



**UNIVERSIDADE DE BRASÍLIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**Tese de Doutorado**

**APLICAÇÃO DE MÉTODOS MULTIVARIADOS PARA A ANÁLISE E CORRELAÇÃO  
DE AMOSTRAS APREENDIDAS DE COCAÍNA**

**Área: Química Analítica**

**Linha de Pesquisa: Métodos de Análise**

**Tatiane Souza Grobério**

**Orientador: Jez Willian Batista Braga**

**Coorientador: Jorge Jardim Zacca**

**Brasília – DF**

**2017**



**UNIVERSIDADE DE BRASÍLIA**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

**Tese de Doutorado**

**APLICAÇÃO DE MÉTODOS MULTIVARIADOS PARA A ANÁLISE E CORRELAÇÃO  
DE AMOSTRAS APREENDIDAS DE COCAÍNA**

*Tese apresentada ao Curso de Doutorado em  
Química, da Universidade de Brasília, como  
requisito para a obtenção do título de Doutor em  
Química.*

**Tatiane Souza Grobério**

**Orientador: Jez Willian Batista Braga**

**Coorientador: Jorge Jardim Zacca**

**Brasília – DF**

**2017**

## Folha de Aprovação

Comunicamos a aprovação da Defesa de Tese de Doutorado do (a) aluno (a) **Tatiane Souza Grobério**, matrícula nº **13/0085979**, intitulada ***“Aplicação de métodos multivariados para a análise e correlação de amostras apreendidas de cocaína”***, apresentada no (a) Auditório Azul do Instituto de Química (IQ) da Universidade de Brasília (UnB) em 23 de janeiro de 2017.

Prof. Dr. Jez Willian Batista Braga  
Presidente de Banca (IQ/UnB)

Prof. Dr. Fábio Moreira da Silva  
Membro Titular (IQ/UnB)

Dra. Lívia de Barros Salum  
Membro Titular (PCDF)

Prof. Dr. Fabrício Machado Silva  
Membro Titular (IQ/UnB)

Prof.<sup>a</sup> Dra. Marly Eiko Osugi  
Membro Titular (IQ/UnB)

Prof.<sup>a</sup> Dra. Valéria Regina Bellotto  
Membro Suplente (IQ/UnB)

Em 23 de janeiro de 2017.

## **Agradecimentos**

A Deus por me fortalecer em momentos difíceis, mostrando-me o caminho nas horas incertas e suprimindo todas as minhas necessidades.

Aos meus pais, Carlos e Ester, que tiveram paciência e me incentivaram durante toda minha formação. Sem eles nada disso seria possível.

Ao meu marido, que traz palavras de ânimo e conforto quando a luz no fim do túnel parece se findar.

Ao meu irmão e cunhada que trouxeram muita alegria com suas filhas Ana Clara, Maria Eduarda e Sofia.

Aos demais familiares que me apoiaram durante essa jornada.

Ao meu orientador, Jez Willian, e coorientador, Jorge Zacca, que com respeito, amor, paciência e dedicação me direcionaram e ensinaram.

Aos peritos, estagiários e colegas mestrando que comigo estiveram trazendo alegrias, companhia e compartilhando experiências. Dentre eles cito o Maurício Leite, o Adriano Maldaner e o Márcio Talhavini, que sempre estiveram presentes me auxiliando nos grandes furos provocados pela falta de experiência e conhecimento.

Ao Conselho Nacional de Pesquisa - CNPq - que me concedeu uma bolsa, durante a realização deste doutorado, fato este que muito contribuiu para a viabilização desta tese.

## Sumário

Introdução .....	1
Capítulo 1: Discriminação e determinação do teor de cocaína em amostras apreendidas de drogas por espectroscopia no infravermelho e regressão por mínimos quadrados parciais .....	4
Objetivos .....	5
1. Referências Bibliográficas .....	5
1.1. Forma de apresentação da cocaína .....	5
1.2. Adulterantes .....	7
1.3. Grau de oxidação .....	8
1.4. Métodos analíticos para análise de cocaína .....	10
2. Fundamentação Teórica .....	12
2.1. Regressão por mínimos quadrados parciais – PLSR .....	12
2.2. Figuras de mérito .....	17
2.3. Transferência de calibração .....	19
3. Materiais e Métodos .....	21
3.1. Amostras e preparação de amostras .....	21
3.1.1. Amostras utilizadas na transferência do modelo de calibração .....	22
3.2. Ensaio de solubilidade .....	22
3.3. Análise de cromatografia gasosa com detector de ionização por chama .....	22
3.4. Medidas de espectroscopia de infravermelho .....	23
3.4.1. Medidas de espectroscopia de infravermelho para a transferência de calibração .....	23
3.5. Análise multivariada .....	23
4. Resultados e Discussão .....	25
4.1. Teste de Solubilidade .....	25
4.2. Análise de cromatografia gasosa com detector de ionização por chama .....	25
4.3. Dados de infravermelho .....	28
4.4. Discriminação de amostras de acordo com a forma de apresentação da cocaína .....	29

4.5. Determinação do teor de cocaína em amostras apreendidas .....	31
4.5.1. Transferência do modelo de determinação do teor de cocaína .....	34
4.6. Determinação do grau de oxidação .....	36
4.7. Determinação de adulterantes em amostras de cocaína apreendidas .....	38
5. Conclusão .....	40
Capítulo 2: Correlação entre amostras de cocaína base livre .....	42
Objetivos .....	43
1. Referências bibliográficas .....	43
1.1. Processo de produção de cocaína .....	43
1.2. Metodologia para análise de solventes residuais em drogas ilícitas .....	45
2. Fundamentação teórica .....	49
2.1. Método ULT .....	49
2.1. Função cosseno .....	52
2.2. Análise hierárquica de agrupamentos – HCA .....	53
2.3. Correlações intra e interclasses .....	55
2.4. Histogramas .....	55
2.5. ROC e AUC .....	57
3. Materiais e métodos .....	60
3.1. Amostras e preparação de amostras .....	60
3.2. Procedimentos experimentais .....	61
3.3. Precisão do Método .....	61
4. Resultados .....	62
4.1. Análises preliminares das amostras .....	62
4.2. Análises dos dados .....	66
4.3. Conjunto controlado de amostras .....	70
4.4. Determinação dos valores limites de $\rho$ .....	75
4.5. Precisão do método .....	81
4.6. Aplicações para casos reais .....	83
4.7. Comparação do método ULT desenvolvido utilizando cocaína sal e cocaína base livre .....	85
5. Conclusão .....	89

Referências .....	91
-------------------	----

### Índice de Figuras

Figura 1: Equilíbrio ácido-base entre cocaína na forma de base livre e sal.....	6
Figura 2: Estrutura química de alguns adulterantes encontrados na cocaína.....	8
Figura 3: Processo de purificação (oxidação) da cocaína com $KMnO_4$ .....	9
Figura 4: Estrutura química da cinamoilcocaína.....	9
Figura 5: Decomposição em variáveis latentes das matrizes <b>X</b> e <b>Y</b> para modelos PLSR.....	13
Figura 6: Fluxograma simplificado com as etapas envolvidas na transferência de calibração.....	21
Figura 7: Passos para a construção de modelos de calibração multivariada.....	24
Figura 8: Histograma de concentrações de cocaína nas amostras apreendidas.....	25
Figura 9: Grau de oxidação de amostras apreendidas: (A) de base livre de cocaína, (B) de cloridrato de cocaína.....	26
Figura 10: Número de adulterantes encontrados em amostras de cocaína apreendidas pela Polícia Federal Brasileira.....	27
Figura 11: Espectros ATR-FTIR de todas as amostras de cocaína analisadas.....	28
Figura 12: Principais diferenças entre os espectros ATR-FTIR de cocaína de base livre e de cocaína sal. Diferenças (A) perceptíveis e (B) não perceptíveis visualmente.....	29
Figura 13: Valores de classe estimados para o modelo de classificação final PLS-DA. Valor de classe = 1: cocaína base livre ( $\square$ ). Valor de classe = 0: cocaína sal ( $\diamond$ ). Amostras de calibração (símbolo vazio) e amostras de validação (símbolo preenchido).....	31
Figura 14: Valores de referência versus valores estimados pelo modelo PLS para a determinação do teor de cocaína. Amostras de calibração ( $\circ$ ) e amostras de validação ( $\blacktriangle$ ).....	33
Figura 15: Distribuição dos erros absolutos (resíduos) para o modelo PLSR versus teor estimado de cocaína. Amostras de calibração ( $\circ$ ) e validação ( $\blacktriangle$ ).....	33
Figura 16: Intervalos de confiança para algumas amostras de validação, com 95% de confiança obtidos para o modelo PLSR.....	34

Figura 17: (A) Valores de referência versus valores estimados do modelo PLSR para o grau de oxidação. (B) Distribuição dos erros absolutos (resíduos) do modelo PLSR para o grau de oxidação. Amostras de calibração (○) e validação (▲).....	37
Figura 18: Valores de referência versus valores estimados para a determinação de adulterantes pelo modelo PLS. Amostras de calibração (○) e validação (▲).....	38
Figura 19: Processo de produção de cocaína.....	44
Figura 20: Fluxograma do Método ULT para determinação dos limites entre amostras correlacionadas e não correlacionadas.....	51
Figura 21: Construção de vetores a partir do cromatograma e comparação de vetores característicos de duas amostras.....	52
Figura 22: Identificação do valor limite de $\rho$ que delimita amostras não correlacionadas de correlacionadas.....	47
Figura 23: Correlações de amostras (A) intra e (B) interclasses.....	55
Figura 24: Histograma de amostras ligadas e não ligadas, erros falsos positivos e falsos negativos, zona de sobreposição entre $\rho_L$ e $\rho_{NL}$ .....	56
Figura 25: (A) Exemplo de curva ROC obtido a partir de um histograma com um número de classes específico e valores progressivos de $\rho$ . 1, 2 e 3 compõem a curva ROC e representam taxa de VP em função de FP para três valores de $\rho$ diferentes. (B) Exemplo de área sob a curva ROC.....	57
Figura 26: (A) Histograma ilustrando os valores limites para a classificação de amostras e (B) caracterização dos falsos positivos (FP), falsos negativos (FN), zona de sobreposição e região de incerteza, onde haverá dúvida sobre a classificação de amostras.....	59
Figura 27: Distribuição das amostras de cocaína base analisadas por HS-CG-EM em relação aos estados do Brasil.....	60
Figura 28: Amostras que apresentaram correlação e foram apreendidas em diferentes ocasiões no Acre em 2010.....	63
Figura 29: Dendrograma identificando os seis grupos de amostras.....	64
Figura 30: Porcentagem de amostras em cada grupo formado pela HCA.....	64
Figura 31: Principais perfis químicos encontrados nos grupos de amostras separadas por HCA.....	65
Figura 32: Cromatograma de uma amostra apreendida em Mato Grosso em 2010.....	66
Figura 33: Cromatogramas de amostras selecionadas para testar a análise de dados.....	68

Figura 34: Valores máximos, mínimos e de média para correlações ( $\rho$ ) entre a amostra de cocaína sal.....	69
Figura 35: Valores máximos, mínimos e de média para correlações ( $\rho$ ) entre duas amostras de cocaína base livre.....	70
Figura 36: (A) Valores de Porcentagem da diferença para o conjunto controlado de amostras mostrando em (B) a variação entre 24 a 40 classes.....	74
Figura 37: Histogramas do conjunto controlado considerando o número de classes igual a 28. Onde a curva vermelha é de correlações entre amostras não correlacionadas e a curva azul é de correlações entre amostras correlacionadas.....	75
Figura 38: Dendrograma obtido pela análise dos espectros de massa acumulados das 619 amostras de cocaína base livre apreendidas pela Polícia Federal.....	76
Figura 39: (A) Valores de Porcentagem da diferença para o conjunto de amostras reais mostrando em (B) a variação entre 140 a 151 classes.....	78
Figura 40: Histogramas para (A) 100, (B) 145 e (C) 180 classes.....	79
Figura 41: Valores Limites com taxa de Falsos Positivos de 0,5% e Falsos Negativos de 2%.....	80
Figura 42: Cromatogramas de amostras selecionadas para avaliar a precisão do método.....	81
Figura 43: Cosseno quadrado ( $\rho$ ) obtido com amostras utilizadas para avaliar a precisão do método.....	83
Figura 44: Caso 1 - Apreensões correlacionadas feitas pela Polícia Federal de amostras de cocaína base livre.....	84
Figura 45: Caso 2 – Cromatogramas das amostras pertencentes às apreensões feitas pela Polícia Federal em Mato Grosso do Sul (na cor cinza) e São Paulo (na cor preta).....	85
Figura 46: Valores de porcentagem da diferença para o conjunto de amostras analisadas pelo método ULT-SAL.....	88

### **Índice de Tabelas**

Tabela 1. Características físico-químicas das formas de apresentações de cocaína.....	6
Tabela 2. Significado do grau de oxidação de uma amostra em termos da proporção do teor de cinamoilcocaínas e cocaína.....	10
Tabela 3. Adulterantes encontrados em amostras de cocaína apreendidas pela Polícia Federal Brasileira.....	27

Tabela 4. Efeito da eliminação de <i>outliers</i> em cada etapa de otimização do modelo de discriminação.....	30
Tabela 5. Efeito da eliminação de <i>outliers</i> em cada etapa de otimização do modelo de determinação do teor de cocaína.....	31
Tabela 6. Figuras de mérito analíticas para os modelos PLSR de determinação do teor de cocaína.....	32
Tabela 7. RMSEP dos modelos construídos com dados obtidos antes e após manutenção do equipamento Nicolet iS10 pertencente ao INC.....	35
Tabela 8. RMSEP dos modelos construídos com dados obtidos dos equipamentos Nicolet iS10 e Jasco 4100.....	36
Tabela 9. Efeito da eliminação de <i>outliers</i> em cada etapa de otimização do modelo de determinação do teor de cocaína.....	36
Tabela 10. Figuras de mérito analíticas do modelo PLSR para o grau de oxidação.....	37
Tabela 11. Figuras de mérito analíticas dos modelos PLSR para a determinação de adulterantes.....	39
Tabela 12. Porcentagem de amostras que podem ser avaliadas pelos modelos considerando o MDC.....	40
Tabela 13. Valores médios e mínimos de $\rho$ entre amostras pertencentes de uma mesma apreensão para o conjunto controlado utilizando o espectro de massas acumulado.....	71
Tabela 14. Valores de AUC para os diversos números de classes estudados no Conjunto Controlado.....	73
Tabela 15. Valores de AUC, percentagem da diferença, $\rho_L$ e $\rho_{NL}$ para os diversos números de classes estudados no conjunto total de amostras analisadas de cocaína base livre.....	77
Tabela 16. Variação de falsos positivos (FP), mantendo $\rho_L=0,903$ , ou seja, nenhum falso negativo.....	79
Tabela 17. Variação de falsos negativos (FN), mantendo $\rho_{NL}=0,997$ , ou seja, nenhum falso positivo.....	80
Tabela 18. Resultados do parâmetro de repetibilidade.....	82
Tabela 19. Resultados do parâmetro de precisão intermediária.....	82

Tabela 20. Valores de AUC, percentagem da diferença,  $\rho_L$  e  $\rho_{NL}$  para os diversos números de classes estudados no conjunto de amostras analisadas pelo método ULT-SAL.....87

Tabela 21. Variação de da taxa de falsos positivos (FP), mantendo  $\rho_L=0,934$ , ou seja, nenhum falso negativo para o conjunto de dados de cocaína sal.....89

Tabela 22. Variação de falsos negativos (FN), mantendo  $\rho_{NL}=0,994$ , ou seja, nenhum falso positivo para o conjunto de dados de cocaína sal.....89

## Lista de Abreviaturas e Acrônimos

- ATR: reflexão total atenuada, do inglês *Attenuated Total Reflectance*
- AUC: área sob a curva ROC, do inglês *area under the curve*
- CG-EM: cromatografia gasosa com detecção por espectrometria de massas
- CG-DIC: cromatografia gasosa com detector de ionização em chama
- COW: correlação deformação otimizada, do inglês *correlation optimized warping*
- CV: Validação Cruzada, do inglês *Cross Validation*
- DPV: desvio padrão de erros de validação
- DS: Padronização Direta, do inglês *Direct Standardization*
- FTIR: espectroscopia de infravermelho por transformada de Fourier, do inglês *Fourier Transform Infrared Spectroscopy*
- FTIR-ATR: espectroscopia de infravermelho por transformada de Fourier com acessório de reflectância total atenuada, do inglês *Fourier Transform Infrared Spectroscopy – Attenuated Total Reflectance*
- FN: falso negativo
- FP: falso positivo
- IC: Intervalo de confiança
- INC: Instituto Nacional de Criminalística
- HCA: análise hierárquica de agrupamentos, do inglês *Hierarchical Cluster Analysis*
- HS-CG-DIC: *headspace* acoplada com cromatografia gasosa com detector de ionização em chama
- HS-CG-EM: *headspace* acoplada com cromatografia gasosa com detecção por espectrometria de massas
- MDC: Concentração mínima detectável, do inglês *minimum detectable concentration*
- MEFS-CG-EM: microextração em fase sólida acoplada com cromatografia gasosa com detecção por espectrometria de massas
- OSC: Correção do Sinal Ortogonal, do inglês: *Orthogonal Signal Correction*
- PCA: Análise de Componentes Principais, do inglês *Principal Component Analysis*
- PDS: Padronização Direta por Partes, do inglês *Piecewise Direct Standardization*
- PEQUI: Projeto de Perfil Químico de Drogas
- PF: Polícia Federal Brasileira
- PLS-DA: regressão por mínimos quadrados parciais para análise discriminante, do inglês *Partial Least Squares for Discriminant Analysis*
- PLSR: regressão por mínimos quadrados parciais, do inglês *Partial Least Squares Regression*

RMSECV: raiz do erro quadrático médio de validação cruzada, do inglês *Root Mean Square Error of Cross Validation*

RMSEP: raiz quadrada do erro médio de previsão, do inglês *Root Mean Square Error of Prediction*

ROC: curva ROC, do inglês *Receiving Operating Characteristics*

SEN: Sensibilidade

SNV: Variação Padrão Normal, do inglês: *Standard Normal Variate*

SWGDRUG: Grupo de Trabalho Científico de Análise de Drogas Apreendidas, do inglês: *Scientific Working Group for the Analysis of Seized Drugs*

TIC: cromatograma de íons totais, do inglês *total ion chromatogram*

TR: tempo de retenção

UNODC: Escritório das Nações Unidas contra as drogas e o crime, do inglês *United Nations Office on Drugs and Crimes*

ULT: método não Supervisionado de Limite de Correlação, do inglês *Unsupervised Linkage Threshold*

UPGMA: método de grupo por par não ponderado com média aritmética, do inglês *unweighted pair group method with arithmetic mean*

UPGMC: método centroide por par não ponderado, do inglês *unweighted pair group method with centroid mean*

VL: variáveis latentes

VN: verdadeiro negativo

VP: verdadeiro positivo

WPGMA: método de grupo por par ponderado com média aritmética, do inglês *weighted pair group method with arithmetic mean*

WPGMC: método centroide por par ponderado, do inglês *weighted pair group method with centroid mean*

## Resumo

É sabido que o abuso de drogas ilícitas constitui um problema social sério e de difícil solução. Dentre as drogas de abuso, a cocaína tem ocupado papel de destaque. A fim de tentar controlar e reduzir a disponibilidade da cocaína no mercado ilícito, diversos organismos de segurança pública nacionais e internacionais buscam estratégias para estabelecer a origem, principais métodos de refino e rotas mais comuns do tráfico de drogas. Tais objetivos podem ser alcançados por meio de investigações policiais e quimicamente através de análises que caracterizem a droga. A caracterização de uma amostra de cocaína é o processo que determina as suas principais propriedades físicas e químicas, revelando similaridades e diferenças entre apreensões.

Neste trabalho, a caracterização química da amostra iniciou-se pela determinação da forma de apresentação da cocaína (sal ou base livre), do teor de cocaína, adulterantes e identificação do grau de oxidação em amostras de drogas utilizando espectroscopia de infravermelho por transformada de Fourier com acessório de reflexão total atenuada (FTIR-ATR, do inglês *Fourier Transform Infrared Spectroscopy – Attenuated Total Reflectance*), que é uma técnica mais simples, rápida, com maior facilidade de aplicação e menor custos que cromatografia gasosa, técnica utilizada em métodos de referência atualmente utilizado pela Polícia Federal Brasileira (PF). Esse método foi desenvolvido e validado com 1085 amostras reais de drogas apreendidas pela PF em todo o Brasil, das quais 500 foram selecionadas para o conjunto de calibração e 585 para o conjunto de validação para uma análise multivariada de regressão por mínimos quadrados parciais (PLSR, do inglês *Partial Least Squares Regression*). Obteve-se um erro médio de previsão de 3,0% (m/m), precisão de 2,0 e concentração mínima detectável de cocaína de 11,8% (m/m), em uma faixa de concentração variando de 24,2 a 99,9% (m/m). Os resultados indicaram que o método desenvolvido é capaz de discriminar entre o cloridrato de cocaína e as amostras de base livre, quantificar o conteúdo de cocaína e estimar a concentração dos principais adulterantes.

Em uma segunda fase do trabalho, a caracterização química das amostras foi utilizada para identificar correlações entre apreensões de cocaína base livre sem o conhecimento *a priori* sobre a origem das amostras. Para isso, buscou-se determinar os solventes residuais com o uso do headspace associado à cromatografia gasosa e detecção por espectrometria de massas (HS-CG-EM). A análise de dados empregou a análise hierárquica de agrupamento (HCA, do inglês *Hierarchical Cluster Analysis*) e o quadrado da função cosseno. Estudos estatísticos apontaram valores limites que puderam

discriminar amostras como sendo “ligadas”, ou seja, correlacionadas em relação à forma de produção, permitindo estabelecer ligações entre apreensões de cocaína efetuadas em todo o Brasil. A metodologia desenvolvida para essa discriminação pôde ser considerada promissora, pois não necessita do conhecimento *a priori* da população de amostras e permite o ajuste dinâmico dos valores limites para diferentes conjuntos.

**Palavras-chave:** cocaína, PLSR, FTIR, HS-CG-EM, correlação.

## Abstract

It is known that illicit drug abuse is a social problem and difficult to solve. Among the drugs of abuse, a cocaine has played a prominent role. In order to try to control and reduce the availability of cocaine in the illicit market, various public and international security agencies are looking for strategies to establish an origin, main refining methods and most common routes of drug trafficking. Such objectives can be seen through software investigations and analyze the aspects that characterize the drug. The characterization of a cocaine sample is the process that determines its main physical and chemical properties, revealing similarities and differences between seizures.

In this work, the chemical characterization of the sample was initiated by determining the form of presentation of cocaine (salt or free base), cocaine content, adulterants and identification of the degree of oxidation in drug samples using infrared spectroscopy (FTIR-ATR), which is a simpler, faster, easier to apply and lower cost than gas chromatography, a technique used in reference methods currently used by the Brazilian Federal Police (BPF). This method was developed and validated with 1085 real drug samples seized by PF throughout Brazil, of which 500 were selected for the calibration set and 585 for the validation set for a multivariate analysis of partial least squares regression (PLSR). An average prediction error of 3.0% (w/w), precision of 2.0 and minimum detectable concentration of cocaine of 11.8% (w/w) was obtained in a concentration range ranging from 24.2 to 99.9% (w/w). The results indicated that the method was able to discriminate between cocaine hydrochloride and free base samples, to quantify the cocaine content and to estimate the concentration of the main adulterants.

In a second phase of the study, the chemical characterization of the samples was used to identify correlations between free base cocaine seizures without a priori knowledge about the origin of the samples. The aim of this study was to determine residual solvents using headspace associated with gas chromatography and mass spectrometry detection (HS-CG-EM). The data analysis employed the hierarchical cluster analysis (HCA) and the square of the cosine function. Statistical studies pointed to limit values that could discriminate samples as being "linked", that is, correlated in relation to the form of production, allowing establishing links between seizures of cocaine made throughout Brazil. The methodology developed for this discrimination could be considered promising since it does not require a priori knowledge of the population of

samples and allows the dynamic adjustment of the limit values for different sets of samples.

**Keywords:** cocaine, PLSR, FTIR, HS-GC-MS, correlation.

# *Introdução*

É sabido que o abuso de drogas ilícitas constitui atualmente um dos problemas sociais mais sérios e de difícil solução para a administração pública. Este fenômeno afeta a segurança nacional e internacional, a estabilidade econômica e social e a saúde da população em geral. Em âmbito mundial, o que se vê é um aumento progressivo do número de usuários dessas substâncias e, conseqüentemente, uma evolução nas formas de produção e distribuição de entorpecentes.

Dentre as drogas de abuso, a cocaína tem ocupado papel de destaque. As estimativas mais recentes encontradas na literatura indicam que, no mundo, 18,8 milhões de pessoas usaram cocaína pelo menos uma vez em 2014, o equivalente a 0,37% da população global com idade entre 15 e 64 anos.<sup>1</sup> No Brasil, uma pesquisa realizada em 2012 constatou que a cocaína base livre é usada por cerca de 370 mil pessoas nas capitais e no Distrito Federal.<sup>2</sup> A fim de tentar controlar e reduzir a disponibilidade da cocaína no mercado ilícito, diversos organismos de segurança pública nacionais e internacionais buscam estratégias para estabelecer a origem, os principais métodos de refino e as rotas mais comuns de tráfico de drogas. Tais objetivos podem ser alcançados por meio de investigações policiais ou através de análises químicas que caracterizem a droga. A caracterização de uma amostra de apreensões contendo cocaína é o processo que determina as suas principais propriedades físicas e químicas, as quais podem ser estudadas para revelar similaridades e diferenças entre apreensões.<sup>3</sup>

A caracterização química de uma amostra de apreensões de cocaína deve abordar diversos aspectos. Dentre eles, a determinação da forma de apresentação da cocaína (sal ou base livre), o teor de cocaína e adulterantes, o grau de oxidação, a identificação do perfil químico de solventes residuais, dentre outros.

Tradicionalmente, a determinação da forma de apresentação da cocaína é feita por meio de ensaios de solubilidade realizados com água e éter etílico, seguida por precipitação com ácidos e bases. No entanto, a diversidade de misturas tais como medicamentos, diluentes e adulterantes presentes nas amostras de cocaína faz com que os resultados apresentados por esses testes sejam, em alguns casos, inconclusivos, sendo necessária a utilização de testes mais eficientes.<sup>4</sup>

A quantificação da cocaína, de adulterantes, da cinamoilcocaína e a determinação do grau de oxidação de uma amostra tem sido feita utilizando a cromatografia em fase gasosa, embora existam muitos métodos propostos para esta finalidade.<sup>5-7</sup> No entanto, apesar dos excelentes resultados, esta é uma técnica destrutiva, que utiliza alguns solventes tóxicos e requer etapas de preparação da amostra. Além disso, a instalação do

equipamento de cromatografia é dispendiosa e exige um treinamento longo para os profissionais que a utilizarão.<sup>8</sup> Por isso, diversos autores propuseram o uso da espectroscopia de infravermelho por transformada de Fourier com acessório de reflectância total atenuada, (FTIR-ATR, do inglês *Fourier Transform Infrared Spectroscopy – Attenuated Total Reflectance*) para determinar a forma de apresentação da cocaína e o teor de cocaína em uma amostra. E estudos recentes consideraram a combinação de espectroscopia vibracional e métodos quimiométricos uma alternativa promissora e simples para obter informação pela análise direta da amostra de cocaína.<sup>9,10</sup> Entretanto, os trabalhos já publicados apresentaram descrições simplistas para essas análises, pois utilizaram conjuntos de dados com um número reduzido de amostras, validações com poucas figuras de mérito e a caracterização das amostras sem incluir a determinação do teor de adulterantes e identificação do grau de oxidação das amostras, sendo que todos esses são fatores importantes para a caracterização das apreensões.

Já a determinação de solventes residuais em amostras de cocaína tem sido feita utilizando a técnica de *headspace* acoplado à cromatografia gasosa com detecção por espectrometria de massas (HS-CG-EM) e esta tem se mostrado eficiente. Os solventes residuais presentes em uma amostra dão indícios sobre as etapas do processo de produção da cocaína, pois esse processo requer o uso de solventes, tais como querosene, gasolina, óleo diesel, *thinner*, e esses deixam vestígios químicos característicos. Assim, cada amostra possui um perfil químico com diferentes tipos e quantidades de solventes residuais.<sup>11-13</sup> Esse perfil químico pode gerar dados de inteligência tática e estratégica para repressão ao tráfico de drogas, tais como: comparação de apreensões de cocaína para determinar se as mesmas foram derivadas de uma fonte comum e avaliação dos precursores e/ou intermediários químicos utilizados nos processos de manufatura/síntese da cocaína.<sup>14-15</sup> A utilização de ferramentas quimiométricas facilita a identificação de padrões e a construção de métodos de classificação em sistemas complexos.<sup>16</sup>

Com base nesses conceitos, esse trabalho foi dividido em dois capítulos. No primeiro, buscou-se apresentar uma metodologia validada e adequada à realidade brasileira que permita a análise rápida e confiável de amostras de cocaína utilizando FTIR-ATR e análise discriminante por mínimos quadrados parciais (PLS-DA, do inglês *Partial Least Squares for Discriminant Analysis*) e regressão por mínimos quadrados parciais (PLSR, do inglês *Partial Least Squares Regression*). O segundo capítulo buscou estabelecer uma metodologia que permita a correlações entre apreensões de cocaína na forma de base livre efetuadas em todo o Brasil utilizando HS-CG-EM.

# *Capítulo 1*

*Discriminação e determinação do teor de cocaína em amostras apreendidas de drogas por espectroscopia no infravermelho e regressão por mínimos quadrados parciais*

## **Discriminação e determinação do teor de cocaína em amostras apreendidas de drogas por espectroscopia no infravermelho e regressão por mínimos quadrados parciais**

A determinação do teor de cocaína, de seus diluentes e adulterantes tem um papel significativo na ciência forense, uma vez que fornece informações importantes sobre a forma de como a droga foi produzida, bem como sobre o modo ilegal de operação das redes de distribuição de drogas em uma determinada área.

Neste sentido, neste trabalho foram desenvolvidos métodos capazes de avaliar qualitativa e quantitativamente amostras reais de drogas de abuso apreendidas pela Polícia Federal Brasileira. Análises foram feitas utilizando FTIR-ATR, técnica mais simples que as atualmente utilizadas nos laboratórios forenses para a caracterização das amostras de cocaína, e análise multivariada.

### **Objetivos**

- Simplificar a caracterização de amostras apreendidas de cocaína em análises forenses de rotina, propondo um método que combine a identificação e quantificação dos principais componentes de amostras de cocaína, bem como a discriminação da forma de apresentação (sal ou base livre) através da determinação por FTIR-ATR, utilizando PLSR e PLS-DA.
- Verificar o funcionamento e robustez do método desenvolvido em duas novas situações: após manutenção do equipamento de FTIR-ATR no qual as amostras foram analisadas e em outro equipamento de FTIR-ATR de marca e modelo diferente. Para isso, utilizou-se a ferramenta de transferência de calibração.

## **1. Referências Bibliográficas**

### **1.1. Forma de Apresentação da Cocaína**

A cocaína pode apresentar-se em duas formas: base livre ou sal, como mostrado na Figura 1. Cada uma dessas formas apresenta propriedades físico-químicas (Tabela 1) e características fisiológicas distintas, mas isso não caracteriza que essas sejam drogas distintas.

A cocaína na forma de base livre pode ser convertida em sal ao reagir com soluções de ácidos orgânicos ou inorgânicos. Ou seja, a forma química de sal ou base livre da cocaína depende do pH do meio, que desloca o equilíbrio químico em direção de

uma forma ou outra. Em meio ácido, o equilíbrio é deslocado para a formação de cocaína sal. Em meio básico, esse equilíbrio encontra-se deslocado para a formação de cocaína base livre.

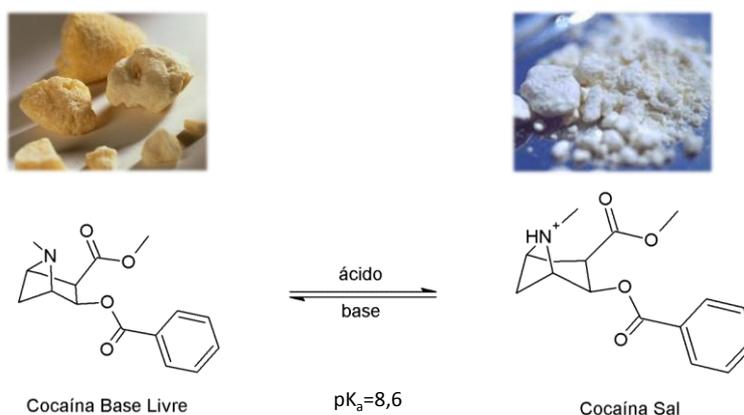


Figura 1: Equilíbrio ácido-base entre cocaína na forma de base livre e sal.

Tabela 1. Características físico-químicas das formas de apresentações de cocaína.

Forma de apresentação	Fórmula molecular	Massa Molar (g.mol <sup>-1</sup> )	P.F. (°C)	Características de Solubilidade			
				H <sub>2</sub> O	MeOH	Et <sub>2</sub> O	CHCl <sub>3</sub>
Cocaína Base Livre	C <sub>17</sub> H <sub>21</sub> NO <sub>4</sub>	303,4	98	I	S	S	S
Cocaína Sal	C <sub>17</sub> H <sub>21</sub> NO <sub>4</sub> .HCl	339,8	157	S	S	I	S

S, solúvel; I, insolúvel

FONTE: adaptada de Vargas & Talhavini.<sup>17</sup>

A cocaína sal é solúvel em água e normalmente obtida na forma de pó branco. Sua administração pode ser por aspiração ou via intravenosa. Já a cocaína base livre, no Brasil, apresenta-se como “pedra” de cor amarelada e é pouco solúvel em água, mas, quando aquecida, permite que seus vapores sejam inalados no ato de fumar.<sup>18</sup> Uma das formas de apresentação mais comuns da base livre no mercado ilícito é o chamado “*crack*”, que se apresenta sob forma de pedras obtidas por um processo de fusão da cocaína seguida de solidificação e fragmentação. A cocaína em pedras é assim vendida ilicitamente de modo a permitir que seja fumada em cachimbos pelos usuários da droga. O nome “*crack*” é originário dos Estados Unidos da América (década de 80) e deriva de uma expressão em inglês que significa “quebra” ou “fratura”. Naquele país, a cocaína é predominantemente traficada sob a forma de sal, de modo que grande parte do *crack* consumido localmente é proveniente da conversão da forma salina para a forma base livre com a utilização de substâncias alcalinas. Nos anos 80, era muito comum a utilização de carbonatos (de cálcio ou de sódio) para esse fim. Quando o *crack* assim produzido era fumado, o aquecimento

dos carbonatos remanescentes na droga propiciava a liberação de gás carbônico (CO<sub>2</sub>), responsável pela ruptura das pedras, e pela produção de sons (estalidos). Acredita-se que a fratura das pedras e a produção de estalos durante o ato de fumar tenha dado origem à denominação “*crack*”. Atualmente essa denominação é aplicada de forma generalizada a toda cocaína base livre sob a forma de pedras destinada ao fumo em cachimbos.

As diferentes formas de administração da cocaína resultam em diferentes velocidades de resposta no organismo humano. Quando aspirada, os efeitos começam em torno de três minutos, injetada os efeitos iniciam-se em aproximadamente um minuto e meio, e fumada os efeitos demoram apenas alguns segundos para aparecer.<sup>19</sup> A meia-vida plasmática (tempo gasto para que a concentração original de um fármaco no plasma sanguíneo se reduza à metade) da cocaína base após fumo é de 10-30 minutos. As aplicações intravenosas e por aspiração de cocaína sal provocam um efeito mais longo que o da cocaína base, cerca de 40 minutos. O declínio das concentrações plasmáticas promove a supressão do efeito e o reaparecimento do desejo de usar a droga.<sup>20</sup>

## **1.2. Adulterantes e Diluentes**

Drogas ilícitas podem ser comercializadas quase puras, adulteradas, diluídas ou contaminadas.<sup>21</sup> A adulteração e a diluição são fenômenos bem conhecidos e, em geral, amplamente difundidos no mercado ilícito de drogas. Têm como propósito o aumento da quantidade do produto final para que os lucros sejam maiores.<sup>22</sup>

Os adulterantes são substâncias que imitam ou potencializam o efeito da droga sobre o indivíduo.<sup>23</sup> Entre os adulterantes mais encontrados estão: cafeína (estimulante), fenacetina (antipirético e analgésico), levamisol (vermífugo que aparentemente potencializa os efeitos da cocaína), procaína (anestésico local), hidroxizina (anti-histamínico), lidocaína e benzocaína (anestésicos locais similares à cocaína).<sup>24</sup> Na Figura 2 são apresentadas as estruturas químicas desses principais adulterantes encontrados em amostras de cocaína.

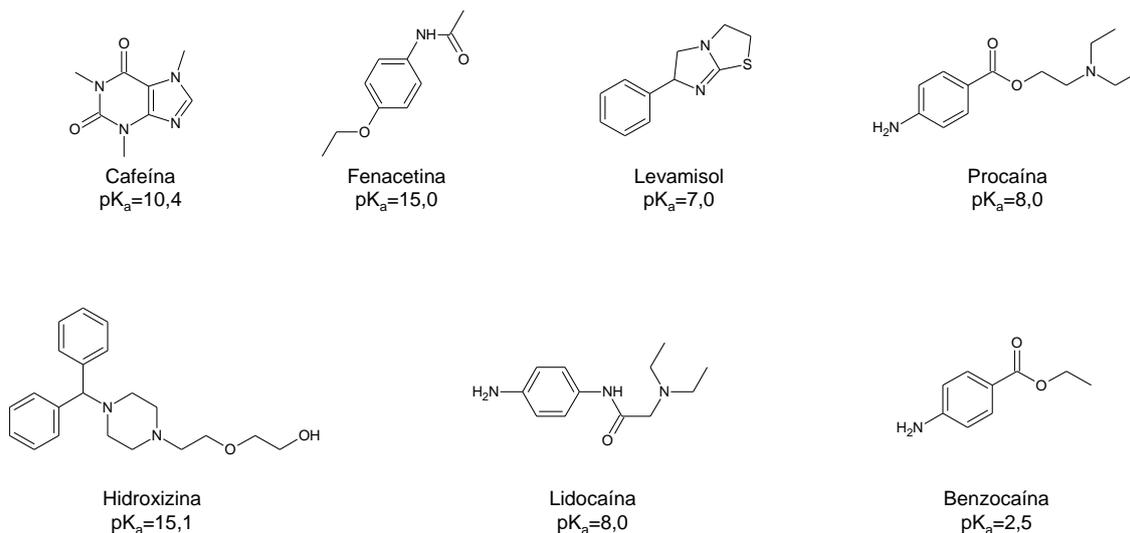


Figura 2: Estrutura química de alguns adulterantes encontrados na cocaína.

Os diluentes, diferentemente dos adulterantes, não possuem sinergismo ou efeitos similares ao efeito da cocaína. Os mais encontrados são amido, açúcares, carbonatos, bicarbonatos, ácido bórico e talco.<sup>25</sup>

A identificação e a quantificação de adulterantes e diluentes presentes em amostras de cocaína têm auxiliado no estabelecimento de ações normativas, a fim de restringir a venda de produtos encontrados em drogas de abuso<sup>26</sup>.

### 1.3. Grau de Oxidação

O grau de oxidação é um parâmetro que dá indícios sobre as etapas de purificação que foram empregadas durante a produção da cocaína. Após o processo de extração, a cocaína é usualmente submetida a processos de clareamento que envolvem limpeza com solventes (ex: lavagem com etanol) ou oxidação por intermédio de reagentes químicos específicos como o permanganato de potássio ( $KMnO_4$ ) – Figura 3. Essa oxidação tem por objetivo tornar o produto mais claro e atraente para os usuários. Nesse processo, há a eliminação de algumas impurezas e a redução do teor de cinamoilcocaínas na amostra.<sup>27</sup>

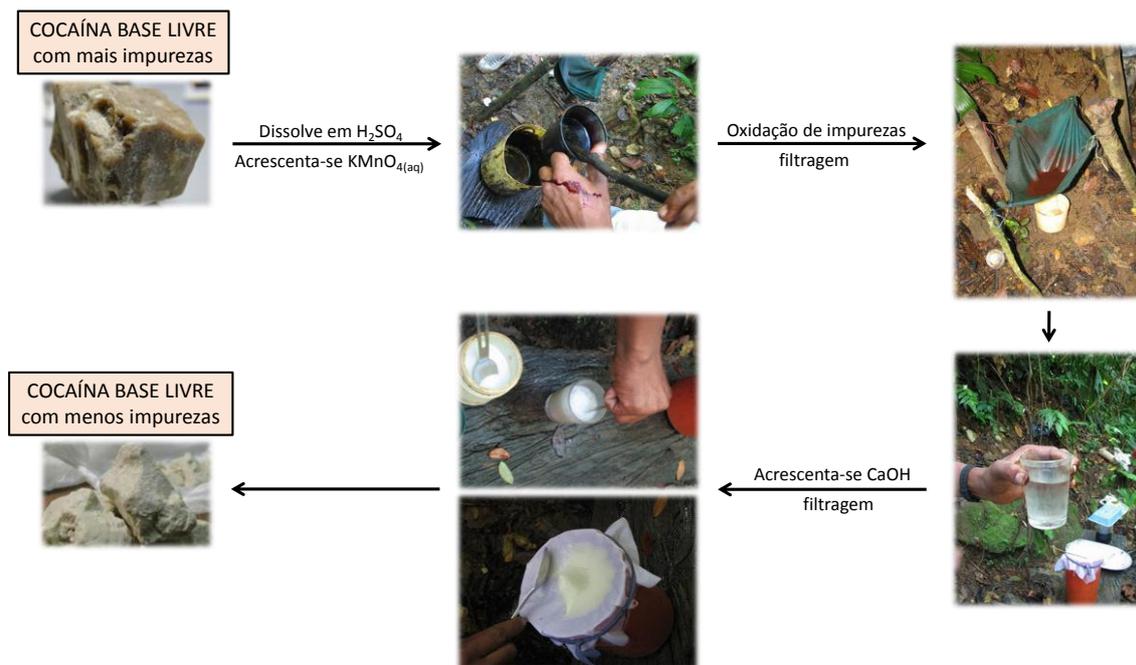
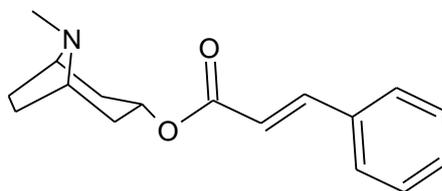


Figura 3: Processo de purificação (oxidação) da cocaína com  $KMnO_4$ .

As cinamoilcocaínas pertencem a um grupo de alcaloides encontrados somente em folhas de coca.<sup>28</sup> Quanto mais uma amostra é oxidada, menor é o teor de cinamoilcocaínas que ela apresenta. A Figura 4 mostra a estrutura da cinamoilcocaína.



Cinamoilcocaína

Figura 4: Estrutura química da cinamoilcocaína.

De acordo com Casale *et al.*<sup>29</sup>, o grau de oxidação da cocaína está relacionado à proporção entre o teor de cinamoilcocaína total (soma da cis e trans-cinamoilcocaína) e o teor de cocaína de uma amostra (Cin/Coc). Quando Cin/Coc é inferior a 2, a amostra apresenta um alto grau de oxidação. Quando este valor estiver entre 2 e 6, diz-se que a amostra apresenta um grau de oxidação médio. Se Cin/Coc é maior que 6, o grau de oxidação é considerado baixo, como mostrado na Tabela 2.

Tabela 2. Significado do grau de oxidação de uma amostra em termos da proporção do teor de cinamoilcocaínas e cocaína.

Cin/Coc	Intervalo de Cin/Coc	Grau de Oxidação
[ <i>cis</i> + <i>trans</i> cinamoilcocaína]	<2	Alto
[cocaína]	2>6	Médio
	>6	Baixo

O trabalho de Janzen<sup>30</sup> demonstrou que o conteúdo de *cis* e *trans* cinamoilcocaínas pode ser usado para fins de comparação de amostras, podendo, portanto, correlacionar apreensões policiais.

#### 1.4. Métodos Analíticos para Análise de Cocaína

Desde os anos 1970, os métodos analíticos utilizados para a determinação de cocaína em amostras de drogas de abuso têm evoluído substancialmente.<sup>31</sup> Diversas técnicas analíticas podem ser empregadas na caracterização química de drogas de abuso, dependendo do tipo de informação que se deseja obter. O uso de FTIR-ATR para determinar a forma de apresentação e o teor de cocaína em uma amostra é desejável pelo fato de a espectroscopia de infravermelho ser uma técnica rápida, sensível e por ser capaz de lidar com vários analitos em uma mistura. Atualmente, essa técnica já tem sido utilizada na análise preliminar de drogas de abuso<sup>32</sup> e na identificação de adulterantes e substâncias inorgânicas<sup>33</sup>. Sua combinação com quimiometria vem sendo considerada uma alternativa eficiente e com grande potencial para extrair informações qualitativas e quantitativas de diversos tipos de dados experimentais.<sup>10</sup>

Apesar do uso frequente da espectroscopia de infravermelho por transformada de Fourier (FTIR, do inglês *Fourier Transform Infrared Spectroscopy*), a cromatografia gasosa ainda ocupa lugar de destaque na análise de amostras de cocaína por fornecer resultados precisos e exatos.<sup>34</sup> Essa é a técnica amplamente utilizada pela criminalística em todo o Brasil e indicada como referência nos guias de análise do Escritório das Nações Unidas contra as drogas e o crime (UNODC, do inglês *United Nations Office on Drugs and Crimes*).<sup>35</sup> No entanto, apesar dos excelentes resultados, essa técnica normalmente requer um procedimento de preparação da amostra relativamente complexo, quando comparada às espectroscopias, resulta em custos elevados e demanda um tempo de análise mais alto. Além disso, a diversidade das amostras exige ajustes e atualizações do método.

Um dos primeiros métodos onde FTIR foi aplicado para a determinação de cocaína foi proposto em 1987 por Ravreby<sup>36</sup>. Nesse caso, as concentrações de cocaína sal

e heroína foram determinadas por regressão univariada. O autor estudou os efeitos de vários adulterantes e diluentes tais como amido, açúcares, manitol, cafeína e procaína. No entanto, considerando a alta complexidade das drogas apreendidas e a variabilidade dos espectros de FTIR obtidos com pastilhas de KBr, o método apresentou limitações para aplicação em análises forenses de rotina.

Em 1999, Ryder *et al.*<sup>37</sup> mostraram que os métodos de análise multivariada combinados com espectroscopia de Raman poderiam ser utilizados como um método analítico para a análise rápida de drogas ilícitas em misturas de dois componentes. Neste estudo, a análise quimiométrica por PLSR permitiu a estimativa da concentração de cocaína em misturas sólidas com glicose.

Em 2009, Maharaj<sup>38</sup> realizou um estudo preliminar comparando a quantificação de cocaína feita por FTIR com outra quantificação feita por cromatografia gasosa com detector de ionização em chama, CG-DIC. Embora apenas algumas amostras tenham sido utilizadas para análise e nenhuma figura de mérito tenha sido apresentada para avaliar o método, o autor afirmou que os resultados obtidos por FTIR e CG-DIC são equivalentes e considerou que a primeira técnica é mais adequada para a análise quantitativa de cocaína.

Em 2013, Rodrigues *et al.*<sup>10</sup> realizaram um estudo exploratório que caracterizou a composição química de 91 amostras de cocaína apreendidas no estado de Minas Gerais entre 2008 e 2010, com base em espectros FTIR e análise quimiométrica. Em seu estudo, tanto a Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*) quanto a PLS-DA foram utilizadas para classificar as amostras de acordo com a sua diluição (abaixo e acima de 15% (m/m)) e a forma química (sal de cocaína ou de base livre), obtendo-se, respectivamente, 83% e 97% dos resultados corretos. Dessa forma, esse estudo mostrou que FTIR aliada a quimiometria é adequada para análise de amostras de cocaína.

Alfonso Pérez *et al.*<sup>39</sup> em 2014 propuseram um método para a determinação de cocaína em amostras ilícitas por medidas de refletância difusa na região espectroscopia de infravermelho próximo. Os resultados obtidos pelos autores mostraram que a concentração de cocaína pode ser determinada em um vasto intervalo de concentrações. No entanto, a validação foi realizada com um conjunto relativamente pequeno de amostras quando comparada à elevada heterogeneidade das amostras que ocorrem em casos forenses reais.

Também em 2014, Conceição *et al.*<sup>23</sup> procuraram diferenciar cocaína sal de seus adulterantes utilizando FTIR-ATR. Isso porque adulterantes e diluentes podem, em alguns casos, fornecer resultados falso-positivos no teste de Scott, teste colorimétrico preliminar padrão utilizado para identificar a cocaína logo após apreensões policiais. Foram utilizadas 39 amostras de misturas binárias de cocaína com: fermento, leite em pó, lidocaína e prometazina (compostos que dão resultados positivos no teste de Scott). As amostras foram separadas aleatoriamente em dois conjuntos: calibração e predição, contendo respectivamente 31 e 8 amostras. Dessa forma, foi realizada a construção do modelo de PLSR. Dos resultados obtidos, foi concluído que o emprego da técnica espectroscópica de FTIR-ATR a misturas binárias seria uma boa alternativa para reduzir resultados inconclusivos ou resolver problemas de falso-positivos em testes preliminares como o Teste de Scott. Além disso, consideraram a utilização dessa técnica aliada à calibração multivariada PLSR como um método promissor para quantificar o teor de cocaína sal em amostras apreendidas pela polícia brasileira.

Os trabalhos apresentados indicam que o uso da combinação FTIR e PLSR traz uma série de vantagens, destacando-se a redução no tempo de análise, boa capacidade de identificar ou caracterizar estruturas complexas e não destruição da amostra. No entanto, os trabalhos já publicados apresentaram algumas limitações com relação ao número de amostras analisadas (no máximo 91 amostras), validação abordando poucas figuras de mérito e caracterização mais simplificada das amostras. Por isso, neste trabalho foi investido tempo e esforços para propor um método mais completo que pudesse ser utilizado em portos, aeroportos e lugares onde houvessem postos da Polícia Federal Brasileira e que necessitassem de caracterizar apreensões de cocaína, mas que não possuem laboratórios com equipamentos especializados para esse fim.

## **2. Fundamentação Teórica**

### **2.1. Regressão por Mínimos Quadrados Parciais – PLSR**

PLSR é um método de análise de dados multivariados e tem por principal finalidade construir um modelo linear capaz de quantificar compostos de interesse em um determinado tipo de amostras a partir de dados instrumentais. Esse modelo pode ser descrito pela igualdade:  $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$ , em que  $\mathbf{Y}$  é uma matriz ( $N \times M$ ), onde  $N$  é o número de amostras e  $M$  o número de propriedades de interesse,  $\mathbf{X}$  é uma matriz ( $N \times K$ ), onde  $K$  são variáveis de previsão,  $\mathbf{B}$  é uma matriz ( $K \times M$ ) dos coeficientes da regressão e  $\mathbf{E}$  é a matriz dos resíduos para o modelo que tem a mesma dimensão de  $\mathbf{Y}$ .<sup>40</sup> Usando como

exemplo o presente trabalho, a matriz  $\mathbf{X}$  seria formada pelos valores de absorvância em diversos comprimentos de onda na região do infravermelho e a matriz  $\mathbf{Y}$  seria formada por valores de teor de cocaína, de adulterantes e grau de oxidação em amostras apreendidas.

A construção desse modelo inicia-se pela decomposição das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  em suas matrizes de escores (do inglês, *scores*) e pesos (do inglês, *loadings*)<sup>41</sup>, Equações 1 e 2:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E}_X \quad (1)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{E}_Y \quad (2)$$

onde  $\mathbf{T}$  e  $\mathbf{U}$  são os *scores* e os elementos de  $\mathbf{P}$  e  $\mathbf{Q}$  são os *loadings* de  $\mathbf{X}$  e  $\mathbf{Y}$ , respectivamente, para cada número de variáveis latentes (VLs) “ $h$ ”. As matrizes  $\mathbf{E}_X$  e  $\mathbf{E}_Y$  correspondem aos resíduos das VLs que não foram utilizadas no modelo, ou seja, as matrizes que contem a parte não modelada.

A Figura 5 ilustra a decomposição das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  no produto das matrizes de *scores* e *loadings*.

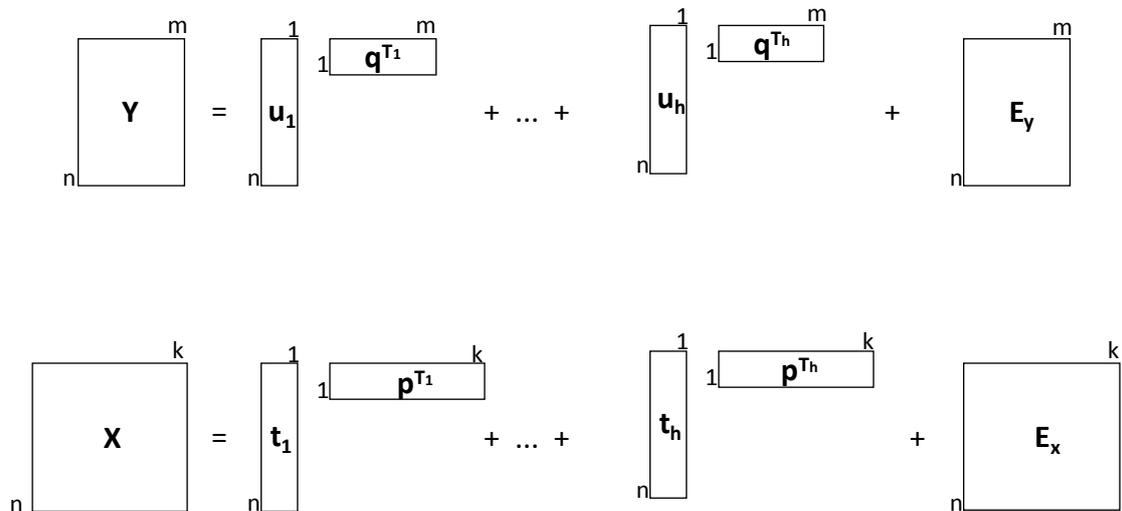


Figura 5: Decomposição em variáveis latentes das matrizes  $\mathbf{X}$  e  $\mathbf{Y}$  para modelos PLSR.

Uma relação linear entre os *scores* de  $\mathbf{X}$  e  $\mathbf{Y}$  é obtida através do vetor coeficiente de regressão linear para cada VL ( $\mathbf{b}_h$ )<sup>42</sup>, como descrito na Equação 3:

$$\mathbf{b}_h = \frac{\mathbf{u}_h}{\mathbf{t}_h} \quad (3)$$

sendo  $\mathbf{u}$  e  $\mathbf{t}$  os elementos das matrizes  $\mathbf{U}$  e  $\mathbf{T}$  das matrizes  $\mathbf{Y}$  e  $\mathbf{X}$ , respectivamente. Os valores de  $\mathbf{b}_h$  são agrupados na matriz diagonal  $\mathbf{B}$ .

O modelo final consiste das matrizes dos *scores* **T** e **U** que são linearmente relacionadas pelo coeficiente **B**. No final do processo, a variância explicada pela primeira VL será maior que a variância explicada pela segunda e a terceira explicará uma variância ainda menor que a explicada pela segunda VL, e assim sucessivamente até o número de VLs escolhido.<sup>43</sup>

O número de VLs utilizadas em um modelo é de fundamental importância nos resultados a serem obtidos. Um modelo contendo um número de VLs inferior ao ideal resultará em subajuste, pois não irá considerar toda a informação relevante no conjunto de dados. Por outro lado, quanto maior o número de VLs, também aumentará o ruído e os erros de modelagem, resultando em um sobreajuste. Deste modo, o número ótimo de VLs corresponde ao ponto no qual a diminuição do erro (produzido pelo aumento da complexidade do modelo) é compensado pelo aumento de erro de superavaliação.<sup>44</sup> O método mais simples de se determinar o número de VLs é o que utiliza o método de Validação Cruzada (CV, do inglês *Cross Validation*).<sup>45</sup>

A validação cruzada pode ser realizada em blocos, ou seja, um número determinado de amostras é deixado de fora no processo de construção do modelo e a seguir essas amostras são previstas pelo modelo construído. O processo é repetido até que todas as amostras tenham sido previstas e a raiz do erro quadrático médio de validação cruzada (RMSECV, do inglês *Root Mean Square Error of Cross Validation*)<sup>46</sup> é calculada através da Equação 4:

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{I_C} (\hat{y}_i - y_i)^2}{I_C}} \quad (4)$$

onde  $\hat{y}_i$  e  $y_i$  são, respectivamente, os valores estimados e de referência para o teor de cocaína e de seus adulterantes e  $I_C$  é o número de amostras do conjunto de calibração.

Os resultados de RMSECV são apresentados em um gráfico em função do número de VLs. O comportamento típico para esses gráficos é a observação de um mínimo ou um patamar, que indica a melhor dimensionalidade do modelo de regressão, ou seja, o melhor número de VLs que produziu o menor RMSECV sem perda significativa da variância dos dados<sup>45</sup>.

Outro critério para avaliar um modelo de regressão, tão importante quanto a determinação de VLs, é a detecção de amostras anômalas (do inglês, *outliers*). Ao verificar a qualidade de um conjunto é necessário identificar esses elementos, assegurando que amostras com erros grosseiros quando comparadas à maioria dos dados

sejam identificadas. Uma amostra anômala é proveniente de dados que diferem de alguma forma do conjunto total de calibração. Seja por causa de ruídos, erros na medição e/ou métodos inadequados para análise.

No entanto, a eliminação de *outliers* deve ser feita com cautela, pois algumas vezes a presença de uma amostra diferente das demais pode conter, ao invés de erros, informações importantes que não são encontradas nos outros dados. Assim sendo, sua presença poderá contribuir para a representatividade do conjunto de dados utilizado no desenvolvimento do modelo de regressão.

Nesta tese, os *outliers* foram identificados segundo a norma ASTM E1655-05<sup>46</sup> e o trabalho descrito por Valderrama *et al.*<sup>47</sup>. Esses critérios baseiam-se na identificação de amostras com alto valor de *leverage* ( $L$ ) – resíduos não modelados em dados espectrais (bloco  $\mathbf{X}$ ) e resíduos não modeladas na variável dependente (bloco  $\mathbf{Y}$ ), tendo em conta os intervalos de confiança de 99%. Para a avaliação dos resíduos no bloco  $\mathbf{X}$ , de acordo com ASTM E1655-00<sup>46</sup>, amostras com  $L_i$  maior que o valor limite ( $L_{limite}$ ), dados pelas equações 5 e 6, devem ser removidas do conjunto de calibração e o modelo deve ser reconstruído. Não é raro, eliminar amostras com alto  $L_i$  em um primeiro modelo e o modelo reconstruído encontrar novos espectros com  $h_i > h_{limite}$ . Quando a aplicação repetitiva dessa regra continua a identificar *outliers*, diz-se que ocorreu o “efeito bola de neve”, o qual pode indicar algum problema com a estrutura do conjunto de dados espectrais. Nestas situações, o teste para verificação de *outliers* pode ser relaxado desde que nenhuma amostra de calibração tenha  $h_i$  superior a 0,5.

$$L_i = \mathbf{t}_{h,i}^T (\mathbf{T}_h^T \mathbf{T})^{-1} \mathbf{t}_{h,i} \quad (5)$$

$$L_{limite} = 3 \frac{h+1}{I_C} \quad (6)$$

sendo  $\mathbf{t}$  elementos da matriz  $\mathbf{T}$  (*scores* de  $\mathbf{X}$ ) e  $h$  o número de VL.

Amostras anômalas no bloco  $\mathbf{Y}$  foram detectadas com base na comparação dos resíduos totais do desvio padrão ( $s(e)$ ) com o desvio padrão de uma amostra específica ( $s(e_i)$ ), definida como<sup>47</sup>:

$$s(e)^2 = \frac{1}{I_C J - J - h_{\max}(J_C, J)} \sum_{i=1}^{I_C} \left( \sum_{j=1}^J (x_{ij} - \hat{x}_{ij})^2 \right) \quad (7)$$

$$s(e_i)^2 = \frac{1}{I_C J - J - h_{\max}(J_C, J)} \sum_{j=1}^J (x_{ij} - \hat{x}_{ij})^2 \quad (8)$$

onde  $J$  é o número de variáveis espectrais,  $x_{i,j}$  é o valor de absorvância da amostra  $i$  no comprimento de onda  $j$  e  $\hat{x}_{i,j}$  é o seu valor estimado com  $h$  variáveis latentes. Se uma amostra apresenta  $s(e_i) > n s(e)$ , onde  $n$  é uma constante que pode variar de 2 a 3, a amostra deve ser removida do conjunto de calibração<sup>48</sup>. Neste trabalho, a constante foi otimizada como 2, por fornecer um bom limite que poderia identificar as amostras que apresentam resíduos significativamente maiores do que os observados a partir das outras amostras.

Sabendo disso, um primeiro modelo de calibração foi construído e os *outliers* foram removidos a partir das amostras de calibração; em seguida, o modelo foi recalculado e tanto a identificação de *outliers* quanto o processo de exclusão foram repetidos. Após duas exclusões, o terceiro modelo foi considerado otimizado. O conjunto de validação foi avaliado com o modelo de calibração otimizado e os *outliers* foram excluídos aplicando os mesmos critérios utilizados para o modelo de calibração.<sup>65</sup>

Por fim, o método PLSR é considerado robusto quando seus parâmetros praticamente não se alteram com a inclusão de novas amostras no conjunto de calibração. Porém, antes da aplicação de um modelo construído com PLSR, o mesmo deve ser validado com o objetivo de se testar a sua capacidade de previsão. A validação consiste em testar o modelo prevendo a variável resposta ( $Y$ ) e determinar suas figuras de mérito, para estabelecer se ele de fato irá refletir o comportamento do analito de interesse.<sup>49</sup>

Uma variante de regressão PLSR é o PLS-DA utilizado para classificação supervisionada, pois é desenvolvido mediante o conhecimento prévio das classes do conjunto de calibração (método supervisionado). A construção de um método por PLS-DA segue o desenvolvimento de um modelo PLSR, no entanto difere na matriz  $Y$ , que ao invés de informações da propriedade de interesse (teor de cocaína, adulterantes e grau de oxidação), são inseridos os números das classes das amostras<sup>50</sup>. Dessa forma, para cada classe a ser prevista, um modelo é construído correlacionando a matriz espectral com um vetor de classe, cujos elementos têm valores de 1 se uma amostra pertence à classe discriminada ou 0 se não pertencer a essa classe. O valor que melhor separa as duas classes é o limite de discriminação, o qual é determinado com base na dispersão das amostras de calibração, de forma a minimizar a ocorrência de erros de classificação<sup>51</sup>. Caso as dispersões das duas classes sejam muito semelhantes, o limite de discriminação fica próximo a 0,5, valor que corresponde ao centro da escala. Caso contrário, há um deslocamento do limite de discriminação, visando minimizar possíveis erros de classificação<sup>50</sup>. Assim, uma amostra que apresente valor de classe superior ao limite de

discriminação é identificada como cocaína base livre (uma das formas de cocaína), enquanto que, se o valor de classe for inferior ao limite de discriminação, esta amostra seria classificada como cocaína sal (outra forma de apresentação de cocaína).

## 2.2. Figuras de Mérito

As figuras de mérito são indicadores quantitativos do desempenho da técnica utilizada na análise de dados. Neste trabalho foram calculados a exatidão, *viés*, precisão, sensibilidade, concentração mínima detectável e intervalo de confiança.

**Exatidão:** informa o grau de concordância entre os valores de referência e aqueles estimados pelo método proposto. Em termos médios, pode ser expressa pela raiz do erro quadrático médio de previsão (RMSEP, do inglês *Root Mean Square Error of Prediction*), Equação 9.<sup>52,53</sup>

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{I_v} (y_i - \hat{y}_i)^2}{I_v}} \quad (9)$$

onde  $I_v$  é o número de amostras do conjunto de validação.

**Viés:** avalia a presença de erros sistemáticos e é definido como a diferença entre a média ( $\bar{y}_v$ ) e o valor verdadeiro, sendo este um parâmetro que deve ser avaliado junto com o RMSEP para atestar a exatidão. É calculado apenas a partir do conjunto de validação, Equação 10.<sup>47</sup>

$$viés = \sum_{i=1}^{I_v} \frac{(\bar{y}_v - y_i)}{I_v} \quad (10)$$

O desvio padrão de erros de validação (DPV) foi calculado pela Equação 11, é utilizado para determinar se as estimativas de validação mostram uma tendência estatisticamente significativa.

$$DPV = \sqrt{\frac{\sum [(y_i - \hat{y}_i) - viés]^2}{I_v - 1}} \quad (11)$$

O valor de t calculado ( $t = |viés| \sqrt{I_v} / DPV$ ) é então comparado com o valor crítico de t com número de  $I_v - 1$  graus de liberdade.

**Precisão:** determinada de acordo com a Equação 12, indica o quanto as medidas repetidas estão próximas umas das outras. Para esse propósito, nessa tese foram feitas, no mesmo dia, 10 replicatas independentes de quatro amostras diferentes, com concentrações regularmente distribuídas ao longo da faixa linear do método.<sup>54,55</sup>

$$Precisão = \sqrt{\frac{\sum_{i=1}^I \sum_{j=1}^m (\hat{y}_{i,j} - \hat{\bar{y}}_i)^2}{I(m-1)}} \quad (12)$$

onde  $I$  é o número de amostras,  $m$  é o número de replicatas,  $\hat{y}_{i,j}$  é a concentração estimada para a amostra  $i$  e réplica  $j$ , e  $\hat{\bar{y}}_i$  é a concentração média de réplicas para amostra  $i$ .

**Sensibilidade (SEN):** determina a fração do sinal analítico devido ao aumento da concentração de um analito particular na unidade de concentração. A SEN foi determinada com base nos coeficientes de regressão do modelo PLSR, de acordo com a Equação 13.<sup>56,57</sup>

$$SEN = \frac{1}{\|\mathbf{b}\|} \quad (13)$$

onde  $\mathbf{b}$  é o vetor do coeficiente de regressão com  $h$  variáveis latentes.

**Concentração mínima detectável (MDC, do inglês *minimum detectable concentration*):** é definida como sendo a menor concentração que pode ser mensurada de forma confiável. Os valores da MDC foram calculados aplicando a norma ISO 11843-2<sup>58</sup>, que segundo Ortiz *et al.*<sup>59</sup>, podem ser obtidos utilizando a Equação 14.

$$MDC = \delta_{\alpha,\beta,v} \frac{s}{b} \sqrt{\frac{1}{m} + \frac{1}{I_C} + \frac{\bar{y}_C^2}{\sum_{i=1}^{I_C} (y_i - \bar{y})^2}} \quad (14)$$

onde  $s$  é o desvio padrão dos resíduos para a regressão linear entre os valores de referência e os valores estimados pelo método proposto,  $\bar{y}_C$  é a concentração média das amostras de calibração,  $\delta_{\alpha,\beta,v}$  é o parâmetro de não centralidade da distribuição  $t$ ,  $\alpha$  e  $\beta$  são as probabilidades de ocorrência de erros falsos negativos e falsos positivos, respectivamente, e  $v = I_C - 2$  graus de liberdade. Neste trabalho, as duas probabilidades  $\alpha$  e  $\beta$  foram consideradas como sendo igual a 0,05 (nível de confiança de 95%).

**Intervalo de confiança (IC):** é definido como um intervalo, com um determinado grau de confiança (isto é, uma certa probabilidade), na qual o valor real da concentração do analito está incluído. Esse intervalo pode ser determinado através da aplicação de um modelo de distribuição residual que segue a distribuição normal e do erro padrão estimado de predição ( $s(\hat{y} - y)$ ), que é determinado pelas Equações 15 e 16, respectivamente.<sup>48,56</sup>

$$IC(y_i) = \hat{y}_i \pm t_{v,1-\alpha/2} s(\hat{y} - y_i) \quad (15)$$

$$IC(y_i) = \hat{y}_i \pm t_{v,1-\alpha/2} \sqrt{s^2 (1 + L_i + 1/I_C)} \quad (16)$$

onde  $\alpha$  é o nível de significância igual a 0,05 (95% de nível de confiança);  $t_{v,1-\alpha/2}$  é o nível crítico correspondente para a distribuição t de Student com  $\nu$  pseudo graus de liberdade, determinado como proposto por Van der Voet<sup>60</sup>;  $s^2$  é uma estimativa do desvio padrão do erro de ajuste para conjunto de calibração, determinado pela equação 17 e  $L_i$  é o *leverage* da amostra.

$$s^2 = MSEC = \frac{\sum_{i=1}^{I_C} (y_i - \hat{y}_i)^2}{\nu} \quad (17)$$

### 2.3. Transferência de Calibração

Como dito na seção anterior, modelos de calibração multivariada geralmente são construídos com o objetivo de quantificar compostos de interesse em um determinado tipo de amostras a partir de dados instrumentais. Porém, quando, por exemplo, o modelo de calibração é desenvolvido em um equipamento, mudanças no sinal instrumental podem ocorrer e com isso tornar os resultados obtidos pelo modelo não mais aceitáveis. Um exemplo desse tipo de mudança é a ocorrência de manutenção ou troca de peças do equipamento.<sup>61</sup>

Uma solução para esse problema seria uma recalibração completa do modelo. No entanto, a calibração multivariada geralmente requer a utilização de grande número de amostras e, dessa forma, torna-se um processo trabalhoso, que demanda tempo e dinheiro.<sup>62</sup> Uma alternativa à recalibração é a transferência de calibração, que pode ser entendida como sendo uma adaptação de um modelo multivariado para atender às novas condições nas quais o equipamento se encontra.

Neste trabalho, a transferência de calibração foi utilizada para testar o funcionamento do método desenvolvido em duas novas situações: após manutenção do equipamento de FTIR-ATR, no qual as amostras foram analisadas, e em outro equipamento de FTIR-ATR de marca e modelo diferente. Isso foi feito para mostrar a capacidade de implementação do método em outras localidades mesmo que não fosse utilizado o mesmo equipamento de FTIR-ATR.

Existem diversos métodos para se fazer a transferência de calibração, sendo que um deles é o chamado Padronização Direta por Partes (PDS, do inglês *Piecewise Direct Standardization*). Nesse caso, busca-se uma modificação do sinal instrumental de uma situação secundária (também chamada de filha ou escrava) para se tornar parecida com a função de resposta da situação primária (pai ou mestre), onde o modelo de calibração foi

desenvolvido. Isto pode ser feito através de correções matemáticas nas respostas espectrais ou nos valores preditos. Este método utiliza medidas espectrais de várias amostras (chamadas de amostras de transferência) registradas nos equipamentos envolvidos na transferência (ou medidas nas duas situações primária e secundária). Se  $\mathbf{X}_p$  ( $N \times P$ ) e  $\mathbf{X}_s$  ( $N \times P$ ) são as matrizes contendo os espectros de  $N$  amostras de transferência registrados nas condições primária e secundária, respectivamente, a padronização das respostas espectrais procura relacioná-las por meio de uma transformação linear expressa pela Equação 18,<sup>62</sup>

$$\mathbf{X}_p = \mathbf{X}_s \mathbf{F} \quad (18)$$

onde  $\mathbf{F}$  é uma matriz de dimensão ( $P \times P$ ), sendo  $P$  o número de comprimentos de onda para os espectros registrados nos equipamentos primário e secundário, respectivamente.

A vantagem da PDS é a minimização do risco de sobreajuste, visto que o número de termos calculados para estimar  $\mathbf{F}$  é reduzido. Em geral, esse método é empregado como referência para avaliação de novas propostas de transferência de calibração, por permitir correção simultânea de diferenças de intensidade, deslocamentos de comprimentos de onda e alteração na largura das bandas.<sup>62</sup>

A avaliação do funcionamento da transferência de calibração é feita através dos valores do RMSEP.<sup>52,53</sup>

O fluxograma da Figura 6 exemplifica uma transferência de calibração mostrando o conjunto de amostras do modelo mestre e escravo com seus respectivo RMSEP.

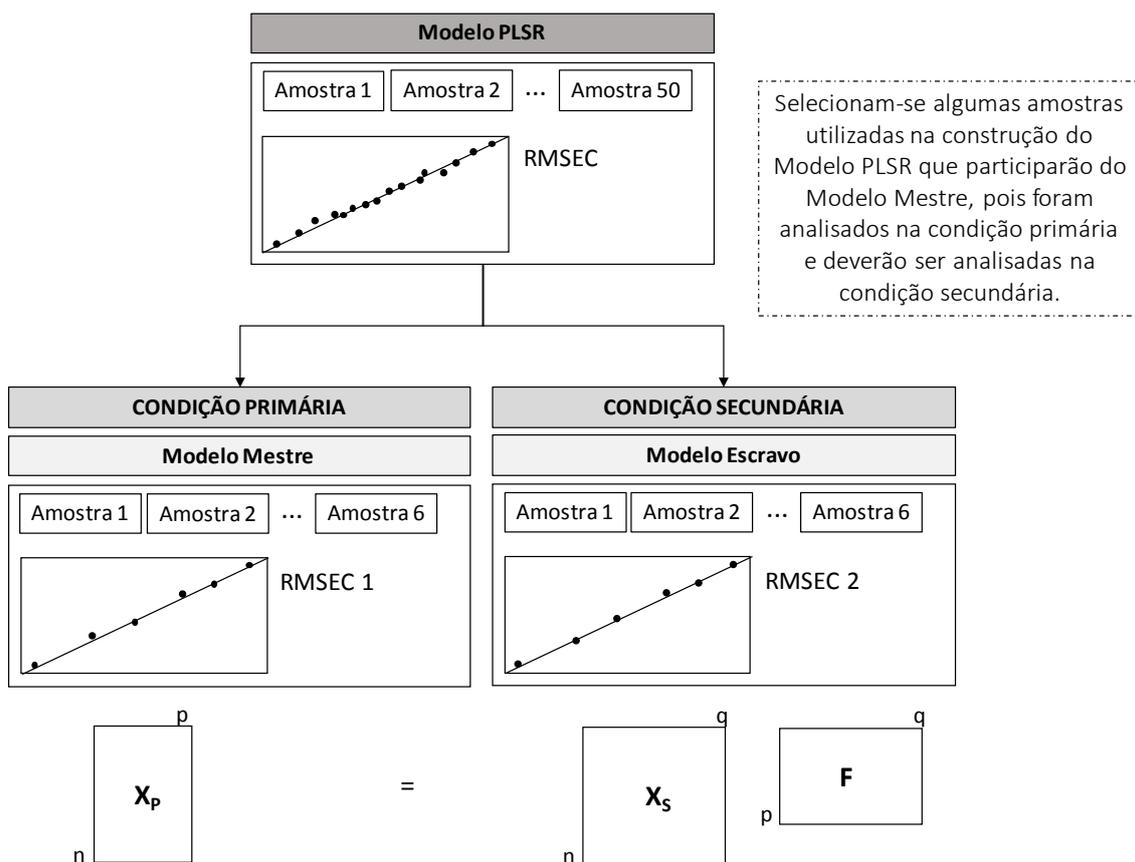


Figura 6: Fluxograma simplificado com as etapas envolvidas na transferência de calibração.

### 3. Materiais e Métodos

#### 3.1. Amostras e Preparação de Amostras

O conjunto de dados para a construção dos modelos de discriminação e determinação do teor de cocaína, adulterantes e grau de oxidação consistiu de 1085 amostras provenientes de cerca de 380 apreensões feitas pela Polícia Federal do Brasil entre 2009 e 2013 nos Estados do Acre, Amazonas, Distrito Federal, Paraná, Mato Grosso, Mato Grosso do Sul e São Paulo. Essas amostras fazem parte do Projeto PEQUI (Perfil Químico de Drogas), em que alíquotas de apreensões com mais de 5 Kg são enviadas ao Laboratório de Química Forense do Instituto Nacional de Criminalística (INC), em Brasília.

Antes da análise instrumental, todas as amostras de cocaína base livre foram cuidadosamente homogeneizadas por maceração manual na presença de nitrogênio líquido. Cerca de 1 g de amostra de cocaína base livre foi colocada num almofariz de porcelana e nitrogênio líquido foi vertido até cobrir a amostra. Com a ajuda de um pistilo, a amostra foi cuidadosamente macerada. Este procedimento criogênico é especialmente

adequado para o tratamento de amostras de cocaína base livre por serem pegajosas quando maceradas sem o nitrogênio.

### **3.1.1. Amostras Utilizadas na Transferência do Modelo de Calibração**

Para avaliar o funcionamento do modelo de calibração em análises após a manutenção do equipamento de FTIR-ATR e em outro equipamento de marca e modelo diferente, foram selecionadas 31 amostras aleatórias, mas representativas que fizeram parte do modelo de calibração PLSR para transferência do método. Dessas amostras, 16 eram de cocaína base livre e 15 de cocaína sal com o teor de cocaína variando de 33,7 a 86,8% (m/m) determinado através do método cromatográfico de referência descrito na seção 3.4.

Seguindo o método de Kennard-Stone<sup>66</sup> 10 dessas amostras foram escolhidas para realizar a transferência do modelo de calibração do equipamento primário (mestre) para o equipamento secundário (escravo) e 21 para a validação desse modelo.

### **3.2. Ensaios de Solubilidade**

Ensaios de solubilidade foram realizados a fim de identificar a forma química da cocaína (sal ou base livre) de acordo com as metodologias recomendadas pela Polícia Federal Brasileira (PF)<sup>63</sup>, Grupo de Trabalho Científico de Análise de Drogas Apreendidas (SWGDRUG, do inglês: *Scientific Working Group for the Analysis of Seized Drugs*)<sup>64</sup> e UNODC<sup>65</sup>.

### **3.3. Análise de Cromatografia Gasosa com Detector de Ionização por Chama**

A quantificação de cocaína, bem como dez dos principais adulterantes (benzocaína, lidocaína, cafeína, procaína, paracetamol, fenacetina, hidroxizina, aminopirina, diltiazem e levamisol), presentes nas amostras, foi determinada por um método de referência<sup>67</sup>, que utiliza CG-DIC, realizado rotineiramente pelo Laboratório do Instituto de Criminalística da Polícia Federal. Nesse método, uma quantidade de 12,00 a 12,50 mg de cada amostra homogeneizada foi misturada com 10,0 mL de uma solução de padrão interno (2,0  $\mu\text{L}\cdot\text{L}^{-1}$  de dietilamina e 512  $\text{mg}\cdot\text{L}^{-1}$  de dipentiltalato preparado em clorofórmio). Um volume de 1 mL desta solução foi então transferido para *vials* de vidro, selados, e submetido a análise cromatográfica.

A análise por CG-DIC foi realizada num cromatógrafo gasoso modelo 6890N (*Agilent Technologies*), equipado com um detector de ionização de chama e um amostrador automático 7683B Series (*Agilent Technologies*). As condições cromatográficas foram as seguintes: volume de injeção de 200  $\mu\text{L}$ ; razão de *split* de 50:1; coluna cromatográfica DB1-MS de metilsiloxano (25 m  $\times$  200  $\mu\text{m}$  [i.d.]  $\times$  0,33  $\mu\text{m}$  de espessura de filme); temperatura do injetor de 280°C e temperatura do detector de 320°C. Utilizou-se hélio como gás transportador com um fluxo de 1,0 mL.min<sup>-1</sup>. A programação de temperatura do forno foi: 150°C durante 2 minutos, rampa de aquecimento de 40°C.min<sup>-1</sup> até 350°C, mantendo-se a 350°C durante 4,5 minutos, resultando em uma corrida cromatográfica de 11,5 minutos.

### **3.4. Medidas de Espectroscopia no Infravermelho**

Os espectros de infravermelho para a construção dos modelos de discriminação e regressão foram obtidos num espectrômetro *Nicolet FTIR iS10* equipado com um acessório de refletância total atenuada SMART ATR com tripla reflexão usando um cristal de diamante pertencente ao Instituto Nacional de Criminalística, INC. As medições foram obtidas no modo de refletância com o acessório preenchido com uma pequena quantidade de cocaína. Os espectros foram coletados entre 4000 e 400 cm<sup>-1</sup> e 16 *scans* foram adquiridos com uma resolução de 4,0 cm<sup>-1</sup>.

#### **3.4.1. Medidas de Espectroscopia no Infravermelho para a Transferência de Calibração**

Como dito anteriormente, a fim de verificar a robustez do modelo, duas novas situações foram testadas. A primeira verificou o funcionamento do modelo após manutenções ocorridas num intervalo de 30 meses espectrômetro *Nicolet FTIR iS10* pertencente ao INC. A segunda situação avaliou o modelo em espectrômetro de infravermelho com transformada de Fourier da marca Jasco modelo 4100 com acessório para medidas por Refletância Total Atenuada (ATR) pertencente ao Instituto de Química da Universidade de Brasília, com espectros coletados entre 4000 e 650 cm<sup>-1</sup>.

### **3.5. Análise Multivariada**

O método de calibração multivariada foi desenvolvido com base em PLSR, para quantificar cocaína e os dez principais adulterantes citados anteriormente, e PLS-DA, que discriminou amostras de cocaína de acordo com sua forma química (base livre ou sal).

Os modelos foram construídos utilizando todo o espectro de infravermelho com dados em escala de absorvância obtida através da relação  $Abs = \log_{10}(1/R)$ , onde R é a reflectância.

As técnicas de pré-processamento testadas foram: Variação Padrão Normal (SNV, do inglês: *Standard Normal Variate*), Correção do Sinal Ortogonal (OSC, do inglês: *Orthogonal Signal Correction*), primeira derivada utilizando o algoritmo de Savitz-Golay com suavização por um polinômio de segundo grau e janela de 15 pontos, centralização dos dados na média (do inglês: *Mean Center*) e combinações binárias desses métodos.<sup>68,69</sup>

Os dados espectrais das amostras foram importados para o *software* MATLAB® (versão 7.12, R2011a) e os modelos de pré-processamento e PLSR foram desenvolvidos utilizando a plataforma PLS Toolbox® (versão 6.5) da Eigenvector Technologies. Para a construção dos modelos, o conjunto de dados foi dividido pelo algoritmo Kennard-Stone<sup>66</sup> em 500 amostras para a calibração e 585 amostras para a validação.

Depois de escolher o método de pré-processamento, os modelos foram otimizados pela eliminação de *outliers* (Figura 6), os quais foram identificados segundo a norma ASTM E1655-05<sup>46</sup> e Valderrama *et al.*<sup>47</sup>.

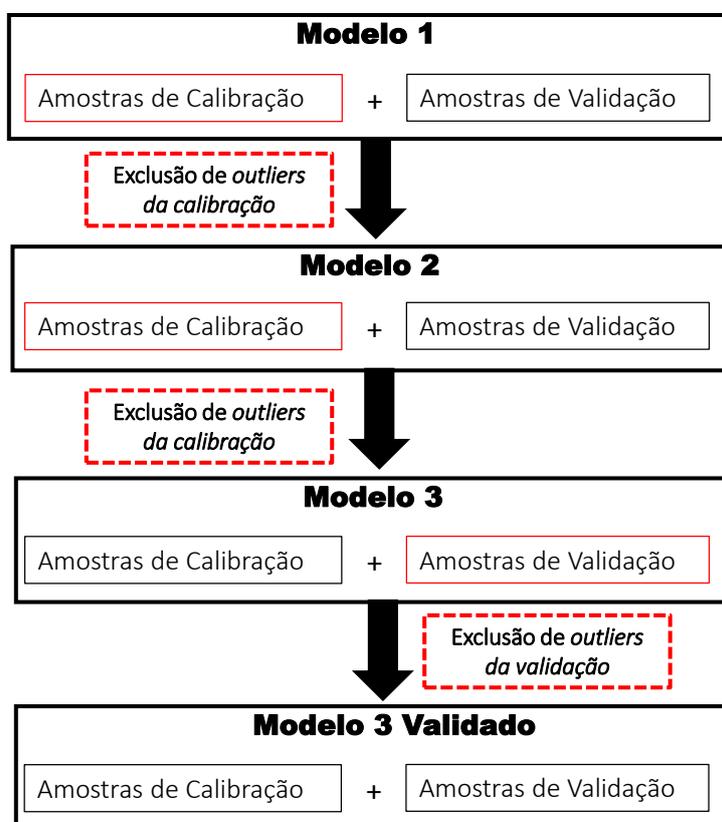


Figura 7: Passos para a construção de modelos de calibração multivariada.

## 4. Resultados e Discussão

### 4.1. Teste de Solubilidade

Com base em testes de solubilidade, 805 amostras (74,2%) foram classificadas como cocaína na forma de base livre e 280 amostras (25,8%) como cocaína sal. Este conjunto de dados representa o cenário das apreensões de drogas ilícitas no Brasil, onde a maioria das apreensões de cocaína é de amostras em forma de base livre.

### 4.2. Análise de Cromatografia Gasosa com Detector de Ionização por Chama

A análise de referência (CG-DIC) determinou o teor de cocaína nas amostras que variou entre 24,2 e 99,9% (m/m), conforme histograma mostrado na Figura 8.

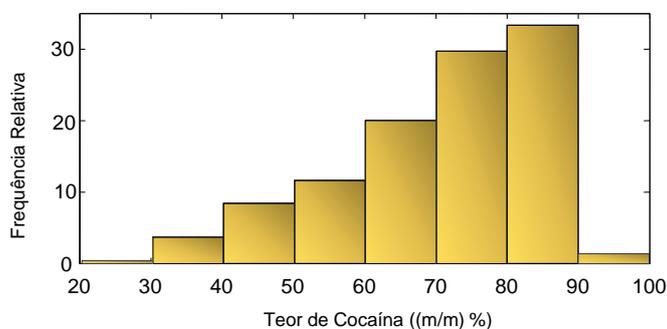


Figura 8: Histograma de concentrações de cocaína nas amostras apreendidas.

Essa análise também revelou que cerca de 73% das amostras de cocaína de base livre apresentaram baixos graus de oxidação, enquanto que cerca de 62% das amostras de cloridrato de cocaína tinham graus médios de oxidação, tal como mostrado na Figura 9. Isso significa que a maioria das amostras de cocaína base livre apreendidas pela Polícia Federal Brasileira não tinham passado por nenhum processo de purificação, diferentemente da cocaína sal que possivelmente foi reprocessada ou produzida a partir da cocaína base livre.

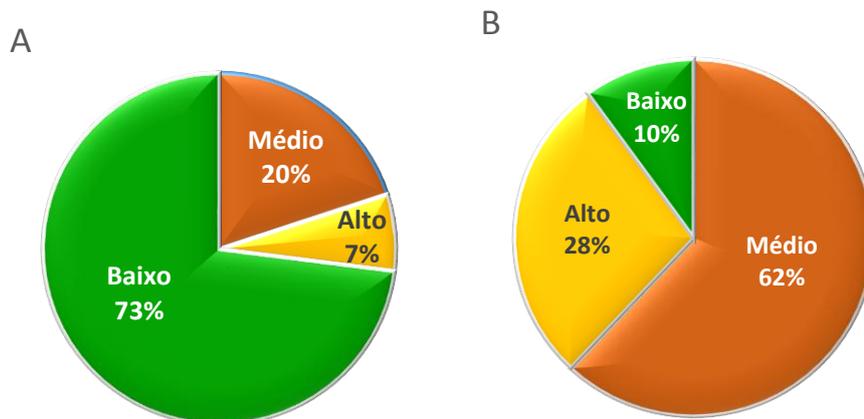


Figura 9: Grau de oxidação de amostras apreendidas: (A) de base livre de cocaína, (B) de cloridrato de cocaína.

Em 63,3% de amostras analisadas por CG-DIC foi detectada a presença de pelo menos um dos dez adulterantes mais comuns citados no UNODC (benzocaína, paracetamol, fenacetina, cafeína, lidocaína, aminopirina, levamisol, procaína, hidroxizina e diltiazem).<sup>9</sup> A Tabela 3 mostra o número de amostras em que esses adulterantes foram encontrados, bem como a sua faixa de concentração.

Observou-se ainda, que o maior número de adulterantes detectados numa amostra foram quatro, mas apenas 0,74% das amostras apresentam esta situação (Figura 10). Para as 1085 amostras estudadas, 36,7% não apresentaram adulterantes, 41,7% apresentaram apenas um adulterante e 21,6% apresentaram mais de 1 adulterante.

Métodos quantitativos de análise multivariada foram avaliados para todos os adulterantes analisados, no entanto, observou-se que apenas para a fenacetina, benzocaína, lidocaína, cafeína, aminopirina e levamisol os modelos puderam ser desenvolvidos e validados com sucesso. Esses adulterantes estavam presentes em pelo menos 45 amostras. Os demais adulterantes estavam presentes em poucas amostras, logo a construção desses modelos foi prejudicada. Diluentes não foram avaliados pelo método de referência, nem pelo método proposto.

Tabela 3. Adulterantes encontrados em amostras de cocaína apreendidas pela Polícia Federal Brasileira.

Forma	Adulterantes	Nº de Amostras	Faixa de Teor (% m/m)
Cocaína Base Livre	Benzocaína	68	0,5 - 14,2
	Paracetamol	13	0,1 - 3,0
	Fenacetina	446	0,5 - 56,8
	Cafeína	18	0,5 - 14,5
	Lidocaína	37	0,4 - 17,2
	Aminopirina	148	0,1 - 41,2
	Levamisol	11	2,7 - 42,5
	Procaína	0	0,0 - 0,0
	Hidroxizina	35	1,3 - 2,7
	Diltiazem	3	0,7 - 1,0
Cocaína Sal	Benzocaína	0	0,0 - 0,0
	Paracetamol	0	0,0 - 0,0
	Fenacetina	26	0,5 - 13,2
	Cafeína	27	0,6 - 25,4
	Lidocaína	36	0,5 - 29,2
	Aminopirina	19	0,1 - 4,1
	Levamisol	86	2,5 - 24,4
	Procaína	6	3,0 - 7,7
	Hidroxizina	0	0,0 - 0,0
	Diltiazem	8	0,2 - 0,7

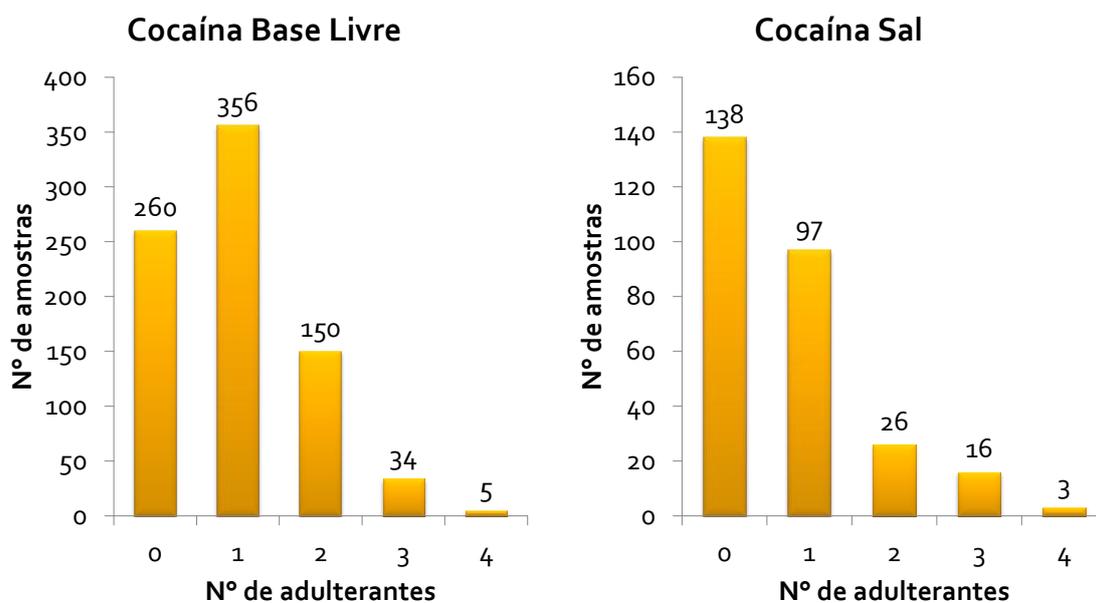


Figura 10: Número de adulterantes encontrados em amostras de cocaína apreendidas pela Polícia Federal Brasileira.

### 4.3. Dados de Infravermelho

Os espectros de infravermelho (Figura 11) e a análise cromatográfica confirmaram a presença de cocaína nas amostras. Alguns dos espectros de infravermelho apresentam diferenças identificáveis facilmente por uma simples análise visual, como mostrou a Figura 12(A). Isso porque no espectro de cocaína sal aparece a banda de absorção localizada cerca de  $2500\text{ cm}^{-1}$  característica da ligação N-H. Já no espectro de cocaína base livre observa-se a presença de bandas de intensidade média entre  $3000$  e  $2800\text{ cm}^{-1}$ . Porém, há também algumas amostras nas quais essa diferença não foi tão significativa (Figura 12(B)). Fenômeno que ocorre com certa frequência devido ao processo de adulteração e diluição da amostra, o que dificulta a identificação da forma da cocaína. Por isso, a importância do uso de análises quimiométricas para a discriminação de amostras de acordo com sua forma de apresentação.

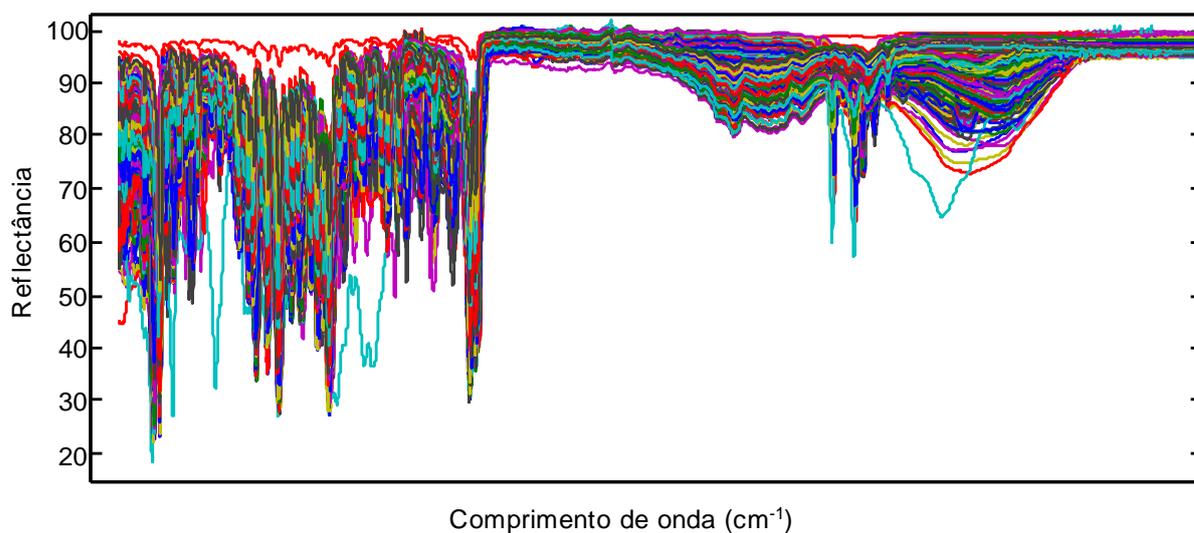


Figura 11: Espectros ATR-FTIR de todas as amostras de cocaína analisadas.

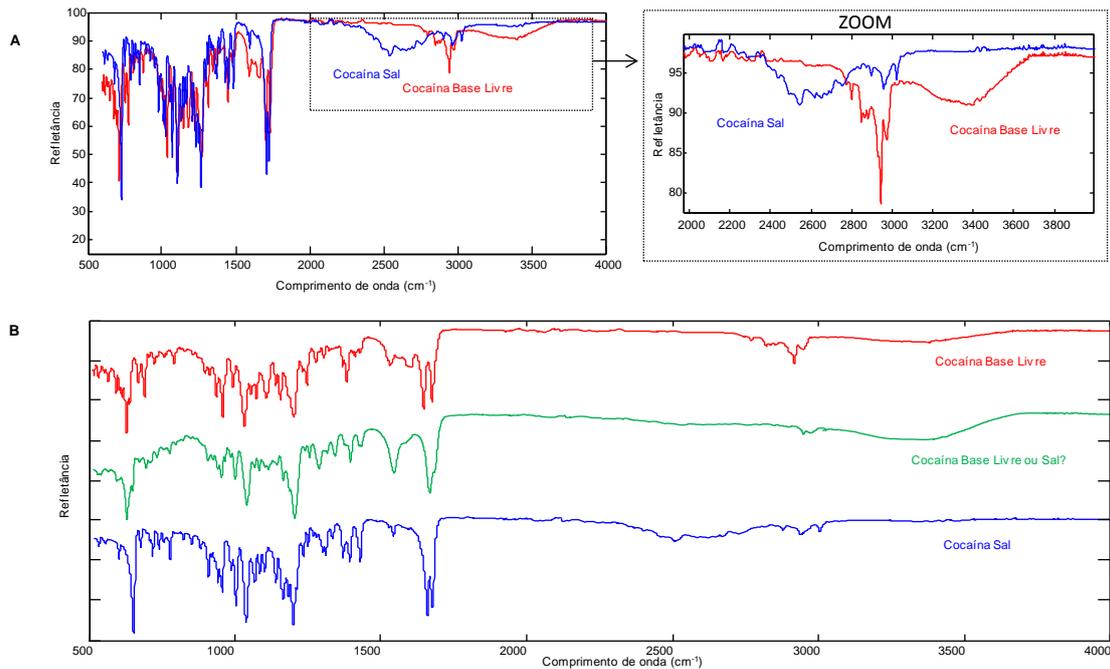


Figura 12: Principais diferenças entre os espectros ATR-FTIR de cocaína de base livre e de cocaína sal. Diferenças (A) perceptíveis e (B) não perceptíveis visualmente.

#### 4.4. Discriminação de amostras de acordo com a forma de apresentação da cocaína

Para o desenvolvimento dos modelos descritos nesse trabalho, os espectros de FTIR foram utilizados em sua totalidade. Dessa forma, o modelo considerou todas as contribuições na discriminação ou quantificação de amostras, levando a resultados mais robustos.

O método de pré-processamento escolhido para a análise de dados foi a primeira derivada com uma janela de 15 pontos, SNC e *mean center*. Essa escolha foi feita com base no RMSECV, obtido por blocos contínuos e 100 *splits*.

Utilizando PLS-DA, construiu-se um modelo de discriminação de amostras de acordo com a forma de apresentação da cocaína que teve um número de *outliers* relativamente pequeno, juntamente com baixos valores de raiz quadrada do erro médio de previsão (RMSEP, do inglês *Root Mean Square Error of Prediction*) que variaram de 0,219 (para o modelo não otimizado) a 0,003 (para o modelo otimizado) (ver Tabela 4).

Tabela 4. Efeito da eliminação de *outliers* em cada etapa de otimização do modelo de discriminação.

<b>Modelos</b>	<b>1º</b>	<b>2º</b>	<b>3º</b>	<b>Final</b>
Nº de amostras de calibração	500	462	427	423
Nº de amostras de validação	585	585	585	559
RMSEC	0,191	0,096	0,038	0,005
RMSEP	0,219	0,219	0,219	0,003

Observando o processo de exclusão de *outliers* apresentado na Seção 1.5, das 585 amostras utilizadas na validação apenas 26 (ou seja, 4,4%) foram excluídas como *outliers*. Notou-se ainda que, se essas mesmas 26 amostras tivessem sido consideradas no modelo, apenas 24 delas teriam sido classificadas erroneamente, o que corresponde a um erro de discriminação de 4,1%. Após eliminação de *outliers*, 100% das amostras de validação foram corretamente identificadas como cocaína base livre ou cocaína sal. Alguma similaridade foi buscada entre as amostras excluídas, tais como ano e local de apreensão, teor de adulterantes, forma de apresentação da cocaína, porém nenhuma semelhança foi encontrada.

Pelo gráfico de dispersão obtido para o modelo final de discriminação de amostras (Figura 13), observou-se que as amostras ficaram bem ajustadas em 0 e 1. Dessa forma, o limite de discriminação foi aproximadamente 0,5. Nesse gráfico, as amostras com um valor de classe próximo a 1 foram identificadas como cocaína base livre, enquanto que amostras com valor de classe próximos a 0 foram identificadas como cocaína sal.

A comparação dos resultados estimados pelo modelo PLS-DA com os obtidos pelo teste de solubilidade mostra que todas as amostras de validação foram classificadas corretamente pelo modelo, com uma pequena dispersão dos resultados, indicando que o modelo é apropriado para discriminar amostras de cocaína de acordo com as suas formas de apresentação.

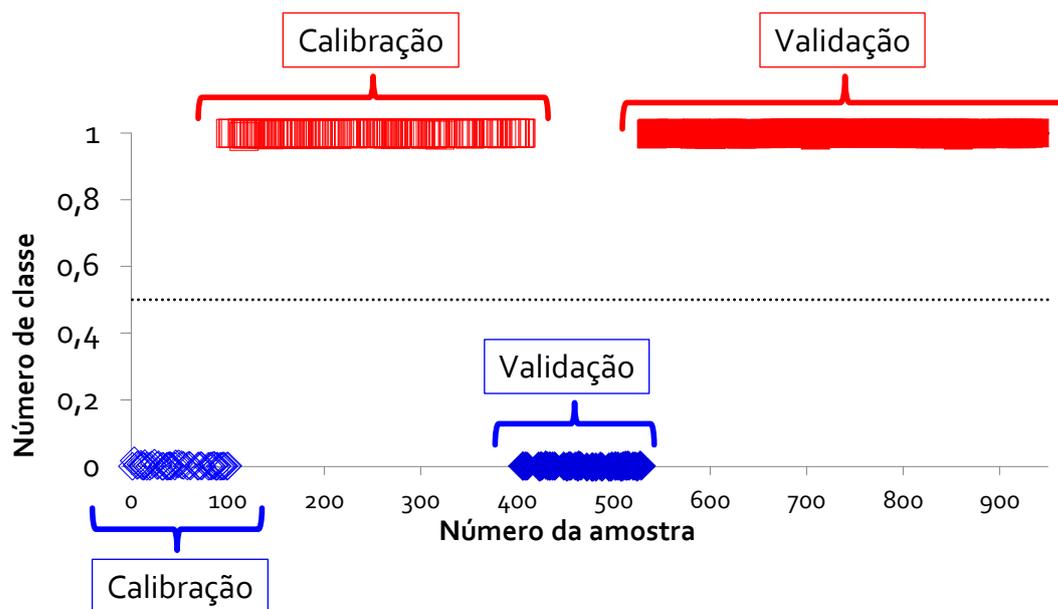


Figura 13: Valores de classe estimados para o modelo de classificação final PLS-DA. Valor de classe = 1: cocaína base livre (□). Valor de classe = 0: cocaína sal (◇). Amostras de calibração (símbolo vazio) e amostras de validação (símbolo preenchido).

#### 4.5. Determinação do teor de cocaína em amostras aprendidas

A otimização do modelo PLSR, utilizado para a determinação do teor de cocaína, feita com a exclusão de *outliers*, mostrou que, mesmo considerando a grande heterogeneidade das amostras de cocaína, o modelo pôde ser aplicado para cerca de 90% das amostras analisadas (Tabela 5). Foram excluídas em média 10,9% das amostras (14,8% das amostras de calibração e 7% das amostras de validação).

Tabela 5. Efeito da eliminação de *outliers* em cada etapa de otimização do modelo de determinação do teor de cocaína.

Modelos	1º	2º	3º	Final
Nº de amostras de calibração	500	470	441	436
Nº de amostras de validação	585	585	585	544
RMSEC	5,43	3,78	3,43	3,02
RMSEP	5,28	5,28	5,28	2,79
VL	8	11	12	13

Para melhor entendimento do efeito da eliminação de *outliers* em cada etapa de otimização do modelo de determinação do teor de cocaína (Tabela 5), é importante ressaltar que, embora algumas amostras tenham sido identificadas como *outliers* com base no critério de *leverage* no terceiro modelo (após a segunda exclusão), elas não foram excluídas do conjunto de dados, uma vez que a norma ASTM E1655-05<sup>46</sup> indica que os

dados podem estar apresentando o “efeito bola de neve”. Nestes casos, a norma sugere que o critério de *leverage* pode ser relaxado desde que nenhuma das amostras de calibração tenham uma influência maior do que 0,5. Como o maior *leverage* observado no modelo otimizado foi de 0,09, essas amostras não foram excluídas. Além disso, as amostras identificadas como *outliers* foram analisadas em detalhe para verificar as razões da sua exclusão. Mais de 60% dessas amostras faziam parte de apreensões realizadas a no mínimo três anos atrás, indicando que pode ter havido alguma alteração nas características químicas dessas amostras. Deve-se destacar que, após a exclusão de *outliers*, houve um decréscimo significativo dos valores RMSEC e RMSEP.

Com valores de RMSEC e RMSEP menores do que 2,80% (m/m), os erros médios mostraram boa concordância entre o teor de cocaína obtido pelo método de referência (CG-DIC) e o teor estimado pelo modelo utilizando FTIR (ver Figura 14). Além disso, o coeficiente de determinação para as amostras de validação gerado foi maior que 0,94 e a precisão do modelo apresentou um valor de 2,29% (m/m), como mostrado na Tabela 6.

Tabela 6. Figuras de mérito analíticas para os modelos PLS de determinação do teor de cocaína.

<b>Figuras de Mérito</b>		
<b>Exatidão</b> (% m/m)	RMSEC	3,02
	RMSEP	2,79
<b>Precisão</b>		2,29
<b>Sensibilidade</b> (% m/m) <sup>-1</sup>		0,100
<b>R<sup>2</sup></b> <sub>validação</sub>		0,946
<b>MDC</b>		10,8
<b>Incerteza média (%)</b>		6,7
<b>Viés</b>		1,91(2,58)*

\*valor de t calculado (valor de t tabelado).

A linearidade do método foi avaliada pelas distribuições dos resíduos do modelo PLS, que é apresentado na Figura 15. Visualmente, por intermédio dessa Figura, é possível verificar que o comportamento destas distribuições é aproximadamente aleatório. No entanto, para verificar a suposição de independência e normalidade dos resíduos, foi aplicado o teste de *viés* e de normalidade de Jarque-Bera<sup>70</sup>. Como resultado desses testes, com 95% de confiança, não existe *viés* significativo nos resíduos e a normalidade das distribuições não pode ser rejeitada.

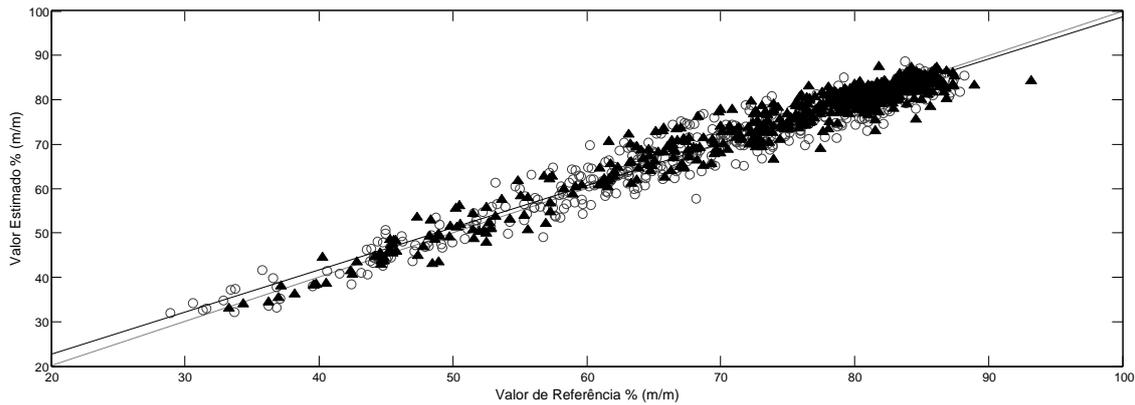


Figura 14: Valores de referência *versus* valores estimados pelo modelo PLS para a determinação do teor de cocaína. Amostras de calibração (○) e amostras de validação (▲).

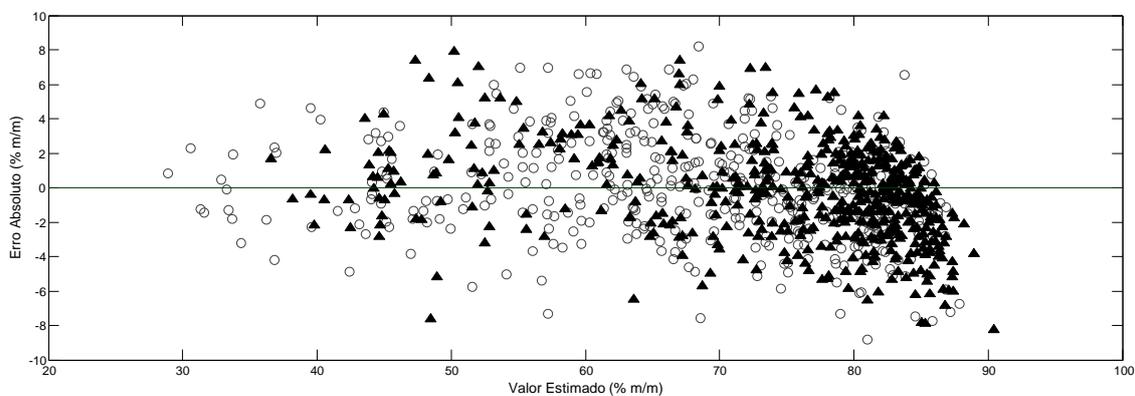


Figura 15: Distribuição dos erros absolutos (resíduos) para o modelo PLS *versus* teor estimado de cocaína. Amostras de calibração (○) e validação (▲).

Conforme a norma ISO 11843-2<sup>58</sup> a MDC representa o menor teor de cocaína que pode ser mensurado de forma confiável pelo método. Os resultados apresentados na Tabela 7 indicam que as amostras contendo teores de cocaína acima de 11% (m/m) podem ser adequadamente estimados pelo método proposto. Este intervalo (>11% m/m) abrange a grande maioria das apreensões da Polícia Federal Brasileira.

A incerteza média do modelo PLSR final estimada de acordo com Pierna *et al.*<sup>48</sup> foi de aproximadamente 7% (m/m) com 95% de confiança. Um gráfico com os intervalos de confiança para algumas amostras de validação é mostrado na Figura 16.

A combinação de todos esses resultados mostrou que a técnica de espectroscopia de ATR-FTIR e PLSR permite a determinação direta do teor de cocaína em amostras de drogas apreendidas. Os baixos erros médios de predição (3,0% (m/m)), precisão de 2,0 e MDC de 11,8% (m/m), em um intervalo variando 24,2 – 99,9% (m/m), os quais

demonstram que o método é capaz de atender às necessidades da Polícia Federal como rotina de triagem a ser utilizada em aeroportos e postos de fronteira.

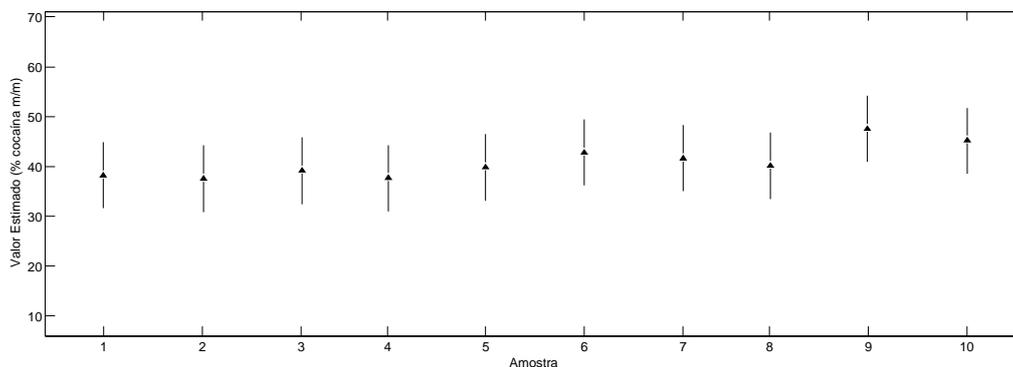


Figura 16: Intervalos de confiança para algumas amostras de validação, com 95% de confiança obtidos para o modelo PLSR.

#### 4.5.1. Transferência do Modelo de Determinação do Teor de Cocaína

A fim de verificar o funcionamento do método em duas novas situações, a transferência do modelo PLSR para determinação do teor de cocaína (Figura 14) foi testada. No primeiro caso, avaliou-se os resultados obtidos pelo mesmo equipamento em que o modelo foi construído (equipamento *Nicolet FTIR iS10* pertencente ao INC) após uma manutenção. No segundo caso, verificou-se o funcionamento do modelo em um equipamento diferente, um espectrômetro de infravermelho com transformada de Fourier da marca *Jasco* modelo 4100 com acessório para medidas por ATR.

Utilizou-se o método de transferência por PDS com o tamanho da janela  $n=5$ . Todos os dados foram pré-processados da mesma maneira em que o modelo original foi construído, com a primeira derivada com uma janela de 15 pontos, SNC e *mean center*.

A avaliação da transferência foi feita com base nos RMSEPs, pois o objetivo da transferência é fazer com que o erro do modelo após a transferência seja o mais próximo possível do erro do modelo original. Para isso, 31 amostras utilizadas na construção do modelo de calibração foram selecionadas. Destas, 10 foram usadas para a transferência dos espectros da condição secundária para o modelo desenvolvido na condição primária. As outras 21 amostras foram utilizadas na validação da transferência. Portanto, as 31 amostras foram analisadas nas condições primária e secundárias (equipamento *Nicolet iS10* após manutenção e equipamento *Jasco 4100*).

Foram apresentados na Tabela 7 os valores de RMSEP para o modelo de calibração que se deseja transferir, ou seja, o modelo estimado com 544 amostras, o

modelo mestre que possui 21 amostras selecionadas do modelo que se deseja transferir e os modelos escravos obtidos pelo equipamento *Nicolet iS10* após manutenção. Pode-se observar que os valores de RMSEP apresentam uma diferença relativamente alta de 2,79% (m/m) para 4,60% (m/m). Essa diferença pode ser justificada pela diferença no número de amostras utilizada para estimar os dois valores e RMSEP.

A Tabela 7 também mostra os valores de RMSEP obtidos para as 21 amostras de validação através da análise pelo modelo PLSR sem e com a transferência na situação em que as medidas foram obtidas pelo equipamento *Nicolet FTIR iS10* após manutenção. Para essa situação, observou-se um RMSEP menor (5,23% (m/m)) para os resultados obtidos sem nenhuma técnica de transferência em comparação ao obtido utilizando a transferência (5,51%). Observa-se que não há diferença significativa entre os valores 4,60; 5,23 e 5,51% (m/m). Portanto, pode-se concluir que após a manutenção do equipamento não foram causadas mudanças significativas no sinal do equipamento, de forma que a transferência não foi necessária.

Tabela 7. RMSEP dos modelos construídos com dados obtidos antes e após manutenção do equipamento *Nicolet iS10* pertencente ao INC.

<b>Modelos</b>	<b>RMSEP % (m/m)</b>
PLSR – <i>Nicolet iS10</i> (com 544 amostras de validação)	2,79
Mestre – <i>Nicolet iS10</i> (com 21 amostras de validação)	4,60
Escravo sem transferência – <i>Nicolet iS10</i> (21 amostras de validação)	5,23
Escravo com transferência – <i>Nicolet iS10</i> (21 amostras de validação)	5,51

Para avaliar a segunda situação, foi necessário excluir a faixa espectral de 400 a 650  $\text{cm}^{-1}$  do espectro do *Nicolet iS10*, pois não havia correspondência nessa região com o espectro obtidos pelo equipamento Jasco 4100. Os equipamentos coletaram os espectros em diferentes faixas espectrais por possuírem cristais de ATR distintos (diamante e ZNSE para *Nicolet iS10* e Jasco 4100, respectivamente) e resoluções distintas. Após essa adequação, o RMSEP foi avaliado com e sem transferência.

Os resultados obtidos mostraram que o modelo construído utilizando a técnica de transferência de calibração apresentou um RMSEP similar ao obtido pelo modelo construído sem transferência, 6,54 e 6,77% (m/m), respectivamente (Tabela 8).

Concluiu-se, portanto, que não houve a necessidade de transferência dos modelos em nenhuma das situações. Dessa forma, o modelo continuou adequado após a manutenção do equipamento. Além disso, a comparação dos RMSEP do modelo original

com os modelos escravos mostrou que a mudança de tipo de equipamento aumentou o erro do modelo. Porém, ainda com essa mudança, foi possível verificar a robustez do modelo possibilitando o compartilhamento de dados entre laboratórios que possuem equipamentos distintos.

Tabela 8. RMSEP dos modelos construídos com dados obtidos dos equipamentos *Nicolet iS10* e *Jasco 4100*.

<b>Modelos</b>	<b>RMSEP % (m/m)</b>
Mestre – <i>Nicolet iS10</i> (com 21 amostras de validação)	4,60
Escravo sem transferência – <i>Jasco 4100</i> (com 21 amostras de validação)	6,77
Escravo com transferência – <i>Jasco 4100</i> (com 21 amostras de validação)	6,54

#### 4.6. Determinação do grau de oxidação

A fim de determinar o grau de oxidação das amostras utilizando FTIR, um modelo para a quantificação da proporção entre o total de cinamoilcoínas e a cocaína das amostras foi construído utilizando PLSR.

Os *outliers* excluídos na otimização do modelo mostraram o modelo podendo ser aplicado para mais de 90,0% das amostras analisadas (Tabela 9).

Tabela 9. Efeito da eliminação de *outliers* em cada etapa de otimização do modelo de determinação do teor de cocaína.

<b>Modelos</b>	<b>1º</b>	<b>2º</b>	<b>3º</b>	<b>Final</b>
Nº de amostras de calibração	500	473	451	444
Nº de amostras de validação	585	585	585	556
RMSEC	1,55	1,45	1,19	1,12
RMSEP	1,51	1,51	1,51	1,12
VL	8	5	6	6

VL, Número de Variáveis Latentes.

Um gráfico com os valores estimados pelo método de FTIR-ATR e PLSR *versus* valores de referência (CG-DIC) é mostrado na Figura 17(A), e apresentou boa concordância do modelo com os resultados experimentais. O gráfico dos resíduos, mostrado na Figura 17(B), indica que não há tendências sistemáticas nos resultados. Os valores de figuras de mérito para o modelo estão apresentados na Tabela 10. A boa concordância do modelo é confirmada pelo coeficiente de determinação nas amostras de validação ( $R^2 = 0,947$ ) com RMSEP de 1,12% (m/m). Este resultado permite concluir que

é possível determinar o grau de oxidação de uma amostra de cocaína diretamente a partir de medidas de FTIR-ATR em associação com um modelo PLSR adequado.

Tabela 10. Figuras de mérito analíticas do modelo PLS para o grau de oxidação.

Figuras de Mérito		
<b>Exatidão</b> (% m/m)	RMSEC	1,12
	RMSEP	1,12
<b>Sensibilidade</b> (% m/m) <sup>-1</sup>		0,764
<b>R<sup>2</sup><sub>validação</sub></b>		0,947
<b>MDC</b>		3,91
<b>Incerteza média (%)</b>		2,3
<b>Viés</b>		2,07(2,58)*

\*valor de t calculado (valor de t tabelado).

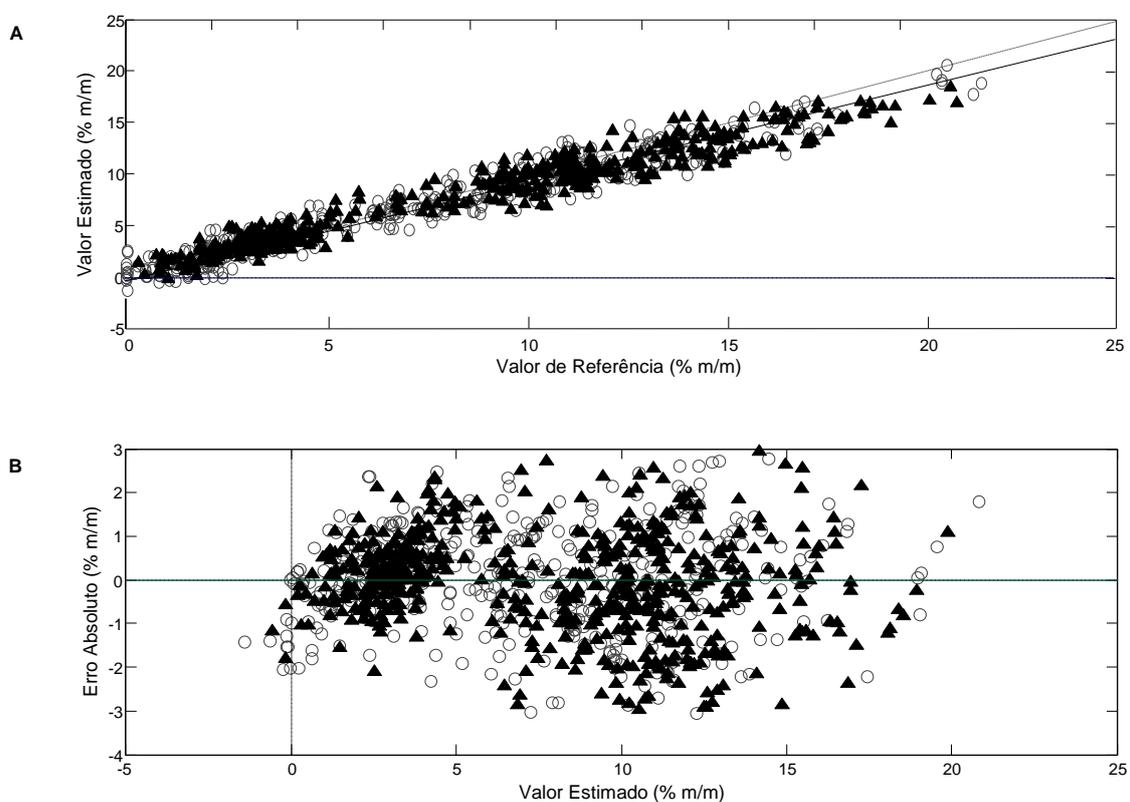


Figura 17: (A) Valores de referência *versus* valores estimados do modelo PLSR para o grau de oxidação. (B) Distribuição dos erros absolutos (resíduos) do modelo PLSR para o grau de oxidação. Amostras de calibração (○) e validação (▲).

#### 4.7. Determinação de adulterantes em amostras de cocaína apreendidas

Tendo em conta o número de amostras que contêm adulterantes, modelos PLSR foram desenvolvidos para determinar o teor dos seguintes adulterantes: fenacetina, benzocaína, cafeína, levamisol, lidocaína e aminopirina. Os gráficos com valores de referência *versus* valores estimados pelo modelo são apresentados na Figura 18. Pela Tabela 11 foi possível observar que foram alcançadas correlações aceitáveis com coeficientes de determinação com valores superiores a 0,83, com exceção do composto levamisol.

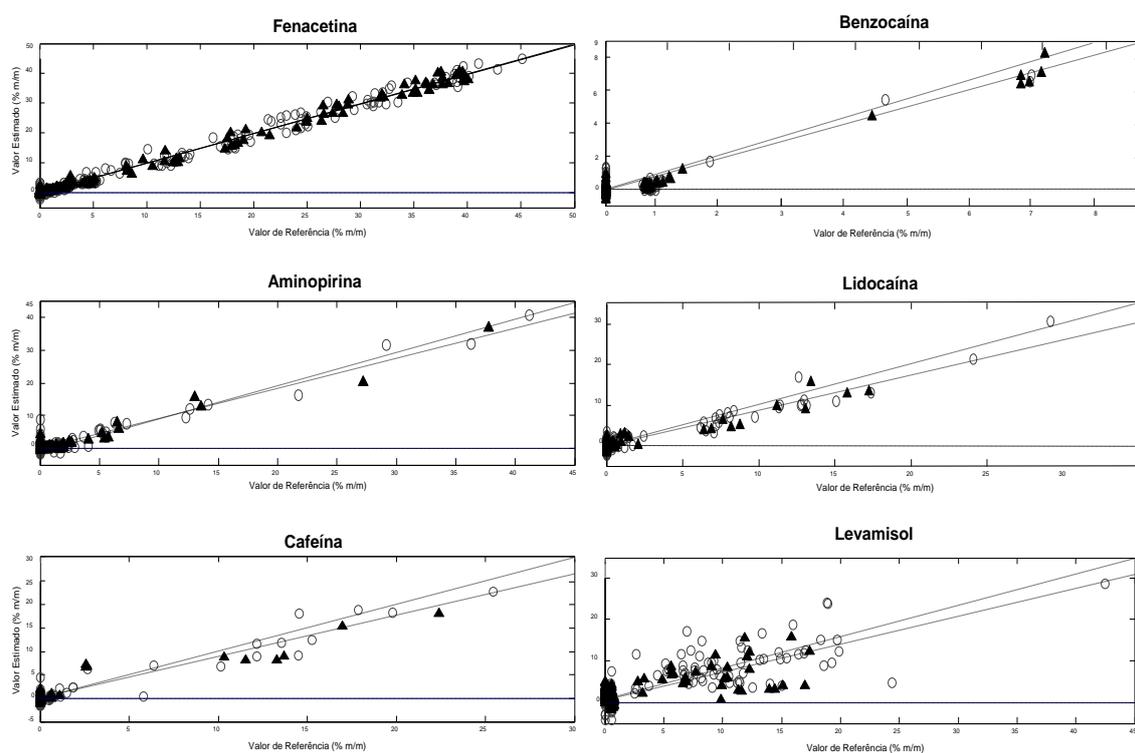


Figura 18: Valores de referência *versus* valores estimados para a determinação de adulterantes pelo modelo PLSR. Amostras de calibração (○) e validação (▲).

Fenacetina foi o adulterante encontrado na maior parte das amostras do conjunto de dados (ver Tabela 3, presente no item 4.3). Como consequência do grande número de amostras disponíveis para o desenvolvimento do modelo PLSR, a fenacetina exibiu os melhores resultados para figuras de mérito quando comparada com os demais adulterantes estudados (Tabela 11).

Tabela 11. Figuras de mérito analíticas dos modelos PLSR para a determinação de adulterantes.

Figuras de Mérito		Fenacetina	Benzocaína	Aminopirina	Lidocaína	Cafeína	Levamisol
Exatidão	RMSEC	1,06	0,39	0,83	1,28	0,75	2,28
(% m/m)	RMSEP	0,73	0,23	0,55	0,72	0,64	1,66
VL*		4	6	4	8	6	8
Sensibilidade (% m/m) <sup>-1</sup>		1,30	2,32	1,77	0,518	1,02	0,46
R <sup>2</sup> validação		0,994	0,926	0,940	0,849	0,834	0,483
MDC (%)		3,5	1,8	2,9	1,1	2,61	10,4
Incerteza média (%)		2,2	0,9	1,8	2,7	2,0	5,0
Vies**		<b>4,36(2,60)</b>	0,17(2,76)	<b>5,42(2,62)</b>	2,68(2,79)	0,38(2,98)	1,94(2,58)

\*Número de variáveis latentes; \*\* valor de t calculado (valor de t tabelado).

Os resultados para benzocaína apresentaram erros médios menores que 0,39% (m/m) e coeficiente de determinação igual a 0,926 (Tabela 11). As principais limitações, neste caso, estão relacionadas com a estreita faixa de concentrações nas amostras do conjunto de dados que apresentaram estreita faixa de variação (de 0 a 6% (m/m)) com distribuição desigual das amostras dentro deste intervalo. Ainda assim, as estimativas do teor de benzocaína fornecidos pelo método FTIR permitiram a detecção desse adulterante nas amostras (até o limite da MDC), fornecendo ao cientista forense uma primeira estimativa sobre o seu conteúdo na amostra (nível de concentração elevado ou baixo). Conclusões análogas podem ser estendidas para o caso da aminopirina, lidocaína e cafeína.

O modelo PLSR para o levamisol apresentou o maior erro de estimativa do teor na amostra. Nesse caso, existe uma grande dispersão entre os valores de referência e os valores estimados, como mostrado na Figura 18. Testes por ANOVA foram realizados para todos os modelos de adulterantes, os quais revelaram que a regressão não foi significativa apenas para o adulterante levamisol. Este é um resultado curioso, pois o levamisol está presente em um número relevante das amostras analisadas (97 amostras) e seu teor nas amostras abrange uma faixa considerável (desde 2,5% a 42,5% (m/m)), logo deveria poder ser quantificado pelo método. No entanto, experiências adicionais envolvendo misturas preparadas com levamisol, cocaína e outros adulterantes demonstraram a absortividade molar relativamente pequena no infravermelho do levamisol com respeito a outros compostos. Este fato pode explicar os elevados erros obtidos para este adulterante.

Os resultados obtidos até aqui, indicam que o método é promissor, mas ainda é necessário investir mais esforços para alcançar a estimativa do teor desses adulterantes

em um maior número de amostras. Isso porque, considerando o MDC, a porcentagem das amostras que podem ser avaliadas pelos modelos construídos para a determinação de adulterantes nas amostras ultrapassa 50% dos casos só para a fenacetina, lidocaína e cafeína, como mostra a Tabela 12.

Tabela 12. Porcentagem de amostras que podem ser avaliadas pelos modelos considerando o MDC.

<b>Adulterantes</b>	<b>% de amostras que podem ser avaliadas pelo modelo considerando o MDC</b>
Fenacetina	59,5
Benzocaína	14,7
Aminopirina	15,6
Lidocaína	57,5
Cafeína	51,1

## 5. Conclusão

Os resultados desse capítulo e o método completo foram publicados na forma de dois artigos científicos. Suas referências foram: “**Quantification of Cocaine Hydrochloride in Seized Drug Samples by Infrared Spectroscopy and PLSR**, Grobério, T.S.; Zacca, J.J.; Talhavini, M.; Braga, J.W.B.; *J. Braz. Chem. Soc.* **2014**, *25*, 1696-1703.” e “**Discrimination and quantification of cocaine and adulterants in seized drug samples by infrared spectroscopy and PLSR**, Grobério, T.S.; Zacca, J.J.; Botelho, É. D.; Talhavini, M. and Braga, J. W. B.; *Forensic Sci. Int.*, **2015**, *257*, 297-306”.

Este Capítulo apresentou uma ampla descrição de um conjunto de dados de amostras de cocaína apreendidas no Brasil, o qual é representativo do cenário de grandes apreensões de cocaína ilícita no país. A maioria das amostras mostrou pureza entre 60 e 90% (m/m), das quais 74,2% foram classificadas como cocaína base livre e 63,3% das amostras analisadas continham de 1 a 4 diferentes tipos de adulterantes.

Também foi demonstrada a utilidade da associação de FTIR com a análise PLS na ampla caracterização de amostras de cocaína, pois o método proposto é capaz de determinar simultaneamente a forma de apresentação da cocaína, o grau de oxidação, o teor de cocaína e de cinco dos adulterantes mais encontrados com valores adequados de figuras de mérito.

O método proposto neste capítulo se mostrou robusto ao longo de análises realizadas por nove meses e após a manutenção do espectrômetro. Além disso, é tão eficaz

quanto os métodos de referência (testes de solubilidade para determinar a forma de apresentação da cocaína e CG-DIC para quantificar cocaína e adulterantes) atualmente utilizados em análises forenses de rotina pela Polícia Federal Brasileira e pode ser útil em termos de redução de tempo e custos de análise.

Dessa forma, concluiu-se que este método é eficiente, versátil e tem capacidade de reduzir significativamente o tempo e o custo de análise em relação à análise cromatográfica. Além disso, não gera resíduos químicos beneficiando o meio ambiente. Apesar de cromatografia gasosa ser frequentemente utilizada por oferecer resultados mais precisos e exatos, a espectroscopia no infravermelho mostrou precisão semelhante, aceitável exatidão e capacidade de detecção numa faixa linear de (35% a 99%(m/m)), preenchendo os requisitos para a sua aplicação em laboratórios forenses.

Além disso, os resultados exibidos neste capítulo empregaram uma grande quantidade de amostras para a previsão e validação do modelo, permitindo uma amostragem realista e representativa de amostras de cocaína apreendidas no Brasil.

# *Capítulo 2*

*Correlação entre amostras de  
cocaína base livre*

## **Correlação entre Amostras de Cocaína Base Livre**

Como destacado anteriormente, a caracterização das amostras de cocaína por meio da determinação do teor de cocaína, adulterantes e grau de oxidação é de grande importância uma vez que fornece informações sobre o modo ilegal de operação das redes de distribuição de drogas em uma determinada área. Ao mesmo tempo, outro aspecto relevante é a identificação de correlações entre investigações policiais as amostras apreendidas para estabelecer ligações entre apreensões, tentar estabelecer a origem das drogas e inferir sobre as rotas mais comuns de tráfico de drogas.

### **Objetivos**

- Estabelecer uma metodologia que permita a correlações entre apreensões de cocaína de cocaína na forma de base livre efetuadas em todo o Brasil.
- Avaliar estratégias para analisar os dados de cromatografia gasosa acoplado com espectrometria de massas (CG-EM) para a correlações entre apreensões de cocaína de cocaína na forma de base livre.
- Propor uma abordagem estatística para estabelecer valores limites que possam decidir sobre a correlação entre amostras de cocaína base sem a necessidade de um conhecimento da população a priori.

## **1. Referências Bibliográficas**

### **1.1. Processo de Produção de Cocaína**

A compreensão de como pode ser feita a identificação de correlações entre amostras é iniciada entendendo o processo de produção ilícita de cocaína, que é composta por várias etapas tais como maceração de folhas de coca, extração por solvente, conversão ácido-base e purificação, como mostrado na Figura 19.

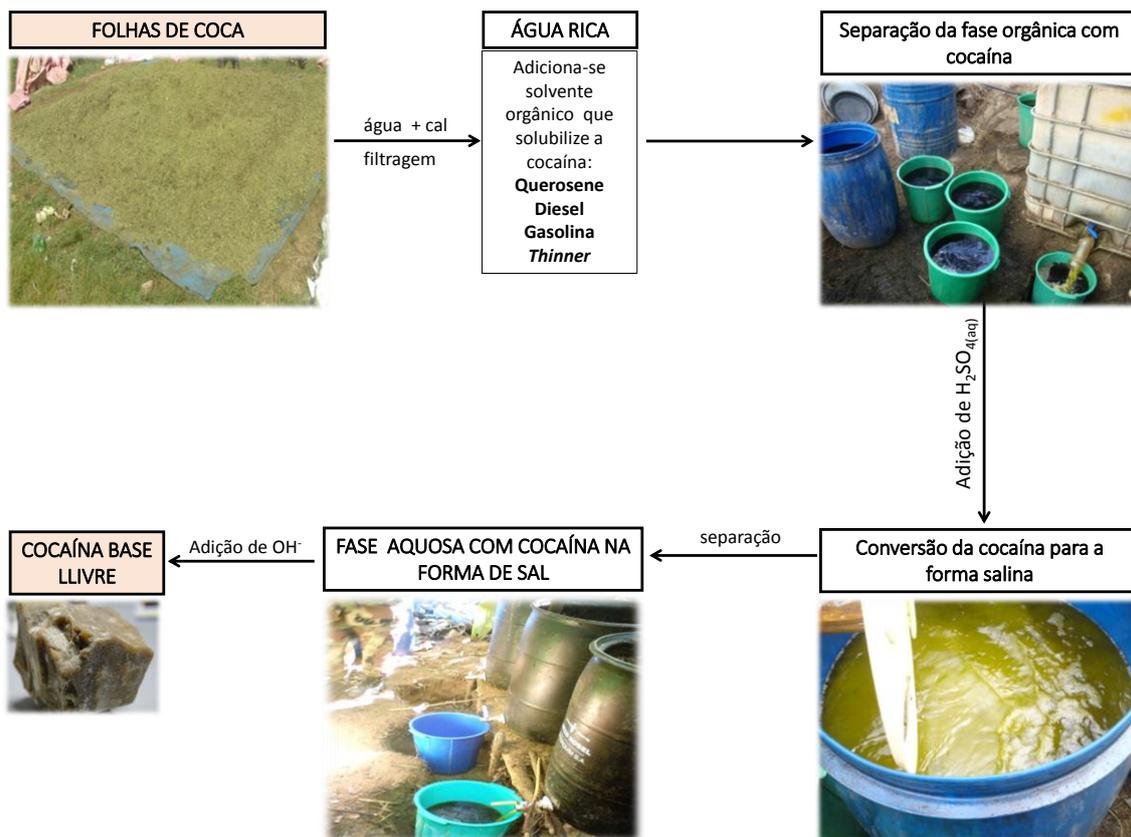


Figura 19: Processo de produção de cocaína.

O processo ilustrado na Figura 19 deixa traços dos solventes utilizados na produção da cocaína, que podem caracterizar cada lote de produção da droga com um perfil químico potencialmente diferente. A análise desses solventes em apreensões de cocaína ilícita tem sido proposta como uma das técnicas eficazes para identificação de perfil químico da droga.<sup>71,73</sup> A partir dele, identificam-se similaridades entre duas amostras quaisquer, revelando quais amostras podem ser consideradas correlacionadas ou não. Poder afirmar, com níveis adequados de confiança estatística que amostras estão correlacionadas significa dizer que estas vieram de um mesmo lote de produção de drogas, ou seja que estão ligadas ou correlacionadas.

A análise de solventes residuais em amostras de cocaína sal já vem sendo estudada e encontra-se relativamente bem descrita na literatura<sup>73,74</sup>. Contudo, até o presente, a identificação de solventes residuais em amostras de cocaína base livre ainda não foi descrita na literatura com clareza; porém, seu estudo é necessário porque a maior casuística de apreensões de cocaína feitas no Brasil pela Polícia Federal é de amostras com essa forma de apresentação.

Dessa forma, este Capítulo procurou aprimorar o método de determinação de solventes residuais em amostras de cocaína base livre apreendidas pela Polícia Federal do Brasil e correlacionar essas amostras através de métodos quimiométricos.

## 1.2. Metodologia para Análise de Solventes Residuais em Drogas Ilícitas

Metodologias que identifiquem os perfis químicos de amostras de drogas têm sido desenvolvidas há cerca de 30 anos.<sup>73</sup> No entanto, ainda existem poucos grupos de pesquisa que trabalham com essas amostras, tendo em vista seu acesso restrito.

Algumas técnicas analíticas que têm sido utilizadas para a identificação do perfil químico de amostras de cocaína são *headspace* acoplada com cromatografia gasosa com detector de ionização em chama (HS-CG-DIC), HS-CG-EM e microextração em fase sólida acoplada com cromatografia gasosa com detecção por espectrometria de massas (MEFS-CG-EM).<sup>73,75,76</sup> Essas técnicas determinam, dentre outras coisas, alcaloides tropânicos<sup>77</sup> e truxilinas<sup>78</sup> presentes nas amostras – que fornecem informações a respeito das condições climáticas de cultivo e das espécies das folhas de coca utilizadas para extrair a cocaína, identificando a possível origem da amostra. E também determinam compostos voláteis (solventes orgânicos residuais) – que revelam o método de refino utilizado na produção da amostra.<sup>74,76,79</sup>

Os pioneiros na busca do perfil químico através de compostos voláteis em amostras de drogas foram Chiarotti e Fucci<sup>73</sup> que iniciaram sua pesquisa com 32 amostras de heroína. Tais amostras foram analisadas com HS-CG-DIC e os resultados mostraram que o perfil químico de compostos voláteis pode ser usado para comparar amostras de rua de heroína. Esses autores consideram que essa é uma análise simples e sensível, que permite sua aplicação com uma pequena quantidade de amostra, além de ser uma importante ferramenta para afirmar ou excluir origens comuns de amostras.

Cartier *et al.*<sup>80</sup> fizeram análise de solventes residuais em amostras de cocaína sal usando HS-CG-DIC e HS-CG-EM e correlacionaram algumas dessas amostras. Observaram-se ainda, algumas tendências e variações geográficas no uso de solventes e com isso, sugeriram recomendações no que diz respeito ao controle de certos solventes frequentemente encontrados em amostras.

Morello e Meyers<sup>74</sup> estabeleceram em seu trabalho a metodologia experimental para a determinação qualitativa e quantitativa de solventes residuais em cocaína sal. A metodologia foi desenvolvida utilizando HS-CG-EM e a liberação dos solventes ocluídos foi otimizada através da completa solubilização das matrizes cristalinas em solução

aquosa de sulfato de sódio 22%. Nesse estudo, 75 amostras de cocaína foram analisadas. Puderam ser identificados e quantificados 25 dos solventes ocluídos mais comumente encontrados em amostras de cocaína sal e heroína.

Outros grupos de pesquisa deram continuidade a esses estudos incluindo além das amostras de drogas<sup>81</sup>, amostras de saliva<sup>82</sup> e cabelo<sup>83</sup>. O foco das pesquisas passou então a ser como tratar essa grande quantidade de dados. Houve um intenso interesse em formalizar uma medida adequada para verificar similaridades entre amostras. Nessa busca, algumas métricas foram testadas, tais como: distancia euclidiana<sup>84</sup>, PCA<sup>76</sup>, função cosseno quadrado<sup>85</sup>.

Janzen *et al.*<sup>84</sup> propuseram um método para comparação de amostras ilícitas de cocaína em que os dados brutos das áreas de picos cromatográficos de quatro alcaloides da cocaína (tropacocaína, norcocaína, cis e trans-cinamoilcocaína) foram utilizados para cada amostra. Esses valores foram compilados em uma base de dados computadorizada e com um programa escrito em *BASIC*. As distâncias euclidianas entre uma amostra de teste e as contidas na biblioteca foram avaliadas para localizar as amostras da biblioteca que estavam mais próximas da amostra de teste. Essa metodologia permitiu a comparação entre amostras e tornou possível obter conclusões em relação ao grupo de origem da amostra.

Em 2002, Chiarotti *et al.*<sup>76</sup> fizeram a avaliação estatística dos resultados obtidos na análise de solventes residuais ocluídos em cristais de cocaína sal utilizando o software UNSCRAMBLER e as amostras foram agrupadas utilizando PCA. A maioria dos grupos obtidos indicaram similaridades de composição química da droga em termos de solventes majoritários e minoritários. Os autores concluíram que essa análise constitui em um indicador útil para agrupar amostras em várias classes de acordo com diferentes tipos de solventes residuais e identificar diversos processos de fabricação clandestinos usados para preparar cocaína ilícita.

Margot *et al.*<sup>85</sup> utilizaram a função cosseno quadrado para a comparação entre amostras. Nessa função, zero é o valor que caracteriza amostras sem nenhuma correlação e 1 o valor que caracteriza uma perfeita correlação, ou seja, as amostras apresentam grande similaridade química. Para o uso dessa função, os cromatogramas de cada amostra foram convertidos em vetores, onde as áreas dos picos cromatográficos dos solventes transformam-se em componentes vetoriais e o cosseno quadrado do ângulo entre esses vetores é utilizado para verificar a similaridade entre as amostras.

A função cosseno quadrado continuou sendo utilizada por outros pesquisadores. Esseiva *et al.*<sup>86</sup> discriminaram amostras de heroína por intermédio de medidas de cosseno quadrado. Consideraram amostras como sendo ligadas quando o valor do cosseno quadrado entre o ângulo dos vetores formados pelo cromatograma multiplicado por um fator de 100, fosse maior ou igual a 99,8.

Dujourdy e Besacier<sup>79</sup>, também utilizaram a função cosseno quadrado e um conjunto de amostras conhecidas, calcularam todas as possíveis correlações dentro dos grupos de amostras ligadas (correlações intraclasse) e todas as correlações entre as amostras consideradas não ligadas (correlações interclasses). Construíram então um histograma e, após análise dos resultados, escolheram 0,994 como valor limite para discriminar amostras de cocaína ligadas e não ligadas.

Lociro *et al.*<sup>87</sup> buscaram a otimização e harmonização da metodologia analítica e estatística que determina solventes residuais em amostras de cocaína (CG-DIC) e que compara essas amostras (função cosseno quadrado). Dois laboratórios forenses participaram dessa pesquisa nas cidades de Lausanne e Lyon. O objetivo era melhorar a análise cromatográfica das amostras de cocaína para criação de perfis químicos e comparar amostras apreendidas e analisadas em dois laboratórios diferentes. Alguns parâmetros de validação importantes foram testados para verificar o método desenvolvido e demonstrar a sua capacidade: a seletividade do método, a reprodutibilidade do tempo de retenção, a escolha de um agente de derivatização que melhorasse a cromatografia, a influência de adulterantes e diluentes (efeito de matriz), a influência das condições de armazenamento da amostra e a quantidade de amostra pesada para análises. Os resultados obtidos mostraram que, uma vez que o método foi otimizado, a metodologia é robusta para poder permitir a comparação interlaboratorial e o controle jurisdicional sobre os dados. Nesse trabalho, foram utilizadas curvas ROC (do inglês *Receiving Operating Characteristics*) e a medida da área sob essas curvas ROC (AUC, do inglês *area under the curve*) foram utilizadas para avaliar a eficiência dos métodos analíticos. A AUC permite determinar os valores de cosseno quadrado que melhor discriminam amostras ligadas das não ligadas considerando a melhor relação entre as taxas de falsos positivos e verdadeiros positivos.

Esseiva *et al.*<sup>88</sup> continuaram o trabalho apresentando reflexões sobre os aspectos estatísticos de perfis de drogas ilícitas e, mais especificamente, sobre o cálculo do limiar para a determinação de apreensões ligadas ou não. Um estudo aprofundado da evolução da taxa de verdadeiro positivo e a taxa de falsos positivos em comparações de amostras

de heroína e de cocaína demonstrou que não existe uma abordagem universal disponível e que os cálculos têm que ser reavaliados para cada nova aplicação específica.

Todos os trabalhos citados anteriormente apresentam como característica comum o uso de conjuntos de amostras controladas ou conhecidas para a análise de correlações e estabelecimento dos limites que determinam se amostras estão ou não ligadas. Essas estratégias têm como desvantagem a necessidade de conjuntos conhecidos, os quais dependem de informações policiais precisas e exatas sobre as amostras apreendidas. Nesse sentido, essas estratégias podem ser consideradas métodos supervisionados. Contudo, em muitas situações reais essas informações não se encontram disponíveis. Além disso, devido a grande mudança observada nos perfis de adulteração ou refino de drogas, um estudo realizado com amostras apreendidas em um determinado ano pode já não ser representativo do perfil das drogas em anos seguintes. Nesse sentido, o uso de métodos não supervisionados e que possam ser facilmente atualizados ao longo do tempo é de grande importância.

Levando em conta os aspectos destacados no parágrafo anterior, em 2013, um esforço de pesquisa entre a PF e a UnB resultou em um novo método não supervisionado para a determinação dos limites de correlação entre amostras de cocaína, sem o uso de qualquer informação policial ou pericial previamente conhecida. Esse método chamado de ULT (Não Supervisionado de Limite de Correlação, do inglês *Unsupervised Linkage Threshold*), baseia-se em análises quimiométricas e foi aplicado para a análise de dados de HS-CG-EM de amostras de cocaína sal apreendidas pela Polícia Federal Brasileira.<sup>89</sup> Encontra-se descrito na dissertação intitulada “*Desenvolvimento de uma metodologia para comparação de amostras de sal de cocaína pela determinação de solventes residuais e análise quimiométrica*” e no artigo “*Correlation of Cocaine Hydrochloride Samples Seized in Brazil Based on Determination of Residual Solvents: An Innovative Chemometric Method for Determination of Linkage Thresholds*” publicado em 2013<sup>90</sup>. O método combinou a análise populacional de coeficientes de correlação obtidos pela função cosseno quadrado de pares de amostras com análise hierárquica de agrupamentos (HCA, do inglês *Hierarchical Cluster Analysis*). O desenvolvimento e a validação do método foram feitos com dois conjuntos de dados específicos. Num primeiro passo, essa abordagem foi aplicada à análise de um conjunto de dados simulados criado a partir de nove classes características de amostras e adição de ruído aleatório. Depois, o método foi aplicado e validado com dados HS-CG-EM obtidos a partir de mais de 250 amostras de cocaína sal apreendidas pela Polícia Federal Brasileira. Em uma aplicação real desse

método, puderam ser estabelecidas conexões significativas entre apreensões feitas em quatro diferentes Estados (Ceará, Pará, Paraná e Goiás), locais onde a correlação era aparentemente improvável. Estas correlações foram confirmadas por informações subsequentes e por análises químicas mais abrangentes. Devido a sua generalidade, os autores argumentaram que esse método também pode ser aplicado a outras áreas da perícia forense, onde o conhecimento disponível da população é limitado e os conjuntos de dados são continuamente modificados com a entrada de nova informação.

Em todo o levantamento bibliográfico as pesquisas mais relevantes foram feitas apenas com cocaína sal. A análise de solventes residuais em cocaína na forma de base livre não tem sido abordada com frequência por representar um percentual menor nas apreensões feitas nos países que mais tem investido em pesquisas nessa área (países como Estados Unidos, França, Suíça, Alemanha). Além disso, esse tipo de amostra frequentemente representa um desafio analítico maior pois é de difícil manuseio e homogeneização e ainda há a ocorrência de misturas complexas utilizadas como solventes (ex: derivados do petróleo como gasolina, diesel e querosene). No entanto, é sabido que, no Brasil, a maior quantidade de casos de apreensões de cocaína é composta de amostras na forma de base livre, justificando o empenho em analisar tais amostras.

## **2. Fundamentação Teórica**

O estabelecimento de valores limites que discriminem amostras correlacionadas e não correlacionadas foi feito utilizando o método ULT descrito a seguir.

### **2.1. Método ULT**

O método ULT, desenvolvido em 2012 e publicado em 2013<sup>90</sup> é utilizado para determinar os valores limites a partir do qual as amostras são consideradas correlacionadas ( $\rho_{NL}$ ) e antes do qual as amostras são consideradas não correlacionadas ( $\rho_L$ ). Esse método envolve diversas etapas. Inicia-se pela determinação de solventes residuais presentes nas amostras utilizando o HS-CG-EM. Os dados obtidos nessas análises são transformados em uma matriz, onde cada linha corresponde a uma amostra e cada coluna apresenta um parâmetro de interesse. Utilizando essa matriz, cada amostra é transformada em um vetor e calcula-se o cosseno quadrado do ângulo entre os vetores característicos entre pares de amostras ( $\rho$ ). Dessa forma, é possível verificar a similaridade entre as amostras matematicamente. No entanto, como não é conhecido, *a priori*, quais amostras estão correlacionadas, o método ULT sugere que as amostras sejam agrupadas utilizando a

HCA para que sejam conhecidas todas as possibilidades agrupamento das amostras. Estuda-se então, os valores de cosseno quadrado em todos os agrupamentos (classes) de amostras entre as amostras consideradas ligadas (intraclasse) e não ligadas (interclasses). Construiu-se histogramas com os valores de  $\rho$  para as amostras consideradas ligadas e outro para as amostras consideradas não ligadas, isso para todas as possibilidades de agrupamentos indicadas pela HCA. Esses histogramas de amostras ligadas e não ligadas se sobrepõem e quanto menor for essa sobreposição, melhor o agrupamento formado pela HCA, ou seja, as amostras que de fato apresentam similaridade estarão reunidas em uma classe. A avaliação da zona de sobreposição entre a curva das amostras ligadas e não ligadas no histograma é avaliada utilizando a curva ROC. A menor sobreposição é determinada calculando a AUC. Quanto maior a AUC, menor a zona de sobreposição e melhor o agrupamento formado pela HCA. Finalmente, identifica-se os valores de  $\rho$  analisando o início do histograma de amostras consideradas ligadas e o fim do histograma de amostras consideradas não ligadas para o agrupamento (número de classes) indicado pela AUC. Um fluxograma completo do método ULT é apresentado na Figura 20. Nas próximas seções foi feito um detalhamento maior do método ULT.

Para o desenvolvimento do método ULT, foram analisadas amostras de cocaína sal, que possuem um perfil químico bem diferente das amostras de cocaína base livre. Por isso, o uso desse método no presente trabalho, teve como objetivo testá-lo em amostras de cocaína base livre, buscando uma generalização do método de forma que o mesmo pudesse ser aplicado em uma gama maior de situações e analitos.

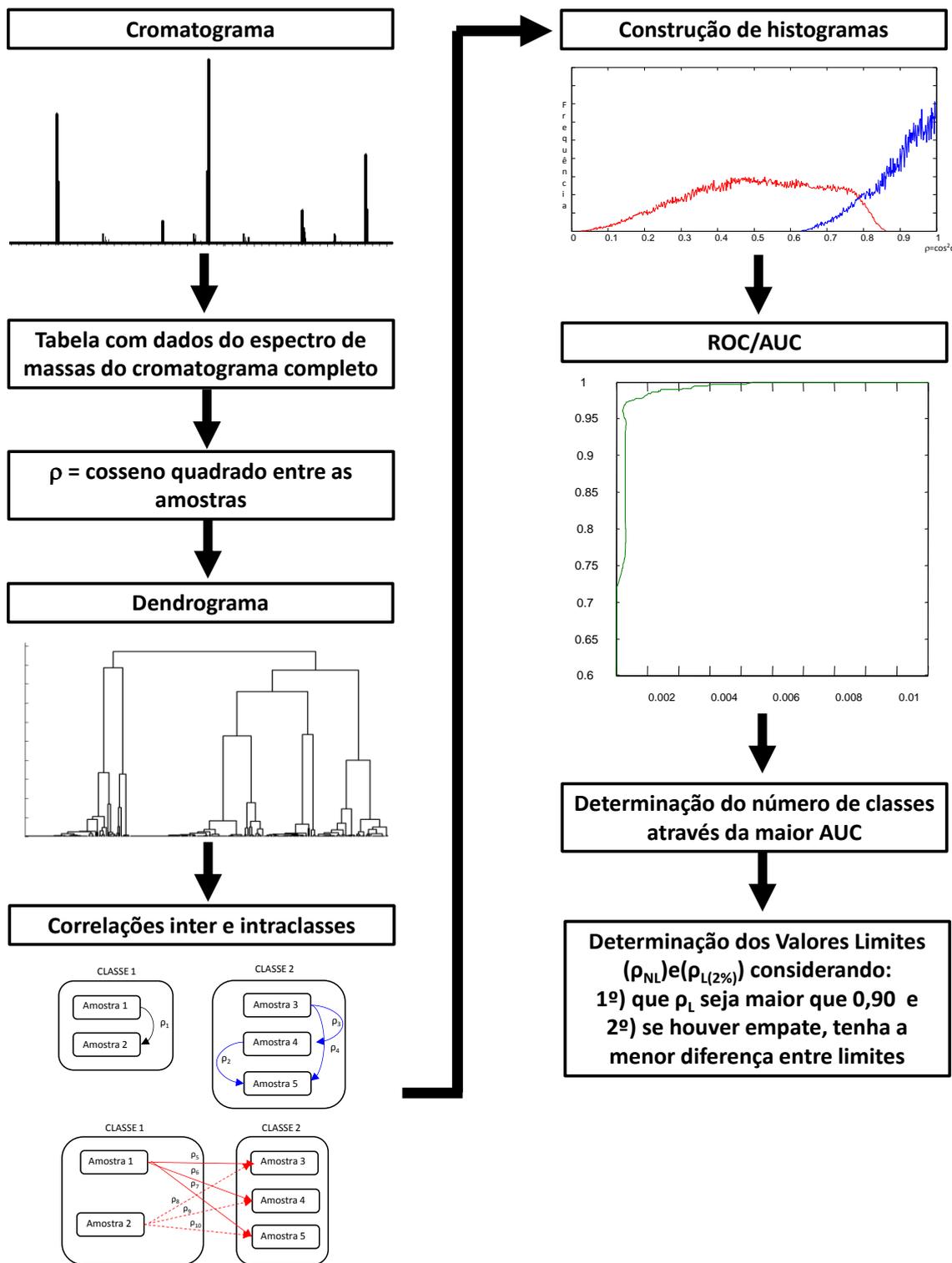


Figura 20: Fluxograma do Método ULT para determinação dos limites entre amostras correlacionadas e não correlacionadas.

## 2.2. Função Cosseno

O cálculo da função cosseno quadrado ( $\rho = \cos^2(\Theta)$ ) considera a área de cada solvente presente numa amostra como componentes de um vetor (Figura 21). Ao comparar dois vetores representantes de um par de amostras é possível obter o cosseno quadrado do ângulo entre eles, que varia de 0 a 1. Quanto mais próximo esse valor estiver de 1, maior a correlação entre amostras. A função cosseno foi escolhida porque independentemente do tamanho do vetor, o ângulo entre dois vetores não varia.

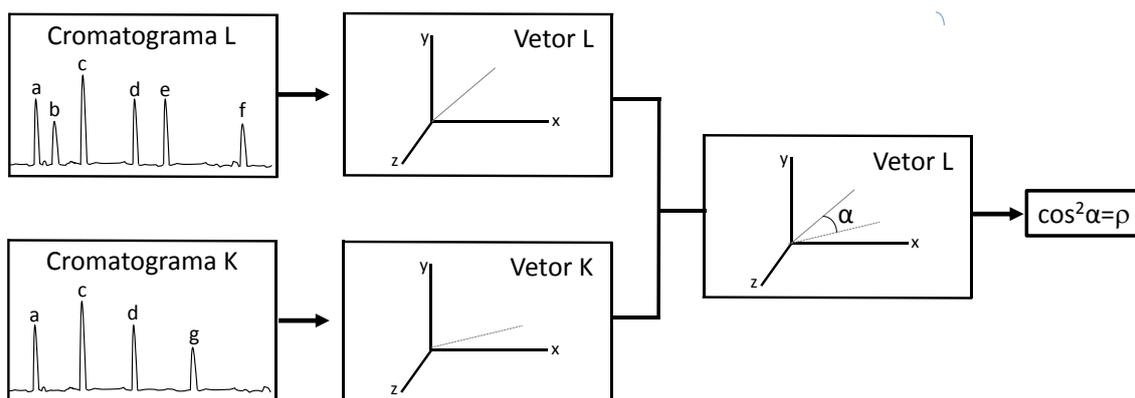


Figura 21: Construção de vetores a partir do cromatograma e comparação de vetores característicos de duas amostras.

Embora a determinação das correlações entre todos os pares de amostras possa ser obtida unicamente com o cálculo de  $\rho$  ( $\cos^2(\Theta)$ ), para que esses valores façam sentido, é necessário estabelecer limites estatisticamente confiáveis capazes de discriminar amostras correlacionadas das não correlacionadas.

A princípio, caso fossem conhecidas quais amostras estão ou não correlacionadas, tais valores limites poderiam ser identificados, mas em um estudo não supervisionado não se dispõe dessa informação *a priori*. Em amostras reais, provenientes de apreensões, tais informações são obtidas somente com o uso de informações policiais e periciais, as quais comumente encontram-se fragmentadas e sem conexão aparente. Portanto, deve-se buscar um método que determine os valores limites de correlação através da composição química da amostra (solventes residuais), permitindo que se possa indagar a respeito de possíveis rotas de tráfico e locais de extração e refino de cocaína.

A solução para os casos reais, onde há a falta de informação *a priori*, foi testar todas as combinações possíveis através de um método sistemático, ou seja, agrupando as amostras em quantas classes forem necessárias para descobrir os valores de  $\rho_L$  e  $\rho_{NL}$  de

acordo com critérios específicos (Figura 22). Esse agrupamento foi feito utilizando a HCA.

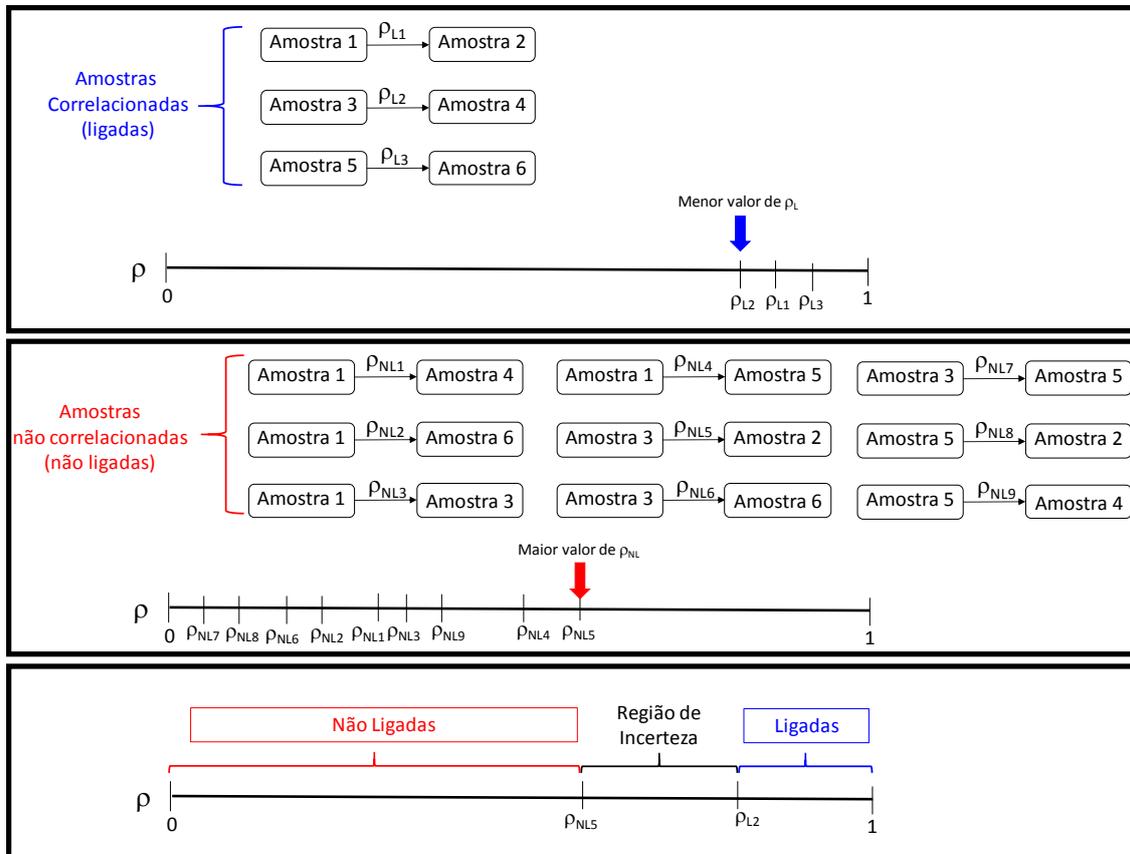


Figura 22: Identificação de  $\rho_L$ ,  $\rho_{NL}$  e região de incerteza, caso se conheça *a priori* quais amostras estão correlacionadas.

### 2.3. Análise Hierárquica de Agrupamentos – HCA

A HCA é utilizada em análises exploratórias dos dados com o intuito de identificar possíveis agrupamentos e o valor provável do número de grupos ou classes.<sup>91</sup> É uma técnica que se baseia em uma medida de proximidade entre as amostras a serem agrupados, normalmente expressa como uma função da distância entre os dados. Essas podem ser medidas de similaridade ou dissimilaridade.

Dentre as medidas de dissimilaridades mais usuais, a distância Euclidiana e a distância de Mahalanobis se destacam. A distância Euclidiana entre duas amostras ( $i$  e  $j$ ) é definida pela raiz quadrada do somatório dos quadrados das diferenças entre os valores das coordenadas das amostras  $i$  e  $j$  para todas as variáveis ( $k = 1, 2, \dots, p$ )<sup>92</sup>:

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (19)$$

onde  $x_{ik}$  representa o valor da medida instrumental da amostra  $i$  para a variável  $k$ ;  $x_{jk}$  representa a característica da amostra  $j$  e variável  $k$ ; e  $p$  representa o número de variáveis nos dados.

A distância de Mahalanobis é uma métrica que difere da distância Euclidiana por levar em consideração a correlação entre os conjuntos de dados. Seu uso corrige algumas das limitações da distância Euclidiana, pois considera a variabilidade dentro de cada unidade amostral, e não somente a medida de tendência central. Com isso, reduz a dependência das unidades de medição e a correlação entre variáveis, sendo, portanto, uma medida mais utilizada quando as variáveis são correlacionadas. De forma semelhante, outras métricas também podem ser empregadas.<sup>76,84</sup> Entre elas está o quadrado da função cosseno entre as amostras, a qual já foi utilizada em alguns trabalhos para a correlação de amostras de drogas.<sup>79,86</sup> Métricas angulares, como  $\rho$ , possuem a vantagem de serem independentes da escala dos vetores envolvidos.

Após a definição da métrica, para a execução da HCA deve ser escolhido ainda o critério de ligação entre amostras, sendo os mais utilizados os seguintes<sup>92</sup>: ligação simples (conhecido também como critério do vizinho mais próximo); ligação completa (conhecido também como critério do vizinho mais distante); ligação média e critério de Ward. A ligação média pode ser definida de diferentes maneiras tais como: método de grupo por par ponderado com média aritmética (WPGMA, do inglês *weighted pair group method with arithmetic mean*), método de grupo por par não ponderado com média aritmética (UPGMA, do inglês *unweighted pair group method with arithmetic mean*), método centroide por par não ponderado (UPGMC, do inglês *unweighted pair group method with centroid mean*) e método centroide por par ponderado (WPGMC, do inglês *weighted pair group method with centroid mean*).<sup>93</sup>

Nesse trabalho, para a identificação das classes onde foram estudados os valores de  $\rho_L$  e  $\rho_{NL}$ , a HCA foi calculada com base no critério da ligação média calculada pelo método WPGMA. Neste método, a distância entre dois grupos é tratada como a média aritmética ponderada das distâncias entre todos os pares de elementos que podem ser formados com os elementos dos dois grupos que estão sendo comparados.

Os resultados da HCA são apresentados na forma de um dendrograma que agrupa amostras em função da similaridade ou dissimilaridade. Neste diagrama, a escala varia de zero (amostras sem similaridades) a um (amostras com elevada similaridade).

Dessa forma, após a identificação dos possíveis agrupamentos, determinam-se os valores de  $\rho_L$  (menor valor intraclasses atribuído para as amostras consideradas ligadas) e  $\rho_{NL}$  (maior valor interclasses atribuído para as amostras consideradas não ligadas).

## 2.4. Correlações Intra e Interclasses

A Figura 23 apresenta um exemplo de correlações hipotéticas intraclasses e interclasses envolvendo 5 amostras divididas em dois agrupamentos, que serão chamados de classes.

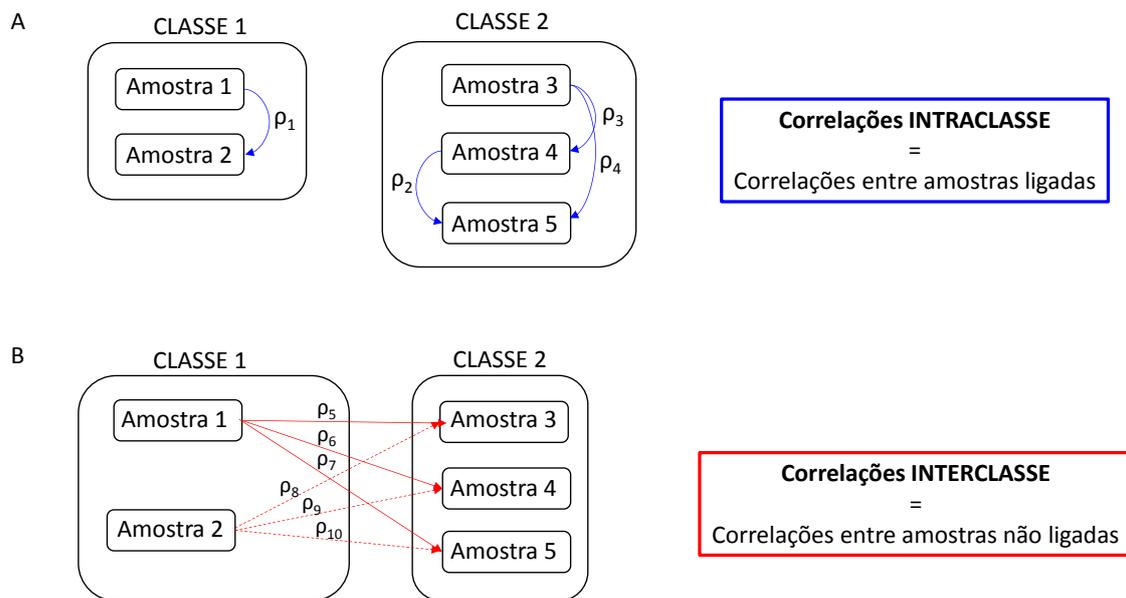


Figura 23: Correlações de amostras (A) intra e (B) interclasses.

## 2.5. Histogramas

Para cada divisão do conjunto de dados em um número específico de classes é gerado um histograma para amostras consideradas ligadas e não ligadas e que apresentam valores característicos de  $\rho_L$  e  $\rho_{NL}$  para esse número de classe. O formato dos histogramas traz diversas informações sobre o comportamento dessa divisão e mostra tendências que indicam se o número de classes analisado é um bom candidato à separação adequada entre as amostras.

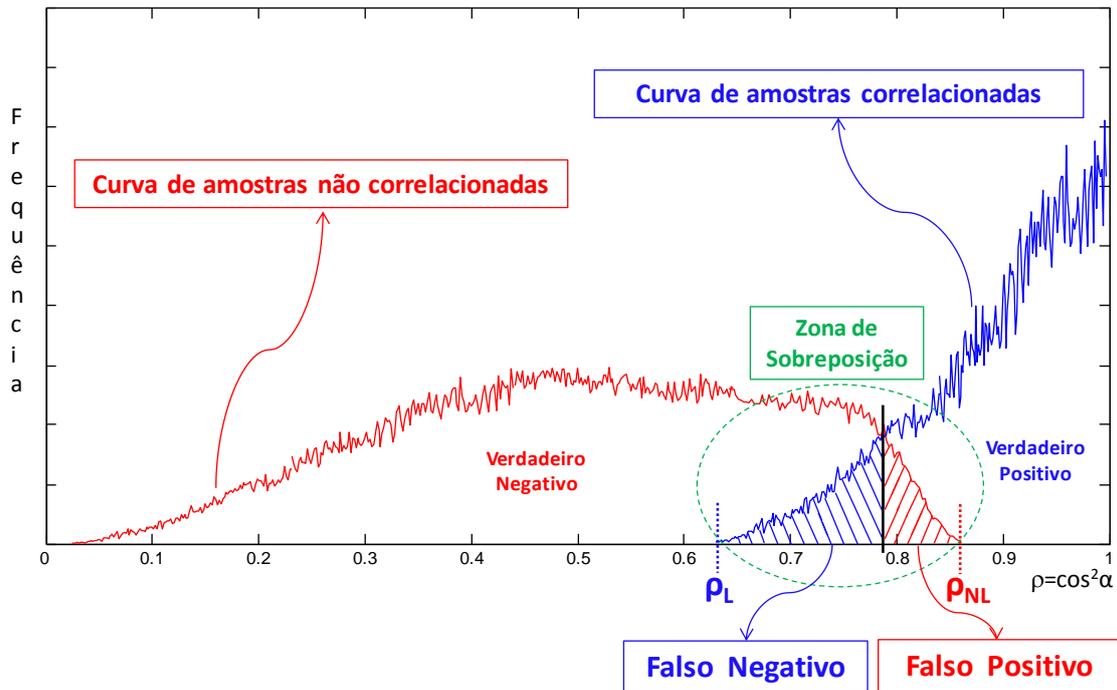


Figura 24: Histograma de amostras ligadas e não ligadas, erros falsos positivos e falsos negativos, zona de sobreposição entre  $\rho_L$  e  $\rho_{NL}$ .

Ao plotar um histograma como o da Figura 24, observam-se quatro possíveis resultados. O primeiro é o chamado verdadeiro positivo (VP), que ocorre quando o cosseno quadrado ( $\rho$ ) entre amostras é maior que o valor de  $\rho_{NL}$ , indicando que as amostras estão correlacionadas, ou seja, a princípio teriam sido produzidas no mesmo laboratório. O segundo resultado ocorre se as amostras não forem correlacionadas e o cosseno quadrado for maior que o valor de  $\rho_{NL}$ , o que resulta em um falso positivo (FP). Um verdadeiro negativo (VN) ocorre quando as amostras possuem  $\rho$  menor que  $\rho_L$  e não estão correlacionadas. Amostras com  $\rho$  abaixo de  $\rho_L$  que estão correlacionadas são classificadas como falso negativo (FN). Quanto menor for a taxa de FN e FP melhor será a divisão das amostras pelo número de classes escolhido. Diminuir a taxa de FP e FN, do ponto de vista judicial, é importante para reduzir o risco de fornecer informações falsas ao processo de persecução penal, não incriminando um inocente nem deixando impune um culpado. Ao diminuir essas taxas, minimiza-se a zona de sobreposição entre as curvas dos histogramas de amostras ligadas e não ligadas. A menor zona de sobreposição ocorre quando o número de classes de amostras agrupa somente amostras que realmente estão correlacionadas, ou seja, amostras que apresentam composição de solventes residuais muito similares e provavelmente teriam sido produzidas em um mesmo lote e laboratório.

## 2.6. ROC e AUC

Um das ferramentas matemáticas que permite a comparação das diversas áreas de sobreposição das curvas dos histogramas para verificar qual o número de classes que melhor divide as amostras correlacionadas em grupos é a curva ROC.<sup>94</sup>

Para um dado número de classes a curva de ROC é construída plotando a taxa de VP em função de FP para um número de classes específico e valores progressivos de  $\rho$  - Figura 25(A).

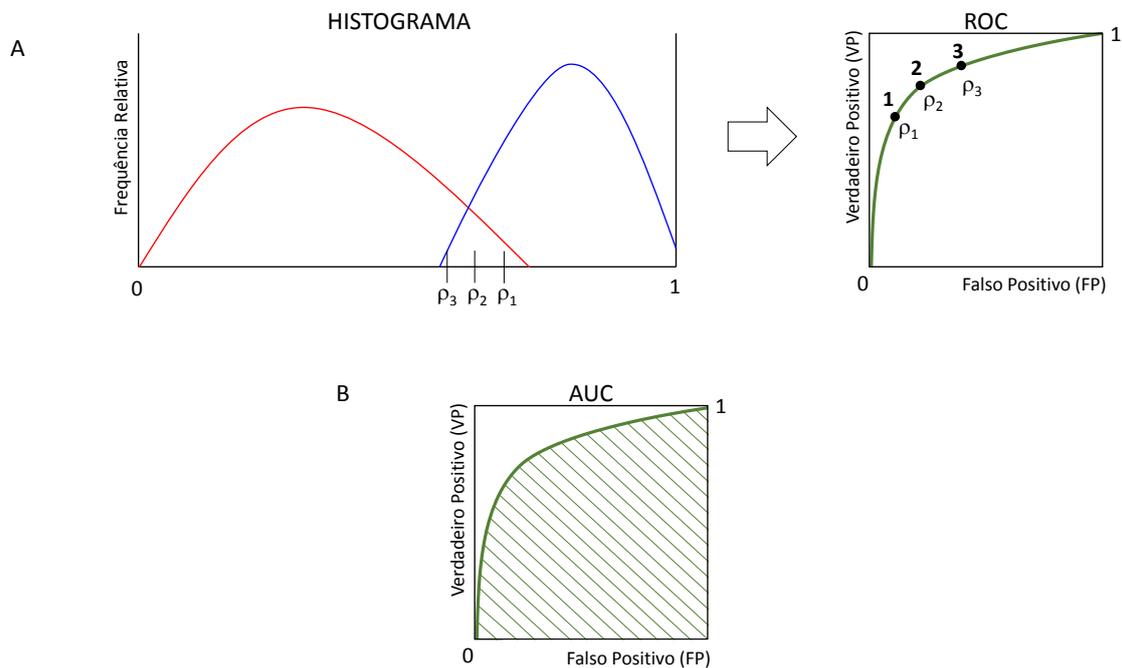


Figura 25: (A) Exemplo de curva ROC obtido a partir de um histograma com um número de classes específico e valores progressivos de  $\rho$ . 1, 2 e 3 compõem a curva ROC e representam taxa de VP em função de FP para três valores de  $\rho$  diferentes. (B) Exemplo de área sob a curva ROC.

Assim, é construída uma curva ROC para cada número de classes testado no agrupamento das amostras e comparam-se as diferentes curvas de ROC para minimizar a zona de sobreposição dos histogramas de amostras correlacionadas e não correlacionadas. Pelos motivos já expostos, deseja-se obter a curva com menor zona de sobreposição (região de incerteza), ou seja, a curva mais próxima ao canto superior esquerdo, mais próximas do ponto (0,1). Uma forma de identificar essa curva é medindo a AUC<sup>94</sup>, mostrada na Figura 25(B).

Existem várias maneiras de calcular a AUC, sendo uma das possibilidades a integração trapezoidal. A integração trapezoidal é uma aproximação da AUC mediante a

área de trapézios. O valor máximo para a AUC é de 1, indicando assim uma separação perfeita das distribuições. Um valor de AUC de 0,5 indica que o número de classes não é discriminativo.

Dessa forma, a AUC é uma medida capaz de determinar o número de classes que agrupe amostras de cocaína similares sem a utilização de conhecimentos *a priori*.

Seguindo esse procedimento, a determinação dos valores de  $\rho_L$  e  $\rho_{NL}$  é possível. Contudo, deve-se levar em conta outros critérios, os quais forneçam resposta ao maior número de casos e minimizem a chance de erros. Portanto, essa determinação deve avaliar a taxa de FP e FN aceitáveis em cada situação e a zona de incerteza do método.

Neste trabalho, os valores limites de  $\rho$  foram escolhidos considerando uma chance de se cometer 2% de FN (valor utilizado no LPS/França<sup>80</sup>) e 0,5% de FP. Assim, a região de incerteza do método, ou seja, a zona de sobreposição no histograma está compreendida entre a probabilidade de 2% da curva das amostras ligadas ( $\rho_{L(2\%)}$ ) e a probabilidade de 99,5% da curva das amostras não ligadas ( $\rho_{NL(0,5\%)}$ ), como mostra a Figura 26. Dessa forma, todas as amostras que possuem  $\rho$  menor que  $\rho_{L(2\%)}$  são consideradas não correlacionadas, amostras com  $\rho$  maior que  $\rho_{NL(0,5\%)}$  são consideradas correlacionadas e amostras com  $\rho_{L(2\%)} < \rho < \rho_{NL(0,5\%)}$  estão na região de incerteza do método em que se necessita de mais informações, como, por exemplo, outras análises laboratoriais ou investigações policiais. Quanto menor for essa região de incerteza, mais aplicável o método será em rotina.

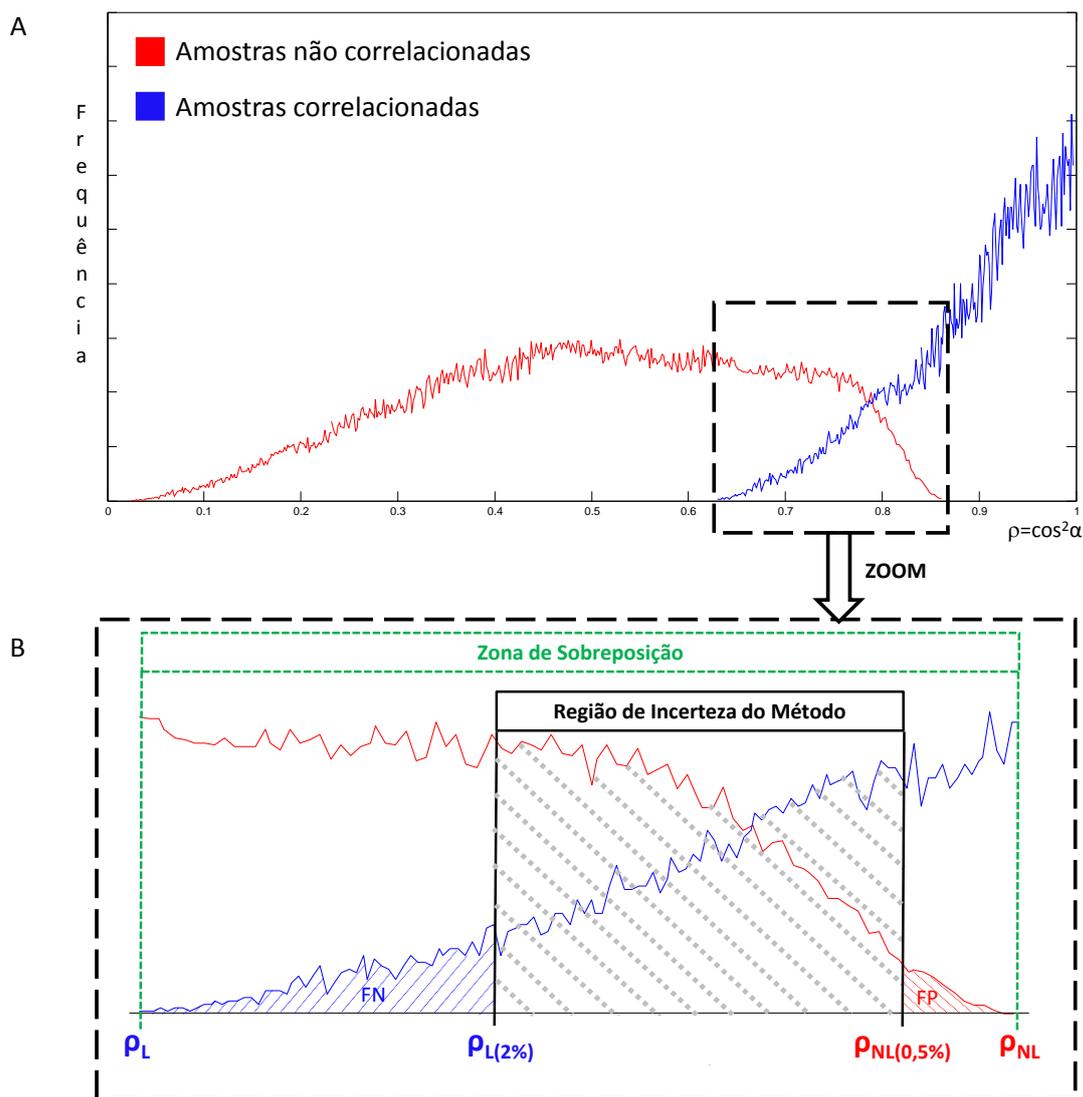


Figura 26: (A) Histograma ilustrando os valores limites para a classificação de amostras e (B) caracterização dos falsos positivos (FP), falsos negativos (FN), zona de sobreposição e região de incerteza, onde haverá dúvida sobre a classificação de amostras.

Em resumo, depois de se construir todas as curvas de ROC para diferentes números de classes, encontra-se qual a maior AUC, identificando o número de classes que melhor agrupa amostras correlacionadas. O próximo passo é verificar se o valor de  $\rho_L$  é maior que 0,90. Esse valor foi determinado empiricamente utilizando informações policiais e através de análises visuais de cromatogramas comparando diversas amostras de uma mesma apreensão e amostras de apreensões feitas no mesmo ano e em um mesmo Estado, as quais mostraram que correlações com valor de  $\rho$  menor que esse valor tem baixa probabilidade de estarem correlacionadas. Dando sequência ao método, caso  $\rho_L$  for menor que 0,90 deve-se procurar o próximo máximo valor de AUC e repetir todo o

procedimento. Se houver empate entre valores máximos de AUC, a melhor classificação é obtida avaliando a menor diferença entre  $\rho_L$  e  $\rho_{NL}$ , a qual fornecerá a menor zona de incerteza.

### 3. Materiais e Métodos

As análises necessárias para o desenvolvimento desse trabalho foram realizadas no Serviço de Perícias em Laboratório e Balística do Instituto Nacional de Criminalista do Departamento de Polícia Federal em Brasília (SEPLAB/INC/PF).

#### 3.1. Amostras e Preparação de Amostras

O conjunto de dados foi composto por 619 amostras de cocaína base livre originado de cerca de 123 apreensões realizadas pela Polícia Federal Brasileira em diferentes partes do Brasil entre os anos de 2009 e 2013, como mostra a Figura 27.

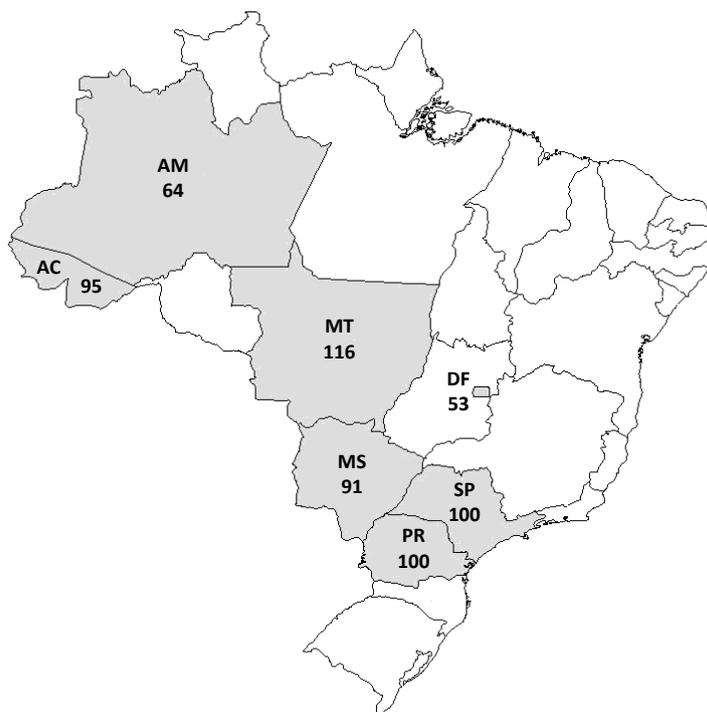


Figura 27: Distribuição das amostras de cocaína base analisadas por HS-CG-EM em relação aos estados do Brasil.

Antes da análise instrumental, as amostras foram cuidadosamente homogeneizadas por maceração manual em um cadinho de cerâmica com nitrogênio líquido suficiente para cobrir a amostra.

Ensaio de solubilidade foram realizados para verificar a forma química da cocaína (sal ou base livre), de acordo com as metodologias recomendadas pela Polícia Federal Brasileira<sup>62</sup>, SWGDRUG<sup>63</sup> e UNODC<sup>64</sup>.

### 3.2. Procedimentos Experimentais

Em frascos de *headspace* (20 mL), foram adicionados 100,00 mg de cocaína base livre, 10,00 mL de uma solução de sulfato de sódio 22% m/v e 10,00 µL de padrão interno (0,25% v/v de solução de bromobenzeno em dimetil-sulfóxido). O frasco foi selado com um septo de silicone e tampa e, posteriormente, colocado no carrossel do HS que estava ligado ao cromatógrafo gasoso. A solução de sulfato de sódio 22% m/v foi preparada utilizando água ultrapura (Millipore Milli-Q) e sulfato de sódio (Cinética).

As análises foram realizadas em um sistema com um amostrador automático de *Headspace* (Agilent Technologies G1888), onde as amostras eram aquecidas a 85 °C durante 14 minutos, com agitação contínua. O frasco era pressurizado por 12 segundos e mantido parado para equilibrar o sistema durante 3 segundos. Depois 3,00 mL de amostras aquecida era transferida ao cromatógrafo por uma linha de transferência em um circuito a 175 °C.

Para realizar as medições de cromatografia gasosa, um dispositivo Agilent 6890N equipado com uma coluna capilar HP-PONA (50m × 200 mm × 0,5 mm) foi utilizado. O hélio foi utilizado como gás carreador; a temperatura do injetor foi mantida a 180 °C, e utilizou-se uma razão de *split* de 20:1. A temperatura inicial do forno com a coluna foi de 30 °C por 10,0 minutos, seguido por um aumento linear de 20 °C/min até 150 °C, e 60 °C/min até 200 °C. O tempo total de corrida cromatográfica foi de 30,0 minutos.

A detecção foi realizada com um espectrômetro de massa de quadrupolo (Agilent 5973 inerte), com faixa de *m/z* de 29-200 amu. Os diferentes compostos foram identificados pela comparação dos espectros experimentais com os presentes na base de dados da biblioteca NIST'05 (NIST / EPA / NIH Mass Spectral Pesquisa Programa, versão 2.0d) do Instituto Nacional de Padrões e Tecnologia.

### 3.3. Precisão do Método

A precisão pode ser definida como sendo o grau de concordância entre os valores obtidos em múltiplas análises de uma mesma amostra em condições pré-definidas. Pode ser determinada no nível de repetibilidade, análise de uma amostra feita em uma mesma condição operacional em curtos intervalos de tempo (num mesmo dia por exemplo), e

pela precisão intermediária, análise de uma amostra feita em uma mesma condição operacional em longos intervalos de tempo ou por diferentes analistas.<sup>52</sup>

Duas amostras de cocaína base livre foram utilizadas para determinar a precisão do método. Tais amostras foram avaliadas seis vezes em um mesmo dia e também ao longo de 12 meses, sendo essas medidas utilizadas para estimar a repetitividade e a precisão intermediária do método, respectivamente. Essas amostras foram escolhidas aleatoriamente com base na quantidade disponível para análises. Elas possuíam perfis químicos de solventes residuais distintos, uma com mais e outra com menos picos cromatográficos.

## **4. Resultados**

### **4.1. Análises preliminares das amostras**

Sabendo que cada pico cromatográfico representa um solvente utilizado no processo de preparo da cocaína e que, em geral, os traficantes não utilizam a mesma proporção e nem sempre o mesmo solvente nesse processo, ao analisar visualmente os dois cromatogramas apresentados na Figura 28, foi possível verificar grande semelhança entre eles. Isso significa que as amostras representadas por esses cromatogramas, mesmo tendo sido apreendidas em épocas e locais distintos, provavelmente passaram por um único processo de preparação, ou seja, estão correlacionadas de alguma forma.

Um artigo publicado em 2012 utilizou a técnica de determinação de solventes residuais para comparar amostras apreendidas pela Polícia Federal e Polícia Civil e puderam confirmar correlações entre as amostras estudadas.<sup>95</sup> No entanto, a comparação binária e visual de várias amostras é uma tarefa árdua. Por isso, a proposta desse trabalho foi criar um método capaz de comparar amostras de forma objetiva e automática, estimando um limiar onde é possível afirmar com segurança que duas amostras estão correlacionadas.

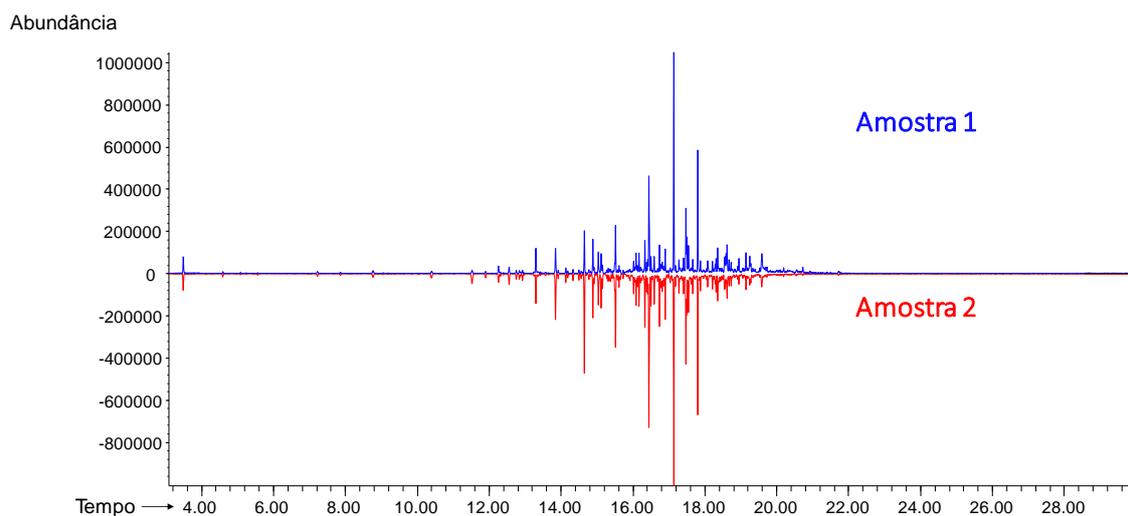


Figura 28: Amostras que apresentaram correlação e foram apreendidas em diferentes ocasiões no Acre em 2010.

Inicialmente, para uma análise exploratória dos dados, foi empregada a comparação visual, a HCA e informações periciais. Os cromatogramas das amostras analisadas passaram por uma triagem a fim de identificar alguns grupos de amostras correlacionadas. Com essa análise foi possível identificar alguns perfis químicos predominantes nas amostras de cocaína base livre apreendidas no Brasil e selecionar um conjunto controlado de amostras que serviram para desenvolver o método proposto.

Seis grupos de amostras foram identificados com o auxílio da HCA obtida utilizando a distância de Mahalanobis (Figuras 29 e 30). Cada um desses grupos exibiu perfis químicos distintos (Figura 31), resultantes dos processos de produção das amostras.

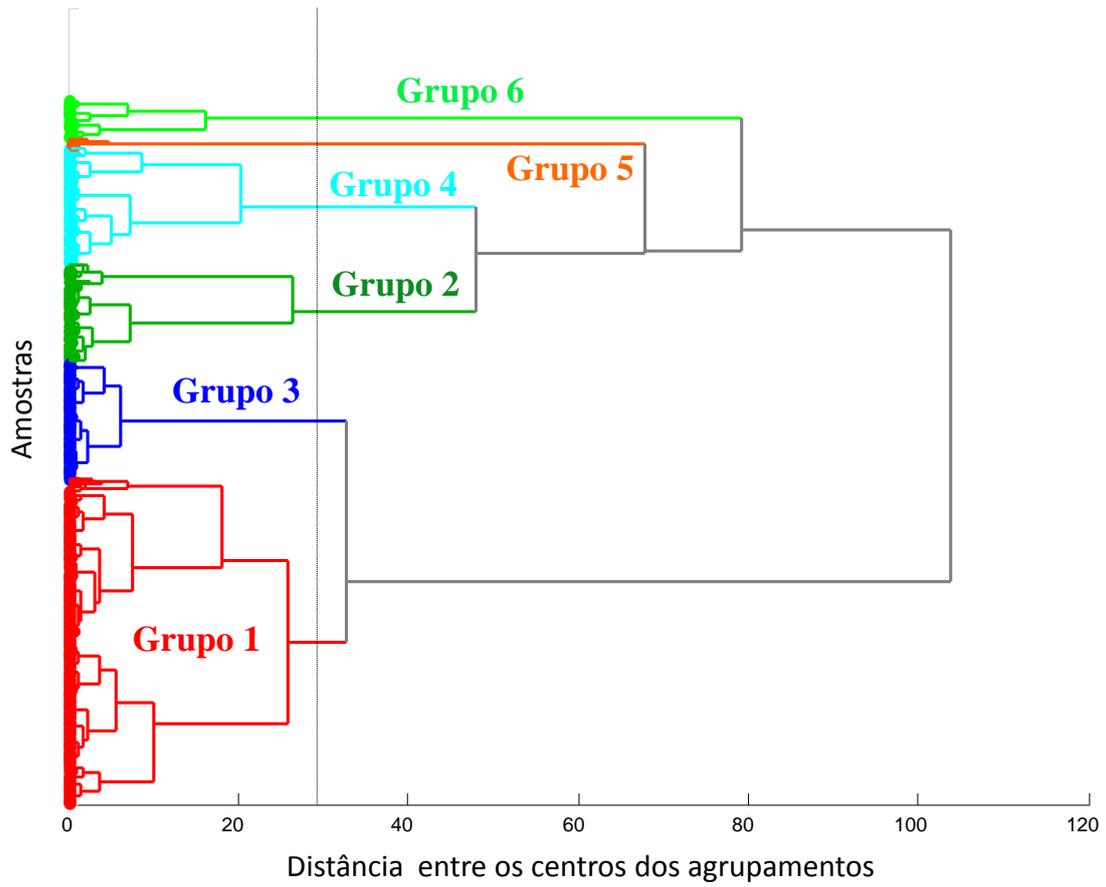


Figura 29: Dendrograma identificando os seis grupos de amostras.

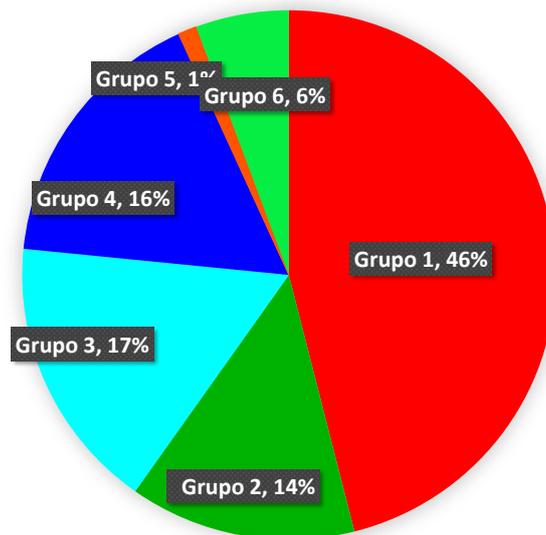


Figura 30: Porcentagem de amostras em cada grupo formado pela HCA.

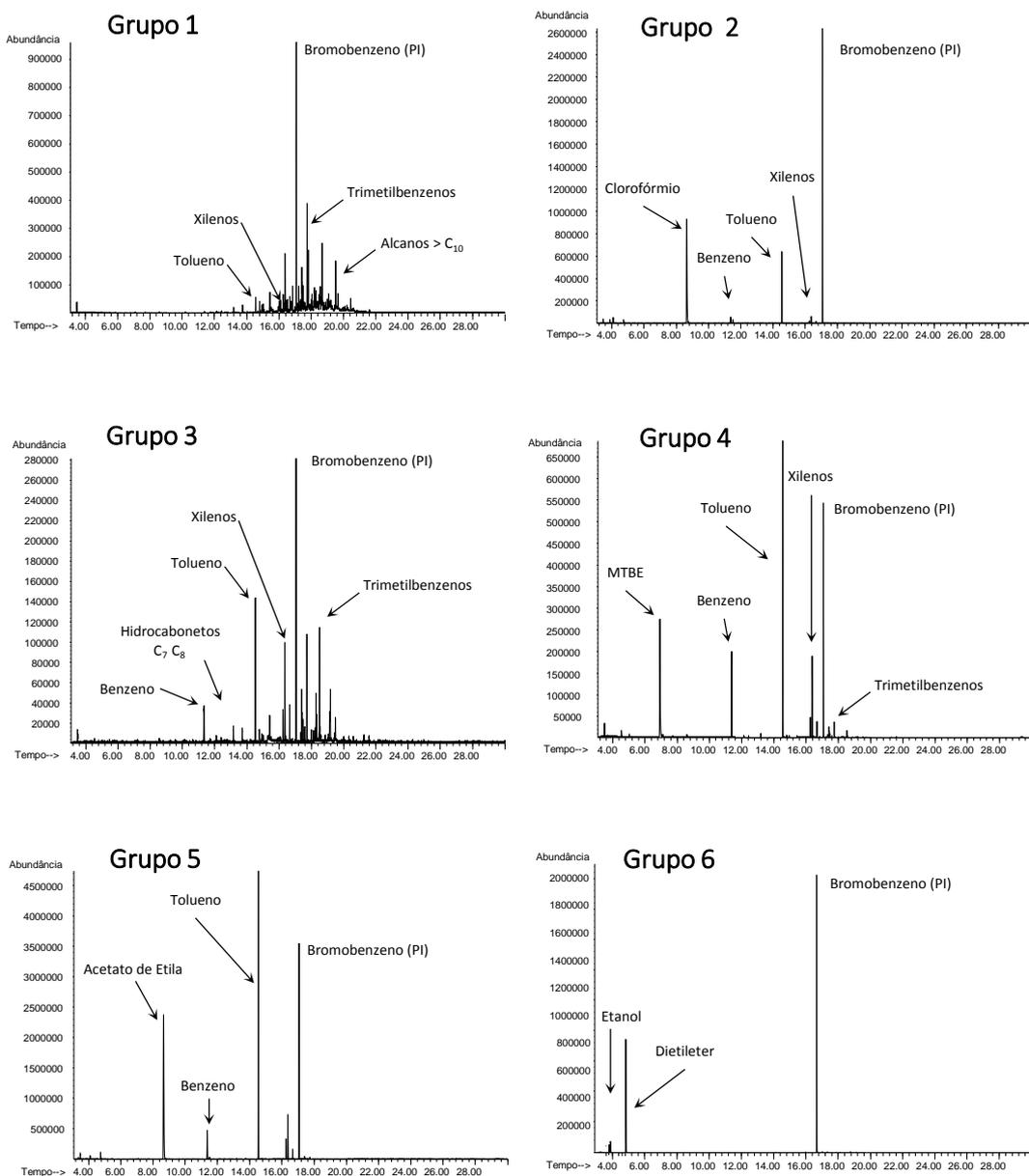


Figura 31: Principais perfis químicos encontrados nos grupos de amostras separadas por HCA.

Grande parte dos solventes identificados na análise cromatográfica foram hidrocarbonetos com perfis similares ao de combustíveis, provenientes do processo de extração da cocaína das folhas de coca, tais como gasolina, querosene, diesel e *thinner*. Devido ao fato da separação cromatográfica desses compostos ser complexa, não foi possível realizar essa separação de forma eficiente com os parâmetros utilizados no método, como mostra o cromatograma da Figura 32.

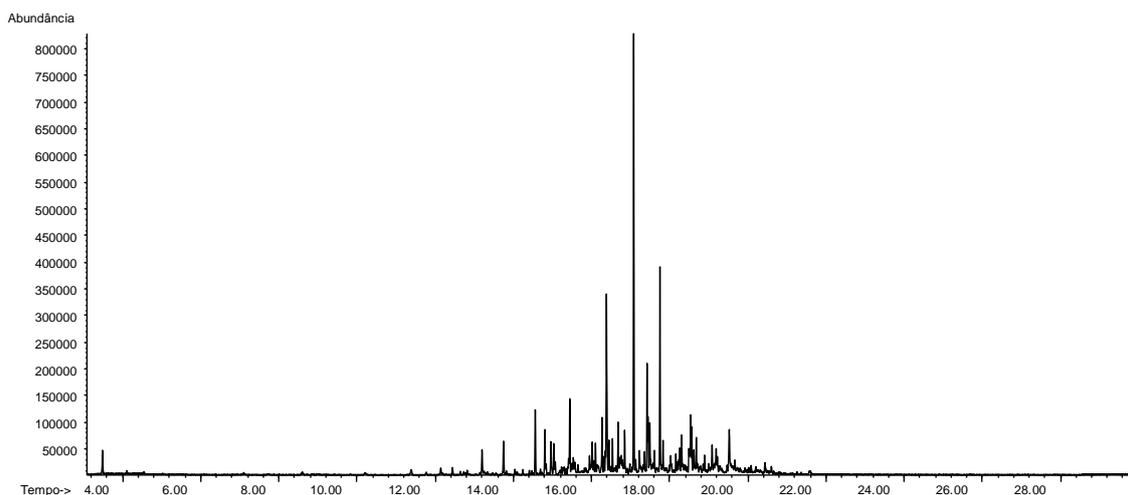


Figura 32: Cromatograma de uma amostra apreendida em Mato Grosso em 2010.

Nesse sentido, a fim de otimizar todo o procedimento experimental, reduzindo custos e tempo de análise, as comparações entre amostras de cocaína base livre foram testadas de duas maneiras: a primeira utilizou o cromatograma de íons totais (TIC, do inglês *total ion chromatogram*) e a segunda, os espectros massas acumulado de cada amostra, descrito a seguir. A comparação desses resultados apontou a melhor métrica para as análises quimiométricas.

#### 4.2. Análises dos dados

No TIC, os cromatogramas foram interpretados diretamente na forma de perfis, o que diferiu do método tradicional que considera as áreas dos picos detectados. A vantagem foi tornar desnecessário o estabelecimento de condições adequadas para uma separação cromatográfica eficiente. Dessa forma, a corrida cromatográfica de 30 minutos foi dividida em 12.166 pontos, nos quais a intensidade do sinal cromatográfico foi medida. Com esses dados, uma matriz foi organizada, onde cada linha correspondia a uma amostra e cada coluna apontava a intensidade do sinal cromatográfico em pontos fixos de tempo de retenção.

Um aspecto importante considerado ao analisar diretamente cromatogramas como perfis foi a variação nos tempos de retenção dos picos, que ocorre com frequência em decorrência de variações na temperatura, idade da coluna, pequenas variações na composição da fase móvel e diferentes maneiras de injeção da amostra. Se o arranjo de dados não estiver disposto de forma que as variáveis observadas expressem os mesmos atributos (cada coluna da matriz deve corresponder ao mesmo tempo em todas as

amostras), o pré-requisito para se utilizar modelos quimiométricos é violado. Dessa forma, aplicou-se um pré-tratamento a fim de corrigir esses deslocamentos: o alinhamento dos picos dos cromatogramas pela técnica de correlação de deformação otimizada (COW, do inglês *correlation optimized warping*), o qual buscou minimizar o número de bandas distintas entre os diversos cromatogramas, utilizando o pico de padrão interno como referência para o alinhamento dos demais cromatogramas em relação à essa referência. Porém, apesar desse mecanismo ser útil, a distorção proporcionada por ele pode forçar que alguns picos relativos a substâncias diferentes fiquem alinhados podendo causar falsos positivos. Devido a esse motivo e buscando melhorar a forma de analisar os dados, comparações entre amostras de cocaína base livre foram feitas também com base nos espectros de massas acumulados, além do TIC com e sem alinhamento.

Os espectros de massas foram registrados com ionização por impacto eletrônico (70 eV) e com aquisição de dados no modo de varredura linear com faixa de  $m/z$  de 29-200 amu. Dessa forma, em espaços fixos de tempo, à medida que ocorria a separação cromatográfica, diversos espectros de massas foram sendo gerados. Assim, cada amostra foi representada por uma matriz onde cada linha representava um tempo de retenção (TR) específico e cada coluna os fragmentos de  $m/z$  obtidos naquele TR. Para ter o chamado espectro de massas acumulado, somou-se todos os elementos de uma linha, obtendo uma nova matriz representativa da amostra. Reuniram-se os dados de todas as amostras em uma nova matriz foi criada, onde cada coluna correspondia a uma amostra e as linhas fragmentos dos espectros de massas. É importante esclarecer que o uso do espectro de massas acumulado é, a princípio, equivalente a utilizar a injeção direta da amostra no espectrômetro de massas, o que pode gerar redução no tempo de análise, uso de solventes e, conseqüentemente, custos de todo o procedimento.

Para se decidir entre uma dessas abordagens para a análise dos dados, foram utilizados os dados de três amostras medidas em replicatas. Duas das três amostras eram de cocaína base citadas na seção 3.3, que avaliaram a precisão do método, e uma amostra de cocaína sal utilizada na validação do método ULT desenvolvido anteriormente. Os resultados obtidos para cada uma dessas abordagens foram avaliados através da média e desvio padrão das correlações obtidas entre três amostras selecionadas para testar o método, sendo duas de cocaína base livre e uma cocaína sal. Na Figura 33 foram apresentados os cromatogramas dessas amostras. A amostra de cocaína sal foi utilizada por já haver sido avaliada em um trabalho anterior que descreveu o método ULT. Nesse

caso, considerou-se a área dos picos cromatográficos de 33 solventes detectados em amostras de cocaína sal.

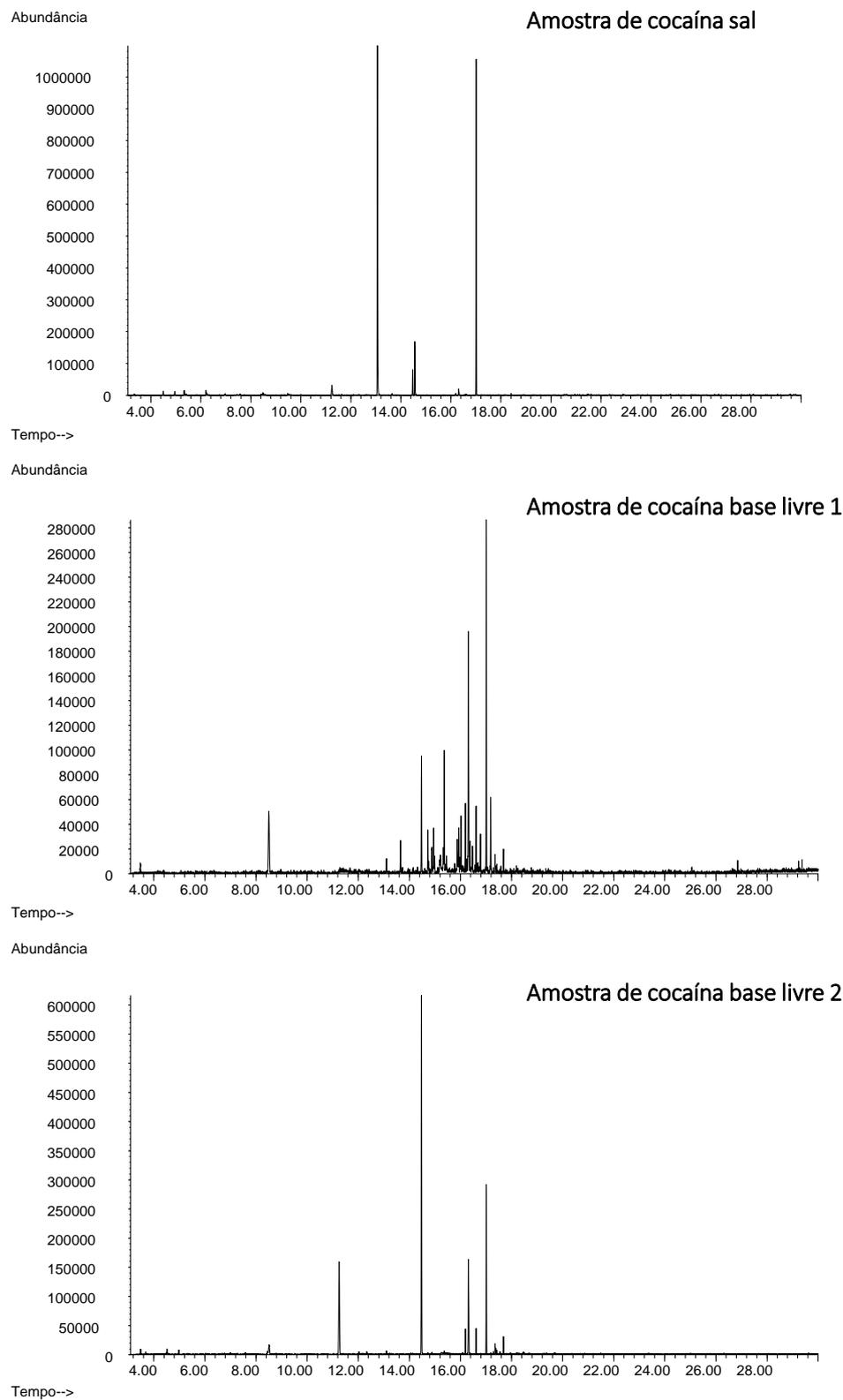


Figura 33: Cromatogramas de amostras selecionadas para testar a análise de dados.

Pela Figura 33, foi possível verificar que cromatogramas de cocaína base apresentaram-se mais complexos que os de cocaína sal, o que gera a necessidade de se avaliar outra maneira de se obter os dados, que não necessite da separação completa dos picos e sua integração individual.

A amostra de cocaína sal foi analisada 23 vezes em dias diferentes pelo mesmo analista, testou-se o cálculo de  $\rho$  entre amostras utilizando integração de picos cromatográficos, TIC, alinhamento do TIC, espectro de massas acumulado e seleção de faixa de massa/carga no espectro de massas acumulado.

Os resultados apresentados na Figura 34 mostraram que, para essa amostra de cocaína sal, depois da integração de picos cromatográficos, o uso do espectro de massas acumulado foi o que gerou os melhores resultados, com maior média entre as correlações, menor desvio padrão e menor coeficiente de variação.

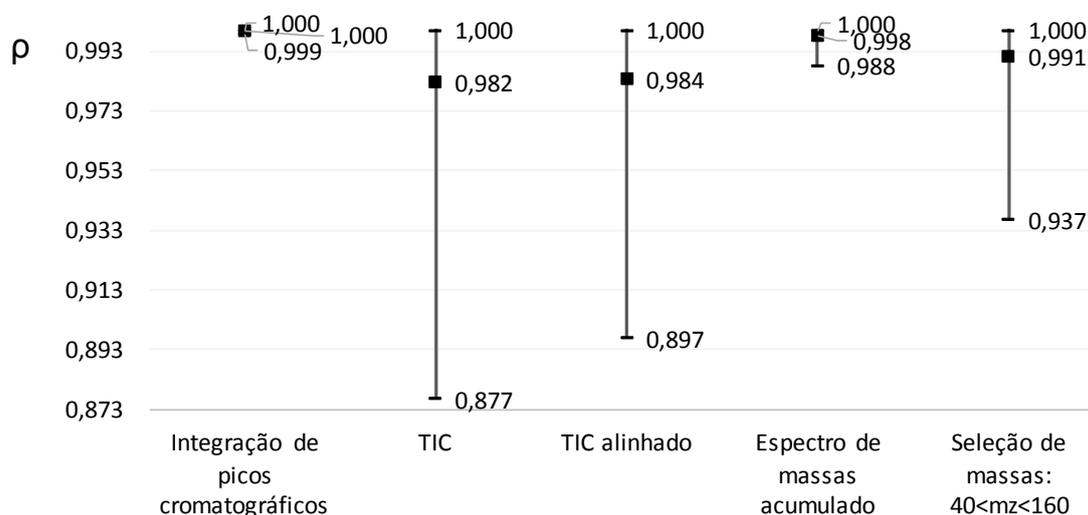


Figura 34: Valores máximos, mínimos e de média para correlações ( $\rho$ ) entre a amostra de cocaína sal.

Após essa análise, duas amostras de cocaína base livre também foram analisadas com 12 replicatas realizadas em dias distintos. Nesse caso, testou-se o TIC e espectro de massas acumulado. Os resultados indicaram que as médias das correlações das replicatas dessas amostras utilizando o espectro de massas acumulado apresentou uma variação significativamente menor que as médias dessas mesmas replicatas utilizando o TIC (Figura 35). Dessa forma, verificou-se que o uso do espectro de massas acumulado foi mais adequado para o cálculo de correlações entre amostras de cocaína base livre e por

isso, essa abordagem foi utilizada para o desenvolvimento do método proposto para correlação de amostras de cocaína base.

Esse resultado é importante por evidenciar a possibilidade de realizar a análise através da introdução direta da amostra na fonte de íons do espectrômetro sem prévia separação cromatográfica. A inserção direta tem como vantagem a eliminação da etapa de separação cromatográfica, o que facilitaria e reduziria o tempo de análise e, ainda, os custos do procedimento.

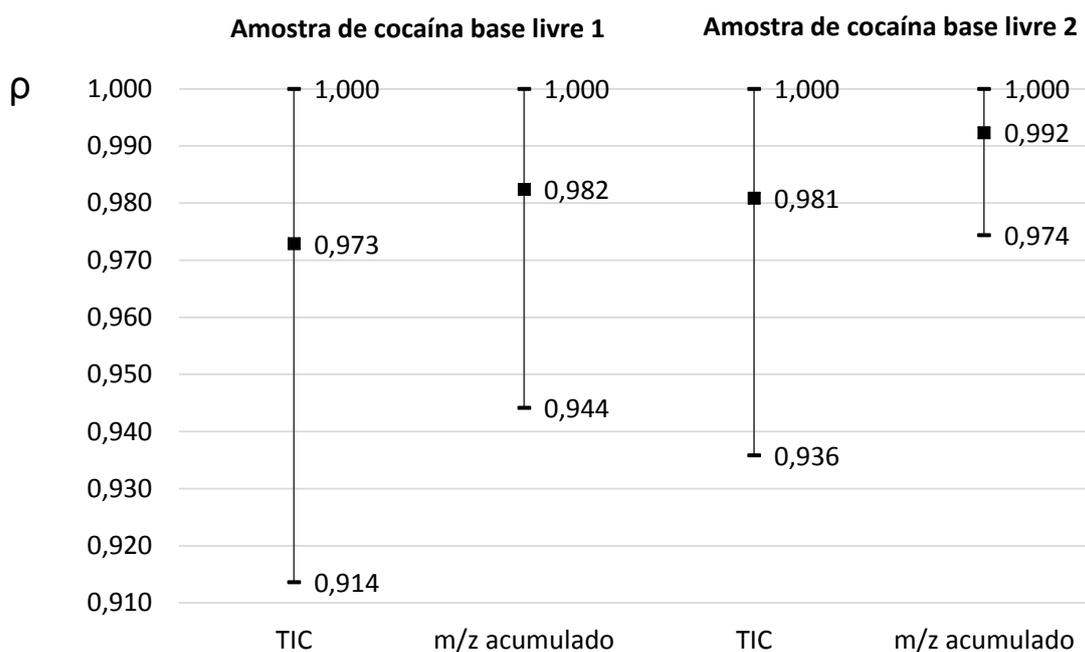


Figura 35: Valores máximos, mínimos e de média para correlações ( $\rho$ ) entre duas amostras de cocaína base livre.

### 4.3. Conjunto controlado de amostras

Ainda a partir dos seis grupos separados pela HCA na seção 4.1 e usando dados cromatográficos, foi possível identificar 28 apreensões que visualmente possuíam perfis químicos semelhantes e foram apreendidas em um mesmo ano e Unidade da Federação. Essas 28 apreensões eram compostas de 193 amostras, sendo utilizadas para compor um conjunto controlado de amostras, o qual foi utilizado para testar o método ULT na determinação de valores  $\rho$  limites.

Para esses dados, foram avaliados o número de amostras em cada uma das apreensões e as correlações entre as amostras de uma mesma apreensão, bem como, a média e os valores mínimos de  $\rho$  dentro dos conjuntos de apreensões e esses resultados

foram apresentados na Tabela 13. Observaram-se variações compatíveis com os testes de precisão apresentados na Figura 36. Outra observação pertinente encontrada foi a que mostrou que os valores de  $\rho$  dentro de uma mesma apreensão foram superiores a 0,90, resultado importante por concordar com um dos critérios do método ULT, o qual exige que  $\rho_L$  seja maior que 0,90 para que o número de classes indicado pela AUC seja aceito.

Tabela 13. Valores médios e mínimos de  $\rho$  entre amostras pertencentes de uma mesma apreensão para o conjunto controlado utilizando o espectro de massas acumulado.

<b>Apreensão</b>	<b>N° de Amostras</b>	<b>Média de <math>\rho</math></b>	<b>Valor Mínimo de <math>\rho</math></b>	<b>N° de Correlações da Classe</b>
1	10	0,992	0,973	55
2	6	0,995	0,982	21
3	6	0,988	0,960	21
4	5	0,983	0,967	15
5	6	0,971	0,910	21
6	8	0,978	0,937	35
7	5	0,982	0,967	15
8	7	0,979	0,948	28
9	7	0,982	0,954	28
10	6	0,992	0,983	21
11	9	0,994	0,980	45
12	6	0,982	0,956	21
13	10	0,984	0,919	55
14	9	0,995	0,986	45
15	4	0,989	0,976	10
16	10	0,974	0,911	55
17	8	0,975	0,928	35
18	6	0,959	0,934	21
19	6	0,987	0,968	21
20	7	0,991	0,974	28
21	7	0,985	0,955	28
22	5	0,976	0,940	15
23	5	0,969	0,931	15
24	8	0,982	0,944	35
25	6	0,978	0,946	21
26	8	0,982	0,952	35
27	8	0,973	0,921	35
28	5	0,968	0,917	15

Ao dar início a determinação dos valores de  $\rho$  pelo método ULT, inicialmente deve-se buscar o maior valor de AUC, verificando se  $\rho_L$  é maior que 0,90 (critério que é satisfeito a partir de 28 classes). No entanto, percebeu-se que, para esse conjunto de dados, os valores de AUC com  $\rho_L$  maior que 0,90 apresentavam variações apenas na

quinta casa decimal (Tabela 14), dificultando a seleção do máximo para AUC. Para lidar com essa situação, adotou-se mais um critério ao método, que consistiu no cálculo da diferença percentual entre ente dois valores de AUC consecutivos, calculado pela Equação 20:

$$Diferença\ percentual = \frac{(AUC_2 - AUC_1)}{AUC_1} \times 100 \quad (20)$$

onde  $AUC_1$  representa um valor anterior ou inicial de AUC e  $AUC_2$  representa o valor de AUC posterior ou final. Uma vez atingido um valor de  $\rho_L$  maior que 0,90 o valor de classes é determinado quando se identifica o primeiro ponto máximo da diferença percentual.

Dessa forma, para o conjunto controlado, os valores da porcentagem da diferença foram apresentados na Tabela 15 e no gráfico da Figura 36 e possibilitaram identificar 28 classes como o número de classes escolhido para o conjunto controlado. Com esse número de classes, os valores de  $\rho_L$  e  $\rho_{NL}$  identificados utilizando o conjunto controlado foram respectivamente 0,900 e 0,986. Pelo histograma construído para esse conjunto com 28 classes (Figura 37) foi possível observar que a região de incerteza do método englobou menos de 9% do total de correlações desse conjunto, sendo que, de acordo com informações policiais, 1,47% dessas correlações de fato pertenciam a amostras correlacionadas.

Tabela 14. Valores de AUC para os diversos números de classes estudados no Conjunto Controlado.

<b>Nº de Classes</b>	<b>AUC</b>	<b>Porcentagem da Diferença</b>	<b><math>\rho_L</math></b>	<b><math>\rho_{NL}</math></b>
4	0,92855	2,22	0,181	0,904
5	0,92966	0,12	0,181	0,904
6	0,95318	2,53	0,327	0,935
7	0,90004	-5,58	0,467	0,968
8	0,94384	4,87	0,605	0,968
9	0,94606	0,24	0,605	0,968
10	0,94995	0,41	0,625	0,968
11	0,95096	0,11	0,625	0,968
12	0,95076	-0,02	0,625	0,968
13	0,95576	0,53	0,653	0,968
14	0,95778	0,21	0,686	0,979
15	0,96961	1,24	0,733	0,979
16	0,97475	0,53	0,752	0,979
17	0,98140	0,68	0,790	0,979
18	0,98417	0,28	0,790	0,979
19	0,98686	0,27	0,818	0,979
20	0,98908	0,22	0,819	0,979
21	0,98962	0,05	0,819	0,979
22	0,98985	0,02	0,819	0,979
23	0,99135	0,15	0,879	0,979
24	0,99337	0,20	0,883	0,979
25	0,99346	0,01	0,883	0,986
26	0,99320	-0,03	0,883	0,986
27	0,99323	0,00	0,883	0,986
28	0,99663	0,34	0,900	0,986
29	0,99666	0,00	0,900	0,986
30	0,99661	-0,01	0,900	0,986
31	0,99674	0,01	0,900	0,986
32	0,99709	0,04	0,900	0,994
33	0,99708	0,00	0,900	0,994
34	0,99710	0,00	0,900	0,994
35	0,99710	0,00	0,900	0,994
36	0,99710	0,00	0,900	0,994
37	0,99863	0,15	0,911	0,994
38	0,99860	0,00	0,911	0,994
39	0,99868	0,01	0,911	0,994
40	0,99890	0,02	0,944	0,994
41	0,99889	0,00	0,944	0,994
42	0,99886	0,00	0,944	0,994
43	0,99886	0,00	0,944	0,994
44	0,99883	0,00	0,948	0,994
45	0,99892	0,01	0,954	0,994

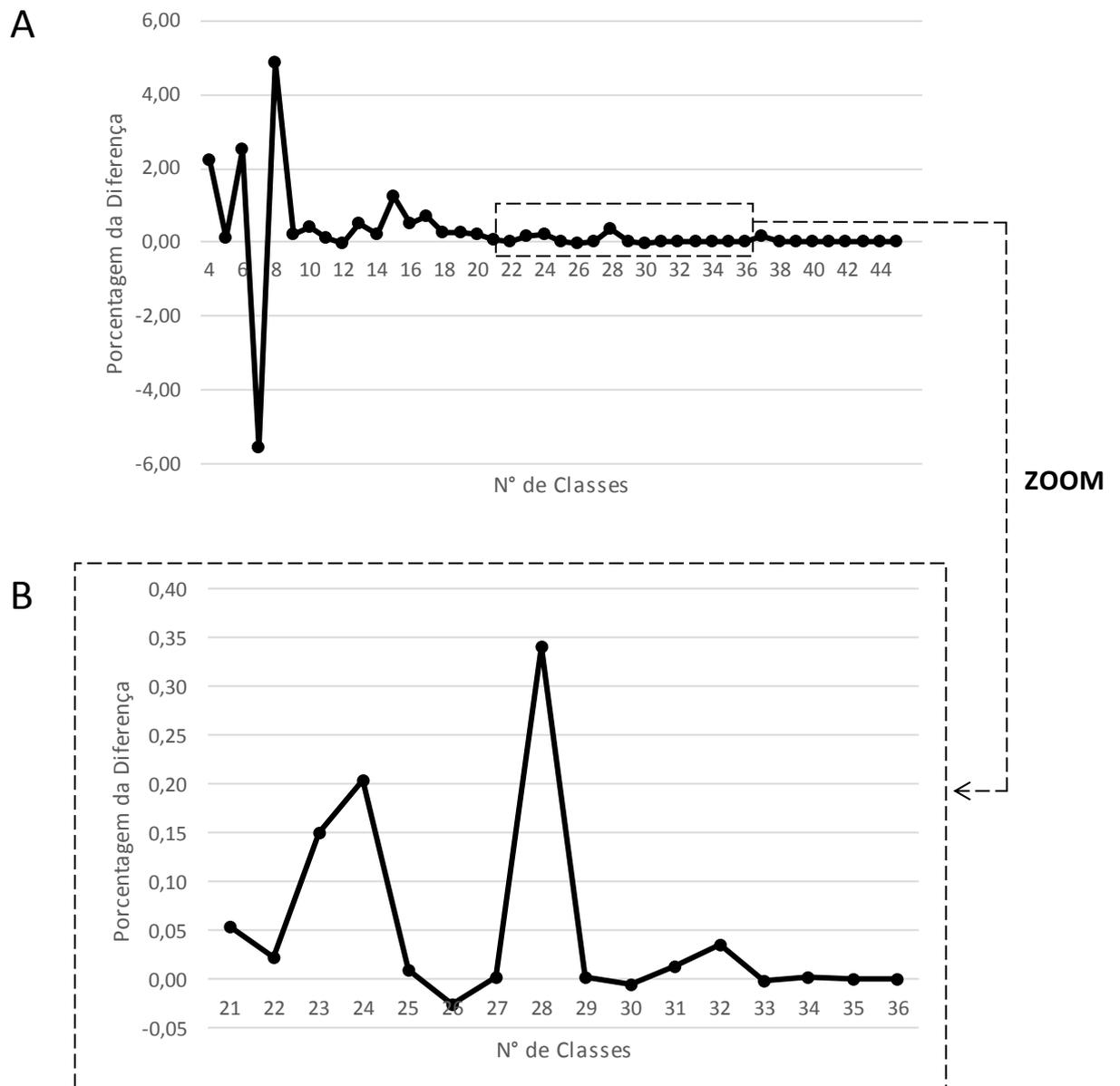


Figura 36: (A) Valores de Porcentagem da diferença para o conjunto controlado de amostras mostrando em (B) a variação entre 24 a 40 classes.

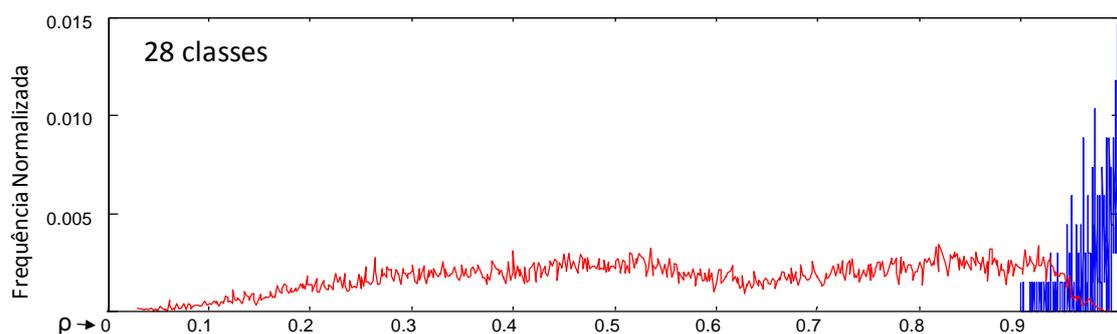


Figura 37: Histograma do conjunto controlado considerando o número de classes igual a 28. Onde a curva vermelha é de correlações entre amostras não correlacionadas e a curva azul é de correlações entre amostras correlacionadas.

Como esperado, notou-se que esse número de classes foi o mesmo determinado através de informações policias e de análises visuais dos cromatogramas. Dessa forma, concluiu-se que a alteração feita para determinar a variação mais relevante no valor de AUC para determinação do número de classes pelo método ULT funcionou para amostras de cocaína base livre.

#### 4.4. Determinação dos valores limites de $\rho$

A análise do conjunto controlado de amostras serviu para embasar a determinação dos valores limites de  $\rho_L$  e  $\rho_{NL}$ , os quais discriminam amostras correlacionadas das não correlacionadas utilizando o conjunto total de amostras analisadas. Essa determinação é feita com o conjunto completo de amostras reais (619 amostras de cocaína base livre).

Seguindo o procedimento adotado para a identificação do número de classes para as amostras reais proposto pelo método ULT e testado anteriormente com o conjunto controlado, construiu-se um dendrograma com HCA utilizando o método da média ponderada (Figura 38), a partir do qual identificou-se as classes de amostras onde foram analisados os valores de  $\rho$  para a construção de curvas ROC. Os resultados obtidos e apresentados na Tabela 15 e na Figura 39 mostraram que o primeiro ponto de máximo local dos valores de porcentagem da diferença com  $\rho_L$  superior a 0,90 foi com 145 classes. De acordo com o método, esse seria o número de classes que melhor agrupa as amostras consideradas correlacionadas. Dessa forma, com esse número de classes, pôde-se determinar os valores de  $\rho_L$  e  $\rho_{NL}$ , que foram respectivamente 0,903 e 0,997.

É importante destacar que os valores de  $\rho_L$  e  $\rho_{NL}$  mantiveram-se constantes mesmo com a variação do número de classes entre 145 a 155, que engloba os cinco máximos

locais observados na diferença de percentual com  $\rho_L$  superior a 0,90 entre esses valores de classe. Isso mostrou que tais limites parecem ser robustos a uma pequena variação do número de classes para as amostras. Portanto, caso o número correto de classes esteja entre 145 e 155 os mesmos limites seriam determinados, o que minimiza as implicações do método indicar um número de classes que seja diferente do correto para o conjunto de amostras em estudo. A manutenção desses limites ocorreu porque mesmo com um aumento do número de classes nem sempre as amostras com menor cosseno quadrado, que a princípio estão agrupadas, são transferidas para a nova classe formada. Pois a formação dos grupos é feita através do dendrograma que calcula a distância entre as amostras utilizando um critério matemático diferente do cosseno quadrado. Esse critério foi especificado anteriormente e foi o WPGMA, o qual calcula a média aritmética ponderada das distâncias entre todos os pares de elementos que podem ser formados com os elementos dos dois conglomerados que estão sendo comparados.

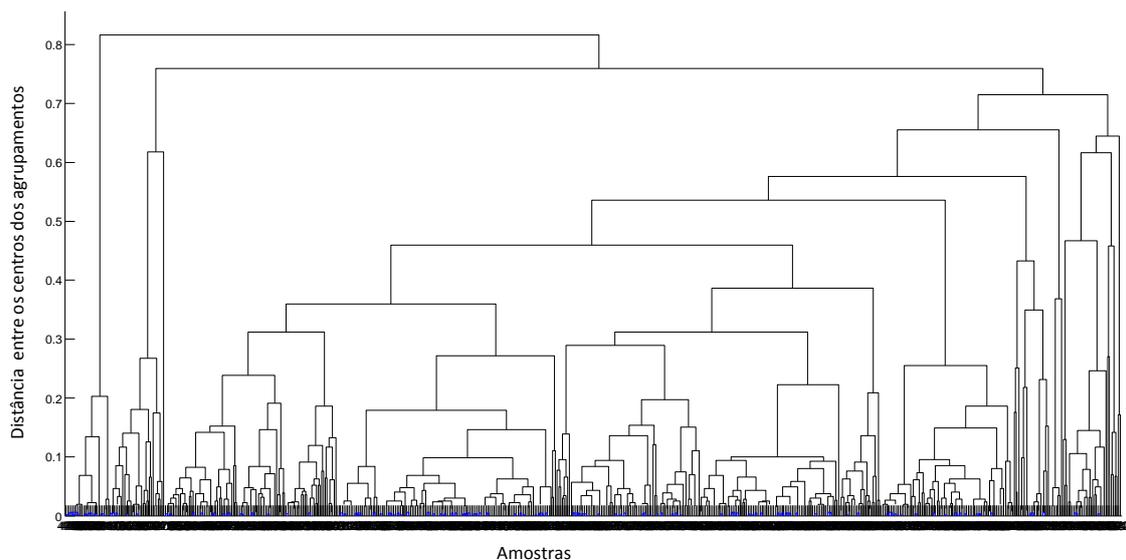


Figura 38: Dendrograma obtido pela análise dos espectros de massa acumulados das 619 amostras de cocaína base livre apreendidas pela Polícia Federal.

Tabela 15. Valores de AUC, percentagem da diferença,  $\rho_L$  e  $\rho_{NL}$  para os diversos números de classes estudados no conjunto total de amostras analisadas de cocaína base livre.

N° de Classe	AUC	% da Diferença	$\rho_L$	$\rho_{NL}$	N° de Classe	AUC	% da Diferença	$\rho_L$	$\rho_{NL}$
85	0,9810	0,199	0,812	0,993	129	0,9931	0,013	0,883	0,997
86	0,9829	0,003	0,823	0,993	130	0,9932	0,001	0,883	0,997
87	0,9829	0,004	0,823	0,993	131	0,9933	0,000	0,883	0,997
88	0,9830	0,091	0,823	0,993	132	0,9933	0,003	0,883	0,997
89	0,9839	0,018	0,823	0,993	133	0,9933	0,000	0,883	0,997
90	0,9841	0,018	0,823	0,993	134	0,9933	0,005	0,883	0,997
91	0,9842	0,001	0,823	0,993	135	0,9933	0,004	0,883	0,997
92	0,9842	0,000	0,823	0,993	136	0,9934	0,001	0,883	0,997
93	0,9842	0,011	0,823	0,993	137	0,9934	0,005	0,883	0,997
94	0,9843	0,001	0,823	0,994	138	0,9934	0,003	0,883	0,997
95	0,9844	0,018	0,823	0,994	139	0,9935	0,008	0,883	0,997
96	0,9845	0,007	0,823	0,994	140	0,9936	0,005	0,883	0,997
97	0,9846	0,004	0,823	0,994	141	0,9936	0,000	0,883	0,997
98	0,9847	0,002	0,823	0,994	142	0,9936	0,008	0,883	0,997
99	0,9847	0,001	0,823	0,994	143	0,9937	0,014	0,883	0,997
100	0,9847	0,000	0,823	0,994	144	0,9938	0,012	0,883	0,997
101	0,9847	0,040	0,823	0,994	145	0,9939	0,037	0,903	0,997
102	0,9851	0,135	0,823	0,994	146	0,9943	0,008	0,903	0,997
103	0,9864	0,041	0,831	0,995	147	0,9944	0,010	0,903	0,997
104	0,9868	0,001	0,831	0,995	148	0,9945	0,019	0,903	0,997
105	0,9868	0,000	0,831	0,995	149	0,9947	0,002	0,903	0,997
106	0,9868	0,000	0,831	0,995	150	0,9947	0,017	0,903	0,997
107	0,9868	0,002	0,831	0,995	151	0,9949	0,001	0,903	0,997
108	0,9868	0,118	0,831	0,995	152	0,9949	0,006	0,903	0,997
109	0,9880	0,000	0,836	0,995	153	0,9949	0,001	0,903	0,997
110	0,9880	0,038	0,836	0,995	154	0,9949	0,007	0,903	0,997
111	0,9884	0,006	0,836	0,995	155	0,9950	0,031	0,903	0,997
112	0,9884	0,011	0,836	0,995	156	0,9953	0,014	0,918	0,997
113	0,9886	0,037	0,836	0,995	157	0,9955	0,018	0,918	0,997
114	0,9889	0,174	0,836	0,995	158	0,9956	0,003	0,918	0,997
115	0,9906	0,004	0,875	0,995	159	0,9957	0,002	0,918	0,997
116	0,9907	0,023	0,875	0,995	160	0,9957	0,002	0,918	0,997
117	0,9909	0,002	0,881	0,995	161	0,9957	0,001	0,918	0,997
118	0,9909	0,005	0,881	0,995	162	0,9957	0,001	0,918	0,997
119	0,9910	0,002	0,881	0,995	163	0,9957	0,003	0,918	0,997
120	0,9910	0,005	0,881	0,995	164	0,9958	0,000	0,918	0,997
121	0,9910	0,002	0,881	0,995	165	0,9958	0,009	0,918	0,997
122	0,9911	0,043	0,881	0,995	166	0,9959	0,000	0,918	0,997
123	0,9915	0,007	0,883	0,995	167	0,9959	0,000	0,918	0,997
124	0,9915	-0,021	0,883	0,995	168	0,9959	0,011	0,918	0,997
125	0,9913	0,124	0,883	0,997	169	0,9960	0,001	0,921	0,997
126	0,9926	0,003	0,883	0,997	170	0,9960	0,001	0,921	0,997
127	0,9926	0,044	0,883	0,997	171	0,9960	0,007	0,921	0,997
128	0,9930	0,007	0,883	0,997	172	0,9961	0,000	0,922	0,997

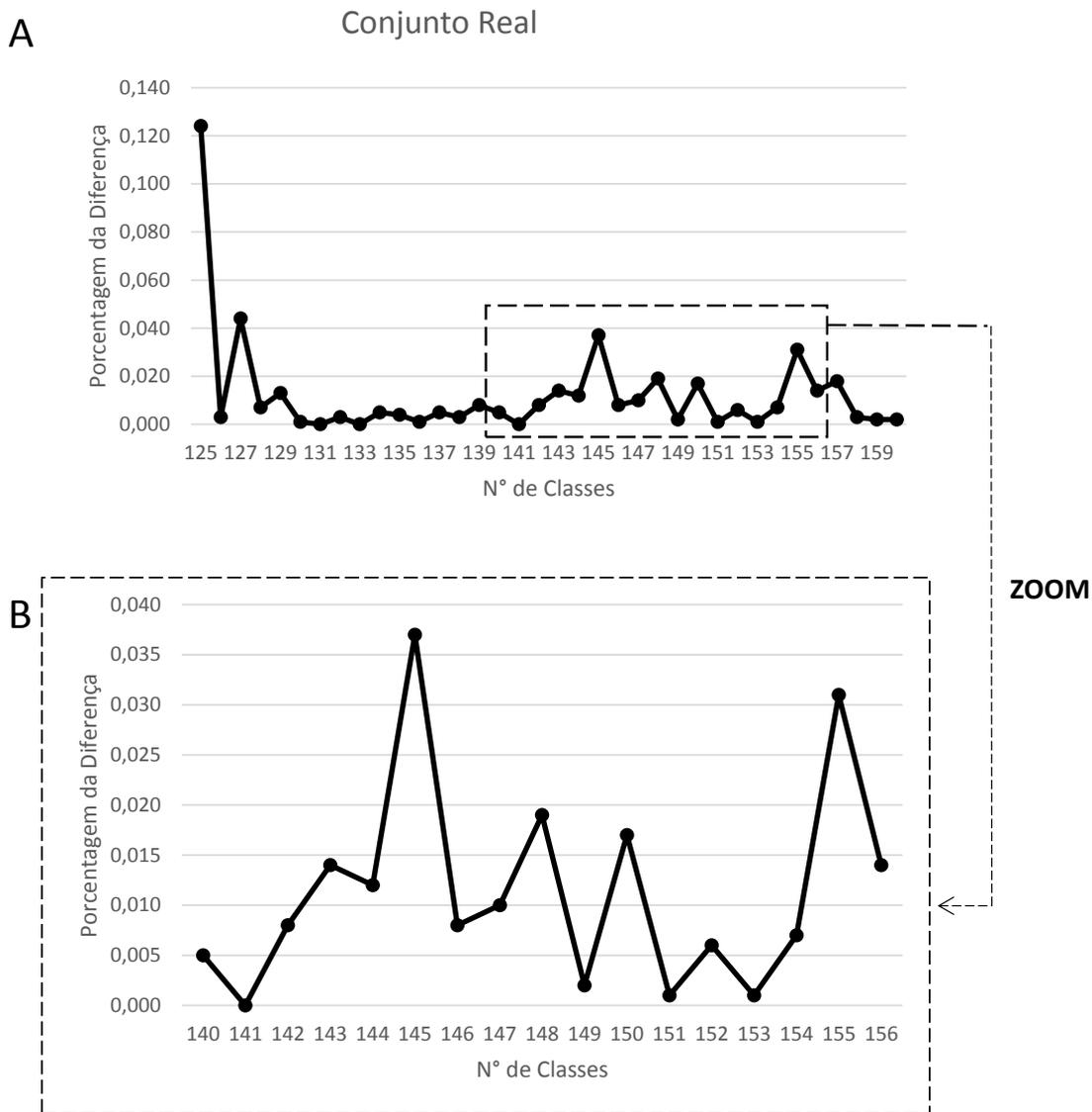


Figura 39: (A) Valores de Porcentagem da diferença para o conjunto de amostras reais mostrando em (B) a variação entre 140 a 151 classes.

Outro ponto relevante notado foi que  $\rho_{NL}$  está bem próximo a 1. Como consequência disso, os histogramas das amostras de cocaína base livre apresentam uma zona de grande sobreposição, como pôde ser apresentado na Figura 40. Essa grande zona de sobreposição faz com que a maioria das correlações do histograma das amostras ligadas acabem se enquadrando na região de incerteza do método.

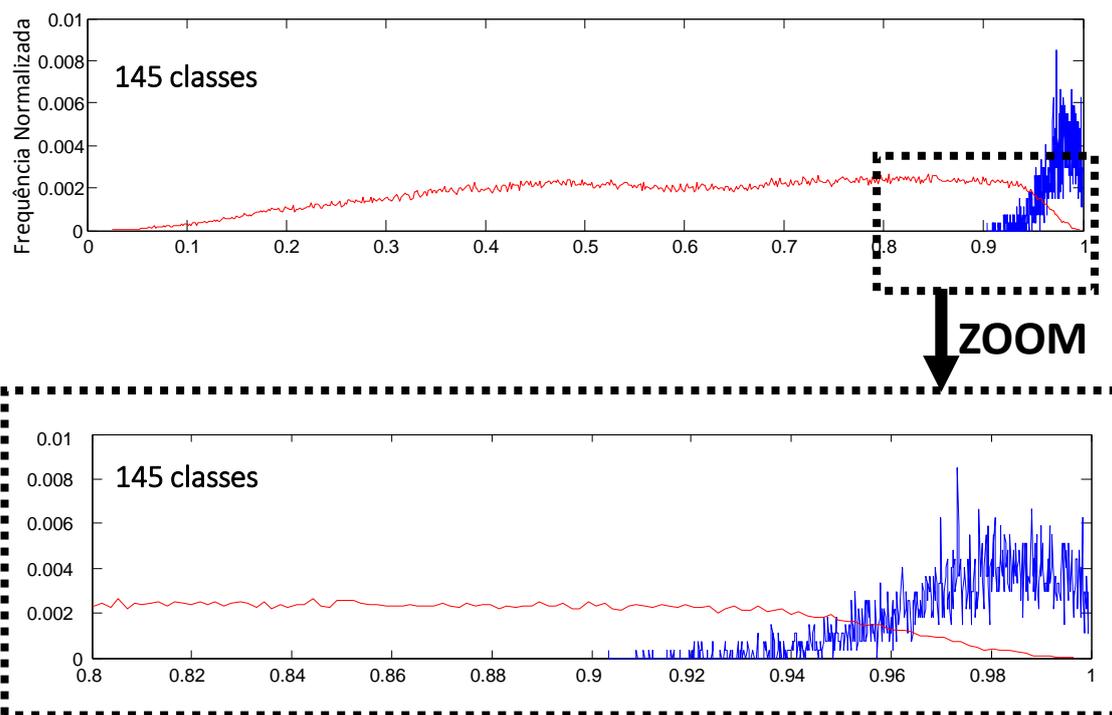


Figura 40: Histogramas para 145 classes.

Uma maneira de minimizar esse problema foi aceitar uma taxa de FP para o método. A Tabela 16 mostrou que a aceitação de uma taxa de FP de apenas 0,5% já reduz significativamente o número de correlações pertencentes ao histograma das amostras ligadas na região de incerteza. A aceitação dessa taxa de FP reduz de 95,2% para 31,7% das correlações de amostras correlacionadas na região de incerteza.

Tabela 16. Variação de falsos positivos (FP), mantendo  $\rho_L=0,903$ , ou seja, nenhum falso negativo.

FP (%)	$\rho_{NL}$	% de amostras ligadas na região de incerteza	% de correlações na região de incerteza
0	0,997	95,2	9,64
0,25	0,977	46,4	8,69
0,5	0,970	31,7	8,18
1	0,962	18,5	7,48
1,5	0,956	12,2	6,89
2	0,950	11,1	6,25

Da mesma forma, a avaliação da percentagem de correlações do histograma das amostras não ligadas em relação à taxa de FN foi feita e observou-se a diminuição do número de correlações da região de incerteza. Contudo, como pode ser visto na Tabela 17, essa diminuição se deu de forma menor que a observada na Tabela 16.

Tabela 17. Variação de falsos negativos (FN), mantendo  $\rho_{NL}=0,997$ , ou seja, nenhum falso positivo.

FN (%)	$\rho_{NL}$	% de amostras não ligadas na região de incerteza	% de correlações na região de incerteza
0	0,903	8,39	9,64
0,5	0,921	5,78	7,13
1	0,927	4,98	6,33
2	0,935	3,88	5,23
3	0,940	3,26	4,59
5	0,945	2,68	3,98

Um trabalho feito pela Polícia da França adotou como aceitável a chance de se obter 2% de erro na taxa de FN<sup>80</sup>, mantendo-se essa taxa de FN e 0,5% de FP, tem-se 3,77% das correlações entre todas as amostras na região de incerteza, o que representa uma incerteza aceitável para o método.

É importante perceber que a aceitação de taxas de FP e FN não altera o número de classes determinado pelo método ULT, que foi 145 classes, apenas redefinem os valores de  $\rho_L$  e  $\rho_{NL}$ . Nesse caso, passam de 0,903 a 0,997 para 0,935 a 0,970 respectivamente (Figura 41).

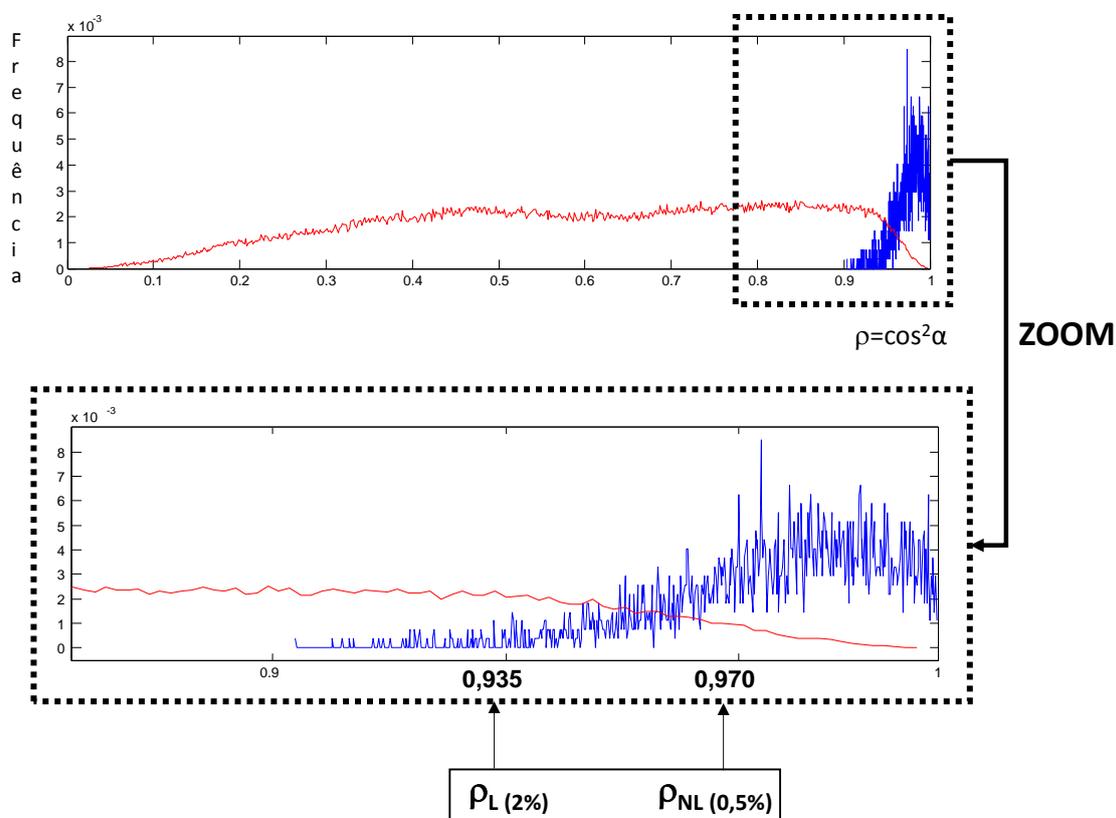


Figura 41: Valores Limites com taxa de Falsos Positivos de 0,5% e Falsos Negativos de 2%.

Os valores de  $\rho_{L2\%}$  (0,935) e  $\rho_{NLO,5\%}$  (0,970) quando confrontados com os valores limite encontrados em outros trabalhos (0,994 em Lion, na França<sup>80</sup> e em Brasília, Brasil<sup>92</sup>) podem parecer baixos. No entanto, esses trabalhos foram feitos utilizando a cocaína sal, uma matriz bem menos complexa que a cocaína base livre aqui analisada. A cocaína sal é um cristal, que durante sua formação pode incorporar pequenas quantidades de solvente em sua rede cristalina, processo chamado de oclusão.

Dessa forma, os solventes utilizados na produção da cocaína sal ficam presos dentro de seus cristais por algum tempo e quando essa amostra é dissolvida em água, tais solventes são liberados. O mesmo não ocorre com a cocaína base livre, pois esta não apresenta uma estrutura cristalina, reduzindo a capacidade de reter solventes residuais. No mais, a homogeneização da cocaína na forma de base livre também é mais complexa. Ainda assim, os resultados obtidos nesse trabalho sinalizaram a capacidade da técnica de *headspace* para determinar os solventes residuais nesse tipo de amostra e do método quimiométrico de comparar essas amostras, dando suporte à inteligência policial com informações que podem auxiliar na identificação de rotas do tráfico.

#### 4.5. Precisão do Método

A precisão do método foi avaliada nos níveis de repetibilidade e precisão intermediária. Para isso, duas amostras de cocaína base livre foram escolhidas como referência, seus cromatogramas foram apresentados na Figura 42.

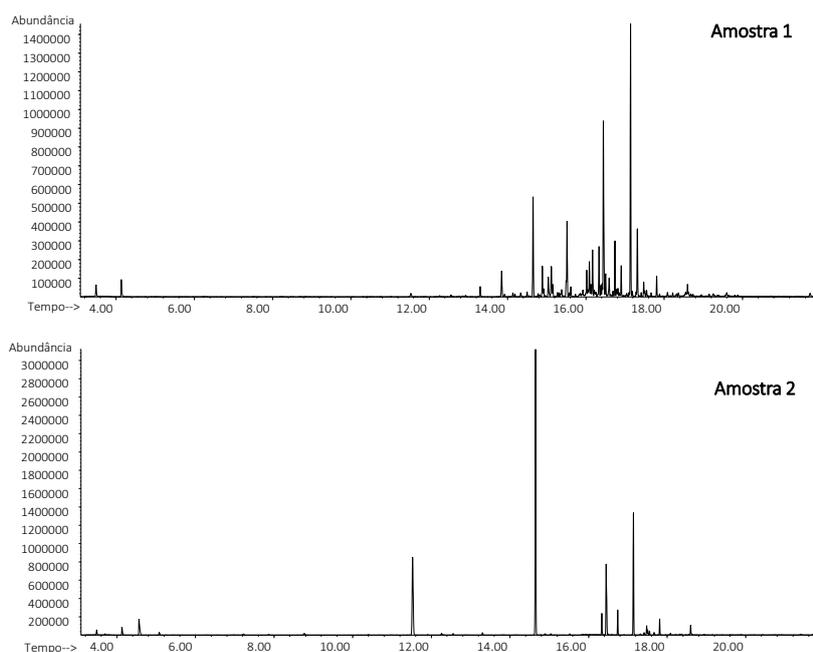


Figura 42: Cromatogramas de amostras selecionadas para avaliar a precisão do método.

Para se avaliar a repetibilidade do método, 6 replicatas de cada amostra foram analisadas, sob as mesmas condições e em um mesmo dia de análise, com o mesmo operador e equipamento sendo então feita a estimativa do desvio padrão (DP) e coeficiente de variação (CV) destas medidas. Os resultados descritos na Tabela 18, mostraram a repetibilidade do método. Foram obtidos coeficientes de variação iguais ou inferiores a 0,5%, os quais evidenciam a boa precisão desse nível.

Tabela 18. Resultados do parâmetro de repetibilidade.

<b>Amostra</b>	<b>Valor Médio</b>	<b>DP*</b>	<b>CV**</b>	<b>Valor Mínimo</b>
<b>1</b>	0,996	0,002	0,2%	0,993
<b>2</b>	0,994	0,005	0,5%	0,987

\*DP, Desvio Padrão; \*\*CV, Coeficiente de Variação.

O nível de precisão intermediária do método foi avaliado a partir da determinação dos solventes residuais das amostras realizada em doze replicatas feitas em dias diferentes por um mesmo analista e seus resultados expressos em média, DP e CV. Os resultados obtidos foram apresentados na Tabela 19 e revelaram boa precisão do método.

Tabela 19. Resultados do parâmetro de precisão intermediária.

<b>Amostra</b>	<b>Valor Médio</b>	<b>DP*</b>	<b>CV**</b>	<b>Valor Mínimo</b>
<b>1</b>	0,996	0,002	0,2	0,990
<b>2</b>	0,991	0,008	0,8	0,974

\*DP, Desvio Padrão; \*\*CV, Coeficiente de Variação.

É importante destacar ainda que os valores de  $\rho$  encontrados na avaliação da precisão do método foram superiores ao limite estabelecido de 0,970 tanto na repetibilidade, quanto na precisão intermediária (Figura 43). Isso indica que, de acordo com os limites estabelecidos pelo método, seria concluído que todas as replicatas das amostras utilizadas no teste de precisão estariam ligadas, ou pertenceriam a uma mesma origem.

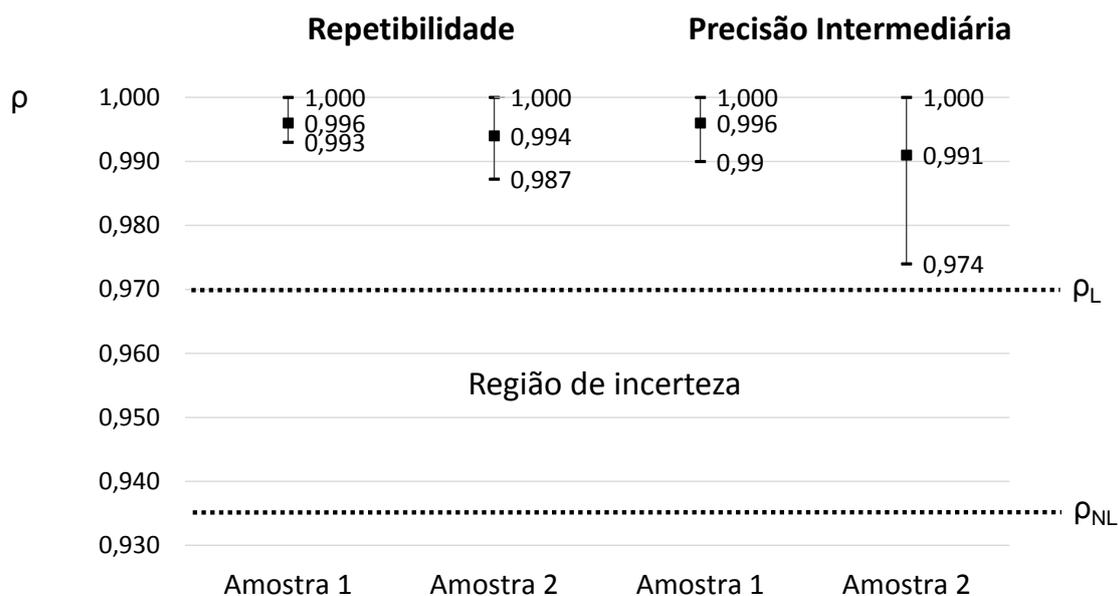


Figura 43: Cosseno quadrado ( $\rho$ ) obtido com amostras utilizadas para avaliar a precisão do método.

#### 4.6. Aplicações para casos reais

A partir dos limites fixados pela aplicação da metodologia proposta nesse trabalho, pôde-se correlacionar diferentes apreensões realizadas pela Polícia Federal.

Em um caso particular (Figura 44), pôde-se inferir que amostras de três diferentes Estados foram produzidas em um único laboratório e distribuídas para dois traficantes que reprocessaram as amostras. Esse segundo processamento das amostras é uma provável explicação para as correlações encontradas entre as amostras apreendidas no Estado do Mato Grosso e no Distrito Federal terem valores dentro da região de incerteza.

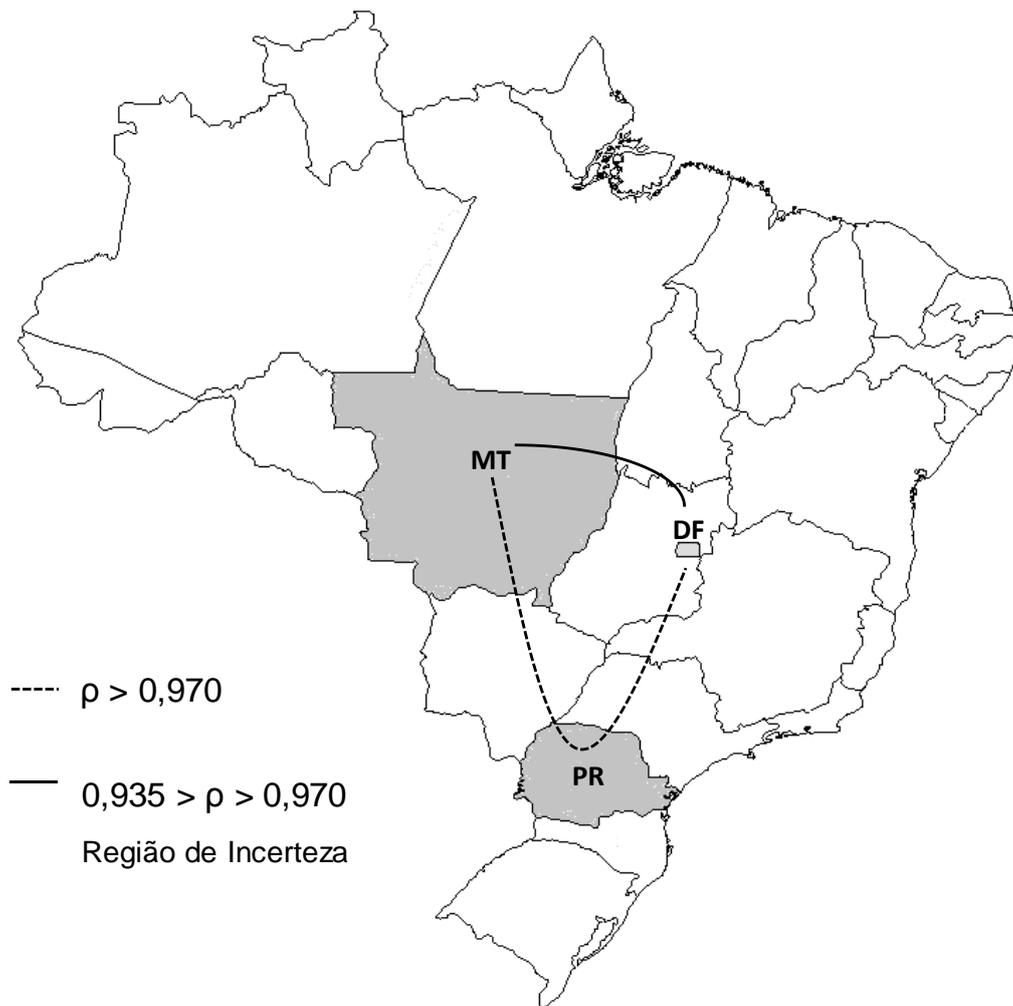


Figura 44: Caso 1 - Apreensões correlacionadas feitas pela Polícia Federal de amostras de cocaína base livre.

Um segundo caso correlacionou amostras do Mato Grosso do Sul e de São Paulo apreendidas em um mesmo ano. Cromatogramas dessas amostras foram apresentados na Figura 45, onde visualmente é possível verificar a similaridade entre tais amostras.

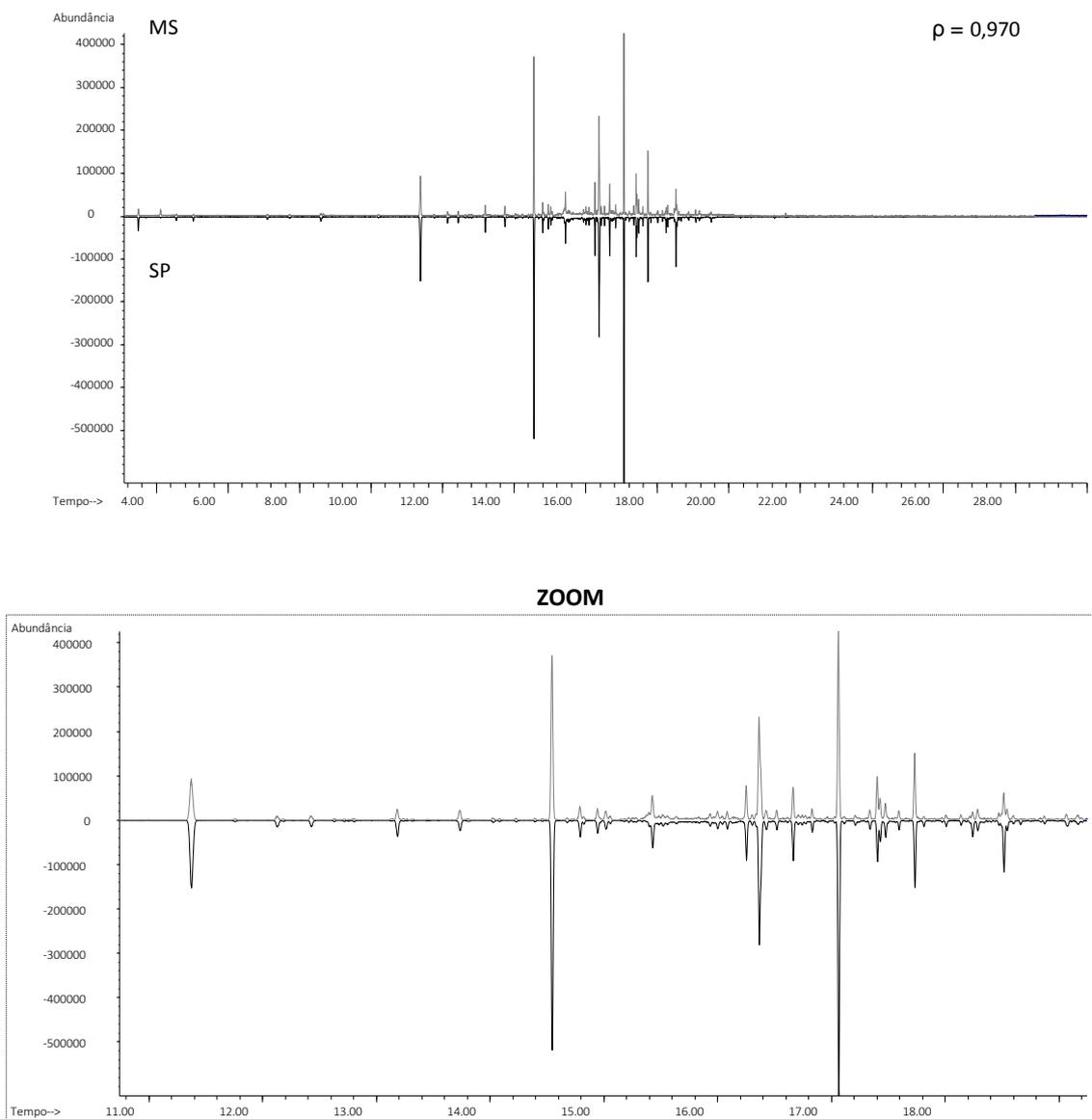


Figura 45: Caso 2 – Cromatogramas das amostras pertencentes às apreensões feitas pela Polícia Federal em Mato Grosso do Sul (na cor cinza) e São Paulo (na cor preta).

#### 4.7. Comparação do Método ULT Desenvolvido Utilizando Cocaína Sal e Cocaína Base Livre

O método ULT original, como descrito na Seção 2.6, foi desenvolvido com o uso de amostras de cocaína sal. Porém, com algumas alterações, também mostrou-se adequado para análise de amostras de cocaína base livre. Uma comparação entre os resultados obtidos por ambas as situações é descrita nessa seção. Na primeira situação, o método ULT foi desenvolvido utilizando a integração de picos cromatográficos. Nesse caso, 271 amostras de apreensões de cocaína sal realizadas pelo Polícia Federal foram analisadas. Para facilitar a compreensão na apresentação dos resultados, esse método foi

chamado de método ULT cocaína sal (ULT-SAL). A segunda situação, foi a descrita nesse trabalho, onde 619 amostras de cocaína base livre foram analisadas e os dados de massas acumuladas foram utilizados para identificar os valores de  $\rho_L$  e  $\rho_{NL}$ . Nesse caso, o método foi chamado de ULT cocaína base livre (ULT-BASE).

Inicialmente, verificou-se o novo critério utilizando a diferença percentual na determinação do número de classes. Os valores de diferença percentual para os diversos números de classes estudados no conjunto de amostras de cocaína sal são apresentados na Tabela 20 e Figura 46. De acordo com esses dados e seguindo o critério do método, deve-se selecionar o primeiro número de classes que apresenta diferença percentual máxima e com  $\rho_L$  maior que 0,90, que ocorreu com o número de classes igual a 23. Esse número de classes não foi o mesmo que o determinado pelo método original – ULT-SAL, o qual avaliava as AUCs. Pelo método ULT-SAL, o número de classes para esse conjunto de dados seriam 28. Contudo, o limite estabelecido para a identificação de amostras correlacionadas manteve-se o mesmo (igual a 0,994) independentemente do número de classes indicado (23 ou 28). Uma pequena variação ocorreu apenas no valor de  $\rho_L$ , sendo observada uma mudança de 0,934 para 0,925 considerando 28 e 21 classes, respectivamente. Tal variação gerou um aumento pouco significativo na região de incerteza do método de 9,7% para 10,2%, incluindo 0,5% das amostras nessa região.

Era previsto que o método ULT-BASE apresentasse um aumento na região de incerteza em comparação com o método ULT-SAL devido ao aumento de informações utilizadas para a comparação entre amostras. Isso é justificado porque quando os picos cromatográficos eram selecionados os compostos não previstos pelos parâmetros de integração não eram utilizados na correlação entre amostras, acarretando possíveis perdas de informação. Dessa forma, embora tenha havido um aumento na região de incerteza teve-se também um aumento na informação química que é utilizada, o que pode gerar benefícios, reduzindo as taxas de FP e FN em algumas situações.

Tabela 20. Valores de AUC, percentagem da diferença,  $\rho_L$  e  $\rho_{NL}$  para os diversos números de classes estudados no conjunto de amostras analisadas pelo método ULT-SAL.

N° de Classe	AUC	% da Diferença	$\rho_L$	$\rho_{NL}$	N° de Classe	AUC	% da Diferença	$\rho_L$	$\rho_{NL}$
8	0,9975	0,20	0,428	0,974	31	0,9934	0,00	0,948	0,997
9	0,9995	0,01	0,531	0,974	32	0,9934	0,00	0,949	0,997
10	0,9996	0,00	0,619	0,974	33	0,9934	0,01	0,951	0,997
11	0,9995	0,00	0,759	0,98	34	0,9935	0,00	0,951	0,997
12	0,9995	0,00	0,759	0,98	35	0,9935	0,04	0,951	0,997
13	0,9995	-0,05	0,767	0,98	36	0,9939	0,07	0,951	0,997
14	0,9991	-0,15	0,808	0,98	37	0,9946	0,00	0,959	0,997
15	0,9976	-0,03	0,812	0,994	38	0,9946	0,00	0,967	0,997
16	0,9973	0,02	0,812	0,994	39	0,9946	0,02	0,967	0,997
17	0,9975	0,14	0,852	0,994	40	0,9948	0,02	0,971	0,997
18	0,9989	0,01	0,885	0,994	41	0,9950	0,00	0,971	0,997
19	0,9989	0,01	0,885	0,994	42	0,9950	0,00	0,971	0,997
20	0,9990	0,04	0,885	0,994	43	0,9950	0,00	0,971	0,997
21	0,9994	0,0003	0,925	0,994	44	0,9950	-0,13	0,971	0,997
22	0,9994	0,0006	0,925	0,994	45	0,9937	0,01	0,975	0,999
23	0,9994	0,0037	0,925	0,994	46	0,9938	0,00	0,975	0,999
24	0,9995	0,0007	0,93	0,994	47	0,9938	0,00	0,975	0,999
25	0,9995	0,0004	0,934	0,994	48	0,9939	0,00	0,975	0,999
26	0,9995	0,0001	0,934	0,994	49	0,9939	0,00	0,975	0,999
27	0,9995	0,0005	0,934	0,994	50	0,9939	0,00	0,975	0,999
28	0,9995	-0,6	0,934	0,994	51	0,9939	-0,02	0,975	0,999
29	0,9931	0,02	0,934	0,997	52	0,9936	0,01	0,977	0,999
30	0,9932	0,02	0,941	0,997					

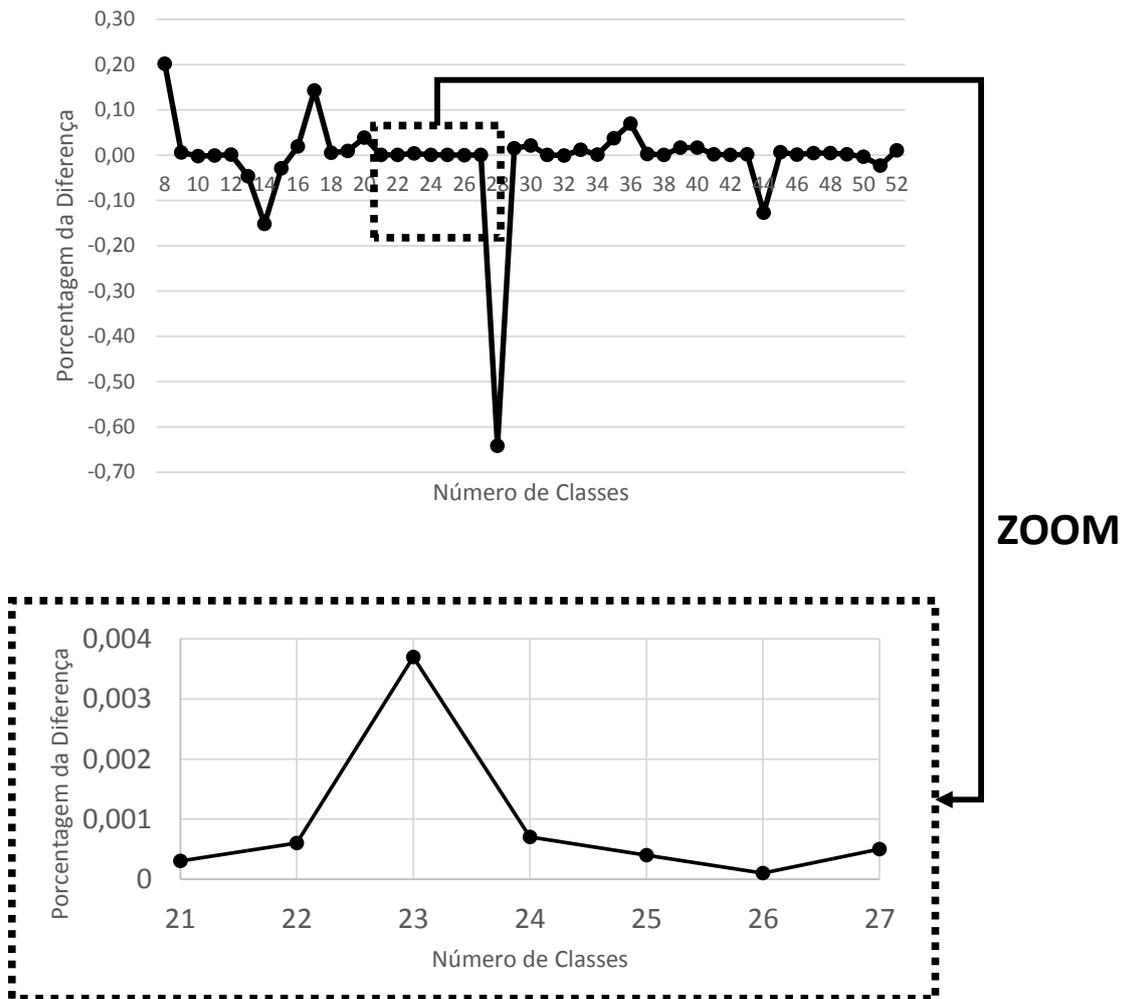


Figura 46: Valores de porcentagem da diferença para o conjunto de amostras analisadas pelo método ULT-SAL.

A avaliação da região de incerteza feita para o método ULT-SAL, mantendo o número de classes determinado pelo método original, ou seja, 28 classes, verificou o número de correlações nessa região com a variação das taxas de FP e FN. A Tabela 21 mostra que a aceitação de uma taxa de FP de 0,25% reduz significativamente o número de correlações entre amostras ligadas na região de incerteza. A aceitação dessa taxa de FP reduz de 53,8% para 4,5% das correlações de amostras correlacionadas na região de incerteza.

Tabela 21. Variação de da taxa de falsos positivos (FP), mantendo  $\rho_L=0,934$ , ou seja, nenhum falso negativo para o conjunto de dados de cocaína sal.

FP (%)	$\rho_{NL}$	% de amostras ligadas na região de incerteza	% de correlações na região de incerteza
0	0,994	53,84	9,67
0,25	0,972	4,50	2,26
0,5	0,964	2,16	1,71
1	0,952	0,52	1,03
1,5	0,940	0,04	0,40
2	0,935	0,02	0,05

Da mesma forma, avaliou-se o número de amostras não correlacionadas na região de incerteza variando a taxa de FN (Tabela 22). Nesse caso, a diminuição do número de correlações na região de incerteza foi menor que a observada na Tabela 21.

Tabela 22. Variação de falsos negativos (FN), mantendo  $\rho_{NL}=0,994$ , ou seja, nenhum falso positivo para o conjunto de dados de cocaína sal.

FN (%)	$\rho_L$	% de amostras não ligadas na região de incerteza	% de correlações na região de incerteza
0	0,934	2,13	9,67
0,5	0,952	1,02	8,64
1	0,957	0,84	8,41
2	0,963	0,55	8,03
3	0,967	0,43	7,78
5	0,973	0,25	7,29

O método ULT-BASE com 0,5% de FP e 2% de FN aceitou uma taxa de 3,77% de correlações na região de incerteza, valor superior ao obtido pelo método ULT-SAL, cerca de 0,35% das correlações com nenhum FP e uma taxa de 2% de FN. Essas comparações forneceram subsídios para concluir que apesar do aumento na região de incerteza, a porcentagem das amostras que puderam ser discriminadas entre correlacionadas e não correlacionadas não variou significativamente, indicando que o uso do espectro de massa acumulado permitiu a análise de uma amostra muito mais complexa sem grandes perdas de eficiência em relação ao uso da seleção e integração dos picos cromatográficos, algo que é extremamente difícil para cocaína base.

## 5. Conclusão

Com este trabalho pôde-se estabelecer um limite capaz de identifique amostras de cocaína base livre correlacionadas e não correlacionadas sem a necessidade de um

conhecimento da população *a priori*, confirmando a funcionalidade do Método ULT em outras análises. O sucesso em estabelecer essa metodologia deveu-se a dois fatores principais: a homogeneização das amostras e o uso do espectro de massas na análise de dados. A homogeneização sempre foi um problema para esse tipo de amostra. A solução apresentada utilizando N<sub>2</sub> líquido ainda não é a mais fácil de ser aplicada em grande escala, mas fornece indícios para novos caminhos.

O uso do espectro de massas na análise de dados foi um grande avanço para o Método ULT pois o generalizou, tornando possível aplica-lo em uma infinidade de problemas que por vezes não são facilmente solucionados com a cromatografia.

Sugere-se para um próximo trabalho a avaliação do método com a análise de amostras por injeção direta no espectrômetro de massas, o que reduziria o tempo de análise e a quantidade de solventes e gases usados. Testes com amostras diferentes também são interessantes para avaliar o funcionamento do método com outras matrizes.

# *Referências*

1. United Nations Office on Drugs and Crimes (UNODC); *World Drug Report*, United Nations: New York, **2013**.
2. Estimativa do número de usuários de crack e/ou similares nas capitais do país, *Casa Civil*, Brasil, **2013**.
3. Chiarotti, M.; Fucci, N.; *Forensic Sci. Int.* **1999**, 733, 127.
4. Elsherbini, S.H.; *Forensic Sci. Rev.* **1998**, 10, 2.
5. Campanella, L.; Colapichchioni, C.; Tomassetti, M.; Dezzi, S.; *J. Pharm. Biomed. Anal.* **1996**, 14, 1047.
6. Moffat, A.C.; Osselton, M.D.; Widdop, B.; *Clarke's Analysis of Drugs and Poisons*, The Pharmaceutical Press, London, England, **2004**.
7. Piñero, L.E.; Casale, J.F.; *Microgram J.* **2006**, 4, 47.
8. Nic Daéid, N.; Waddell, R.J.H.; *Talanta* **2005**, 67, 280.
9. Grobério, T.S.; Zacca, J.J.; Talhavini, M.; Braga, J.W.B.; *J. Braz. Chem. Soc.* **2014**, 25, 1696.
10. Rodrigues, N.V.S.; Cardoso, E.M.; Andrade, M.V.O.; Donnicia, C.L.; Sena, M.M.; *J. Braz. Chem. Soc.* **2013**, 24, 507.
11. Rossi, R.C.; *Determinação de componentes voláteis minoritários em amostras de cocaína empregando extração por headspace estático e cromatografia a gás*. Brasil: Nova Hamburgo, **2009**.
12. Chiarotti, M.; Fucci, N.; *Forensic Sci. Int.* **1988**, 37, 47.
13. Oropesa Ortuno L.; Oropesa Garcia J.C.; Oropesa Garcia F.; *Manual internacional de investigation y control quimico antonarcoticos*, Ed. FTM, Bolivia: La Paz, **2005**.
14. Moore, J.M.; Casale, J.F.; *J. Chromatogr A.* **1994**, 674, 165.
15. United Nations International Drug Control Programme (UNODCCP); *Drug characterization/Impurity Profiling – Background and Concepts*, United Nations: New York, **2001**.
16. Barros Neto, B.; Scarminio, I.S.; Bruns, R.; *Quím. Nova* **2006**, 9, 1401.
17. Vargas, R.M.; Talhavini, M.; *Cocaína. Relatório de atividades do projeto FAPDF no 193.000.360/99*, Fundação de Apoio a Pesquisa do Distrito Federal, Brasil: Brasília, **2000**.
18. Romão, W.; Schwab, N.V.; Bueno, M.I.M.S.; Sparrapan, R.; Eberlin, M. N.; Martiny, A.; Sabino B.D.; Maldaner, A.O.; *Quim. Nova*, **2011**, 34, 1717.
19. Siqueira, L.P.; Fabri, A.C.O.C.; Fabri, R.L.; *Rev. Elet. Farm.* **2011**, 8, 75.

20. Oga, S; Camargo, MMA; Batistuzzo, JAO; *Fundamentos de Toxicologia*: São Paulo: Atheneu, **2008**.
21. Fucci, N. & De Giovanni, N.; *Forensic Sci. Int.* **1998**, *95*, 247. Brunt, T.; Rigter, S.; Hoek, J.; Vogels, N.; Van Dijk, P.; Niesink, R.; *Addiction*, **2009**, *104*, 798.
22. Evrard, I.; Legleye, S.; Taïrou, A.C.; *Int. J. Drug Policy* **2010**, *21*, 399.
23. Conceição, V.N.; Souza, L.M.; Merlo, B.B.; Filgueiras, P.R.; Poppi, R.J.; Romão, W.; *Quim. Nova* **2014**, *37*, 1538.
24. United Nations Office on Drugs and Crimes (UNODC); *World Drug Report*, United Nations: New York, **2011**.
25. Carvalho, D.G.; Determinação dos componentes do cloridrato de cocaína ilegalmente comercializado na região metropolitana de São Paulo no ano de 1997. Dissertação de mestrado, Universidade de São Paulo, **2010**.
26. Bruni, A.T.; Velho, J.A.; Oliveira, M.F.; *Fundamentos de Química Forense – Uma análise prática da química que soluciona crimes*. Milenium Editora, Brasil: Campinas, **2012**.
27. Maiette, S.; Castagna, F.; Molin, L.; Ferrara, S.D.; Traldi, P.; *J. Mass Spectrom.* **2009**, *44*, 1124.; Zacca, J.J.; Botelho, E.D.; Vieira, M.L.; Ferreira, L.S.; Maldaner, A.O.; *Sci. Justice* **2014**, *54*, 300.
28. Bacchi, E.M.; *Alcalóides tropânicos*, em: Simões, C.M.O.; Schenkel, E.P.; Gosmann, G. (Eds.), *Farmacognosia: da planta ao medicamento*, Editora da UFRGS/Editora da UFSC, Brasil: Porto Alegre/Florianópolis, **2004**.
29. Casale, J.; Hays, P.A.; Toske, S.G.; Berrier, A.L.; *J. Forensic Sci.*, **2007**, *52*, 860.
30. Jansen, K.E.; Walter, L.; Fernando, A.R.; *J. Forensic Sci.*, **1992**, *37*, 436.
31. Chasin, A.A.M; Mídio, A.F.; *Rev. Bras. de Toxicol.*, **1991**, *4*, 23.
32. Wielbo, D.; Tebbett, I.R.; *Scientific & Technical*, **1993**, *33*, 25.
33. Decker, W.J.; *Introduction and history*, In: Haley, J.J.; Berndt, W.O. (Eds.) *Handbook of Toxicology*, Hemisphere Publishing Corporation, United Nations: New York, **1987**.; Vargas, R.M.; *Determinação do DNA da cocaine, importante ferramenta investigativa*. Perícia Federal, **2002**.
34. Moore, J.M.; Casale, J.F.; *J. Chromatogr. A* **1994**, *674*, 165.; Floriani, G.; Gasparetto, J.C.; Pontarolo, R.; Gonçalves, A.G.; *Forensic Sci. Int.* **2014**, *235*, 32.
35. Buján, L.; Fernández, P.; Lafuente, N.; Aldonza, M.; Bermejo, A.M.; *Anal. Chem. Lett.* **2001**, *34*, 2263.; Piñero, E.L.; Casale, J.F.; *Microgram J.* **2006**, *4*, 47.
36. Ravreby, M; *J. Forensic Sci.* **1987**, *32*, 20.

37. Ryder, A.G.; O'Connor, G.M.; Glynn T.J.; *J. Forensic Sci.* **1999**, *44*, 1013.
38. Maharaj, R.; *J. of Third World Med.* **2009**, *7*.
39. Pérez-Alfonso, C; Galipienso, N.; Garrigues, S.; Guardia, M.; *Forensic Sci. Int.* **2014**, *237*, 70.
40. Walczak, B.; Massart, D.L.; *Chemom. Intell. Lab. Syst.* **1998**, *41*, 1.
41. Næs, T.; Isaksson, T.; Fearn, T.; Davies, T.; A user-friendly guide to Multivariate Calibration and Classification, NIR Publications: Chichester, **2004**.
42. Neto, W.B.; Parâmetros de qualidade de lubrificantes e óleo de oliva através de espectroscopia vibracional, calibração multivariada e seleção de variáveis. Tese de Doutorado, Universidade Estadual de Campinas, SP, **2005**.
43. CENTNER, V.; MASSART, D. Elimination of uninformative variables for multivariate calibration, *Anal. Chem.*, v.68, n.21, 3851-3858, 1996.
44. Parisotto, G.; Determinação do número de acidez total em resíduo de destilação atmosférica e de vácuo do petróleo empregando a espectroscopia no infravermelho (ATR-FTIR) e calibração multivariada. Dissertação de mestrado, Universidade Federal de Santa Maria, RS, **2007**.
45. Araujo, T. P.; Emprego de espectroscopia no infravermelho e métodos quimiométricos para a análise direta de tetracilinas em leite bovino. Dissertação de Mestrado, Universidade Estadual de Campinas, SP, **2007**.
46. Annual Book of ASTM Standarts, Standarts Practices for Infrared, Multivariate, Qualitative Analysis, ASTM International E1655-05, West Conshohocken, USA: Pennsylvania, **2005**.
47. Valderrama, P.; Poppi, R.J.; Braga, J.W.B.; *J. Agric. Food Chem.* **2007**, *55*, 8331.
48. Pierna, J.A.F.; Jin, L.; Wahl, F.; Faber, N.M.; Massart, D.L.; *Chemom. Intell. Lab. Syst.* **2003**, *65*, 281.
49. Souza, A. M.; Poppi, R. J.; *Quim. Nova.* **2012**, *35*, 223.; Braga, J. W. B.; Pastore, T. C. M.; Coradin, V. T. R.; Camargos, J. A. A.; Silva, A. R.; *IWA J.* **2011**, *32*, 285.
50. Oliveira, E.M.; Braga, J.W.B.; Costa, A.F.; *Quím. Nova*, **2015**, *38*(9), 1176.
51. Ferreira, M.M.C.; *Quimiometria: conceitos, métodos e aplicações*. Editora Unicamp, Brasil: Campinas, **2015**.
52. INMETRO, Vocabulário Internacional de Metrologia – Conceitos fundamentais e gerais e termos associados, Brasil: Brasília, **2012**.
53. Valderrama, P.; Braga, J.W.B.; Poppi, R.J.; *Quim. Nova*, **2009**, *32*, 1278.

54. INMETRO, Orientações sobre Validação de Métodos de Ensaio Químicos, DOQ-CGCRE-008, Brasil: Rio de Janeiro, **2007**.
55. MAPA, Guia de validação e controle de qualidade analítica: fármacos em produtos para alimentação e medicamentos veterinários, Brasil: Brasília, **2011**.
56. Olivieri, A.C.; Faber, M.N.; Ferré, J.; Boqué, R.; Kalivas, J.H.; Mark, H.; *Pure Appl. Chem.*, **2006**, 78, 633.
57. Lorber, A.; *Anal. Chem.*, **1986**, 58, 1167.
58. ISO, 11843-2, Capability of Detection, International Standards Organization, Switzerland: Geneva, **2000**.
59. Ortiz, M.C.; Sarabia, L.A.; Herrero, A.; Sánchez, M.S.; Sanz, M.B.; Rueda, D.; Giménez, M.E.; Meléndez, M.E.; *Chemom. Intell. Lab. Syst.*, **2003**, 69, 21.
60. Van Der Voet, H. ; *J. Chem.*, **1999**, 13, 195.
61. Lima, F.S.G.; Borges, L.E.P.; *J. Near Infrared Spectrosc.*, **2002**, 10, 269.
62. Noord, O.E.; *Chemom. Intell. Lab. Syst.*, **1994**, 25, 85.; Honorato, F.A.; Neto, B.B.; Martins, M.N.; Galvão, R.K.H.; Pimentel, M.F.; *Quim. Nova*, **2007**, 5, 1301.
63. Departamento de Polícia Federal, Instrução Técnica No 456 006/2006 – GAB/DITEC, Brasil: Brasília, **2006**.
64. Scientific Working Group for the Analysis of Seized Drugs (SWGDRUG) Recommendations, United States Department of Justice, United Nations: Washington, **2011**.
65. United Nations Division of Narcotic Drugs (UNODC), Recommended methods for testing cocaine, Manual for Use by National Narcotics Laboratories, United Nations: New York, **1986**.
66. Kennard, R.W.; Stone, L.A.; *Technometrics*, 1969, 11, 137.
67. Maldaner, A.O.; Botelho, E.D.; Zacca, J.J.; Camargo, M.A.; Braga, J.W.B.; Grobério, T.S.; *J. Braz. Chem. Soc.*, **2015**, 26(6), 1227.
68. Brereton, R.G.; *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, Wiley, United Nations: Chichester, **2003**.
69. Martens, H.; Naes, T.; *Multivariate Calibration*, John Wiley & Sons, United Nations: New York, **1989**.
70. Jarque, C.M.; Bera, A.K.; *Int. Stat. Rev.*, **1987**, 55, 163.
71. Cole, M.D.; *Forensic Sci. Rev.*, **1998**, 10, 113.
72. Otero, R.; Carrera, G.; Dulsat, J.F.; Fabrega, J.L.; Claramunt, J.; *J. Chromatogr., A*, **2004**, 1057, 193.
73. Chiarotti, M.; Fucci, N.; *Forensic Sci. Int.*, **1988**, 37, 47.
74. Morello, D.R.; Meyers, R.P.; *J. Forensic Sci.*, **1995**, 40, 957.

75. Esseiva, P.; Dujourdy, L.; Besacier, F.; Margot, P.; Lociciro, S.; Hayoz, P.; *Forensic Sci. Int.*, **2007**, *167*, 220.
76. Chiarotti, M.; Marsili, R.; Moreda-Piñero, A.; *J. Chromatogr.*, **2002**, *772*, 249.
77. Lydon, J.; Casale, J. F.; Kong, H. Sullivan, J.H.; Daughtry, C. S. T.; Bailey, B.; *Photochem. Photobiol.*, **2009**, *85*, 1156.
78. Moore, J. M.; Casale, J. F.; Cooper, D. A.; *J Chromatogr.*, **1996**, *756*, 193.
79. Dujourdy, L.; Besacier, F.; *Forensic Sci. Int.*, **2008**, *179*, 111-122.
80. Cartier, J.; Gueniat, O.; Cole, M.D.; *Science & Justice*. **1997**, *37*, 175.
81. Garzón, W.F.; Parada, F.; Florián, N.M.; *Vitae*. **2009**, *16*, 228.
82. Yonamine, M.; Tawil, N.; Moreau, R.L.M.; Silva, O.A.; *J. Chromatog. B*. **2003**, *789*, 73.
83. Sporkert, F.; Pragst, F.; *Forensic Sci. Int.*, **2000**, *107*, 129.
84. Janzen, K.E.; Walter, L.; Fernando, A.R.; *J. Forensic Sci.*, **1992**, *37*, 436.
85. Margot, P.; Dujourdy, L.; Barbati, G.; Taroni, F.; Guéniat, O.; Esseiva, P.; Anglada, F.; *Forensic Sci. Int.*, **2003**, *131*, 171.
86. Esseiva, P.; Anglada, F.; Dujourdy, L.; Taroni, F.; Margot, P.; Du Pasquier, E.; Dawson, M.; Roux, C.; Doble, P.; *Talanta.*, **2005**, *67*, 360.
87. Lociciro, S.; Hayoz, P.; Esseiva, P.; Dujourdy, L.; Besacier, F.; Margot, P.; *Forensic Sci. Int.*, **2007**, *167*, 220.; Lociciro, S.; Hayoz, P.; Esseiva, P.; Dujourdy, L.; Besacier, F.; Margot, P.; *Forensic Sci. Int.*, **2008**, *177*, 199.
88. Esseiva, P.; Gaste, L.; Alvarez, D.; Anglada, F.; *Forensic Sci. Int.*, **2011**, *207*, 27.
89. Grobério, T.S.; Desenvolvimento de uma metodologia para comparação de amostras de sal de cocaína pela determinação de solventes residuais e análise quimiométrica. Dissertação de Mestrado, Universidade de Brasília, DF, **2012**.
90. Zacca, J.J.; Groberio, T. Maldaner, A.O.; Vieira, M.L.; Braga, J.W.B.; *Anal Chem.*, **2013**, *85*, 2457.
91. Prado, B.Q.M.; Análise de agrupamentos das taxas de incidência de dengue nos estados brasileiros. Trabalho de Conclusão de Curso, Universidade Federal de Uberlândia, MG, **2015**.
92. [https://www.mathworks.com/help/bioinfo/ref/seqlinkage.html?s\\_tid=srchtitle](https://www.mathworks.com/help/bioinfo/ref/seqlinkage.html?s_tid=srchtitle) Acessado em 29/11/2016.
93. <https://arxiv.org/pdf/1109.2378.pdf> , acessado em 09/12/2016.
94. Thornton, J.I.; *J. Forensic Sci.* **1975**, *19(4)*, 873.; Obuchowski, N.A.; *American Journal of Roentgenology*. **2005**, *184(2)*, 364-372.
95. Silva Junior, R.C.; Gomes, C.S.; Goulart Júnior, S.S.; Almeida, F.V.; Grobério, T.S.; Braga, J.W.B.B.; Zacca, J.J.; Vieira, M.L.; Botelho, E.D.; Maldaner, A.O.; *Forensic Sci. Int.*, **2012**, *221*, 113.