} \qquad (71)$$

*3.3.4.2 Deformação Dinâmica do Tempo*

O método de Deformação Dinâmica do Tempo (DTW) (do acrônimo e inglês *Dynamic Time Warping)*, foi originalmente concebido para alinhar espectros de frequências de palavras pronunciadas por diferentes emissores de som para propósitos de reconhecimento.

No reconhecimento de fala, a velocidade com que um indivíduo pronuncia uma palavra, ou uma sentença composta por diferentes palavras, limita a possibilidade do reconhecimento automático do texto falado e a geração de sinais correspondentes. Assim, faz-se necessária uma nova amostra para adequar-se ao sinal registrado.

A fim de resolver este problema, o método chamado de *time warping*, que faz uso da DP, foi implementado. Métodos de deformação esticam e comprimem o eixo do tempo ou frequência, resultando numa duração longa ou curta respectivamente. Já o DTW é um método flexível que estica ou encolhe o eixo do tempo ou da frequência de um sinal, para se adequar o melhor possível em um sinal de referência. O objetivo é transformar a trajetória de dois sinais, de forma que os eventos semelhantes correspondam. A medida para isso pode ser a distância mínima, a alta correlação, ou qualquer outra medida de igualdade que se adapte aos dados sob investigação (JELLEMA, 2009).

O caminho de deformação $W$ procura transformar o eixo do tempo, ou de frequência, de uma amostra de sinal $x$,de comprimento $L_x$, de tal forma que encaixe em um sinal de referência $y$, de comprimento $L_y$. Um conjunto comum de pontos $k = 1,2 \dots K$ é utilizado para o eixo comum em ambos os sinais. Nestes eixos, os pontos de tempo correspondentes, $n(k)$ e $m(k)$, representam os índices para a amostra de sinal $x$ e a referência de sinal $y$ respectivamente. O $k$-ésimo elemento de $W$, $[m(k), n(k)]$, contém os índices para a amostra e para a referência no $k$-ésimo ponto em comum (deformado) no eixo do tempo (JELLEMA, 2009). Isto é ilustrado na equação 72.

Com DTW, uma nova sequencia de pontos é construída, conforme mostra a equação 73. Cada ponto $c(k)$ é um par de índices $m(k)$ e $n(k)$que remetem aos sinais originais $x$ e $y$.

$$W = \langle [m(k), n(k)] | k = 1, \dots, K \rangle \tag{72}$$

$$c(k) = [m(k), n(k)] \tag{73}$$

$$W = \{c(1), c(2), \dots, c(k), \dots c(K)\} \tag{74}$$

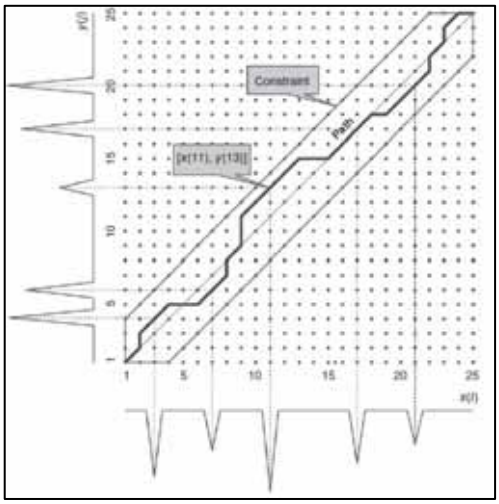
A equação 74 é o caminho deformado, resultado combinação das equações 72 e 73. Cada valor de $W$ mostra qual elemento de $x$ e de $y$, ou seja $m(k)$ e $n(k)$ combinados em $c(k)$, é posicionado no mesmo elemento. O k-ésimo elemento de $W$, $[m(k), n(k)]$, contém os índices para a referência da amostra no k-ésimo ponto do eixo comum ou deformado (JELLEMA, 2009).

O sinal de referência é esboçado ao longo do eixo $y$, enquanto que o sinal da amostra ao longo do eixo $x$. Se há um alinhamento perfeito, a representação da função de deformação será diagonal. Havendo desalinhamentos, a função desvia da diagonal. Na figura 23, tem-se que o pico com tempo igual a 11 da amostra $x$ não corresponde a nenhum pico no mesmo tempo na referência $y$. Adequando os índices 11 e 13 no caminho de deformação, o sinal da amostra é adequado ao sinal de referência. A porção horizontal do caminho representa que a mesma intensidade de referência é tomada para duas intensidades consecutivas da amostra (repetindo o valor), ao passo que as transições verticais são feitas de forma contrária (JELLEMA, 2009).

A fim de encontrar o caminho ótimo através da grade de possibilidades mostrada na figura 23, são necessárias restrições para a obtenção de resultados

sensíveis. A primeira é onde começam e terminam os pontos a serem fixados, correspondendo a $c(1)$ igual a $(1,1)$ e $c(K) = (L_x, L_y)$. Para restringir o caminho de deformação, indo de trás para frente, uma restrição de continuidade local é imposta. Isto assegura que o caminho é sempre igual com uma inclinação não negativa, expressa por $m(k+1) \geq m(k)$ e $n(k+1) \geq n(k)$. A restrição global é representada na figura 23, em que as linhas limitam uma área de pontos de grade permitidos para a busca de um caminho ótimo. Desvios maiores que $\pm M$ pontos de grade da diagonal ou caminho linear não são permitidos. A escolha de $M$ depende da troca máxima esperada para o sinal. Um valor de $M$ muito baixo restringirá os picos a serem adequados com os picos correspondentes da amostra de referência, enquanto que um valor de $M$ muito alto resultará em uma combinação incorreta. Após a construção das restrições, o caminho ótimo é feito através da DP (JELLEMA, 2009).

**Figura 23** - Grade mostrando os pontos da amostra original $x$ e o sinal de referência $y$ e suas novas posições nos eixos de deformação.




**Fonte**: JELLEMA, R. H., 2009.

### *3.3.4.3 Correlação de Deformação Otimizada*

O método COW (do acrônimo em inglês *Correlation Optimized Warping*) é uma extensão da DTW. Enquanto que a DTW trata cada ponto de tempo individualmente, o algoritmo COW trabalha com segmentos de dados, mantendo pontos

consecutivos próximos. A forma de operação, semelhante à DTW, é feita em dados vetoriais completos e, assim, não necessita de um algoritmo para detecção de picos para decompor o sinal. DTW sem restrições é muito flexível e, portanto, inadequada para o pré-processamento de dados cromatográficos.

Em COW, dois parâmetros precisam ser estabelecidos: o *comprimento de segmento* $s$ e o *comprimento de slack* (ou folga) $l$. Dado o tamanho de $s$, os sinais de referência e de amostra são subdivididos em regiões locais, ou janelas, que são iterativamente esticados e comprimidos para maximizar $l$ via interpolação, maximizando, portanto, a correlação entre os sinais da amostra e da referência. A seleção do comprimento de segmento e de folga pode ser realizado através de um esquema de otimização.

O algoritmo COW adequa todos os sinais em torno de um sinal de referência comum. Como apenas um sinal de referência é envolvido, é importante a escolha de um sinal com muitas semelhanças entre as amostras.

O sinal da amostra $x$, de tamanho $L_x$, é dividido em seções de comprimento $s$, em que o número de seções $N$ é dado pela equação 75.

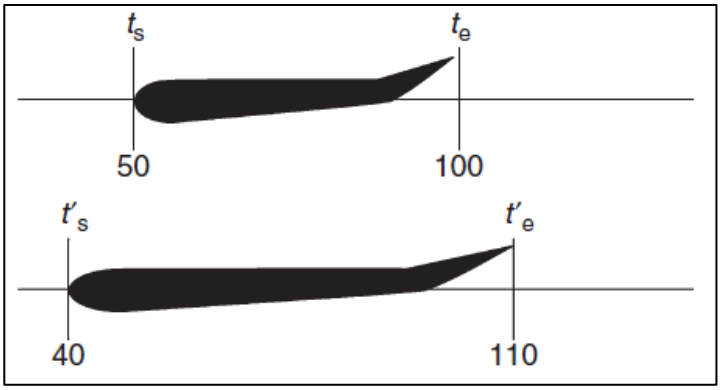
$$N = \left[\frac{L_x}{s}\right] \tag{75}$$

Em que $N$ é um valor inteiro positivo. O sinal de referência $y$, de tamanho $L_y$, é dividido em $N$ seções também. Os segmentos são deformados em comprimentos maiores ou menores, de tal forma que melhor se encaixarão no sinal de referência. Com COW, a interpolação linear é utilizada para realizar a deformação. A interpolação linear de um segmento com posição inicial $t_s$ e posição final $t_e$ para um novo conjunto de posições $t'_s$ e $t'_e$ é feita através da equação:

$$w_j = \frac{j}{t'_e - t'_s}(t_e - t_s) + t'_s \tag{76}$$

Em que $j = 0, \ldots, t'_e - t'_s$. A figura 24 mostra o resultado de um segmento esticado da posição 50 a 100 até a posição 40 a 110. O ponto original $t_s = 50$ é deformado para a posição $t'_s = 40$, de acordo com o cálculo $w_j = [0/(110 - 40)](100 - 50) + 40 = 40$.

**Figura 24**. Exemplo de alongamento de um segmento na direção de uma nova linha do tempo. Este é um exemplo exagerado para mostrar que COW foi desenvolvido de tal forma que segmentos que estão ligados, permanecem conectados. Um segmento que requer encurtamento ou deslocamento para a esquerda para melhor adequação com o sinal de referência precisa harmonizar-se com seu segmento vizinho, que também precisará tornar-se maior ou deslocar-se para a esquerda.




**Fonte**: JELLEMA, 2009.

Os pontos finais de cada segmento são chamados *nodos*, em que a posição do ponto inicial do segmento $i$, depois de deformado, é denotado $t_i$. O nodo 0 é o ponto inicial da primeira seção e o nodo $M$ é o ponto final do sinal completo. Já que o sinal da amostra é deformado através de um sinal de referência, após o alinhamento, ambos os sinais terão o mesmo comprimento $L_y$ do sinal de referência $y$ e, portanto, o comprimento $x'_M$ é igual a $L_y$.

Considera-se que apenas certa quantidade de alongamento e compressão em cada segmento é necessária e é controlada pelo tamanho do *slack* $l$. A restrição nessa quantidade acelera o cálculo, enquanto que limitam as possíveis instâncias de deformação. Da mesma forma, alongamentos e compressões extremas são contidos, o que resultaria em grandes deformações dos picos.

A qualidade de alinhamento de cada segmento depende da escolha da quantidade de alongamento e compressão. A qualidade de alinhamento, $\rho$, é uma correlação entre os sinais da amostra e de referência. Seu cálculo é realizado para todas as possíveis combinações de segmentos deformados, permitidas pelos parâmetros de deformação $s$ e $l$. Essa relação é mostrada na equação 77.

$$\rho(\mathbf{n}) = \frac{\mathrm{cov}[y(\mathbf{n}),x\prime(\mathbf{n})]}{\sqrt{\mathrm{var}[y(\mathbf{n})]\mathrm{var}[x\prime(\mathbf{n})]}} \qquad (77)$$

O vetor **n** contém os pontos do tempo, selecionados do sinal de referência original, $y$, e do sinal deformado, $x'$. Uma vez que as correlações são obtidas para todos os possíveis segmentos, a função de deformação pode ser determinada para um valor global ótimo, em que a teoria da DP é usada. No caso da COW, o método de encontrar o caminho ótimo começa ao se determinar o caminho ótimo de deformação para o último segmento e caminhar até o primeiro segmento. Essa variante é chamada de DP inversa. Para cada segmento, as combinações subótimas de caminhos de deformação são descartadas. Por fim, as funções de deformação locais são combinadas, a fim de produzir um caminho global ótimo de deformação para $x$, resultando no sinal deformado $x'$.

Por exemplo, se um tamanho de *slack* de 2 é escolhido, haverão cinco possíveis pontos iniciais para o segmento, definidos por $-2, -1, 0, 1, 2$. Após usar interpolação linear para corrigir o comprimento dos segmentos esticados $(2,1)$ e comprimidos $(-2, -1)$ do sinal da amostra $x$, o coeficiente de correlação entre o sinal deformado e o segmento de referência correspondente do sinal $y$ é computado. Cinco diferentes coeficientes de correlação são encontrados. No segundo passo, os segmentos $N - 1$ em $x$ e em $y$ são considerados e a mesma aproximação é usada para a deformação nesta seção. Assim, para as 5 novas condições, existirão 25 coeficientes de correlação. Os princípios da DP são empregados novamente como na técnica de DTW.

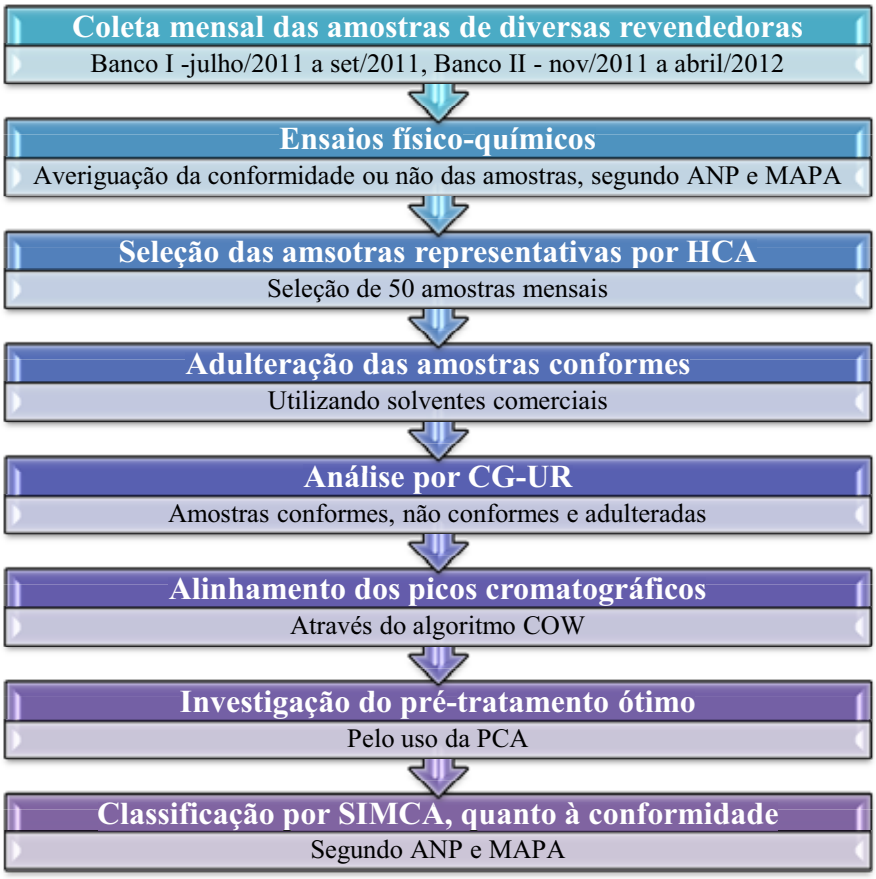
O ótimo para um segmento deformado predecessor é igual ao ótimo global. O caminho ótimo é encontrado pelo cálculo do ótimo global e de todos os seus predecessores. As correlações ótimas calculadas podem ser somadas (seja soma simples, seja soma ponderada), ou multiplicadas. O caminho ótimo das posições de fronteira dos segmentos é então aplicado e os pontos entre os limites são interpolados, resultando na melhor adequação da amostra pré-processada para os pré-determinados tamanhos de *slack* e segmento (JELLEMA, 2009).

# *4 MATERIAIS E MÉTODOS*

## *4.1 Resumo do Procedimento Experimental*

Para rápido entendimento, a figura 25 representa o resumo do procedimento experimental realizado neste trabalho.

**Figura 25** - Resumo do procedimento experimental realizado neste trabalho.




**Fonte**: Elaborado pelo autor (2013).

### *4.2 Armazenamento*

As diversas amostras de gasolina utilizadas foram coletadas mensalmente de vários postos revendedores de combustíveis na região centro-oeste de São Paulo pelo CEMPEQC, localizado no Instituto de Química da Unesp de Araraquara.

As gasolinas foram coletadas em frascos PET âmbar e brancos com 1L de capacidade e armazenadas a 10ºC ao abrigo da luz. Após a realização dos ensaios físico-químicos, o volume remanescente foi adulterado, se a amostra fora classificada como não conforme, com pelo menos um solvente entre os descritos na seção *Adulterantes*, e, então, foram armazenados em frascos âmbar de 100mL a fim de se repetirem as análises se necessário. Para amostras com volume superior a 1L, usou-se um galão, ou bomba, com 20L de capacidade e este foi devidamente etiquetado.

**Figura 26** - Frascos de armazenamento de amostras de gasolina.


**Fonte**: Elaborado pelo autor (2013).

### *4.3 Solventes e Adulterantes*

Os solventes usados para análise cromatográfica estão relacionados a seguir:

- n-heptano 99,5% P.A. Marca: Vetec Química Fina. Frasco de vidro âmbar de 1L;

- n-hexano 95% P.A. Marca: Vetec Química Fina. Frasco de vidro âmbar de 1L;
- octano 99,0% P.A. Marca: Fluka Analytical. Frasco âmbar de 1500mL;
- n-tridecano 99,0%. Marca: Acros Organics. Frasco âmbar de 100mL.

Os solventes utilizados para adulteração da gasolina estão relacionados a seguir.

- Solvente ecológico A-90. Marca: Carbosolv. Frasco de vidro âmbar de 1L (Figura 27a);
- Aguarrás. Marca: Thinsol Química. Frasco de alumínio de 900mL (Figura 27b);
- Solvente alifático leve DT. Marca: Carbosolv. Frasco de vidro âmbar de 1L (Figura 27c);
- Solvente alifático leve SB. Marca: Carbosolv. Frasco de vidro âmbar de 1L (Figura 27d);
- Solvente alifático médio AR. Marca: Carbosolv. Frasco de vidro âmbar de 1L (Figura 27e);
- Solvente aromático AB-9. Marca: Carbosolv. Frasco de vidro âmbar de 1L (Figura 27f);
- Benzolum Crystallisabile P.A. Marca: Reagentes ECIBRAS. Frasco âmbar de 1L (Figura 27g);
- Solvente A 4070. Marca: Carbosolv. Frasco de vidro âmbar de 1L (Figura 27h);
- Cicloexano 99% P.A. Marca: Vetec Química Fina. Frasco de vidro âmbar de 1L (Figura 27i);
- Éter de petróleo (30-60) P.A. Marca: Isofar. Frasco de vidro âmbar de 1L (Figura 27j);
- Etilbenzeno 99,80%. Marca: Acros Organic. Frasco de vidro âmbar de 1L (Figura 27k);
- Etilenoglicol 99,5% P.A. Marca: Carlo Erba. Frasco de vidro âmbar de 1L (Figura 27l);
- n-heptano 99,5% P.A. Marca: Vetec Química Fina. Frasco de vidro âmbar de 1L (Figura 27m);
- n-hexano 95% P.A. Marca: Vetec Química Fina. Frasco de vidro âmbar de 1L (Figura 27n);
- NAFTA (solvente cedido por fornecedores). Frasco de vidro âmbar de 1L (Figura 27o);

- Querosene (solvente cedido por fornecedores). Frasco de vidro âmbar de 1L (Figura 27p);
- Solvente para borracha (solvente cedido por fornecedores). Frasco de vidro âmbar de 1L (Figura 27q);
- Tolueno 99,90%. Marca: J.T. Baker. Frasco âmbar de 1L (Figura 27r);
- Xileno 98%. Marca: Vetec Química Fina. Frasco âmbar de 1L (Figura 27s);

**Figura 27** - Solventes utilizados para adulteração das amostras de gasolina: (a) Solvente ecológico A-90, (b) Aguarrás, (c) Solvente alifático leve DT, (d) Solvente alifático leve SB, (e) Solvente alifático médio AR, (f) Solvente aromático AB-9, (g) Benzolum Crystallisabile, (h) Solvente A 4070, (i) Cicloexano, (j) Éter de petróleo, (k) Etilbenzeno, (l) Etilenoglicol, (m) n-heptano, (n) n-hexano, (o) NAFTA, (p) Querosene, (q) Solvente para borracha, (r) Tolueno, (s) Xileno.


**Fonte**: Elaborado pelo autor (2013).

## *4.4 Equipamentos*

Os equipamentos utilizados estão relacionados a seguir.

- Destilador automático. AD-6. Tanaka Scientific Limited (Figura 28a);
- Densímetro eletrônico. DMA 4500. Anton Paar (Figura 28b);
- Infravermelho. Irox 2000. Grabner Instruments (Figura 28c);
- Cromatógrafo a gás com módulo ultrarrápido acoplado e detector FID. Trace CG Ultra. Thermo Fisher Scientific Inc. (Figura 28d);
- Amostrador automático. TriPlus. Thermo Fisher Scientific Inc. (Figura 28e).

**Figura 28** - Equipamentos utilizados para análises das amostras: (a) Destilador automático, (b) Densímetro eletrônico, (c) Irox, (d) Cromatógrafo a gás com módulo ultrarrápido acoplado, (e) Amostrador automático.


**Fonte**: Elaborado pelo autor (2013).

## *4.5* Softwares *Utilizados*

Os softwares utilizados neste trabalho foram o ChromQuest®, para análises cromatográficas, o MATLAB®, para alinhamento dos picos cromatográficos e o Pirouette®, análises quimiométricas.

- ChromQuest® 5.0. Version 3.2.1. Build 3.2.1.837. © 2008 Thermo Fisher Scientific Inc.;
- MATLAB®. Version 7.5.0.342. 1984-2007, ©The MathWorks Inc.;
- Microsoft Excel®. Version 14.0.4760.1000 (32 bits)
- Microsoft Word®. Version 14.0.4760.1000 (32 bits);
- Pirouette®. Version 3.11. © 1990-2003, ©Infometrix, Inc.

## *4.6 Análises Cromatográficas*

A Figura 29 representa o equipamento utilizado na análise cromatográfica de gasolina comercial. A coluna cromatográfica capilar empregada foi a PH5, da Thermo Fisher Scientific Inc., cujas características são: 5,00m de comprimento, 0,40μm de espessura do filme e 0,10mm de diâmetro interno. O detector usado foi o FID e sua temperatura foi fixada em 270ºC. Os gases utilizados pelo detector foram o hidrogênio, com fluxo de 45mL $min^{-1}$, gás nitrogênio, 30mL $min^{-1}$,como *make-up* e ar sintético, 350mL $min^{-1}$. O injetor empregado foi o *Split/Splitless* e sua temperatura foi mantida em 270ºC. As amostras foram injetadas com uma seringa de capacidade de 10μL e o volume de amostra injetado foi de 0,1 μL.

As análises cromatográficas foram realizadas com temperatura inicial de 40 ºC, mantida por 0,60 min, e elevada até 250ºC, mantido por 0,15min, em uma taxa de 100 ºC $min^{-1}$. O fluxo de gás de arraste (hidrogênio) foi mantido constante em 0,5mL $min^{-1}$ e a taxa de *split* foi de 1:500.

**Figura 29** - Módulo ultrarrápido acoplado em um cromatógrafo a gás.

**Fonte**: Elaborado pelo autor (2013).

### *4.7 Análises Quimiométricas*

O alinhamento de picos cromatográficos pelo algoritmo COW foi realizado no *software* MATLAB®. Version 7.5.0.342. As análises quimiométricas HCA, PCA e SIMCA foram realizadas no *software* Pirouette®. Version 3.11. Por limitação da versão do programa, os valores de intensidade foram divididos por 1000.

### *4.8 Descarte*

Os resíduos gerados foram rotulados como *descarte gasolina*, adulterada ou não, e *descarte solventes* e encaminhados para o Depósito de Resíduos do Instituto de Química da Unesp de Araraquara/SP. Posteriormente, foram levados para incineração com contínua emissão de gases de pequenas quantidades, realizada pela empresa AMBICAMP®.

# *5 RESULTADOS E DISCUSSÕES*

## *5.1 Análise Físico-química das Amostras Coletadas*

As análises físico-químicas foram realizadas no mesmo mês de coleta das amostras de gasolinas comerciais das distribuidoras. As análises realizadas foram realizadas foram:

- Teor de etanol, segundo a ABNT NBR 13992 (ASSOCIAÇÃO..., 2003);
- Massa específica, pelo Densímetro eletrônico. DMA 4500, da ©*Anton Paar*;
- Temperatura de 10%, 50% e 90% gasolina destilada, ponto final de ebulição e resíduo de destilação, pelo Destilador automático. AD-6, da ©*Tanaka Scientific Limited*;
- Índice de MON, RON, IAD, quantidade hidrocarbonetos saturados, olefínicos, aromáticos e de benzeno, por infravermelho pelo Irox 2000, da ©*Grabner Instruments*.

Todas as análises foram realizadas segundo as especificações da Resolução nº 57 de 2011 da ANP e as Portarias nº 143 e 678 do MAPA. Relembrando que, após o mês de setembro de 2011, a percentagem de AEAC na gasolina comercial brasileira C passou a ser de 20±1%. Por isso, as amostras coletadas de julho de 2011 a setembro de 2011 foram denominadas de Banco I e as amostras coletadas de novembro 2011 a abril de 2012 de Banco II. No mês de outubro de 2011 não houve coleta de amostras.

Desta forma, os resultados de cada amostra dos dois bancos foram comparados com as especificações da ANP e do MAPA, tabela 1 (AGÊNCIA..., 2013a; BRASIL, 2013a; BRASIL, 2013b). Caso houvesse um ou mais parâmetros diferentes dos especificados, a amostra era classificada como não conforme e seu código era acrescido por '–NC'. Por exemplo, a amostra 1107231 possui um valor de MON de 81,8, por isso é classificada como não conforme e seu código passa a ser 1107207–NC. Se a amostra estiver em concordância com os parâmetros da tabela 1 (AGÊNCIA..., 2013a e BRASIL, 2013a; BRASIL, 2013b), é classificada como conforme e seu código é acrescido por '–C'.

### *5.2 Classificação das Amostras Quanto à Conformidade*

Os resultados das análises físico-químicas estão apresentados nas tabelas em anexo deste trabalho (arquivo Ensaios_Físico-químicos.xlsx).

Como se pode observar, a maioria das amostras não conformes possui apenas o parâmetro AEAC em discordância com a tabela 1 (AGÊNCIA..., 2013a e BRASIL, 2013a; BRASIL, 2013b). Deste modo, fez-se necessária a adulteração das amostras conformes selecionadas com os solventes descritos na tabela 4. Por conseguinte, as amostras adulteradas possuem um novo código, substituindo o '–C' pelo '–A'. Apesar de terem esta nova classificação, as amostras adulteradas parte do grupo das amostras não conformes.

Por fim, para as amostras adulteradas provindas da mesma matriz de gasolina, seu código é distinto das demais e está descrito na tabela 4. Lembrando que a amostra codificada como JF S 00 é conforme e não possui nenhum adulterante em sua composição.

**Tabela 4** - Características e codificação das amostras provindas de uma mesma matriz de gasolina.

| Solventes | Código do solvente | Percentagem / % | Código da amostra |
|---|---|---|---|
| A-90 | S 01 | 2 | **JF S 01 2%** |
| | | 5 | **JF S 01 5%** |
| | | 10 | **JF S 01 10%** |
| aguarrás | S 02 | 2 | **JF S 02 2%** |
| | | 5 | **JF S 02 5%** |
| | | 10 | **JF S 02 10%** |
| alifático leve DT | S 03 | 2 | **JF S 03 2%** |
| | | 5 | **JF S 03 5%** |
| | | 10 | **JF S 03 10%** |
| alifático leve SB | S 04 | 2 | **JF S 04 2%** |
| | | 5 | **JF S 04 5%** |
| | | 10 | **JF S 04 10%** |
| alifático médio AR | S 05 | 2 | **JF S 05 2%** |
| | | 5 | **JF S 05 5%** |
| | | 10 | **JF S 05 10%** |
| aromático AB-9 | S 06 | 2 | **JF S 06 2%** |
| | | 5 | **JF S 06 5%** |
| | | 10 | **JF S 06 10%** |
| benzeno | S 07 | 2 | **JF S 07 2%** |
| | | 5 | **JF S 07 5%** |
| | | 10 | **JF S 07 10%** |
| Carbosolv A4070 | S 08 | 2 | **JF S 08 2%** |
| | | 5 | **JF S 08 5%** |
| | | 10 | **JF S 08 10%** |
| cicloexano | S 09 | 2 | **JF S 09 2%** |
| | | 5 | **JF S 09 5%** |
| | | 10 | **JF S 09 10%** |
| éter de pertóleo | S 10 | 2 | **JF S 10 2%** |
| | | 5 | **JF S 10 5%** |
| | | 10 | **JF S 10 10%** |
| etilbenzeno | S 11 | 2 | **JF S 11 2%** |
| | | 5 | **JF S 11 5%** |
| | | 10 | **JF S 11 10%** |
| etilenoglicol | S 12 | 2 | **JF S 12 2%** |
| | | 5 | **JF S 12 5%** |
| | | 10 | **JF S 12 10%** |
| heptano | S 13 | 2 | **JF S 13 2%** |
| | | 5 | **JF S 13 5%** |
| | | 10 | **JF S 13 10%** |
| hexano | S 14 | 2 | **JF S 14 2%** |
| | | 5 | **JF S 14 5%** |
| | | 10 | **JF S 14 10%** |
| NAFTA | S 15 | 2 | **JF S 15 2%** |
| | | 5 | **JF S 15 5%** |
| | | 10 | **JF S 15 10%** |
| querosene | S 16 | 2 | **JF S 16 2%** |
| | | 5 | **JF S 16 5%** |
| | | 10 | **JF S 16 10%** |
| solvente para borracha | S 17 | 2 | **JF S 17 2%** |
| | | 5 | **JF S 17 5%** |
| | | 10 | **JF S 17 10%** |
| tolueno | S 18 | 2 | **JF S 18 2%** |
| | | 5 | **JF S 18 5%** |
| | | 10 | **JF S 18 10%** |
| xileno | S 19 | 2 | **JF S 19 2%** |
| | | 5 | **JF S 19 5%** |
| | | 10 | **JF S 19 10%** |

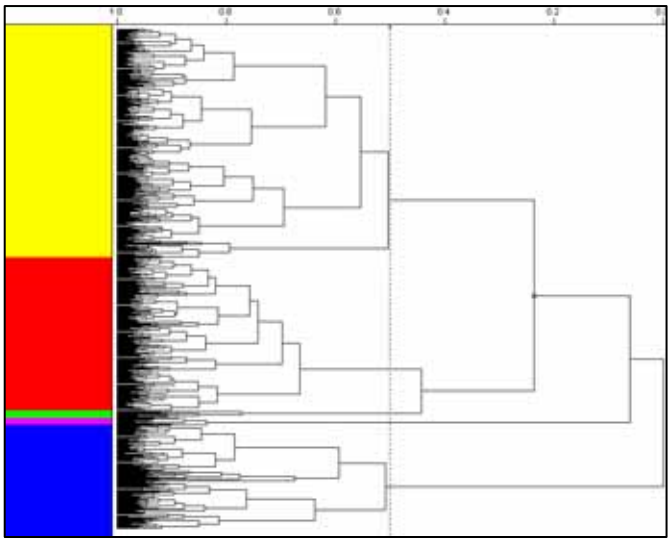
**Fonte**: Elaborado pelo autor (2013).

### *5.3 Seleção de Amostras*

A seleção das amostras foi realizada através da ferramenta quimiométrica HCA, utilizando-se distância Euclidiana, visto que é o cálculo mais apropriado para vetores, mantendo suas características, e conexão incremental (conexão centróide que emprega uma soma de quadrados no cálculo da distância entre grupos), com a finalidade de formar o maior número de agrupamentos possíveis.

Um dendograma característico pode ser observado na figura 30. É possível observar a separação do conjunto de amostras em cerca de 6 subconjuntos, cada qual com cor distinta, obtido através da fixação da similaridade em 0,5. Neste valor, há um equilíbrio entre a obtenção de muitos agrupamentos (obtidos com similaridade menor que 0,5) e pouca distinção entre as amostras (poucos grupos, obtidos com similaridade maior que 0,5).

**Figura 30** - Dendograma das amostras de gasolina comercial brasileira tipo C coletadas em julho de 2011.



**Fonte**: Elaborado pelo autor (2013).

Do total mensal, foram selecionadas 50 amostras. O número de amostras a serem escolhidas por grupo é encontrado através de um cálculo proporcional:

$$amostras\ selecionadas\ no\ grupo = \frac{50.quantidade\ de\ amostras\ no\ grupo}{quantidade\ de\ amostras\ totais\ no\ mês} \quad (78)$$

Por exemplo, no mês de julho de 2011 foram obtidas 584 amostras de diferentes distribuidoras de gasolina comercial. Após a construção do dendograma, encontraram-se 5 grupos. No grupo 1, existem 266 amostras similares. Desta forma, a quantidade de amostras a serem selecionadas será:

$$amostras\ do\ grupo\ 1 = \frac{50.266}{584} = 22{,}7 \approx 23\ amostras \quad (79)$$

Realizando este cálculo, obtêm-se para os grupos 2 (181 amostras), 3 (9 amostras), 4 (8 amostras) e 5 (120 amostras) têm-se $15$, $1$, $1$ e $10$ amostras selecionadas respectivamente. Totalizando em $23 + 15 + 1 + 1 + 10 = 50$ amostras selecionadas no mês de julho de 2011.

De semelhante forma, este cálculo foi realizado para todos os meses, de julho de 2011 a abril de 2012.

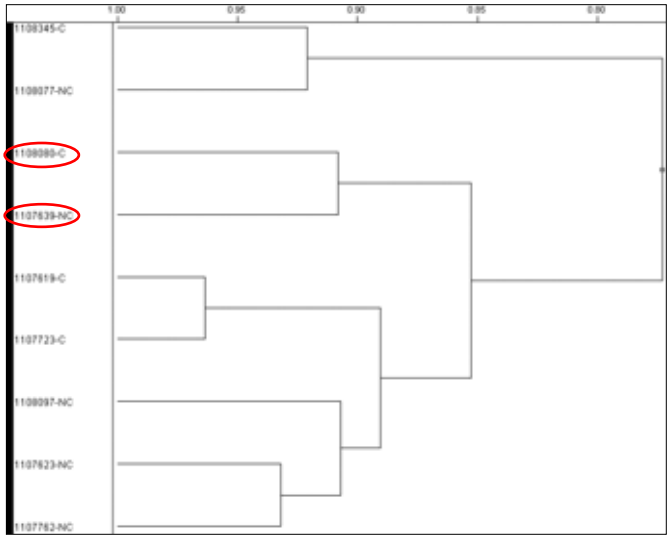
É importante lembrar que as amostras escolhidas são representativas do próprio grupo. Ou seja, as amostras selecionadas possuem as mesmas informações que o grupo como um todo.

Além disso, a seleção foi realizada, de tal modo que fossem obtidas as amostras mais distintas entre si no grupo. Prosseguindo com o exemplo dado acima, o grupo 3 do dendograma do mês de julho de 2011 é mostrado na figura 31. A amostra selecionada será uma das duas destacadas em vermelho. Deste modo, a amostra escolhida é a mais distinta, porém mantém as mesmas características do grupo a que pertencem.

Uma das ferramentas a ser considerada para a seleção das amostras representativas foi a PCA. Contudo, conforme mostra a figura 32, não há uma boa separação de grupos. Isto dificulta a escolha das amostras representativas. Desta forma, por ser uma técnica mais visual, a HCA mostrou-se superior nesta etapa.

Vale ressaltar que, as amostras selecionadas (inclusive as conformes adulteradas) entre os meses de julho de 2011 e setembro de 2011 passaram a se chamar de Banco I e entre novembro de 2011 e abril de 2012, bem como as amostras citadas na tabela 4, de Banco II.

**Figura 31** - Exemplo de seleção de amostras em um mesmo grupo. Dendograma de julho/2011. As amostras selecionadas estão circuladas em vermelho.



**Fonte**: Elaborado pelo autor (2013).

**Figura 32** - Distinção entre HCA (a) e PCA, sendo (b) PC1 e PC2, (c) PC1 e PC3 e (d) PC2 e PC3, ambos de julho de 2011, quanto à seleção de amostras.

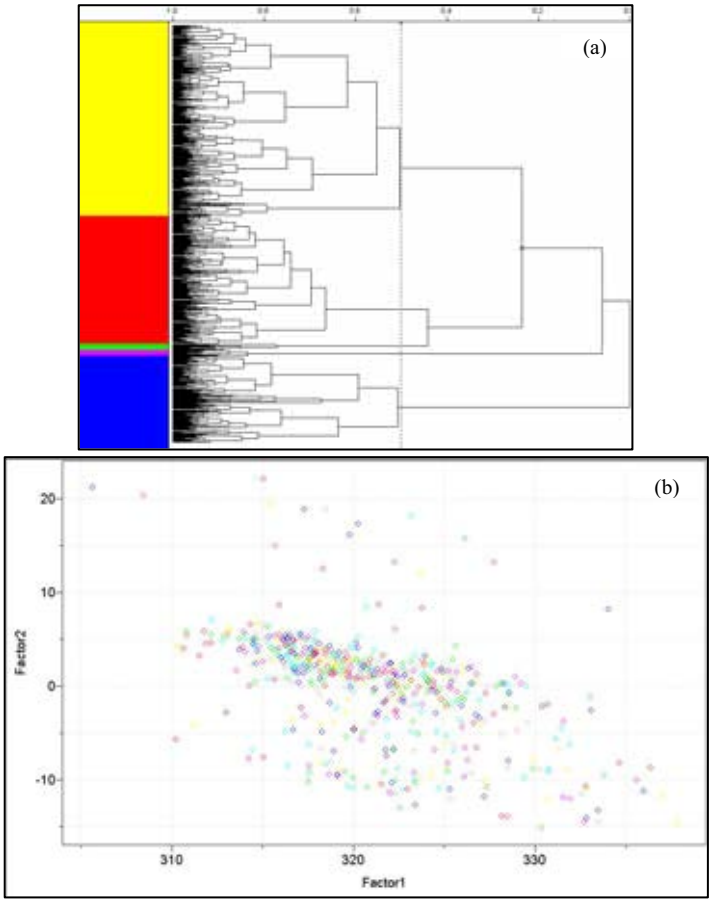
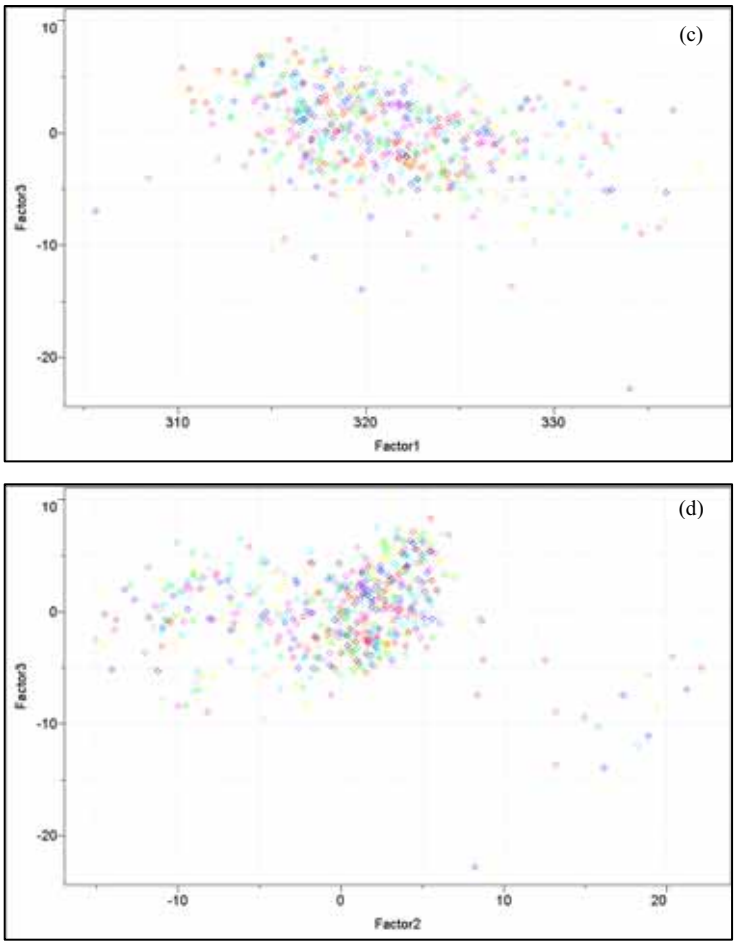


**Figura 32** - Distinção entre HCA (a) e PCA, sendo (b) PC1 e PC2, (c) PC1 e PC3 e (d) PC2 e PC3, ambos de julho de 2011, quanto à seleção de amostras (continuação).




**Fonte**: Elaborado pelo autor (2013).

Por fim, os dendogramas dos meses de julho de 2011 a abril de 2012 estão exibidos na figura 33. As relações das amostras que compõem os Bancos I e II são apresentadas nos quadros 3 e 4 respectivamente.

**Figura 33** - Dendogramas obtidos através da coleta mensal de amostras de gasolinas comerciais brasileiras do tipo C na região centro-oeste do estado de São Paulo, sendo (a) agosto/2011, (b) setembro/2011, (c) novembro/2011, (d) dezembro/2011, (e) janeiro/2012, (f) fevereiro/2012, (g) março/2012, (h) abril/2012.

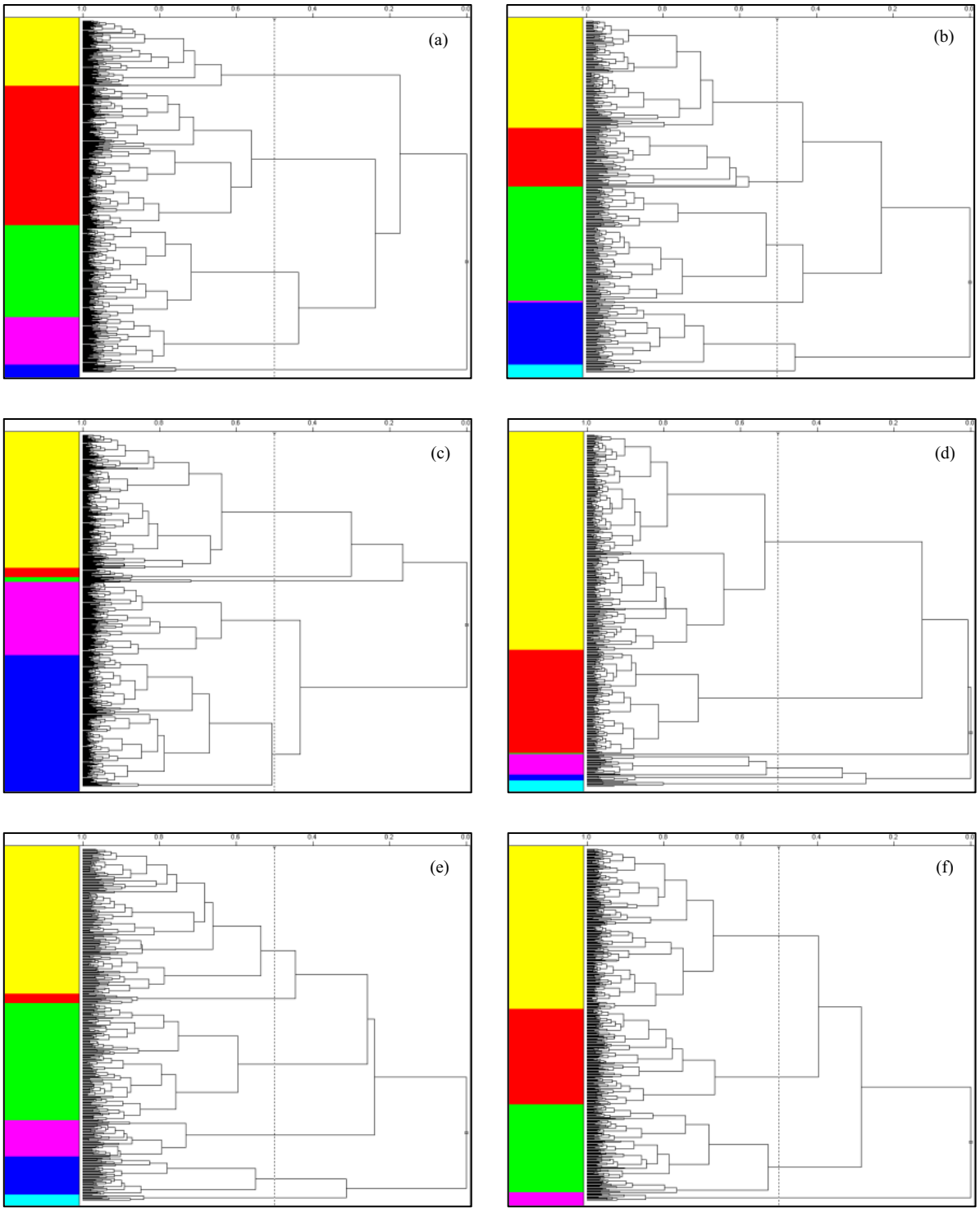
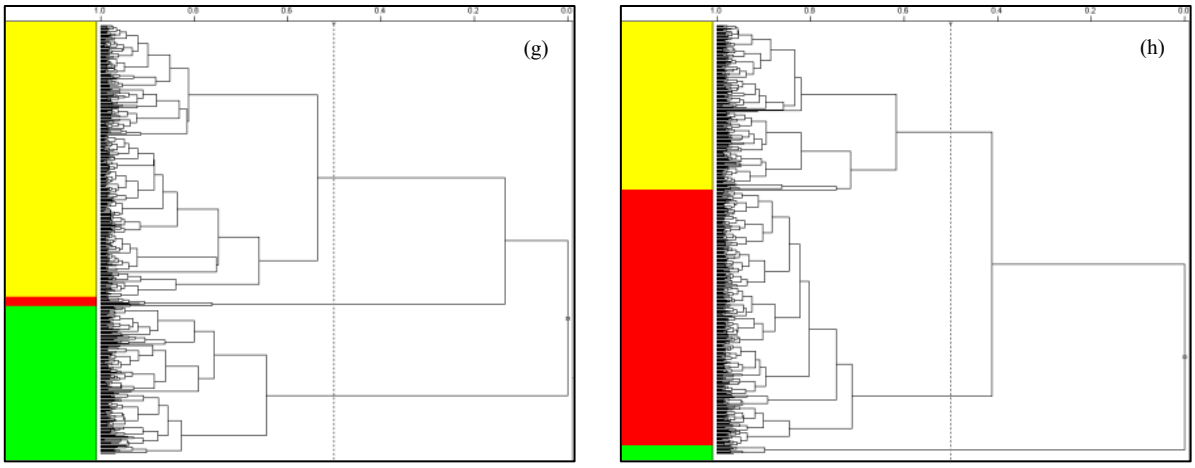


**Figura 33** - Dendogramas obtidos através da coleta mensal de amostras de gasolinas comerciais brasileiras do tipo C na região centro-oeste do estado de São Paulo, sendo (a) agosto/2011, (b) setembro/2011, (c) novembro/2011, (d) dezembro/2011, (e) janeiro/2012, (f) fevereiro/2012, (g) março/2012, (h) abril/2012 (continuação).




**Fonte**: Elaborado pelo autor (2013).

**Quadro 3** - Relação das amostras do Banco I, com 25±1% de AEAC e selecionadas de julho/2011 a setembro/2011.

| Banco I | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1107204-C | 1108070-NC | 1108892-NC | 1109488-C | 1110283-NC | 1110807-NC | 1107738-A | 1110138-A |
| 1107240-C | 1108097-NC | 1108925-NC | 1109514-NC | 1110299-NC | 1110810-C | 1107996-A | 1110207-A |
| 1107247-C | 1108140-C | 1108968-NC | 1109520-C | 1110308-C | 1110854-C | 1108140-A | 1110222-A |
| 1107249-NC | 1108160-NC | 1108993-C | 1109581-NC | 1110318-NC | 1110884-C | 1108311-A | 1110308-A |
| 1107259-C | 1108178-NC | 1109029-NC | 1109646-NC | 1110322-NC | 1110908-NC | 1108383-A | 1110513-A |
| 1107274-C | 1108192-NC | 1109042-C | 1109670-C | 1110392-C | 1110930-C | 1108540-A | 1110604-A |
| 1107280-C | 1108207-NC | 1109044-NC | 1109679-NC | 1110430-NC | 1110934-NC | 1108576-A | 1110665-A |
| 1107304-NC | 1108228-NC | 1109047-NC | 1109814-NC | 1110446-C | 1110956-NC | 1108803-A | 1110671-A |
| 1107404-C | 1108245-NC | 1109077-NC | 1109861-C | 1110458-NC | 1110964-C | 1108993-A | 1110714-A |
| 1107451-C | 1108257-C | 1109081-C | 1109873-C | 1110473-NC | 1110974-C | 1109119-A | 1110734-A |
| 1107489-C | 1108269-NC | 1109119-C | 1109883-C | 1110507-NC | 1110990-C | 1109143-A | 1110776-A |
| 1107494-NC | 1108311-C | 1109143-C | 1109891-C | 1110509-NC | 1111003-NC | 1109339-A | 1110803-A |
| 1107497-NC | 1108345-C | 1109196-NC | 1109908-C | 1110513-C | 1111011-NC | 1109352-A | 1110810-A |
| 1107540-NC | 1108351-NC | 1109214-NC | 1109928-C | 1110552-NC | 1111037-NC | 1109377-A | 1110854-A |
| 1107543-C | 1108367-NC | 1109248-NC | 1109932-NC | 1110604-C | 1111047-NC | 1109440-A | 1110884-A |
| 1107580-C | 1108383-C | 1109309-NC | 1109946-C | 1110641-NC | 1111059-C | 1109488-A | 1110930-A |
| 1107607-C | 1108442-NC | 1109314-C | 1109954-NC | 1110665-C | 1111079-C | 1109520-A | 1110976-A |
| 1107621-NC | 1108540-C | 1109324-C | 1109956-NC | 1110671-C | 1111085-NC | 1109670-A | 1110990-A |
| 1107720-NC | 1108576-C | 1109339-C | 1110035-NC | 1110674-C | 1107204-A | 1109861-A | 1111059-A |
| 1107726-C | 1108594-C | 1109341-NC | 1110105-C | 1110705-NC | 1107240-A | 1109883-A | 1111079-A |
| 1107738-C | 1108635-NC | 1109352-C | 1110121-C | 1110714-C | 1107247-A | 1109891-A | |
| 1107755-NC | 1108637-C | 1109374-NC | 1110138-C | 1110734-C | 1107280-A | 1109908-A | |
| 1107787-NC | 1108644-NC | 1109375-NC | 1110207-C | 1110736-NC | 1107451-A | 1109928-A | |
| 1107911-NC | 1108652-NC | 1109377-C | 1110222-C | 1110773-NC | 1107489-A | 1109946-A | |
| 1107996-C | 1108803-C | 1109440-C | 1110247-NC | 1110776-C | 1107580-A | 1110105-A | |
| 1108017-NC | 1108825-NC | 1109452-NC | 1110254-NC | 1110803-C | 1107726-A | 1110121-A | |

**Fonte**: Elaborado pelo autor (2013).

**Quadro 4** - Relação das amostras do Banco II, com 20±1% de AEAC e selecionadas de novembro/2011 a abril/2012, incluindo as amostras adulteradas e o branco da mesma matriz de gasolina.

| Banco II | | | | | | | |
|---|---|---|---|---|---|---|---|
| **1113508-NC** | **1114966-C** | **120590-NC** | **121907-NC** | **123365-C** | **1113250-A** | **122748-A** | **JF S 03 10%** |
| **1113561-NC** | **1114970-NC** | **120618-C** | **121922-NC** | **123400-NC** | **1113287-A** | **122777-A** | **JF S 04 2%** |
| **1113569-C** | **1114994-C** | **120633-C** | **121938-NC** | **123405-NC** | **1113374-A** | **122790-A** | **JF S 04 5%** |
| **1113578-C** | **1115080-C** | **120705-NC** | **121988-NC** | **123417-NC** | **1113569-A** | **122813-A** | **JF S 04 10%** |
| **1113593-NC** | **1115105-NC** | **120736-NC** | **122070-NC** | **123435-C** | **1113578-A** | **122822-A** | **JF S 05 2%** |
| **1113631-NC** | **1115109-C** | **120737-NC** | **122077-NC** | **123489-C** | **1113655-A** | **123022-A** | **JF S 05 5%** |
| **1113655-C** | **1115130-NC** | **120758-NC** | **122084-NC** | **123524-NC** | **1113664-A** | **123077-A** | **JF S 05 10%** |
| **1113664-C** | **1115140-NC** | **120777-C** | **122092-NC** | **123537-C** | **1113706-A** | **123080-A** | **JF S 06 2%** |
| **1113706-C** | **1115175-NC** | **120780-C** | **122102-NC** | **123576-NC** | **1113717-A** | **123097-A** | **JF S 06 5%** |
| **1113717-C** | **1115182-NC** | **120782-NC** | **122292-C** | **123579-NC** | **1113833-A** | **123153-A** | **JF S 06 10%** |
| **1113819-NC** | **1115204-C** | **120794-NC** | **122321-NC** | **123584-NC** | **1113855-A** | **123216-A** | **JF S 07 2%** |
| **1113833-C** | **1115228-C** | **120795-NC** | **122348-NC** | **123585-NC** | **1113903-A** | **123225-A** | **JF S 07 5%** |
| **1113835-NC** | **1115237-NC** | **120802-NC** | **122350-NC** | **123620-C** | **1113977-A** | **123251-A** | **JF S 07 10%** |
| **1113851-C** | **1115248-NC** | **120831-C** | **122354-NC** | **123628-NC** | **1113979-A** | **123255-A** | **JF S 08 2%** |
| **1113855-C** | **1115258-C** | **120850-C** | **122372-NC** | **123707-NC** | **1113988-A** | **123329-A** | **JF S 08 5%** |
| **1113884-C** | **1115265-NC** | **120853-C** | **122380-NC** | **123710-C** | **1114019-A** | **123362-A** | **JF S 08 10%** |
| **1113903-C** | **1115267-C** | **120856-C** | **122386-C** | **123719-C** | **1114101-A** | **123365-A** | **JF S 09 2%** |
| **1113944-NC** | **1115271-NC** | **120871-C** | **122395-C** | **123785-C** | **1114124-A** | **123400-A** | **JF S 09 5%** |
| **1113956-NC** | **1115309-C** | **120892-C** | **122411-NC** | **123833-NC** | **1114178-A** | **123405-A** | **JF S 09 10%** |
| **1113971-NC** | **1115324-NC** | **120899-C** | **122427-NC** | **123855-NC** | **1114413-A** | **123435-A** | **JF S 10 2%** |
| **1113977-C** | **1115327-NC** | **120934-NC** | **122459-NC** | **123862-C** | **1114484-A** | **123489-A** | **JF S 10 5%** |
| **1113979-C** | **1115359-C** | **120937-C** | **122474-NC** | **123871-C** | **1114538-A** | **123537-A** | **JF S 10 10%** |
| **1113988-C** | **1115386-C** | **120942-C** | **122491-C** | **123875-NC** | **1114600-A** | **123620-A** | **JF S 11 2%** |
| **1114012-NC** | **1115394-NC** | **120971-C** | **122563-NC** | **124004-NC** | **1114649-A** | **123710-A** | **JF S 11 5%** |
| **1114019-C** | **1115396-C** | **120979-C** | **122576-NC** | **124062-C** | **1114700-A** | **123719-A** | **JF S 11 10%** |
| **1114061-NC** | **1115398-NC** | **120995-C** | **122595-C** | **124140-C** | **1114752-A** | **123785-A** | **JF S 12 2%** |
| **1114101-C** | **1115456-C** | **121055-C** | **122614-C** | **124183-C** | **1114878-A** | **123862-A** | **JF S 12 5%** |
| **1114124-C** | **1115490-NC** | **121060-NC** | **122635-NC** | **124214-C** | **1114889-A** | **123871-A** | **JF S 12 10%** |
| **1114167-NC** | **1115492-NC** | **121080-C** | **122641-NC** | **124239-C** | **1114907-A** | **124062-A** | **JF S 13 2%** |
| **1114178-C** | **1115504-NC** | **121183-C** | **122653-NC** | **124244-C** | **1114966-A** | **124140-A** | **JF S 13 5%** |
| **1114288-NC** | **120034-NC** | **121242-C** | **122660-NC** | **124294-C** | **1114994-A** | **124183-A** | **JF S 13 10%** |
| **1114317-NC** | **120064-NC** | **121246-NC** | **122686-C** | **124319-NC** | **1115109-A** | **124214-A** | **JF S 14 2%** |
| **1114332-NC** | **120095-NC** | **121300-NC** | **122730-C** | **124329-NC** | **1115204-A** | **124239-A** | **JF S 14 5%** |
| **1114341-NC** | **120126-C** | **121325-C** | **122748-C** | **124334-C** | **1115228-A** | **124244-A** | **JF S 14 10%** |
| **1114413-C** | **120128-C** | **121332-C** | **122762-NC** | **124364-C** | **121041-A** | **124294-A** | **JF S 15 2%** |
| **1114417-NC** | **120161-NC** | **121388-NC** | **122777-C** | **124369-NC** | **121044-A** | **124334-A** | **JF S 15 5%** |
| **1114469-NC** | **120184-NC** | **121421-C** | **122790-C** | **124376-C** | **121055-A** | **124364-A** | **JF S 15 10%** |
| **1114481-NC** | **120194-NC** | **121452-C** | **122791-NC** | **124393-C** | **121468-A** | **124376-A** | **JF S 16 2%** |
| **1114484-C** | **120200-NC** | **121460-C** | **122813-C** | **124434-C** | **121559-A** | **124393-A** | **JF S 16 5%** |
| **1114493-NC** | **120212-C** | **121468-C** | **122816-NC** | **124448-NC** | **121610-A** | **124434-A** | **JF S 16 10%** |
| **1114538-C** | **120222-NC** | **121505-NC** | **122822-C** | **124460-NC** | **121611-A** | **124481-A** | **JF S 17 2%** |
| **1114549-NC** | **120235-C** | **121532-NC** | **122870-NC** | **124481-C** | **121657-A** | **124489-A** | **JF S 17 5%** |
| **1114568-NC** | **120249-C** | **121559-C** | **122971-NC** | **124485-NC** | **121663-A** | **124498-A** | **JF S 17 10%** |
| **1114595-NC** | **120282-C** | **121610-C** | **123022-C** | **124489-C** | **121790-A** | **124531-A** | **JF S 18 2%** |
| **1114600-C** | **120291-C** | **121611-C** | **123077-C** | **124498-C** | **121808-A** | **124546-A** | **JF S 18 5%** |
| **1114649-C** | **120294-C** | **121631-NC** | **123080-C** | **124517-NC** | **121843-A** | **124573-A** | **JF S 18 10%** |
| **1114653-NC** | **120305-NC** | **121657-C** | **123090-NC** | **124531-C** | **122169-A** | **124593-A** | **JF S 19 2%** |
| **1114700-C** | **120341-C** | **121660-NC** | **123097-C** | **124543-NC** | **122194-A** | **JF S 00** | **JF S 19 5%** |
| **1114752-C** | **120360-C** | **121663-C** | **123146-NC** | **124545-NC** | **122292-A** | **JF S 01 2%** | **JF S 19 10%** |
| **1114797-NC** | **120377-C** | **121679-NC** | **123153-C** | **124546-C** | **122386-A** | **JF S 01 5%** | |
| **1114872-NC** | **120428-C** | **121736-NC** | **123216-C** | **124573-C** | **122395-A** | **JF S 01 10%** | |
| **1114876-NC** | **120486-NC** | **121757-NC** | **123225-C** | **124593-C** | **122491-A** | **JF S 02 2%** | |
| **1114878-C** | **120515-C** | **121790-C** | **123251-NC** | **124613-NC** | **122595-A** | **JF S 02 5%** | |
| **1114889-C** | **120533-NC** | **121808-C** | **123255-C** | **1113025-A** | **122614-A** | **JF S 02 10%** | |
| **1114899-C** | **120566-C** | **121843-C** | **123329-NC** | **1113140-A** | **122686-A** | **JF S 03 2%** | |
| **1114907-C** | **120587-NC** | **121859-NC** | **123362-C** | **1113225-A** | **122730-A** | **JF S 03 5%** | |

**Fonte**: Elaborado pelo autor (2013).

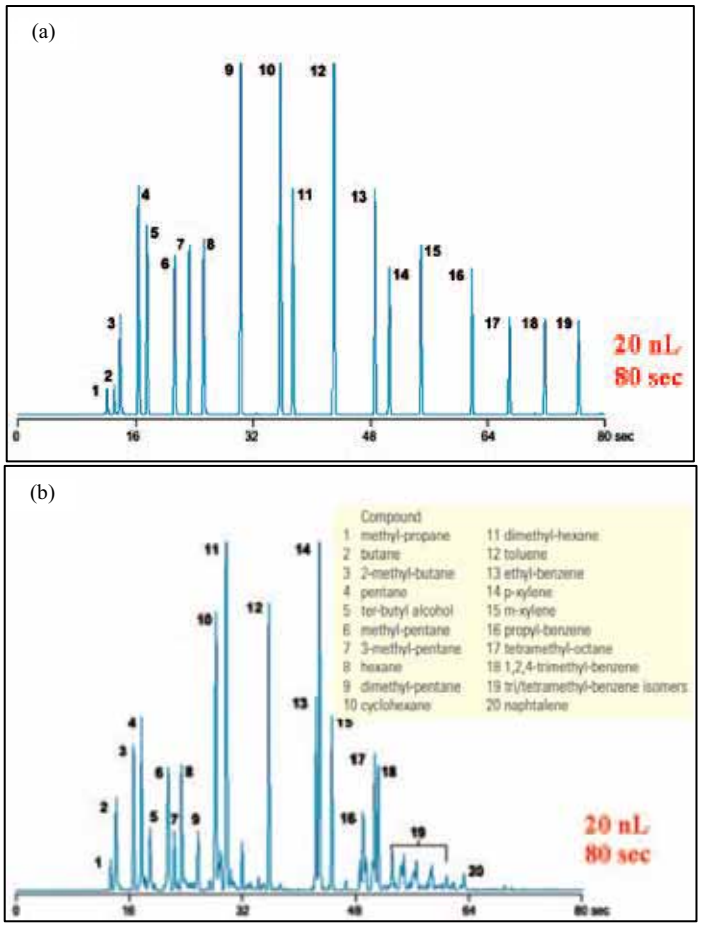
## *5.4 Análise Cromatográfica Ultrarrápida*

Após a devida seleção, foram avaliadas as condições cromatográficas necessárias para a análise das amostras, sendo estas executadas para todas as amostras de gasolina comercial brasileira do tipo C.

### 5.4.1 Avaliação das Condições Cromatográficas

Utilizou-se como base para encontrar os parâmetros a *Application Note* 10131 da *Thermo Fisher Scientific Inc* de autoria de CADOPPI e FACCHETTI. Neste trabalho, é feita a separação cromatográfica de uma mistura de hidrocarbonetos, desde propano até n-pentadecano, como mostra a figura 34. A análise foi feita em um cromatógrafo a gás com modelo ultrarrápido acoplado, cuja coluna capilar era RTX-1 $1{,}5\text{m} \times 0{,}32\text{mm} \times 0{,}1\text{mm}$, injetor *split*/*splitless*, mantido a 225ºC, detector FID, mantido a 320ºC, e usando hélio como gás de arraste, com 0,5mL $\text{min}^{-1}$ a fluxo constante. A corrida cromatográfica foi realizada com temperatura programada de 40ºC, mantendo por 6s, até 300º, mantendo também por 6s, com taxa de 180ºC $\text{min}^{-1}$. O volume de amostra injetado foi de 20nL e a taxa de *split* foi de 1:1000. Duas análises foram realizadas neste trabalho: a primeira foi mistura padrão, descrita pela ASTM D3710 de 2009, e a segunda foi a amostra não diluída de gasolina (CADOPPI, 2004) (AMERICAN..., 2009). Ambas estão representadas na figura 34.

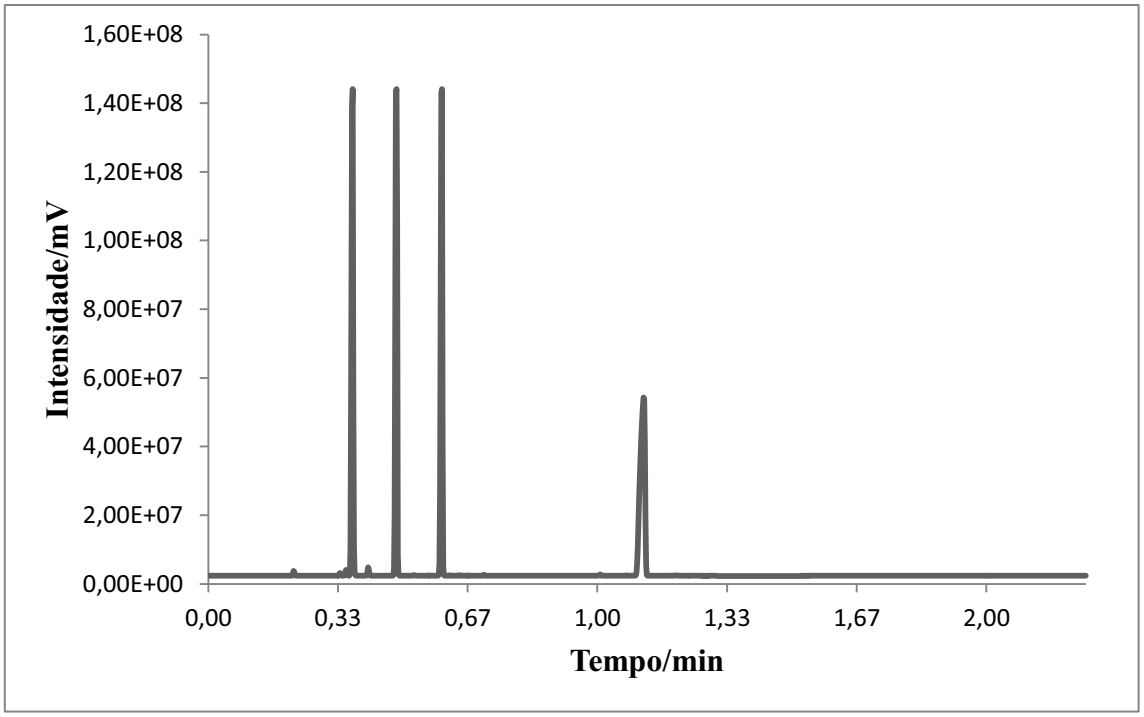
**Figura 34** - Separação cromatográfica da mistura padrão descrita pela ASTM D3710 (a) e de uma amostra de gasolina não diluída (b) obtidas no módulo *Ultra Fast* (injeção de 20nL).




**Fonte**: CADOPPI, A.; FACCHETTI, R., 2004.

Baseado neste método, realizou-se uma análise cromatográfica ultrarrápida de uma mistura de padrões de hidrocarbonetos: n-hexano, n-heptano, n-octano, n-decano. As condições foram as mesmas da *Application Note*, exceto que o volume de amostra injetado foi de 0,1µL, a coluna utilizada foi a PH5, e o gás de arraste utilizado foi o hidrogênio. A figura 35 mostra a análise feita destes padrões.

**Figura 35** - Análise cromatográfica ultrarrápida dos padrões n-hexano (0,370min), n-heptano (0,482min), n-octano (0,598min) e n-decano (1,113min).




**Fonte**: Elaborado pelo autor (2013).

Para análise cromatográfica de gasolina, selecionou-se uma amostra aleatória conforme (14889-C), não diluída, no sentido de avaliar os parâmetros cromatográficos. Embora, tenha havido uma boa resolução entre os picos dos padrões, o mesmo não foi observado para a amostra de gasolina. Como mostra a figura 36(a).

Assim, alguns parâmetros foram alterados. A taxa de Split foi reduzida de 1:1000 para 1:500, a taxa de aquecimento foi reduzida de 180ºC $min^{-1}$ para 100ºC $min^{-1}$. A temperatura final também foi reduzida: de 300ºC, mantida por 6s, para 250ºC, por 9s. Esta redução foi feita, ao se perceber que a todos os constituintes do padrão eram separados antes de se alcançar a temperatura final de 300ºC. A fim de se evitar gradientes de temperatura, tanto o FID quanto o injetor foram mantidos a 270ºC.

Feitos estes ajustes, obteve-se o cromatograma da amostra conforme, como mostrado na figura 36(b). Nota-se uma melhor resolução entre os picos após as modificações descritas acima. O tempo total de análise foi de 2,85 minutos (171 segundos).

Portanto, as condições cromatográficas encontradas (também descritas na seção 4.6) são citadas a seguir.

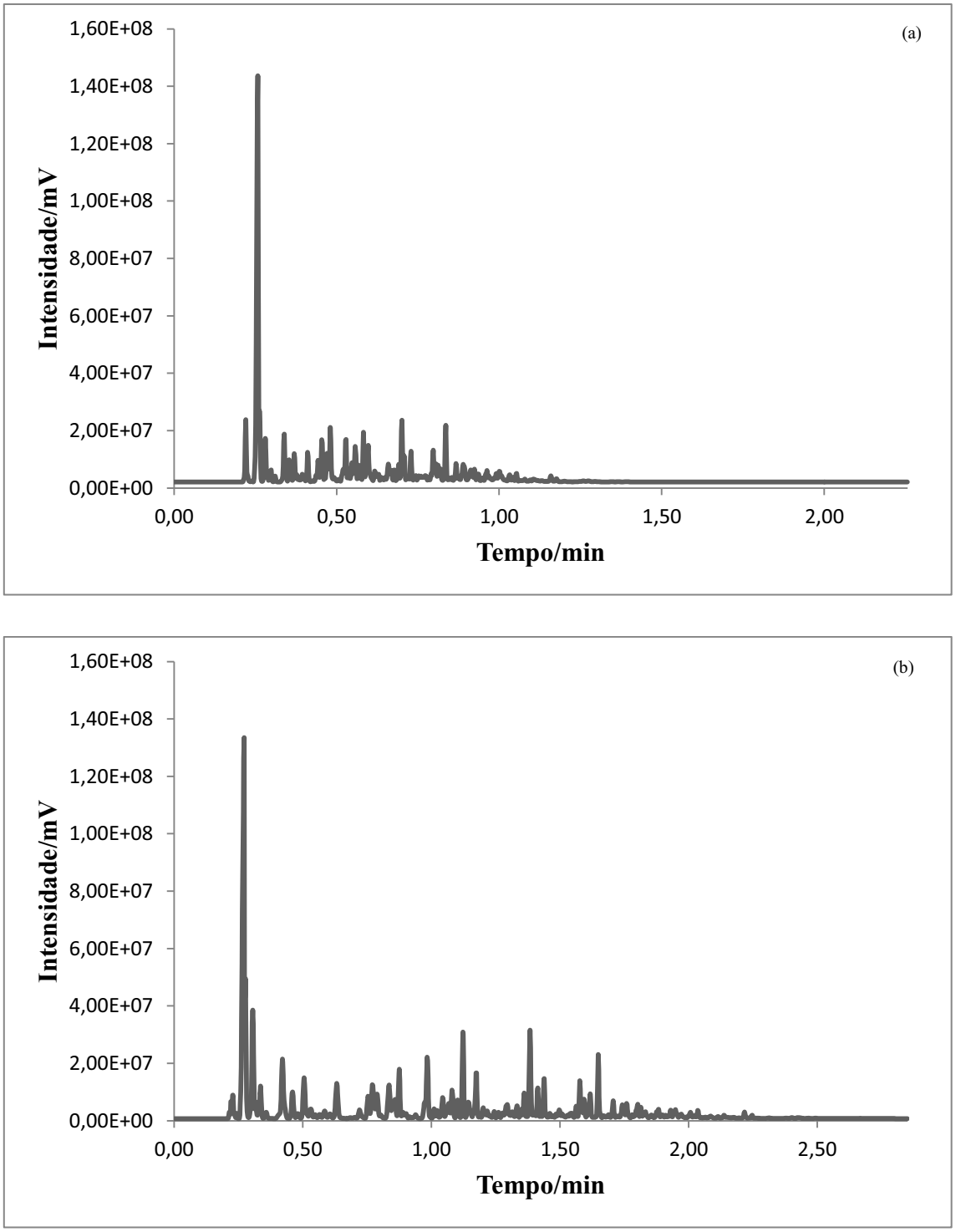
- Coluna capilar PH5 da *Thermo Fisher Scientific Inc*. Dimensões: 5,00m de comprimento, 0,40μm de espessura de filme, 0,10mm de diâmetro interno;
- $H_2$ como gás de arraste, com fluxo de 0,5 mL $min^{-1}$;
- Detector FID com temperatura fixada em 270ºC e os gases utilizados foram o $H_2$ a 45mL $min^{-1}$, o $N_2$ como make-up a 30mL $min^{-1}$ e o ar sintético a 350mL $min^{-1}$;
- Injetor *Split*/*Splitless* com temperatura fixada em 270ºC;
- Injeção de 0,1μL de amostra através de uma seringa com capacidade de 10μL, com taxa de *Split* de 1:500.

Durante as análises cromatográficas, a temperatura inicial do forno foi de 40 ºC, mantida por 0,60 min, e elevada até 250ºC, mantido por 0,15min, em uma taxa de 100 ºC $min^{-1}$. O fluxo de gás foi mantido constante em 0,5 mL $min^{-1}$ e a taxa de *split* foi de 1:500.

Com estas condições em mãos, realizaram-se as análises cromatográficas para todas as amostras. Por se tratar de uma matriz muito complexa e o intuito deste trabalho é a análise qualitativa da gasolina comercial brasileira do tipo C, não serão identificados todos os picos.

Para melhor visualização da figura 36, ver o anexo Comparação_entre_Parâmetros_Cromatográficos.xlsx, o qual compara os perfis cromatográficos com os parâmetros da Application Note e os otimizados descritos acima. Os cromatogramas das amostras de gasolina estão no anexo Cromatogramas_das_Amostras_de_Gasolinas.xlsx (Amostras com asterisco apresentaram perfis cromatográficos distinto dos demais).

**Figura 36** - Análise cromatográfica ultrarrápida de uma amostra aleatória conforme com (a) parâmetros da *Application Note* e com (b) parâmetros ajustados para máxima resolução entre constituintes.




**Fonte**: Elaborado pelo autor (2013).

### *5.5 Alinhamento dos Picos Cromatográficos*

Para o alinhamento dos picos cromatográficos, foi utilizado o software MATLAB®. O comando utilizado está descrito a seguir.

$$[Warping, XWarped, Diagnos] = cow(T, X, Seg, Slack, Options)$$

Em que $X$ é a matriz de dados com 643 amostras e 1712 variáveis, $XWarped$ é a matriz de dados alinhada, $T$ é a matriz contendo apenas as 1712 variáveis, $Seg$ é o tamanho do segmento, $Slack$ é o tamanho do *slack*, $Diagnos$ e $Options$ são as opções para uso no MATLAB® e, por isso, não serão descritos aqui (nas linhas abaixo, $Options$ está representado por [], que significa não alteração das opções).

Antes de se calcular a matriz $XWarped$, precisou-se definir os valores de $Seg$ e $Slack$. Para isso, realizou-se um teste com diferentes valores de tamanho de segmento, mantendo o tamanho de *slack* fixo. O comando inserido foi:

(1) for $\mathrm{i} = 1{:}\,20$;

(2) $[Warping, XWarped, Diagnos] = cow(X(:,2)', X(:,2{,}644)', 50, \mathrm{i}, [])$;

(3) $\mathrm{s} = \mathrm{svd}(XWarped, 0)$;

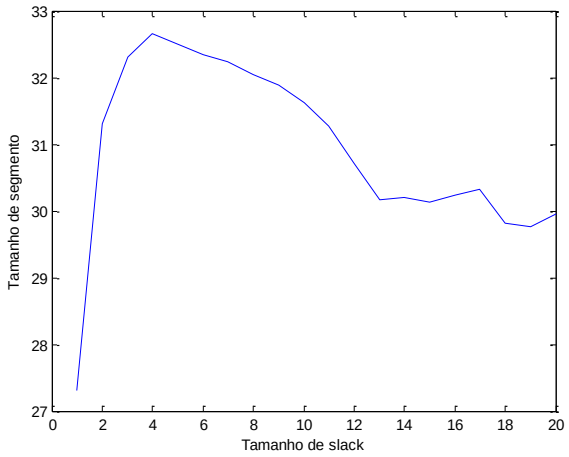
(4) $sv(i,:) = \mathrm{s}./\mathrm{sum}(s) * 100$;

(5) $\mathrm{plot}(\mathrm{sum}(\mathrm{sv}(:,1{:}\,2)'))$;

(6) end

Em que $X(:,2)'$ representa apenas as variáveis, $X(:,2{,}644)'$ representa o conjunto de dados completo, $s$ é o desvio padrão. Assim, foi obtido um gráfico de tamanho de segmento por tamanho de slack, como mostra a figura 37.

**Figura 37** - Pesquisa dos valores ótimos de tamanho de segmento e de tamanho de *slack* para alinhamento dos cromatogramas.




**Fonte**: Elaborado pelo autor (2013).

Do gráfico, é possível perceber que os tamanhos ótimos de segmento e de slack são, respectivamente, iguais a 4 e 33. Com estes dados em mãos, realizou-se um novo cálculo:

(1) $[Warping, XWarped, Diagnos] = cow(X(:,2)', X(:,2{,}644)', 33{,}4, [])$;

(2) plot$(XWarped')$.

Deste modo, os novos valores dos cromatogramas alinhados são dados pela matriz $XWarped$. A figura 38 mostra alguns picos antes e após o alinhamento pelo algoritmo COW.

**Figura 38** - Cromatogramas das amostras de gasolina antes (a) e depois (b) do alinhamento pelo algortimo COW.

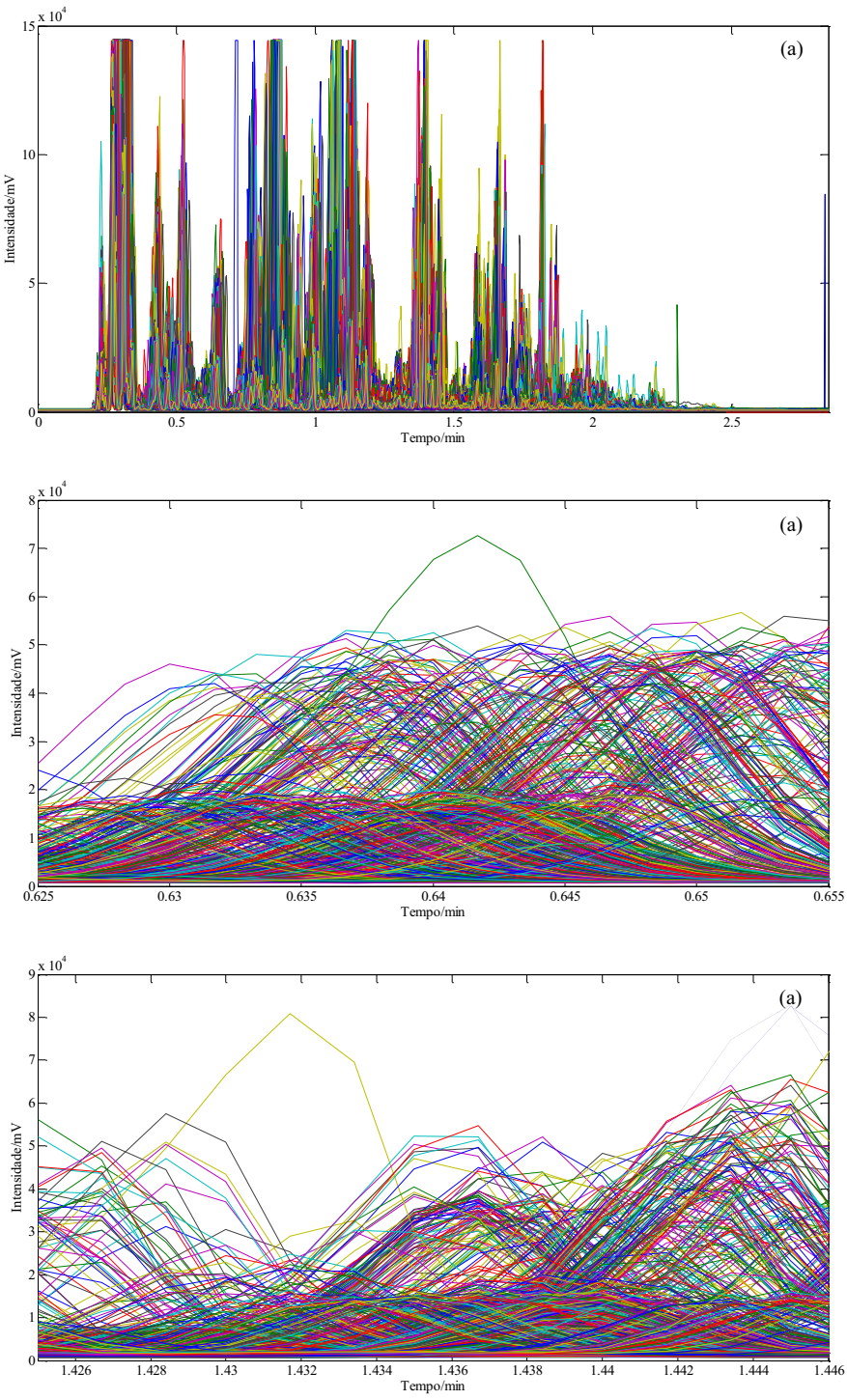
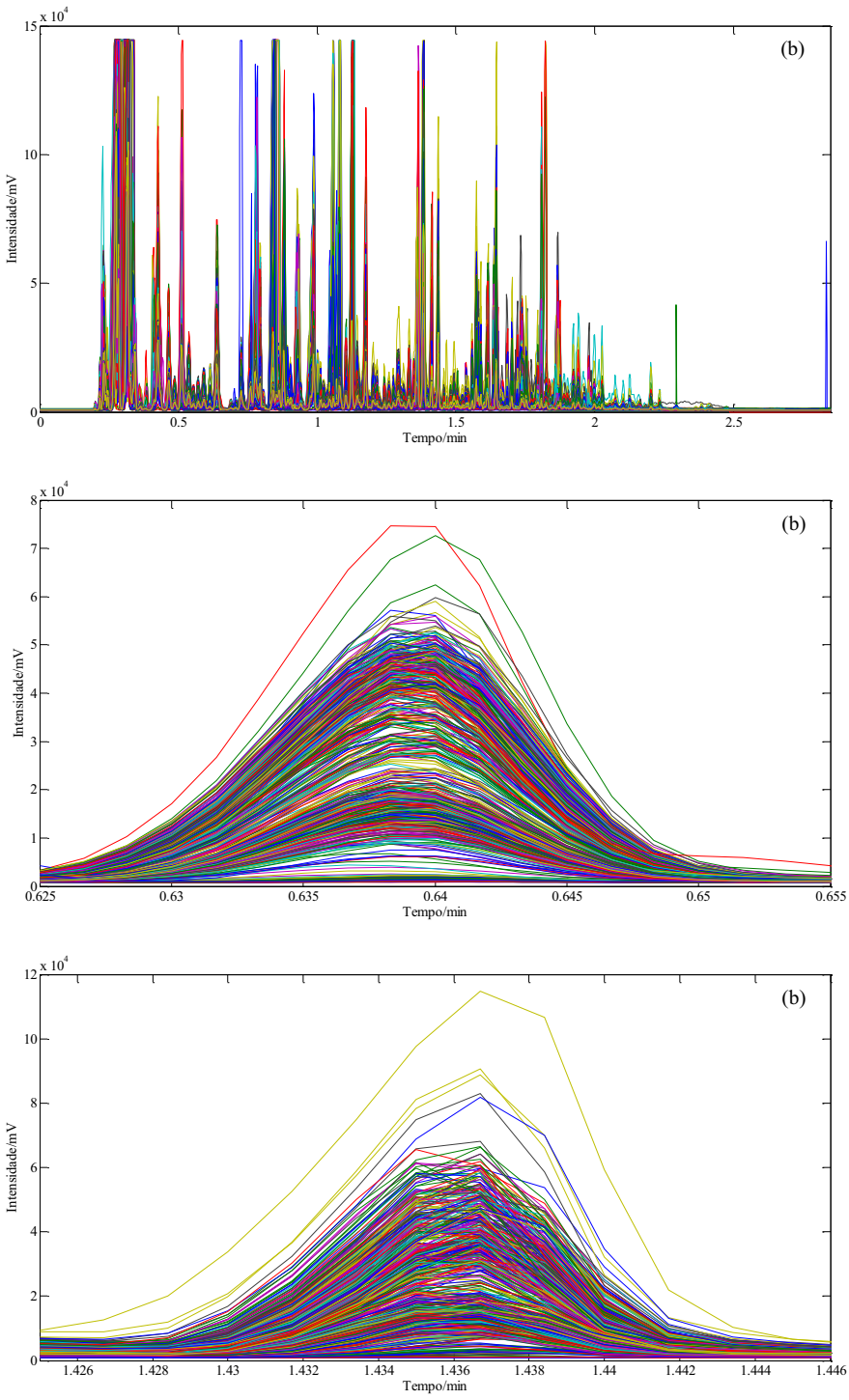


**Figura 38** - Cromatogramas das amostras de gasolina antes (a) e depois (b) do alinhamento pelo algortimo COW (continuacão).




**Fonte**: Elaborado pelo autor (2013).

## *5.6 Análise de Componentes Principais dos Bancos I e II*

Antes de se realizar a classificação pelo método quimiométrico SIMCA, fez-se necessária a investigação dos pré-processamentos e transformadas ótimos. Para isso, realizou-se a Análise de Componentes Principais, para os Bancos I (amostras com 25% de AEAC e coletadas de julho/2011 a setembro/2011) e II (amostras com 20% de AEAC e coletadas de novembro/2011 a abril/2012 e, além das amostras descritas no tabela 4), e averiguaram-se quais pré-tratamentos corresponderiam a uma melhor descrição do respectivo banco de dados e com o menor número de componentes principais.

### 5.6.1 Avaliação do Pré-processamento e Transformação para o Banco I

Para o Banco I, realizou-se a PCA, verificando a descrição do conjunto de dados por 3, 7 e 10 componentes principais, com os seguintes pré-tratamentos:

- Pré-processamento: nenhum, auto-escalado, escalado pela média, escalado pela amplitude e escalado pela variância;
- Transformadas: nenhuma, primeira e segunda derivadas, logaritmo na base 10, MSC, SNV e *smoothing*.

Os resultados estão descritos na tabela 5.

Em seguida, fez-se uma nova PCA com validação cruzada, removendo 5 vetores de amostras para cada análise, com os cinco pré-tratamentos que melhor descreveram a matriz original de dados. As tabelas 6 e 7 fornecem o resultado dessa análise.

Pela tabela 5, é possível visualizar que o conjunto de dados pode ser descrito por apenas 3 componentes principais sem a necessidade de pré-tratamento. Isto se deve ao fato da realização do alinhamento dos cromatogramas, descrito na seção 5.5. Além disso, o pré-processamento de melhor resultado foi o escalamento pela variância, uma vez que esse dá a mesma influência no modelo para todas as variáveis.

Já as transformadas ótimas foram por logaritmo na base 10, por SNV e por *smoothing*. Conforme visto na seção de pré-tratamentos, a transformada por logaritmo na base 10 foi utilizada com o intuito de enfatizar os picos cromatográficos de baixa intensidade. O SNV minimiza as interferências causadas pelas diferenças de intensidade entre os mesmos picos cromatográficos. O *smoothing* aumenta a relação sinal/ruído dos cromatogramas.

Com isso, os cinco pré-tratamentos que melhor descrevem o Banco I são: sem pré-processamento e transformada por logaritmo na base 10, escalada pela variância e sem transformada, escalada pela variância e transformada por logaritmo na base 10, escalada pela variância e transformada por SNV e escalada pela variância e transformada por *smoothing*.

**Tabela 5** - Investigação do melhor pré-tratamento, através da Análise de Componentes Principais, do Banco I. Os cinco pré-tratamentos que obtiveram as melhores respostas estão em destaque.

| **Pré-processamento** | **Transformadas** | **3PC / %** | **7PC / %** | **10PC / %** |
|---|---|---|---|---|
| Nenhum | Nenhuma | 95,2140 | 97,3723 | 98,0676 |
| | Primeira Derivada | 92,3084 | 96,0189 | 97,1155 |
| | Segunda Derivada | 87,3897 | 93,1324 | 94,9121 |
| | Logaritmo Base 10 | 99,9260 | 99,9671 | 99,9760 |
| | MSC | 92,0097 | 96,1508 | 97,4299 |
| | SNV | 93,1279 | 96,0201 | 97,2545 |
| | *Smoothing* | 95,8252 | 97,8532 | 98,5008 |
| Auto-escalado | Nenhuma | 76,8575 | 88,1745 | 91,6472 |
| | Primeira Derivada | 50,2456 | 66,5258 | 73,3614 |
| | Segunda Derivada | 47,1199 | 61,9562 | 68,2104 |
| | Logaritmo Base 10 | 84,4812 | 92,6640 | 94,7289 |
| | MSC | 69,0031 | 81,6165 | 85,9970 |
| | SNV | 73,6387 | 84,1018 | 87,3156 |
| | *Smoothing* | 76,8538 | 88,1985 | 91,6801 |
| Centrado na Média | Nenhuma | 75,9186 | 85,7987 | 89,4855 |
| | Primeira Derivada | 68,9718 | 82,2358 | 87,2360 |
| | Segunda Derivada | 62,7903 | 78,0406 | 83,3964 |
| | Logaritmo Base 10 | 84,6985 | 91,4549 | 93,4146 |
| | MSC | 61,1807 | 77,8147 | 85,1123 |
| | SNV | 54,5786 | 72,9776 | 81,4684 |
| | *Smoothing* | 78,5891 | 88,0996 | 91,6324 |
| Escala de Amplitude | Nenhuma | 95,3870 | 97,4725 | 98,1749 |
| | Primeira Derivada | 95,9433 | 97,3365 | 97,8275 |
| | Segunda Derivada | 95,2830 | 96,6785 | 97,2028 |
| | Logaritmo Base 10 | 97,9373 | 99,0457 | 99,3390 |
| | MSC | 96,2511 | 98,0778 | 98,5522 |
| | SNV | 94,7679 | 97,2243 | 97,8121 |
| | *Smoothing* | 95,4754 | 97,5547 | 98,2799 |
| *Escala de Variância* | Nenhuma | 98,7804 | 99,4130 | 99,5837 |
| | Primeira Derivada | 82,3867 | 87,9050 | 90,5625 |
| | Segunda Derivada | 80,5524 | 85,8149 | 88,2501 |
| | Logaritmo Base 10 | 99,9825 | 99,9921 | 99,9943 |
| | MSC | 97,1051 | 98,5012 | 98,8613 |
| | SNV | 98,9916 | 99,4087 | 99,5364 |
| | *Smoothing* | 98,7746 | 99,4130 | 99,5846 |

**Fonte**: Elaborado pelo autor (2013).

**Tabela 6** - Porcentagem da informação descrita para cada componente principal na investigação do pré-tratamento ótimo para a PCA, realizada com validação cruzada removendo 5 vetores de amostras, do Banco I. O pré-tratamento com a melhor resposta está em destaque.

| **Pré-processamento** | Nenhum | Escala de Variância | Escala de Variância | Escala de Variância | Escala de Variância |
|---|---|---|---|---|---|
| **Transformada** | Log 10 | Nenhuma | Log 10 | SNV | *Smoothing* |
| **1PC / %** | 99,8328 | 97,3510 | 99,9498 | 98,1325 | 97,3367 |
| **2PC / %** | 0,0642 | 1,0642 | 0,0257 | 0,6112 | 1,0716 |
| **3PC / %** | 0,0290 | 0,3654 | 0,0071 | 0,2479 | 0,3663 |
| **4PC / %** | 0,0231 | 0,2616 | 0,0043 | 0,1411 | 0,2657 |
| **5PC / %** | 0,0074 | 0,1634 | 0,0029 | 0,1075 | 0,1635 |
| **6PC / %** | 0,0061 | 0,1129 | 0,0013 | 0,0838 | 0,1140 |
| **7PC / %** | 0,0044 | 0,0947 | 0,0011 | 0,0847 | 0,0952 |
| **8PC / %** | 0,0043 | 0,0790 | 0,0010 | 0,0484 | 0,0795 |
| **9PC / %** | 0,0026 | 0,0559 | 0,0007 | 0,0436 | 0,0558 |
| **10PC / %** | 0,0019 | 0,0358 | 0,0006 | 0,0357 | 0,0363 |

**Fonte**: Elaborado pelo autor (2013).

**Tabela 7** - Percentagem cumulativa da informação descrita para cada componente principal na investigação do pré-tratamento ótimo para a PCA, realizada com validação cruzada removendo 5 vetores de amostras, do Banco I. O pré-tratamento com a melhor resposta está em destaque.

| **Pré-processamento** | Nenhum | Escala de Variância | Escala de Variância | Escala de Variância | Escala de Variância |
|---|---|---|---|---|---|
| **Transformada** | Log 10 | Nenhuma | Log 10 | SNV | *Smoothing* |
| **1PC / %** | 99,8328 | 97,3510 | 99,9498 | 98,1325 | 97,3367 |
| **2PC / %** | 99,8970 | 98,4151 | 99,9754 | 98,7437 | 98,4083 |
| **3PC / %** | 99,9260 | 98,7804 | 99,9825 | 98,9916 | 98,7746 |
| **4PC / %** | 99,9491 | 99,0421 | 99,9867 | 99,1327 | 99,0404 |
| **5PC / %** | 99,9565 | 99,2055 | 99,9896 | 99,2402 | 99,2038 |
| **6PC / %** | 99,9626 | 99,3184 | 99,9910 | 99,3240 | 99,3178 |
| **7PC / %** | 99,9671 | 99,4130 | 99,9921 | 99,4087 | 99,4130 |
| **8PC / %** | 99,9714 | 99,4920 | 99,9930 | 99,4570 | 99,4925 |
| **9PC / %** | 99,9740 | 99,5479 | 99,9938 | 99,5006 | 99,5482 |
| **10PC / %** | 99,9760 | 99,5837 | 99,9943 | 99,5364 | 99,5846 |

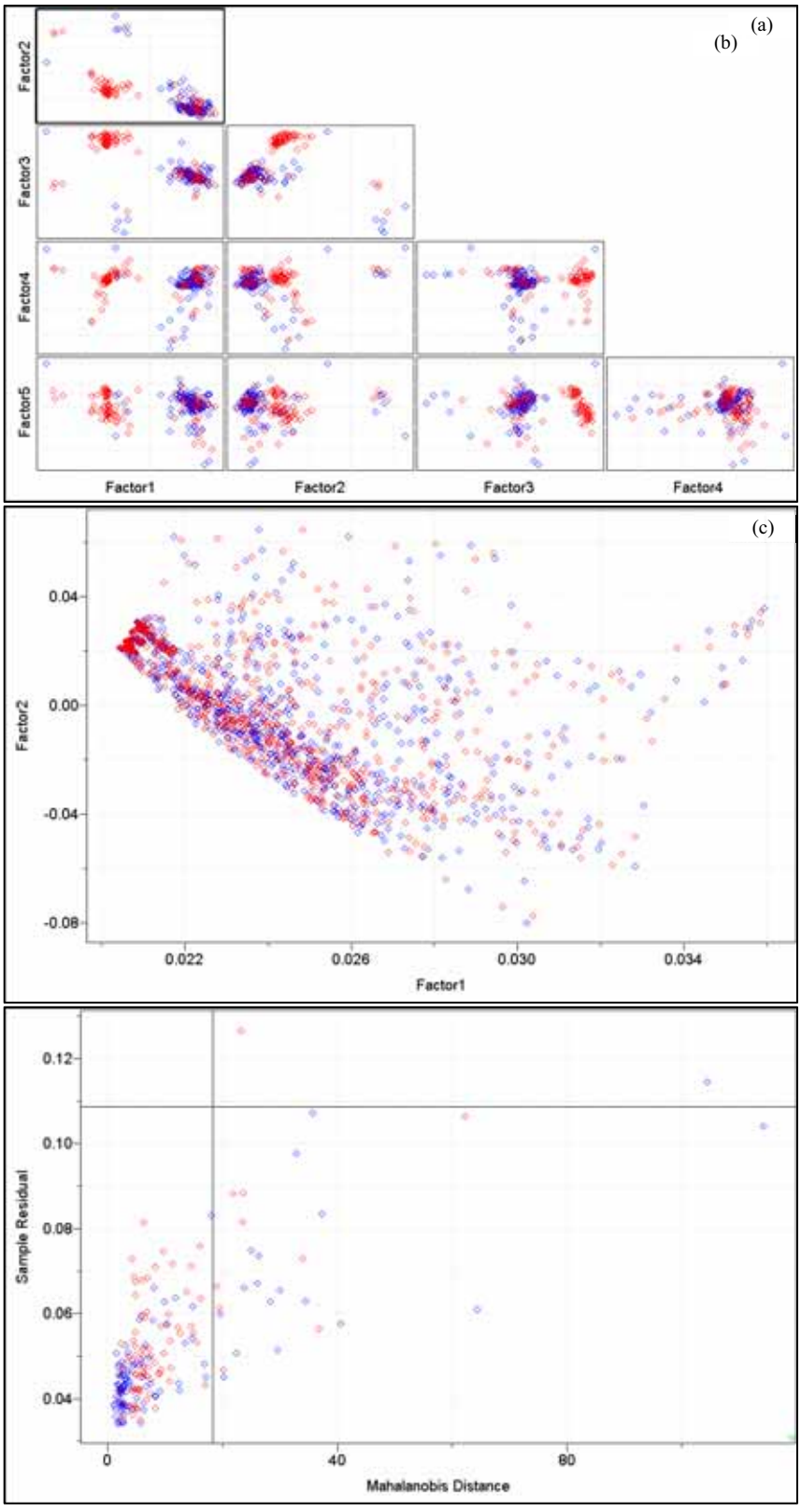
**Fonte**: Elaborado pelo autor (2013).

Realizando a PCA com os cinco pré-tratamentos citados acima e com validação cruzada, obtiveram-se as figuras de 39 a 43, que mostram os *scores*, os *loadings* e os resíduos de cada análise. Os pontos azuis representam amostras conformes e os vermelhos, amostras não conformes.

De acordo com a tabela 6 e 7, o pré-tratamentos ótimo é escalado pela variância e com transformação matemática por logaritmo na base 10. Isto se deve ao fato do escalamento pela variância igualar os efeitos de todas as variáveis, enquanto que a transformada por logaritmo enfatiza os picos de baixa intensidade e lineariza as amostras sem afetar o resultado final.

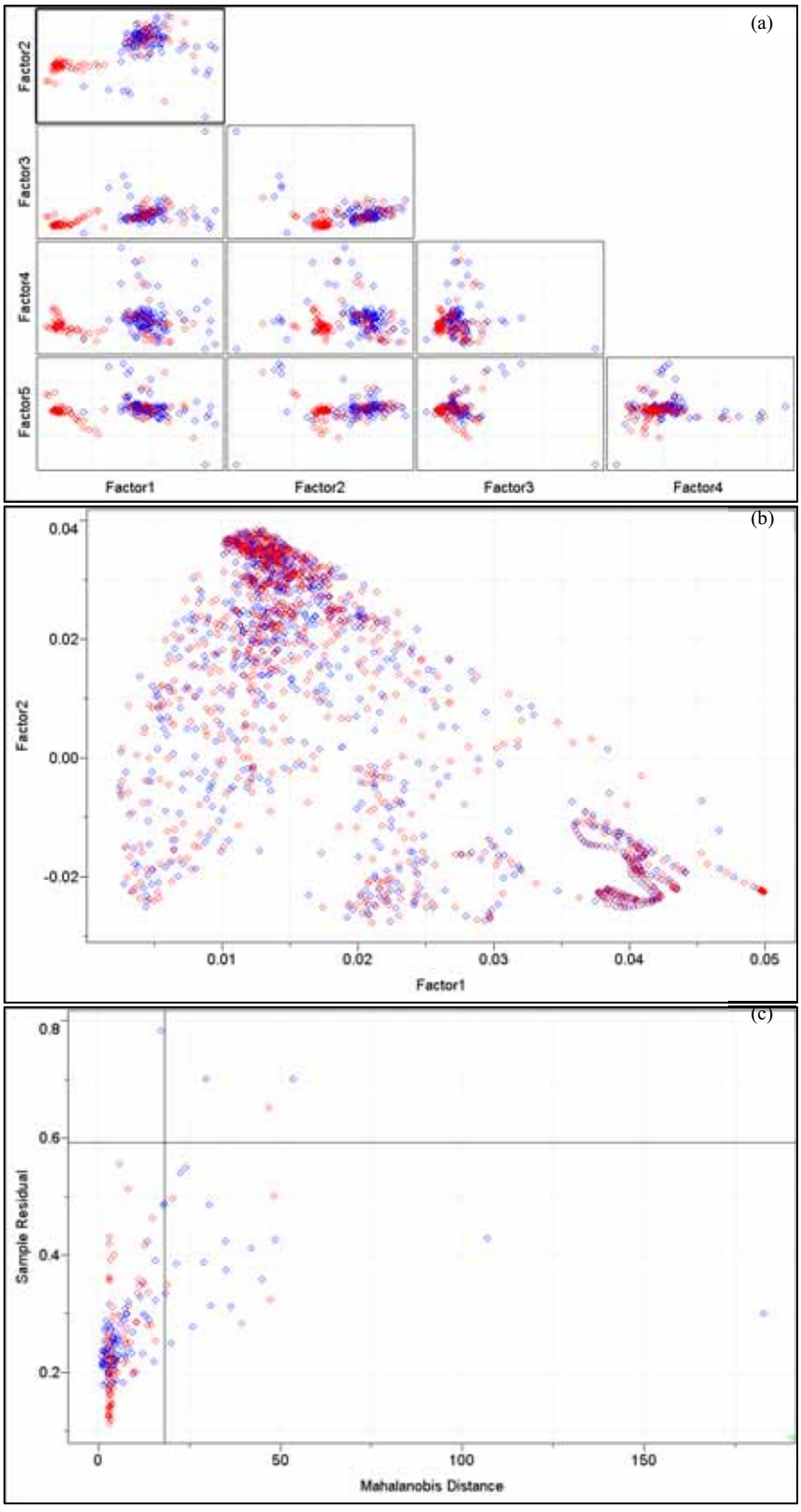
Nota-se que no quarto caso, escalado pela variância e transformado por SNV, há *loadings* de valores nulos, tornando possível a remoção destas variáveis e, por conseguinte, não alterando o resultado da classificação do Banco I pelo SIMCA. Esta eliminação está representada na figura 44 e a alteração no cromatograma, usando a amostra 1111079-C como exemplo, na figura 45.

**Figura 39** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I, sem pré-processamento e transformada por logaritmo na base 10.



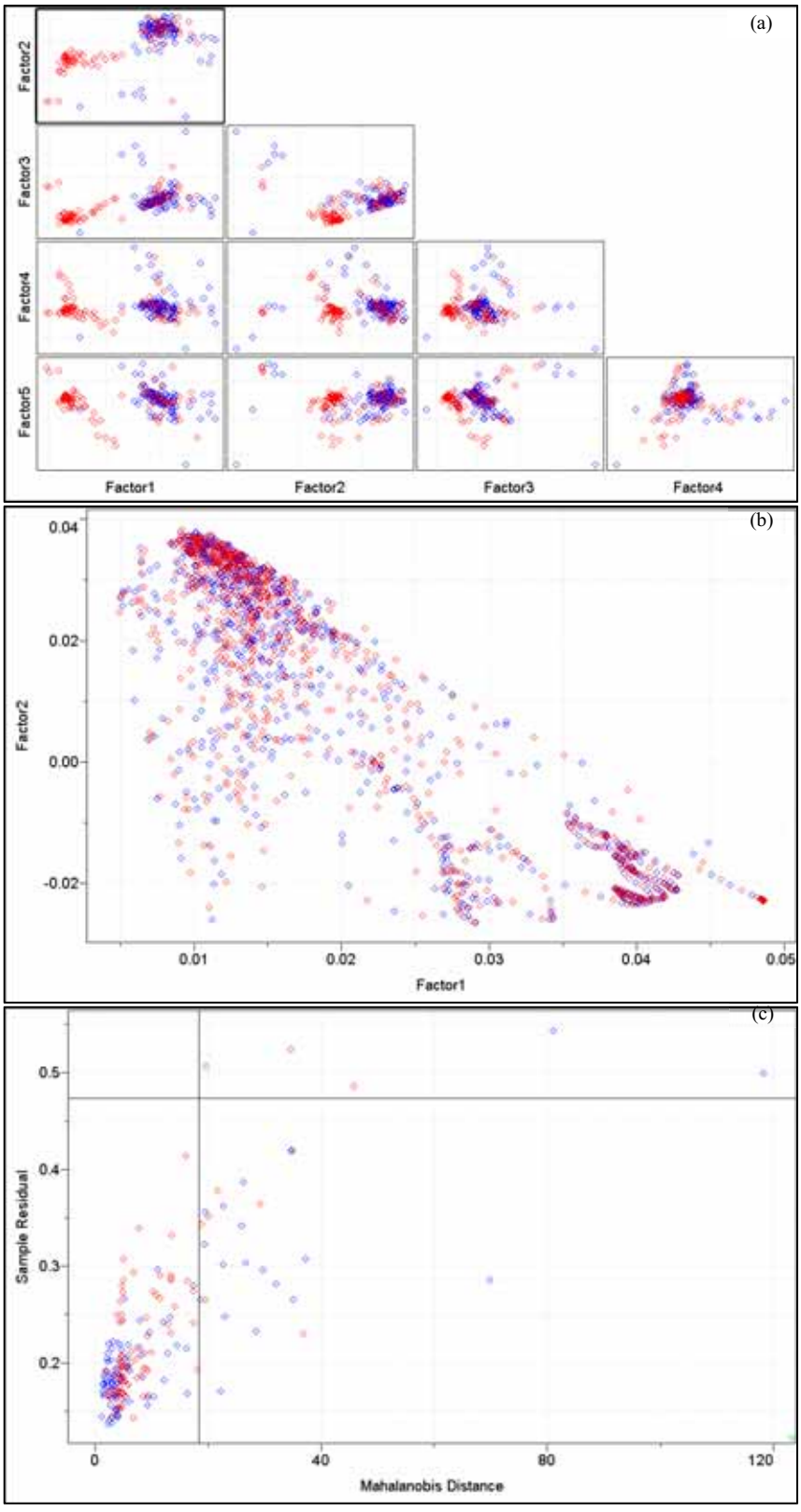

**Fonte**: Elaborado pelo autor (2013).

**Figura 40** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I, escalada pela variância e sem transformada.



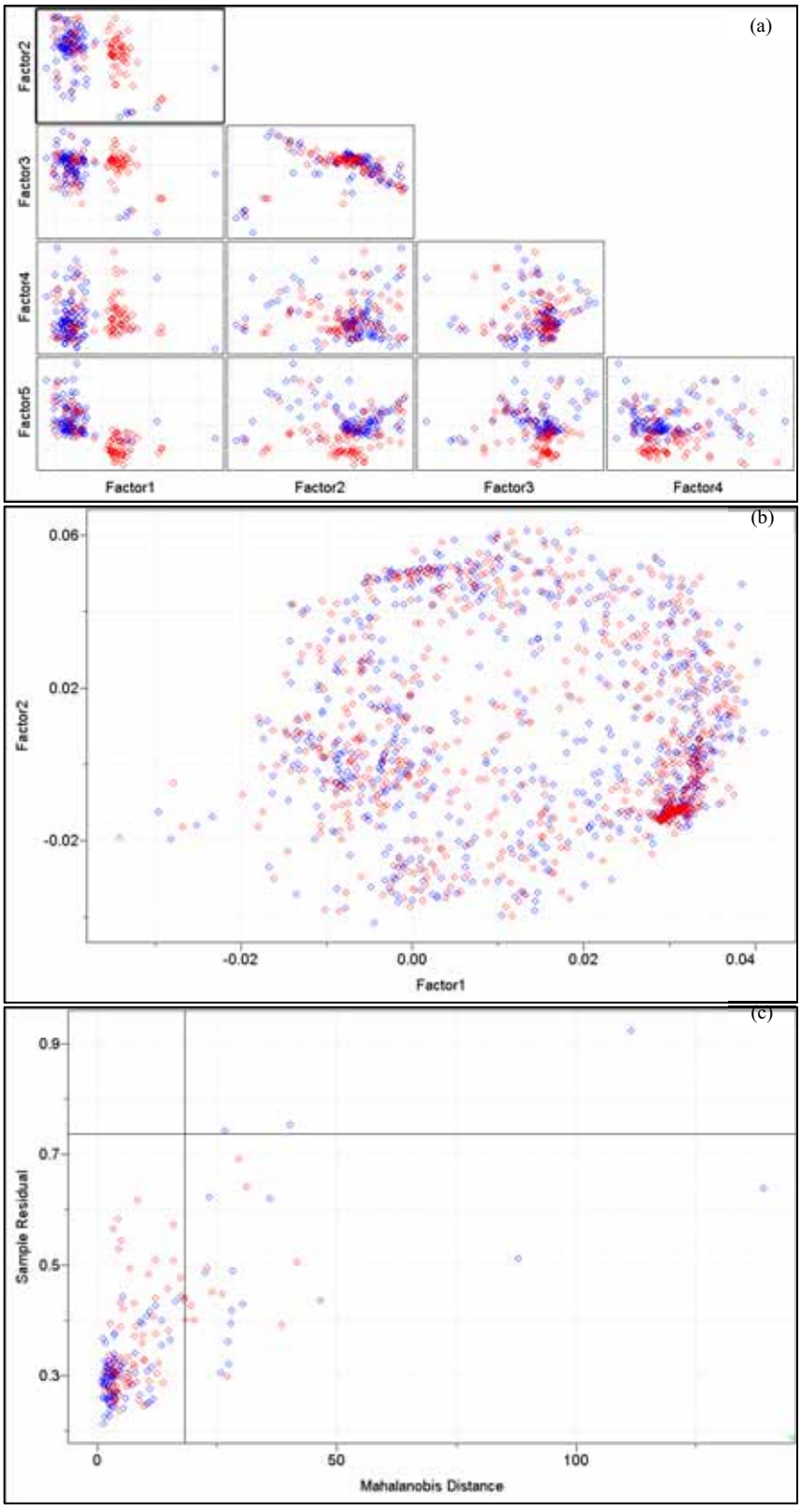

**Fonte**: Elaborado pelo autor (2013).

**Figura 41** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I, escalada pela variância e transformada por logaritmo na base 10.



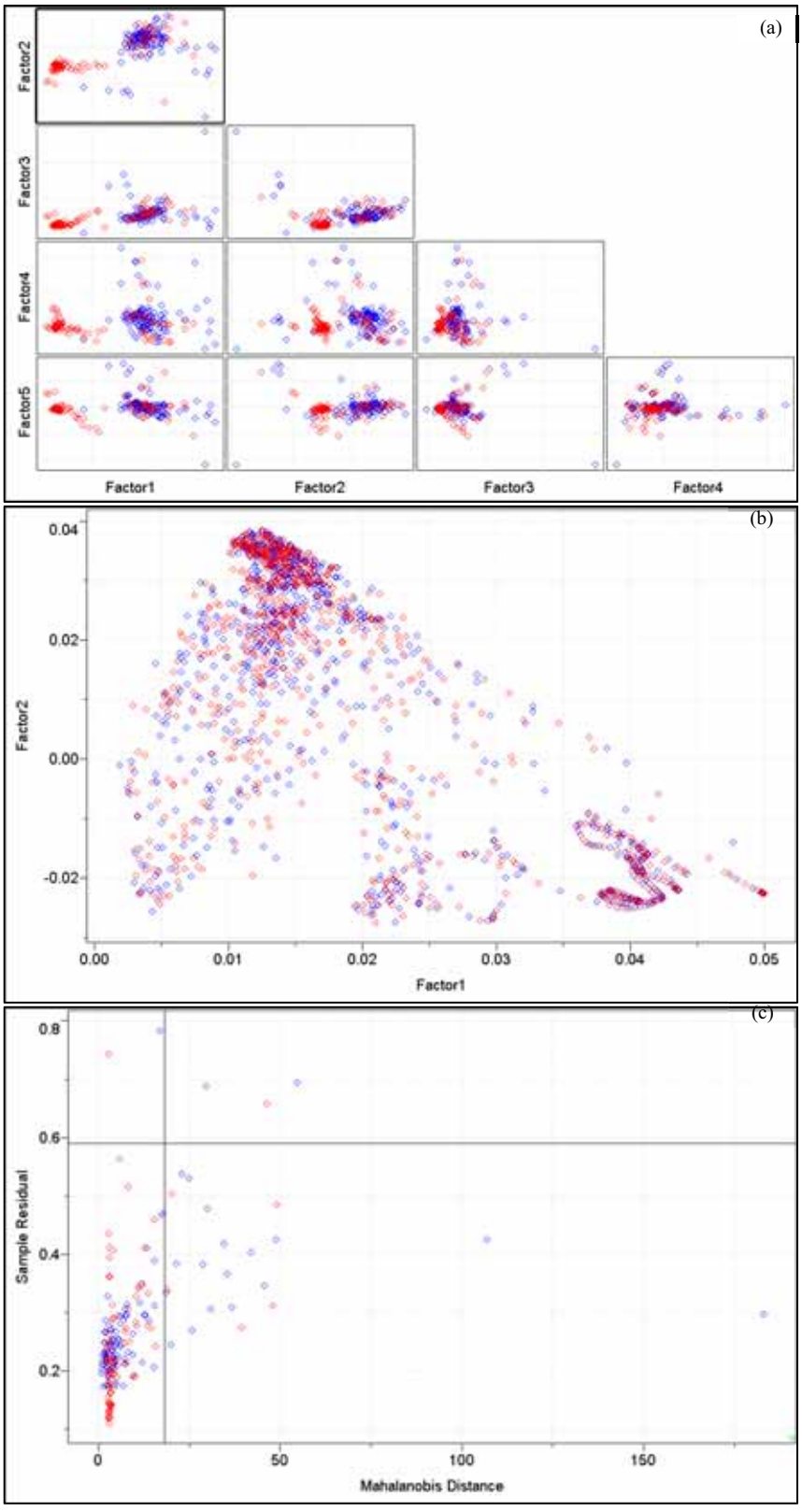

**Fonte**: Elaborado pelo autor (2013).

**Figura 42** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I, escalada pela variância e transformada por SNV.



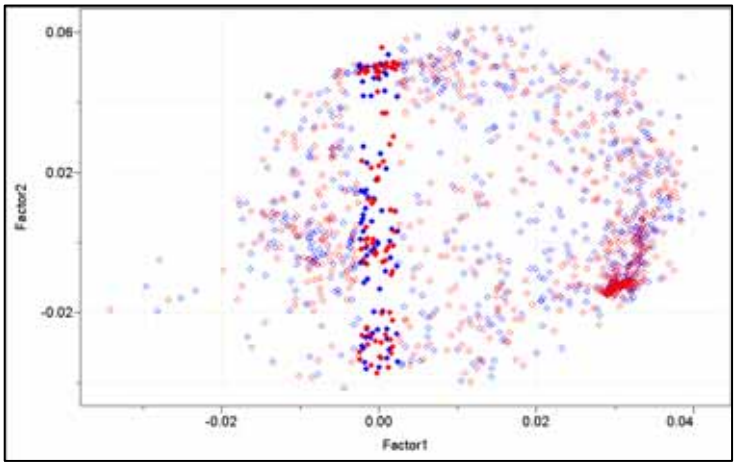

**Fonte**: Elaborado pelo autor (2013).

**Figura 43** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I, escalada pela variância e transformada por *smoothing*.




**Fonte**: Elaborado pelo autor (2013).

**Figura 44** - Remoção das variáveis, em destaque, com peso nulo da análise de componentes principais, escalada pela variância e transformada por SNV, do Banco I.




**Fonte**: Elaborado pelo autor (2013).

**Figura 45** - Cromatograma da amostra 1111079-C antes (a) e após a remoção de loadings nulos (b). As áreas cinza representam as variáveis a serem removidas.

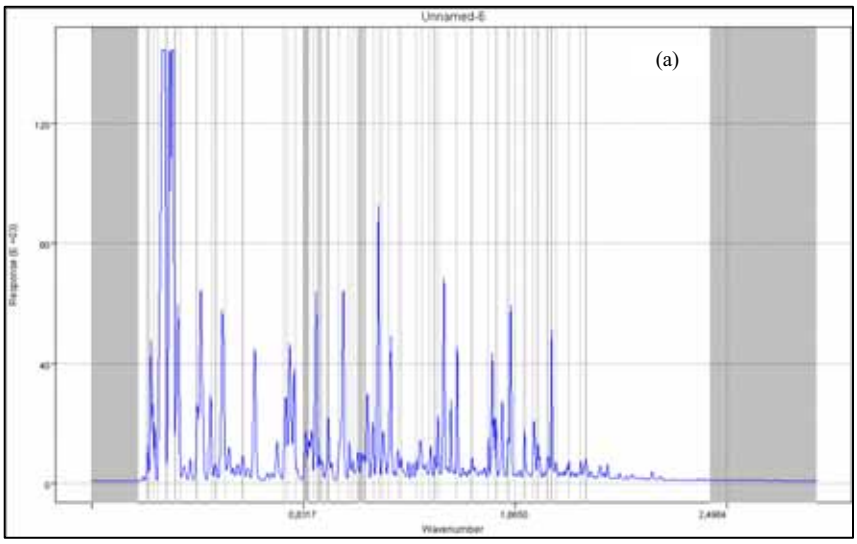
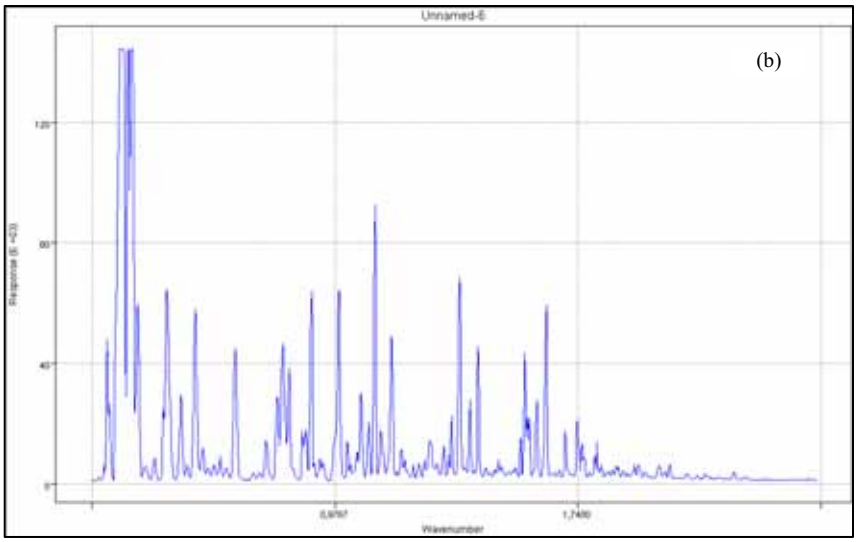


**Figura 45** - Cromatograma da amostra 1111079-C antes (a) e após a remoção de loadings nulos (b). As áreas cinza representam as variáveis a serem removidas (continuação).




**Fonte**: Elaborado pelo autor (2013).

### 5.6.2 Avaliação do Pré-processamento e Transformação para o Banco II

De semelhante modo, foi realizada a investigação do melhor pré-tratamento para a PCA do Banco II. A tabela 8 mostra os valores das percentagens cumulativas para 3, 7 e 10 componentes principais. Posteriormente, para os cinco pré-tratamentos que melhor descreveram o a matriz original de dados, realizou-se a PCA com validação cruzada, removendo 5 vetores de amostras para cada análise, a fim de se confirmarem os resultados obtidos anteriormente. Resultado exibido nas tabelas 9 e 10.

Similar ao Banco I, a PCA sem pré-processamento e escalada pela variância apresentaram descrição máxima do Banco II.

Além disso, as transformadas foram similares às do Banco I. Os pré-tratamentos que melhor descrevem o Banco II são: sem pré-processamento e transformada por logaritmo na base 10, escalada pela variância e sem transformada, escalada pela variância e transformada por logaritmo na base 10, escalada pela variância e transformada por SNV e escalada pela variância e transformada por *smoothing*.

**Tabela 8** - Investigação do melhor pré-tratamento, através da Análise de Componentes Principais, do Banco II. Os cinco pré-tratamentos que obtiveram as melhores respostas estão em destaque.

| Pré-processamento | Transformadas | 3PC / % | 7PC / % | 10PC / % |
|---|---|---|---|---|
| *Nenhum* | Nenhuma | 92,2495 | 95,7526 | 97,0336 |
| | Primeira Derivada | 88,7961 | 93,7344 | 95,5806 |
| | Segunda Derivada | 84,1407 | 90,1146 | 92,8035 |
| | Logaritmo Base 10 | 99,8706 | 99,9463 | 99,9617 |
| | MSC | 89,7747 | 94,8129 | 96,2317 |
| | SNV | 90,6518 | 94,9927 | 96,3205 |
| | *Smoothing* | 92,8841 | 96,3269 | 97,5725 |
| Auto-escalado | Nenhuma | 71,6082 | 85,2098 | 88,9539 |
| | Primeira Derivada | 49,8034 | 63,9254 | 69,5284 |
| | Segunda Derivada | 48,2559 | 60,7288 | 65,4968 |
| | Logaritmo Base 10 | 75,0925 | 89,8129 | 92,5905 |
| | MSC | 68,9496 | 81,4445 | 84,7411 |
| | SNV | 69,2784 | 81,7509 | 84,7695 |
| | *Smoothing* | 71,3837 | 85,0467 | 88,7893 |
| Centrado na Média | Nenhuma | 77,2875 | 86,8120 | 90,2421 |
| | Primeira Derivada | 69,6863 | 82,4103 | 87,1905 |
| | Segunda Derivada | 62,8209 | 76,2981 | 82,3099 |
| | Logaritmo Base 10 | 85,0681 | 92,5025 | 94,3382 |
| | MSC | 56,6787 | 73,6197 | 80,9628 |
| | SNV | 57,4469 | 72,9512 | 79,7974 |
| | *Smoothing* | 79,0000 | 88,5335 | 91,8584 |
| Escala de Amplitude | Nenhuma | 94,6062 | 97,5452 | 98,2465 |
| | Primeira Derivada | 97,7270 | 98,4772 | 98,6956 |
| | Segunda Derivada | 97,3394 | 98,1052 | 98,3259 |
| | Logaritmo Base 10 | 97,7436 | 99,0958 | 99,3505 |
| | MSC | 98,6289 | 99,2727 | 99,4421 |
| | SNV | 95,8922 | 97,8034 | 98,3118 |
| | *Smoothing* | 94,7761 | 97,6643 | 98,3561 |
| *Escala de Variância* | Nenhuma | 98,8094 | 99,3961 | 99,5488 |
| | Primeira Derivada | 72,2240 | 80,4804 | 83,7581 |
| | Segunda Derivada | 71,1535 | 78,6288 | 81,4547 |
| | Logaritmo Base 10 | 99,9694 | 99,9879 | 99,9912 |
| | MSC | 95,6382 | 97,6257 | 98,1059 |
| | SNV | 98,2839 | 98,9986 | 99,1754 |
| | *Smoothing* | 98,7865 | 99,3832 | 99,5373 |

**Fonte**: Elaborado pelo autor (2013).

**Tabela 9** - Percentagem da informação descrita para cada componente principal na investigação do pré-tratamento ótimo para a PCA, realizada com validação cruzada removendo 5 vetores de amostras, do Banco II. O pré-tratamento com a melhor resposta está em destaque.

| **Pré-processamento** | Nenhum | Escala de Variância | Escala de Variância | Escala de Variância | Escala de Variância |
|---|---|---|---|---|---|
| **Transformada** | Log 10 | Nenhum | Log 10 | SNV | *Smoothing* |
| **1PC / %** | 99,5656 | 96,9332 | 99,9074 | 97,2976 | 96,8930 |
| **2PC / %** | 0,2409 | 1,5873 | 0,0467 | 0,6563 | 1,6014 |
| **3PC / %** | 0,0641 | 0,2888 | 0,0152 | 0,3299 | 0,2921 |
| **4PC / %** | 0,0358 | 0,2149 | 0,0078 | 0,2739 | 0,2196 |
| **5PC / %** | 0,0172 | 0,1514 | 0,0055 | 0,1804 | 0,1560 |
| **6PC / %** | 0,0145 | 0,1314 | 0,0033 | 0,1659 | 0,1315 |
| **7PC / %** | 0,0081 | 0,089 | 0,0018 | 0,0945 | 0,0958 |
| **8PC / %** | 0,0066 | 0,0604 | 0,0012 | 0,0815 | 0,0611 |
| **9PC / %** | 0,0045 | 0,0507 | 0,0011 | 0,0501 | 0,0512 |
| **10PC / %** | 0,0042 | 0,0415 | 0,0009 | 0,0452 | 0,0418 |

**Fonte**: Elaborado pelo autor (2013).

**Tabela 10** - Percentagem cumulativa da informação descrita para cada componente principal na investigação do pré-tratamento ótimo para a PCA, realizada com validação cruzada removendo 5 vetores de amostras, do Banco II. O pré-tratamento com a melhor resposta está em destaque.

| **Pré-processamento** | Nenhum | Escala de Variância | Escala de Variância | Escala de Variância | Escala de Variância |
|---|---|---|---|---|---|
| **Transformada** | Log 10 | Nenhum | Log 10 | SNV | *Smoothing* |
| **1PC / %** | 99,5656 | 96,9332 | 99,9074 | 97,2976 | 96,8930 |
| **2PC / %** | 99,8065 | 98,5206 | 99,9542 | 97,9539 | 98,4943 |
| **3PC / %** | 99,8706 | 98,8094 | 99,9694 | 98,2839 | 98,7865 |
| **4PC / %** | 99,9064 | 99,0243 | 99,9772 | 98,5578 | 99,0061 |
| **5PC / %** | 99,9237 | 99,1757 | 99,9827 | 98,7382 | 99,1621 |
| **6PC / %** | 99,9382 | 99,3071 | 99,9861 | 98,9041 | 99,2936 |
| **7PC / %** | 99,9463 | 99,3961 | 99,9879 | 98,9986 | 99,3832 |
| **8PC / %** | 99,9529 | 99,4566 | 99,9891 | 99,0801 | 99,4443 |
| **9PC / %** | 99,9574 | 99,5072 | 99,9902 | 99,1302 | 99,4955 |
| **10PC / %** | 99,9617 | 99,5488 | 99,9912 | 99,1754 | 99,5373 |

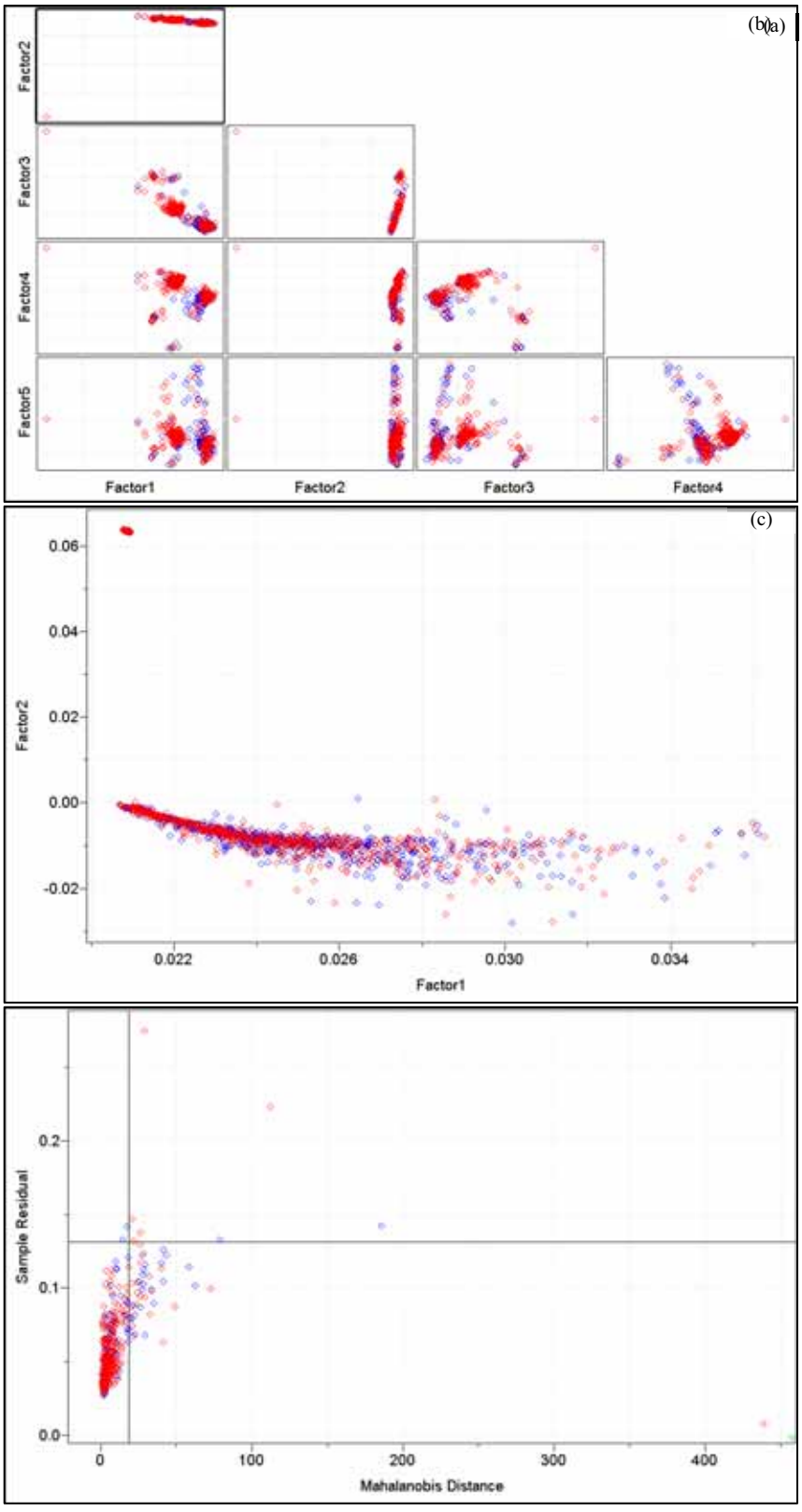
**Fonte**: Elaborado pelo autor (2013).

Semelhante ao Banco I, as análises com os cinco pré-tratamentos citados acima e com validação cruzada são ilustradas nas figuras de 46 a 50, que mostram os *scores*, os *loadings* e os resíduos de cada análise. E, de modo similar, ao se escalar pela variância e transformar por SNV, o Banco II obteve *loadings* de valores nulos. Por isso, também se fez uma remoção destas variáveis. A eliminação está descrita na figura 51 e a alteração no cromatograma, usando a amostra 114889-C como exemplo, na figura 52.

De acordo com as tabelas 9 e 10, o pré-tratamentos ótimo é escalado pela variância e com transformação matemática por logaritmo na base 10. Isto se deve ao fato do escalamento pela variância igualar os efeitos de todas as variáveis, enquanto que

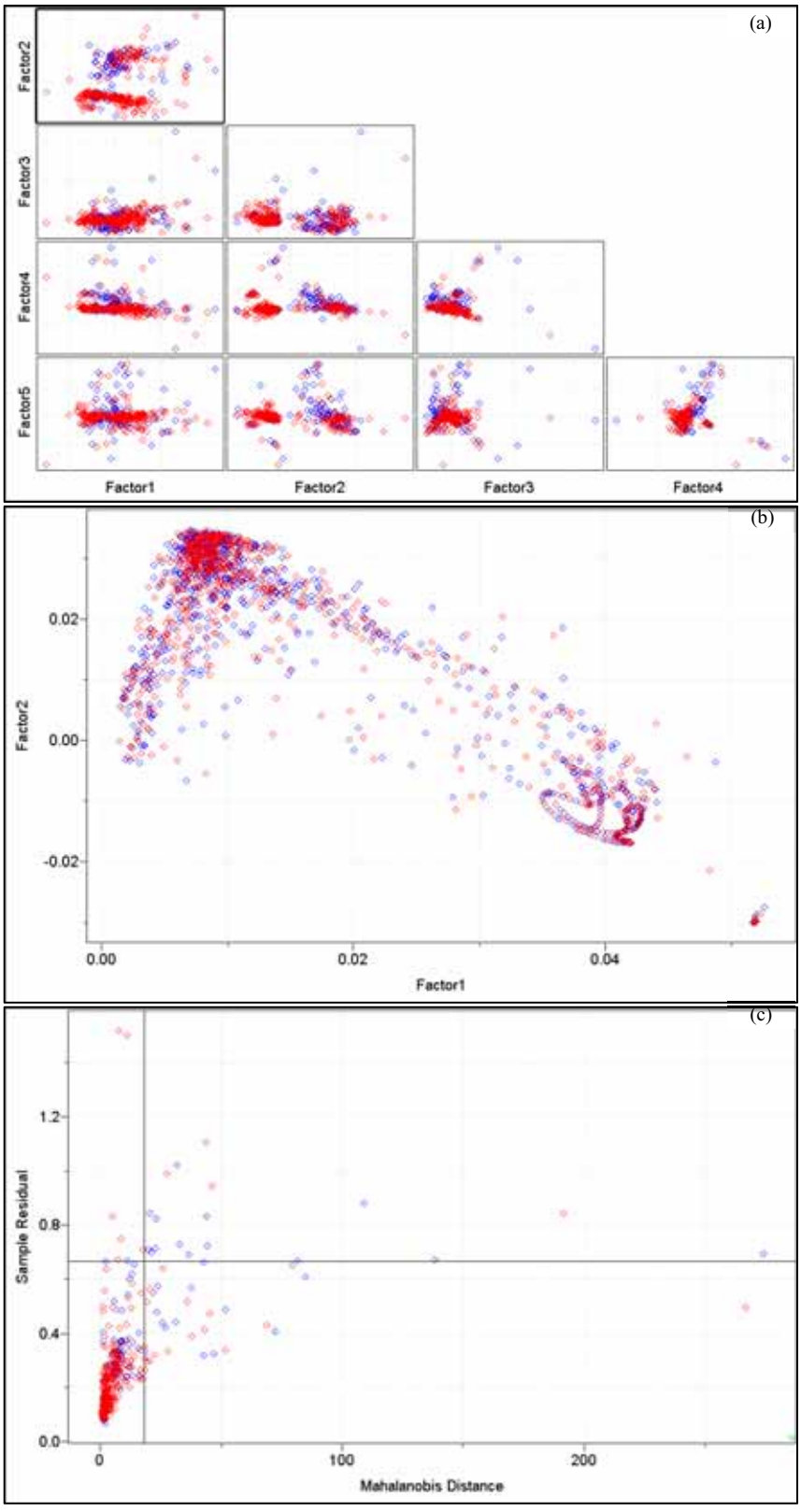
a transformada por logaritmo enfatiza os picos de baixa intensidade e lineariza as amostras sem afetar o resultado final.

**Figura 46** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II, sem pré-processamento e transformada por logaritmo na base 10.



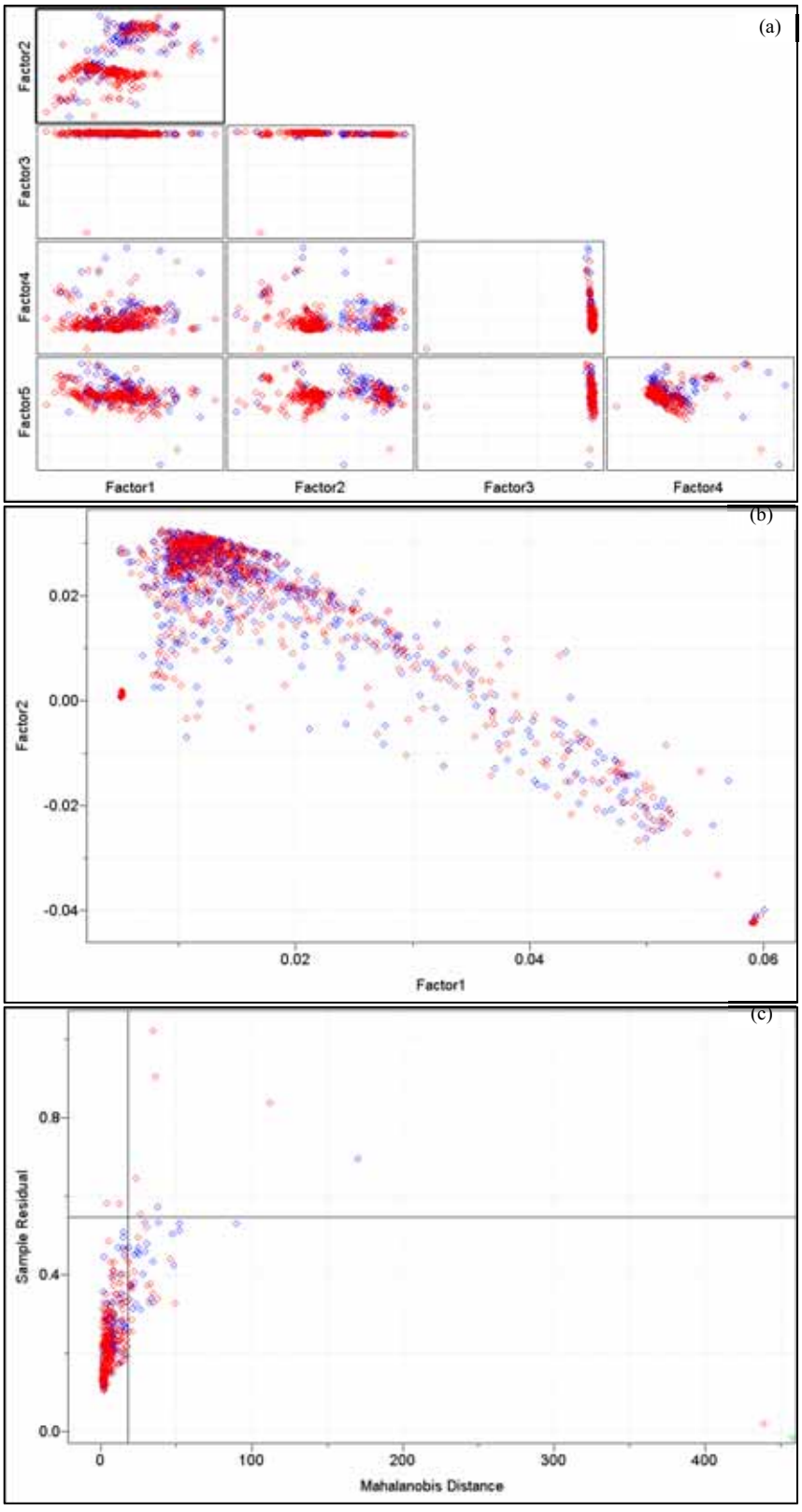

**Fonte**: Elaborado pelo autor (2013).

**Figura 47** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II, escalada pela variância e sem transformada.



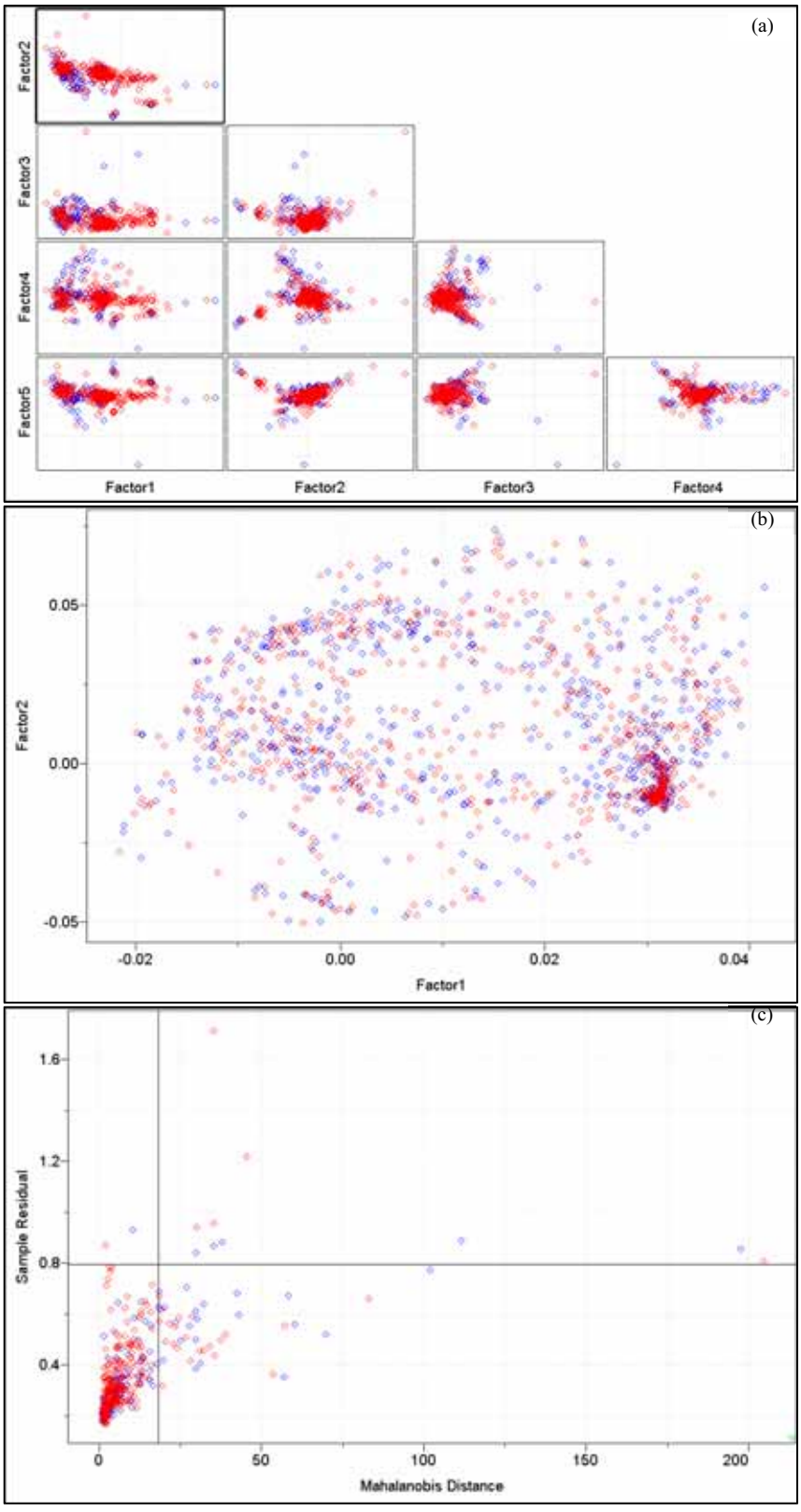

**Fonte**: Elaborado pelo autor (2013).

**Figura 48** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II, escalada pela variância e transformada por logaritmo na base 10.



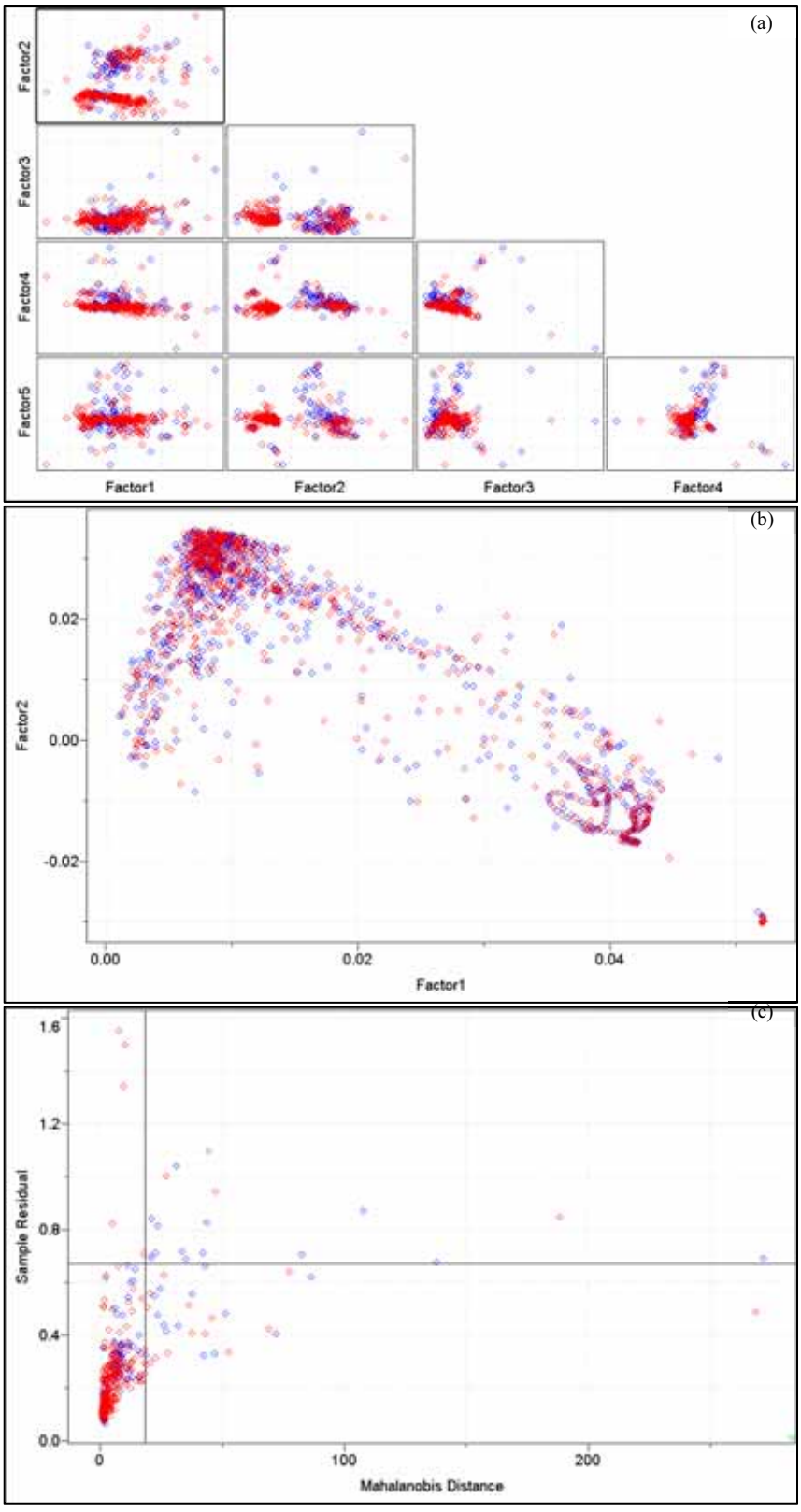

**Fonte**: Elaborado pelo autor (2013).

**Figura 49** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II, escalada pela variância e transformada por SNV.



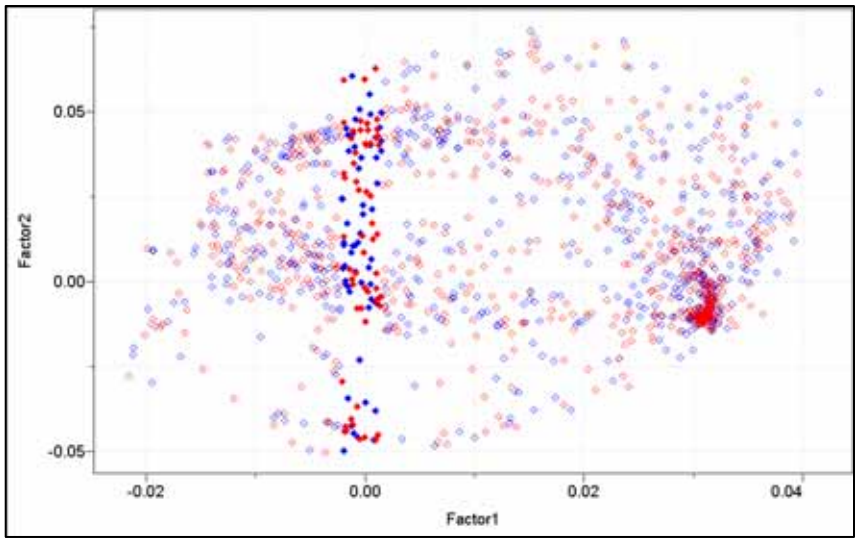

**Fonte**: Elaborado pelo autor (2013).

**Figura 50** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II, escalada pela variância e transformada por *smoothing*.




**Fonte**: Elaborado pelo autor (2013).

**Figura 51** - Remoção das variáveis, em destaque, com peso nulo da análise de componentes principais, escalada pela variância e transformada por SNV, do Banco II.




**Fonte**: Elaborado pelo autor (2013).

**Figura 52** - Cromatograma da amostra 114889-C antes (a) e após a remoção de loadings nulos (b). As áreas cinza representam as variáveis a serem removidas.

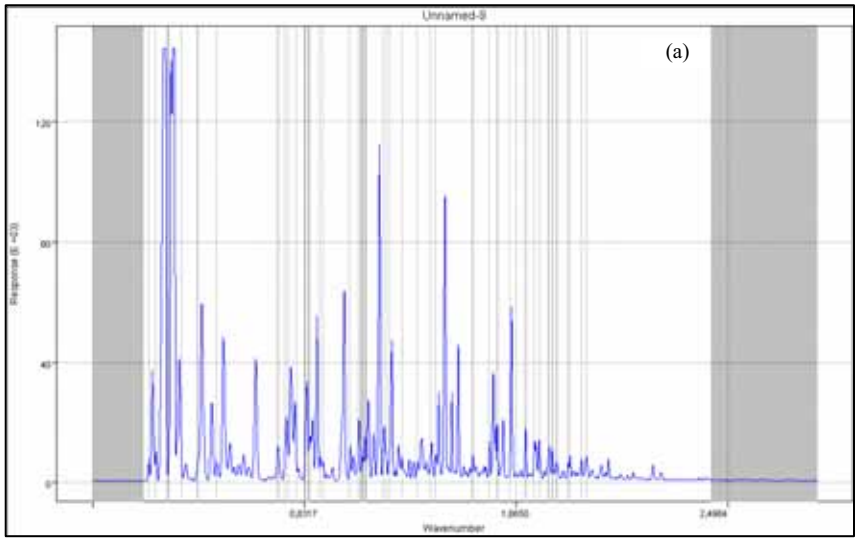
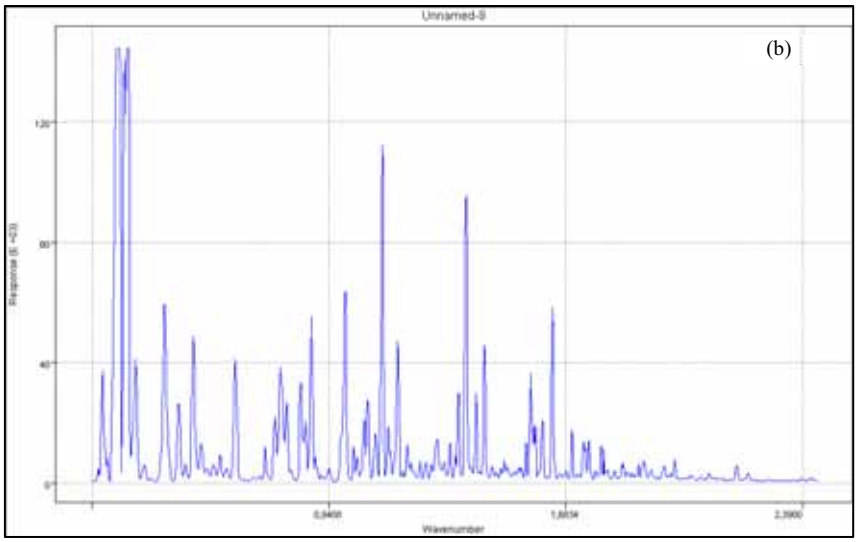


**Figura 52** - Cromatograma da amostra 114889-C antes (a) e após a remoção de loadings nulos (b). As áreas cinza representam as variáveis a serem removidas (continuação).




**Fonte**: Elaborado pelo autor (2013).

## *5.7 Classificação quanto à Conformidade dos Bancos I e II por SIMCA*

Por fim, realizou-se a classificação das amostras de gasolinas comerciais do tipo C quanto à conformidade ou não segundo a Resolução nº 57 de 2011 da ANP e a Portaria nº 143 e 678 do MAPA.

De semelhante modo à PCA, a SIMCA foi realizada para os Bancos I e II separadamente. A análise foi feita com os cinco pré-tratamentos, averiguando qual o melhor método de classificação das amostras.

Antes da análise classificatória, cada banco foi divido em dois grupos: *grupo de treinamento*, consistindo em 80% das amostras e *grupo de previsão*, 20% das amostras. O modelo de classificação é calculado através das características das amostras do grupo de treinamento e validado através do grupo de previsão.

A porcentagem de cada grupo é proporcional à quantidade de amostras conformes e não conformes existentes no banco. O cálculo da quantidade de amostra em cada grupo para cada banco é descrito a seguir.

No Banco I, há um total de 202 amostras, sendo 112 conformes e 90 não conformes. Desta forma, no grupo de treinamento há 162 amostras e no de previsão há 40 amostras.

Assim sendo, somando-se 80% do número de amostras conformes e 80% das não conformes obtém-se o total do grupo de treinamento:

$$n_{-C\,BI}.0{,}8 + n_{-NC\,BI}.0{,}8 = 112.0{,}8 + 90.0{,}8 = 89{,}6 + 72 \cong 162 \text{ amostras} \quad (80)$$

Em que $n_{-C\,BI}$ é o número de amostras conformes para o Banco I e $n_{-NC\,BI}$ é o número de amostras não conformes para o Banco I. Assim, das 112 amostras conformes 90 farão parte do grupo de treinamento e $112 - 90 = 22$ do grupo de previsão e das 90 não conformes, 72 serão do grupo de treinamento e $90 - 72 = 18$ do grupo de previsão.

De semelhante modo, no Banco II há um total de 441 amostras, 142 conformes e 299 não conformes. Desta maneira, no grupo de treinamento há 353 amostras e no de previsão 88 amostras. Portanto:

$$n_{-C\,BII}.0{,}8 + n_{-NC\,BII}.0{,}8 = 142.0{,}8 + 299.0{,}8 = 113{,}6 + 239 \cong 353\ amostras \quad (81)$$

Em que $n_{-C\,BII}$ é o número de amostras conformes para o Banco II e $n_{-NC\,BII}$ é o número de amostras não conformes para o Banco II. Deste modo, das 142 amostras conformes do Banco II 114 farão parte do grupo de treinamento e $142 - 114 = 28$ do grupo de previsão e das 299 amostras não conformes do Banco II, 239 estão contidas no grupo de treinamento e $299 - 239 = 60$ amostras estão contidas no grupo de previsão.

### 5.7.1 Análise Classificatória por SIMCA para o Banco I

A tabela 11 mostra o resultado da classificação pela SIMCA para o grupo de treinamento do Banco I. As predições para a classe conforme mostraram um acerto de 94% (sem pré-tratamento e transformada por logaritmo na base 10), enquanto que para a classe não conforme o acerto foi de 90% (escalado pela variância e transformada por

SNV). Isto se deve à similaridade entre algumas amostras não conformes e as amostras conformes.

A tabela 12 mostra o resultado da classificação pela SIMCA para o grupo de previsão do Banco I. As percentagens de acerto de classificação foram de 59%, para as amostras conformes (sem pré-processamento e transformada por logaritmo na base 10), e de 72% para as amostras não conformes (sem pré-processamento e transformada por logaritmo na base 10).

Como fora dito, houve uma necessidade de adulteração das amostras conformes para ampliação e variabilidade do banco de dados totais, uma vez que algumas das amostras que foram classificadas como não conformes, segundo os resultados das análises físico-químicas, por possuírem percentagem de AEAC próxima da permitida, como, por exemplo, 27% no caso do Banco I, ou 22% no caso do Banco II. Além disso, os métodos de pré-tratamentos e algoritmos de alinhamento de picos cromatográficos podem reduzir ainda mais as diferenças entre as classes, devido a aproximações matemáticas.

Entretanto, a classificação é eficiente para amostras adulteradas por solventes, principalmente os comerciais de fácil acesso público, como alguns dos listados na tabela 4. A figura 53 mostra a classificação pela SIMCA, com os cinco pré-tratamentos, do grupo de treinamento do Banco I. É difícil distinguir amostras conformes de amostras não conformes, entretanto é visível a separação entre amostras adulteradas e o aglomerado de amostras conformes e não conformes.

Outrossim, no grupo de previsão é notada a tendência para a classificação de não conformidade. Poucas amostras são preditas como conformes, devido à menor quantidade destas e às pequenas diferenças, quase próximas, das características entre as duas classes. A figura 54 exibe a classificação pela SIMCA, com os cinco pré-tratamentos, do grupo de previsão do Banco I. Desta vez, é possível distinguir as amostras conformes e não conformes. Isto se deve a maior presença de amostras adulteradas no grupo das amostras não conformes.

A tabela 13 exibe os resultados da SIMCA, escalada pela variância e transformada por SNV, com seleção de variáveis. Nota-se que não houve muita

diferença entre a mesma SIMCA sem remoção. Isto corrobora com a possibilidade da eliminação das variáveis de peso nulo.

**Tabela 11** - Resultado da classificação pela SIMCA para grupo de treinamento (90 amostras conformes e 72 amostras não conformes) para o Banco I. As melhores classificações estão destacadas.

| Pré-processamentos | Transformadas | Classe[3] | Previsão C | Previsão NC | Sem classe | Acerto/% |
|---|---|---|---|---|---|---|
| Nenhum | Log10 | C | 85 | 5 | 0 | 94 |
| | | NC | 10 | 62 | 0 | 86 |
| Escala de Variância | Nenhuma | C | 81 | 9 | 0 | 90 |
| | | NC | 9 | 63 | 0 | 87 |
| Escala de Variância | Log10 | C | 78 | 12 | 0 | 87 |
| | | NC | 10 | 62 | 0 | 86 |
| Escala de Variância | SNV | C | 82 | 8 | 0 | 91 |
| | | NC | 7 | 65 | 0 | 90 |
| Escala de Variância | *Smoothing* | C | 81 | 9 | 0 | 90 |
| | | NC | 9 | 63 | 0 | 87 |

[3] C – amostras conformes, NC – amostras não conformes

**Fonte**: Elaborado pelo autor (2013).

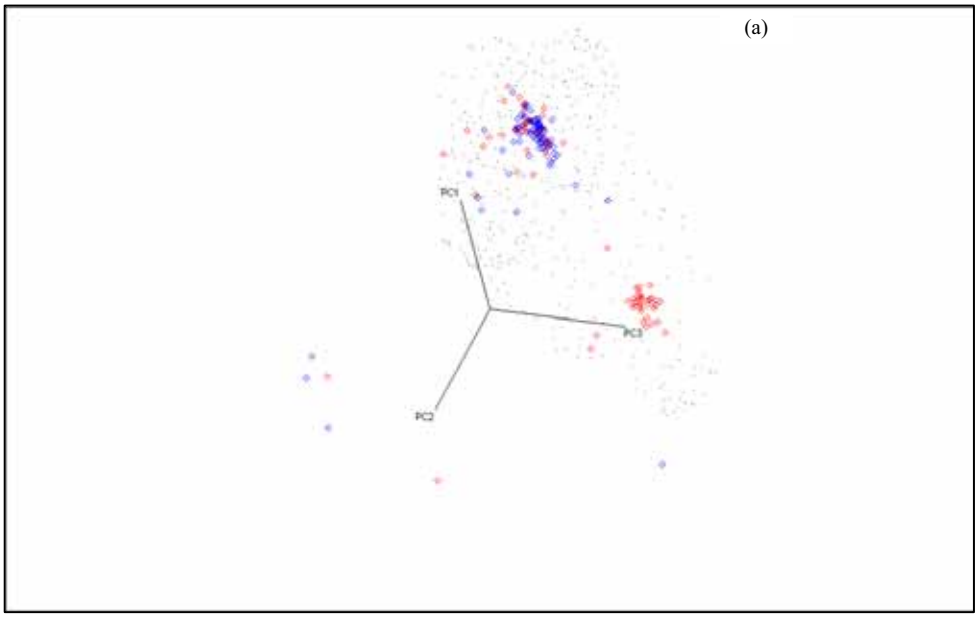
**Tabela 12** - Resultado da classificação pela SIMCA para grupo de previsão (22 amostras conformes e 18 amostras não conformes) para o Banco I. As melhores classificações estão destacadas.

| Pré-processamentos | Transformadas | Classe[4] | Previsão C | Previsão NC | Sem classe | Acerto/% |
|---|---|---|---|---|---|---|
| Nenhum | Log10 | C | 13 | 7 | 2 | 59 |
| | | NC | 0 | 13 | 5 | 72 |
| Escala de Variância | Nenhuma | C | 9 | 6 | 7 | 41 |
| | | NC | 0 | 10 | 8 | 56 |
| Escala de Variância | Log10 | C | 12 | 8 | 2 | 54 |
| | | NC | 0 | 12 | 6 | 67 |
| Escala de Variância | SNV | C | 10 | 6 | 6 | 45 |
| | | NC | 1 | 4 | 13 | 22 |
| Escala de Variância | *Smoothing* | C | 9 | 6 | 7 | 41 |
| | | NC | 0 | 10 | 8 | 56 |

[4] C – amostras conformes, NC – amostras não conformes

**Fonte**: Elaborado pelo autor (2013).

**Figura 53** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco I, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes.



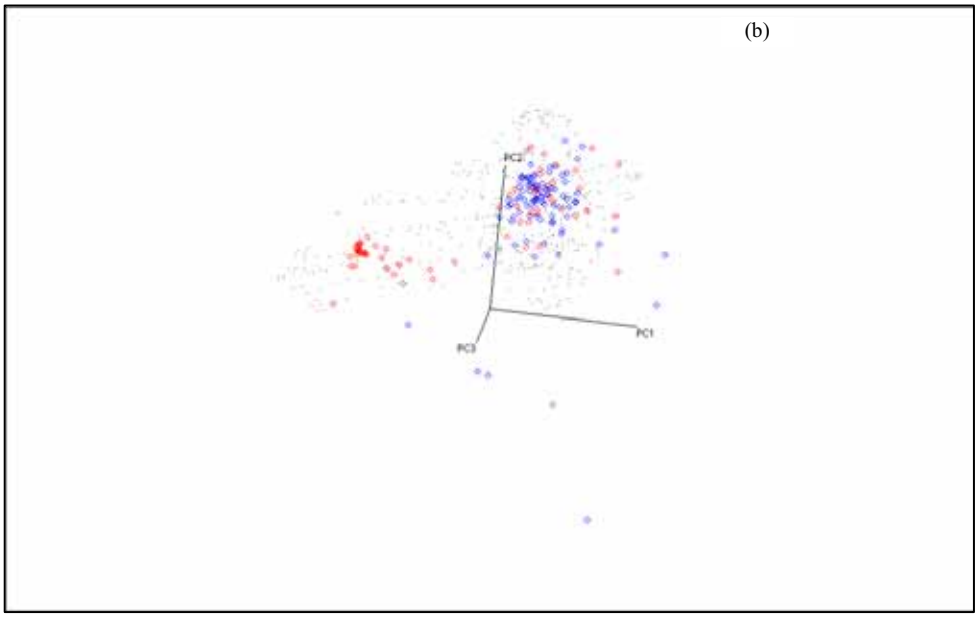
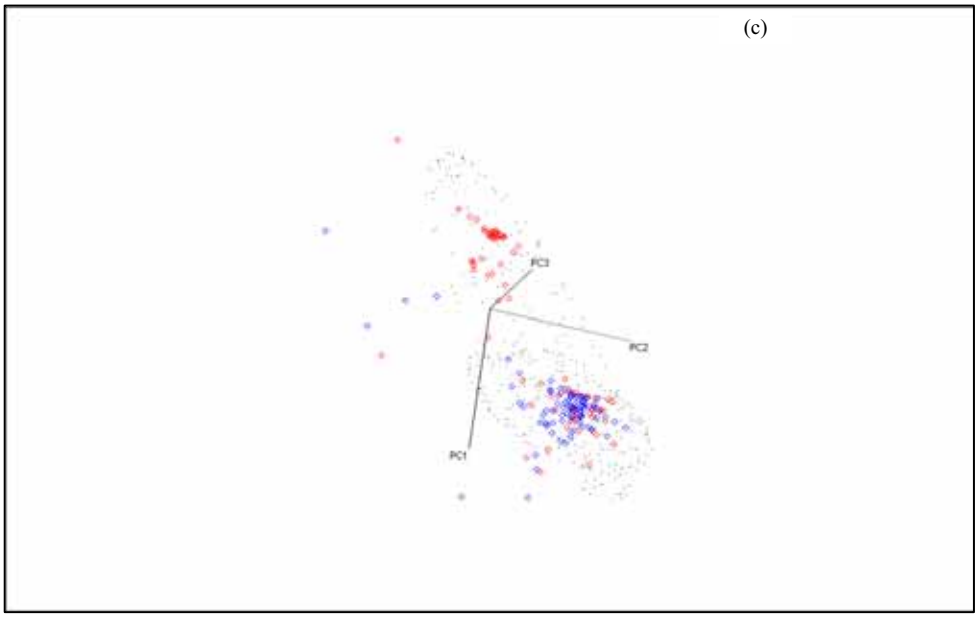



**Figura 53** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco I, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (continuação).



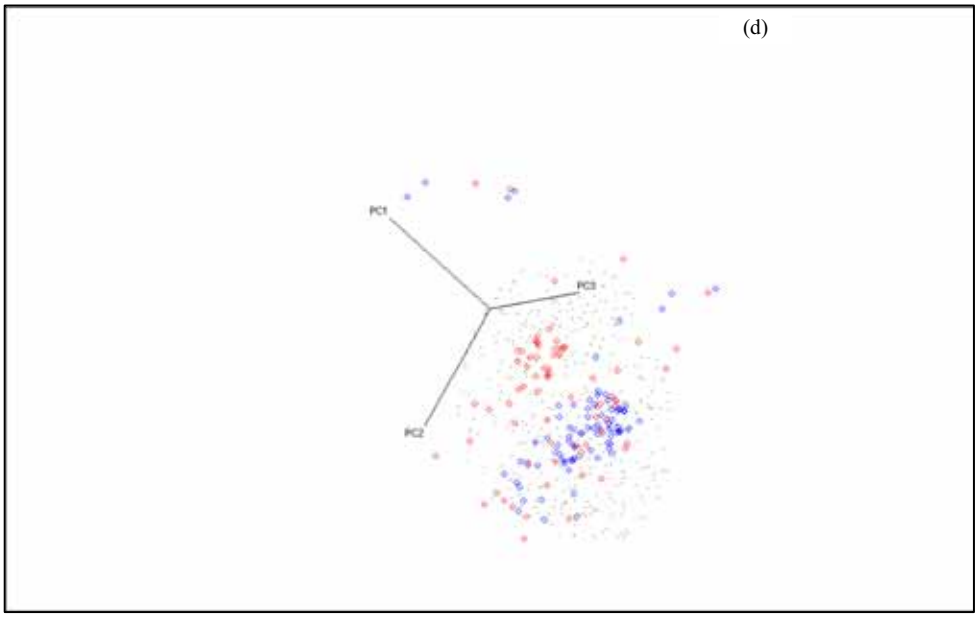
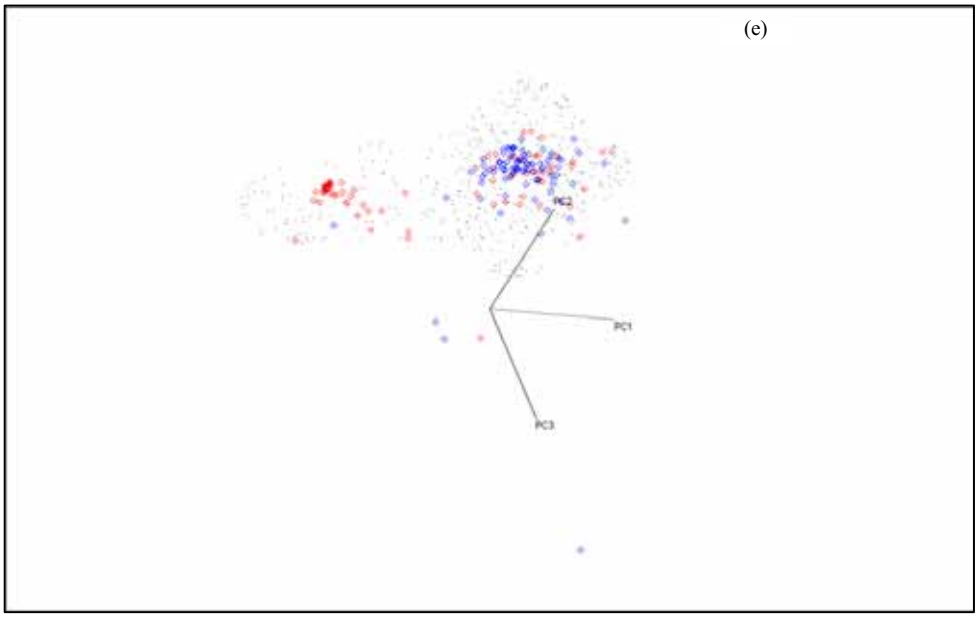



**Figura 53** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco I, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (conclusão).




**Fonte**: Elaborado pelo autor (2013).

**Figura 54** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco I, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes.

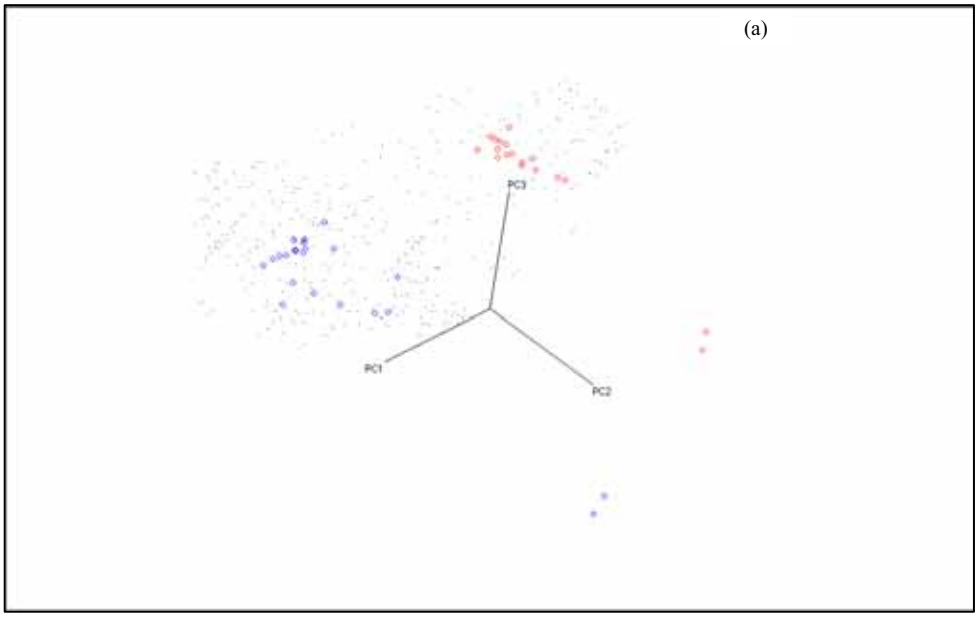
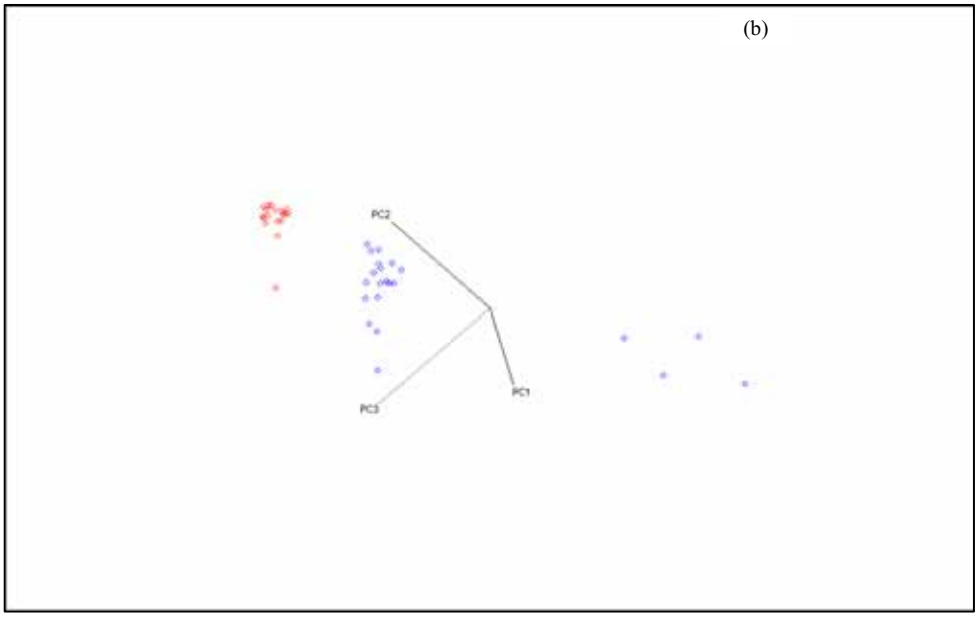


**Figura 54** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco I, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (continuação).



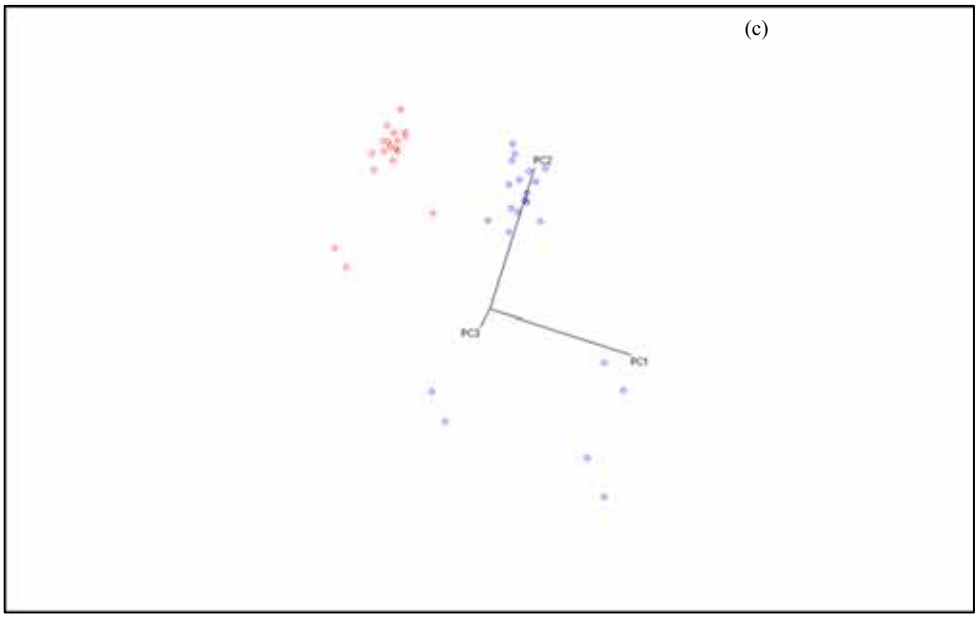
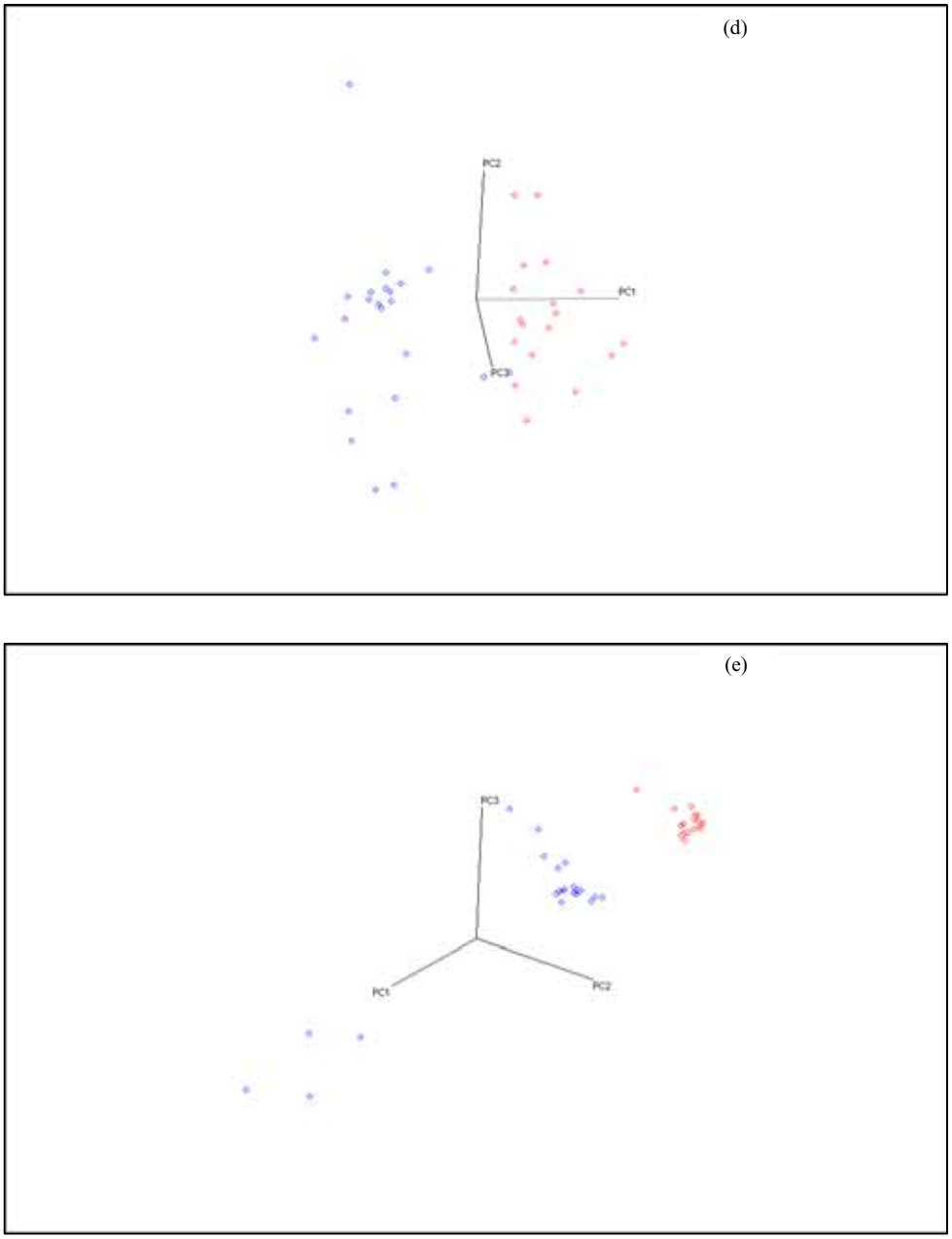



**Figura 54** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco I, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (conclusão).




**Fonte**: Elaborado pelo autor (2013).

**Tabela 13** - Resultado da classificação pela SIMCA, escalada pela variância e transformada por SNV, do Banco I com remoção de *loadings* iguais a zero.

| Grupo | Classe[5] | Previsão C | Previsão NC | Sem classe | Acerto/% |
|---|---|---|---|---|---|
| Treinamento | C | 82 | 8 | 0 | 91 |
| | NC | 7 | 65 | 0 | 90 |
| Previsão | C | 10 | 8 | 4 | 45 |
| | NC | 0 | 6 | 12 | 33 |

[5] C – amostras conformes, NC – amostras não conformes

**Fonte**: Elaborado pelo autor (2013).

### 5.7.2 Análise Classificatória por SIMCA para o Banco II

Por fim, os resultados das SIMCA para os grupos de treinamento para o Banco II são exibidos na tabela 14. Para o grupo de treinamento, a predição de acerto foram de 75% para as amostras conformes (sem pré-processamento e transformada por logaritmo na base 10) e 92% para as amostras não conformes (escalada pela variância e transformada por SNV).

Já os resultados das SIMCA para os grupos de previsão para o Banco II são exibidos na tabela 15. Para o grupo de treinamento, a predição de acerto foram de 14% para as amostras conformes (sem pré-processamento e transformada por logaritmo na base 10) e 95% para as amostras não conformes (escalada pela variância e transformada por SNV). E, semelhante ao Banco I, há um grande erro na classificação entre conforme e não conforme. A tabela 16, análise escalada pela variância e transformada por SNV, também corrobora o que fora descrito para o Banco I, quanto à eliminação de variáveis.

A figura 55, para o grupo de treinamento do Banco II, é similar à figura 53. As amostras adulteradas estão visivelmente separadas do aglomerado conforme/não conforme. No grupo de previsão, na figura 56, é mais visível a separação das classes, posto que mais amostras adulteradas que não conformes foram usadas no cálculo, semelhante ao descrito para a figura 54.

**Tabela 14** - Resultado da classificação pela SIMCA para grupo de treinamento (114 amostras conformes e 239 amostras não conformes) para o Banco II. As melhores classificações estão destacadas.

| Pré-processamentos | Transformadas | Classe[6] | Previsão C | Previsão NC | Sem classe | Acerto/% |
|---|---|---|---|---|---|---|
| Nenhum | Log10 | C | 86 | 28 | 0 | 75 |
| | | NC | 66 | 171 | 3 | 72 |
| Escala de Variância | Nenhuma | C | 67 | 47 | 0 | 59 |
| | | NC | 27 | 208 | 5 | 87 |
| Escala de Variância | Log10 | C | 79 | 35 | 0 | 69 |
| | | NC | 38 | 197 | 5 | 82 |
| Escala de Variância | SNV | C | 62 | 51 | 1 | 54 |
| | | NC | 18 | 219 | 3 | 92 |
| Escala de Variância | Smoothing | C | 67 | 47 | 0 | 59 |
| | | NC | 26 | 208 | 6 | 87 |

[6] C – amostras conformes, NC – amostras não conformes

**Fonte**: Elaborado pelo autor (2013).

**Tabela 15** - Resultado da classificação pela SIMCA para grupo de previsão (28 amostras conformes e 60 amostras não conformes) para o Banco II. As melhores classificações estão destacadas.

| Pré-processamentos | Transformadas | Classe[7] | Previsão C | Previsão NC | Sem classe | Acerto/% |
|---|---|---|---|---|---|---|
| Nenhum | Log10 | C | 4 | 23 | 0 | 14 |
| | | NC | 4 | 55 | 0 | 92 |
| Escala de Variância | Nenhuma | C | 1 | 26 | 0 | 4 |
| | | NC | 3 | 55 | 1 | 92 |
| Escala de Variância | Log10 | C | 1 | 26 | 0 | 4 |
| | | NC | 1 | 57 | 1 | 95 |
| Escala de Variância | SNV | C | 0 | 27 | 0 | 0 |
| | | NC | 6 | 47 | 6 | 78 |
| Escala de Variância | *Smoothing* | C | 1 | 26 | 0 | 4 |
| | | NC | 3 | 56 | 0 | 93 |

[7] C – amostras conformes, NC – amostras não conformes

**Fonte**: Elaborado pelo autor (2013).

**Tabela 16** - Resultado da classificação pela SIMCA, escalada pela variância e transformada por SNV, do Banco II com remoção de *loadings* iguais a zero.

| Grupo | Classe[8] | Previsão C | Previsão NC | Sem classe | Acerto/% |
|---|---|---|---|---|---|
| Treinamento | C | 70 | 43 | 1 | 61 |
| | NC | 32 | 204 | 3 | 85 |
| Previsão | C | 0 | 27 | 1 | 0 |
| | NC | 5 | 51 | 5 | 85 |

[8] C – amostras conformes, NC – amostras não conformes

**Fonte**: Elaborado pelo autor (2013).

**Figura 55** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco II, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes.

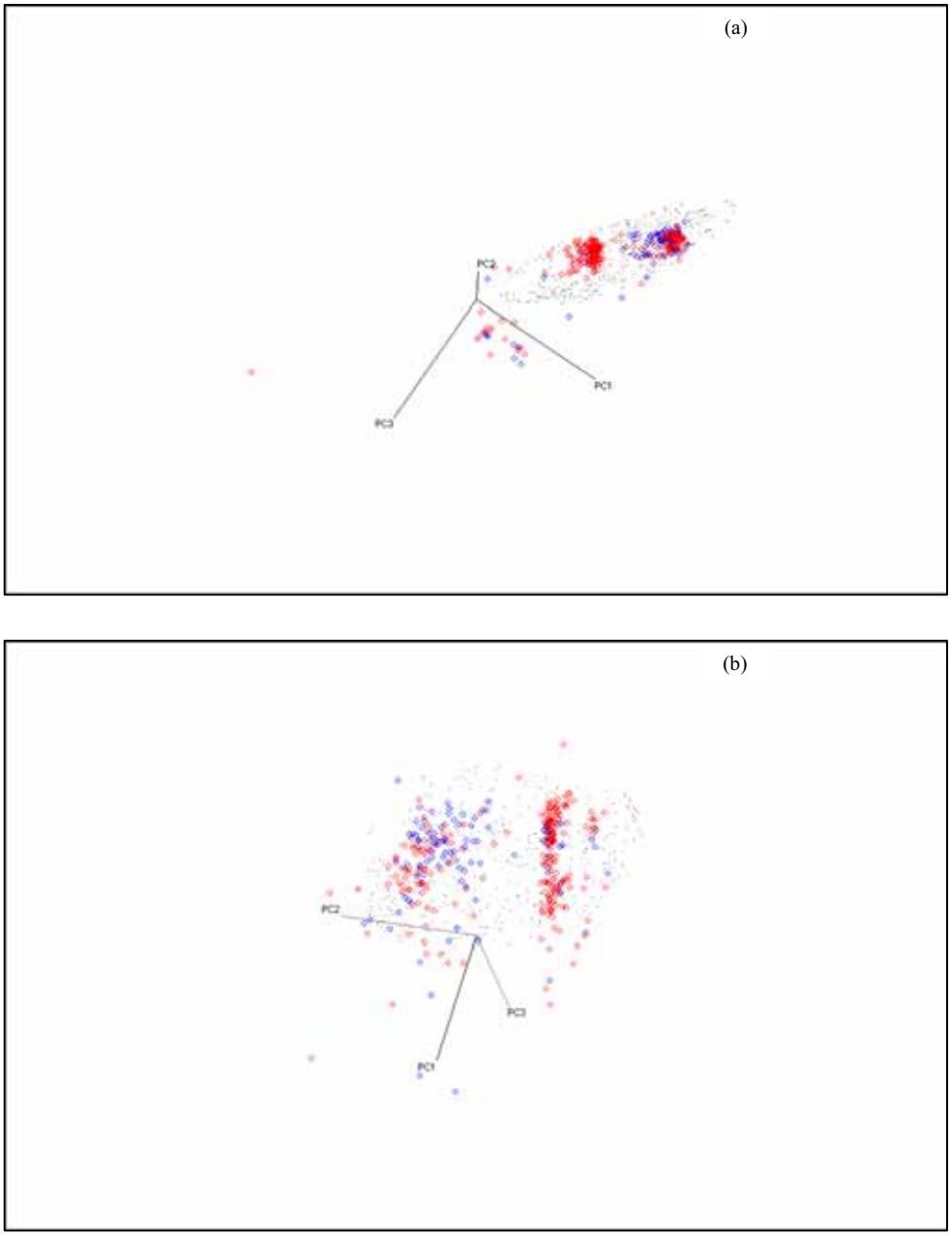
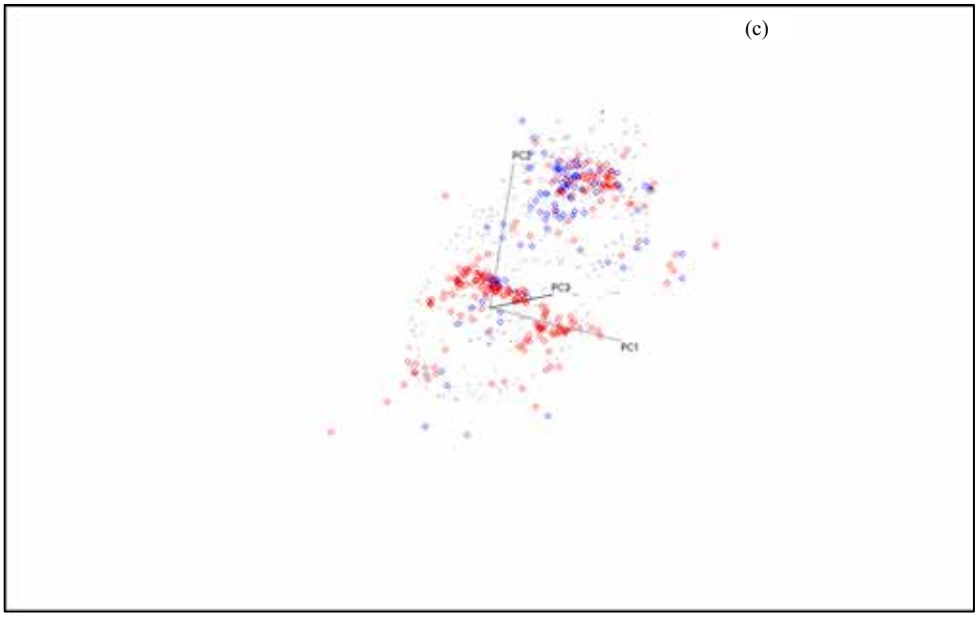


**Figura 55** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco II, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (continuação).



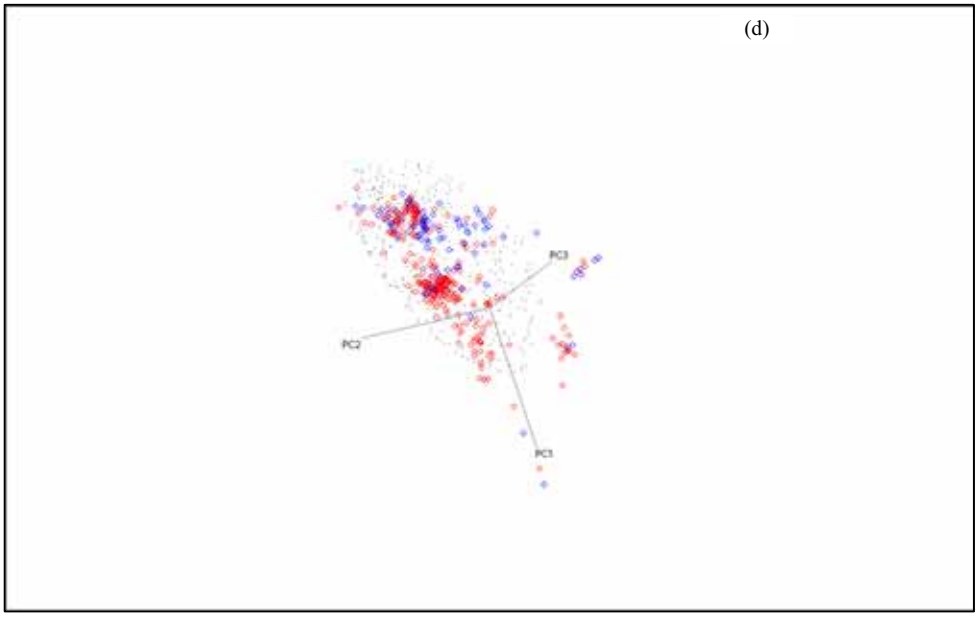
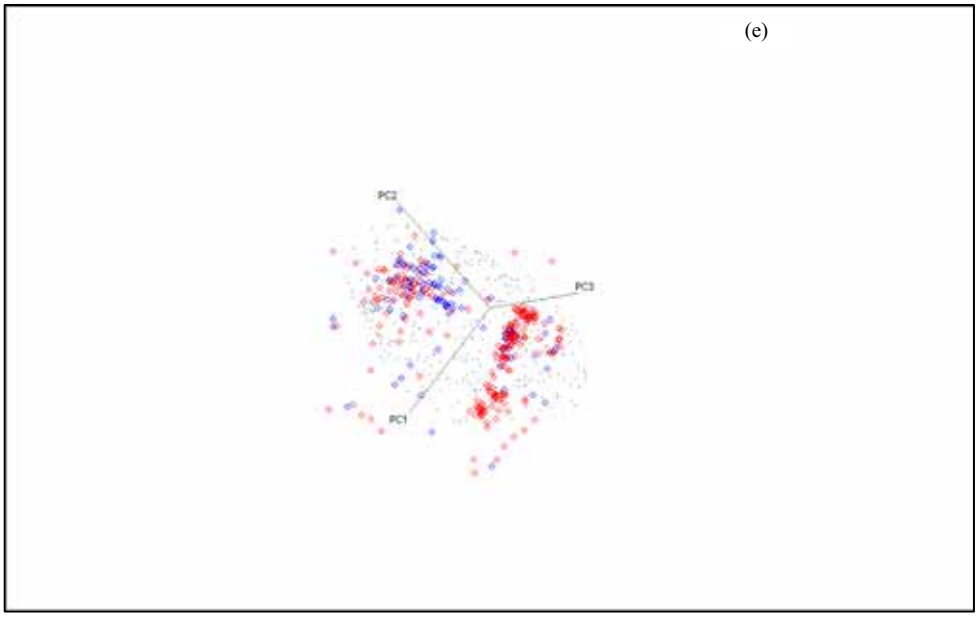



**Figura 55** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco II, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (conclusão).




**Fonte**: Elaborado pelo autor (2013).

**Figura 56** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco II, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes.

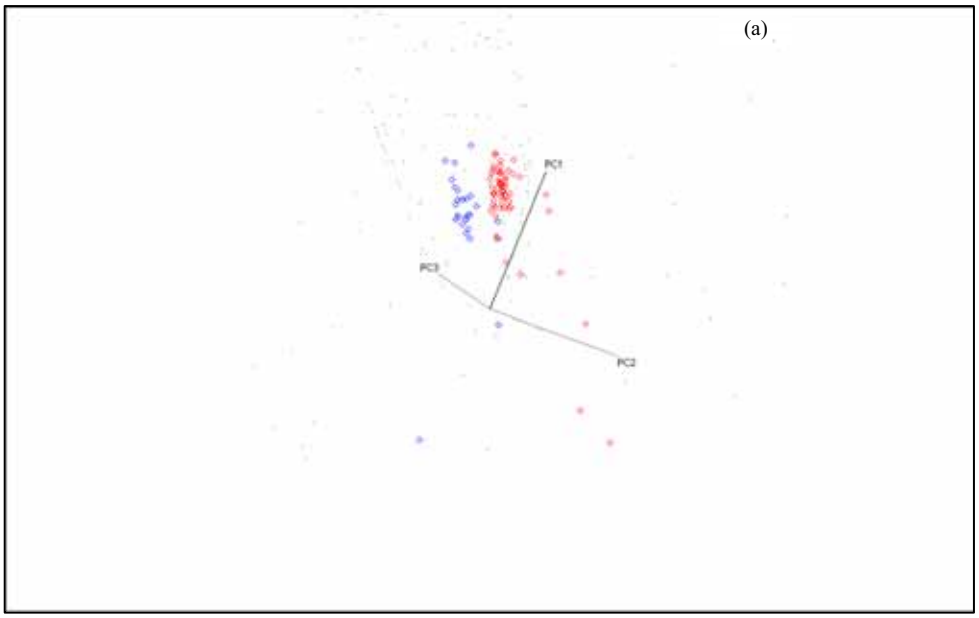
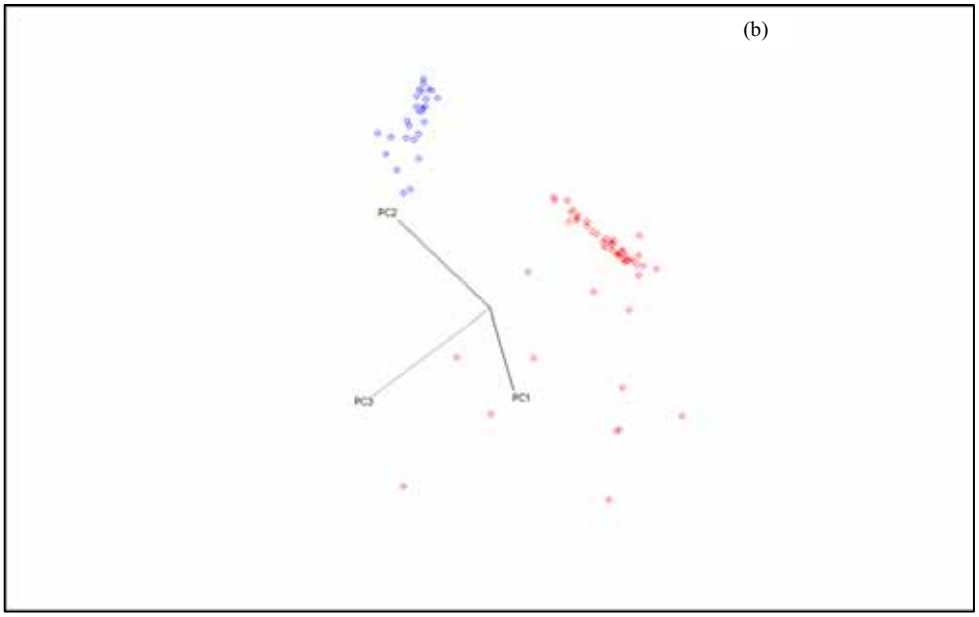


**Figura 56** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco II, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (continuação).



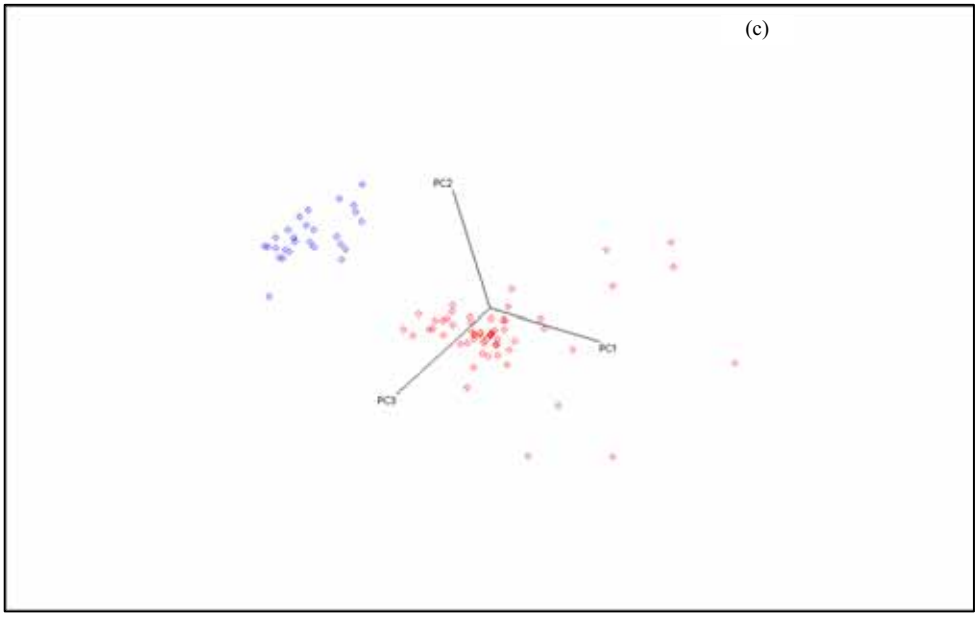
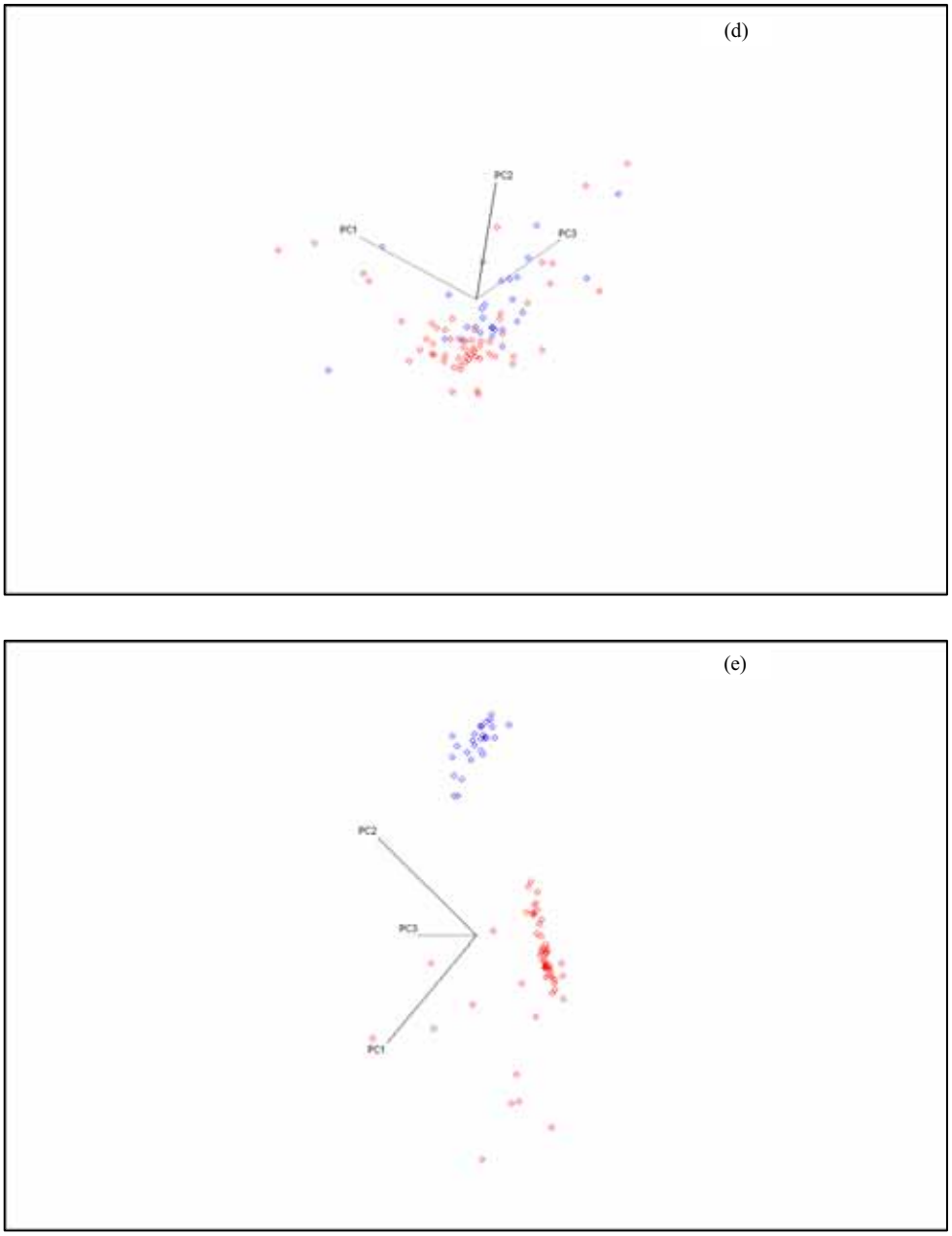



**Figura 56** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco II, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (conclusão).




**Fonte**: Elaborado pelo autor (2013).

## *5.8 Análise Classificatória das Amostras Conformes e Adulteradas*

Conforme fora explicitado na seção 5.7, as amostras conformes e não conformes diferem muito pouco em suas características físico-químicas. Portanto, as amostras não conformes foram removidas dos Bancos I e II e a as análises quimiométricas, PCA e SIMCA, foram aplicadas somente as amostras conformes e adulteradas de cada grupo. Para fácil entendimento, o Banco I, excluindo suas amostras não conformes, chamar-se-á Banco I-A e o Banco II, de semelhante modo, de Banco II-A.

### 5.8.1 Análise de Componentes Principais das Amostras Conformes e Adulteradas

Em ambos os Bancos, realizou-se a investigação dos pré-tratamentos ótimos pela PCA, assim como fora descrito na seção 5.6.

Os cinco pré-processamentos e transformadas que apresentaram maior percentagem de descrição do Banco I-A foram: sem pré-processamento e transformada por logaritmo na base 10, escalada pela variância e sem transformada, escalada pela variância e transformada por logaritmo na base 10, escalada pela variância e transformada por SNV e escalada pela variância e transformada por *smoothing*. Estes mesmos cinco pré-tratamentos também melhor descreveram o Banco II-A. Isto é comprovado pelas tabelas 17, Banco I-A, e 18, Banco II-A.

**Tabela 17** - Investigação do melhor pré-tratamento, através da Análise de Componentes Principais, do Banco I-A. Os cinco pré-tratamentos que obtiveram as melhores respostas estão em destaque.

| Pré-processamento | Transformada | PC3/% | PC7/% | PC10/% |
|---|---|---|---|---|
| Nenhum | Nenhuma | 94,5292 | 97,0665 | 97,9351 |
| | Primeira Derivada | 91,4644 | 95,5824 | 96,9774 |
| | Segunda Derivada | 85,9990 | 92,6852 | 94,8791 |
| | Log10 | 99,9220 | 99,9671 | 99,9759 |
| | MSC | 91,0711 | 96,1318 | 97,6306 |
| | SNV | 91,8539 | 95,6954 | 97,3115 |
| | *Smoothing* | 95,2000 | 97,6263 | 98,3938 |
| Auto-escalado | Nenhuma | 82,1564 | 90,9518 | 93,4233 |
| | Primeira Derivada | 56,2217 | 71,5229 | 76,9687 |
| | Segunda Derivada | 52,4858 | 66,6593 | 72,2563 |
| | Log10 | 87,0220 | 93,9748 | 95,6248 |
| | MSC | 72,6645 | 84,0202 | 87,9595 |
| | SNV | 76,0313 | 85,3226 | 88,6081 |
| | *Smoothing* | 82,0885 | 90,8465 | 93,3247 |
| Escalado pela Média | Nenhuma | 79,8312 | 88,2472 | 91,6012 |
| | Primeira Derivada | 71,8590 | 84,4429 | 89,3488 |
| | Segunda Derivada | 66,5136 | 81,1046 | 86,4145 |
| | Log10 | 86,8769 | 92,4418 | 94,3782 |
| | MSC | 63,7230 | 81,8231 | 88,8867 |
| | SNV | 55,2233 | 76,1734 | 85,1900 |
| | *Smoothing* | 66,5065 | 90,3186 | 93,3448 |
| Escalado pela Amplitude | Nenhuma | 95,0542 | 97,5200 | 98,2511 |
| | Primeira Derivada | 95,4065 | 97,1581 | 97,7289 |
| | Segunda Derivada | 94,7603 | 96,4569 | 97,0708 |
| | Log10 | 97,5823 | 98,9759 | 99,2865 |
| | MSC | 96,4971 | 98,2685 | 98,7208 |
| | SNV | 94,7907 | 97,2700 | 97,8988 |
| | *Smoothing* | 95,1020 | 97,6049 | 98,3393 |
| *Escalado pela Variância* | Nenhuma | 99,0050 | 99,5311 | 99,6570 |
| | Primeira Derivada | 80,9794 | 87,8961 | 90,4519 |
| | Segunda Derivada | 78,8902 | 85,3251 | 87,9310 |
| | Log10 | 99,9840 | 99,9930 | 99,9950 |
| | MSC | 97,3482 | 98,6320 | 98,9665 |
| | SNV | 98,8785 | 99,3554 | 99,5023 |
| | *Smoothing* | 98,9984 | 99,5246 | 99,6516 |

**Fonte**: Elaborado pelo autor (2013).

**Tabela 18** - Investigação do melhor pré-tratamento, através da Análise de Componentes Principais, do Banco II-A. Os cinco pré-tratamentos que obtiveram as melhores respostas estão em destaque.

| Pré-processamento | Transformada | PC3/% | PC7/% | PC10/% |
|---|---|---|---|---|
| Nenhum | Nenhuma | 91,5125 | 95,5459 | 96,8557 |
| | Primeira Derivada | 87,8130 | 93,2610 | 95,4506 |
| | Segunda Derivada | 82,9940 | 89,4697 | 92,4281 |
| | Log10 | 99,9060 | 99,9534 | 99,9659 |
| | MSC | 89,3117 | 94,9161 | 96,3879 |
| | SNV | 89,9742 | 94,8736 | 96,3050 |
| | Smoothing | 92,2029 | 96,1721 | 97,3934 |
| Auto-escalado | Nenhuma | 73,1530 | 86,9220 | 90,0130 |
| | Primeira Derivada | 49,1871 | 65,0289 | 71,2106 |
| | Segunda Derivada | 47,7737 | 61,3788 | 66,9370 |
| | Log10 | 79,0951 | 89,7885 | 92,2026 |
| | MSC | 69,1488 | 81,5258 | 84,8364 |
| | SNV | 69,4274 | 81,7025 | 84,9669 |
| | Smoothing | 73,0727 | 86,9717 | 90,0802 |
| Escalado pela Média | Nenhuma | 76,7350 | 86,5356 | 90,1304 |
| | Primeira Derivada | 68,7253 | 81,7855 | 87,3492 |
| | Segunda Derivada | 61,0042 | 75,1928 | 82,1163 |
| | Log10 | 77,3838 | 87,8984 | 90,9140 |
| | MSC | 56,4425 | 74,8207 | 82,0977 |
| | SNV | 56,6192 | 73,2824 | 80,5522 |
| | Smoothing | 78,4866 | 88,3325 | 91,6708 |
| Escalado pela Amplitude | Nenhuma | 92,9117 | 96,5321 | 97,4139 |
| | Primeira Derivada | 97,3649 | 98,2476 | 98,5154 |
| | Segunda Derivada | 96,9659 | 97,8705 | 98,1284 |
| | Log10 | 96,5501 | 98,5203 | 98,9379 |
| | MSC | 96,5541 | 98,1749 | 98,5913 |
| | SNV | 95,3617 | 97,4454 | 98,0192 |
| | Smoothing | 93,0797 | 96,6791 | 97,5444 |
| Escalado pela Variância | Nenhuma | 98,9334 | 99,4911 | 99,6120 |
| | Primeira Derivada | 70,9055 | 80,3420 | 84,1795 |
| | Segunda Derivada | 69,9662 | 78,4329 | 81,7145 |
| | Log10 | 99,9834 | 99,9922 | 99,9941 |
| | MSC | 96,0013 | 97,8704 | 98,2690 |
| | SNV | 98,3369 | 99,0443 | 99,2198 |
| | *Smoothing* | 98,9264 | 99,4917 | 99,6138 |

**Fonte**: Elaborado pelo autor (2013).

Do mesmo modo que descrito na seção 5.6, fez-se uma nova PCA com validação cruzada, removendo 5 vetores de amostras para cada análise, com os cinco pré-tratamentos que apresentaram que melhor descreveram os Bancos I-A e II-A. As respostas destas análises estão discorridas nas tabelas 19 e 20, para o Banco I-A, e nas tabelas 21 e 22, Banco II-A.

**Tabela 19** - Percentagem da informação descrita para cada componente principal na investigação do pré-tratamento ótimo para a PCA, realizada com validação cruzada removendo 5 vetores de amostras, do Banco I-A. O pré-tratamento com a melhor resposta está em destaque.

| **Pré-processamento** | Nenhum | Escala de Variância | Escala de Variância | Escala de Variância | Escala de Variância |
|---|---|---|---|---|---|
| **Transformada** | Log 10 | Nenhum | Log 10 | SNV | *Smoothing* |
| **1 PC / %** | 99,8164 | 97,5068 | 99,9506 | 97,8835 | 97,4875 |
| **2 PC / %** | 0,0674 | 1,1485 | 0,0262 | 0,7250 | 1,1568 |
| **3 PC / %** | 0,0382 | 0,3496 | 0,0072 | 0,2701 | 0,3541 |
| **4 PC / %** | 0,0277 | 0,2500 | 0,0045 | 0,1614 | 0,2528 |
| **5 PC / %** | 0,0071 | 0,1316 | 0,0026 | 0,1486 | 0,1282 |
| **6 PC / %** | 0,0056 | 0,0777 | 0,0010 | 0,0962 | 0,0778 |
| **7 PC / %** | 0,0047 | 0,0668 | 0,0009 | 0,0707 | 0,0674 |
| **8 PC / %** | 0,0032 | 0,0575 | 0,0008 | 0,0535 | 0,0579 |
| **9 PC / %** | 0,0030 | 0,0395 | 0,0007 | 0,0501 | 0,0399 |
| **10 PC / %** | 0,0026 | 0,0290 | 0,0006 | 0,0433 | 0,0292 |

**Fonte**: Elaborado pelo autor (2013).

**Tabela 20** - Percentagem cumulativa da informação descrita para cada componente principal na investigação do pré-tratamento ótimo para a PCA, realizada com validação cruzada removendo 5 vetores de amostras, do Banco I-A. O pré-tratamento com a melhor resposta está em destaque.

| **Pré-processamento** | Nenhum | Escala de Variância | Escala de Variância | Escala de Variância | Escala de Variância |
|---|---|---|---|---|---|
| **Transformada** | Log 10 | Nenhum | Log 10 | SNV | *Smoothing* |
| **1 PC / %** | 99,8164 | 97,5068 | 99,9506 | 97,8835 | 97,4875 |
| **2 PC / %** | 99,8838 | 98,6553 | 99,9768 | 98,6085 | 98,6443 |
| **3 PC / %** | 99,9220 | 99,0050 | 99,9840 | 98,8785 | 98,9984 |
| **4 PC / %** | 99,9496 | 99,2550 | 99,9885 | 99,0399 | 99,2513 |
| **5 PC / %** | 99,9568 | 99,3866 | 99,9911 | 99,1884 | 99,3794 |
| **6 PC / %** | 99,9624 | 99,4643 | 99,9921 | 99,2847 | 99,4572 |
| **7 PC / %** | 99,9671 | 99,5311 | 99,9930 | 99,3554 | 99,5246 |
| **8 PC / %** | 99,9703 | 99,5886 | 99,9938 | 99,4089 | 99,5825 |
| **9 PC / %** | 99,9733 | 99,6281 | 99,9945 | 99,4590 | 99,6224 |
| **10 PC / %** | 99,9759 | 99,6570 | 99,9950 | 99,5023 | 99,6516 |

**Fonte**: Elaborado pelo autor (2013).

**Tabela 21** - Percentagem da informação descrita para cada componente principal na investigação do pré-tratamento ótimo para a PCA, realizada com validação cruzada removendo 5 vetores de amostras, do Banco II-A. O pré-tratamento com a melhor resposta está em destaque.

| **Pré-processamento** | Nenhum | Escala de Variância | Escala de Variância | Escala de Variância | Escala de Variância |
|---|---|---|---|---|---|
| **Transformada** | Log 10 | Nenhum | Log 10 | SNV | *Smoothing* |
| **1 PC / %** | 99,7884 | 97,0863 | 99,9461 | 97,0774 | 97,4226 |
| **2 PC / %** | 0,0780 | 1,5162 | 0,0315 | 1,5170 | 0,6103 |
| **3 PC / %** | 0,0397 | 0,3310 | 0,0058 | 0,3320 | 0,3040 |
| **4 PC / %** | 0,0177 | 0,2143 | 0,0040 | 0,2177 | 0,2895 |
| **5 PC / %** | 0,0146 | 0,1502 | 0,0024 | 0,1541 | 0,1735 |
| **6 PC / %** | 0,0080 | 0,1181 | 0,0015 | 0,1183 | 0,1575 |
| **7 PC / %** | 0,0071 | 0,0751 | 0,0010 | 0,0753 | 0,0868 |
| **8 PC / %** | 0,0053 | 0,0506 | 0,0007 | 0,0511 | 0,0705 |
| **9 PC / %** | 0,0043 | 0,0378 | 0,0007 | 0,0382 | 0,0564 |
| **10 PC / %** | 0,0029 | 0,0326 | 0,0005 | 0,0328 | 0,0487 |

**Fonte**: Elaborado pelo autor (2013).

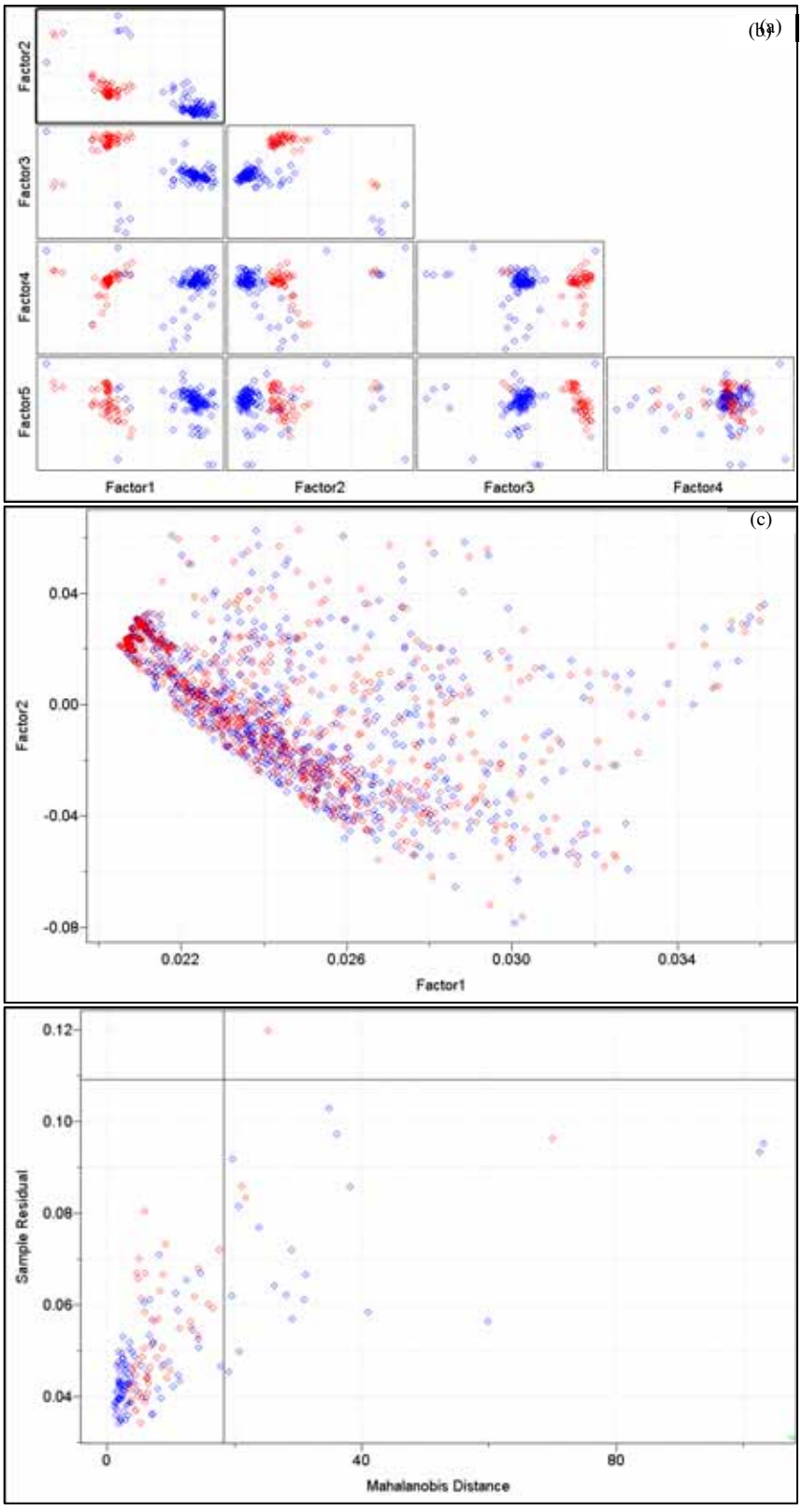
**Tabela 22** - Percentagem cumulativa da informação descrita para cada componente principal na investigação do pré-tratamento ótimo para a PCA, realizada com validação cruzada removendo 5 vetores de amostras, do Banco II-A. O pré-tratamento com a melhor resposta está em destaque.

| **Pré-processamento** | Nenhum | Escala de Variância | Escala de Variância | Escala de Variância | Escala de Variância |
|---|---|---|---|---|---|
| **Transformada** | Log 10 | Nenhum | Log 10 | SNV | *Smoothing* |
| **1 PC / %** | 99,7884 | 97,0863 | 99,9461 | 97,0774 | 97,4226 |
| **2 PC / %** | 99,8663 | 98,6024 | 99,9776 | 98,5944 | 98,0329 |
| **3 PC / %** | 99,9060 | 98,9334 | 99,9834 | 98,9264 | 98,3369 |
| **4 PC / %** | 99,9237 | 99,1478 | 99,9873 | 99,1441 | 98,6265 |
| **5 PC / %** | 99,9383 | 99,2979 | 99,9897 | 99,2982 | 98,8000 |
| **6 PC / %** | 99,9463 | 99,4160 | 99,9912 | 99,4165 | 98,9575 |
| **7 PC / %** | 99,9534 | 99,4911 | 99,9922 | 99,4917 | 99,0443 |
| **8 PC / %** | 99,9587 | 99,5416 | 99,9930 | 99,5428 | 99,1147 |
| **9 PC / %** | 99,9630 | 99,5795 | 99,9936 | 99,5810 | 99,1712 |
| **10 PC / %** | 99,9659 | 99,6120 | 99,9941 | 99,6138 | 99,2198 |

**Fonte**: Elaborado pelo autor (2013).

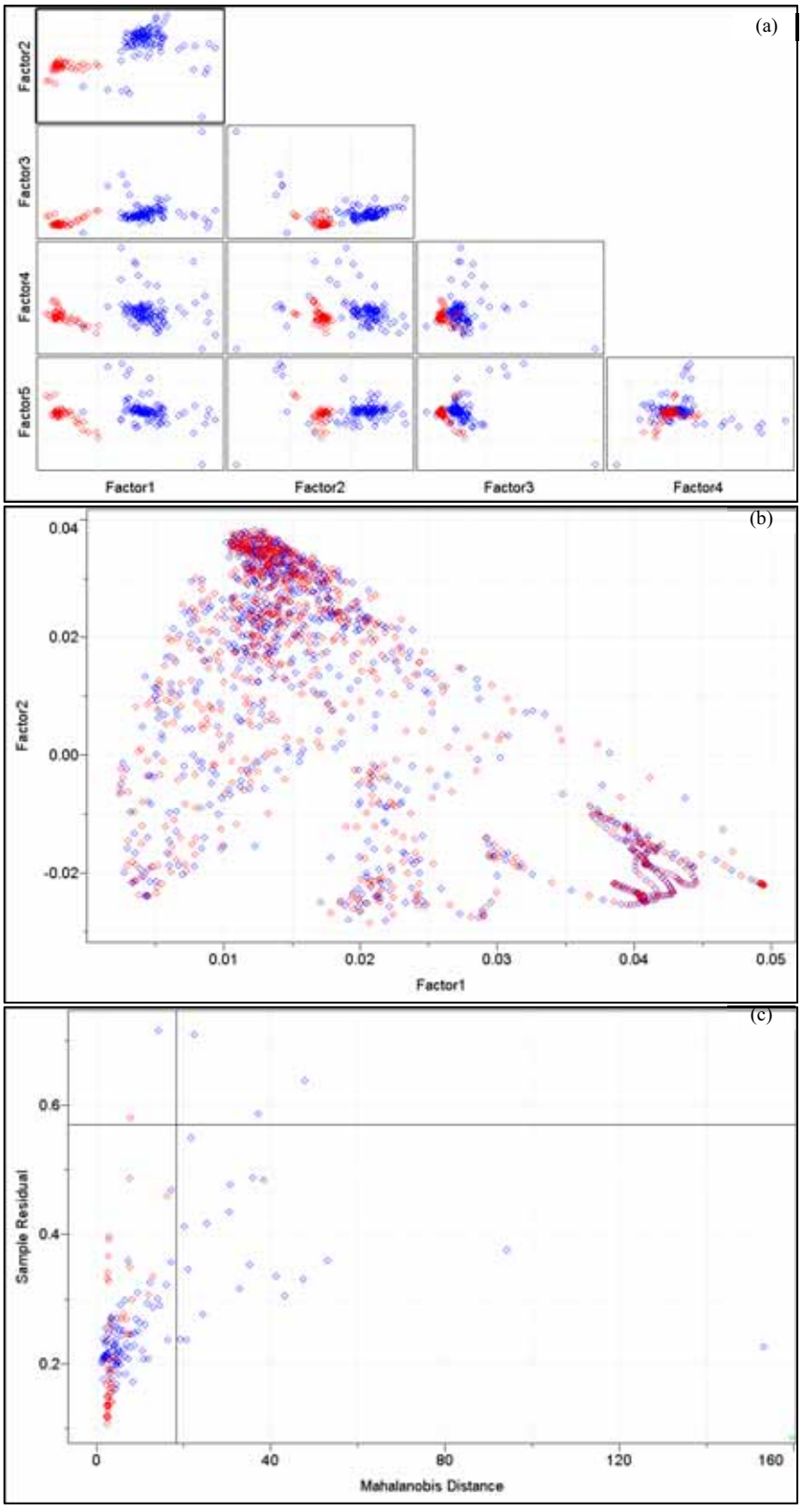
A partir destas análises, obtiveram-se as figuras 57 a 61. Estas representam os *scores*, *loadings* e resíduos para as PCA do Banco I-A, com seus cinco pré-tratamentos ótimos. Os pontos azuis representam as amostras conformes e os vermelhos, as amostras adulteradas. Pelos *scores*, nota-se uma melhor distinção entre os grupos de amostras do que o observado na PCA do Banco I. Já os *loadings* e os resíduos são semelhantes aos vistos nas figuras 37 a 42.

**Figura 57** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I-A, sem pré-processamento e transformada por logaritmo na base 10.



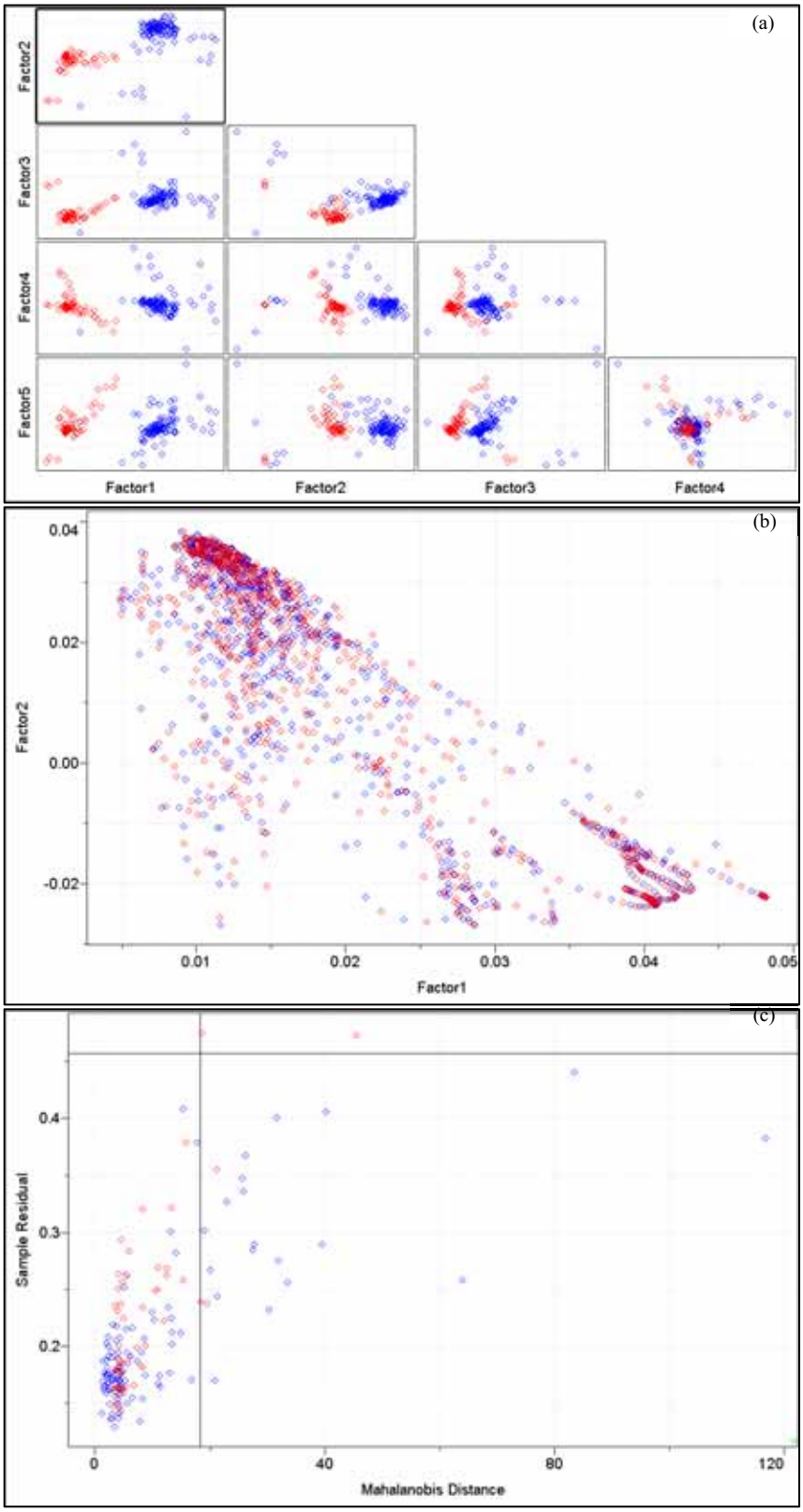

**Fonte**: Elaborado pelo autor (2013).

**Figura 58** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I-A, escalada pela variância e sem transformada.



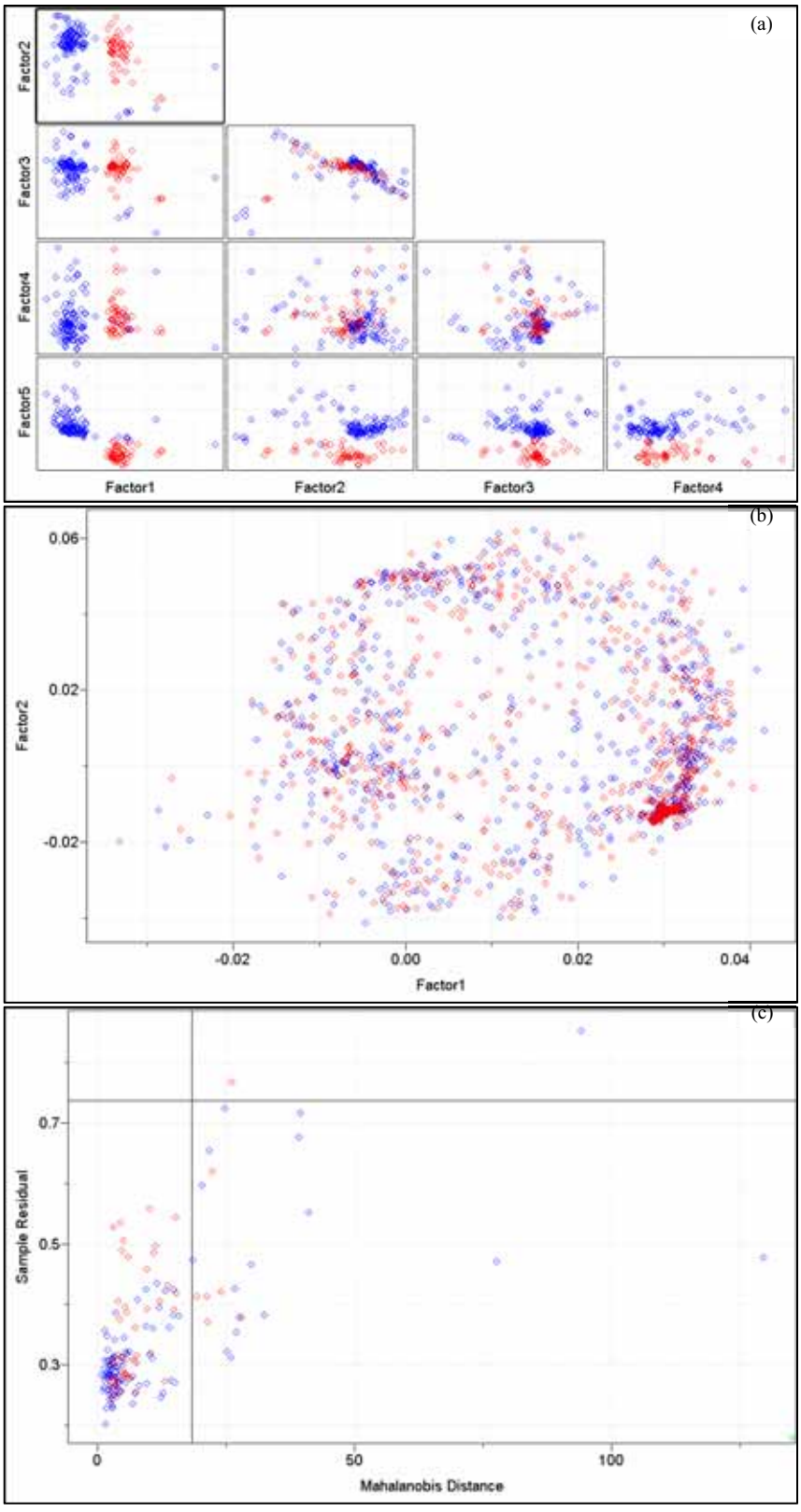

**Fonte**: Elaborado pelo autor (2013).

**Figura 59** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I-A, escalada pela variância e transformada por logaritmo na base 10.



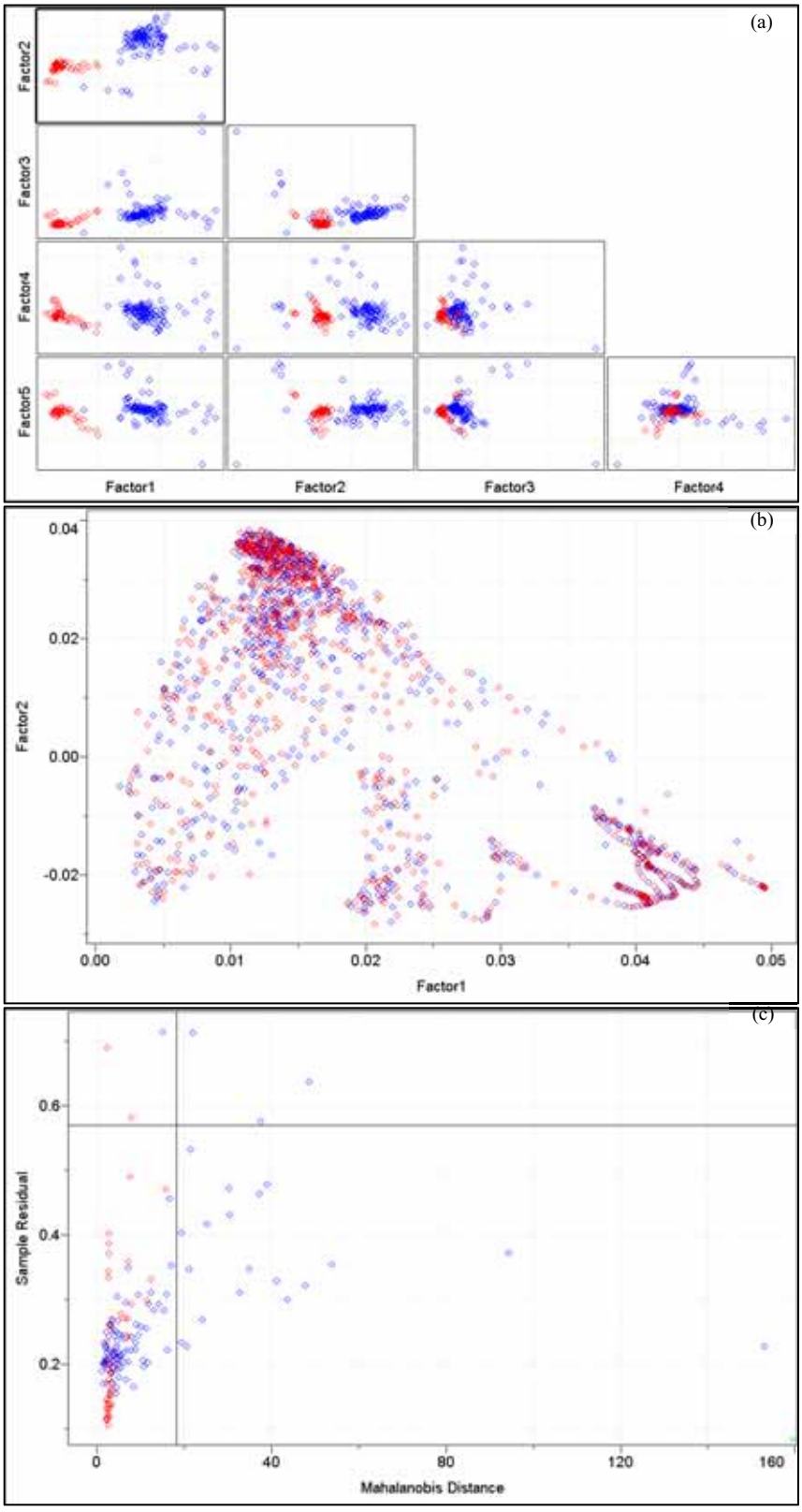

**Fonte**: Elaborado pelo autor (2013).

**Figura 60** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I-A, escalada pela variância e transformada por SNV.




**Fonte**: Elaborado pelo autor (2013).

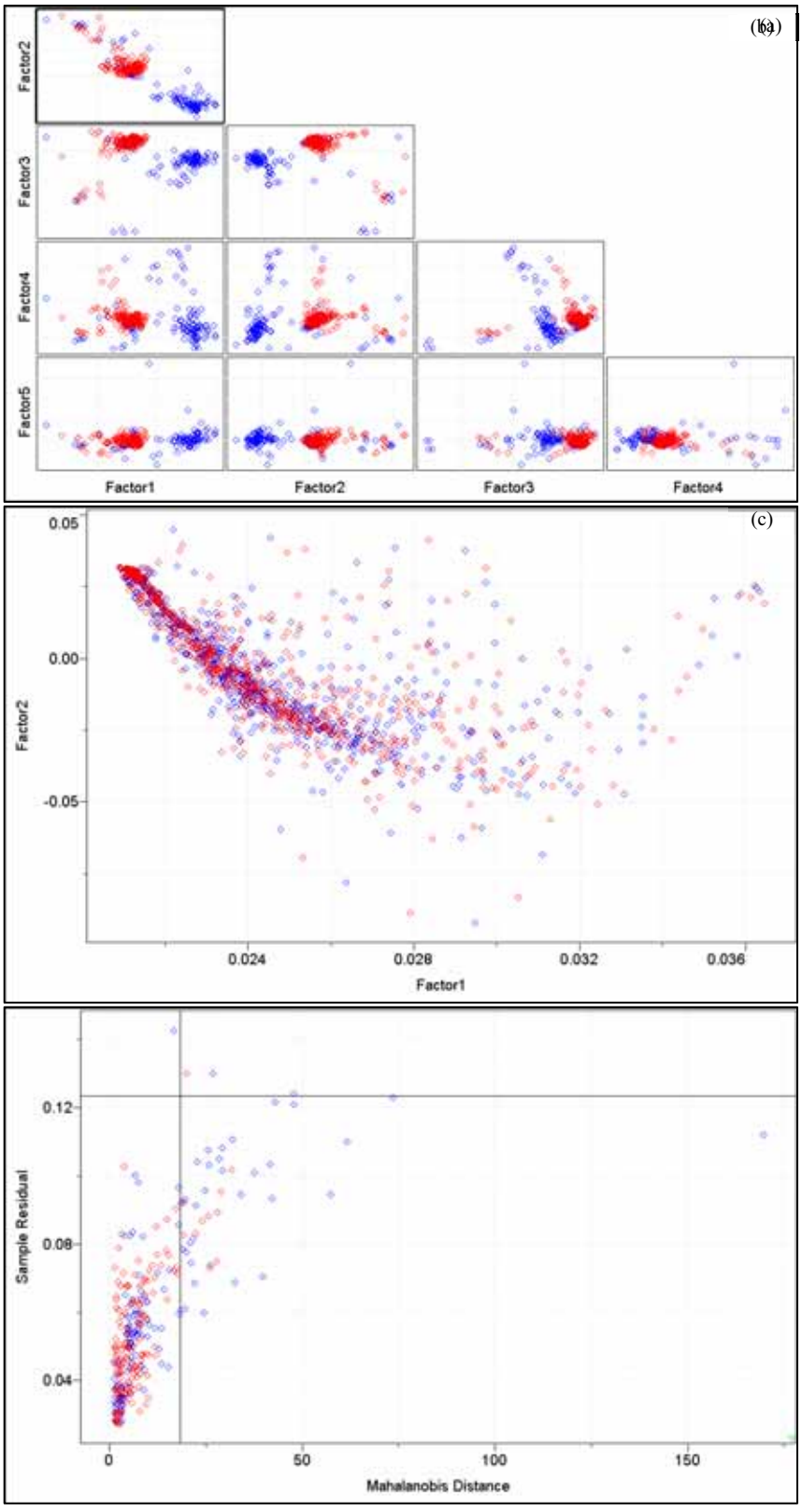
**Figura 61** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco I-A, escalada pela variância e transformada por *smoothing*.




**Fonte**: Elaborado pelo autor (2013).

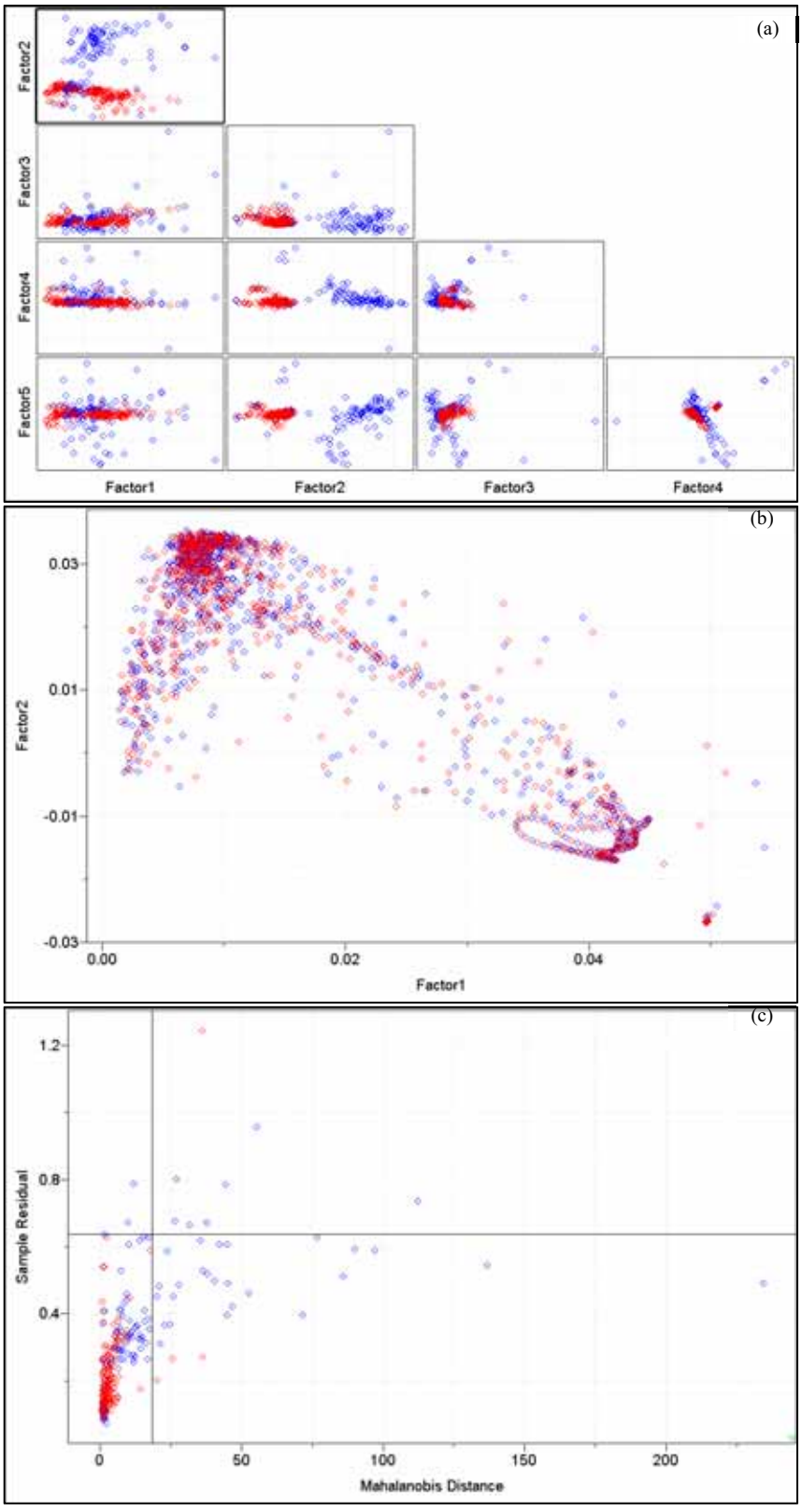
De mesmo modo, geraram-se as figuras 62 a 66, que indicam os *scores*, *loadings* e resíduos para as PCA do Banco II-A, com seus cinco pré-tratamentos ótimos. Os pontos azuis representam as amostras conformes e os vermelhos, as amostras adulteradas. Semelhante ao Banco I-A, os *scores* apresentam uma melhor distinção entre os grupos de amostras do que o observado na PCA do Banco II. Tanto os *loadings* como os resíduos são semelhantes aos vistos nas figuras 44 a 48.

**Figura 62** - Representação de (a) *scores*, (b) *loadings* e (c) *resíduos* da PCA do Banco II-A, sem pré-processamento e transformada por logaritmo na base 10.



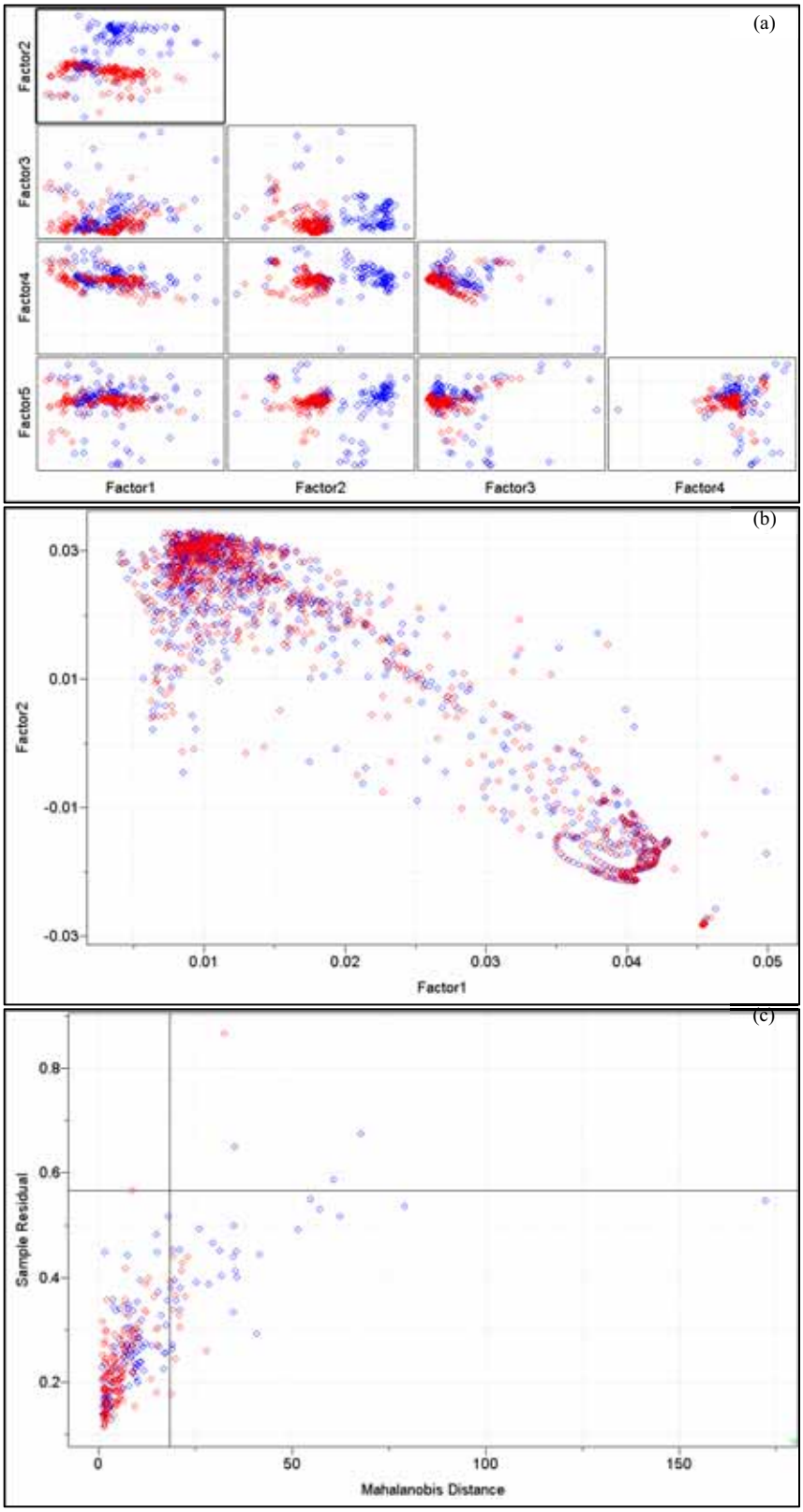

**Fonte**: Elaborado pelo autor (2013).

**Figura 63** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II-A, escalada pela variância e sem transformada.



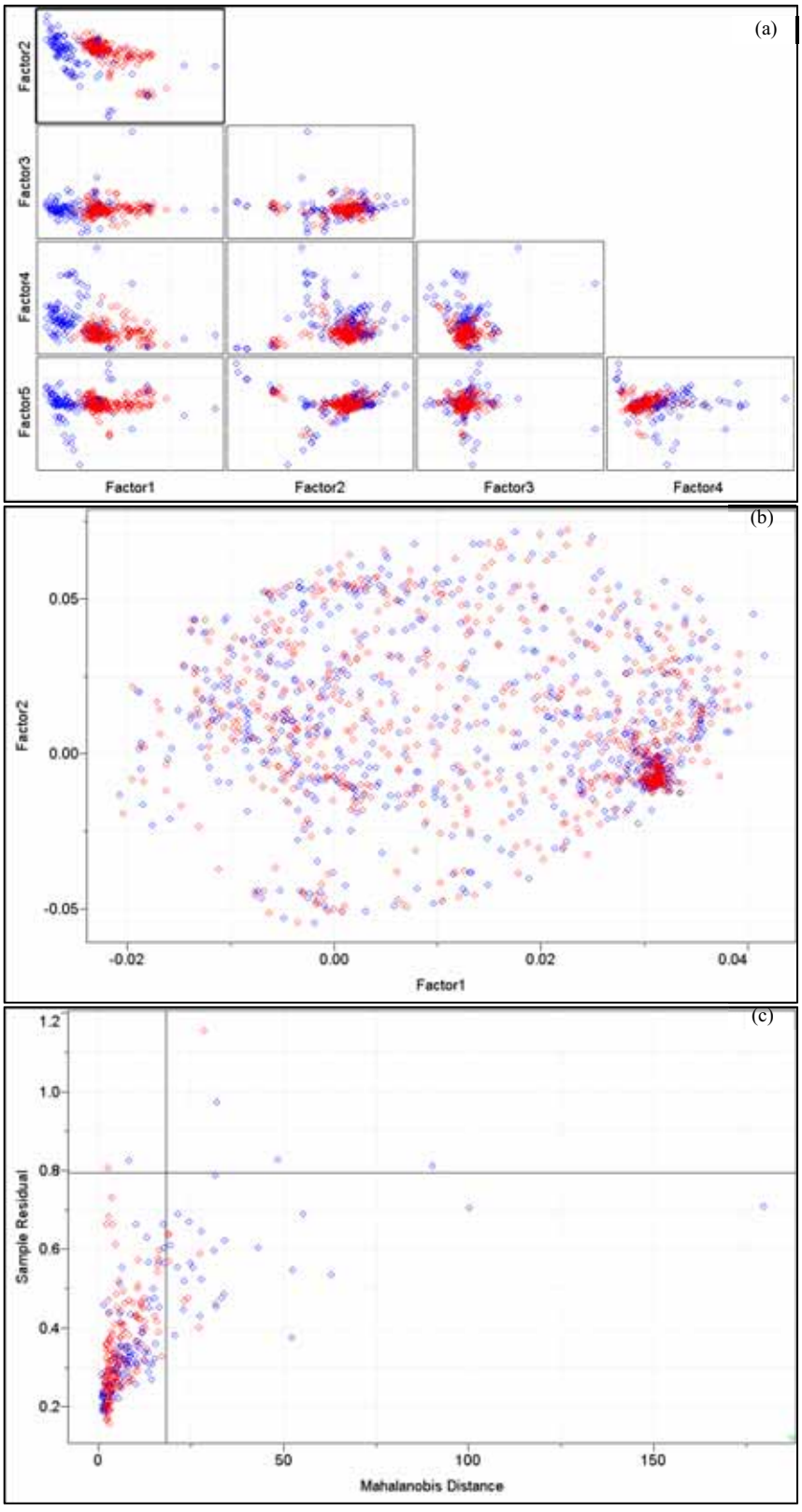

**Fonte**: Elaborado pelo autor (2013).

**Figura 64** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II-A, escalada pela variância e transformada por logaritmo na base 10.



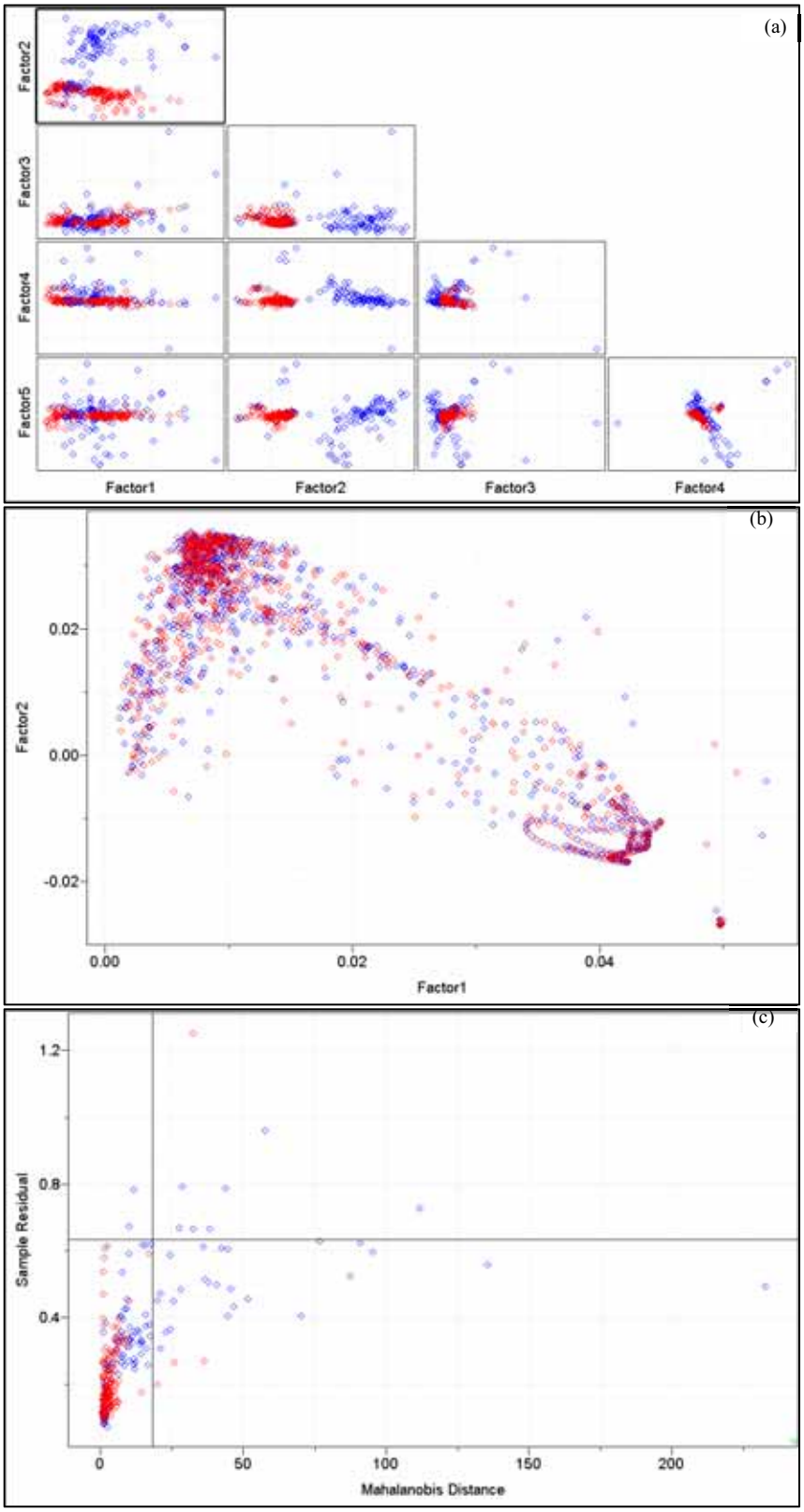

**Fonte**: Elaborado pelo autor (2013).

**Figura 65** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II-A, escalada pela variância e transformada por SNV.




**Fonte**: Elaborado pelo autor (2013).

**Figura 66** - Representação de (a) *scores*, (b) *loadings* e (c) resíduos da PCA do Banco II-A, escalada pela variância e transformada por *smoothing*.




**Fonte**: Elaborado pelo autor (2013).

### 5.8.2 Classificação das Amostras Conformes e Adulteradas por SIMCA

Similar ao que fora descrito na seção 5.7, a análise pela SIMCA foi feita para os Bancos I-A e II-A. Os pré-tratamentos utilizados foram os cinco descritos na seção 5.8.1.

Os grupos de treinamento e previsão foram calculados baseados nas equações 80 e 81.

Para o Banco I-A, o total de amostras é de $166$, sendo $112$ conformes e $54$ adulteradas. Assim, o grupo de treinamento será composto de $112.0{,}8 = 89{,}6 \cong 90$ amostras conformes e $54.0{,}8 = 43{,}2 = 43$ amostras adulteradas, totalizando $90 + 43 = 112.0{,}8 + 54.0{,}8 = 133$ amostras. Já o grupo de previsão contém $112.0{,}2 = 22{,}4 \cong 22$ amostras conformes e $54.0{,}2 = 10{,}8 \cong 11$ amostras adulteradas, totalizando $22 + 11 = 112.0{,}2 + 54.0{,}2 = 33$ amostras.

Já para o Banco II-A, o total de amostras é de $305$, sendo $142$ conformes e $163$ adulteradas. Deste modo, o grupo de treinamento será composto de $142.0{,}8 = 113{,}6 \cong 114$ amostras conformes e $163.0{,}8 = 130{,}4 = 130$ amostras adulteradas, totalizando $114 + 130 = 142.0{,}8 + 163.0{,}8 = 244$ amostras. Já o grupo de previsão contém $142.0{,}2 = 28{,}4 \cong 28$ amostras conformes e $163.0{,}2 = 32{,}6 \cong 33$ amostras adulteradas, totalizando $28 + 33 = 142.0{,}2 + 163.0{,}2 = 61$ amostras.

As tabelas 23 e 24 exibem o resultado da classificação pela SIMCA dos grupos de treinamento dos Bancos I-A e II-A respectivamente. Nota-se que o erro de previsão no Banco I-A é menor que no Banco I, devido à remoção das amostras não conformes, as quais apresentavam grande influencia na classificação, por serem semelhantes às amostras conformes.

Apesar de maior que o apresentado pelo Banco I-A, o erro gerado na classificação do Banco II-A ainda é menor que o do Banco II pelo mesmo motivo.

As figuras 67 e 68 mostram uma melhor distinção de classes para o grupo de treinamento dos Bancos I-A e II-A, do que o exposto pelas figuras 53, Banco I, e 55, Banco II. Comprovando, assim a alta similaridade com as amostras conformes, o que gerou o erro na classificação pela SIMCA.

**Tabela 23** - Resultado da classificação pela SIMCA para grupo de treinamento (90 amostras conformes e 43 amostras adulteradas) para o Banco I-A. As melhores classificações estão destacadas.

| Pré-processamentos | Transformadas | Classe[9] | Previsão C | Previsão A | Sem classe | Acerto/% |
|---|---|---|---|---|---|---|
| Nenhum | Log10 | C | 88 | 1 | 1 | 98 |
| | | A | 0 | 43 | 0 | 100 |
| Escala de Variância | Nenhuma | C | 89 | 0 | 1 | 99 |
| | | A | 0 | 43 | 0 | 100 |
| Escala de Variância | Log10 | C | 88 | 1 | 1 | 98 |
| | | A | 0 | 43 | 0 | 100 |
| Escala de Variância | SNV | C | 89 | 0 | 1 | 99 |
| | | A | 0 | 43 | 0 | 100 |
| Escala de Variância | Smoothing | C | 89 | 0 | 1 | 99 |
| | | A | 0 | 43 | 0 | 100 |

[9] C – amostras conformes, A – amostras adulteradas

**Fonte**: Elaborado pelo autor (2013).

**Tabela 24** - Resultado da classificação pela SIMCA para grupo de treinamento (114 amostras conformes e 130 amostras adulteradas) para o Banco II-A. As melhores classificações estão destacadas.

| Pré-processamentos | Transformadas | Classe[10] | Previsão C | Previsão A | Sem classe | Acerto/% |
|---|---|---|---|---|---|---|
| Nenhum | Log10 | C | 93 | 21 | 0 | 82 |
| | | A | 2 | 128 | 0 | 98 |
| Escala de Variância | Nenhuma | C | 93 | 21 | 0 | 82 |
| | | A | 2 | 128 | 0 | 98 |
| Escala de Variância | Log10 | C | 95 | 19 | 0 | 83 |
| | | A | 2 | 128 | 0 | 98 |
| Escala de Variância | SNV | C | 98 | 15 | 1 | 86 |
| | | A | 5 | 125 | 0 | 96 |
| Escala de Variância | Smoothing | C | 92 | 22 | 0 | 81 |
| | | A | 3 | 127 | 0 | 98 |

[10] C – amostras conformes, A – amostras adulteradas

**Fonte**: Elaborado pelo autor (2013).

**Figura 67** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco I-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes.

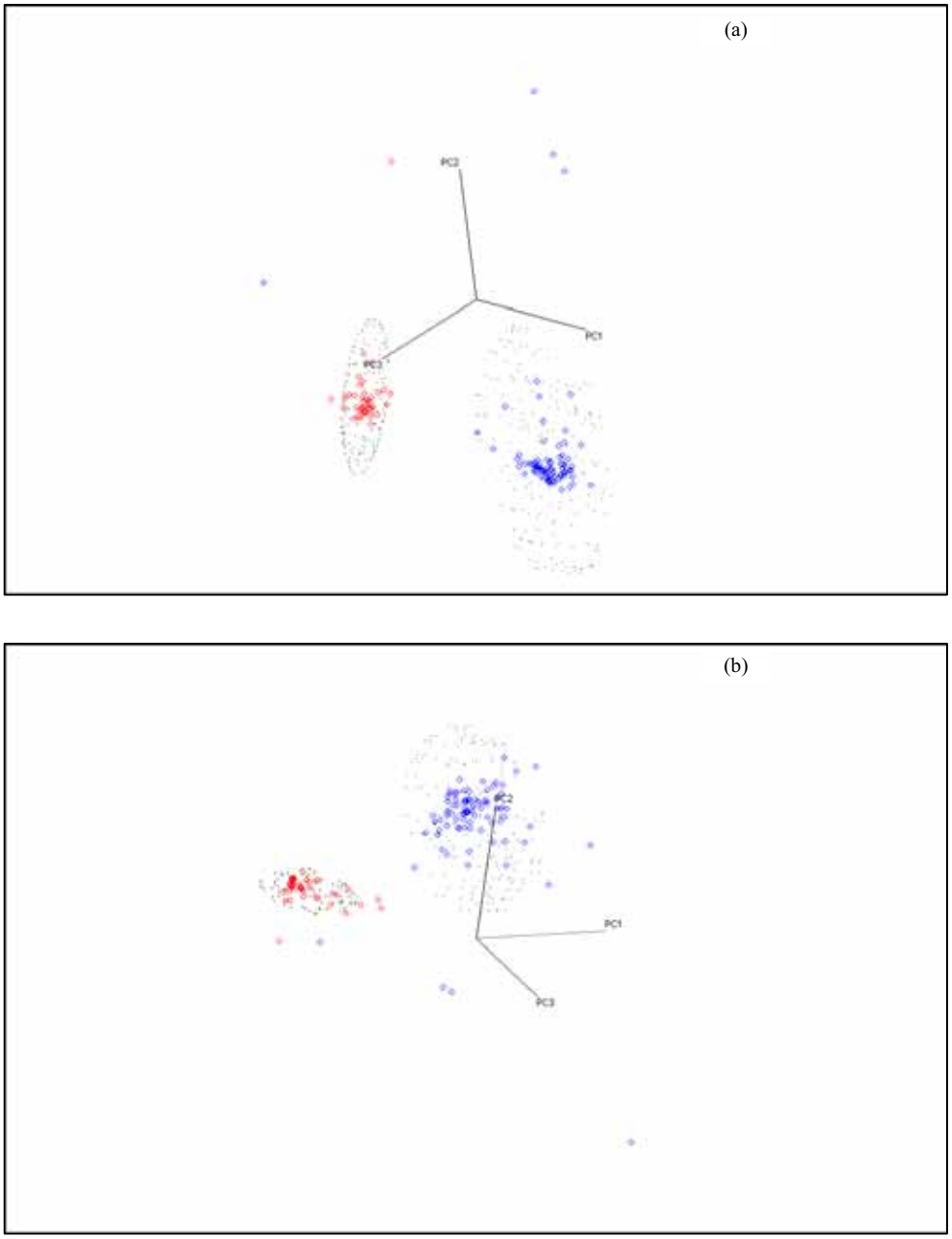
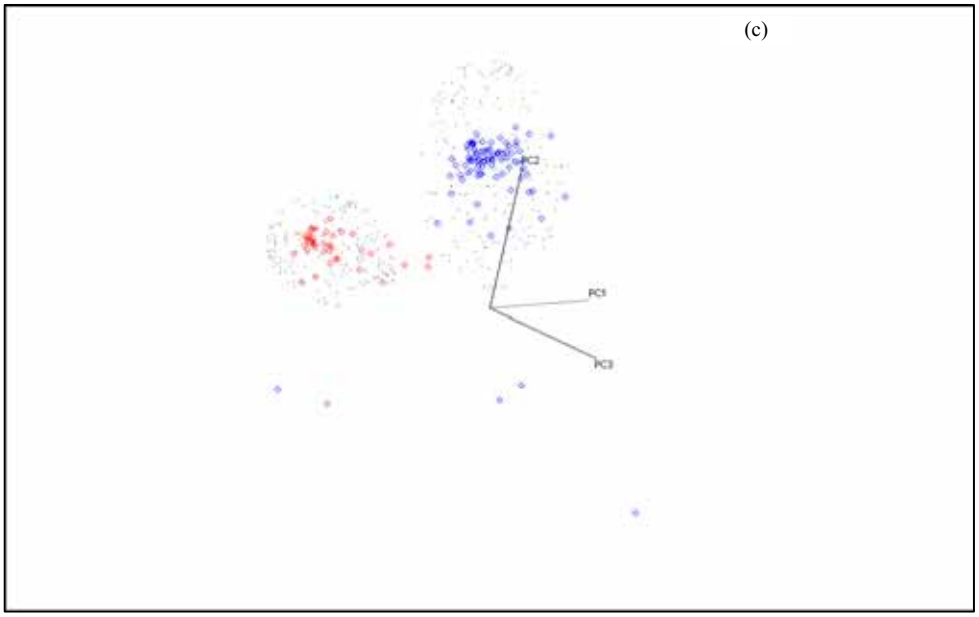


**Figura 67** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco I-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (continuação).



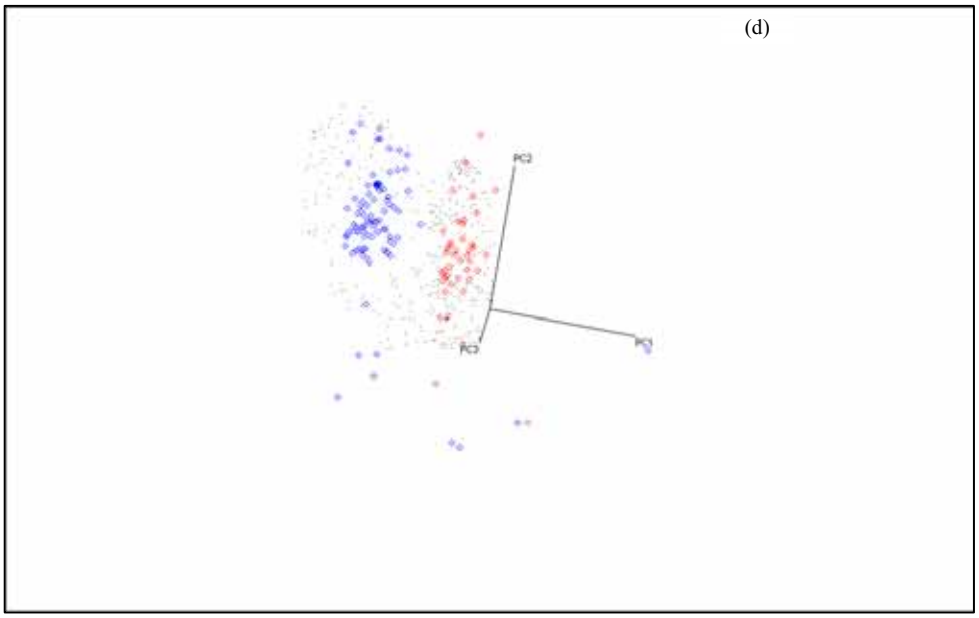
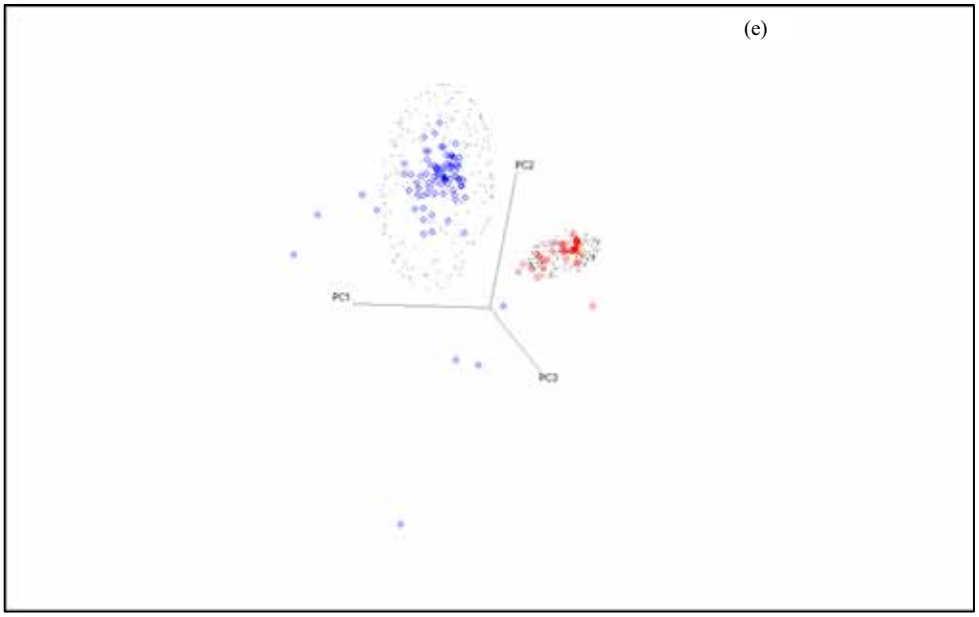
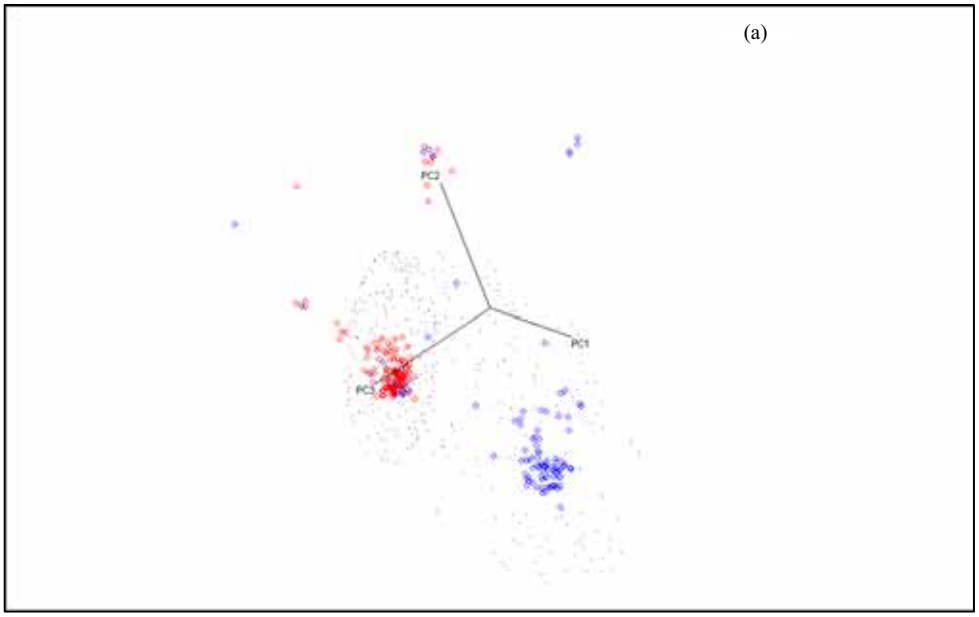



**Figura 67** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco I-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (conclusão).




**Fonte**: Elaborado pelo autor (2013).

**Figura 68** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco II-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes.



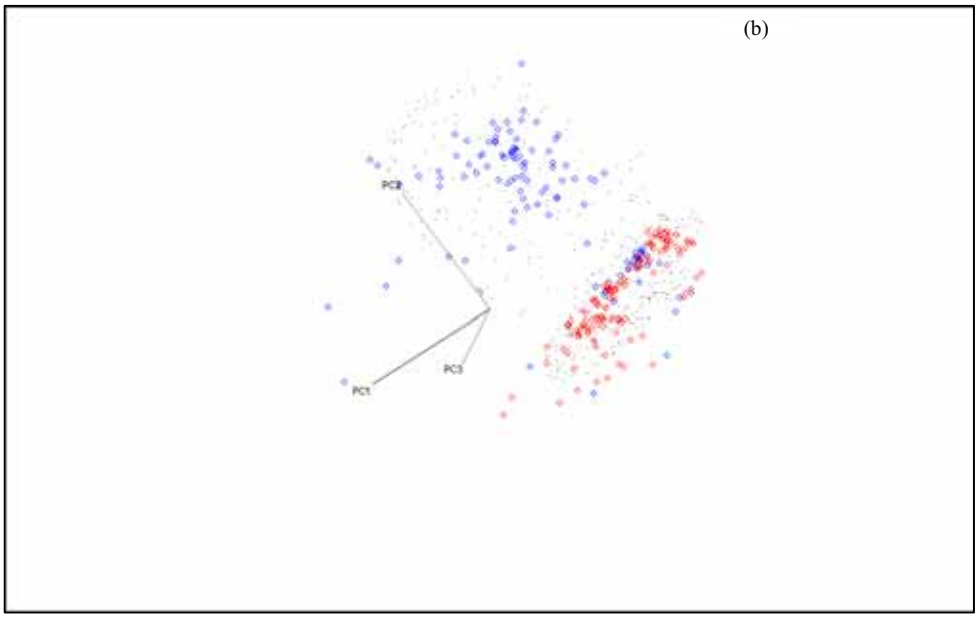
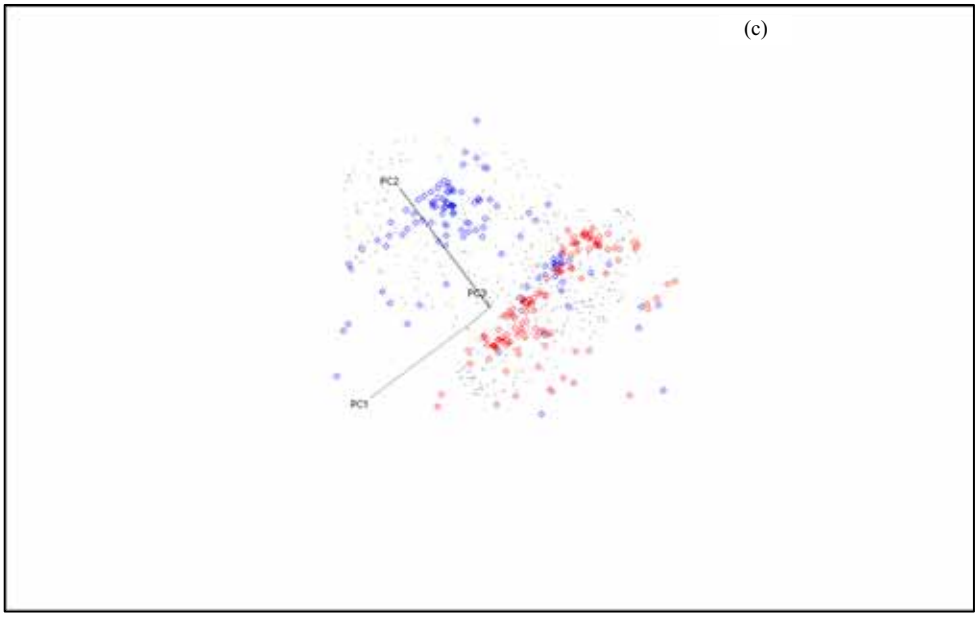



**Figura 68** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco II-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (continuação).



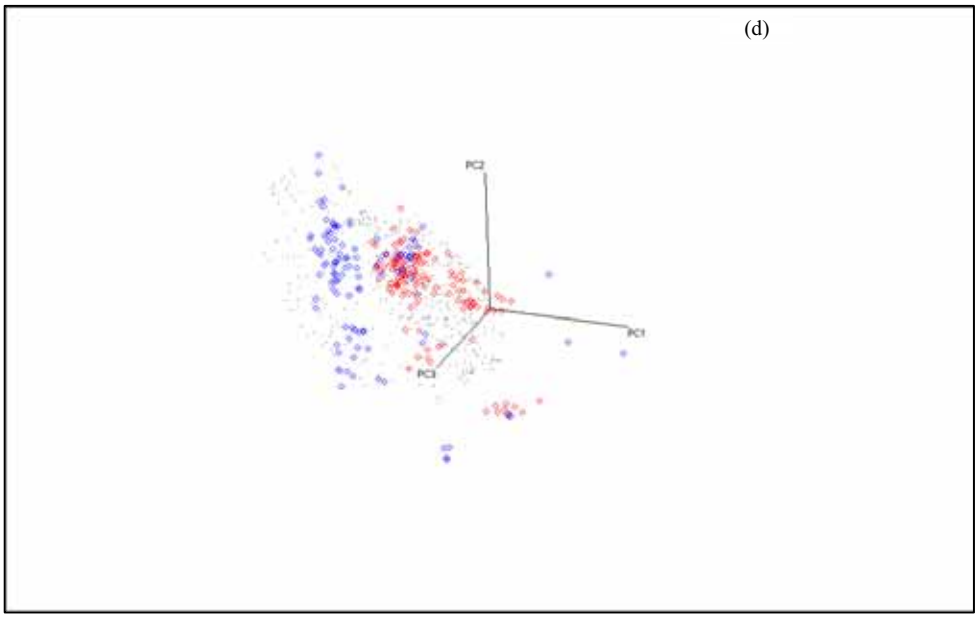
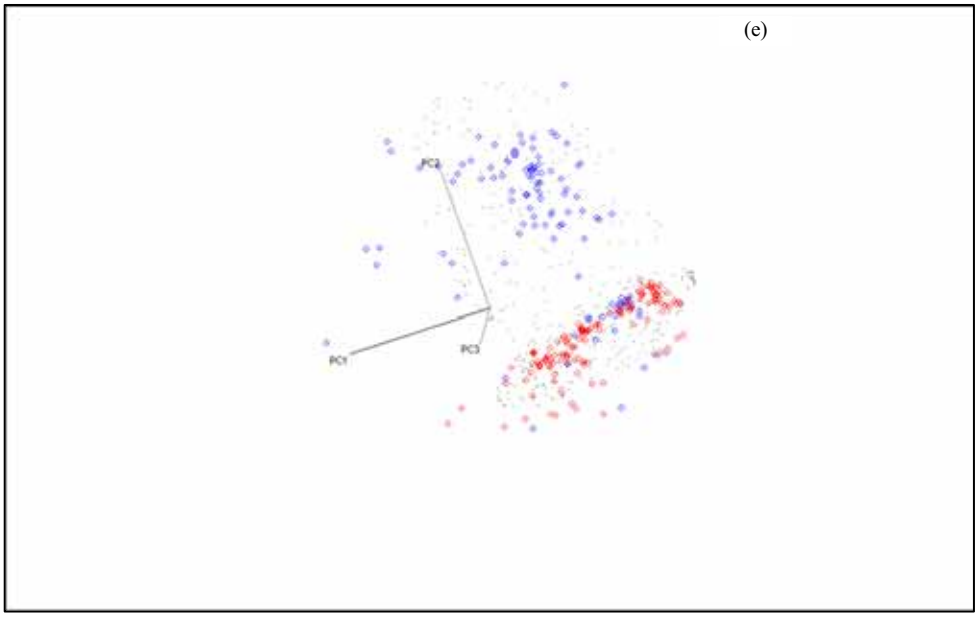



**Figura 68** - Representação tridimensional da classificação do grupo de treinamento, quanto à conformidade para o Banco II-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (conclusão).




**Fonte**: Elaborado pelo autor (2013).

Por fim, as tabelas 25 e 26 exibem o resultado da classificação pela SIMCA dos grupos de previsão dos Bancos I-A e II-A respectivamente. Segundo a tabela 25, para o grupo de treinamento, a predição de acerto foi de 81% para as amostras conformes (sem pré-processamento e transformada por logaritmo na base 10, escalada pela variância e transformada por logaritmo na base 10) e 54% para as amostras não conformes (sem pré-processamento e transformada por logaritmo na base 10).

Para o Banco I-A, o grupo de previsão gerou resposta, apesar das não classificações altas. Isto pode ter sido gerado por meio de amostras *outliers*, ou seja, de valores atípicos às do mesmo grupo. Desta forma, por haverem poucas amostras no grupo de previsão, o erro gerado pode se tornar elevado.

No caso do Banco II-A, tabela 26, para o grupo de treinamento, a predição de acerto foi de 18% para as amostras conformes (escalada pela variância e transformada por SNV) e 97% para as amostras não conformes (sem pré-processamento e transformada por logaritmo na base 10, escalada pela variância e transformada por logaritmo na base 10).

Os resultados apresentados na tabela 26 foram semelhantes aos do grupo de previsão do Banco II. Isto se deve a semelhança das amostras utilizadas no mesmo cálculo. A maioria das amostras não conformes pertencentes ao grupo de previsão do Banco II eram adulteradas de mesma matriz (ver tabela 4). Deste modo, o erro gerado deve provir destas amostras.

Apesar disso, é possível distinguir amostras conformes de adulteradas nos grupos de previsão, de acordo com as figuras 69, Banco I-A, e 70, Banco II-A. Embora, neste último, seja impossível distinguir no caso 'a' (sem pré-processamento e com transformada por logaritmo na base 10).

**Tabela 25** - Resultado da classificação pela SIMCA para grupo de previsão (22 amostras conformes e 11 amostras adulteradas) para o Banco I-A. As melhores classificações estão destacadas.

| Pré-processamentos | Transformadas | Classe[11] | Previsão C | Previsão A | Sem classe | Acerto/% |
|---|---|---|---|---|---|---|
| Nenhum | Log10 | C | 18 | 0 | 4 | 81 |
| | | A | 0 | 5 | 6 | 54 |
| Escala de Variância | Nenhuma | C | 13 | 0 | 9 | 59 |
| | | A | 0 | 2 | 9 | 18 |
| Escala de Variância | Log10 | C | 18 | 0 | 4 | 81 |
| | | A | 0 | 5 | 6 | 45 |
| Escala de Variância | SNV | C | 16 | 0 | 6 | 72 |
| | | A | 0 | 3 | 8 | 27 |
| Escala de Variância | Smoothing | C | 13 | 0 | 9 | 59 |
| | | A | 1 | 1 | 9 | 9 |

[11] C – amostras conformes, A – amostras adulteradas

**Fonte**: Elaborado pelo autor (2013).

**Tabela 26** - Resultado da classificação pela SIMCA para grupo de previsão (28 amostras conformes e 33 amostras adulteradas) para o Banco II-A. As melhores classificações estão destacadas.

| Pré-processamentos | Transformadas | Classe[12] | Previsão C | Previsão A | Sem classe | Acerto/% |
|---|---|---|---|---|---|---|
| Nenhum | Log10 | C | 3 | 25 | 0 | 11 |
| | | A | 1 | 32 | 0 | 97 |
| Escala de Variância | Nenhuma | C | 4 | 24 | 0 | 14 |
| | | A | 4 | 29 | 0 | 89 |
| Escala de Variância | Log10 | C | 2 | 26 | 0 | 7 |
| | | A | 1 | 32 | 0 | 97 |
| Escala de Variância | SNV | C | 5 | 23 | 0 | 18 |
| | | A | 2 | 29 | 2 | 89 |
| Escala de Variância | Smoothing | C | 4 | 24 | 0 | 14 |
| | | A | 4 | 29 | 0 | 89 |

[12] C – amostras conformes, A – amostras adulteradas

**Fonte**: Elaborado pelo autor (2013).

**Figura 69** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco I-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes.

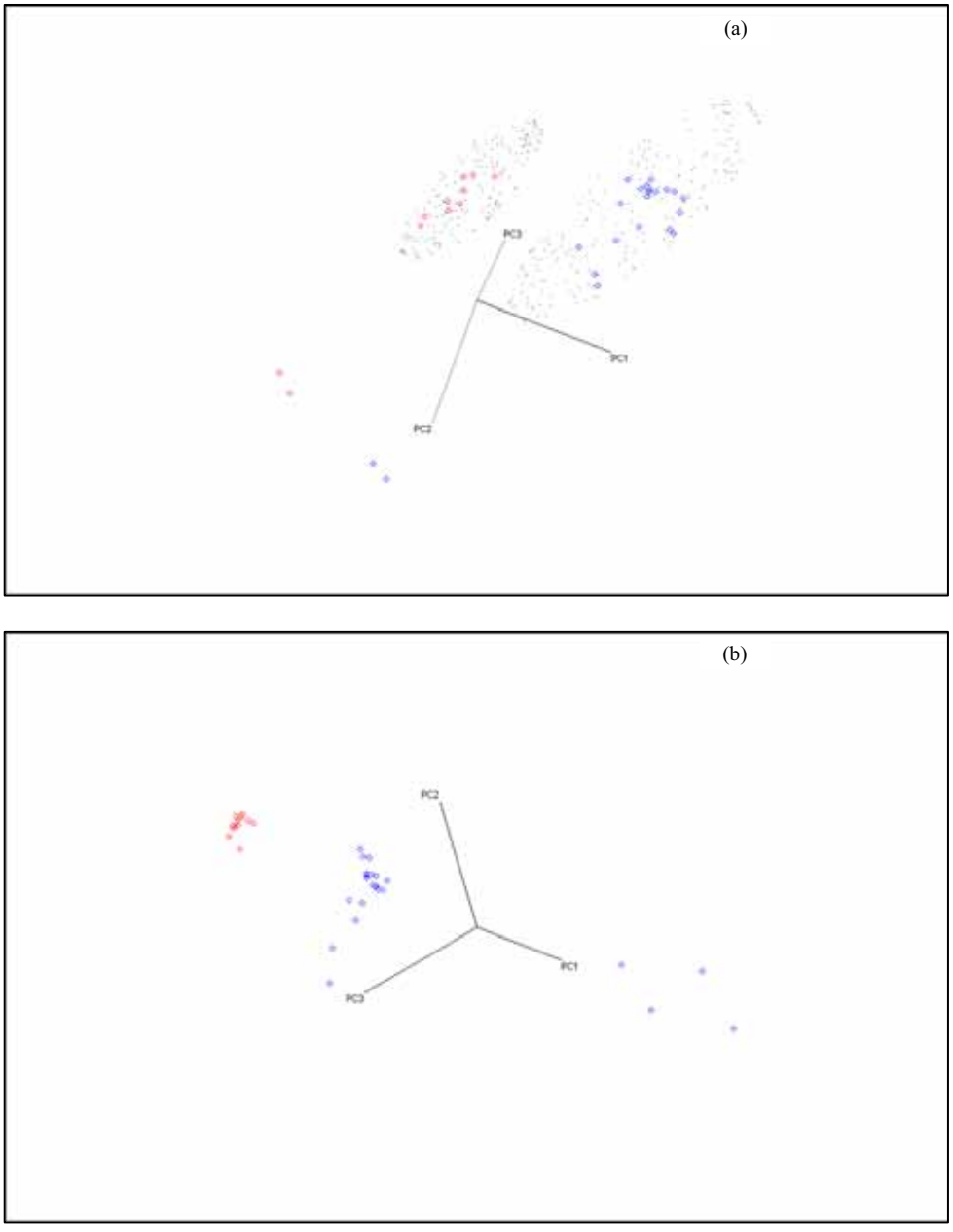
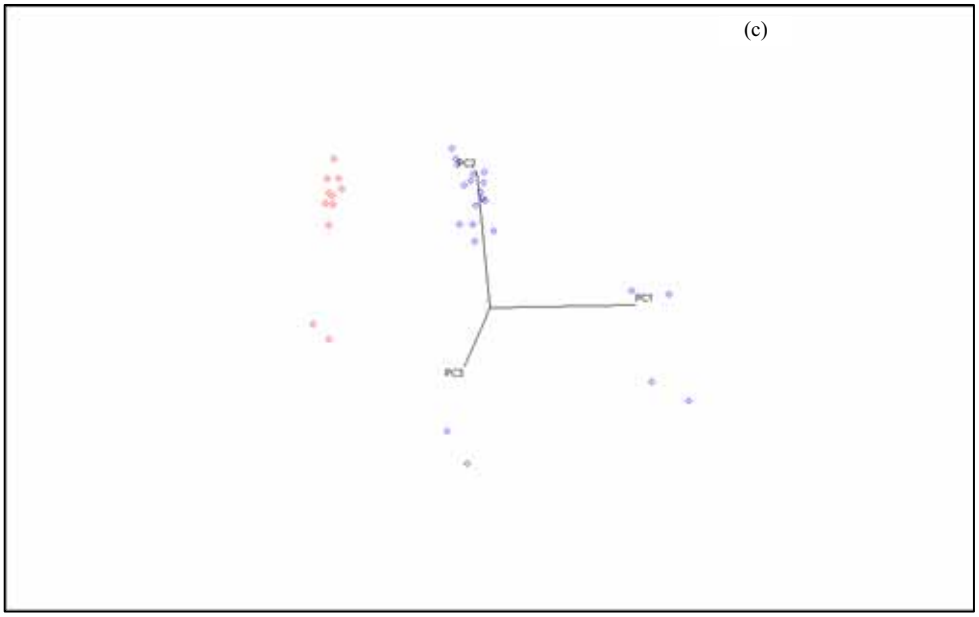


**Figura 69** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco I-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (continuação).



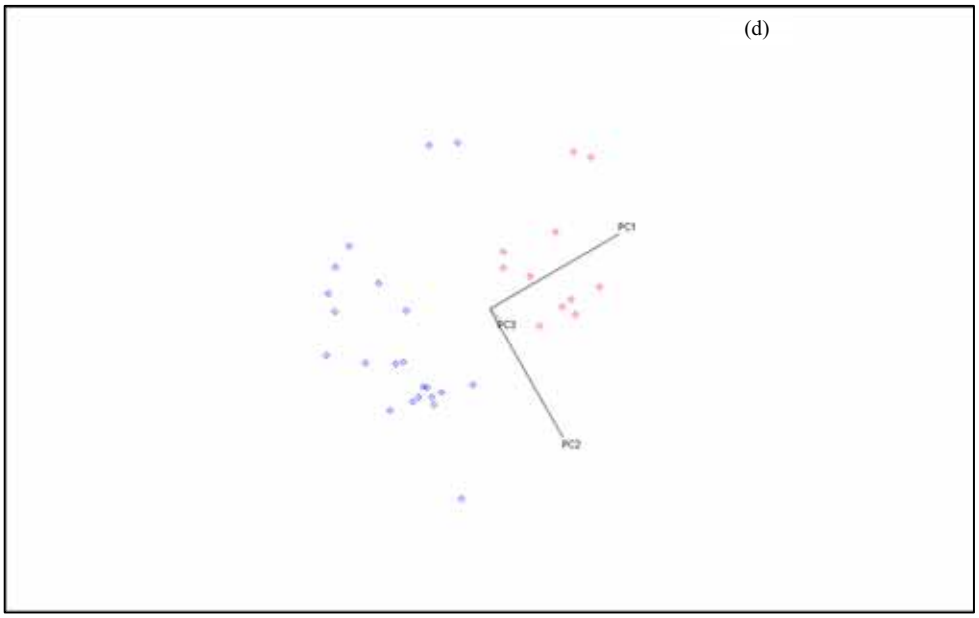
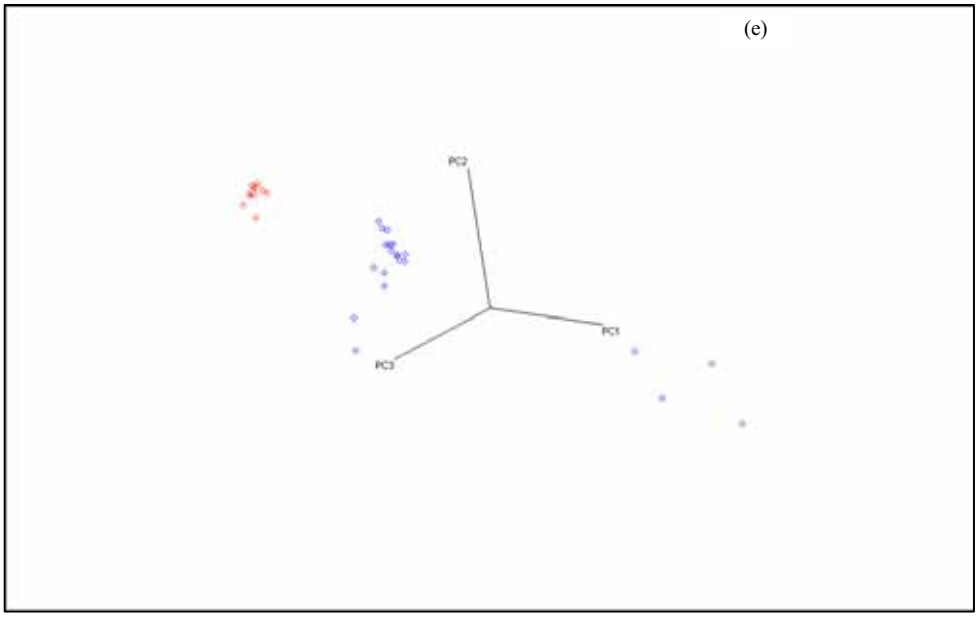
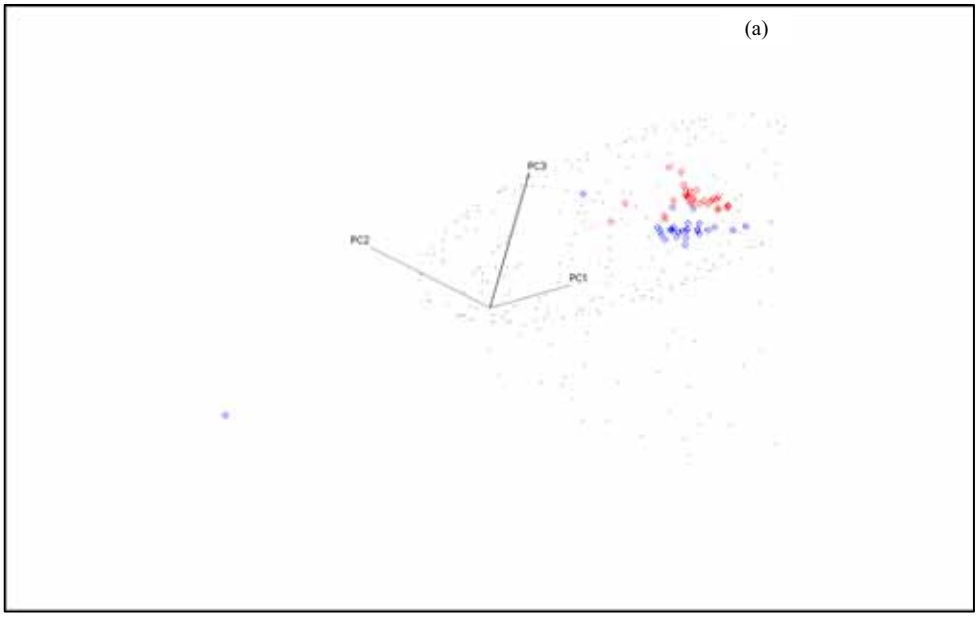



**Figura 69** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco I-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (conclusão).




**Fonte**: Elaborado pelo autor (2013).

**Figura 70** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco II-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes.



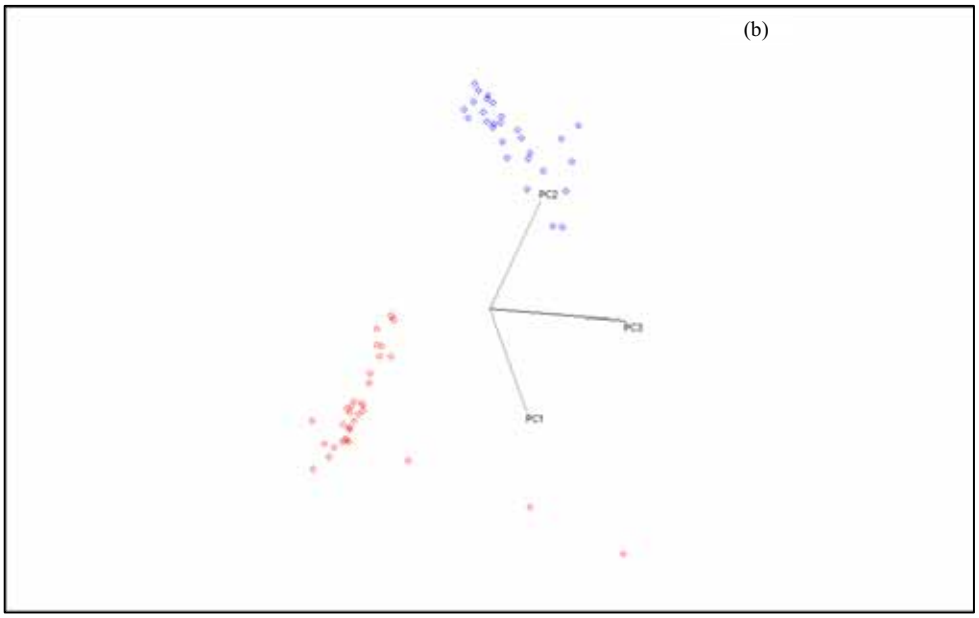
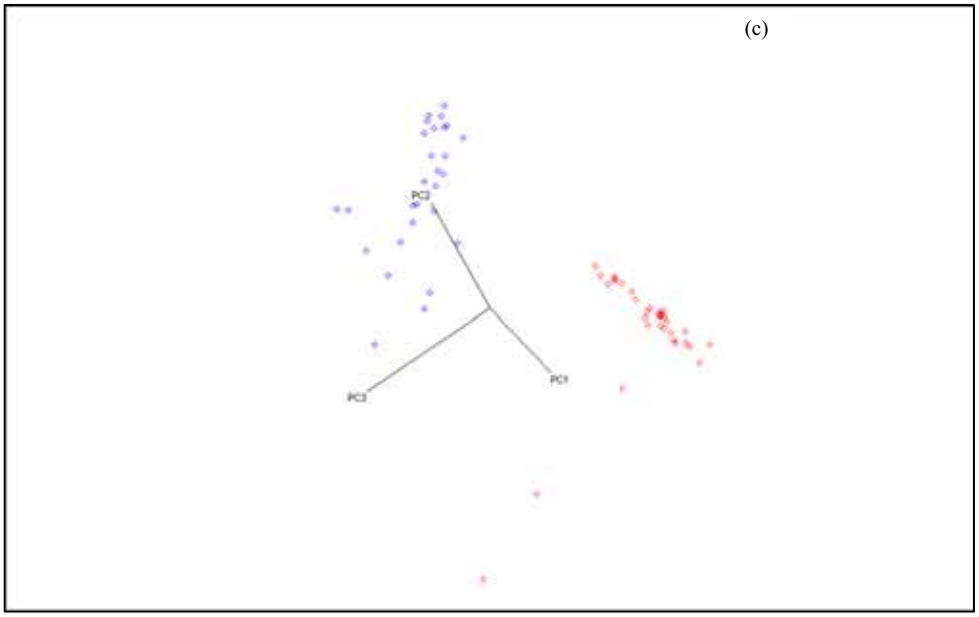



**Figura 70** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco II-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (continuação).



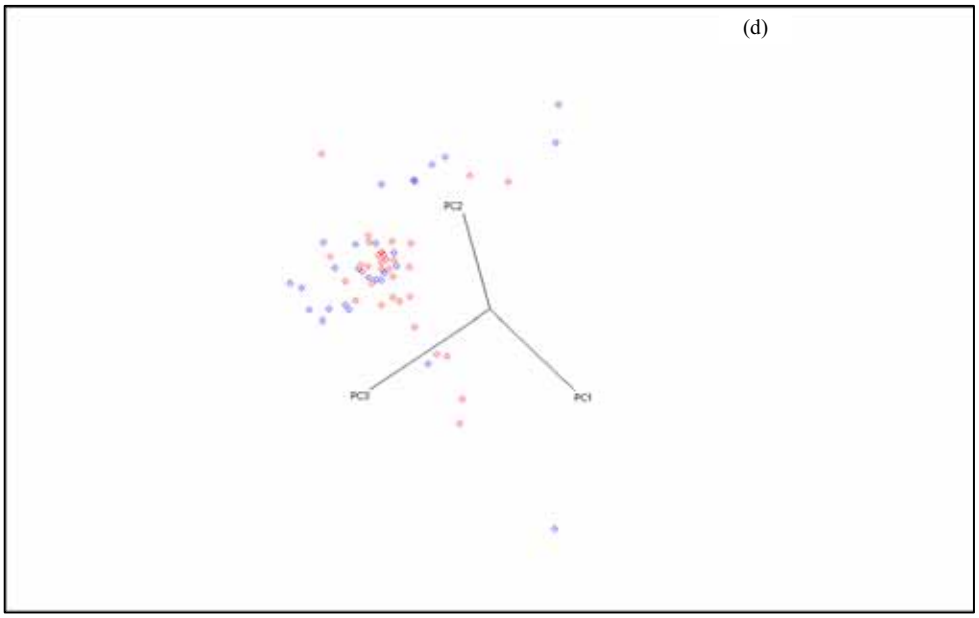
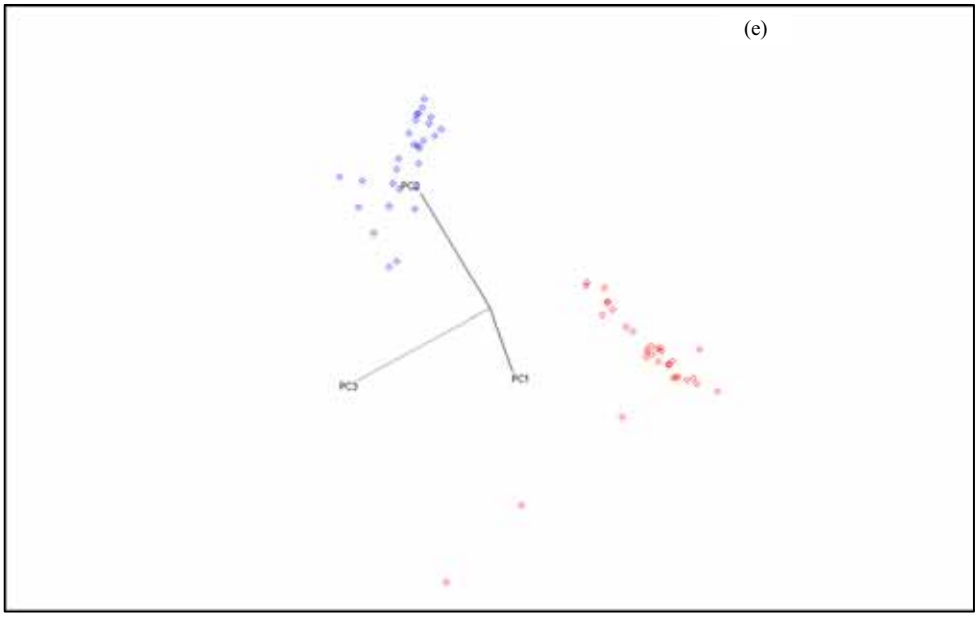



**Figura 70** - Representação tridimensional da classificação do grupo de previsão, quanto à conformidade para o Banco II-A, sendo (a) sem pré-processamento e Log10 (b) escalada pela variância e sem transformada (c) escalada pela variância e Log10 (d) escalada pela variância e SNV (e) escalada pela variância e *smoothing*. Os pontos azuis representam amostras conformes e pontos vermelhos indicam amostras não conformes (conclusão).




**Fonte**: Elaborado pelo autor (2013).

## *6 CONCLUSÃO*

A HCA mostrou-se superior na seleção mensal de amostras de gasolinas, quando comparado com a PCA. A HCA apresentou uma melhor distinção visual de cada elemento.

A CG-UR tornou possível análises cromatográficas com tempo de duração de 2,85 minutos, 171 segundos. Foi observada uma clara distinção dos picos cromatográficos dos constituintes da mistura, mostrando boa resolução do cromatograma final. Com isso, é possível realizar uma corrida semelhante à obtida em um cromatógrafo a gás convencional em tempo muito reduzido.

O algoritmo COW é uma ferramenta poderosa para o alinhamento dos picos cromatográficos. Assim, os possíveis erros que seriam gerados pela PCA foram minimizados. Através da PCA foi possível descrever os bancos de dados I e II com apenas duas componentes principais, sendo que a análise escalada pela variância e transformada por logaritmo na base 10 obteve a maior percentagem de descrição do banco de dados. A análise sem pré-processamento obteve uma resposta acima de 98%, reafirmando o uso do COW.

Através da SIMCA, não foi possível distinguir completamente as amostras conformes e não conformes dos Bancos I e II, de acordo com os resultados obtidos pelos ensaios e com a Resolução nº 57 de 2011 da ANP e com as Portarias nº 143 e 678 do MAPA, visto que estas amostras eram muito similares entre si e, que durante o alinhamento e os pré-tratamentos realizados, as aproximações matemáticas reduziram ainda mais suas diferenças. Deste modo foram observados erros nas classificações das amostras. Contudo, a classificação foi satisfatória para os Bancos I-A e II-A, com apenas amostras conformes e amostras adulteradas com solventes comerciais. Desta forma, foi possível a separação e classificação de amostras de gasolinas através da ferramenta SIMCA, quanto à conformidade com as especificações da ANP e do MAPA.

# REFERÊNCIAS

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVIES. Resolução ANP nº 3 de 19 de janeiro de 2011. Institui o Programa de Marcação Compulsória de Produtos em todo território nacional e regulamentados os termos e condições dispostos no § 4º do art. 5º da Lei n.º 10.336, de 2001, que determina a identificação mediante marcação dos hidrocarbonetos líquidos não destinados a formulação de gasolina ou óleo diesel. **Diário Oficial da União**, Brasília, DF, 20 jan. 2011. Disponível em: <http://nxt.anp.gov.br/nxt/gateway.dll/leg/resolucoes_anp/2011/janeiro/ranp%203%20-%202011.xml>. Acesso em: 18 abr. 2013a.

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVIES. Resolução ANP nº 57 de 20 de outubro de 2011. Regulamenta as especificações das gasolinas de uso automotivo e as obrigações quanto ao controle da qualidade a serem atendidas pelos diversos agentes econômicos que comercializam o produto em todo o território nacional. **Diário Oficial da União**, Brasília, DF, 21 out. 2011. Disponível em: <http://nxt.anp.gov.br/nxt/gateway.dll/leg/resolucoes_anp/2011/outubro/ranp%2057%20-%202011.xml >. Acesso em: 16 abr. 2013b.

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVIES. **Manual de procedimentos e orientações do Programa Nacional do Monitoramento da Qualidade de Combustíveis – PMQC**. 2. ed. rev. Brasília, DF, 2010.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **D4953**: standard test method for vapor pressure of gasoline and gasoline-oxygenate blends (dry method). West Conshohocken, 2006. 8 p.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **D1266**: standard test method for sulfur in petroleum products (lamp method). West Conshohocken, 2007. 12 p.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **D3710**: Standard test method for boiling range distribution of gasoline and gasoline fractions by gas chromatography. West Conshohocken, 2009. 12 p.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **D130**: standard test method for corrosiveness to copper from petroleum products by copper strip test. West Conshohocken, 2012a. 10 p.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **D381**: standard test method for gum content in fuels by jet evaporation. West Conshohocken, 2012b. 6 p.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **D2700**: standard test method for motor octane number for spark-ignition engine fuel. West Conshohocken, 2012c. 58 p.

AMERICAN SOCIETY FOR TESTING AND MATERIALS. **D3237**: standard test method for lead in gasoline by atomic absorption spectroscopy. West Conshohocken, 2012d. 4 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 13992**: gasolina automotiva – determinação do teor de álcool etílico anidro combustível (AEAC). Rio de Janeiro, 2003. 2 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 16041**: etanol combustível – determinação dos teores de metanol e etanol por cromatografia gasosa. Rio de Janeiro, 2012. 10 p.

BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. 25 anos de quimiometria no Brasil. **Química Nova**, v. 29, n. 6, p. 1401-1406, 2006.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Portaria MAPA n. 143, de 27 de junho de 2007. O ministro de Estado da agricultura, pecuária e abastecimento fixa em 25 por cento o percentual obrigatório de adição de álcool etanol anidro combustível à gasolina. **Diário Oficial da União**, Brasília, DF, n. 124, 29 jun. 2007. Seção 1, p. 9. Disponível em: < http://www.jusbrasil.com.br/ diarios/ 625313/dou-secao-1-29-06-2007-pg-9>. Acesso em: 21 fev. 2013a.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Portaria MAPA n. 678, de 31 de agosto de 2011. O ministro de Estado da agricultura, pecuária e abastecimento fixa em 20 por cento o percentual obrigatório de adição de álcool etanol anidro combustível à gasolina. **Diário Oficial da União**, Brasília, DF, n. 169, 1º set. 2011. Seção 1, p. 3. Disponível em: < http://www.jusbrasil.com.br/ diarios/ 30187428/dou-secao-1-01-09-2011-pg-3>. Acesso em: 21 fev. 2013b.

BYLUND, D. et al. Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modeling of liquid chromatography-mass spectrometry data. **Journal of Chromatography A**, v. 961, p. 237-244, May 2002.

CADOPPI, A.; FACCHETTI, R. **Simulated distillation of petroleum products (ASTM D2887) by GC in less than two minutes**. Milan, 2004. Thermo Fisher Scientific. Application note: 10074.

CAMARGO, N. L. **Avaliação da qualidade das gasolinas comerciais brasileiras através da aplicação do método quimiométrico SIMCA do reconhecimento de padrões nos perfis cromatográficos obtidos por cromatografia gasosa rápida (Fast GC)**. 2009. 43 f. Trabalho de Conclusão de Curso (Bacharelado em Química) - Instituto de Química, Universidade Estadual Paulista, Araraquara, 2009.

CONAWAY, C. F. **The petroleum industry**: a nontechnical guide. Tulsa: Penn Well, 1999. 289 p.

DONI, M. V. **Análise de clusters**: métodos hierárquicos e de particionamento. 2004. 93 f. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) - Universidade Presbiteriana Mackenzie, São Paulo, 2004.

FERREIRA, M. M. C. et al. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, n. 5, p. 724-731, jan. 1999.

FERREIRA, M. M. C. **Quimiometria**. Disponível em: <http://lqta.iqm.unicamp.br/portugues/downloads/Introducao.pdf>. Acesso em: 11 fev. 2013.

FLUMIGNAN, D. L. **Caracterização da qualidade e interpretação das propriedades físico-químicas de gasolinas C brasileiras através de cromatografia gasosa e métodos quimiométricos**. 2006. 140 f. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Estadual Paulista, Araraquara, 2006.

FLUMIGNAN, D. L. et al. Multivariate calibrations in gas chromatographic profiles for prediction of several physicochemical parameters of Brazilian commercial gasoline. **Chemometrics and Intelligents Laboratory Systems**, v. 92, p. 53-60, Dec. 2007.

FOO-TIM, C. et al. **Chemometrics from basics to wavelet transform**. Hoboken: Wiley-Interscience. 2004. 316 p.

FSPANERO. **Análise de Componentes Principais – PCA**. Disponível em: <fspanero.wordpress.com/2009/12/30/analise-de-componente-principais-pca/>. Acesso em: 23. jan. 2013.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regretion: a tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17. Jan. 1986.

GOLAY, M. J. E. Gas chromatographic terms and definitions. **Nature**, v. 182, n. 4643, p. 1146-1147, Oct. 1958.

HADDEN, N. The Chromatographic technique. In: HADDEN, N. et al. **Basic liquid chromatography**. Walnut Creek: Varian Aerograph, 1971.

HASWELL, S. J. **Practical guide to chemometrics**. New York: Marcel Dekker, 1992. 324 p.

HAWKES, S. J. Modernization of the van Deemter Equation for chromatographic zone dispersion. **Journal of Chemical Education**, v. 60, n. 5, p. 393-398, May 1983.

HOW STUFF WORKS. **Como funciona o refino de petróleo**. Disponível em: <ciência.hsw.uol.com.br/refino-de-petroleo5.htm>. Acesso em: 15 out. 2012.

INFOMETRIX INC. **Multivariate data analysis**. Bothell, 2008. 506 p. Disponível em: < http://pt.scribd.com/doc/95819644/Pirouette >. Acesso em: 21 fev. 2013.

JELLEMA, R. H. Variable shift and alignment. In: BROWN, S. D.; TAULER, R.; WALCZAK, B. **Comprehensive chemometrics**. Amsterdam: Elsevier, 2009. v. 2, cap. 2, p. 85-108.

LANÇAS, F. M. **Cromatografia em fase gasosa**. São Carlos: ACTA, 1993. 254 p.

McNAIR, H. M.; MILLER, J. M. **Basic gas chromatography**. 2nd ed. New Jersey: John Wiley & Sons, 2009. 239 p.

PARREIRA, T. F. **Utilização de métodos quimiométricos em dados de natureza multivariada**. 2003. 106 f. Dissertação (Mestrado em Química) - Instituto de Química, Universidade Estadual de Campinas, Campinas, 2003.

PETROBRAS. **Gasolina automotiva**. 4. ed. [Betim], 2000. 80 p. (Produtos Petrobras).

PINTO, M. R.; SILVA, E. C. D. O brilho da bandeira branca: concorrência no mercado de combustíveis no Brasil. **Planejamento e Políticas Públicas**, n. 31, p. 37-66, jun. 2008.

SANTOS, M. F. P. **Desenvolvimento e validação de métodos de espectroscopia no infravermelho próximo e médio para caracterização de lamas de ETAR para uso agrícola**. 2007. 87 f. Dissertação (Mestrado em Engenharia Biológica) – Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, 2007.

SEQUINEL, R. et al. Cromatografia gasosa ultrarrápida: uma visão geral sobre parâmetros, instrumentação e aplicações. **Química Nova**, v. 33, n. 10, p. 2226-2232, Out. 2010.

SHARAF, M. A.; ILLMAN, D. L.; KOWALSKI, B. R. Exploratory data analysis. In: _____. **Chemometrics**. New York: John Wiley & Sons, 1986. v. 82, cap. 6, p. 219-227.

SKOV, T. et al. Automated alignment of chromatographic data. **Journal of Chemometrics,** v. 20, p. 484-497, Mar. 2006.

VASCONCELOS, S. **Análise de Componentes Principais (PCA)**. Disponível em: <http://www.ic.uff.br/~aconci/PCA-ACP.pdf>. Acesso em: 15 fev. 2013.

WOLD, S. Principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 2, p. 37-52. 1987.

# BIBLIOGRAFIA CONSULTADA

ALMEIDA, F. M. N. **Espectroscopia de infravermelho próximo com Trasnformada de Fourier (FT-NIR) na caracterização de farinhas para alimentação pueril**. 2009. 84 f. Dissertação (Engenharia Biológica) – Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, 2009.

BICCHI, C. et al. Direct resistively heated column gas chromatography (Ultrafast module-GC) for high-speed analysis of essential oils of differing complexities. **Journal of Chromatography A**, v. 1024, p. 195-207, Oct. 2004.

BONFIM, R. R.; ALVES, M. I. R.; ANTONIOSI FILHO, N. R. Fast-HRGC method for quantitative determination of benzene in gasoline. **Fuel**, v. 99, p. 165-169, Aug. 2012

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

BROWN, S. D.; TAULER, R.; WALCAZK, B. **Comprehnsive chemometrics**: chemical and biochemical data analysis. Amsterdam: Elsevier, 2009. v. 2, 736 p.

DEGANI, A. L.; CASS, Q. B.; VIEIRA, P. C. Cromatografia um breve ensaio. **Química Nova na Escola**, v. 7, p. 21-25, maio 1998.

FLUMIGNAN, D. L. et al. Screening Brazilian automotive gasoline quality through quantification of saturated hydrocarbons and anhydrous ethanol by gas chromatography and exploratory data analysis. **Chromatographia**, v. 65, p. 617-623, Feb. 2007.

FLUMIGNAN, D. L. et al. Development, optimization and validation of gas chromatography fingerprint of Brazilian commercial gasoline for quality control. **Journal of Chromatography A**, v. 1202, p. 181-188. 2008.

FLUMIGNAN, D. L.; BORALLE, N.; OLIVEIRA, J. E. Screening Brazilian commercial gasoline quality by hydrogen nuclear magnetic resonance spectroscopic fingerprints and pattern-recognition multivariate chemometric analysis. **Talanta**, v. 82, p. 99-105. 2010.

MOREIRA, L. S.; D'AVILA, L. A.; AZEVEDO, D. A. Automotive gasoline quality analysis by gas chromatography: study of adulteration. **Chromatographia**, v. 58, n.7, p. 501-505, Oct. 2003.

PEREIRA, R. C. C. et al. Determination of gasoline adulteration by principal components analysis-linear discriminant analysis applied to FTIR spectra. **Energy & Fuels**, v. 20, n. 3, p. 1097-1102, Jan. 2006.

POVOLO, M.; PELIZZOLA, V.; CONTARINI, G. Directly resistively heated-column gas chromatography for the evaluation of cow milk fat purity. **European Journal of Lipid Science and Technology**, v. 110, p. 1050-1057, May 2008.

SILVA, F. L. N. S. et al. Determinação de benzeno, tolueno, etilbenzeno e xilenos em gasolina comercializada nos postos do estado do Piauí. **Química Nova**, v. 32, n. 1, p. 56-60, ago. 2009.

SKROBOT, V. L. et al. Use of principal component analysis (PCA) and linear discriminant analysis (LDA) in gas chromatographic (GC) data in the investigation of gasoline adulteration. **Energy & Fuels**, v. 21, p. 3394-3400, Aug. 2007.

WIEDEMANN, L. S. M.; D'AVILA L. A.; AZEVEDO, D. A. Brazilian gasoline quality: study of adulteration by statistical analysis and gas chromatography. **Sociedade Brasileira de Química**, v. 16, n. 2, p. 139-146, 2005.