

**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

**INSTITUTO DE CIÊNCIAS EXATAS**

**DEPARTAMENTO DE QUÍMICA**

**Kele Cristina Ferreira Dantas**

**Desenvolvimento de método quimiométrico baseado em  
análise de imagens digitais para a quantificação dos  
corantes amarelo crepúsculo e tartrazina em bebidas**

**Belo Horizonte**

**2015**

**UFMF/ICEx/DQ. 1116<sup>a</sup>**

**D 614<sup>a</sup>**

**KELE CRISTINA FERREIRA DANTAS**

**DESENVOLVIMENTO DE MÉTODO QUIMIOMÉTRICO  
BASEADO EM ANÁLISE DE IMAGENS DIGITAIS PARA A  
QUANTIFICAÇÃO DOS CORANTES AMARELO  
CREPÚSCULO E TARTRAZINA EM BEBIDAS**

Dissertação apresentada ao Departamento de Química do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial para obtenção do grau de Mestre em Química – Química Analítica.

D192d Dantas, Kele Cristina Ferreira  
2015 Desenvolvimento de método quimiométrico baseado em  
D análise de imagens digitais para a quantificação dos  
corantes amarelo crepúsculo e tartrazina em bebidas  
[manuscrito] / Kele Cristina Ferreira Dantas. 2015.  
[xii], 94 f. : il.

Orientador: Marcelo Martins de Sena.  
Coorientadora: Lúcia Maria L. de Alencar Auler.

Dissertação (mestrado) - Universidade Federal de  
Minas Gerais - Departamento de Química.  
Inclui bibliografia.

1. Química analítica - Teses 2. Análise  
multivariada - Teses 3. Calibração - Teses 4. Corantes  
sintéticos - Teses 5. Processamento de imagens -  
Técnicas digitais - Teses 6. Alimentos - Análise -  
Teses I. Sena, Marcelo Martins de, Orientador II.  
Auler, Lúcia Maria L. de Alencar, Coorientadora III.  
Título.

CDU 043

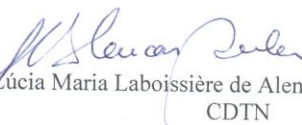
**"Desenvolvimento de Método Quimiométrico Baseado em Análise de Imagens Digitais para a Quantificação dos Corantes Amarelo Crepúsculo e Tartrazina em Bebidas"**

**Kele Cristina Ferreira Dantas**

Dissertação aprovada pela banca examinadora constituída pelos Professores:



Prof. Marcelo Martins de Sena - Orientador  
UFMG



Profa. Lúcia Maria Laboissière de Alencar Auler - Coorientadora  
CDTN



Profa. Isabel Cristina Pereira Fortes  
UFMG



Prof. Ricardo Mathias Orlando  
UFMG

Belo Horizonte, 15 de outubro de 2015.

Orientador: Prof. Dr. Marcelo Martins de Sena

Co-orientadora: Dr<sup>a</sup>. Lúcia Maria L. de Alencar Auler

Área de concentração: Química Analítica

Linha de pesquisa: Quimiometria; Desenvolvimento de método analítico.

Laboratório: 214

## **Agradecimentos**

Ao autor da minha vida; Deus, a Ele todo louvor e toda honra.

À minha família pelo carinho de sempre e muito necessário nestes momentos. Em especial à minha mãe e ao meu pai por se disporem tão inteiramente, na minha ausência, no cuidado da netinha.

Ao meu esposo, Thiago, pelo incentivo e companheirismo e por cuidar tão bem da nossa pequena, Mariana.

Aos meus amigos, parentes, colegas que ao lerem esses agradecimentos terão certeza das suas contribuições para a conclusão desta dissertação.

Ao meu orientador e Prof. Marcelo Sena por disponibilizar seus conhecimentos e parcerias sempre em favor de um desenvolvimento de trabalho ainda mais pleno.

À co-orientadora, mais mãe que já vi, Lúcia Auler. Obrigada pela dedicação a esse trabalho.

À Núbia e a Dani pela dedicação as análises cromatográficas.

Ao CDTN pela disponibilidade da estrutura e recursos para as análises cromatográficas.

Aos meus colegas do grupo de quimiometria pelas discussões e colaborações: Em especial ao Bruno Botelho e ao Leandro Pereira pela paciência no ensino das rotinas de Matlab.

À Universidade Federal de Minas Gerais pela estrutura física para desenvolvimento do trabalho

À pós-graduação pela oportunidade de desenvolvimento do trabalho no departamento de química.

Ao CNPQ e a Capes pelo apoio financeiro para que esse trabalho fosse possível.

## Resumo

Esta dissertação teve como principal objetivo o desenvolvimento e a validação de uma metodologia quimiométrica para quantificação simultânea dos corantes artificiais amarelo crepúsculo (AC) e tartrazina (TA) em bebidas não alcoólicas, como refrigerantes, isotônicos e sucos artificiais. Para isso, foi necessário previamente otimizar um método baseado em cromatografia líquida de alta eficiência (CLAE) e sua validação a fim de obter valores de referência para a calibração multivariada. Para o desenvolvimento deste método foi usada uma coluna C-18, fase móvel composta por acetonitrila/metanol 80:20 (v/v) e solução de acetato de amônio (pH = 6,7), e detecção espectrofotométrica. Foram analisadas 123 amostras de bebidas de diversos tipos, sabores e marcas. Modelos quantitativos foram construídos utilizando os valores de referência cromatográficos, análise de imagens digitais e as ferramentas quimiométricas PLS (*Partial Least Squares*, mínimos quadrados parciais) e OPS (*Ordered Predictors Selection*, seleção de preditores ordenados). As imagens foram obtidas através de um simples escâner de mesa, usando ultrassom como único pré-tratamento das amostras. Histogramas de frequência RGB das imagens digitais foram utilizados como sinal analítico. Os analitos foram determinados pelo método quimiométrico nas faixas de 2,3 a 41,1 e 0,1 a 15,1 mg.L<sup>-1</sup> para AC e TA, respectivamente. Os melhores modelos PLS apresentaram valores de raiz quadrada do erro médio quadrático de previsão (RMSEP, *root mean square error of prediction*) de 2,8 e 2,6 mg.L<sup>-1</sup> para AC e TA, respectivamente. Uma seleção de variáveis usando OPS permitiu reduzir o número de variáveis utilizadas na construção dos modelos de 768 para 100, produzindo previsões similares às do modelo anterior, com valores de RMSEP de 2,6 e 2,7 mg.L<sup>-1</sup> para AC e TA, respectivamente. Os modelos foram submetidos a um processo de validação analítica multivariada com a estimativa de figuras de mérito adequadas. Os métodos quimiométricos desenvolvidos apontam para a possibilidade de desenvolvimento de procedimentos analíticos mais rápidos, de menor custo, com mínimo preparo de amostra, sem uso de solventes e geração de resíduos.

*Palavras Chaves:* Quimiometria; Calibração multivariada; Corantes artificiais; CLAE; Análise de imagens digitais; Análise de Alimentos.

## Abstract

Development of a chemometric method based on digital image analysis for quantifying sunset yellow and tartrazine in soft beverages

The main goal of this dissertation was the development and validation of a chemometric methodology for the simultaneous determination of two artificial food dyes, sunset yellow (SY) and tartrazine (TA), in non-alcoholic beverages, such as soft drinks, isotonic and artificial juices. For this goal, it was necessary a previous optimization of a high performance liquid chromatography (HPLC) method and its validation in order to obtain reference values for the multivariate calibration method. The HPLC method was developed using a C-18 column, a mobile phase composed of acetonitrile/methanol 80:20 (v/v) and ammonium acetate (pH = 6.7) solution, and spectrophotometric detection. A number of 123 different samples of several types, brands and flavors were analyzed. Quantitative models were built with the reference HPLC values, digital image analysis and chemometric tools, such as PLS (*Partial Least Squares*) and OPS (*Ordered Predictors Selection*). Digital Images were generated with a simple scanner, using ultrasonic bath as the only sample pretreatment. RGB histograms obtained from digital images were used as analytical signals. The analytes were determined in the range from 2.3 to 41.1 and from 0.1 to 15.1 mg.L<sup>-1</sup> for SY and TA, respectively. The best PLS models provided RMSEP (root mean square error of prediction) of 2.8 and 2.6 mg.L<sup>-1</sup> for SY and TA, respectively. A variable selection with OPS allowed reducing the number of variables used in models construction from 768 to 100, providing predictions similar to the previous models, with RMSEP of 2.6 and 2.7 mg/L for SY and TA, respectively. All the models were validated through the estimate of appropriate figures of merit. The developed chemometric methods were rapid, of low cost, requiring a minimum sample pretreatment, and clean, not consuming reagents nor generating residues.

**Keywords:** Chemometrics; Multivariate calibration; artificial food dyes; HPLC; Digital image analysis; Food analysis.

## Lista de figuras

Figura 1. Estrutura química do Amarelo Crepúsculo .....	7
Figura 2. Estrutura química da Tartrazina .....	8
Figura 3. Determinação do fator de assimetria e alargamento do pico. <sup>44</sup> .....	14
Figura 4. . Representação gráfica do sistema RGB.....	16
Figura 5. Formação de tons no sistema RGB.....	17
Figura 6. Diagrama representando uma calibração multivariada <sup>65</sup> .....	20
Figura 7. Matriz das variáveis independentes.....	20
Figura 8. Decomposição em VL das matrizes X e Y para modelos PLS.....	22
Figura 9. Etapas da seleção de variáveis usando o OPS <sup>72</sup> .....	27
Figura 10. Representação vetorial do NAS (adaptado de Valderrama) <sup>77</sup> .....	29
Figura 11. Espectro de absorção molecular UV-Vis do padrão amarelo crepúsculo 30 ppm.....	39
Figura 12. Espectro de absorção molecular na região UV-Vis do padrão tartrazina 30 ppm.....	39
Figura 13. Cromatograma (FR): FM: acetato de amônio. 20 mmol L-1 (A) / ACN:MeOH 80:20 (B); vazão: 1,2 mL.min-1. Volume de injeção: 20 µL. Detecção: UV (428 nm); Padrão: 1-TA-(5,0mg.L-1) t <sub>R</sub> =0,7min ; 2-AC-(5,0mg.L-1) t <sub>R</sub> =7,9 min; 3-AP (20 mg.L-1) t <sub>R</sub> =11,5min. ....	40
Figura 14. Cromatograma (FR): FM: acetato de amônio. 20 mmol L-1(A) / ACN:MeOH 80:20 (B); vazão: 1,2 mL.min-1. Volume de injeção: 20 µL . Detecção: UV (428 nm); Padrão: 1 – TA (5,0mg.L-1) - t <sub>R</sub> =2,5min; 2 – AC (5,0mg.L-1)- t <sub>R</sub> =7,8min ; 3 – AP (20 mg.L-1)- t <sub>R</sub> =11,5min. ....	41
Figura 15. Cromatograma (FR): FM: acetato de amônio. 20mmol L-1 (A)/ ACN:MeOH 80:20 (B); vazão: 1,2 mL.min-1. Volume de injeção: 20 µL. Detecção: UV (428 nm); Amostra 8: refrigerante de maracujá ; 1-TA-t <sub>R</sub> =2,4min; 2-AC-t <sub>R</sub> =7,7min; 3-AP (20 mg.L-1)-t <sub>R</sub> =11,3min. ....	42
Figura 16. Cromatograma em (FR): FM: acetato de amônio. 20mmol L-1 (A)/ ACN:MeOH 80:20 (B); vazão: 1,2 mL.min-1. Volume de injeção: 20 µL . Detecção: UV (428 nm); Padrão: 1-TA (5,0 mg.L-1) t <sub>R</sub> =9,0 min; 2-AC- (5,0mg.L-1) t <sub>R</sub> =11,6 min; 3-AP(20 mg.L-1) t <sub>R</sub> =15,1 min. ....	43
Figura 17. Cromatograma em (FR). FM: acetato de amônio 20 mmol L-1(A) / ACN:MeOH 80:20 (B); vazão:1,2 mL.min-1. Volume de injeção: 20 µL . Detecção UV (428 nm); Amostra 4: bebida de frutas cítricas . 1-TA-t <sub>R</sub> =7,4 min; 2-AC-t <sub>R</sub> =10,1 min; 3-AP-(20 mg.L-1)-t <sub>R</sub> =13,6min.....	43
Figura 18. Cromatograma em (FR): FM: acetato de amônio. 20mmol L-1(A) / ACN:MeOH 80:20 (B); vazão: 1,2 mL.min-1. Volume de injeção: 20µL . Detecção: UV (428 nm); Amostra 8: refrigerante de maracujá ; 1-TA-t <sub>R</sub> =5,7 min; 2-AC-t <sub>R</sub> =8,6 min; 3-AP-(20 mg.L-1) t <sub>R</sub> =11,9 min.....	44
Figura 19. Cromatograma em (FR).FM:acetato de amônio. 20mmol L-1 (A) / ACN:MeOH 80:20(B); vazão:1,2 mL.min-1. Volume de injeção: 20 µL . Detecção: UV (428 nm); Amostra 4: bebida de frutas cítricas.1-TA-t <sub>R</sub> =3,4 min; 2-AC-t <sub>R</sub> =8,0 min; 3-AP(20 mg.L-1)-t <sub>R</sub> =11,6 min. ....	44
Figura 20. Cromatograma em (FR): FM: acetato de amônio. 20mmol L-1 (A) / ACN:MeOH 80:20 (B); vazão: 1,2 mL.min-1. Volume de injeção: 20 µL. Detecção: UV (428 nm); Padrão: 1-TA-(5,0mg.L-1)- t <sub>R</sub> =6,7 min; 2-AC-(5,0mg.L-1)-t <sub>R</sub> =8,3 min; 3-AP-(20 mg.L-1)-t <sub>R</sub> =11,8 min .....	45
Figura 21. Cromatograma em (FR). FM: acetato de amônio 20mmol L-1 (A) / ACN:MeOH 80:20 (B); vazão:1,2 mL.min-1. Volume de injeção: 20µL . Detecção UV (428 nm); Amostra 4: bebida de frutas cítricas . 1-TA-t <sub>R</sub> =6,6 min; 2-AC-t <sub>R</sub> =8,3 min; 3-AP-(20 mg.L-1) t <sub>R</sub> =11,7 min .....	46
Figura 22. Curva analítica referente ao AC, equação da reta e R <sup>2</sup> .....	50
Figura 23. Curva analítica referente ao TA, equação da reta e R <sup>2</sup> .....	51

Figura 24. Escâner CanoScan LiDE 110 com o anteparo .....	57
Figura 25. Imagem adquirida de uma amostra de isotônico sabor laranja marca Power ADE.....	58
Figura 26. Histograma de frequência das amostras na faixa de concentração de referência para AC	60
Figura 27. Histograma de frequência das amostras na faixa de concentração de referência para TA.	60
Figura 28. Histogramas(RGB) de frequência para amostras do modelo AC .....	61
Figura 29. Histogramas(RGB) de frequência para amostras do modelo TA.....	61
Figura 30. Coeficientes de regressão para o modelo AC .....	63
Figura 31. Coeficiente de regressão para o modelo TA .....	63
Figura 32. Vip Scores para o modelo AC .....	64
Figura 33. Vip Scores para o modelo TA .....	64
Figura 34. Canais selecionados pelo OPS no modelo AC .....	65
Figura 35. . Canais selecionados pelo OPS no modelo TA.....	66
Figura 36. Gráfico de valores de referência vs. valores preditos para o modelo AC. Calibração (círculos) e validação (triângulos) .....	68
Figura 37. Gráfico de valores de referência vs. valores preditos para o modelo TA. Calibração (círculos) e validação (triângulos) .....	68
Figura 38. Gráfico de resíduos para as amostras de calibração versus valores preditos pelo modelo AC .....	69
Figura 39. Gráfico de resíduos para as amostras de calibração versus valores preditos pelo modelo TA .....	69
Figura 40. Avaliação da normalidade dos resíduos para modelo AC.....	70
Figura 41. Avaliação da normalidade dos resíduos para modelo TA .....	71

## Lista de Tabelas

Tabela 1. Condições do método cromatográfico otimizado .....	47
Tabela 2. Estudo de linearidade para o método cromatográfico .....	52
Tabela 3. Avaliação da repetibilidade para a determinação de TA nas amostras usando o método cromatográfico .....	52
Tabela 4. Avaliação da repetibilidade para a determinação de AC nas amostras usando o método cromatográfico .....	53
Tabela 5. Adições de padrões à amostra de suco para estudo de veracidade do método cromatográfico para TA.....	54
Tabela 6. Adições de padrões à amostra de suco para estudo de veracidade do método cromatográfico para AC .....	54
Tabela 7. Resultados da otimização dos modelos PLS e dos modelos PLS-OPS .....	66
Tabela 8. Parâmetros para avaliação das FOM no modelo proposto para a determinação de AC em bebidas não alcoólicas .....	67
Tabela 9. Parâmetros para avaliação das FOM no modelo proposto para a determinação de TA em bebidas não alcoólicas .....	67
Tabela 10. Valores de R críticos e Req para o modelo AC.....	70
Tabela 11. Valores de R críticos e Req para o modelo TA.....	71
Tabela 12. Valores previstos e de referência para as amostras de validação do modelo AC.....	73
Tabela 13. Valores previstos e de referência para as amostras de validação do modelo TA.....	74
Tabela 14. Estudo de repetibilidade e precisão intermediária .....	75

## Lista de abreviaturas e siglas

AC	Amarelo Crepúsculo
ANVISA	Agência Nacional de Vigilância Sanitária
CLAE	Cromatografia Líquida de Alta Eficiência
AP	Azul Patente V
CLS	<i>Classical Least Squares</i>
CV	Coefficiente de variação
DP	Desvio Padrão
DPR	Desvio Padrão Relativo
FAO	<i>Food and Agriculture Organization</i>
FE	Fase Estacionária
FM	Fase Móvel
FOM	Figuras de Mérito ( <i>Figures of Merit</i> )
FR	Fase Reversa
ICS	<i>International Chemometrics Society</i>
IDA	Ingestão Diária Aceitável
INMETRO	Instituto Nacional de Metrologia, Normalização e Qualidade Industrial
MLR	Regressão Linear Múltipla ( <i>Multiple Linear Regression</i> )
NAS	Sinal Analítico Líquido ( <i>Net Analyte Signal</i> )
OMS	Organização Mundial da Saúde
PCA	Análise de Componentes Principais ( <i>Principal Component Analysis</i> )
PCR	Regressão por Componentes Principais ( <i>Principal Component Regression</i> )
PLS	Mínimos Quadrados Parciais ( <i>Partial Least Squares</i> )
RMSEC	Raiz Quadrada dos Erros Quadráticos Médios de Calibração ( <i>Root Mean Square Error of Calibration</i> )
RMSECV	Raiz Quadrada dos Erros Quadráticos Médios de Calibração Cruzada ( <i>Root Mean Square Error of Cross-Validation</i> )
RMSEP	Raiz Quadrada dos Erros Quadráticos Médios de Validação ( <i>Root Mean Square Error of Prediction</i> )

RPD	Relação de Desempenho do Desvio ( <i>Residual Prediction Deviation</i> ou <i>Relative Predictive Determinant</i> )
SEL	Seletividade
SEN	Sensibilidade
TA	Tartrazina
UV-Vis	Ultravioleta-Visível
VIP Scores	Escores de importância na projeção das variáveis ( <i>Variable Importance Projection Scores</i> )
VL	Variável latente

## Sumário

1. Introdução .....	1
2. Referencial teórico .....	4
2.1 O consumo de bebidas .....	4
2.2 Corantes artificiais em bebidas .....	5
2.2.1 Amarelo Crepúsculo e Tartrazina .....	7
2.3 Métodos para a determinação de AC e TA em bebidas .....	9
2.4 Cromatografia líquida de alta eficiência .....	10
2.4.1. Parâmetros Cromatográficos .....	13
2.5 Análise de imagens digitais .....	15
2.6 Quimiometria .....	18
2.7 Calibração multivariada.....	19
2.8 Ferramentas quimiométricas .....	21
2.8.1 <i>Partial Least Squares</i> (PLS) .....	21
2.8.2 <i>Ordered Predictors Selection</i> (OPS) .....	25
2.9 Validação multivariada.....	27
2.9.1. Sinal Analítico Líquido – NAS.....	28
2.9.2 Figuras de mérito .....	29
3. Otimização e validação do método Cromatográfico.....	35
3.1 Parte Experimental.....	35
3.1.1 Equipamentos .....	35
3.1.2 Reagentes e Solventes .....	35
3.1.3 Coluna cromatográfica .....	36
3.1.4 Preparo da coluna cromatográfica.....	36
3.1.5 Preparo das Fases Móveis (FM) .....	36
3.1.6 Padrões.....	37
3.1.7 Amostras .....	37
3.1.8 Otimização do método.....	38
3.1.9 Validação do método .....	47
3.2 Resultados e Discussão .....	50
3.3 Conclusão Parcial .....	55
4. Análise de imagens.....	56
4.1 Parte experimental.....	56

4.1.1 Equipamentos e softwares.....	56
4.1.2 Amostras .....	56
4.1.3 Procedimento de análise.....	56
4.2 Resultados e Discussão .....	58
4.2.1 Construção dos modelos.....	58
4.2.2 Validação Analítica Multivariada.....	66
4.3 Conclusão Parcial .....	76
5. Conclusão geral .....	78
6. Referências.....	79

## 1. Introdução

O uso considerável de corantes artificiais em alimentos levou ao estudo de seus possíveis efeitos à saúde. Não há ainda um consenso em relação à segurança do uso de corantes artificiais alimentícios, mesmo se utilizados dentro dos limites estabelecidos pelas normas vigentes. Mas, pesquisas significativas apontam para o seu possível potencial maléfico. O consumo de bebidas não alcoólicas, como refrigerantes, sucos artificiais e isotônicos, é crescente, o que aumenta a exposição da população, principalmente crianças, aos possíveis efeitos negativos da ingestão de corantes artificiais. O uso desses corantes nesta matriz tem como objetivo melhorar o aspecto sensorial/visual dos produtos.

A química analítica neste contexto tem importante papel no controle de qualidade dessas bebidas e na quantificação dos corantes alimentícios. Diversos métodos para a análise de corantes artificiais são bem estabelecidos, dentre eles, os baseados em cromatografia líquida de alta eficiência (CLAE, ou HPLC – *High Performance Liquid Chromatography*). Essa técnica é atualmente uma das mais usadas na análise de corantes do tipo azo, o principal tipo de corante artificial alimentar.

Porém, é de suma importância a preocupação em desenvolver métodos e ferramentas analíticas mais simples, rápidas e limpas e não geradoras de resíduos. De um modo geral, a CLAE apresenta vantagens de eficiência na separação, robustez e maior reconhecimento pelos órgãos reguladores em comparação a métodos espectrofotométricos. Quando comparada a métodos espectrofotométricos-quimiométricos, a CLAE apresenta desvantagens como o alto custo do equipamento e manutenção, a falta de um detector universal, tempo de análise elevado, necessidade de experiência do operador e de preparo de amostras e fases móveis, consumindo solventes e gerando resíduos. A geração de resíduos de solventes orgânicos e a exposição do operador aos mesmos são comumente problemas da análise por CLAE.

A análise de imagens digitais associada à quimiometria não é recente e vem sendo estudada há vários anos. Um número considerável de novos artigos teóricos e também de aplicações são publicados atualmente. Diversas substâncias

em diferentes matrizes têm sido objetos de estudos e há várias possibilidades de estratégias para o tratamento dos dados. A aquisição de imagens digitais atualmente se torna cada vez mais rápida e simples. Instrumentos diversos podem ser usados, como telefones celulares, câmaras web (*webcams*) e escâneres de escritório, os quais têm se tornado cada vez mais simples e baratos. Para a extração de informações de imagens digitais, as ferramentas analíticas univariadas, muitas vezes, não são suficientes, demandando então o uso de métodos multivariados/quimiométricos.

Com isso, a quimiometria, que trata da análise de dados químicos de natureza multivariada, torna-se necessária. As ferramentas usadas são matemáticas, estatísticas e computacionais. A interpretação dos modelos desenvolvidos, entretanto, requer conhecimentos químicos sobre o objeto estudado. Embora a quimiometria quando combinada a técnicas espectroscópicas ou de aquisição de imagens digitais tenha grande potencial para o desenvolvimento de métodos mais simples, rápidos e baratos, a utilização de métodos quimiométricos como oficiais ainda é escassa. A grande maioria dos guias de validação não contempla as particularidades da análise multivariada nem reconhecem métodos quimiométricos.

Considerando a possibilidade de uso de equipamentos de baixo custo para a aquisição de imagens digitais, os quais chegam a custar frações do preço de um espectrofotômetro, o desenvolvimento de métodos que combinem análise de imagens e ferramentas quimiométricas, tanto qualitativos quanto quantitativos, se torna promissor. É certo que se caminha para uma capacidade maior de gerar dados a partir de um só instrumento e que com a quimiometria mais informações significativas poderão ser extraídas e conclusões mais completas poderão ser obtidas sobre a amostra.

Assim sendo, este trabalho teve como objetivo geral determinar a concentração dos corantes artificiais alimentares tartrazina (TA) e amarelo crepúsculo (AC) em bebidas não alcoólicas, refrigerantes, isotônicos e sucos artificiais. Os objetivos específicos foram: (i) otimizar metodologia analítica por CLAE para a determinação quantitativa dos corantes em bebidas não alcoólicas e validá-la como método de referência. (ii) determinar o perfil e os teores dos corantes em estudo nas bebidas não alcoólicas. (iii) desenvolver e validar um método analítico

multivariado simples e rápido utilizando imagens digitais obtidas em um escâner de mesa comercial.

Esta dissertação foi organizada da seguinte forma. Após esta Introdução, no capítulo 2 serão apresentados os fundamentos teóricos de todos os aspectos envolvidos neste trabalho. No nosso planejamento inicial, com a utilização da metodologia cromatográfica, visava-se apenas obter valores de referência para a construção dos modelos quimiométricos. Mas, considerando todo o esforço experimental e teórico envolvido no desenvolvimento desse método, decidiu-se organizar sua descrição em um capítulo autônomo, destacando-se desta forma a relevância de seu conteúdo. Assim, os dois próximos capítulos, 3 e 4, tratarão de forma independente do desenvolvimento das duas metodologias analíticas utilizadas nesta dissertação, cromatográfica e quimiométrica/análise de imagens, respectivamente. Cada um destes dois capítulos terá suas próprias seções experimentais, de resultados e discussão e conclusões. Finalmente, o último capítulo trará uma conclusão geral.

## 2. Referencial teórico

### 2.1 O consumo de bebidas

As indústrias de alimentos e bebidas são consideradas do mesmo ramo por terem como objetivo comum o fornecimento de produtos para a nutrição humana.

Segundo dados de 2006, estima-se que o consumo médio de alimentos líquidos de uma pessoa seja em torno de 730 litros por ano; 484 litros de água e 246 litros entre café, refrigerantes, cerveja, água envasada, chás, bebidas alcoólicas, sucos e outros.<sup>1</sup>

O setor de bebidas possui os seguintes segmentos: água envasada (potável, mineral, mineralizada), bebidas tradicionais (café, chá, chocolate), bebidas alcoólicas (cervejas, vinhos, destilados) e o segmento de bebidas não alcoólicas industrializadas (refrigerantes, sucos, isotônicos, energéticos).<sup>1</sup> Esse último segmento é objeto de estudo dessa dissertação.

De acordo com dados do Euromonitor<sup>2,3</sup>, o Brasil conquistou em anos recentes (2005-2011) a terceira posição na lista dos maiores consumidores mundiais de cervejas e refrigerantes. O consumo brasileiro é inferior apenas aos verificados nos Estados Unidos da América (EUA) e na China. Em relação aos refrigerantes, o Brasil consome cerca de 85 litros/habitante/ano, o que o faz ocupar a 12ª posição do *ranking* mundial, em cujas primeiras posições estão os EUA (170 litros/habitante/ano), o México (146 litros/habitante/ano) e o Chile (127 litros/habitante/ano). Devido a uma maior preocupação dos consumidores com a saúde, o consumo de água envasada, bebidas isotônicas e sucos vêm aumentando.<sup>1</sup> Porém, a presença de aditivos como corantes artificiais em sucos e bebidas isotônicas coloca em dúvida a segurança na escolha dessas bebidas como uma alternativa correta quanto à saúde.

Os sucos podem ser classificados em 5 categorias: naturais, em pó (desidratação da fruta), sucos concentrados, sucos de polpa (congelamento in

natura) e os sucos prontos para beber (fabricados a partir de extrato de suco, água e uma série de aditivos).<sup>1</sup> Essa última categoria faz parte do grupo de interesse deste trabalho.

A tendência mundial é que a cada dia haja maior demanda por produtos saudáveis e alimentos naturais. As bebidas industrializadas, portanto, não serão bem vistas, pois devido aos altos teores de açúcar e aditivos presentes podem estar associadas às doenças como a obesidade e diabetes, entre outras.<sup>4</sup>

## **2.2 Corantes artificiais em bebidas**

A importância dos corantes em alimentos está relacionada com os interesses das empresas em atender e ganhar o mercado consumidor, tornando os alimentos mais atrativos.

As civilizações antigas já possuíam o hábito de adicionar cor aos seus alimentos. As substâncias usadas eram de origem animal, vegetal ou mineral e eram especiarias e condimentos que conferiam cor. A partir do século XVIII, com a descoberta de corantes sintéticos que conferem mais cores aos alimentos e são mais estáveis, os corantes naturais ficaram em segundo plano.<sup>5</sup>

Os aditivos são considerados ainda inofensivos à saúde se obedecidos percentuais máximos estabelecidos pela ANVISA (Agência Nacional de Vigilância Sanitária). A Ingestão diária aceitável (IDA) já é estabelecida para todos os corantes artificiais permitidos pela legislação brasileira. Pelas resoluções nº 382 e 388 da ANVISA, no Brasil são permitidos 11 corantes artificiais (amaranto, vermelho de eritrosina, vermelho 40, ponceau 4R, amarelo crepúsculo, amarelo tartrazina, azul de indigotina, azul brilhante, azorrubina, verde rápido e azul patente V).<sup>6,7</sup>

A FAO (*Food And Agriculture Organization*) e a OMS (Organização Mundial da Saúde) recomendam que os países verifiquem o consumo total de aditivos permitidos, a partir de seus dados sobre a dieta populacional, assegurando que a ingestão total não ultrapasse os valores determinados na quantidade de Ingestão Diária Aceitável (IDA).<sup>5</sup>

Os corantes compreendem dois principais componentes: o grupo cromóforo que é responsável pela cor, com absorção característica na região do ultravioleta e/ou do visível e o grupo funcional que está diretamente ligado à fixação do corante na matriz de interesse, neste trabalho, o alimento. Na síntese de corantes, estão sendo utilizados vários grupos cromóforos. Em geral, em alimentos, os corantes mais representativos e amplamente utilizados pertencem à família dos corantes azóicos, que se caracterizam por apresentarem um ou mais grupamentos – N=N– ligados a sistemas aromáticos.<sup>8</sup>

Os corantes artificiais constituem uma classe de aditivos sem nenhum valor nutricional, que são adicionados aos alimentos para fornecer uma coloração que “seduz” as pessoas pela visão, ou seja, a sua utilidade está ligada unicamente ao aspecto sensorial/visual do produto. Eles apresentam inúmeras vantagens comparativas em relação aos corantes naturais, tais como maior estabilidade à exposição ao oxigênio, à luz e a mudanças de pH, alta solubilidade em água e menor custo de produção.<sup>9</sup> Existem algumas exceções, como o corante natural caramelo IV, largamente usado em refrigerantes do tipo “Cola”.

Entre os possíveis efeitos negativos dos corantes alimentícios artificiais, especialmente quando consumidos em excesso, têm sido relatadas respostas alérgicas, asma, urticária e imunossupressão.<sup>9,10</sup> Sob este aspecto, vários estudos têm sido realizados sobre o impacto na saúde humana do uso desses corantes nos alimentos. A inocuidade dos corantes artificiais bem como seus efeitos adversos à saúde ainda estão em debate e são contraditórios, gerando em diversos países uma diferença de aceitação e estabelecimento dos limites máximos permissíveis. Isso gera diferenças entre os corantes que são permitidos nos diversos países e também sobre seus limites máximos aceitáveis.

Os corantes alimentícios podem causar reações alérgicas em alguns consumidores e por isso o controle de suas concentrações em alimentos é de suma importância. Os corantes sintéticos mais utilizados na indústria alimentícia, farmacêutica e de cosméticos são a tartrazina e o amarelo crepúsculo. Em alimentos, estes estão presentes nos sucos em pó, refrigerantes e balas.<sup>11</sup> Daí, a importância em controlar e quantificar esses corantes artificiais.

Os alimentos possuem matrizes diferentes e complexas, sendo relevante separar os alimentos em grupos para fins de análise. A análise de corantes

artificiais em bebidas não alcoólicas (sucos, refrigerantes e isotônicos) é relevante, pois se trata de um grupo de alimentos consideravelmente presente na vida das pessoas, sendo importante controlar o seu consumo para que o teor de corantes não ultrapasse os limites segundo a IDA.<sup>6,7</sup>

### 2.2.1 Amarelo Crepúsculo e Tartrazina

O Amarelo Crepúsculo (AC), também conhecido como *Sunset Yellow*, Amarelo Entardecer, E110 ou Amarelo Comestível 3, é um dos corantes azo mais utilizado em alimentos (Figura 1). Ele possui uma coloração alaranjada e é utilizado principalmente em produtos com sabor de frutas, como refrigerantes, sucos, doces e sorvetes. O seu valor de pKa é igual a 10,4. O AC também possui uma ampla utilização nas indústrias farmacêuticas e de cosméticos. Entretanto, o seu consumo pode causar alguns efeitos colaterais em humanos, e a sua ingestão tem sido associada a riscos de falência renal e danos hepatocelulares.<sup>9</sup>

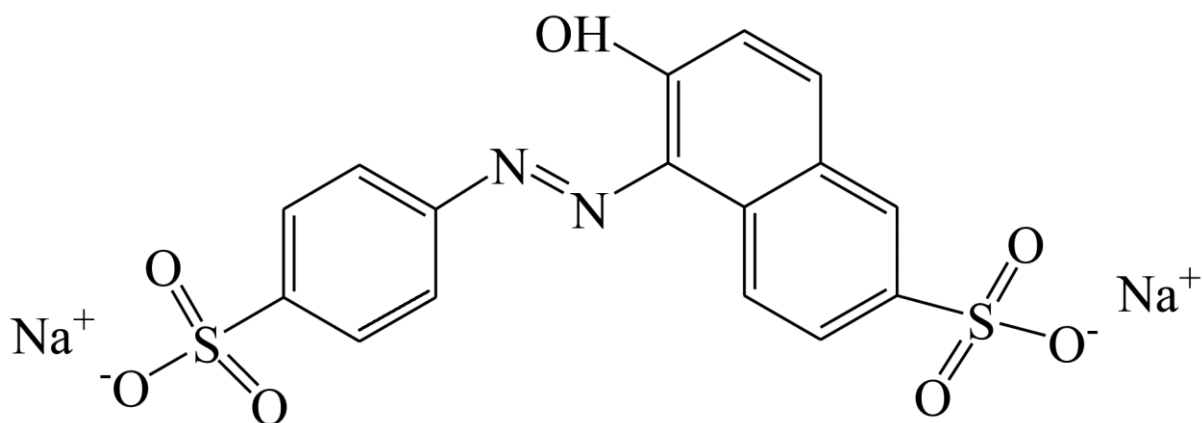


Figura 1. Estrutura química do Amarelo Crepúsculo

No Brasil, a ANVISA é o órgão governamental responsável pela regulação de alimentos e que estabelece os limites para utilização de corantes artificiais em diferentes produtos. De acordo com a resolução R05/07 (ANVISA), o limite para a concentração de amarelo crepúsculo em bebidas não alcoólicas é de 100 mg.L<sup>-1</sup>.<sup>12</sup>

A Tartrazina (TA) é sintetizada a partir da tinta do alcatrão de carvão, também denominada Amarelo Tartrazina ou E102, com nomenclatura oficial (IUPAC)

de sal trissódico 5-hidroxi-1-(4-sulfonatofenil)-4-(4-sulfonato-fenilazo)-H-pirazol-3-carboxilato (Figura 2). Ela faz parte da classe dos monoazos, possuindo fórmula  $C_{16}H_9N_4Na_3O_9S_2$ , massa molar  $534,36 \text{ g mol}^{-1}$ , solubilidade a  $25^\circ\text{C}$  em água de  $20\text{g}/100\text{mL}$ , em glicerina  $18 \text{ g}/100 \text{ mL}$ , em propileno  $7 \text{ g}/100\text{mL}$  e em etanol  $<0,1 \text{ g}/100 \text{ mL}$ . A TA apresenta um valor de pKa igual a 9,4. A IDA é de  $7,5 \text{ mg.kg}^{-1}$  de peso corpóreo.<sup>11,13</sup>

De acordo com a resolução R05/07, o limite para a concentração de TA em bebidas não alcoólicas é de  $100 \text{ mg.L}^{-1}$ .<sup>12</sup>

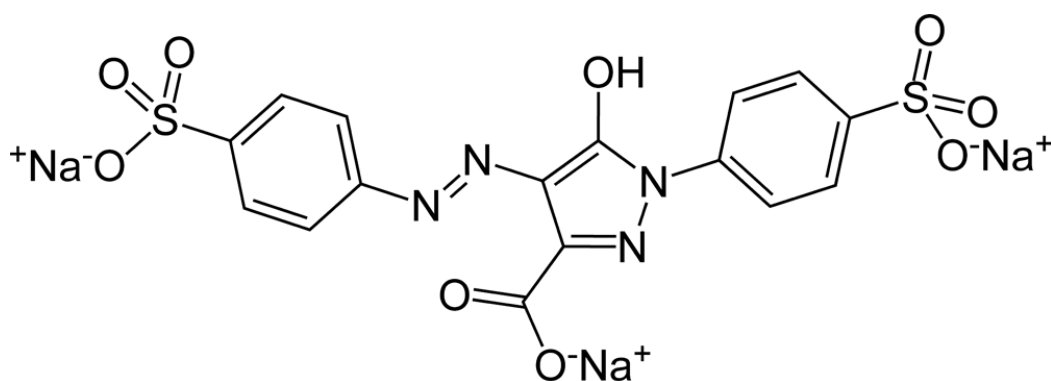


Figura 2. Estrutura química da Tartrazina

A tartrazina, por ser um composto nitroso (classe azo), é reduzida no organismo e tem sido alvo de estudos de mutagênese por produzir como todos os corantes azóicos uma amina aromática que provoca sensibilidade no organismo do indivíduo. O ácido sulfanílico foi o principal metabólito identificado.<sup>13,14</sup>

Dentre os corantes azo, a tartrazina tem despertado maior atenção dos toxicologistas e alergistas. Ela é apontada como responsável por várias reações adversas, causando desde urticária até asma. Estima-se que uma em cada 10 mil pessoas apresente reações a este corante.<sup>15</sup>

Os corantes amarelo crepúsculo e amarelo tartrazina são espécies com absorção espectral na região do visível. Em misturas eles apresentam uma intensa sobreposição espectral das bandas de absorção, que impede a aplicação da espectrofotometria direta para a quantificação simultânea desses corantes.

### **2.3 Métodos para a determinação de AC e TA em bebidas**

Os corantes azo são o maior grupo de corantes artificiais alimentícios, correspondendo a cerca de 60-70% dos utilizados. O método oficial para determinação de corantes azo em alimentos nas formas sólida ou semi-sólida baseia-se em espectroscopia de absorção no UV/Vis, o qual requer extrações líquido-líquido sequenciais com metanol contendo 5% de hidróxido de amônio.<sup>16</sup> Diversos outros métodos analíticos têm sido utilizados para a determinação destes corantes em alimentos e bebidas, individualmente ou em misturas. Esses métodos se baseiam em técnicas tais como: cromatografia em camada delgada,<sup>17</sup> cromatografia líquida de alta eficiência (CLAE),<sup>5,18,19</sup> voltametria adsortiva,<sup>20</sup> potenciometria,<sup>21</sup> eletroforese capilar,<sup>22</sup> espectrofluorimetria,<sup>23</sup> sensores baseados em polímeros impressos molecularmente,<sup>24</sup> imunoenaios<sup>9</sup> e extração no ponto nuvem com determinação espectrofotométrica.<sup>25</sup>

A determinação simultânea de misturas de vários corantes numa mesma matriz pode ser feita por métodos que envolvem etapas laboriosas, como microextração usando líquidos iônicos seguida por CLAE<sup>26</sup> ou sensores de óxido de grafeno funcionalizados com nanopartículas de ouro.<sup>27</sup> Para o desenvolvimento de metodologias mais simples e rápidas, estratégias quimiométricas também têm sido aplicadas na determinação simultânea de corantes em alimentos, principalmente usando espectrofotometria de absorção molecular na região do UV/Vis.<sup>28-31</sup> As cromatografias em papel e em camada delgada são técnicas rápidas, porém, com baixa exatidão e precisão. A eletroforese apresenta desvantagens semelhantes às da CLAE, gasto de solventes, geração de resíduos e longo tempo de corrida, e apesar das poucas aplicações na quantificação de corantes, essa técnica vem sendo apontada como promissora. A CLAE é atualmente uma das principais técnicas para análises de corantes em alimentos,<sup>32,33</sup> porém apresenta ainda as desvantagens de requerer etapas de extração e o alto custo do equipamento. Em todos esses métodos, a preparação de amostra é uma etapa muito importante e laboriosa para a determinação de corantes alimentares. Extraí-los de diferentes tipos de matrizes não é uma tarefa simples.<sup>5,34</sup>

A análise de imagens associada a métodos quimiométricos de calibração multivariada é uma alternativa para o desenvolvimento de metodologias mais rápidas, baratas, não destrutivas e ambientalmente corretas. Métodos deste tipo possuem vantagens similares às metodologias quimiométricas baseadas em espectros de absorção no UV/Vis,<sup>28-31</sup> entretanto possuem uma grande vantagem adicional, dado o custo extremamente baixo da instrumentação utilizada, pois um simples escâner de mesa pode custar em torno de R\$ 100. Em 2014, um método baseado em análise de imagens e calibração multivariada foi desenvolvido em nosso grupo de pesquisa para a determinação de AC em refrigerantes e isotônicos.<sup>35</sup> Nesta dissertação, um novo método foi desenvolvido para a determinação simultânea de dois corantes, numa situação analítica mais desafiadora que a anterior, uma vez que envolve produtos de composição mais complexa e variada, com maior sobreposição de sinais analíticos.

## **2.4 Cromatografia líquida de alta eficiência**<sup>34,36-42</sup>

Diversas são as técnicas cromatográficas, sendo que, dentre elas, a Cromatografia Líquida de Alta Eficiência, CLAE, também conhecida como de Cromatografia Líquida de Alta Velocidade, ou de Alta Pressão, vem se destacando cada vez mais e o seu emprego em vários laboratórios é considerado atualmente indispensável. As razões para o crescimento do uso desta técnica podem ser atribuídas à sua sensibilidade, sua pronta adaptabilidade para determinações quantitativas e à separação de diferentes analitos em uma matriz. Além disso, a CLAE possui uma variedade de mecanismos de separação possíveis podendo dessa forma, ser aplicada tanto para compostos orgânicos como inorgânicos. A única exigência é que a amostra seja solúvel na fase móvel (FM) ou eluente. Assim sendo, amostras líquidas ou sólidas, iônicas ou covalentes, como também sólidos de alta ou baixa massa molecular podem ser separados por essa técnica.<sup>34,36-38</sup>

A CLAE é definida como uma técnica físico-química de separação dos componentes de uma mistura, realizada através da distribuição destes entre duas fases. Uma das fases permanece estacionária, enquanto a outra se move através

dela. Devido à diferença nos coeficientes de distribuição de cada um dos analitos nas respectivas fases, obtêm-se velocidades de eluições diferentes, de tal modo que o composto menos retido na fase estacionária é eluído primeiramente.<sup>36</sup>

Na CLAE, a fase móvel (FM) é bombeada sob alta pressão a uma vazão controlada. Uma pequena quantidade de amostra (da ordem de  $\mu\text{L}$ ) é introduzida por meio de uma válvula de injeção. A amostra é então arrastada pela FM, passando pela coluna, onde ocorre a separação cromatográfica, e então os componentes da mistura chegam ao detector, no qual um sinal proporcional à concentração do analito é enviado ao computador. As colunas utilizadas são recheadas com fases estacionárias (FE), as quais são geralmente constituídas de partículas cujo tamanho varia de 1,5 a 10  $\mu\text{m}$ .<sup>36</sup>

Na cromatografia diversos parâmetros são importantes para a otimização ou desenvolvimento de um método analítico para fins qualitativos e/ou quantitativos. A separação dos diversos componentes que constituem a amostra depende dos diferentes graus de retenção de cada um desses componentes na coluna, sendo este o parâmetro mais importante em uma separação cromatográfica. Por definição, o *tempo de retenção* é o tempo decorrido entre a injeção da amostra e o máximo do pico referente à espécie em estudo. A certeza da presença do analito pode ser confirmada pela injeção de um padrão da mesma espécie de interesse, de concentração conhecida. Após a análise da amostra através da técnica de CLAE, obtém-se um cromatograma que consiste em um gráfico que exhibe a resposta do detector (tamanho do sinal) em função do tempo ou do volume de eluição. Cada um desses componentes eluídos pode ser identificado qualitativamente através do tempo de retenção ( $t_R$ ) na coluna cromatográfica e dos dados espectrais, e quantificado através de padrões previamente injetados no cromatógrafo.<sup>39</sup> A área obtida sob o pico no cromatograma é proporcional à concentração do componente na amostra original em análise, dentro de uma faixa de linearidade previamente definida. Desta forma, deve-se determinar antecipadamente uma curva analítica, traçando-se uma representação gráfica da área dos picos em função da concentração de uma série de padrões analisados.<sup>40</sup>

Quanto à eluição, esta pode ser de dois tipos: isocrática ou por gradiente. Na eluição isocrática, a concentração da fase móvel é mantida constante durante toda a corrida cromatográfica. Na eluição por gradiente, a força cromatográfica da

FM varia ao longo do processo, ou seja, a composição da FM varia de forma que a força aumente gradativamente. Este tipo de eluição gera maior simetria de picos, melhor resolução e detectabilidade, e menor tempo de análise.<sup>36</sup>

Sempre que possível, deve-se preferir a eluição isocrática, que é mais simples e de menor custo. Na eluição por gradiente, faz-se necessária a regeneração da coluna antes da próxima análise, condicionando-a as condições iniciais. Na maioria dos casos, um gradiente composto por uma mistura binária já é suficiente para solucionar os problemas de separação. Gradientes ternários e quaternários podem ser úteis, mas são de difícil programação e interpretação.

Atualmente, vários métodos envolvendo a CLAE estão sendo usados para determinar qualitativa e quantitativamente os corantes alimentares.<sup>5,32,33</sup> O modo de separação mais usado é o modo em fase reversa (CLAE-FR), tanto para a determinação de corantes, como de um modo geral, correspondendo à aproximadamente 80% das aplicações em áreas industriais e acadêmicas, pois ele permite separações analíticas e preparativas eficientes para uma ampla faixa de compostos.

Exemplos típicos de fases estacionárias para CLAE-FR são hidrocarbonetos (C8, C18) ligados quimicamente à sílica. Exemplos de fase móvel para CLAE-FR incluem solventes relativamente polares como água, metanol, acetonitrila ou tetrahydrofurano. Neste modo de separação, o mecanismo de retenção de moléculas de soluto (analitos) em CLAE-FR ocorre em uma região interfacial extremamente fina entre a estrutura sólida do suporte da FE, composta principalmente de sílica porosa e a FM.<sup>41</sup> Além do mais, o modo de separação em fases reversas apresenta curto tempo de equilíbrio entre análises, pela possibilidade de usar eluição na forma de gradiente e, principalmente, pelo uso de fase móvel rica em fase aquosa.<sup>42</sup>

Vários corantes permitidos no Brasil e presentes em alguns tipos de guloseimas como balas, gomas de mascar, confeitos de chocolate coloridos artificialmente e cereais matinais de vários sabores, foram determinados simultaneamente, utilizando o método de CLAE-FR.<sup>5,32,33</sup> Como ponto de partida para o desenvolvimento de um método é necessário obter informações sobre a composição da amostra e suas propriedades, tais como o número de compostos presentes, suas estruturas químicas, massas molares, valores de pKa, espectros de

absorbâncias no UV, intervalos de concentração dos analitos de interesse e solubilidades. Deve-se ter um objetivo claro em relação à separação, qualitativa ou quantitativa, sobre quantos analitos deseja-se isolar.

Nesta dissertação, uma metodologia por CLAE foi otimizada e validada para quantificação simultânea dos corantes amarelo crepúsculo e tartrazina em bebidas não alcoólicas. O objetivo inicial foi a obtenção de valores de referência para o desenvolvimento do método quimiométrico a partir da análise de imagens digitais. No entanto, considerando tanto o esforço experimental quanto intelectual envolvido no desenvolvimento deste método original, optou-se por dar maior destaque a esta parte da dissertação, cujos resultados serão descritos em um capítulo autônomo (Capítulo 3).

#### **2.4.1. Parâmetros Cromatográficos**

Uma coluna deve ser avaliada por meio de parâmetros cromatográficos, a fim de avaliar o seu potencial para futuras separações cromatográficas e esses parâmetros devem ser monitorados periodicamente. Existem vários parâmetros importantes na avaliação das colunas cromatográficas e dos analitos em uma corrida cromatográfica. Neste trabalho os parâmetros considerados em uma separação foram o fator de retenção ( $k$ ), a resolução ( $R_s$ ), o número de pratos teóricos ( $N$ ) e o fator de assimetria ( $A_s$ ).

O fator de retenção,  $k$ , é determinado pela razão dos tempos em que as moléculas ou íons do analito ficam retidas na fase estacionária ou percorrendo a coluna através da fase móvel (equação 1). Valores de  $2 < k < 10$  são considerados ideais, mas se aceita  $1 < k < 20$ .<sup>38,43</sup>

$$k = \frac{(t_R - t_M)}{t_M} \quad (1)$$

A resolução,  $R_s$ , é a medida quantitativa do grau de separação entre dois picos adjacentes e pode ser calculada de acordo com a equação 2:

$$R_s = 2 \left[ \frac{(t_{R2} - t_{R1})}{W_{b1} + W_{b2}} \right] = 1,177 \left[ \frac{(t_{R2} - t_{R1})}{W_{h1} + W_{h2}} \right] \quad (2)$$

Onde  $t_{R2}$  e  $t_{R1}$  são os tempos de retenção de dois analitos,  $w_{b1}$  e  $w_{b2}$  são as larguras dos respectivos picos na linha de base e  $w_{h1}$  e  $w_{h2}$  são as larguras dos picos a meia altura, com a largura medida em unidades de tempo.

A eficiência, medida pelo número de pratos teóricos,  $N$ , representa o número de etapas de equilíbrio do analito entre a fase móvel e a fase estacionária, calculada por meio da equação 3:

$$N = 16 \left[ \frac{t_R}{W_b} \right]^2 = 5,545 \left[ \frac{t_R}{W_h} \right]^2 \quad (3)$$

A simetria dos picos é avaliada pelo fator de assimetria ( $A_s$ ), que é a medida da proporção entre as duas partes de um pico cromatográfico no sentido horizontal a 10 % da sua altura. Os valores de  $A_s$  devem estar entre 0,95 e 1,3, sendo admitidos valores de até 1,5. Outra maneira de definir a forma do pico é através do fator de alargamento (TF) a 5 % da sua altura. Nessa abordagem a assimetria é medida a 5 % da altura do pico (Figura 3).<sup>44</sup>

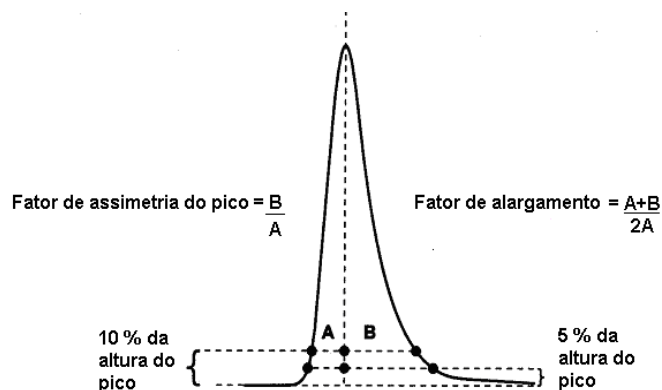


Figura 3. Determinação do fator de assimetria e alargamento do pico.<sup>44</sup>

## **2.5 Análise de imagens digitais**

Pode-se definir imagem como uma reprodução de um objeto real ou cena. Para fins científicos, as imagens são obtidas para expressar propriedades do objeto observado.<sup>45</sup> A imagem pode ser expressa como uma função matemática. No processo de digitalização, uma imagem digital é obtida, consistindo em uma estrutura na qual cada unidade básica recebe o nome de pixel.

As imagens digitais podem ter duas ou mais dimensões. Na aquisição de imagens científicas podem ser usadas as técnicas de projeção, varredura, tomográfica ou a combinação dentre estas.<sup>45</sup> A análise de imagens digitais vem sendo crescentemente utilizada em muitas áreas da ciência e da indústria. Como técnica analítica, ela é uma ferramenta não destrutiva e não invasiva, que possibilita obter informações quantitativas e qualitativas sobre as amostras observadas. Além disso, sua maior vantagem reside no fato de os dados serem obtidos de forma rápida e com baixíssimo custo por análise.

Na maioria das vezes, as informações visuais não são tratadas com o mesmo rigor científico que os dados numéricos. A análise de imagens em química muitas vezes é baseada em algumas inspeções visuais ao olho humano, que podem ser subjetivas e limitadas em termos de tomada de decisão.<sup>45</sup> A maioria das imagens contém uma grande quantidade de dados, dependendo de sua resolução. Para tratar esses dados de forma a transformá-los em informações químicas qualitativas e/ou quantitativas, a quimiometria tem um papel fundamental e indispensável.

A análise de imagens encontrou uso na indústria, por exemplo, na análise de propriedades de partículas, na avaliação de homogeneidade de misturas e em muitas outras aplicações.<sup>46</sup> A análise de imagens associada com a quimiometria é definida por alguns autores pela sigla MIA (*Multivariate Image Analysis*, análise de imagens multivariada),<sup>45</sup> podendo ser aplicada tanto no processamento de imagens digitais quanto hiperespectrais. Considerando a natureza multivariada intrínseca de imagens digitais, a maioria dos trabalhos utilizou métodos quimiométricos de análise exploratória ou classificação não supervisionada, como a PCA (*Principal Component Analysis*, análise de componentes principais), para extrair informações qualitativas. Para extrair informações quantitativas de imagens digitais, é necessário o uso de

métodos de calibração multivariada, como a regressão em componentes principais (PCR, *Principal Component Regression*) e os mínimos quadrados parciais (PLS, *Partial Least Squares*), dando origem ao termo MIR (*Multivariate Image Regression*, regressão multivariada de imagens).<sup>47</sup>

Imagens digitais já vêm sendo utilizadas como fonte de informação analítica desde o século passado. O primeiro trabalho que descreveu a utilização de imagens digitais empregou um equipamento rudimentar similar a um escâner para converter exames médicos em dados digitais, que serviram como base para diagnósticos mais rápidos e exatos.<sup>48</sup> Mais de vinte anos depois, em 1986, houve a publicação do primeiro artigo que empregou um tratamento quimiométrico de imagens digitais.<sup>49</sup> Desde então, principalmente nos últimos dez anos, diversos artigos têm sido publicados utilizando diferentes instrumentos para a aquisição de imagens, tais como celulares, *webcams*, escâneres de mesa e câmeras digitais.<sup>5-5</sup>

A maneira mais comum de se extrair informações de imagens digitais é a sua decomposição em um sistema de cores, como o RGB. O sistema RGB é um sistema que utiliza a combinação das cores Vermelho (*Red*), Verde (*Green*) e Azul (*Blue*) para formar uma ampla variedade de tons (Figura 4).

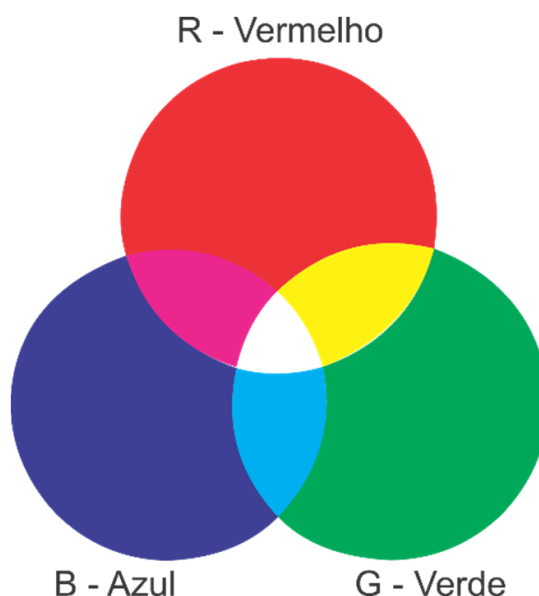


Figura 4. . Representação gráfica do sistema RGB

Fonte: <http://www.jbprint.com.br/blog/imgs/rgb.jpg>

Cada pixel, unidade formadora básica de uma imagem digital, é formado pela combinação dessas três cores. A intensidade de cada cor no sistema RGB é medida em canais (256 para cada cor). O canal 0 significa completa ausência de uma cor, enquanto o canal 255 representa essa mesma cor em sua intensidade máxima. A combinação dos canais do sistema RGB fornece  $256^3$  possíveis tonalidades de cor (Figura 5).

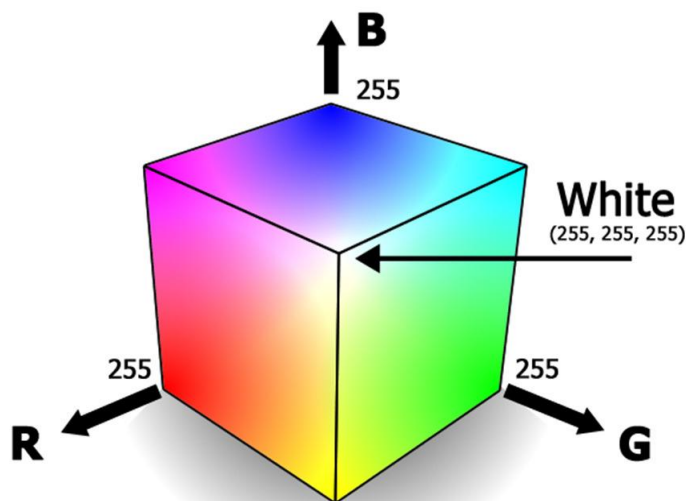


Figura 5. Formação de tons no sistema RGB.

Fonte: <https://fperrotti.wikispaces.com/file/view/cuborgb.jpg/418292514/27x223/cuborgb.jpg>

Após a decomposição de todos os pixels da imagem, a frequência de cada canal é contada, gerando um histograma de frequências. Este histograma contém a frequência em função dos canais, um tipo de informação que é proporcional à composição da amostra em termos de compostos que apresentem cor. Dessa maneira, um conjunto de histogramas para uma série de amostras pode ser usado para construir uma matriz, que será tratada usando métodos quimiométricos de calibração multivariada ou classificação. Neste trabalho os histogramas serão tratados usando o método PLS para a quantificação de dois corantes. Existem ainda outras estratégias para o tratamento quimiométrico de imagens digitais, que envolvem o uso de outros sistemas de cor ou a combinação das variáveis RGB com outros parâmetros, como saturação, intensidade e matiz, resultando nos chamados colorgramas.<sup>53,55,57</sup> Na presença de fortes variações texturais nas imagens (variações de pixel para pixel, ou seja, heterogeneidade das

imagens), existem alternativas mais complexas em termos quimiométricos, como o pré-processamento por transformada de Fourier, seguido da construção de uma matriz tridimensional (pixel em x *versus* pixel em y *versus* intensidade) e o uso de métodos de ordem superior, como o PLS multilinear (N-PLS).<sup>56,58</sup> Como no presente trabalho serão tratadas amostras que apresentam cor homogênea sem variação de textura (as bebidas depois de submetidas ao ultrassom), a estratégia mais simples e parcimoniosa será adotada, tratando os histogramas como se fossem espectros na construção de um modelo de calibração multivariada.

## **2.6 Quimiometria**

A quimiometria nasceu da necessidade de extrair informação química da quantidade cada vez maior de dados gerados pelos instrumentos e programas computacionais modernos. Ela utiliza ferramentas matemáticas e estatísticas para extrair de conjuntos de dados químicos informação útil.

Historicamente, o artigo de Kowalski e Bender, publicado em 1972, teve grande impacto e foi considerado um marco do início das pesquisas em quimiometria.<sup>59</sup> Em 1976, o Simpósio *Chemometrics: Theory and Application*, patrocinado pela Divisão *Computers in Chemistry* da *American Chemical Society*, resultou no primeiro livro voltado especificamente para a quimiometria.<sup>60,61</sup>

Em 1980 a revista *Analytical Chemistry* substituiu o título *Statistical and Mathematical Methods in Analytical Chemistry* por “*Chemometrics*” em sua revisão bianual, consolidando o uso deste termo em inglês. E, finalmente, em 1987 dois jornais exclusivos da área de quimiometria foram lançados: o *Journal of Chemometrics* e o *Chemometrics and Intelligent Laboratory Systems*.<sup>60</sup>

De modo geral, no Brasil, o desenvolvimento da quimiometria pode ser dividido em três grandes momentos: antes do microcomputador, depois do micro e atualmente. No final da década de 70, os microcomputadores ainda não tinham chegado ao Brasil, sendo grande a dificuldade em processar cálculos extensos. Em 1980, o professor Kowalski ministrou no Instituto de Química da Unicamp o primeiro curso no país sobre quimiometria. Nesta mesma década, os primeiros

microcomputadores de 8 bits começaram a chegar aos laboratórios brasileiros. Problemas de incompatibilidade entre programas desenvolvidos e as máquinas eram comuns. O grande avanço se deu por volta de 1985 com os primeiros microcomputadores de 16 bits, tornando possível executar em um determinado micro um programa compilado em outro computador, abrindo possibilidade da quimiometria ser implantada como disciplina acadêmica. Resultados de trabalhos realizados pelo grupo pioneiro de pesquisa geraram o primeiro livro sobre quimiometria publicado no Brasil, em 1995.<sup>62</sup> Uma versão atualizada e ampliada foi lançada, em 2001.<sup>63</sup> Ao longo de 25 anos (1980-2005), as áreas consolidadas no Brasil podem ser agrupadas da seguinte maneira: planejamento e otimização de experimentos, reconhecimento de padrões (métodos de análise exploratória e classificação) e calibração multivariada.<sup>62</sup> Essa última área é de interesse desta dissertação.

Há várias definições da palavra quimiometria, dentre elas está a definição original da *International Chemometrics Society*: A quimiometria é a aplicação de ferramentas matemáticas e estatísticas à química.<sup>64</sup>

A quimiometria é uma ferramenta essencial na química analítica, mas, não está presa a essa área; ela possui natureza multidisciplinar e tem marcado presença em indústrias químicas, farmacêuticas, de alimentos, além de estar crescendo seu uso em outros campos, como na área médica, no tratamento de imagens e nos estudos de metabolômica.

## **2.7 Calibração multivariada**<sup>60,65</sup>

A calibração multivariada visa à construção de modelos quantitativos a partir de uma grande quantidade de variáveis. Esses modelos relacionam dois conjuntos de dados: Um deles contém normalmente medidas instrumentais, representadas na matriz  $\mathbf{X}$ , de variáveis independentes. O outro conjunto contém os valores de teores de analitos ou propriedades de interesse das amostras, representados no vetor  $\mathbf{y}$  (se for uma única variável), ou na matriz  $\mathbf{Y}$ , (se forem

várias), de variáveis dependentes. O resultado de uma calibração pode ser representado pelo esquema da Figura 6:

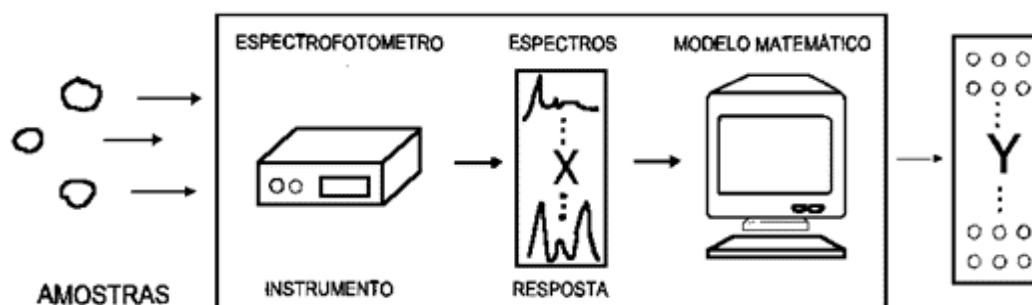


Figura 6. Diagrama representando uma calibração multivariada<sup>65</sup>

O modelo estimado poderá ser usado para prever a concentração do analito de interesse em amostras de composição desconhecida, usando as respostas instrumentais das mesmas. Os dados das variáveis independentes (espectros, imagens digitais ou outro tipo de sinal analítico) são organizados numa matriz,  $\mathbf{X}$  ( $n \times p$ ), na qual cada linha representa uma amostra e cada coluna representa uma variável medida (Figura 7).

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1p} \\ X_{21} & X_{22} & \dots & X_{2p} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{np} \end{bmatrix}$$

Figura 7. Matriz das variáveis independentes

O outro conjunto de dados é constituído das variáveis dependentes e organizado na matriz  $\mathbf{Y}$ , caso haja mais de um analito de interesse, ou no vetor  $\mathbf{y}$ , no caso de um único analito determinado. O total de elementos deste vetor é igual a  $n$ , isto é, o número de amostras analisadas.

O processo geral de calibração consiste de duas etapas: a modelagem/calibração, que estabelece uma relação matemática entre  $\mathbf{X}$  e  $\mathbf{Y}$  para um grupo de amostras representativas contidas no conjunto de calibração, e a previsão, em que o modelo desenvolvido é usado para a previsão de novas amostras. Uma vez concluída a calibração, o sistema (instrumento físico + modelo

matemático) representado esquematicamente no diagrama da Figura 5 está apto a ser utilizado.

Os métodos mais consagrados na literatura de calibração multivariada, todos de calibração inversa, são a regressão linear múltipla (RLM), a regressão de componentes principais (PCR), os quadrados mínimos parciais (PLS), as redes neurais artificiais (ANN) e as máquinas de suporte de vetores (SVM). Dentre estes, os mais utilizados são o PCR e o PLS. Estes dois métodos são capazes de lidar com grande presença de colinearidade e ruídos nos dados, assim como permitem manipular arranjos de dados com mais variáveis que amostras (mais colunas do que linhas na matriz de dados), o que é uma limitação para a RLM.

## **2.8 Ferramentas quimiométricas**

### **2.8.1 Partial Least Squares (PLS)<sup>65-67</sup>**

O PLS é uma regressão multivariada baseada na decomposição dos dados através de uma análise de fatores. Trata-se do método de calibração multivariada mais popular em quimiometria. Nele, ao contrário da PCR, as variáveis independentes e dependentes são decompostas simultaneamente, antes que se estabeleça a regressão entre elas. Neste caso, os fatores são denominados variáveis latentes (VL), ao invés de componentes principais, pois a etapa de decomposição simultânea impede a restrição de ortogonalidade. As variáveis latentes são definidas de maneira tal que expliquem simultaneamente a variância em **X** e a variância em **Y**.

Na construção de modelos PLS, **X** e **Y** são decompostas simultaneamente em uma soma de  $h$  variáveis latentes. **X** é uma matriz  $n \times p$ ; sendo  $n$  o número de amostras e  $p$  o número de variáveis medidas; **Y** pode ser um vetor (**y**)  $n \times 1$ , quando o modelo PLS prediz apenas um analito de cada vez (PLS1), ou uma matriz  $n \times k$ , quando o PLS prediz  $k$  analitos simultaneamente (PLS2). É importante

ressaltar que a denominação PLS1 ou PLS2 não está relacionada com a quantidade de analitos preditos, e sim com a ordem do tensor dos valores de referência, pois um vetor é um tensor de primeira ordem, enquanto uma matriz é um tensor de segunda ordem.

A base estrutural do modelo PLS é dada pelas equações 4 e 5.

$$X = TP^t + E = \sum t_h p_h^t + E \quad (4)$$

$$Y = UQ^t + F = \sum u_h q_h^t + F \quad (5)$$

Onde **T** e **U** são as matrizes de escores das matrizes **X** e **Y**, respectivamente, **P** e **Q** são as matrizes de pesos (*loadings*) de **X** e **Y**, respectivamente, e **E** e **F** são os resíduos de **X** e **Y**, respectivamente, conforme mostrado na Figura 8.

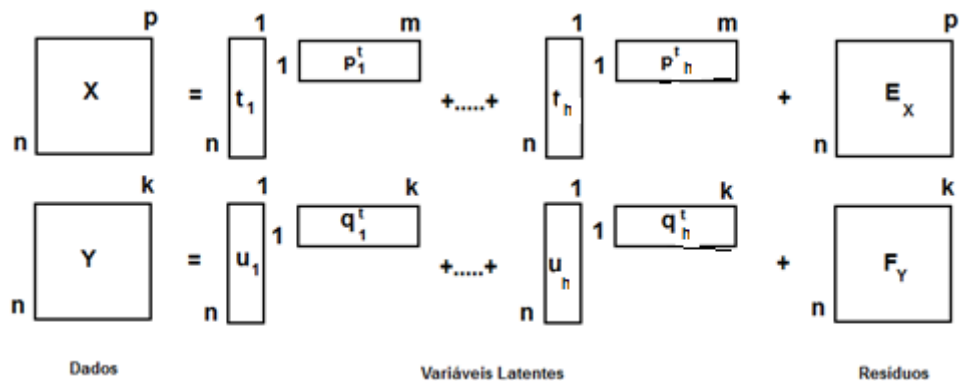


Figura 8. Decomposição em VL das matrizes X e Y para modelos PLS

O vetor de coeficientes de regressão linear **b** correlaciona os blocos **X** e **Y** de forma linear, para *h* VL, de acordo com a equação 6.

$$u_h = b_h \times t_h \quad (6)$$

Os valores de *b<sub>h</sub>* são agrupados na matriz diagonal **B**, que contém os coeficientes de regressão entre as matrizes de escores **T** de **X** e **U** de **Y**. A melhor relação linear possível entre os escores desses dois blocos é obtida através de

pequenas rotações das variáveis latentes dos blocos **X** e **Y**. Os valores preditos para novas amostras ( $\hat{Y}$ ) podem ser então calculados de acordo com a equação 7, com base nos seus escores  $T^*$ .

$$\hat{Y}=T^*BQ \quad (7)$$

Para a construção dos modelos de calibração é necessário que as amostras sejam divididas em conjuntos de calibração e de validação, de maneira que o primeiro conjunto seja representativo de toda a variância que se deseja modelar. O conjunto de validação deve conter amostras homogeneamente distribuídas dentro da faixa de composição, sem extrapolar o conjunto de calibração. Em situação de calibração natural, como neste trabalho, em que não há controle da composição da matriz analítica, ou seja, não se podem preparar as amostras de acordo com um planejamento prévio, deve-se usar algum método que garanta a seleção sistemática das amostras mais representativas no conjunto de calibração. Para este fim, foi usado o algoritmo de Kennard-Stone.

#### ***2.8.1.1 Algoritmo de seleção Kennard-Stone***

O algoritmo de seleção Kennard-Stone utiliza inicialmente a distância euclidiana para definir as duas amostras mais distantes entre si. Em seguida, o algoritmo seleciona a amostra mais distante das duas amostras inicialmente selecionadas. Este processo é repetido até que a quantidade de amostras a ser selecionada, previamente definida pelo analista, seja alcançada. Desta forma, o algoritmo Kennard-Stone garante a presença de amostras representativas do modelo, homogeneamente distribuídas, no conjunto de calibração.<sup>68</sup>

### **2.8.1.2 Validação cruzada**

Uma etapa importante é a validação cruzada para a escolha do número de VL a ser usado na construção de modelo PLS, em que se separa uma parte (ou apenas uma) das amostras de calibração e constrói-se o modelo com as restantes. Existem vários tipos de validação cruzada, variando a forma de como a amostra, ou o subconjunto de amostras, é retirada dos dados. Os mais comuns são *leave-one-out*, blocos contínuos, subconjuntos aleatórios e venezianas (*venetian blinds*). Para conjuntos de dados grandes (usualmente com mais que 20 amostras) e em que as amostras estão organizadas em ordem aleatória é recomendado a validação por blocos contínuos.<sup>60</sup>

Posteriormente, os erros de previsão são estimados para as amostras que foram separadas, utilizando diferentes números de VL. Esse processo é repetido para outras amostras, até que todas tenham ficado de fora. Em aplicações de calibração multivariada usando PLS, o critério para a escolha do número de VL é o menor valor da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, *root mean square error of cross-validation*).<sup>69</sup> Em toda esta dissertação usou-se a validação cruzada do tipo blocos contínuos.

### **2.8.1.3 VIP Scores**

A interpretação espectral (variáveis independentes) de modelos PLS é usualmente realizada baseando-se na análise dos coeficientes de regressão do modelo. Entretanto, esta interpretação não deve se ater somente nestes coeficientes, pois eles são dependentes da composição das amostras no conjunto de calibração, da covariância implícita entre os componentes dessas amostras e da relação sinal/ruído dos dados analíticos.<sup>70</sup>

Uma ferramenta mais eficiente para a interpretação espectral são os gráficos de Importância das Variáveis na Projeção dos escores (VIP Scores – Variable Importance in Projection). Eles estimam a importância de cada variável na

projeção utilizada pelo modelo PLS através dos coeficientes em cada componente, juntamente com a significância de cada componente na regressão em módulo.<sup>71</sup> Ou seja, os VIP Scores são mais robustos que os coeficientes de regressão para identificar quais variáveis são mais significativas para o modelo, apresentando maior contribuição. Porém, diferentemente dos coeficientes de regressão, os VIP Scores fornecem valores em módulo.

Os VIP Scores podem ser utilizados, também, para a seleção de variáveis. O critério para a seleção da variável  $j$  é a média dos quadrados dos VIP Scores serem maiores ou iguais a 1,0. A importância da variável de previsão de ordem  $j$  com base no modelo com  $h$  VL pode ser calculada por equação 8:

$$VIP_j = \sqrt{p \sum_{k=1}^h \left( SS(b_k t_k) (w_{jk} / \|w_k\|)^2 \right) / \sum_{k=1}^h SS(b_k t_k)}, \quad (8)$$

Onde  $J$  é o número de variáveis previstas,  $w_{hj}$  é a importância do peso (*weight*) da  $j$ ésima variável no  $h$ ésimo fator PLS e  $SS(\mathbf{b}_h \mathbf{t}_h)$  é a porcentagem de  $Y$  explicada pelas  $h$  VL.<sup>71</sup>

### 2.8.2 Ordered Predictors Selection (OPS)<sup>72,73</sup>

Os quimiometristas costumam visualizar gráficos de vetores informativos para extrair informações desejadas de modelos multivariados. Os vetores informativos são aqueles obtidos a partir do modelo matemático dos dados, que tornam mais visíveis padrões de comportamentos das variáveis independentes na sua relação com as variáveis dependentes. Uma característica desejável em um método de seleção de variáveis é sua característica universal. O uso de vetores prognósticos, por exemplo, é universal, visto que a informação extraída por eles é intrínseca das variáveis, independente de sua natureza. Muitos autores<sup>70,71,74,75</sup> e o programa Pirouette<sup>TM</sup><sup>76</sup> apresentam o vetor de regressão como potencial ferramenta informativa para auxiliar na seleção de variáveis em calibração multivariada.

Variáveis com pequenos coeficientes de regressão não contribuem significativamente para a previsão e, portanto, podem ser eliminadas. O vetor de regressão pode ser considerado ainda, como uma soma ponderada dos *loadings* incluídos no modelo.<sup>76</sup> Portanto, o formato do vetor muda dramaticamente com o número de componentes quando informações importantes são incluídas e poucas mudanças ocorrem quando ruídos passam a ser modelados e toda informação já está descrita nas componentes. Outros vetores informativos em calibração multivariada para seleção de variáveis são aqueles que relacionam, de maneira univariada, cada variável independente com a variável dependente e retiram parâmetros desta relação, como por exemplo, a correlação. Normalmente, variáveis com pobres correlações não trazem informações importantes e podem ser excluídas. Entretanto, uma atenção maior deve ser dada, visto que este vetor traz informações univariadas e não multivariadas. Além desses, o OPS pode iniciar a otimização também através de outros vetores informativos, tais como o vetor contendo os resíduos de cada variável, o vetor do sinal analítico líquido, os VIP Scores e o vetor contendo a razão sinal/ruído para cada variável, ou através de combinações deles.

O *Ordered Predictors Selection* (OPS), ou Seleção de Preditores Ordenados, é uma estratégia simples e intuitiva para seleção de variáveis, baseada nos vetores informativos e nas combinações entre eles. O ponto central está na obtenção de um vetor que contenha a informação da localização das melhores variáveis independentes para previsão. Logicamente, a dimensão do vetor deve ser igual ao número de variáveis independentes e cada posição do vetor deve estar alinhada com a sua variável independente correspondente.

Em uma segunda etapa, a partir do vetor prognóstico gerado, as variáveis independentes originais (colunas da matriz **X**) são decrescentemente ordenadas de acordo com a ordem dos seus valores correspondentes, contidos no vetor. Os maiores valores no vetor informativo indicam a posição das variáveis independentes originais mais importantes (Figura 9).

Na terceira etapa, as variáveis ordenadas são avaliadas usando alguma estratégia de validação cruzada com algum método de regressão multivariada. Os modelos são construídos e avaliados sobre uma janela de variáveis e, posteriormente, sobre a janela acrescida de um incremento fixo de variáveis, e assim sucessivamente, até que todas as variáveis ou uma percentagem do total

sejam adicionadas à janela. Parâmetros de qualidade dos modelos são obtidos a cada avaliação, para posterior comparação.

Na quarta e última etapa, os conjuntos de variáveis avaliados (janela inicial e janela mais incrementos) são comparados utilizando os parâmetros de qualidade calculados como o R (coeficiente de correlação) e o RMSECV. O conjunto de variáveis que apresentar os melhores parâmetros de qualidade contém as variáveis que apresentam a melhor capacidade de previsão para o modelo construído e, portanto, estas serão as variáveis selecionadas.

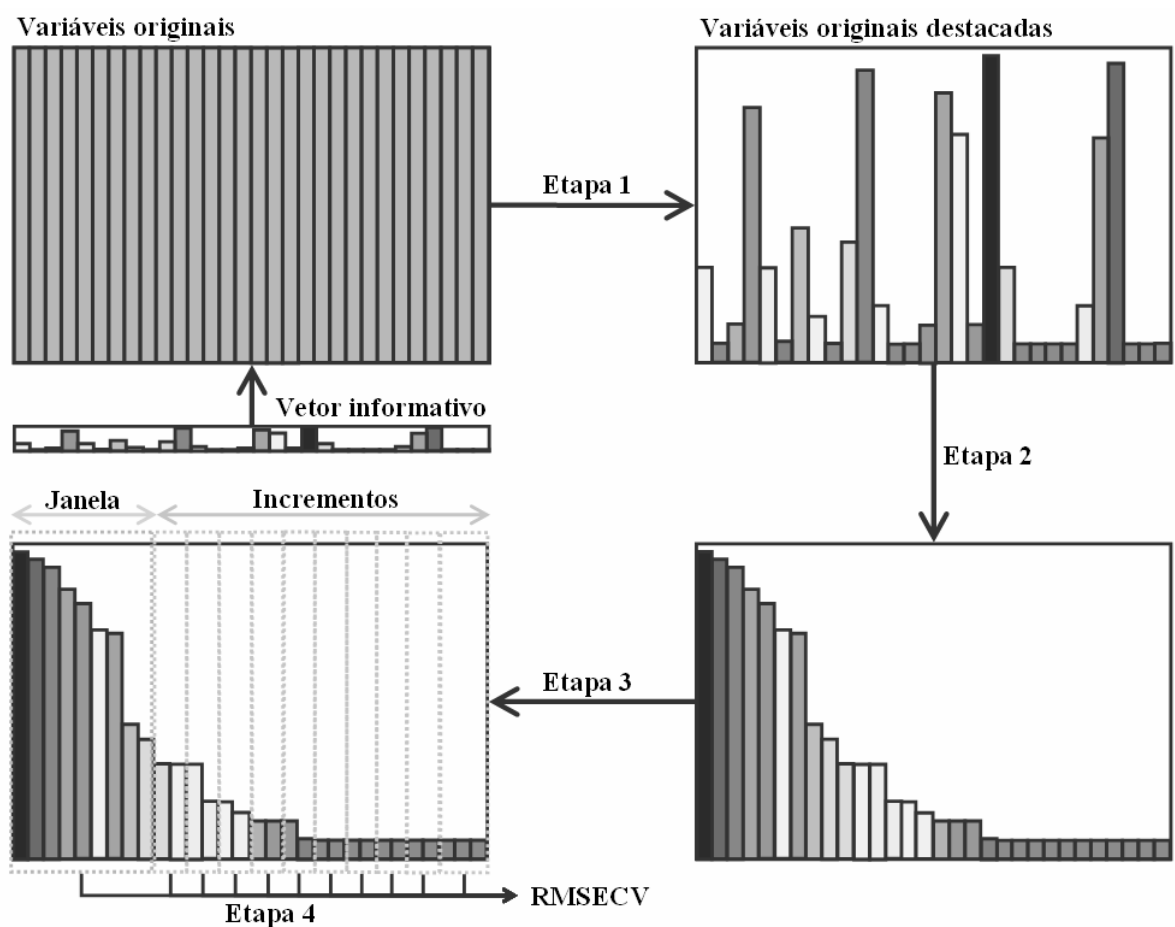


Figura 9. Etapas da seleção de variáveis usando o OPS<sup>72</sup>

## 2.9 Validação multivariada<sup>75-79</sup>

O desenvolvimento de novos métodos analíticos requer a averiguação da confiabilidade, com a orientação e regulamentação por órgãos fiscalizadores, como

a Agência Nacional de Vigilância Sanitária (ANVISA) e o Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (INMETRO). É importante demonstrar que o método é adequado e confiável para a finalidade à qual se destina. A isso se dá o nome de validação analítica.

A fim de otimizar um método de calibração multivariada e que esse seja robusto às novas amostras, é necessário que ele seja calibrado com amostras representativas. Visando isso, são feitas a seleção de amostras de calibração e a remoção de amostras anômalas (*outliers*). São consideradas amostras anômalas aquelas que: 1) apresentam alta influência (*leverage*) no modelo; 2) apresentam grandes resíduos no sinal analítico, portanto, não estão bem modeladas; e 3) apresentam grandes resíduos na resposta analítica, portanto, não são bem previstas.

Os parâmetros de validação são importantes para apontar se um método atende satisfatoriamente à sua finalidade, com um nível desejado de confiança estatística. A maneira que esses parâmetros são calculados é estabelecida em normas específicas para determinados métodos ou guias de validação de modo geral. Os parâmetros são calculados a partir da calibração entre um sinal analítico, referente ao analito, e o valor de referência. Porém, em análises multivariadas o sinal analítico é referente tanto ao analito como à matriz. Para suprir a necessidade de um sinal analítico multivariado referente ao analito, desenvolveu-se o conceito de Sinal Analítico Líquido (NAS – *Net Analyte Signal*).

### **2.9.1. Sinal Analítico Líquido – NAS**

O Sinal Analítico Líquido, proposto originalmente por Lorber,<sup>80</sup> é a parte do sinal analítico que advém apenas do analito. Por cálculos matriciais, pode-se fazer uma separação entre um vetor do sinal dos interferentes e um vetor NAS, ortogonal ao primeiro e, portanto, independente, como mostrado na Figura 11. O método de cálculo original do NAS foi proposto para modelos de mínimos quadrados clássicos (CLS – *classical least squares*), de calibração clássica/direta. Posteriormente, o método de cálculo foi modificado,<sup>81</sup> permitindo o seu uso em

calibração inversa, como nos métodos PLS e PCR, de aplicação muito mais ampla em calibração multivariada.

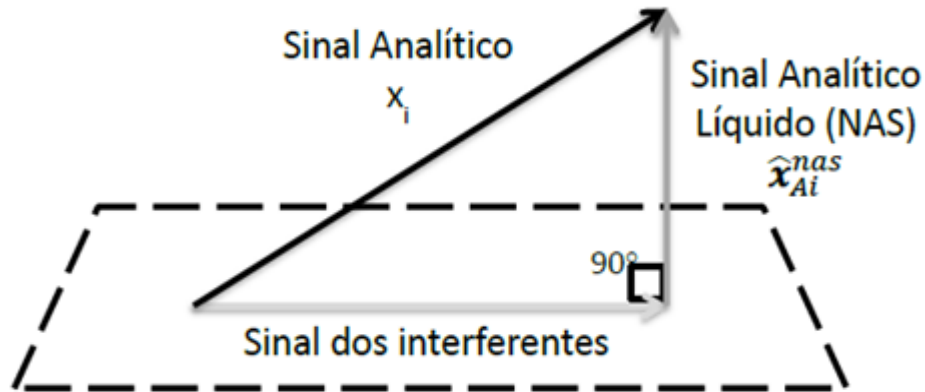


Figura 10. Representação vetorial do NAS (adaptado de Valderrama) <sup>77</sup>

O vetor  $\hat{x}_{Ai}^{nas}$  pode ser calculado para cada amostra  $i$ , a partir dos coeficientes de regressão  $b$  de um modelo PLS com  $h$  variáveis latentes, de acordo com a equação 9.

$$\hat{x}_{Ai}^{nas} = b(b^T b)^{-1} b^T x_i \quad (9)$$

A norma do vetor  $\hat{x}_{Ai}^{nas}$  fornece um valor escalar  $\widehat{nas}_i$  (análogo ao sinal analítico seletivo de uma análise univariada) para cada uma das amostras que são organizados em um vetor  $\widehat{nas}$ . O NAS pode ser usado tanto para a estimativa de uma curva pseudo-univariada,<sup>82</sup> que é uma representação univariada de um sinal analítico multivariado, quanto para figuras de mérito especificamente usadas em validação multivariada, como sensibilidade, seletividade e limite de detecção.

### 2.9.2 Figuras de mérito

Algumas figuras de mérito são calculadas de maneira similar tanto para a calibração multivariada como para a calibração univariada. Entretanto, muitas das

figuras de mérito apresentam diferenças significativas nesses cálculos, exigindo novos conceitos, como o de NAS.

A veracidade é a estimativa do erro entre o valor previsto e o valor de referência. Apesar dessa figura mérito poder ser avaliada como nos métodos univariados, é comum considerar para avaliar esse parâmetro as raízes quadráticas dos erros médios quadráticos de calibração (RMSEC) e de previsão (RMSEP). Sendo que o RMSEP é calculado a partir das amostras de validação não utilizadas na construção dos modelos e, por isso, possui maior robustez para avaliação da veracidade.

A precisão é uma estimativa da dispersão dos valores de uma série de medidas, em níveis de repetibilidade, precisão intermediária ou reprodutividade. Usa-se como parâmetro o desvio padrão relativo (DPR).

A linearidade pode ser verificada por meio do coeficiente de correlação  $r$  da reta estimada para os valores de referência versus os valores previstos, além de assegurar a aleatoriedade dos resíduos da regressão, que deve ser comprovada de maneira objetiva, por meio de testes estatísticos.<sup>83</sup> Esses testes avaliam a normalidade (Ryan-Joiner), a homocedasticidade (Brown-Forsythe) e a independência (Durbin-Watson) dos resíduos e o seu uso foi recentemente estendido a métodos multivariados.<sup>84,85</sup>

A normalidade dos resíduos deve ser avaliada através do teste de Ryan-Joiner (RJ), no qual os resíduos ( $e_i$ ) são ordenados de forma crescente e plotados contra a porcentagem de pontos correspondentes ( $q_i$ ) a partir da distribuição normal padrão (gráfico quantil-quantil). Se os resíduos apresentam comportamento normal, os dados se aproximam do ajuste de uma linha reta. Caso eles sigam uma distribuição alternativa, o gráfico exibirá algum grau de curvatura. O coeficiente de correlação ( $R_{eq}$ ) para os dados normalmente distribuídos é, então, comparado com os valores de  $R_{crítico}$  calculados, para a rejeição ou aceitação da hipótese nula de normalidade. Os resíduos são considerados normalmente distribuídos se  $R_{eq} > R_{crítico}$ .

O teste Brown-Forsythe (BF), ou Levene modificado, é usado na avaliação da homocedasticidade dos resíduos (variação absoluta constante de todos os pontos da faixa de trabalho). Os resíduos devem ser divididos em dois grupos de tamanho  $n_1$  e  $n_2$ . Os resíduos correspondentes aos valores da metade inferior da

faixa ficam no grupo 1 e os resíduos correspondentes à metade superior da faixa ficam no grupo 2. As medianas ( $\tilde{e}_1$  e  $\tilde{e}_2$ ) de cada grupo são calculadas. Em seguida, os valores absolutos das diferenças entre os resíduos e as medianas de cada grupo  $d_{i1}=|e_i - \tilde{e}_1|$  e  $d_{i2}=|e_i - \tilde{e}_2|$  devem ser calculados. Por fim, a soma dos desvios quadrados de  $d_j$  (SSD1 e SSD2) de cada grupo devem ser estimadas. A estatística de  $t_L$  (t de Levene) é avaliada pela equação 10

$$t_L = \frac{\bar{d}_1 - \bar{d}_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_p^2}} \quad (10)$$

na qual é a média de  $d_{ij}$  de cada grupo,  $n_j$  é o número de amostras de cada grupo e  $s_p^2 = \text{SSD1} + \text{SSD2}/(n_1 + n_2 - 2)$ . Os resíduos são considerados homocedásticos se  $t_L < t_{\text{critico}}$ .

A independência dos resíduos é avaliada com o teste de Durbin-Watson (DW). Este teste é definido pela equação 11

$$d = \frac{\sum_{i=1}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (11)$$

A ausência de autocorrelação é verificada através da comparação entre os desvios padrão dos resíduos ordenados (denominador da Eq. 11) e das diferenças sucessivas desses mesmos resíduos (numerador dessa equação). Para cada conjunto de dados, existem dois limites para  $d$  ( $d_L$  é o limite inferior e  $d_U$  é o limite superior). Se o valor de  $d$  estimado para um conjunto de dados estiver fora destes limites, o teste é inconclusivo ou existe autocorrelação (pode-se também calcular limites para estas duas situações). Os valores de DW variam de 0 a 4 com média 2. Se o valor calculado convergir para 2 significa que não há autocorrelação.

A seletividade (SEL) é calculada para cada amostra como a razão entre a norma do vetor NAS e norma do vetor de sinal analítico ( $\mathbf{x}_i$ ). A seletividade do

método é calculada como a média das seletividades individuais das amostras. Para calibração multivariada a seletividade não tem a mesma relevância como na univariada. Na calibração univariada é necessário que haja seletividade de sinal próxima a 100%, enquanto na calibração multivariada é a falta de seletividade que torna necessário o uso de quimiometria. O símbolo “ $\| \cdot \|$ ” indica a norma euclidiana (ou módulo) de um vetor (equação 12).

$$SEL = \frac{\widehat{nas}_i}{\|x_i\|} \quad (12)$$

A sensibilidade (SEN) é a capacidade do método variar a resposta analítica por unidade de sinal analítico. Para os métodos multivariados, a sensibilidade líquida de uma amostra  $i$  pode ser calculada a partir do vetor NAS  $\widehat{x}_{Ai}^{nas}$  (equação 13) e a sensibilidade do método pela equação 14.

$$S_i^{nas} = \left\| \frac{\widehat{x}_{Ai}^{nas}}{y} \right\| \quad (13)$$

$$SEN = \|S_i^{nas}\| \quad (14)$$

A sensibilidade é dependente da técnica analítica. O parâmetro sensibilidade analítica ( $\gamma$ ) é mais útil para fins de comparação entre técnicas distintas e pode ser calculado conforme a equação 15.

$$\gamma = \frac{SEN}{\varepsilon} \quad (15)$$

Onde  $\varepsilon$  é o ruído instrumental, estimado como o desvio padrão combinado de repetidas medidas do branco. O inverso da sensibilidade analítica,  $(\gamma)^{-1}$ , indica a menor concentração discernível pelo método, considerando o ruído instrumental aleatório como única fonte significativa de erro, e indica o número de algarismos significativos com que os resultados devem ser expressos.

O viés ou *bias*, definido como a diferença entre o valor de uma medida (valor previsto) e o valor de referência, é a medida de erros sistemáticos no modelo, calculado apenas com as amostras de validação. A norma ASTM<sup>86</sup> propõe como calcular essa figura de mérito para modelos multivariados.

$$viés = \sum_{i=1}^{n_v} \frac{(y_i^{ref} - \hat{y}_i)}{n_v} \quad (16)$$

Em que  $Y_i^{ref}$  é a medida de referência e  $\hat{Y}_i$  o valor previsto para cada amostra. O número de amostras do conjunto da validação é  $n_v$ .

O desvio padrão de erros das amostras de validação (SDV) é usado para avaliar através de um teste t de Student com  $n_v$  graus de liberdades, se o viés é significativamente diferente de zero.

$$SDV = \sqrt{\frac{\sum[(y_i^{ref} - \hat{y}_i) - bias]^2}{n_v - 1}} \quad (17)$$

$$t_{calc} = \frac{|viés| \sqrt{n_v}}{SDV} \quad (18)$$

A Relação de Desempenho do Desvio (RPD)<sup>87</sup> é uma figura de mérito utilizada para verificar a eficiência do método em relação à variação natural das amostras. Este parâmetro é utilizado para comparação entre modelos de diferentes faixas analíticas ou em termos absolutos. O RPD é calculado separadamente para os conjuntos de calibração e validação conforme equações 19 e 20.

$$RPD_{Cal} = \frac{DP_{cal}}{RMSECV} \quad (19)$$

$$RPD_{Val} = \frac{DP_{val}}{RMSEP} \quad (20)$$

Em que DP é o desvio padrão dos valores de referência para as amostras do conjunto de calibração (DPcal) e de validação (DPval).

### 3. Otimização e Validação do Método Cromatográfico

#### 3.1 Parte Experimental

##### 3.1.1 Equipamentos

Sistema de purificação de água (Milli-Q Millipore Corporation<sup>®</sup>), sistema de filtração a vácuo (Marconi<sup>®</sup>), pHmetro Methron modelo 744 e banho de ultrassom Soniclean 2 PS. Cromatógrafo a líquido, marca Shimadzu, utilizado nas separações por fase reversa, consistindo dos seguintes componentes:

- ✓ Bomba de alta pressão tipo pistão, recíproca, cabeça dupla, modelo LC-6AD;
- ✓ autoinjeter SIL20A;
- ✓ Forno para coluna modelo CTO-20A;
- ✓ Interface modelo CBM 20A;
- ✓ Detector espectrofotométrico de arranjo de diodos UV-Vis modelo SPD-M20A,
- ✓ Injetor Rheodyne com alça amostradora de 20µL, modelo 725;
- ✓ *Software class* LC-Solution;

##### 3.1.2 Reagentes e Solventes

- Os padrões utilizados foram: Azul Patente V (ALDRICH<sup>®</sup> - lote BCBF7733V), Tartrazina PA (Dinâmica<sup>®</sup> - lote 45098), Amarelo Crepúsculo (ALDRICH<sup>®</sup> - lote MKBC5362).
- Acetato de Amônia PA (ISO FAR).
- Metanol grau HPLC (J.T.Baker<sup>®</sup> - lote M14CO3).

- Acetonitrila grau HPLC (JTBaker<sup>®</sup> - lote L46C63).

### 3.1.3 Coluna cromatográfica

A coluna cromatográfica utilizada foi a de nome comercial Luna<sup>®</sup>, marca Phenomenex<sup>®</sup>, de fase estacionária C18, (100 x 3,00 mm) e tamanho de partícula de 5 µm.

### 3.1.4 Preparo da coluna cromatográfica

Antes das separações cromatográficas é preciso condicionar a coluna com uma FM previamente definida. Em geral, o próprio manual do fabricante especifica como fazer este condicionamento. Inicialmente, a coluna foi conectada à bomba do cromatógrafo a líquido para ser condicionada, sem conectar a sua extremidade de saída ao detector, para evitar que partículas residuais decorrentes do processo de enchimento da coluna migrem pela linha de fluxo. Para o uso da coluna Luna<sup>®</sup> C18, foram usados pelo menos 20 volumes de coluna com uma fase móvel de metanol:H<sub>2</sub>O 80:20 (v/v), durante 2h, em uma vazão de 0,5 mL.min<sup>-1</sup>. Este condicionamento é necessário para que ocorra um equilíbrio entre a FE e a FM, evitando assim a não repetibilidade dos tempos de retenção dos compostos eluídos.<sup>88</sup> Após o período de condicionamento da coluna, esta foi conectada ao detector UV/VIS, iniciando-se a avaliação cromatográfica.

### 3.1.5 Preparo das Fases Móveis (FM)

Fase Móvel A (Solução de acetato de amônio 0,02 mol.L<sup>-1</sup>): O preparo consistiu na pesagem de massa referente a 10 vezes a concentração desejada (solução estoque 0,20 mol.L<sup>-1</sup>), solubilização e volume completado com água ultrapura. A solução estoque foi armazenada em geladeira. Para uso diário, uma solução de 0,02 mol.L<sup>-1</sup> era preparada a partir da solução concentrada. Essa solução

era filtrada à vácuo em filtro de membrana de nylon Millipore e utilizada imediatamente para o preparo da FM.

Fase Móvel B (ACN:MeOH 80:20, (v/v)): A FM foi preparada medindo-se os volumes de metanol e acetonitrila separadamente, por meio do uso de uma proveta. Então, a solução foi filtrada à vácuo em filtro de membrana de nylon Millipore e colocada em recipiente de vidro. A FM foi preparada diariamente durante os experimentos e quando necessário foi guardada em geladeira.

### **3.1.6 Padrões**

Foram preparadas soluções padrão estoque de TA e AC na concentração de 500 mg.L<sup>-1</sup>. Para o azul patente a solução estoque foi de 125 mg.L<sup>-1</sup>. Os padrões foram diluídos em água ultrapura Milli-Q (18 MΩ). Alíquotas destas soluções padrão foram pipetadas em volumes previamente calculados e transferidas diretamente para vials de 1,5 mL com micropipeta calibrada, totalizando um volume no vial de 1,0 mL. Para a quantificação das amostras, foram feitas soluções padrão nas seguintes concentrações: 0,5; 1,0; 2,0; 5,0; 10,0; 20,0; 30,0; 40,0; 50,0 e 100 mg.L<sup>-1</sup>. O padrão interno azul patente V foi adicionado às soluções de padrões e amostras em uma concentração fixa de 20 mg.L<sup>-1</sup>. Esse valor foi estabelecido como um ponto intermediário na curva analítica.

### **3.1.7 Amostras**

Foram analisadas no total 123 amostras de sucos artificiais, refrigerantes e isotônicos, de diferentes marcas, cores e sabores (ANEXO A). As amostras foram adquiridas majoritariamente no comércio da região metropolitana de Belo Horizonte, mas também em outros estados brasileiros. O preparo das amostras consistiu na degaseificação por ultrassom, filtração das mesmas em filtro PTFE hidrofílico 13 mm, com tamanho de poro 0,45 µm (marca Analítica), e dopagem com padrão

interno azul patente. Alíquotas das amostras foram pipetadas e transferidas diretamente para vials de 1,5 mL, totalizando um volume no vial de 1,0 mL. O padrão interno foi adicionado às amostras com micropipeta calibrada, em um volume previamente calculado correspondente à concentração de 20 mg.L<sup>-1</sup>. O passo seguinte foi transferir os vials para o injetor automático e registrar o *batch* das análises no *software*.

### 3.1.8 Otimização do método

Inicialmente, o desenvolvimento deste método se baseou numa metodologia publicada em uma nota técnica de um fabricante de equipamentos cromatográficos (Perkin Elmer)<sup>89</sup>. Esta referência utilizou UHPLC para a determinação de oito corantes artificiais em refrigerantes, incluindo AC, mas não TA. Em um trabalho anterior em nosso grupo de pesquisa, esta metodologia já havia sido adaptada para a determinação de AC em bebidas não alcoólicas usando HPLC.<sup>35</sup> No entanto, a adaptação do método para a determinação simultânea de AC e TA exigiu um esforço muito maior, tanto do ponto de vista experimental quanto intelectual, o que justificou uma maior ênfase para a otimização deste método cromatográfico nesta dissertação.

Neste trabalho, para todas as determinações foi usada a metodologia de separação por CLAE-FR, modo gradiente, com detecção espectrofotométrica nos comprimentos de onda de maior absorção, 428 e 483 nm para TA e AC, respectivamente (Figuras 11 e 12). Nestas condições,<sup>89</sup> a corrida cromatográfica foi iniciada com 95% de uma solução de acetato de amônio 0,02 mol.L<sup>-1</sup> (pH=6,7) e 5% de uma Fase B composta de acetonitrila:metanol 80:20 (v/v) (pH=8,0). Foi feito um teste inicial com padrões e amostras. Como um primeiro resultado, verificou-se que a TA praticamente não foi retida na coluna cromatográfica, eluindo com um tempo de retenção  $t_R=0,66$  min (Figura 11).

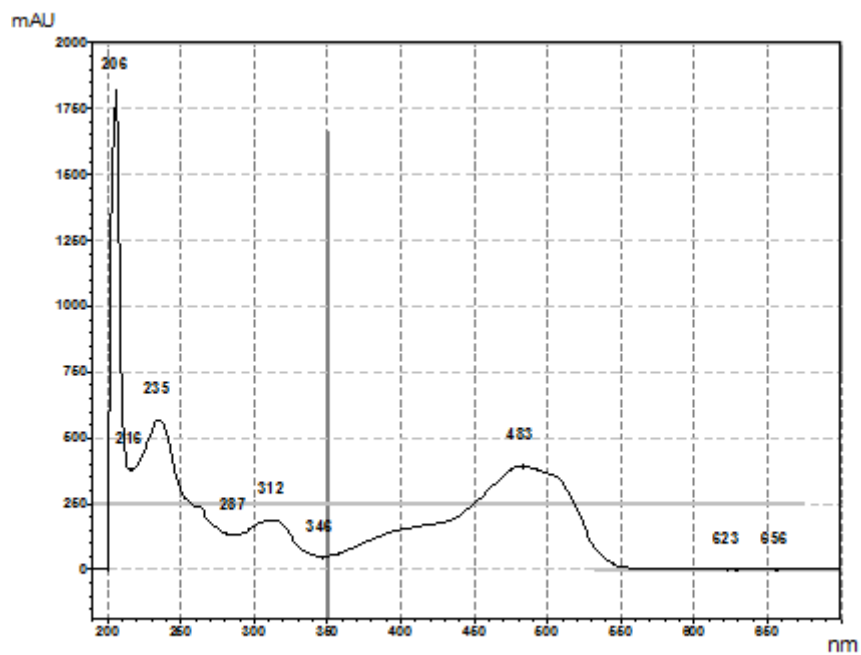


Figura 11. Espectro de absorção molecular UV-Vis do padrão amarelo crepúsculo 30 ppm

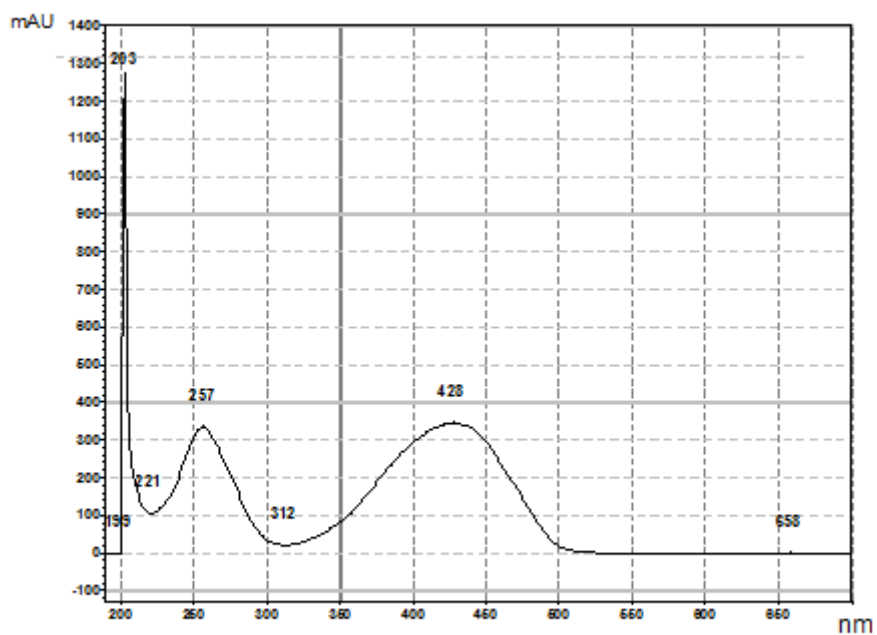


Figura 12. Espectro de absorção molecular na região UV-Vis do padrão tartrazina 30 ppm

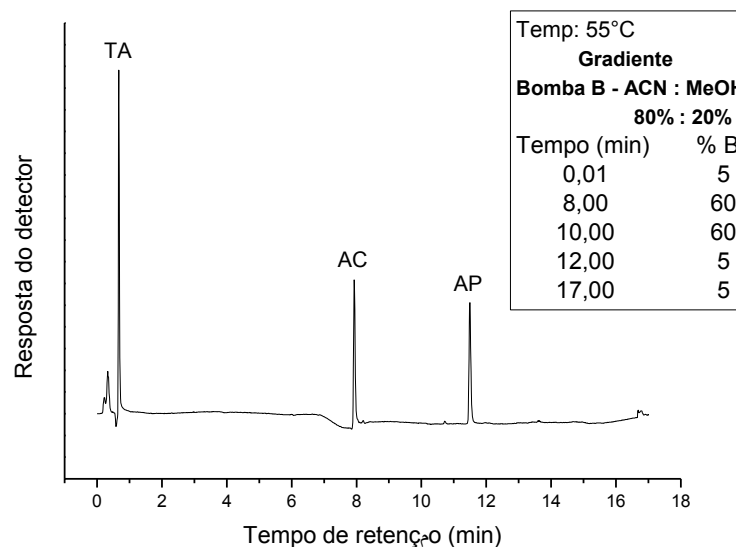


Figura 13. Cromatograma (FR): FM: acetato de amônio. 20 mmol L<sup>-1</sup> (A) / ACN:MeOH 80:20 (B); vazão: 1,2 mL.min<sup>-1</sup>. Volume de injeção: 20 µL. Detecção: UV (428 nm); Padrão: 1-TA-(5,0mg.L<sup>-1</sup>) t<sub>R</sub>=0,7min ; 2-AC-(5,0mg.L<sup>-1</sup>) t<sub>R</sub>=7,9 min; 3-AP (20 mg.L<sup>-1</sup>) t<sub>R</sub>=11,5min.

A partir deste resultado, foram feitas alterações na metodologia variando-se alguns parâmetros em função dos materiais comercialmente disponíveis e do sistema cromatográfico utilizado para os experimentos. A temperatura do forno de coluna (33, 40 e 55 °C), assim como a proporção inicial no gradiente da FM B, ACN:MeOH, 80:20 (v/v) (1, 2, 3, 4 e 5%) e a vazão (1,0 e 1,2 mL.min<sup>-1</sup>), foram parâmetros avaliados na separação cromatográfica dos corantes. O tempo de condicionamento da coluna para o restabelecimento do equilíbrio da mesma entre uma injeção e outra, ou o tempo para iniciar um novo gradiente, foi aumentado de 3 minutos nas condições iniciais, para 7 e 10 minutos nos testes subsequentes. O volume de injeção dos padrões e amostras (20 µL) e a composição da FM não foram mudados ao longo de todas as determinações.

A força cromatográfica da FM adequada é selecionada através da análise do fator de retenção k. Um valor de k baixo significa pouca interação do soluto com a fase estacionária, que é o recheio da coluna cromatográfica. Este efeito foi evidenciado no cromatograma da Figura 11. Em CLAE-FR, na qual a fase estacionária é mais apolar que a fase móvel, solventes fortes são os mais apolares e fracos são os mais polares. Portanto, fases móveis fortes em CLAE-FR possuem maior proporção do solvente orgânico.<sup>36</sup> A diminuição da proporção da FM B, solvente mais forte no início do gradiente, significa uma redução na força

cromatográfica da FM, permitindo uma retenção maior dos compostos eluídos. Isto é evidenciado no cromatograma da Figura 12. A tartrazina, cujo pico inicialmente estava bem próximo ao tempo morto, foi deslocada para um tempo de retenção de 2,5 min.

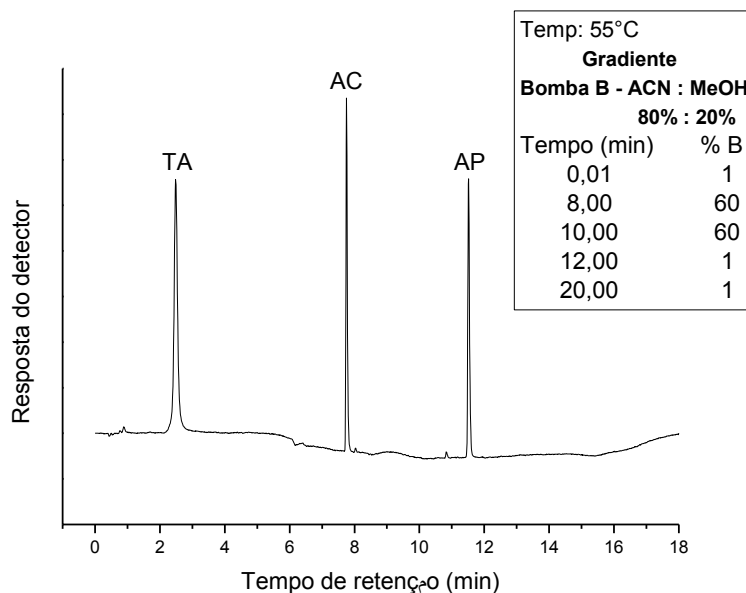


Figura 14. Cromatograma (FR): FM: acetato de amônio. 20 mmol L<sup>-1</sup>(A) / ACN:MeOH 80:20 (B); vazão: 1,2 mL.min<sup>-1</sup>. Volume de injeção: 20 µL . Detecção: UV (428 nm); Padrão: 1 – TA (5,0mg.L<sup>-1</sup>) - t<sub>R</sub>=2,5min; 2 – AC (5,0mg.L<sup>-1</sup>)- t<sub>R</sub>=7,8min ;3 – AP (20 mg.L<sup>-1</sup>)- t<sub>R</sub>=11,5min.

Para os padrões, esta condição de separação foi ótima, entretanto, em relação às amostras (Figura 13) observaram-se impurezas que co-eluíram com o corante tartrazina. Essas interferências foram visualizadas utilizando-se um recurso do *software* LC-solution Shimadzu. A falta de seletividade sugere que à temperatura de 55 °C alguma espécie presente na amostra analisada coeluiu com o corante.

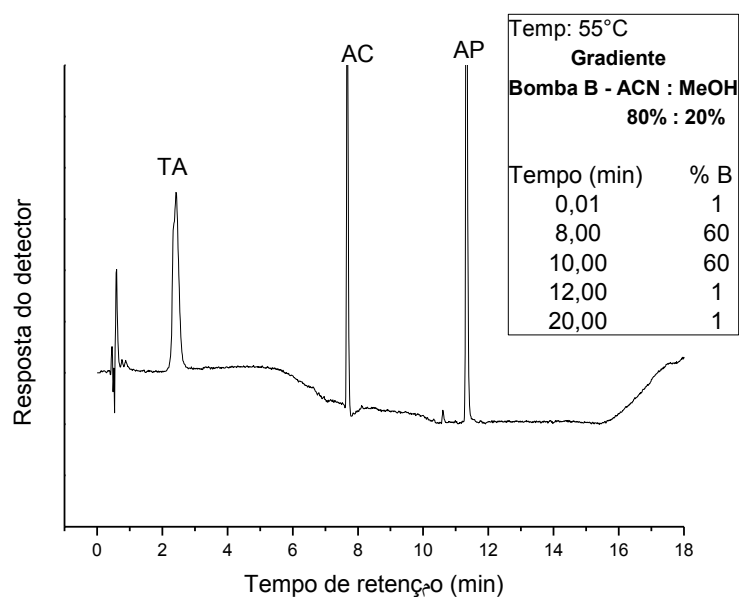


Figura 15. Cromatograma (FR): FM: acetato de amônio. 20mmol L<sup>-1</sup> (A)/ ACN:MeOH 80:20 (B); vazão: 1,2 mL.min<sup>-1</sup>. Volume de injeção: 20 µL. Detecção: UV (428 nm); Amostra 8: refrigerante de maracujá; 1-TA- $t_R$ =2,4min; 2-AC- $t_R$ =7,7min; 3-AP (20 mg.L<sup>-1</sup>)- $t_R$ =11,3min.

Foram feitos também testes à temperatura ambiente e os resultados obtidos não foram confiáveis, porque os tempos de retenção dos corantes TA, AC e AP para os padrões e amostras não foram reproduzíveis (Figuras 14-16). Além disso, houve alargamento do pico do corante TA. Há diversas causas de alargamento dos picos considerando processos de transporte de massa que alteram a velocidade das moléculas de uma mesma espécie dentro da coluna, diminuindo a eficiência de separação do analito.<sup>90</sup>

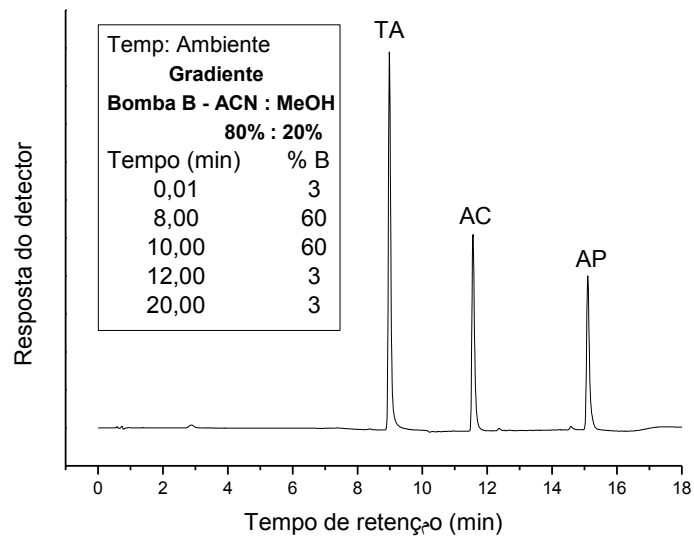


Figura 16. Cromatograma em (FR): FM: acetato de amônio. 20mmol L<sup>-1</sup> (A)/ ACN:MeOH 80:20 (B); vazão: 1,2 mL.min<sup>-1</sup>. Volume de injeção: 20 µL . Detecção: UV (428 nm); Padrão: 1-TA (5,0 mg.L<sup>-1</sup>) t<sub>R</sub>=9,0 min; 2-AC- (5,0mg.L<sup>-1</sup>) t<sub>R</sub>=11,6 min; 3-AP(20 mg.L<sup>-1</sup>) t<sub>R</sub>=15,1 min.

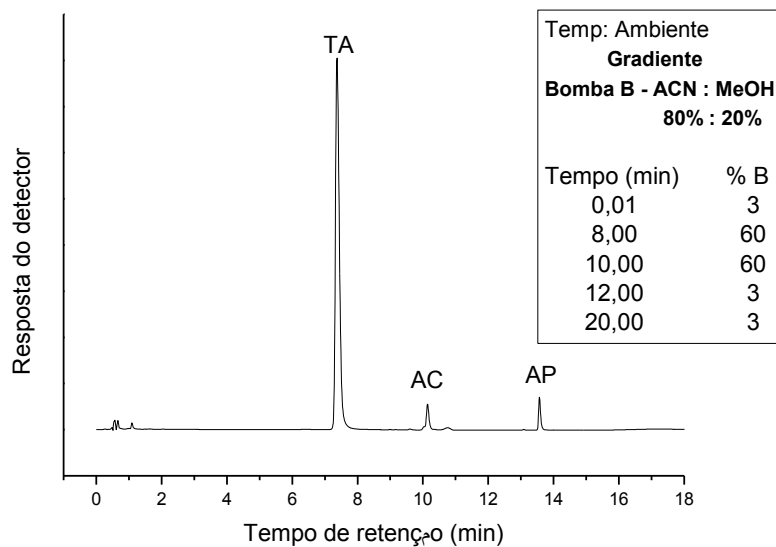


Figura 17. Cromatograma em (FR). FM: acetato de amônio 20 mmol L<sup>-1</sup>(A) / ACN:MeOH 80:20 (B); vazão:1,2 mL.min<sup>-1</sup>. Volume de injeção: 20 µL . Detecção UV (428 nm); Amostra 4: bebida de frutas cítricas. 1-TA-t<sub>R</sub>=7,4 min; 2-AC-t<sub>R</sub>=10,1 min; 3-AP-(20 mg.L<sup>-1</sup>)-t<sub>R</sub>=13,6min

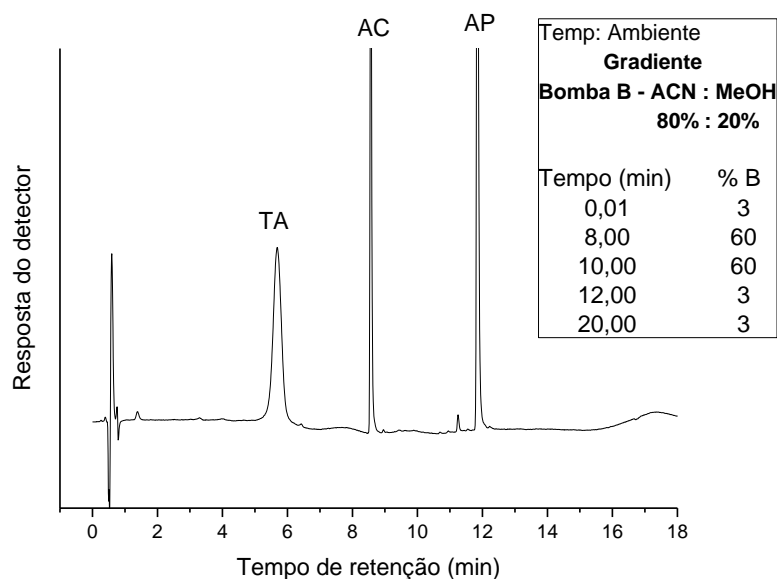


Figura 18. Cromatograma em (FR): FM: acetato de amônio. 20mmol L-1(A) / ACN:MeOH 80:20 (B); vazão: 1,2 mL.min-1. Volume de injeção: 20 $\mu$ L . Detecção: UV (428 nm); Amostra 8: refrigerante de maracujá; 1-TA- $t_R$ =5,7 min; 2-AC- $t_R$ =8,6 min; 3-AP-(20 mg.L-1)  $t_R$ =11,9 min

Este efeito também pôde ser observado no cromatograma da amostra 4 (Figura 17) em um outro experimento, cujo gradiente foi iniciado com 4% da fase móvel B, mantendo-se as mesmas condições em relação aos outros parâmetros.

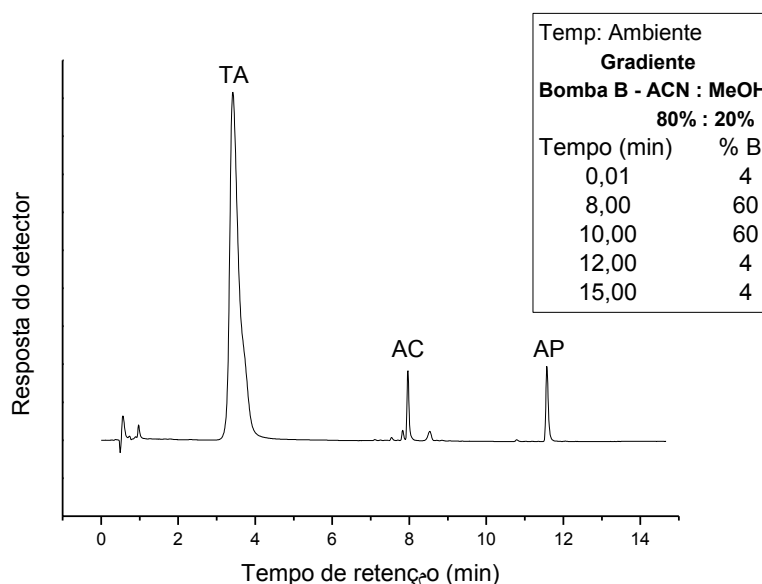


Figura 19. Cromatograma em (FR). FM:acetato de amônio. 20mmol L-1 (A) / ACN:MeOH 80:20(B); vazão:1,2 mL.min-1. Volume de injeção: 20  $\mu$ L . Detecção: UV (428 nm); Amostra 4: bebida de frutas cítricas.1-TA- $t_R$ =3,4 min; 2-AC- $t_R$ =8,0 min; 3-AP(20 mg.L-1)- $t_R$ =11,6 min.

Várias separações foram também realizadas nas temperaturas de 33°C e 40°C, modificando-se a proporção inicial do gradiente da FM B, ACN:MeOH, 80:20 (v/v). Outras pequenas variações da proporção da FM B no início do gradiente não provocaram mudanças significativas na separação dos três corantes, TA, AC e padrão interno. Nestas duas últimas temperaturas (33°C e 40°C) não foram observadas diferenças significativas nas áreas dos picos para os corantes estudados.

A melhor vazão para a obtenção da separação cromatográfica foi de 1,2 mL.min<sup>-1</sup>. A alteração da proporção inicial da fase móvel B foi importante para uma maior retenção da TA, quando comparada às condições iniciais.<sup>89</sup>

Desta forma, definiu-se a metodologia a ser validada, conforme as condições descritas no cromatograma da Figura 18, escolhendo-se a menor temperatura na qual houve estabilização do forno da coluna e gradiente com proporção inicial de fase B igual a 1%.. Nestas condições, a separação de padrões e amostras foi eficiente, com boa resolução e simetria de todos picos (Figuras 18 e 19).

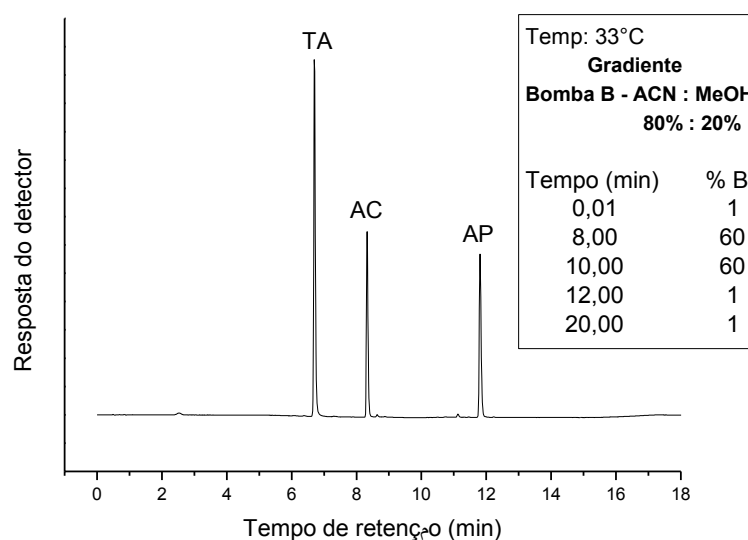


Figura 20. Cromatograma em (FR): FM: acetato de amônio. 20mmol L<sup>-1</sup> (A) / ACN:MeOH 80:20 (B); vazão: 1,2 mL.min<sup>-1</sup>. Volume de injeção: 20 µL. Detecção: UV (428 nm); Padrão: 1-TA-(5,0mg.L<sup>-1</sup>)-t<sub>R</sub>=6,7 min; 2-AC-(5,0mg.L<sup>-1</sup>)-t<sub>R</sub>=8,3 min; 3-AP-(20 mg.L<sup>-1</sup>)-t<sub>R</sub>=11,8 min

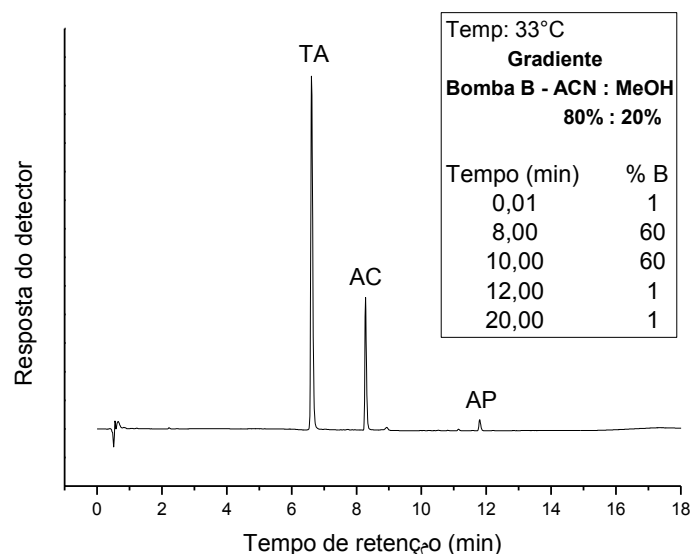


Figura 21. Cromatograma em (FR). FM: acetato de amônio 20mmol L<sup>-1</sup> (A) / ACN:MeOH 80:20 (B); vazão:1,2 mL.min<sup>-1</sup>. Volume de injeção: 20 $\mu$ L . Detecção UV (428 nm); Amostra 4: bebida de frutas cítricas. 1-TA- $t_R$ =6,6 min; 2-AC- $t_R$ =8,3 min; 3-AP-(20 mg.L<sup>-1</sup>)  $t_R$ =11,7 min

Para a quantificação das amostras, foram construídas as curvas analíticas com os padrões dos corantes em estudo e do padrão interno na faixa de 0,1 mg.L<sup>-1</sup> a 100,0 mg.L<sup>-1</sup>. Definidas as melhores condições cromatográficas para a separação dos corantes de interesse, as 123 amostras de sucos, refrigerantes e isotônicos, de diferentes marcas, cores e sabores, foram analisadas e quantificadas.

Para escolha do método de quantificação adequado foram testadas: padronização externa e padronização interna. O método de padronização externa compara a área da substância a ser quantificada na amostra com as áreas obtidas a partir de soluções padrão e concentrações conhecidas. O método de padronização interna corrige variações no sinal analítico devido a mudanças nas condições de análise.<sup>91</sup>

Na Tabela 1 são apresentadas as condições otimizadas do método.

Tabela 1. Condições do método cromatográfico otimizado

Sistema	Cromatógrafo a líquido com detector de arranjo de diodos (DAD), Shimadzu®
Coluna	Luna, C18, 5 µm, fase reversa, (100 x 3.00mm) Phenomenex®
Modo de eluição	Gradiente
Fase móvel	A: solução aquosa de acetato de amônio (CH <sub>3</sub> COONH <sub>4</sub> ) 20 mmol.L <sup>-1</sup> B: ACN : MeOH 80:20
Tempo(min)	% B
0-8	1-60
8-10	60
10-12	60-1
12-20	1
vazão	1.2 mL.min <sup>-1</sup>
temperatura	33°C
Volume de injeção	20 µL (autoinjeter)
Tempo da análise	12 minutos

Nessas condições uma separação eficiente dos corantes foi obtida e os parâmetros cromatográficos calculados foram: tempo morto (igual a 0,5 min), fator de retenção para AC igual a 15,6 e para TA igual a 12,3, fator de assimetria igual a 1,5 para ambos os corantes, resolução de pico maior que 16 e número de pratos teóricos de aproximadamente 6111 e 3864 para AC e TA respectivamente.

### 3.1.9 Validação do método<sup>91-94</sup>

Visando assegurar a confiabilidade dos resultados, o método desenvolvido foi devidamente validado, através da estimativa das figuras de mérito seletividade, linearidade, precisão e exatidão,<sup>91,92</sup> conforme preconizados pelas diretrizes do Ministério da Agricultura, Pecuária e Abastecimento<sup>93</sup> e pelo documento do INMETRO DOQ-CGCRE-008.<sup>94</sup> Neste trabalho as substâncias estudadas TA e AC, estão presentes nas amostras a níveis de traços, abaixo de 0,01%.<sup>94</sup> Isso deve

ser levado em consideração ao consultar os guias de validação em relação aos limites máximos para as figuras de mérito.

Os experimentos para validação do método foram realizados após a definição das melhores condições cromatográficas de análise, as quais foram mostradas na Tabela 1.

A seletividade do método foi determinada pela completa separação dos picos correspondentes aos corantes e análise dos seguintes parâmetros cromatográficos: resolução e fator de retenção. Através do uso de um detector de arranjo de diodos (DAD), foi possível comparar o espectro eletrônico do pico obtido na separação dos corantes nas amostras com o obtido na leitura do padrão, observando-se que não houve interferentes co-eluído com os analitos de interesse.

A linearidade de um método analítico é a capacidade deste em demonstrar que os resultados obtidos são diretamente proporcionais à concentração do analito na amostra, dentro de um intervalo especificado.

Julga-se satisfatória a linearidade do gráfico quando o coeficiente de correlação (a raiz quadrada do coeficiente de determinação) da reta obtida não é estatisticamente diferente da unidade. No caso, considera-se.<sup>93,94</sup>

- R = 1 Correlação perfeita
- 0,91 < R < 0,99 Correlação fortíssima
- 0,61 < R < 0,91 Correlação forte
- 0,31 < R < 0,60 Correlação média
- 0,01 < R < 0,30 Correlação fraca
- R = zero Correlação nula

Como o coeficiente de correlação por si só não garante a adequação do modelo, a linearidade pode ser avaliada ainda por meio de testes estatísticos que confirmem a aleatoriedade dos resíduos.<sup>83</sup>

A precisão do método pode ser avaliada nos níveis de precisão instrumental, repetibilidade e precisão intermediária (três dias diferentes).<sup>93,94</sup> Esses parâmetros são estimados através do cálculo de desvio padrão relativo (DPR) conforme equação 21.

$$\text{DPR (\%)} = s/x * 100 \quad (21)$$

Onde s é a estimativa do desvio padrão absoluto, e x é a média das medidas.

A veracidade de um método analítico é a proximidade dos resultados obtidos pelo método em estudo em relação ao valor verdadeiro. Ela é expressa pela relação entre a concentração média determinada experimentalmente e a concentração teórica correspondente, de acordo com a equação 22.

$$\text{Veracidade} = \frac{\text{Concentração média experimental}}{\text{Concentração teórica}} \times 100 \quad (22)$$

O limite de detecção (LD) representa a menor concentração da substância em exame que pode ser detectada, mas não necessariamente quantificada, utilizando um determinado procedimento experimental. O limite de quantificação inferior (LIQ) representa a menor concentração da substância em exame que pode ser medida, utilizando um determinado procedimento experimental.<sup>91</sup>

Neste trabalho, o método considerado para o cálculo desses parâmetros foi a relação entre a estimativa do desvio padrão da resposta do coeficiente linear da equação e a inclinação da curva analítica (S), em níveis próximos ao LIQ, segundo as equações 23 e 24.<sup>91</sup>

$$\text{LD} = 3 \frac{\text{DP } b}{a} \quad (23)$$

$$\text{LIQ} = 10 \frac{\text{DP } b}{a} \quad (24)$$

### 3.2 Resultados e Discussão

A preparação da curva de calibração para estudo de linearidade foi realizada a partir de 14 concentrações diferentes, em triplicata, variando-se de 0,1 a 100 mg.L<sup>-1</sup>, seguida do tratamento estatístico dos dados para a determinação da equação da reta e do coeficiente de determinação R<sup>2</sup> (Figuras 20 e 21). Este último parâmetro permite uma estimativa da qualidade da curva obtida, uma vez que, quanto mais próximo de 1,0, menor a dispersão do conjunto de pontos experimentais e menor a incerteza dos coeficientes de regressão estimados.<sup>93,94</sup>

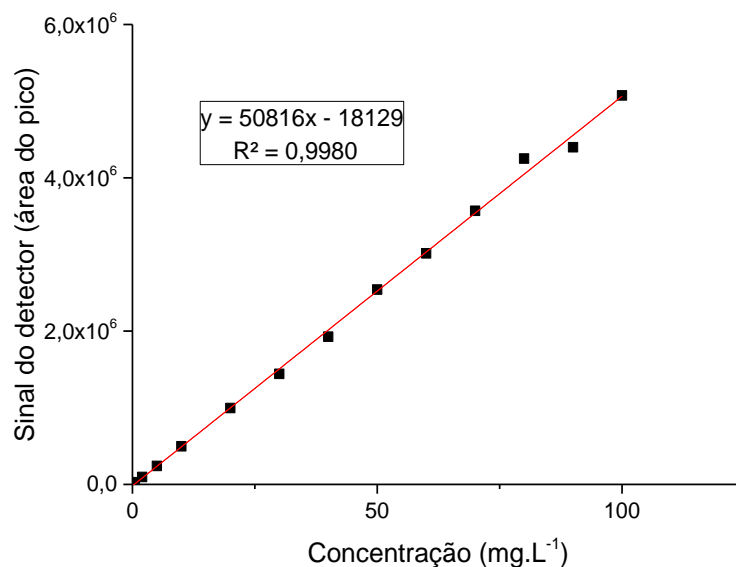


Figura 22. Curva analítica referente ao AC, equação da reta e R<sup>2</sup>

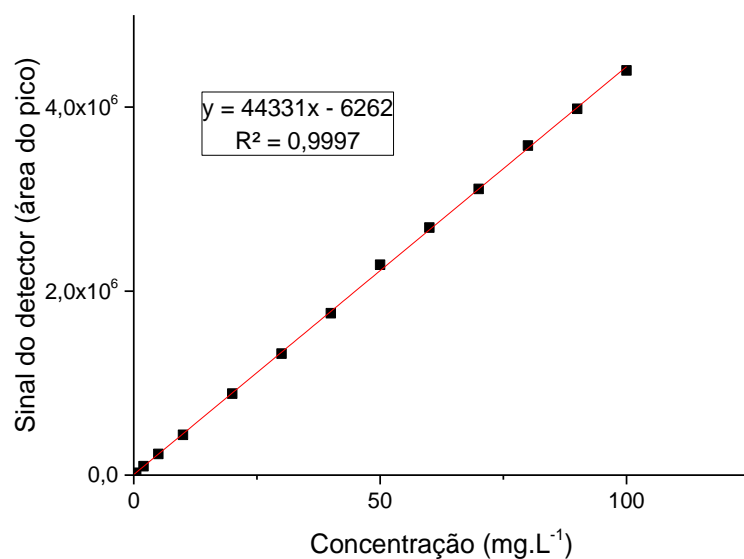


Figura 23. Curva analítica referente ao TA, equação da reta e  $R^2$

O método mostrou grande sensibilidade, pois os coeficientes angulares das retas apresentaram valores elevados. Isso significa que com pequenas variações na concentração obtêm-se grandes variações nos sinais medidos, garantindo a diferenciação entre duas concentrações bastante próximas.

A linearidade foi também avaliada quanto a homocedasticidade, aleatoriedade e a não correlação dos resíduos inicialmente com 42 pontos no total (14 concentrações diferentes em triplicata). Os *outliers* foram detectados com 95% de confiança utilizando-se o teste de resíduos *Jackknife*.<sup>83</sup> Os pontos anômalos podem ser removidos até um limite de 22% do número total de pontos.<sup>93</sup> Neste teste foram excluídos 12% e 17% para TA e AC, respectivamente. Testes estatísticos foram aplicados em relação à normalidade (Ryan-Joiner), ausência de correlação dos resíduos (Durbin-Watson), ambos a 95% de confiança e homocedasticidade (Brown-Forsythe a 95% para TA e 99% para AC).<sup>83</sup> O estudo de linearidade realizado para o intervalo de concentração considerado mostrou uma boa relação linear com o sinal analítico (resposta do detector).

Na Tabela 2 são mostrados os coeficientes de determinação das retas e os respectivos resultados encontrados para os testes estatísticos.<sup>83</sup>

Tabela 2. Estudo de linearidade para o método cromatográfico

TESTES	TA	AC
R <sup>2</sup>	0,999	0,998
*Durbin-Watson	2,45	1,81
**Teste de Brown Forsythe	1,27	2,57
***Ryan-Joiner	0,980	0,972

\*Valores entre os limites inferior e superior de 1,53 a 2,47 para TA e 1,52 a 2,48 para AC

\*\* Valores de t calculado < t tabelado de 2,03 a 95% de confiança para TA e 2,73 a 99% de confiança para AC.

\*\*\*Valor de R<sub>eq</sub> calculado maior que R<sub>crítico</sub> a 95% de confiança (0,968 para AC e 0,970 para TA).

A aplicação dos testes estatísticos e os resultados obtidos referentes à faixa estudada sugerem que os resíduos seguem a normalidade, são independentes e não possuem autocorrelação. A faixa linear então considerada para os dois corantes foi de 0,1 a 100 mg.L<sup>-1</sup>.

Nos testes de precisão instrumental e repetibilidade (de uma mesma preparação), uma nova amostra de isotônico sabor maracujá que continha os dois corantes (AC e TA) foi utilizada. A precisão instrumental foi calculada através de dez injeções consecutivas de uma única preparação e apresentou um DPR de 1,16% para a TA e de 2,05% para o AC. O teste de repetibilidade realizado com 12 preparações da mesma amostra teve resultados de DPR, para TA e AC, de 1,65 e 1,88%, respectivamente. A repetibilidade foi também avaliada considerando diferentes níveis de concentração através de outras cinco diferentes amostras reais e os resultados são apresentados nas Tabelas 3 e 4.

Tabela 3. Avaliação da repetibilidade para a determinação de TA nas amostras usando o método cromatográfico

Amostras reais	Média das concentrações (mg.L <sup>-1</sup> )	DPR(%)
Isotônico de tangerina	0,94	2,69
Isotônico de maracujá	4,18	0,63
Isotônico de tangerina	12,94	3,29
Suco de fruta adoçada tangerina	0,93	1,34
Refrigerante Maracujá	2,76	7,32

Tabela 4. Avaliação da repetibilidade para a determinação de AC nas amostras usando o método cromatográfico

Amostras reais	Média das concentrações (mg.L <sup>-1</sup> )	DPR (%)
Isotônico de tangerina	10,38	2,47
Isotônico de maracujá	2,90	0,32
Isotônico de tangerina	11,77	1,45
Suco de fruta adoçada tangerina	10,57	0,52
Refrigerante Maracujá	3,68	3,86

A precisão intermediária foi avaliada com amostras reais contendo os dois corantes simultaneamente. Para a TA a faixa de concentração de 0,7 a 13 mg.L<sup>-1</sup> foi avaliada e o DPR foi estimado entre 1,35 a 5,59%. Para o AC a faixa de concentração de 2,8 a 40 mg.L<sup>-1</sup> foi avaliada e o coeficiente de variação ficou entre 0,34 a 3,97%.

Os testes de precisão instrumental, repetibilidade e precisão intermediária estão de acordo com o máximo de 13% de DPR (2/3 de 20%) estabelecido no item II.6.5 das diretrizes do Ministério da Agricultura, Pecuária e Abastecimento.<sup>93</sup>

A veracidade do método foi avaliada por meio de ensaio de recuperação, devido à ausência de materiais de referência certificados. Para tanto, uma amostra de suco de frutas cítricas foi dopada com os padrões nas concentrações 2,5; 5,0; 10,0; 15,0; 20,0; 25,0 e 100,0 mg.L<sup>-1</sup> de TA e AC. A amostra sem adição do padrão e cada uma das amostras com o padrão adicionado foram analisadas e as quantidades medidas relacionadas com a quantidade adicionada. Os resultados de veracidade do método para TA (Tabela 5) e AC (Tabela 6) foram aceitáveis (80-110%)<sup>93,94</sup> e são apresentados a seguir.

Tabela 5. Adições de padrões à amostra de suco para estudo de veracidade do método cromatográfico para TA

Adições (mg.L <sup>-1</sup> )	Concentração encontrada mg.L <sup>-1</sup>	Concentração esperada	Veracidade
0,0	25,25	-	-
2,5	28,13	27,75	101,39
5,0	29,95	30,25	99,02
10,0	37,25	35,25	105,67
15,0	40,20	40,25	99,88
20,0	45,20	45,25	99,89
25,0	51,28	50,25	102,05
100,0	128,12	125,25	102,29

Tabela 6. Adições de padrões à amostra de suco para estudo de veracidade do método cromatográfico para AC

Adições (mg.L <sup>-1</sup> )	Concentração encontrada mg.L <sup>-1</sup>	Concentração esperada	Veracidade
0,0	10,12	-	-
2,5	12,62	12,62	99,98
5,0	14,69	15,12	97,16
10,0	20,36	20,12	101,23
15,0	24,15	25,12	96,15
20,0	29,10	30,12	96,63
25,0	34,23	35,12	97,47
100,0	103,51	110,12	94,00

No estudo de LD e LQ uma amostra isenta de corantes artificiais (isotônico de limão) foi dopada e utilizada para estudo. Adições de padrões dos corantes AC e TA na concentração de 0,05 a 5 ppm foram realizadas a fim de que uma curva pudesse ser analisada e os limites de detecção e quantificação pudessem ser calculados. Os valores de LD e LIQ calculados para TA foram 0,02 e 0,07 mg.L<sup>-1</sup> e para AC 0,05 e 0,14 mg.L<sup>-1</sup>, respectivamente.

O estudo de robustez realizado foi simplificado determinando as faixas de variação de alguns fatores considerados.<sup>93</sup> A robustez foi avaliada em relação ao fluxo, gradiente, e temperatura. Variações em torno de 2% em relação aos valores do método a ser validado foram realizadas. Os maiores DPR (%) foram observados para a mudança em relação à temperatura. Para o corante TA, os DPR (%) encontrados para estudo de robustez em relação ao fluxo, gradiente e temperatura

foram, respectivamente, 0,62, 0,43 e 2,48 %. Para o corante AC, os DPR (%) encontrados em relação ao fluxo, gradiente e temperatura foram, respectivamente, 0,87; 1,64 e 2,19 %. Desta maneira, pode-se observar que os valores se encontram abaixo do limite de 13% para a faixa de concentração considerada. Sendo assim, o método foi considerado robusto para a determinação dos dois corantes em relação as alterações nos fatores (fluxo, gradiente, e temperatura).

### **3.3 Conclusão Parcial**

Um método CLAE para a determinação de dois corantes artificiais em bebidas não alcoólicas foi desenvolvido e apresentou boa seletividade, resolução (>16) e simetria de picos ( $A_s=1,5$ ). Os tempos de retenção e comprimentos de onda máximos para a TA e AC foram 6,6 min e 428nm, e 8,2 min e 483 nm, respectivamente. Todas as 123 amostras foram determinadas e os valores de concentrações obtidos foram usados como referência para o desenvolvimento do método quimiométrico, o qual será discutido no próximo capítulo. O método desenvolvido foi submetido à validação analítica e está de acordo com os guias normativos,<sup>93,94</sup> demonstrando que o método é adequado para quantificação simultânea de AC e TA em refrigerantes, sucos artificiais e isotônicos. Esse método pode, portanto, ser utilizado para a quantificação simultânea de AC e TA na matriz estudada.

## **4. Análise de imagens**

### **4.1 Parte experimental**

#### **4.1.1 Equipamentos e softwares**

As imagens foram obtidas utilizando um escâner de mesa comercial CanoScan LiDE 110 (Tóquio, Japão). Os dados foram tratados usando o *software* MATLAB, versão 7.13 (MathWorks, Natick, EUA) juntamente com o pacote PLS Toolbox versão 6.7 (Eigenvector Technologies, Manson, EUA). Para o processamento das imagens foi usado o pacote Image Processing Toolbox, versão 8.0 (MathWorks). Para a detecção de amostras anômalas utilizou-se uma rotina desenvolvida no Matlab por Jez W. B. Braga (IQ-UnB).<sup>95</sup>

#### **4.1.2 Amostras**

As mesmas 123 amostras mencionadas na seção 3.1.7 e no Anexo A foram analisadas para a construção do modelo quimiométrico. Elas foram armazenadas em geladeira a 4°C até o momento das análises. Após abertas, tanto a aquisição das imagens quanto as análises cromatográficas das amostras foram executadas dentro de no máximo dois dias.

#### **4.1.3 Procedimento de análise**

As amostras foram deixadas em descanso sobre a bancada do laboratório, para atingirem o equilíbrio térmico antes do início dos procedimentos. Cada amostra foi coletada em um béquer de 50 mL e levada ao banho de ultrassom,

para desgaseificação, por 5 minutos. Após esta etapa, 5 mL foram utilizados para a aquisição das imagens.

A aquisição das imagens foi realizada utilizando placas de Petri pequenas (5,0 cm de diâmetro x 1,5 cm de altura), nas quais as amostras a serem lidas eram vertidas, e as placas posicionadas no canto inferior esquerdo do escâner. Um anteparo de madeira preto foi utilizado para evitar a entrada de luz externa (Figuras 23 e 24).

Todas as imagens foram digitalizadas no sistema RGB 24-bit, com 16,8 milhões de cores e resolução de 300 ppp (pixels por polegadas), no formato “.tif”. A conversão das imagens em histogramas RGB foi realizada em ambiente MatLab. Primeiramente, uma área de 100x100 pixels de cada imagem foi selecionada de uma área central da placa de Petri. A área selecionada foi tratada com um filtro digital (*unsharp*) para redução de ruído, e então decomposta em histogramas RGB. Após todos os tratamentos, um histograma contendo a frequência de 768 canais (256 para cada cor RGB) foi obtido para cada amostra. Cada amostra foi escaneada três vezes, e a média dos histogramas de cada amostra foi utilizada para a construção do modelo PLS.



Figura 24. Escâner CanoScan LiDE 110 com o anteparo

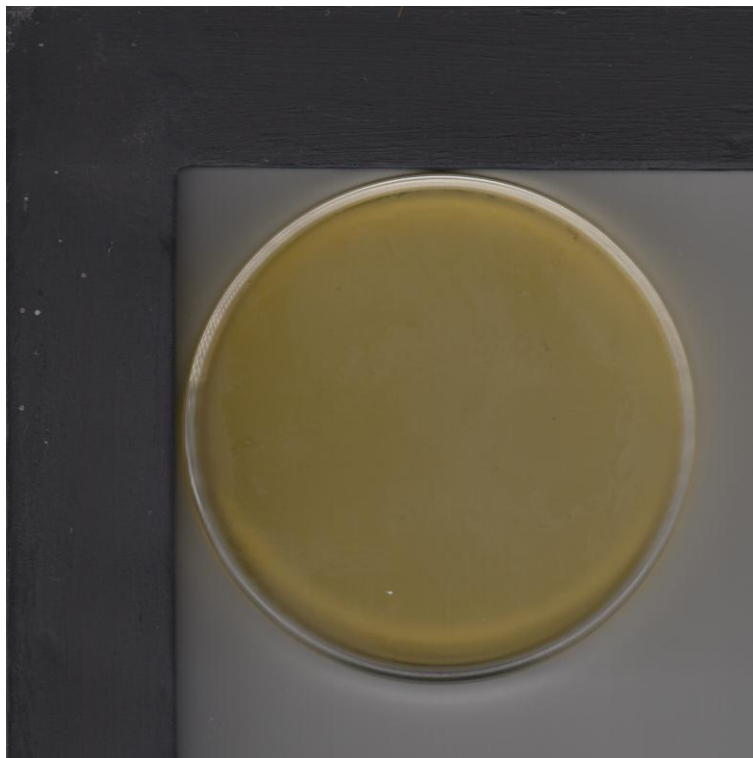


Figura 25. Imagem adquirida de uma amostra de isotônico sabor laranja marca Power ADE

## **4.2 Resultados e Discussão**

### **4.2.1 Construção dos modelos**

Dentre as 123 amostras analisadas há grande diversidade de marcas, sabores (laranja, cítrico, tangerina, limão, maracujá, etc.) e composição. Todas as amostras continham somente AC e/ou TA como corantes artificiais mencionados em suas formulações.

Foram construídos dois modelos de calibração multivariada usando o método PLS1 (uma variável dependente prevista de cada vez), um para cada analito/corante. Na construção do modelo para a quantificação de AC foram utilizadas as amostras que continham somente este corante (40) mais as que continham também TA (55), totalizando 95. Na construção do modelo para a quantificação de TA foram usadas as amostras que continham somente este corante (27) mais as que continham também AC (55), totalizando 82. Os modelos foram

construídos utilizando como variáveis independentes as freqüências dos canais de cor dos histogramas RGB obtidos das imagens digitais de refrigerantes, isotônicos e sucos e, como variáveis dependentes, os respectivos valores de referência das concentrações dos corantes, obtidos através do método cromatográfico.

De acordo com a norma ASTM,<sup>86</sup> o conjunto de amostras utilizado para a construção dos modelos deve representar toda a variabilidade química e física normalmente encontrada na rotina de análises. Além disso, esta variabilidade deve estar distribuída uniformemente ao longo de ampla faixa de concentração, permitindo a construção de modelos com considerável faixa de trabalho.

Nos histogramas das Figuras 25 e 26 estão representadas as faixas de concentração dos corantes obtidas pela metodologia de referência para as amostras utilizadas, bem como a distribuição dessas concentrações entre as amostras. Para a construção dos modelos, as amostras foram agrupadas em dois conjuntos, de calibração e de validação. Os histogramas das amostras foram divididos em dois terços para o conjunto de calibração e um terço para o conjunto de validação usando o algoritmo de Kennard-Stone (seção 2.8.1.1), garantindo ao máximo a distribuição das amostras de calibração ao longo de toda a faixa analítica, de maneira homogênea.

Pode-se inferir, a partir das Figuras 24 e 25 que os valores de concentrações não estão distribuídos de maneira totalmente homogênea, em ampla faixa de concentração, e que há algumas lacunas quando se avalia a faixa de concentração do conjunto selecionado para validação se comparada à faixa de concentração do conjunto de calibração. No entanto, esta limitação está ligada ao fato de trabalharmos com uma situação de calibração natural, na qual não é possível obter as amostras com concentrações previamente determinadas, de maneira planejada.

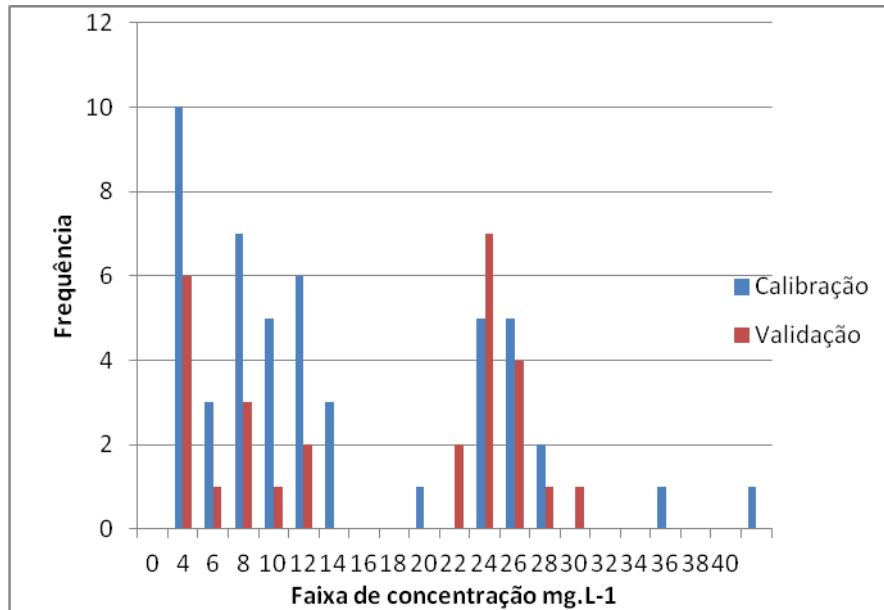


Figura 26. Histograma de frequência das amostras na faixa de concentração de referência para AC

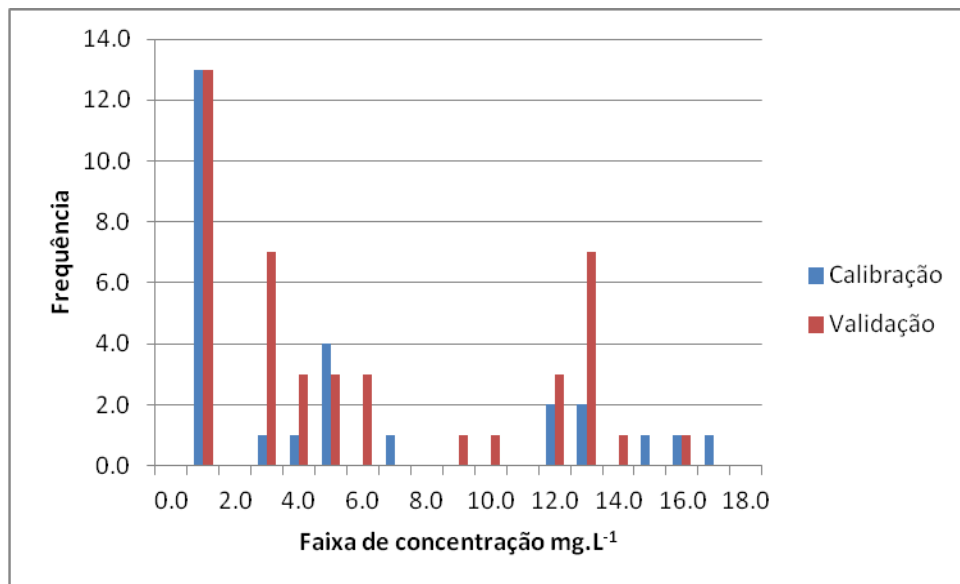


Figura 27. Histograma de frequência das amostras na faixa de concentração de referência para TA

Os histogramas foram somente centrados na média, sem nenhum outro pré-processamento. Nas figuras 26 e 27, são apresentados os histogramas RGB para as 95 amostras do modelo AC e para as 82 amostras do modelo TA, respectivamente.

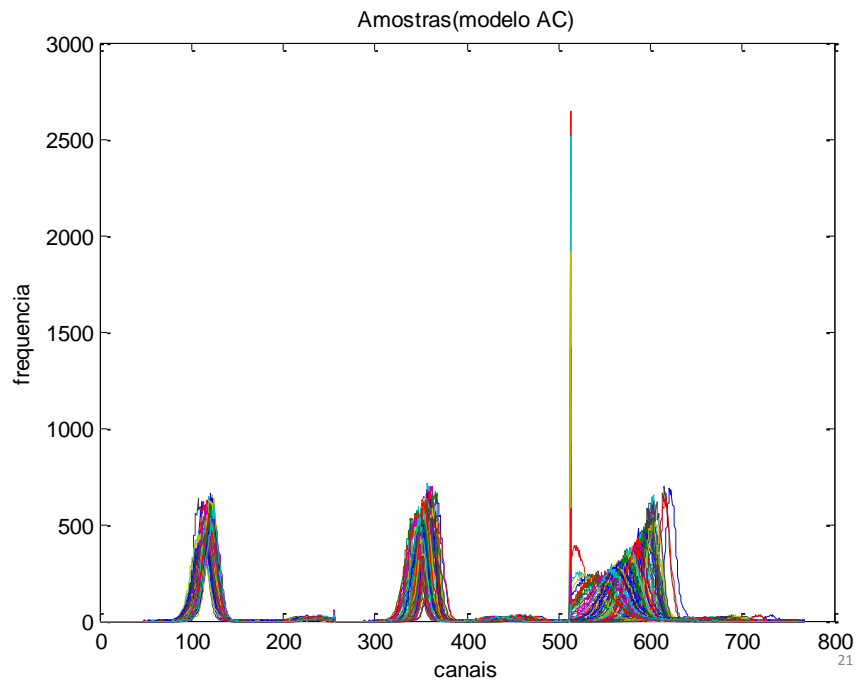


Figura 28. Histogramas(RGB) de frequência para amostras do modelo AC

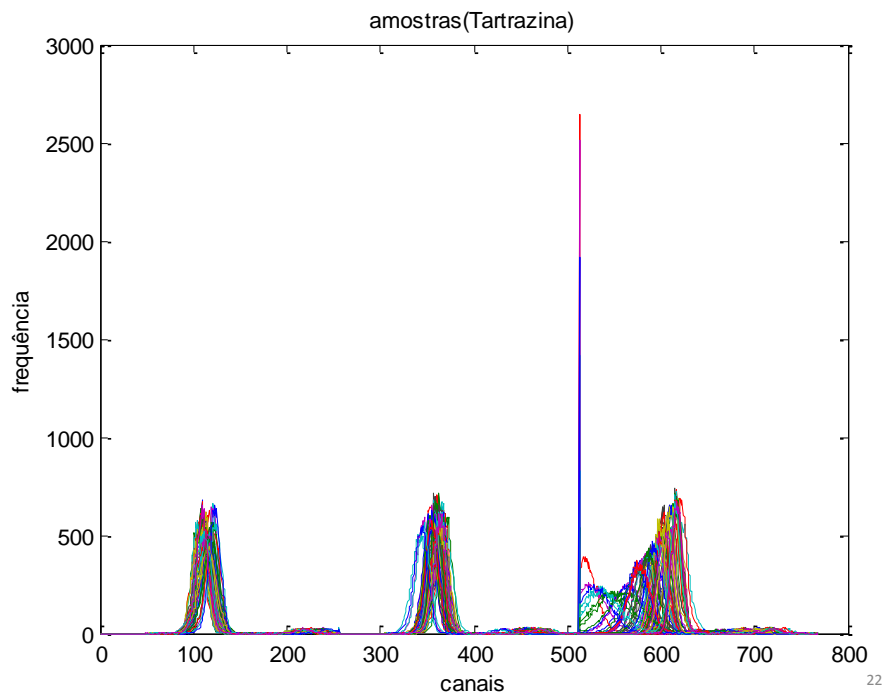


Figura 29. Histogramas(RGB) de frequência para amostras do modelo TA

A detecção de amostras anômalas (*outliers*) nos modelos foi realizada no nível de confiança de 95%, baseada na influência (*leverage*), nos resíduos espectrais (resíduos Q) e nos resíduos de previsão (resíduos em Y). O modelo AC

continha 63 amostras de calibração e 32 de validação (antes da detecção de *outliers*). Com duas rodadas de detecção para o grupo calibração e uma rodada para o grupo de validação, um número de 14 amostras do grupo de calibração (22%) e 4 amostras do grupo de validação (12%) foram detectadas como anômalas, permanecendo 49 amostras de calibração e 28 de validação, totalizando 77 amostras no modelo final. O modelo TA continha 55 amostras de calibração e 27 de validação (antes da detecção de *outliers*). Com duas rodadas de detecção para o grupo calibração e uma rodada para o grupo de validação 12 amostras do grupo de calibração (22%) e nenhuma amostra no grupo de validação foram consideradas anômalas. O modelo final TA foi construído com 43 amostras de calibração e 27 de validação, totalizando 70 amostras. Nos dois modelos, a otimização do conjunto de validação foi realizada após o final da otimização da calibração.

O número de VL foi selecionado com base no menor valor de RMSECV, estimado através da validação cruzada do tipo blocos contínuos (divisão em 7 blocos). O modelo PLS para AC foi construído com 6 VL, e explicou 94,1% da variância em X e 92,7% em Y. O modelo PLS para TA foi construído com 5 VL, e explicou 91,0% da variância em X e 72,7% em Y.

Os sinais de frequência mais intensos não são, necessariamente, os mais importantes para a predição. Vetores informativos obtidos dos modelos PLS, como os coeficientes de regressão e os VIP Scores, são as ferramentas para avaliar a importância das variáveis para a predição.

Os coeficientes de regressão e os VIP Scores para os modelos PLS desenvolvidos são mostrados nas Figuras 28-31.

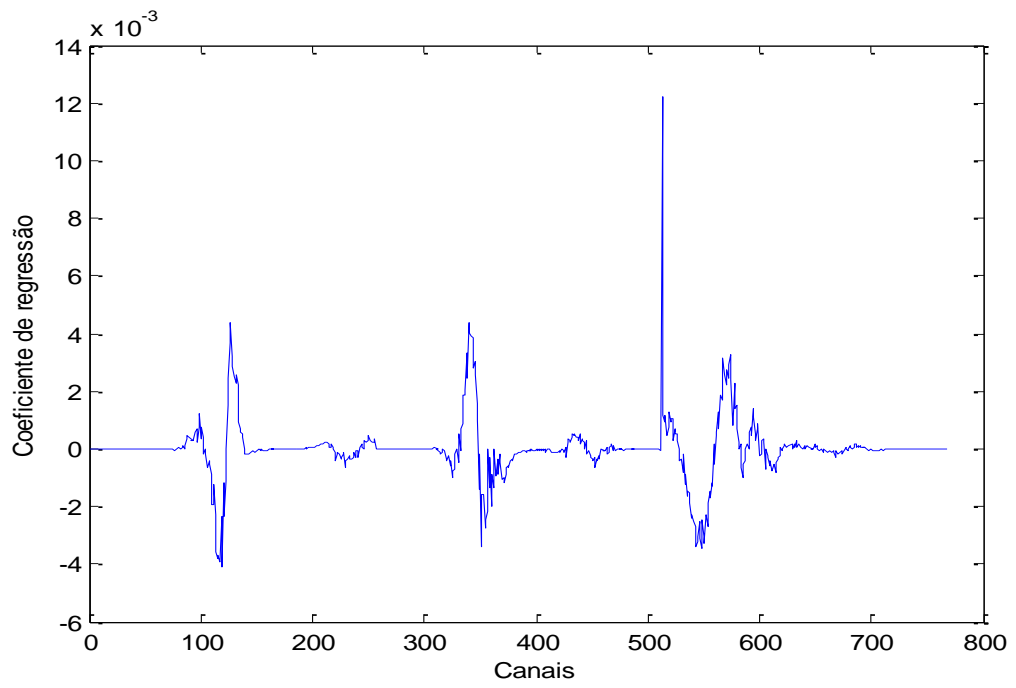


Figura 30. Coeficientes de regressão para o modelo AC

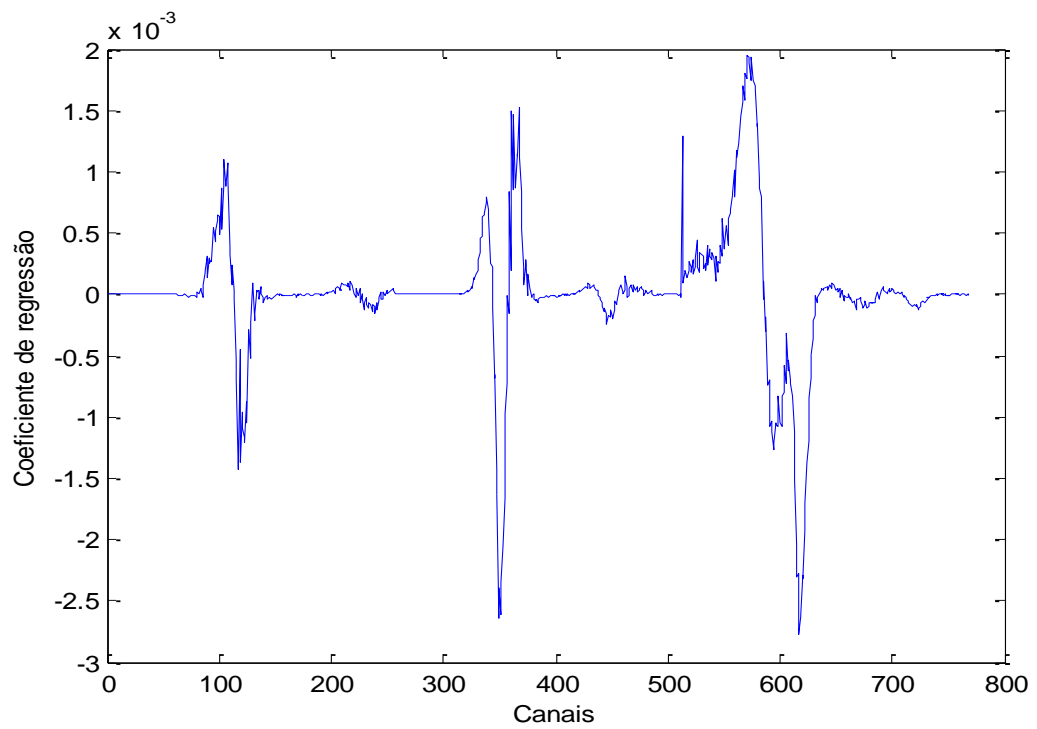


Figura 31. Coeficiente de regressão para o modelo TA

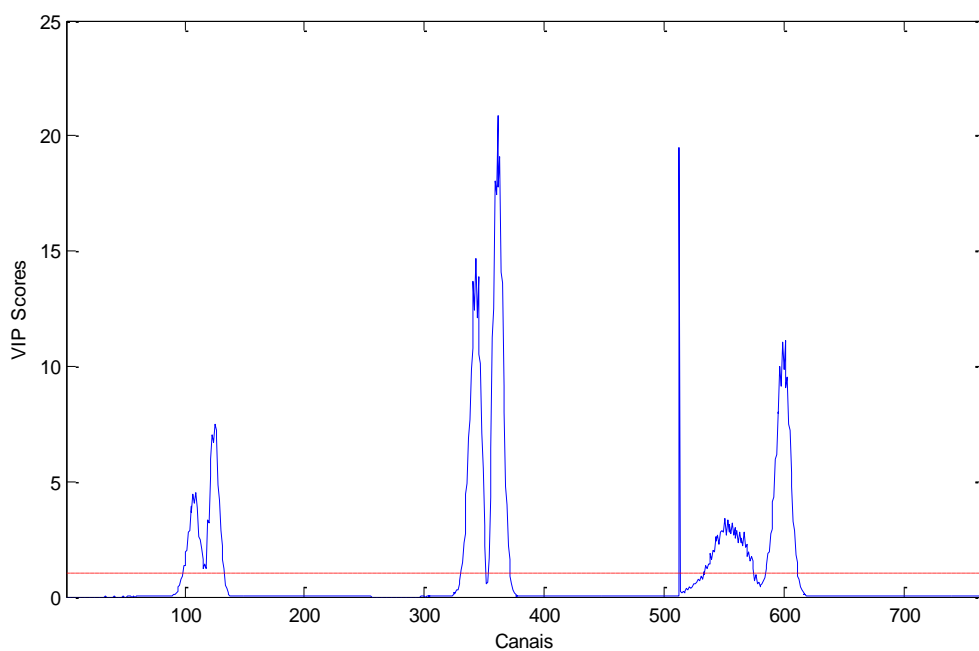


Figura 32. Vip Scores para o modelo AC

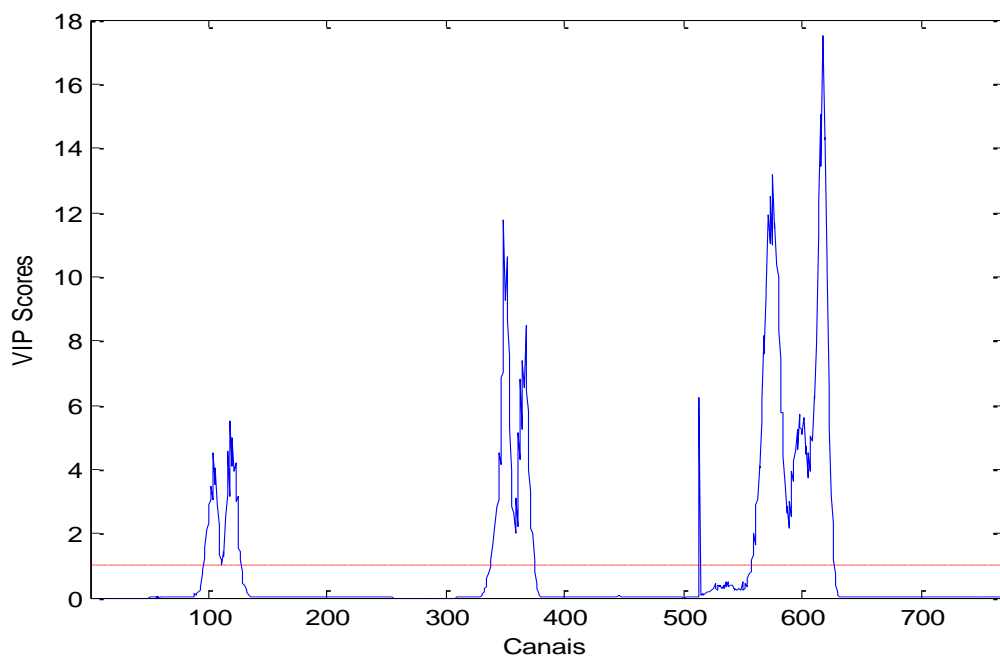


Figura 33. Vip Scores para o modelo TA

Os coeficientes de regressão mais significativos para o modelo AC estão nos canais 119 (contribuição negativa) e 513 (contribuição positiva). Os coeficientes de regressão mais significativos para o modelo TA estão nos canais 571 (contribuição positiva) e 617 (contribuição negativa). De um modo geral, os

coeficientes de regressão mais positivos associados ao AC mostraram contribuição negativa para o modelo TA, e vice-versa, evidenciando o comportamento de que os canais de cor que mais contribuem para prever um corante são interferentes na previsão do outro corante. Isto faz sentido, considerando que os dois corantes possuem tonalidades diferentes e que um é o principal interferente no modelo para prever o outro. Por outro lado, os VIP scores mais significativos, acima de 1,0, indicam de maneira mais robusta os canais mais importantes para a previsão de cada corante, mas não indicam se essa contribuição é positiva ou negativa.

A seleção de variáveis usando OPS foi avaliada e os resultados são mostrados nas Figuras 32 e 33. De um modo geral, as previsões das amostras e a linearidade dos modelos permaneceram similares (Tabela 7). Entretanto, os modelos PLS foram simplificados, pois, ao invés das 768 variáveis originais, foram usadas apenas 100 em cada modelo. Os canais selecionados coincidem, na maioria das vezes, com os mesmos canais que haviam apresentado VIP Scores significativos para cada modelo PLS (Figuras 30 e 31). Sendo assim, todo processo de validação multivariada que será discutido na próxima seção foi aplicado somente aos modelos PLS com seleção de variáveis (PLS-OPS).

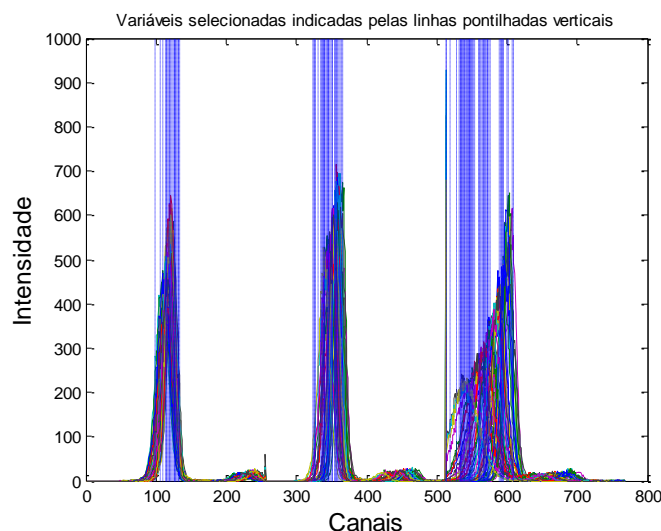


Figura 34. Canais selecionados pelo OPS no modelo AC

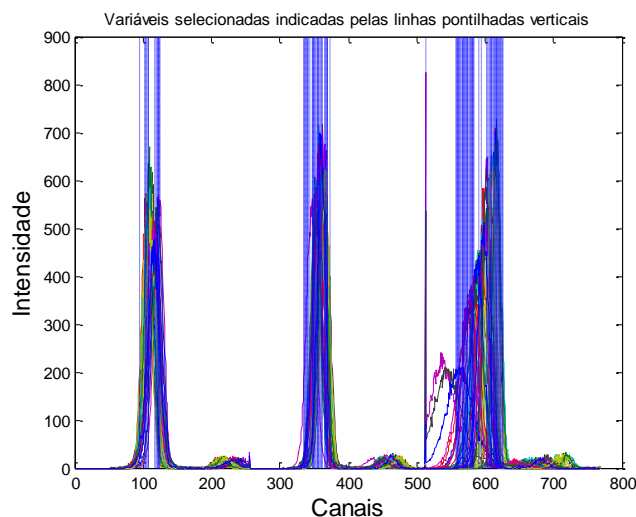


Figura 35. . Canais selecionados pelo OPS no modelo TA

Tabela 7. Resultados da otimização dos modelos PLS e dos modelos PLS-OPS

	Amarelo Crepúsculo 6VL		Tartrazina 5VL	
	PLS	OPS-PLS	PLS	OPS-PLS
$N_{vars}$	768	100	768	100
RMSEC	2,6	2,4	2,5	2,3
RMSECV	4,2	3,5	4,0	3,4
RMSEP	2,8	2,6	2,6	2,7
$r_{cal}$	0,9630	0,9683	0,8527	0,8738
$r_{val}$	0,9547	0,9586	0,8749	0,8646

O modelo PLS-OPS para o AC foi construído também com 6 VL, explicando 94,4% da variância em X e 93,8% em Y. O modelo PLS-OPS para a TA também foi construído com 5 VL, o mesmo número usado no modelo PLS anterior, explicando 92,1%% da variância em X e 76,4% em Y. Esses números de VL podem ser considerados adequados, tendo em vista a complexidade e a variedade de composição das amostras analisadas.

#### 4.2.2 Validação Analítica Multivariada

A validação analítica multivariada<sup>35,77-79</sup> foi aplicada somente aos modelos PLS com seleção de variáveis (OPS), tendo em vista que esses modelos forneceram

os melhores resultados. As Tabelas 8 e 9 apresentam as FOM estimadas para os modelos AC e TA. Os gráficos de valores preditos versus valores de referência são mostrados nas Figuras 35 e 36.

Tabela 8. Parâmetros para avaliação das FOM no modelo proposto para a determinação de AC em bebidas não alcoólicas

Amarelo Crepúsculo		
Figuras de mérito	Parâmetros	Valores
		12,3%
	DPR Repetibilidade	15,7%
Precisão		2,0%
		9,8%
	DPR precisão intermediária	16,9%
		2,5%
Linearidade	Teste de Durbin-Watson	1,9%
	Inclinação/intercepto	0,94/0,80
	Coefficiente de correlação	0,97
Faixa de trabalho		2,3 a 41,1 mg.L <sup>-1</sup>
Seletividade		0,11
Sensibilidade Analítica (Y)		4,86 L.mg <sup>-1</sup>
1/Y		0,2 mg.L <sup>-1</sup>
Viés (bias)		0,2±0,3
LD/LIQ		0,6/2,0 (mg.L <sup>-1</sup> )

Tabela 9. Parâmetros para avaliação das FOM no modelo proposto para a determinação de TA em bebidas não alcoólicas

Tartrazina		
Figuras de mérito	Parâmetros	Valores
		1,6%
	DPR Repetibilidade	4,1%
Precisão		11,1%
		2,1%
	DPR precisão intermediária	3,6%
		11,7%
Linearidade	Teste de Durbin-Watson	1,9%
	Inclinação/intercepto	0,76/1,31
	Coefficiente de correlação	0,87
Faixa de trabalho		0,13 a 15,1 mg.L <sup>-1</sup>
Seletividade		0,20
Sensibilidade Analítica (Y)		3,95 L.mg <sup>-1</sup>
1/Y		0,2 mg.L <sup>-1</sup>
Viés (bias)		-0,1±0,2
LD/LIQ		0,3/1,1 (mg.L <sup>-1</sup> )

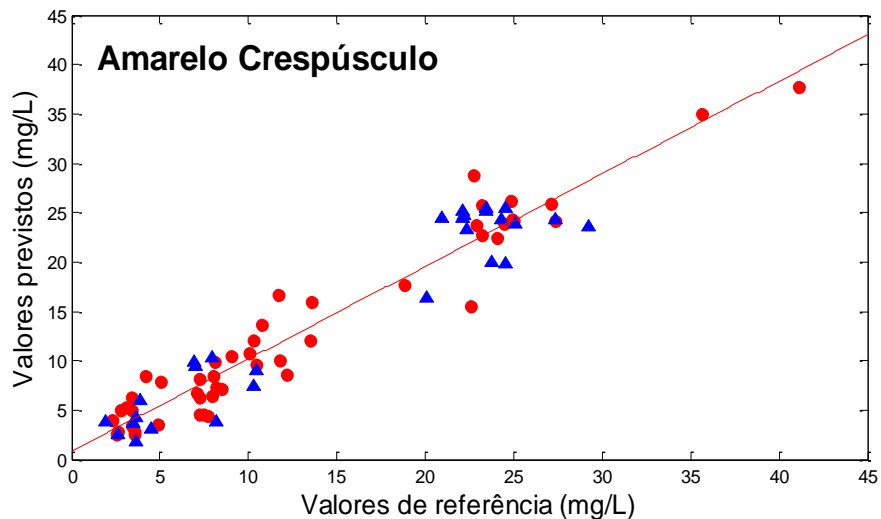


Figura 36. Gráfico de valores de referência vs. valores preditos para o modelo AC. Calibração (círculos) e validação (triângulos)

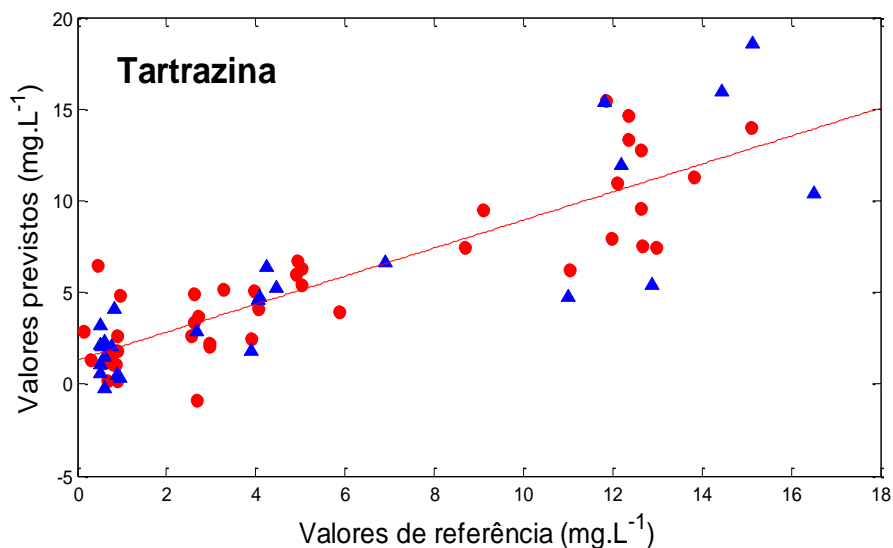


Figura 37. Gráfico de valores de referência vs. valores preditos para o modelo TA. Calibração (círculos) e validação (triângulos)

Para a comprovação da linearidade, a aleatoriedade dos resíduos para as amostras da calibração pode ser avaliada de forma qualitativa através dos gráficos de resíduos em função dos valores preditos, que são mostrados nas Figuras 36 e 37. Nenhuma tendência sistemática foi observada nas distribuições residuais. Ainda assim, o comportamento aleatório desses resíduos foi avaliado através de testes estatísticos (Tabelas 8 e 9).

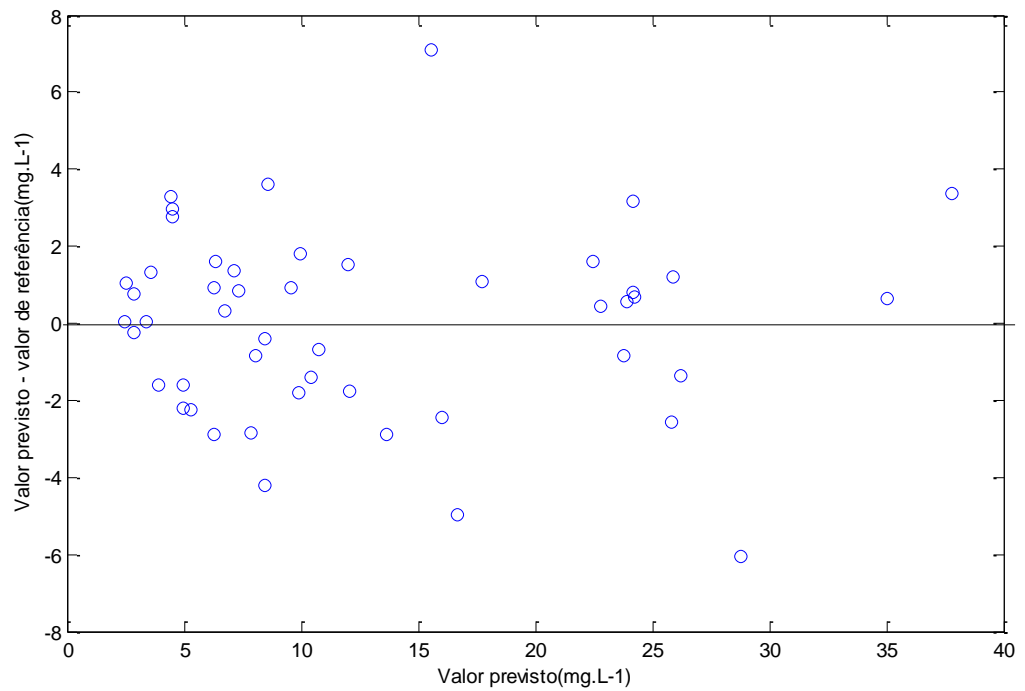


Figura 38. Gráfico de resíduos para as amostras de calibração versus valores preditos pelo modelo AC

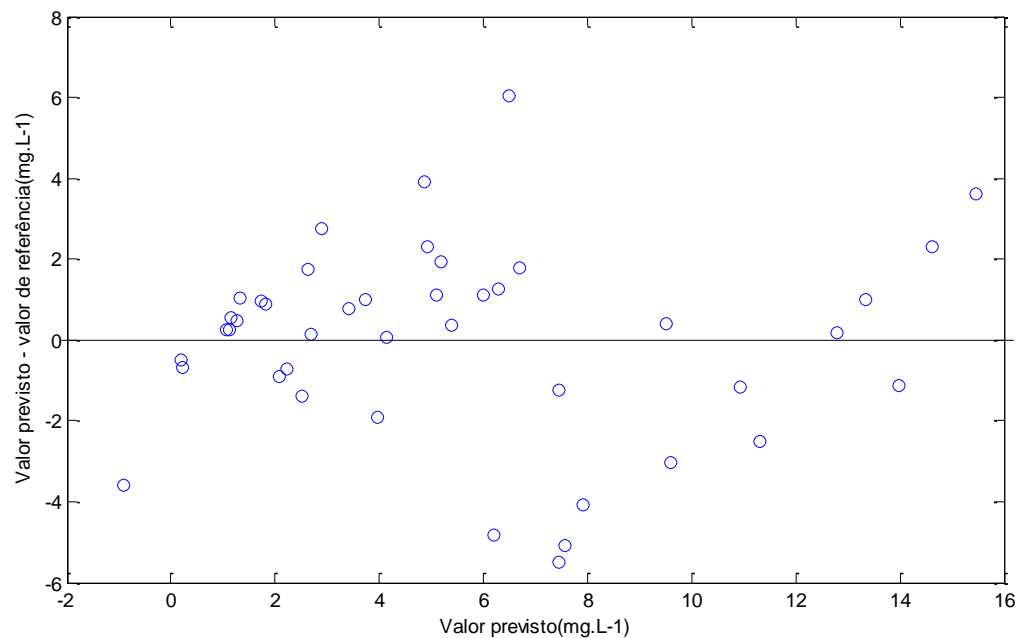


Figura 39. Gráfico de resíduos para as amostras de calibração versus valores preditos pelo modelo TA

Para complementar a avaliação da linearidade, foram usados testes estatísticos apropriados, a fim de verificar a normalidade (Ryan-Joiner),a

homocedasticidade (Brown-Forsythe) e a ausência de correlação dos resíduos (Durbin-Watson), a 95% de confiança.

A normalidade dos resíduos foi avaliada através do teste de Ryan-Joiner (RJ), no qual os resíduos ( $e_i$ ) foram ordenados de forma crescente e plotados contra a porcentagem de pontos correspondentes ( $q_i$ ) a partir da distribuição normal padrão (gráfico quantil-quantil), conforme mostrado nas Figuras 39 e 40. Os resíduos tiveram comportamento normal para o modelo AC, pois os dados se aproximaram do ajuste de uma linha reta. O mesmo não pôde ser concluído para o modelo TA. O coeficiente de correlação ( $R_{eq}$ ) para os dados normalmente distribuídos foi comparado com os valores de  $R_{crítico}$  calculados, para a rejeição ou aceitação da hipótese nula de normalidade. Os resíduos foram considerados normalmente distribuídos quando  $R_{eq} > R_{crítico}$  (Tabelas 10 e 11).

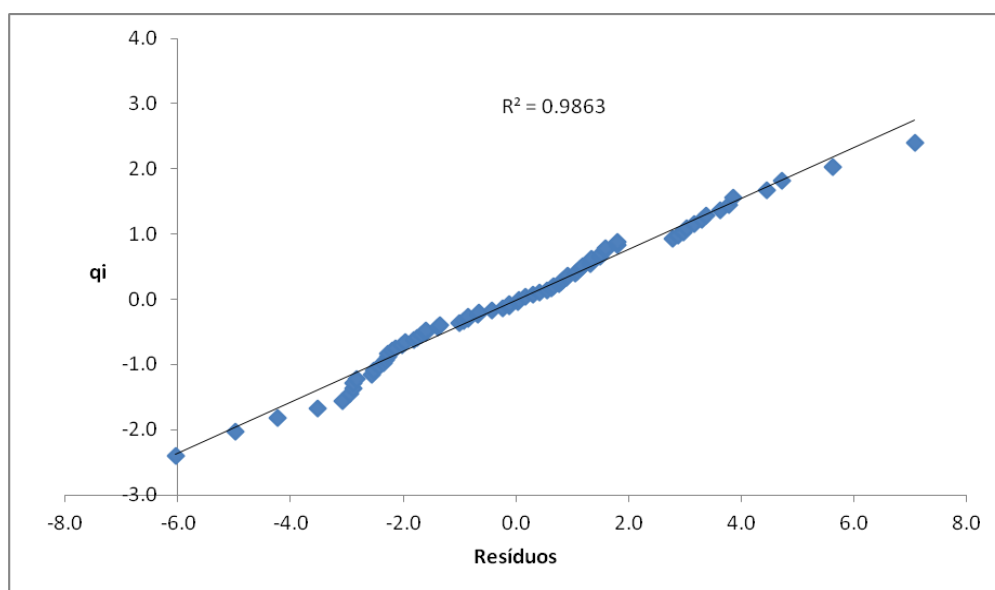


Figura 40. Avaliação da normalidade dos resíduos para modelo AC

Tabela 10. Valores de R críticos e  $R_{eq}$  para o modelo AC

$R_{eq}$	Valores Críticos de R observando confiança considerada	
0,9931	Rcrit. (90%) =	0,9915
	Rcrit. (95%) =	0,9839
	Rcrit. (99%) =	0,9761

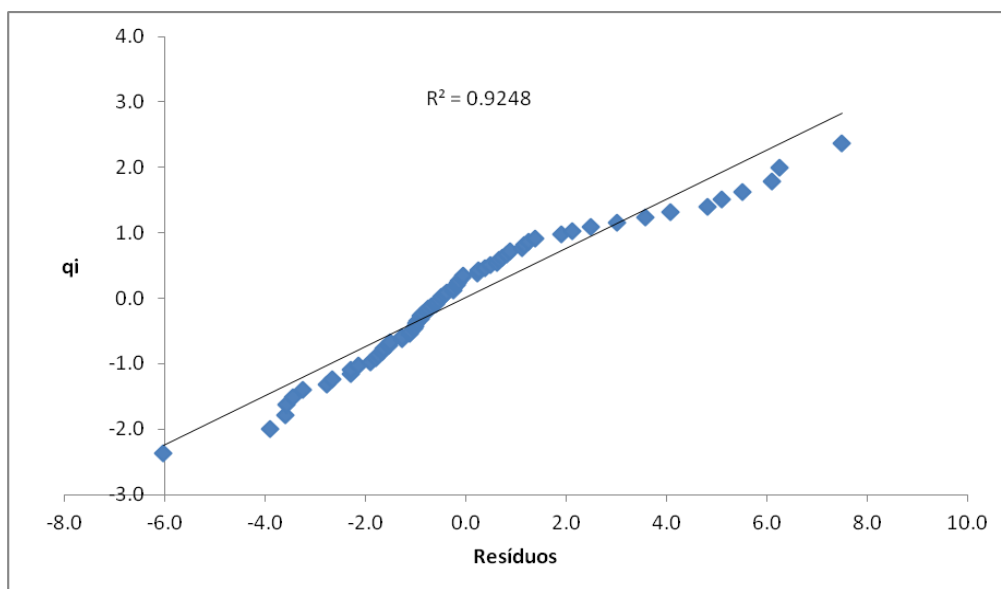


Figura 41. Avaliação da normalidade dos resíduos para modelo TA

Tabela 11. Valores de R críticos e Req para o modelo TA

Req	Valores Críticos de R observando confiança considerada
0,9617	Rcrit. (90%) = 0,9907 Rcrit. (95%) = 0,9824 Rcrit. (99%) = 0,9743

O teste Brown-Forsythe (BF), ou Levene modificado, foi usado na avaliação da homocedastidade dos resíduos, ou seja, a variação absoluta constante de todos os pontos da faixa de trabalho. Os resíduos do modelo AC foram divididos em dois grupos e foram considerados homocedásticos, pois  $t_{\text{calculado}}$  foi 0,10 e se encontra abaixo do valor de  $t_{\text{crítico}}$  para 75 graus de liberdade a 95% de confiança (1,9921). Os resíduos do modelo TA também foram divididos em dois grupos, a 95% de confiança os valores não foram considerados homocedásticos, pois  $t_{\text{calculado}}$  foi 2,26, valor acima do valor de  $t_{\text{crítico}}$  para 68 graus de liberdade (1,9955). Se considerarmos  $t_{\text{crítico}}$  para 68 graus de liberdade a 98% (2,3824) os resíduos são considerados homocedásticos.

A independência dos resíduos foi avaliada com o teste de Durbin-Watson (DW). Para cada conjunto de dados, existem dois limites para d ( $dL$  é o limite inferior e  $dU$  é o limite superior). Se o valor de d estimado para um conjunto de dados

estiver fora destes limites, o teste é inconclusivo ou existe autocorrelação. Os valores d de DW encontrados pra os modelos AC e TA foram, respectivamente, 1,90 e 1,87, valores dentro dos limites de aceitação superior e inferior (1,6 – 2,3), confirmando a independência dos resíduos a 95% de confiança.

Os resultados dos testes confirmam a linearidade do modelo AC. Entretanto, para o modelo TA o teste de Ryan-Joiner não foi satisfatório.

Ainda assim para os dois modelos os parâmetros para a regressão linear foram estimados, incluindo o coeficiente de correlação entre valores preditos e valores de referência (r) de 0,986(modelo AC) e 0,924(modelo TA).

A veracidade foi estimada através de parâmetros que avaliam os erros absolutos, como o RMSEP de 2,6 e 2,7 mg.L<sup>-1</sup> e o RMSEC de 2,4 e 2,3 mg.L<sup>-1</sup> para AC e TA, respectivamente. Os valores previstos e de referência para cada amostra teste (amostra de validação) são mostrados nas Tabelas 12 e 13.

Tabela 12. Valores previstos e de referência para as amostras de validação do modelo AC

<b>Valor referencia</b>	<b>Valor previsto</b>
3,6	1,8
4,5	3,2
3,8	6,0
3,6	4,3
20,1	16,3
24,6	25,5
8,2	3,8
8,0	10,2
1,9	3,9
10,5	9,0
2,6	2,5
23,8	19,9
24,3	24,4
6,9	9,8
7,0	9,4
10,3	7,4
24,5	19,8
23,4	25,1
3,5	3,6
29,2	23,6
23,4	25,4
27,4	24,3
20,9	24,4
25,1	23,9
22,2	24,7
22,2	24,4
22,1	25,2
22,4	23,4

Tabela 13. Valores previstos e de referência para as amostras de validação do modelo TA

Valor referencia	Valor previsto
0,6	2,3
0,6	2,1
0,5	1,1
4,0	4,6
11,0	4,8
0,8	4,1
3,9	1,8
0,5	3,2
0,5	0,6
0,5	2,2
1,0	0,3
0,6	-0,2
16,5	10,4
12,9	5,4
14,4	16,0
15,1	18,6
6,9	6,6
4,1	4,8
4,3	6,4
4,4	5,2
11,8	15,4
12,2	12,0
0,9	0,5
2,7	2,9
0,8	2,0
0,6	1,5
0,6	1,5

Os valores de precisão sob condições de repetibilidade e precisão intermediária, expressos como desvio padrão relativo ( $DPR_r$  e  $DPR_R$ , respectivamente), foram estimados para 18 replicatas de amostras em três níveis de concentração distintos.<sup>98</sup> Os valores de  $DPR_r$  e  $DPR_R$  encontrados para os modelos AC e TA para cada nível e limites máximos calculado através da equação de Horwitz<sup>94</sup> são apresentados na Tabela 14.

Tabela 14. Estudo de repetibilidade e precisão intermediária

	AC			TA		
Concentração(mg.L <sup>-1</sup> )	3,7	11,7	39,3	0,7	4,1	12,9
DPR <sub>r</sub>	12,3	15,7	2,0	6,3	5,2	4,7
DPR <sub>R</sub>	9,8	16,9	2,5	2,1	3,6	11,7
DPR limite <sup>94</sup>	6,6	5,5	4,5	5,4	5,4	5,1

Os resultados de veracidade e precisão não corroboram com a afirmação de que o método pode ser considerado exato. A faixa de trabalho para o primeiro modelo foi de 2.3 a 41.1 mg.L<sup>-1</sup> de AC. A faixa de trabalho para o segundo modelo foi de 0,13 a 15,11 mg.L<sup>-1</sup> de TA. Porém, avaliando o LD e LIQ (Tabela 9) pode-se inferir que o modelo TA não se aplica a valores de concentrações de TA menores que 1,1 mg.L<sup>-1</sup>; o que gera uma questão relevante, tendo em vista que algumas amostras foram determinadas pelo método de referência com concentrações abaixo desse limite (anexo A)

A SEN e a SEL também foram estimadas, com base no NAS. A necessidade de um valor mínimo de SEL não tem sentido prático na validação multivariada, uma vez que, diferente da validação univariada, métodos multivariados podem ser exatos mesmo com valores de SEL baixos. A SEL estimada indica que apenas 11% do sinal analítico original foi utilizado no modelo AC e 20% no modelo TA. A sensibilidade analítica foi estimada através do ruído instrumental (5,95), obtido da medida de dez replicatas de imagens de uma placa de Petri vazia. O seu inverso ( $\gamma^{-1}$ ), 0,2 mg.L<sup>-1</sup> (modelo AC) e 0,1 mg.L<sup>-1</sup> (modelo TA), indica a quantidade mínima discernível pelo método, considerando o erro instrumental aleatório como única fonte de erro, e também define o número de casas decimais com as quais os resultados devem ser expressos.

O viés, estimado apenas para o conjunto de validação, apresenta um valor de t igual a 0,3287 (modelo AC) e 0,1062 (modelo TA), ambos menores que os valores críticos (2,052, com 27 graus de liberdade (modelo TA) e 2,048 com 28 graus de liberdade (modelo AC), e 95% de confiança), o que indica a ausência de erro sistemático nas predições do modelo.

De acordo com a literatura,<sup>87</sup> bons modelos de calibração devem apresentar valores de RPD acima de 2,4, enquanto valores entre 2,4 e 1,5 são considerados aceitáveis. Considerando os valores de RPD<sub>cal</sub> e RPD<sub>val</sub> para o modelo

AC (2,7 e 3,5, respectivamente) e para o modelo TA (1,4 e 2,0, respectivamente); pode se concluir que os valores para o modelo AC indicam um bom modelo de calibração, porém para o modelo TA os valores foram apenas razoáveis.

### **4.3 Conclusão Parcial**

Um método de calibração multivariada simples, baseado em histogramas RGB obtidos a partir de imagens digitais, foi desenvolvido para a determinação simultânea de amarelo crepúsculo e tartrazina em bebidas não alcoólicas.

Avaliando a confiabilidade dos modelos através da validação analítica, percebeu-se que para o modelo AC alguns critérios não foram satisfatórios, como a precisão e erros individuais de previsão das amostras altos. Entretanto, com relação a outros parâmetros o modelo foi satisfatório, apresentando linearidade, sensibilidade, ausência de viés e RPD aceitáveis. Para o modelo TA, os problemas encontrados foram maiores. Houve desvios de linearidade (teste Ryan-Joiner), altos erros, valores fora dos limites estabelecidos para precisão, limite de quantificação inferior (LIQ) superior ao encontrado em várias amostras pelo método de referência e RPDcal insatisfatório. Previamente, não havia nenhum levantamento sobre a possível faixa de concentração de TA nas amostras, ao contrário da concentração de AC (já estudada em outro trabalho).<sup>35</sup> Foram determinados valores de concentrações de TA muito baixos e faixa bastante ampla (0,13 a 15,1 mg.L<sup>-1</sup>), uma diferença da menor para a maior concentração de mais de 100 vezes. Além do número de amostras reduzidas, contendo esse corante, se comparado ao número de amostras contendo AC. Entretanto, ainda assim o modelo TA apresentou boa sensibilidade, ausência de viés e apesar de não passar nos teste Ryan-Joiner, passou no teste de Durbin-Watson (DW) a 95% e Brown-Forsythe (BF) a 98%, indicando a independência dos resíduos (não houve autocorrelação) e a homocedasticidade.

Este método utilizou um equipamento de baixo custo (escâner de mesa comercial), não necessitou de pré-tratamento das amostras, além de banho de ultrassom, e uso de solventes, e apresentando-se muito mais rápido que o método de referência (poucos segundos contra 12 minutos da corrida cromatográfica).

O método proposto pode ser voltado para uma finalidade produtiva em automatização direta (*on-line*) de processos industriais, com uma frequência de amostragem elevada, ou em equipamentos portáteis.

O método não foi completamente validado de acordo com guias brasileiros e internacionais de validação. O cálculo das figuras de mérito para os modelos apontaram para a necessidade de melhoramento dos mesmos.

Percebe-se a complexidade de tais amostras e a possibilidade de um melhor modelo com inclusão de um número maior de amostras e de amostras representativas de cada marca e tipo. Além disso, métodos de calibração multivariada não linear, como máquinas de suporte de vetores (SVM, *Support Vector Machines*)<sup>99</sup> poderiam também ser utilizadas, principalmente para o modelo TA, que apresenta ampla faixa.

## 5. Conclusão geral

O objetivo principal desta dissertação foi o desenvolvimento de um método simples, rápido e barato para a determinação simultânea dos corantes artificiais amarelo crepúsculo e tartrazina em bebidas não alcoólicas. Este método se baseou na análise de imagens digitais adquiridas por um escâner de mesa, tratadas por métodos quimiométricos de calibração multivariada e seleção de variáveis. Para este objetivo, foi necessário previamente desenvolver e validar um método baseado em CLAE de fase reversa para a quantificação destes dois corantes nas amostras e obtenção dos valores de referência.

O método quimiométrico apresentou bons resultados para a quantificação do amarelo crepúsculo, atendendo a maioria dos requisitos para sua validação analítica. Por outro lado, os resultados para o modelo de previsão da tartrazina não foram tão bons, embora este tenha apresentado uma razoável linearidade. A pior qualidade deste modelo está ligada à maior amplitude de sua faixa analítica e à presença de várias amostras com concentrações muito baixas, próximas ao limite de sensibilidade do método. Para melhorar este modelo, seria recomendável a obtenção de mais amostras. Além disso, poderia ser testado o uso de métodos de calibração multivariada não lineares, como SVM,<sup>99</sup> que são recomendados justamente para faixas analíticas muito amplas. Uma outra opção de tratamento de dados para estas imagens seria o uso de uma estratégia quimiométrica mais complexa, empregando pré-processamento por transformada de Fourier e gerando um arranjo multidimensional, que precisaria ser modelado por métodos de calibração de ordem superior, como o PLS multilinear (N-PLS).<sup>56,58</sup>

Finalmente, os dois modelos de calibração multivariada foram otimizados usando OPS e reduzindo o número de variáveis/canais empregadas em sua construção cerca de 8 vezes (de 768 para 100).

## 6. Referências

- [1] Rosa, S.E.S.; Cosenza, J.P.; Leão, L.T.S. “Panorama do setor de bebidas no Brasil”. BNDES Setorial, Rio de Janeiro, 23, 101-150 (2006).
- [2] Cervieri Júnior, O. ; Teixeira Junior, J.R. ; Galinari, R.; Rawet, E.L.; Silveira, C.T.J. “O setor de bebidas no Brasil”. BNDES Setorial 40, 93-130 (2014). Disponível em: [www.bndes.gov.br/biblioteca\\_digital](http://www.bndes.gov.br/biblioteca_digital) . Acesso em 08/2015.
- [3] Euromonitor Internacional, “Ranking mundial de consumo per capita de refrigerantes”. Disponível em: [www.euromonitor.com/brazil](http://www.euromonitor.com/brazil). Acesso em 08/2015.
- [4] Brownell, K.D.; Thomas R.F. “Ounces of prevention- The public policy case for taxes on sugared beverages”. The New England Journal of Medicine. 360, 1805-1808 (2009).
- [5] Prado, M.A.; Godoy, H.T. “Corantes artificiais em alimentos”. Alimentos e Nutrição, 14, 237-250 (2003).
- [6] ANVISA. Agência Nacional de Vigilância Sanitária. Resolução Nº 382, de 05 de Agosto de 1999.
- [7] ANVISA. Agência Nacional de Vigilância Sanitária. Resolução Nº388, de 05 de Agosto de 1999.
- [8] Kunz, A.; Peralta-Zamora, P.; Moraes, S.G.; Duran, N. “Novas tendências no tratamento de efluentes têxteis”. Química Nova, 25, 78-82 (2002).
- [9] Xing, Y.; Meng, M.; Xue, H.; Zhang, T.; Yin, Y.; Xi, R. “Development of a polyclonal antibody-based enzyme-linked 117 immunosorbent assay (ELISA) for detection of Sunset Yellow FCF in food samples”. Talanta, 99, 125–131 (2012).
- [10] Yadav, A.; Kumar, A.; Tripathi, A.; Das, M. “Sunset yellow FCF, a permitted food

dye, alters functional responses of splenocytes at non-cytotoxic dose”. *Toxicology Letters*, 217, 197–204 (2013).

[11] Freitas, A.S. “Tartrazina: uma revisão das propriedades e análises de quantificação”. *Acta Tecnológica*, 7, 65-72 (2012).

[12] ANVISA. Agência Nacional de Vigilância Sanitária. “Compêndio da Legislação Brasileira de Aditivos Alimentares”, Brasília, 163 (2011).

[13] Moutinho, I.L.S.; Bertges, L.C.; Assis, R.V.C. “Prolonged use of food dye tartrazine (FD&C yellow nº5) and its effects on the gastric mucosa of Wistar rats”. *Brazilian Journal of Biology*, 67, 141-145 (2007).

[14] Sasaki, Y.F.; Kawaguchi, S.; Kamaya, A.; Ohshita, M.; Kabasawa, K.; Iwama, K.; Taniguchi, K.; Tsuda, S. “The comet with 8 mouse organs: results with 39 currently used food additives”. *Mutation Research*, 519, 103-119 (2002).

[15] Vogler Ingredients. “Dossiê corantes”. *Food Ingredients Brasil*, 9, 40-59, 2009. Disponível em: [www.revista-fi.com](http://www.revista-fi.com). Acesso em 08/2015.

[16] IAL. Instituto Adolfo Lutz. “Métodos Físicos-Químicos para Análise de Alimentos”, 4ª edição. Instituto Adolfo Lutz (IAL), São Paulo (2005).

[17] Tang, T. X.; Xu, X.J.; Wang, D.M.; Zhao, Z.M.; Zhu, L.P. ; Yang, D.P. “A rapid and green limit test method for five synthetic colorants in foods using polyamide thin-layer chromatography”. *Food Analytical Methods*. 8, 459-466 (2015).

[18] Bonan, S.; Fedrizzi, G.; Menotta, S.; Elisabetta, C. “Simultaneous determination of synthetic dyes in foodstuffs and beverages by high performance liquid chromatography coupled with diode-array detector”. *Dyes and Pigments*, 99, 36–40 (2013).

[19]. Vidotti, E.C.; Costa, W.F.; Oliveira, C.C. “Development of a green

chromatographic method for determination of colorants in food samples". *Talanta*, 68, 516–521 (2006).

[20] Nevado, J.J.B.; Flores, J.R.; Llerena, M.J.V. "Square wave adsorptive voltametric determination of sunset yellow". *Talanta* 44, 467–474 (1997).

[21] Ghoreishi, S.M.; Behpour, M.; Golestaneh, M. "Simultaneous determination of Sunset yellow and Tartrazine in soft drinks using gold nanoparticles carbon paste electrode". *Food Chemistry*, 132, 637–641 (2012).

[22] Dossi, N.; Piccin, E.; Bontempelli, G.; Carrilho, E.; Wang, J. "Rapid analysis of azo-dyes in food by microchip electrophoresis with electrochemical detection". *Electrophoresis*, 4240-4246 (2007).

[23] Blanco, C.C.; Campana, A.M. G.; Barrero, F.A. "Derivative spectrophotometric resolution of mixtures of the food colourants Tartrazine, Amaranth and Curcumin in a micellar medium". *Talanta*, 43, 1019-1027 (1996).

[24] Jiang, S.H.; Xu, J.F.; Xu, P.D.; Liu, L.J.; Chen, Y.; Qiao, C.S.; Yang, S.F. ; Sha, Z.L.; Zhang, J.K. "A novel molecularly imprinted sensor for direct tartrazine detection". *Analytical Letters*, 47, 323-330 (2014).

[25] El-Shahawi, M.S.; Hamza, A.; Al-Sibaai, A.A.; Bashammakh, A.S.; Al-Saidi, H.M. "A new method for analysis of sunset yellow in food samples based on cloud point extraction prior to spectrophotometric determination". *Journal of Industrial and Engineering Chemistry*, 19, 529–535 (2013).

[26] Sha, O.; Zhu, X.S.; Feng, Y.L.; Ma, W.X. "Aqueous two-phase based on ionic liquid liquid-liquid microextraction for simultaneous determination of five synthetic food colourants in different food samples by high-performance liquid chromatography". *Food Chemistry*, 174, 380-386 (2015).

[27] Wang, M.L.; Zhao, J.W. "Facile synthesis of Au supported on ionic liquid

functionalized reduced graphene oxide for simultaneous determination of Sunset yellow and Tartrazine in drinks”. *Sensors and Actuators B-Chemical*, 216, 578-585 (2015).

[28] Nevado, J.J.B.; Flores, J.R.; Cabanillas, C.G.; Llerena, M.J.V.; Salcedo, A.C. “Resolution of ternary mixtures of Tartrazine, Sunset yellow and Ponceau 4R by derivative spectrophotometric ratio spectrum-zero crossing method in commercial foods”. *Talanta*, 46, 933–942 (1998).

[29] Nevado, J.J.B.; Flores, J.R.; Llerena, M.J.V.; Fariñas, N.R. “Simultaneous spectrophotometric determination of tartrazine, patent blue V, and indigo carmine in commercial products by partial least squares and principal component regression methods”. *Talanta*, 48, 895–903 (1999).

[30] El-Sheikh, A.H.; Al-Degs, Y.S. “Spectrophotometric determination of food dyes in soft drinks by second order multivariate calibration of the absorbance spectra-pH data matrices”. *Dyes and Pigments*. 97, 330–339 (2013).

[31] Santos, M.E.; Demiate, I.M.; Nagata, N. “Determinação simultânea de amarelo tartrazina e amarelo crepúsculo em alimentos via espectrofotometria UV-VIS e métodos de calibração multivariada”. *Ciência e Tecnologia de Alimentos*, 30, 903-909 (2010).

[32] Prado, M.A.; Godoy, H.T. “Determinação de corantes artificiais por Cromatografia Líquida de Alta Eficiência (CLAE) em pó para gelatina”. *Química Nova*, 27, 22-26 (2004).

[33] Prado, M.A.; Godoy, H.T. “Teores de corantes artificiais em alimentos determinados por cromatografia líquida de alta eficiência”. *Química Nova*, 30, 268-273 (2007).

[34] Kucharska, M.; Jan, G. “A review of chromatographic methods for determination of synthetic food dyes”. *Talanta*, 80, 1045-1051 (2010).

- [35] Botelho, B.G.; Assis L.P.; Sena M.M. "Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages". *Food Chemistry*, 159, 175-180 (2014).
- [36] Collins, C.H.; Braga, G.L.; Bonato, P.S. "Fundamentos de Cromatografia". Editora da Unicamp. Campinas, (2006).
- [37] Gjerde, D.T.; Fritz, J.S. "Ion Chromatography", Wiley-VCH. Heidelberg, 283 (1987)
- [38] Kirkland, J.J. "Modern Practice of Liquid Chromatography". Wiley-Interscience, New York, 454 (1971).
- [39] Ciola, R. "Fundamentos da Cromatografia a Líquido de Alto Desempenho". Editora Blucher, São Paulo (1998).
- [40] Skoog, D.A.; West, D.M.; Holler, F.J.; Crouch, S.R. "Fundamentos de Química Analítica". 1ª edição, Cengage Learning (2007).
- [41] Schure, M.R.; Rafferty, J.L.; Zhang L.; Siepmann, J.L. "How reversed-phase liquid chromatography works", *LC GC North America*, 31, 630-637 (2013).
- [42] Poole, C.F.; Schuette, S.A. "Contemporary Practice of Chromatography". Elsevier, Amsterdam (1984).
- [43] Neue, U.D. "HPLC Columns: Theory, Technology and Practice". Wiley-VCH, New York, 416 (1997).
- [44] Snyder, L.R., Kirkland, J.J., Glajch, J.L. "Practical HPLC Method Development" 2<sup>nd</sup> edition. Wiley, New York (1997).
- [45] Geladi, P.; Grahn, H. "Multivariate Image Analysis". 1<sup>st</sup> Edition. John Wiley &

Sons. New York (1996).

[46] Matos, G.D.; Pereira-Filho, E.R.; Poppi, R.J.; Arruda, M.A.Z. “Análise exploratória em química analítica com emprego de quimiometria: PCA e PCA de imagens”. *Revista Analytica*, 06, 38-50 (2003).

[47] Kucheryavski, S. “Extracting useful information from images”. *Chemometrics and Intelligent Laboratory Systems*, 108, 2-12 (2011).

[48] Lied, T.T.; Esbensen, K.H. “Principles of MIR, multivariate image regression I: Regression typology and representative application studies”. *Chemometrics and Intelligent Laboratory Systems*, 58, 213-226 (2001).

[49] Ledley, R.S. “High-speed automatic analysis of biomedical pictures”. *Science*, 146, 216–223 (1964).

[50] Geladi, P.; Wold, S.; Esbensen, K. “Image analysis and chemical information in images”. *Analytica Chimica Acta*, 191, 473–480 (1986).

[51] Foca, G.; Masino, F.; Antonelli, A.; Ulrici, A. “Prediction of compositional and sensory characteristics using RGB digital images and multivariate calibration techniques”. *Analytica Chimica Acta*, 706, 238–245 (2011).

[52] Iqbal, Z.; Bjorklund, R. B. “Assessment of a mobile phone for use as a spectroscopic analytical tool for foods and beverages”. *International Journal of Food Science & Technology*, 46, 2428–2436 (2011).

[53] Santos, P.M.; Wentzell, P.D.; Pereira-Filho, E.R. “Scanner digital images combined with color parameters: A case study to detect adulterations in liquid cow’s milk”. *Food Analytical Methods*, 5, 89–95 (2012).

[54] Borin, A.; Ferrão, M.F.; Mello, C.; Cordi, L.; Pataca, L.C.; Durán, N.; Poppi, R.J. “Quantification of lactobacillus in fermented milk by multivariate image analysis with

least-squares support-vector machines”. *Analytical and Bioanalytical Chemistry*, 387, 1105–1112 (2007).

[55] Acevedo, C.A.; Skurtys, O.; Young, M.E.; Enrione, J.; Pedreschi, F.; Osorio, F. “A non-destructive digital imaging method to predict immobilized yeast-biomass”. *LWT - Food Science and Technology*, 42, 1444–1449 (2009).

[56] Godinho, M.S.; Oliveira, A.E.; Sena, M.M. “Determination of interfacial tension of insulating oils by using image analysis and multi-way calibration. *Microchemical Journal*”, 96, 42–45 (2010).

[57] Oliveira, L.F.; Canevari, N.T.; Guerra, M.B.B.; Pereira, F.M.V.; Schaefer, C.E.G.R.; Pereira-Filho, E.R. “Proposition of a simple method for chromium (VI) determination in soils from remote places applying digital images: A case study from Brazilian Antarctic Station”. *Microchemical Journal*, 109, 165–169, (2013).

[58] Huang, J.; Wium, H.; Qvist, K.B.; Esbensen, K.H. “Multi-way methods in image analysis—relationships and applications”. *Chemometrics and Intelligent Laboratory Systems*, 66, 141–158 (2003).

[59] Kowalski, B.R.; Bender, C. “Patern recognition. A powerful approach to interpreting chemical data”. *Journal of the American Chemical Society*, 94, 5632-5639 (1972).

[60] Ferreira, M.M.C. “Quimiometria - Conceitos, Métodos e Aplicações”. Editora da Unicamp, Campinas (2015).

[61] Kowalski, B.R. “Chemometrics: Theory and Application”. American Chemical Society, Washington (1977).

[62] Barros Neto, B.; Scarminio, I.S.; Bruns, R.E. “25 anos de quimiometria no Brasil”. *Química Nova*, 29, 1401–1406 (2006).

[63] Neto, B.B.; Scarminio, I. S.; Bruns, R.E. “Como Fazer Experimentos – Pesquisa e Desenvolvimento na Ciência e na Indústria”, Editora da Unicamp, Campinas (2001).

[64] Kowalski, B.R. “Chemometrics: Views and propositions”. *Journal of Chemical Information and Computer Sciences*, 15, 201-203 (1975).

[65] Ferreira, M.M.C.; Antunes, A.M.; Melgo, M.S.; Volpe, P.L.O. “Quimiometria I: Calibração multivariada, um tutorial”. *Química Nova*, 22, 724–731 (1999).

[66] Geladi, P.; Kowalski, B. “Partial Least Square Regression: a tutorial”. *Analytica Chimica Acta*, 186, 1-17 (1986).

[67] Brereton, R.G. “Applied Chemometrics for Scientists”, Chichester. Wiley, 2007.

[68] Kennard, R.W.; Stone, L.A. “Computer aided design of experiments”. *Technometrics*, 11, 137-148 (1969).

[69] Wise, B.W.; Gallagher, N.B.; Bro, R.; Shaver, J.M.; Windig, W.; Koch, R.S. “PLS-Toolbox 4.0 for use with Matlab™”. EigenVector Research Inc, Manson (2006).

[70] Brown, C.D.; Green, R.L. “Critical factors limiting the interpretation of regression vectors in multivariate calibration”. *TrAC - Trends in Analytical Chemistry*, 28, 506-514 (2009).

[71] Chong, I.-G.; Jun, C.-H. “Performance of some variable selection methods when multicollinearity is present”. *Chemometrics and Intelligent Laboratory Systems*, 78, 103-112 (2005).

[72] Teófilo, R.F.” Métodos quimiométricos em estudos eletroquímicos de fenóis sobre filmes de diamante dopado com boro”. Tese de doutorado. Instituto de Química, Universidade Estadual de Campinas (2007).

- [73] Teófilo, R.F.; Martins, J.P.A.; Ferreira, M.M.C. "Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression". *Journal of Chemometrics*, 23, 32-48 (2009).
- [74] Helland, I.S. "On the structure of partial least squares regression". *Communications in Statistics - Simulation and Computation*, 17, 581-607 (1988).
- [75] Williams, R.P.; Swinkels, A.J.; Maeder, M. "Identification and application of a prognostic vector for use in multivariate calibration and prediction". *Chemometrics and Intelligent Laboratory Systems*, 15, 185-193 (1992).
- [76] Pirouette<sup>TM</sup>. *Pirouette: Multivariate data analysis*, 3.11; Infometrix, Woodinville (2003).
- [77] Valderrama, P.; Braga, J.W.B.; Poppi, R.J. "Estado da arte de figuras de mérito em calibração multivariada". *Química Nova*, 32, 1278-1287 (2009).
- [78] Valderrama, P.; Braga, J.W.B.; Poppi, R.J. "Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy". *Journal of Agricultural and Food Chemistry*, 55, 8331-8338 (2007).
- [79] Silva, M.A.M.; Ferreira M.H.; Braga, J.W.B.; Sena, M.M. "Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy". *Talanta* 89, 342-351 (2012).
- [80] Lorber, A. "Error propagation and figures of merit for quantification by solving matrix equations". *Analytical Chemistry*, 58, 1167-1172 (1986).
- [81] Lorber, A.; Faber, K.; Kowalski, B.R. "Net analyte signal calculation in multivariate calibration". *Analytical Chemistry*, 69, 1620-1626 (1997).

[82] Faber, N.M. "Exact presentation of multivariate calibration model as univariate calibration graph". *Chemometrics and Intelligent Laboratory Systems*, 50, 107-114 (2000).

[83] Souza, S.V.C.; Junqueira, R.G. "A procedure to assess linearity by ordinary least squares method". *Analytica Chimica Acta*, 552, 25-35 (2005).

[84] Botelho, B.G.; Mendes, B.A.P.; Sena, M.M. "Development and analytical validation of robust near-infrared multivariate calibration models for the quality inspection control of mozzarella cheese". *Food Analytical Methods*, 6, 881-891 (2013).

[85] Guimarães, C.C.; Simeone, M.L.F.; Parrella, R.A.C.; Sena, M.M. "Use of NIRS to predict composition and bioethanol yield from cell wall structural components of sweet sorghum biomass". *Microchemical Journal*, 117, 194-201 (2014).

[86] ASTM International. ASTM E1655-05. "Standard Practices for Infrared Multivariate Quantitative Analysis" (2012).

[87] Williams, P.; Norris, K. "Near-infrared technology in the agricultural and food industries", 1<sup>st</sup> Edition. American Association of Cereal Chemists Inc. St. Paul (2001).

[88] Lopes, N. P. "Fases Estacionárias de Sílica e Polibutadieno para Cromatografia Líquida de Alta Eficiência". Tese de doutorado em Ciências, Instituto de Química, Universidade Estadual de Campinas (2004).

[89] Pedjie, N. "Analysis of color additives in beverages with the PerkinElmer Flexar FX-15 system equipped with a PDA detector". Application Note. PerkinElmer, Waltham, Disponível em: [http://www.perkinelmer.com/CMSResources/Images/44-134896APP\\_Food\\_Additives\\_in\\_Beverages.pdf](http://www.perkinelmer.com/CMSResources/Images/44-134896APP_Food_Additives_in_Beverages.pdf) . Acesso em 09/2015.

- [90] Snyder, L.R.; Kirkland, J.J. "Introduction to Modern Liquid Chromatography"; 2<sup>nd</sup> edition. Wiley Interscience, New York (1979).
- [91] Ribani, M.; Bottoli, C.B.; Collins, C.H.; Jardim, I.C.S.F.; Melo, L.F.C. "Validação em métodos de separação". Química Nova, 27, 771-780 (2004).
- [92] Lanças, F.M. "Validação de Métodos Cromatográficos de Análise". Editora RiMa, São Carlos (2004).
- [93] Secretaria de Defesa Agropecuária. Ministério da Agricultura, Pecuária e Abastecimento. "Manual de Garantia da Qualidade Analítica: Resíduos e Contaminantes em Alimentos" (2011).
- [94] Instituto Nacional de Metrologia, Normalização e Qualidade Industrial - INMETRO. "Orientações sobre validação de métodos de ensaios químicos". DOQ-CGCRE-008, Revisão 04, 20. Rio de Janeiro (2011).
- [95] Ferreira, M.H.; Braga, J.W.B.; Sena, M.M. "Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy". Microchemical Journal, 109, 158–164 (2013).
- [96] Food and Agriculture Organization. "Validation of analytical methods for food control". Report of a Joint FAO/IAEA Expert Consultation, FAO Food and Nutrition Paper n° 68, Roma, Itália (2000).
- [97] Thompson, M.; Ellison, S.L.R.; Wood, R. "Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report)". Pure and Applied Chemistry, 74, 835–855 (2002).
- [98] AOAC (Association of Official Analytical Chemists). AOAC Official Methods of analysis. Guidelines for Standard Method Performance Requirements. Washington: AOAC (2012).

[99] Ferrão, M.F.; Mello, C.; Borin, A.; Maretto, D.A.; Poppi, R.J. "LS-SVM: Uma nova ferramenta quimiométrica para regressão multivariada. Comparação de modelos de regressão LS-SVM e PLS na quantificação de adulterantes em leite em pó empregando NIR". *Química Nova*, 30, 852-859 (2007).

## Anexo A

Especificação comercial das 123 amostras de bebidas analisadas e valores de concentrações de AC e TA determinados por CLAE.

Tipo	AC (mg.L <sup>-1</sup> )	TA (mg.L <sup>-1</sup> )
Isotônico tangerina-cravo TAEQ	8,90	0,00
Isotônico laranja Power ADE	6,34	10,68
Isotônico limão Power ADE	0,00	4,00
Bebida adoçada de frutas cítricas MATE COURO	3,14	33,53
Bebida de limão adoçada DEL VALLE FRUT	0,00	0,59
Bebida mista de maçã tangerina DEL VALLE FRUT	8,63	0,82
Refrigerante laranja-mini SCHIN	16,26	0,00
Refrigerante maracuja FANTA	3,30	2,33
Refrigerante cítrico de baixa caloria- SCHWEPPE	0,00	0,62
Refrigerante cítrico de baixa caloria- SCHWEPPE	0,00	0,61
Refrigerante cítrico - ANTARCTICA	0,00	0,54
Refrigerante cítrico CIASSIC DILLAR´s	0,00	0,39
Refrigerante laranja MATE COURO	39,34	0,00
Refrigerante Laranja DEL REIZINHO	36,70	8,71
Refrigerante laranja FANTA	20,18	0,00
Refrigerante laranja SUKITA	20,89	0,00
Refrigerante laranja MINALBA	17,75	16,30
Refrigerante Maracujá FANTA	3,61	2,73
Isotônico limão POWER ADE	0,00	4,00
Isotônico limão POWER ADE	0,00	4,14
Isotônico Tangerina GATORADE	10,41	0,00
Isotônico Maracujá GATORADE	1,82	3,41
Isotônico Maracujá GATORADE	1,45	3,09
Isotônico laranja GATORADE	3,30	9,43
Isotônico Maracujá leve POWER ADE	2,82	2,31
Isotônico Maracujá leve POWER ADE	2,25	2,05
Isotônico Tangerina leve POWER ADE	8,04	0,66
Isotônico Tangerina leve POWER ADE	8,54	9,56
Isotônico limão siciliano ATHLÉTICA	0,00	0,07
Isotônico com água de coco-DUCOCO	1,20	4,50
Isotônico laranja POWER ADE	5,40	10,56

Isotônico Maracujá leve POWER ADE	2,96	2,60
Isotônico limão POWER ADE	0,00	2,94
Isotônico Maracujá MARATHON	1,82	5,83
Refrigerante Maracujá FANTA	2,85	2,59
Refrigerante citrus ANTARTICA	0,00	0,41
Refrigerante citrus SCHWEPPEES	0,00	0,48
Refrigerante citrus SCHWEPPEES	0,00	0,47
Bebida de limão adoçada DEL VALLE FRUT	0,00	0,44
Bebida Mista abacaxi e hortelã adoçada- GRAPETTICO	0,00	3,38
Isotônico limão POWER ADE	0,00	4,65
Isotônico limão POWER ADE	0,00	5,10
Refrigerante Citrus SCHWEPPEES	0,00	0,74
Refrigerante Citrus SCHWEPPEES	0,00	0,56
Refrigerante Citrus SCHWEPPEES	0,00	0,47
Refrigerante Citrus SCHWEPPEES	0,00	0,55
Refrigerante Citrus ANTARCTICA	0,00	0,38
Refrigerante sabor laranja MINI-SCHIN	18,77	0,00
Refrigerante sabor laranja SUKITA	21,95	0,00
Refrigerante sabor laranja SUKITA	22,99	0,00
Refrigerante misto levemente gaseificado H2OH!	7,12	0,00
Isotônico Tangerina GATORADE	12,47	0,00
Isotônico maracujá GATORADE	2,07	3,80
Isotônico laranja POWER ADE	6,74	16,03
Isotônico maracujá leve BY POWER ADE	2,70	2,23
Isotônico laranja GATORADE	4,11	10,39
Isotônico laranja POWER ADE	6,39	11,76
Isotônico laranja POWER ADE	7,10	12,14
Isotônico com água de coco- DUCOCO SPORT	0,34	2,40
Isotônico laranja POWER ADE	6,76	12,41
Isotônico laranja POWER ADE	7,15	14,11
Isotônico frutas cítricas FRUKITO	6,86	14,02
Bebida mista BIOLEVE	3,55	14,70
Bebida mista adoçada -TAMPICO	8,95	23,56
Bebida mista adoçada -FRUTY SUK	1,31	6,26
Isotônico Tangerina leve BY POWER ADE	9,53	0,72
Isotônico Tangerina leve BY POWER ADE	9,56	0,70
Isotônico Tangerina leve BY POWER ADE	9,46	0,66
Isotônico laranja POWER ADE	6,99	12,73
Isotônico maracujá GATORADE	2,15	3,93

Isotônico maracujá GATORADE	2,20	3,66
Isotônico limão Power ADE	0,00	4,25
Isotônico limão Power ADE	0,00	4,26
Refrigerante laranja FANTA	24,02	0,00
Refrigerante laranja SUKITA	23,34	0,00
Refrigerante laranja SUKITA	22,53	0,00
Refrigerante laranja SUKITA	21,94	0,00
Refrigerante misto limão e laranja H2OH!	7,46	0,00
Refrigerante misto limão e laranja H2OH!	7,17	0,00
Isotônico citrus SARANDI	4,24	13,19
Isotônico Laranja POWER ADE	6,30	11,80
Isotônico tangerina POWER ADE	11,43	12,61
Isotônico Laranja POWER ADE	6,71	12,35
Isotônico Laranja POWER ADE	6,39	12,05
Isotônico tangerina POWER ADE	11,00	12,61
Isotônico Tangerina leve BY POWER ADE	2,63	0,71
Isotônico Laranja POWER ADE	7,41	12,13
Isotônico Tangerina leve BY POWER ADE	9,11	0,65
Isotônico Laranja POWER ADE	6,62	12,14
Refrigerante Laranja SUKITA	23,33	0,00
Refrigerante Laranja MINI SCHIN	26,14	0,00
Refrigerante Laranja DEL REY REIZINHO	33,94	8,68
Refrigerante Laranja SUKITA	22,49	0,00
Refrigerante Laranja MINI SCHIN	21,27	0,00
Refrigerante Laranja DEL REY REIZINHO	33,48	8,44
Refrigerante Laranja MATE COURO	38,17	0,00
Refrigerante Laranja SUKITA	21,83	0,00
Refrigerante Laranja SUKITA	21,69	0,00
Refrigerante maracujá FANTA	2,80	2,36
Refrigerante maracujá FANTA	2,83	2,58
Refrigerante citrus light SCHWEPES	0,00	0,57
Refrigerante de laranja zero FANTA	27,89	0,00
Refrigerante de laranja SUKITA	25,33	0,00
Refrigerante de laranja SUKITA	23,15	0,00
Refrigerante de laranja FANTA	23,29	0,00
Refrigerante de laranja FANTA	25,48	0,00
Refrigerante de laranja DEL REY	36,01	9,46
Refrigerante citrus ANTARCTICA	0,00	0,41
Refrigerante citrus ANTARCTICA	0,00	0,42

Refrigerante citrus SCHWEPPEES	0,00	0,63
Refrigerante citrus ANTARCTICA	0,00	0,42
Refrigerante de laranja FANTA	20,21	0,00
Refrigerante de laranja SUKITA	24,04	0,00
Refrigerante de laranja FANTA	20,97	0,00
Refrigerante de laranja FANTA	22,85	0,00
Refrigerante de laranja SUKITA	6,16	0,00
Refrigerante de laranja FANTA	20,98	0,00
Refrigerante de laranja FANTA	22,02	0,00
Refrigerante de laranja SUKITA	21,92	0,00
Refrigerante de laranja SUKITA	20,54	0,00
Refrigerante de laranja FANTA	21,73	0,00
Refrigerante de laranja FANTA	20,87	0,00
Refrigerante de laranja SUKITA	18,68	0,00