

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE QUÍMICA

CARLOS ALBERTO RIOS

**ESTUDO DE MANCHAS DE SANGUE: UMA ABORDAGEM FORENSE
EMPREGANDO ESPECTROSCOPIA RAMAN, IMAGENS DIGITAIS E
FERRAMENTAS QUIMIOMÉTRICAS**

Belo Horizonte

2018

UFMG / ICEx / DQ. 1294^a

D. 709^a

CARLOS ALBERTO RIOS

**ESTUDO DE MANCHAS DE SANGUE: UMA ABORDAGEM FORENSE
EMPREGANDO ESPECTROSCOPIA RAMAN, IMAGENS DIGITAIS E
FERRAMENTAS QUIMIOMÉTRICAS**

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

Belo Horizonte

2018

Ficha Catalográfica

R586e Rios, Carlos Alberto
2018 Estudos de manchas de sangue [manuscrito] : uma
D abordagem forense empregando espectroscopia Raman,
imagens digitais e ferramentas quimiométricas / Carlos
Alberto Rios. 2018.
[xviii], 145 f. : il.

Orientadora: Mariana Ramos de Almeida.

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

Inclui bibliografia.

1. Química analítica - Teses 2. Sangue - Análise e
química - Teses 3. Raman, Espectroscopia de - Teses 4.
Processamento de imagens - Técnicas digitais - Teses
5. Química legal - Teses 6. Análise de regressão -
Teses 7. Mínimos quadrados - Teses I. Almeida, Mariana
Ramos de, Orientadora II. Título.

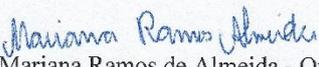
CDU 043

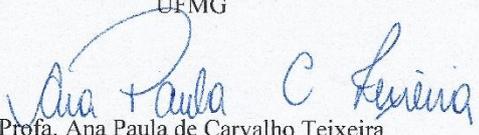


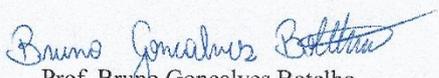
"Estudo de Manchas de Sangue: Uma Abordagem Forense Empregando Espectroscopia Raman, Imagens Digitais e Ferramentas Quimiométricas"

Carlos Alberto Rios

Dissertação aprovada pela banca examinadora constituída pelos Professores:


Profa. Mariana Ramos de Almeida - Orientadora
UFMG


Profa. Ana Paula de Carvalho Teixeira
UFMG


Prof. Bruno Gonçalves Botelho
UFMG

Belo Horizonte, 30 de julho de 2018.

A Eva Victória Rios (in memoriam), que com sua simplicidade, me ensinou que as
pessoas demonstram seu caráter através de pequenos gestos;
A Francisco, meu pai; a Marilene minha mãe, pelo vínculo sublime na minha vida, e
pelos exemplos que me impulsionam e me trazem esperança para seguir em frente em
busca de um futuro melhor;
A Lázaro, meu avô; a Maria dos Anjos, minha avó, que sempre trabalharam no
ambiente rural e com simplicidade me instruíram a ser sempre verdadeiro e honesto;

DEDICO

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus pela oportunidade de ingressar em uma pós-graduação e por mais uma vitória em minha vida;

Aos meus pais pelo exemplo de vida;

À Universidade Federal de Minas Gerais (UFMG) pela estrutura e incentivo durante todo o mestrado;

Ao Departamento de Química da UFMG (DQI-UFMG), com destaque especial aos professores pelos ensinamentos e dedicação durante as disciplinas;

A minha orientadora Prof. Dra. Mariana Ramos de Almeida pela orientação, paciência e dedicação durante o mestrado, em especial por seus ensinamentos e esforço que foram de grande importância para a conclusão deste estudo;

Ao professor da Faculdade de Medicina Leonardo Vasconcellos pelo auxílio com os trâmites no Comitê de Ética em Pesquisa (COEP-UFMG);

Aos professores que disponibilizaram equipamentos e o espaço em seus laboratórios para a execução da parte experimental do projeto: Prof. Dr. Cristiano Fantini (Física), Prof. Dr. Rodinei Augusti (Química) e a Prof. Dra. Ana Paula de Carvalho Teixeira (Química);

Por fim, um agradecimento aos órgãos financiadores: CAPES, FAPEMIG e CNPq;

“Viver é enfrentar um problema atrás do outro. O modo como você o encara é que faz a diferença.” **Benjamin Franklin**

RESUMO

O estudo das evidências e vestígios encontrados no local de um crime é uma das etapas essenciais de uma investigação criminal. Dessa forma, a proposta deste trabalho foi realizar um estudo com manchas de sangue empregando espectroscopia Raman e imagens digitais juntamente com ferramentas quimiométricas. As manchas de sangue foram analisadas ao longo do tempo e em diferentes condições ambientes de exposição. No estudo foram coletadas amostras de 10 doadores voluntários, homens e mulheres com idade entre 20 e 50 anos. Uma análise exploratória dos resultados obtidos foi realizada empregando a análise por componentes principais (PCA). Os modelos de regressão foram construídos usando o método de regressão por mínimos quadrados parciais (PLS) para estimar o tempo das manchas. Na metodologia empregando espectroscopia Raman, o intervalo de tempo estudado foi de 269 horas e o melhor modelo de regressão obtido foi construído usando 7 variáveis latentes com uma variância explicada na matriz \mathbf{X} de 84,5% e no vetor \mathbf{y} de 98,49%. O erro médio obtido foi de 19h e o coeficiente de correlação foi de 0,91. A banda mais importante dos espectros Raman obtidos foi atribuído à desoxihemoglobina (meta-Hb) em 377 cm^{-1} , uma vez que essa banda está relacionada com as mudanças químicas que ocorrem no sangue ao longo do tempo de exposição. As imagens digitais foram obtidas com um *smartphone Galaxi S8* e 9 sistemas de cores foram usados para a construção dos modelos de regressão, assim como diferentes substratos e condições ambientes. Para o melhor modelo de regressão foram selecionadas 6 variáveis latentes com 98,26% da variância explicada na matriz \mathbf{X} e de 89,19% no vetor \mathbf{y} , que descreve o tempo de exposição em escala logarítmica dos minutos. O erro médio de predição do modelo foi de 0,43 e o intervalo de tempo monitorado foi de 3967 horas. Os resultados obtidos nesse trabalho mostram que as condições ambientes influenciam a velocidade das reações químicas envolvidas no envelhecimento de uma mancha de sangue e podem causar altos erros de predição, por isso, estas informações devem ser consideradas no momento da construção dos modelos de regressão. Por fim, as metodologias propostas apresentam potencial para acompanhar as mudanças químicas e físicas que ocorrem em manchas de sangue em função do tempo de exposição e podem ser aplicadas como testes de triagem para fins forenses.

Palavras chave: manchas de sangue; espectroscopia Raman, imagens digitais, análises forense; regressão pelo método dos mínimos quadrados parciais.

ABSTRACT

BLOODSTAIN STUDY: A FORENSIC APPROACH EMPLOYING RAMAN SPECTROSCOPY, DIGITAL IMAGES AND CHEMOMETRICS TOOLS

The study of evidence and traces found at the scene of a crime is one of the essential steps in a criminal investigation. Thus, the purpose of this work was to perform a bloodstain study using Raman spectroscopy and digital images, along with chemometric tools. Bloodstains were analyzed over time and the different environmental exposure conditions. In the study, samples were collected from 10 voluntary donors, men and women aged between 20 and 50 years. The chemometric tools were used for exploratory analysis, in which case the principal component analysis (PCA) was employed. Multivariate regression models were constructed using partial least squares regression (PLS) to estimate the exposure time of the stains. In the methodology using Raman spectroscopy, the time interval studied was 269 hours and the regression model was constructed employing 7 latent variables with a variance explained in the \mathbf{X} matrix of 84.5% and in the vector \mathbf{y} of 98.49%, which provides exposure time information in hours. The mean error of prediction was 19h and the correlation coefficient of 0.91. The most important Raman spectrum band for the regression model were assigned to deoxyhemoglobin (met-Hb) at 377 cm^{-1} , since this band are related to chemical changes occurring in the blood over the exposure time. The digital images were obtained with a *smartphone Galaxy S8* and 9 color systems were used for the construction of the regression models, as well as different substrates and environmental conditions. For the best regression model, we selected 6 latent variables with 98.26% of the variance explained in matrix \mathbf{X} and 89.19% in vector \mathbf{y} , which describes the time of logarithmic minute exposure. The mean prediction error of the model was 0.43 and the time interval monitored was 3967 hours. The results obtained in this work show that environmental conditions influence the speed of the chemical reactions that occur in the bloodstain and cause high prediction errors, therefore this information should be considered when the regression models are constructed. Finally, the proposed methodologies present potential to accompany the chemical and physical changes that occur in bloodstains as a function of the exposure time and can be applied as screening tests for forensic purposes.

Keywords: bloodstain; Raman spectroscopy; digital images; forensic analysis; partial least squares.

LISTA DE FIGURAS

Figura 1. Esquema da estrutura da hemoglobina de acordo com a ref. 28.....	24
Figura 2. Alterações químicas da mancha em função do tempo de exposição baseada na ref. 28	25
Figura 3. Interferentes na análise de manchas de sangue.....	25
Figura 4. Interpretação geométrica da PCA. (dir.)- Amostras em função das variáveis e (esq.) – Escores de PC1 contra PC2 baseada na ref. 53	30
Figura 5. Representação da decomposição da matriz de dados de acordo com o PCA, baseada na referência 53	31
Figura 6. Representação geométrica do PLS, baseada na ref. 53.....	34
Figura 7. Esquema ilustrativo para o pré-processamento das variáveis baseada na ref. 20.....	38
Figura 8. Diagramas de energia envolvidos nos processos de espalhamento Raman e Rayleigh. ($h\nu_0$ energia incidente, e_ν é a diferença de energia entre os dois estados vibracionais) baseada na ref. 88	43
Figura 9. Metodologia de coleta até a recuperação das manchas de sangue.....	46
Figura 10. Espectrômetro Senterra Raman (Bruker), detector CCD e microscópio Olympus...	47
Figura 11. Composição de um espectro Raman de uma mancha de sangue e os seus possíveis interferentes.....	48
Figura 12. Espectro da mancha de sangue sem (acima) e após (abaixo) o pré-processamento usando mínimos quadrados ponderados – WLS (2º grau) e normalização vetorial	50
Figura 13. Espectros obtidos incidindo diretamente o laser na mancha. I-Polinômio de 1º ordem com uma janela de 101 pontos; II – Polinômio de 1º ordem com uma janela de 51 pontos; III- Espectro sem alisamento; IV – Polinômio de 6º ordem com uma janela de 9 pontos; V – Polinômio de 1º ordem com uma janela de 9 pontos	51
Figura 14. Espectros de uma mancha de sangue usando o método de recuperação. I-Polinômio de 1º ordem com uma janela de 101 pontos; II – Polinômio de 1º ordem com uma janela de 51 pontos; III- Espectro sem alisamento; IV – Polinômio de 6º ordem com uma janela de 9 pontos; V – Polinômio de 1º ordem com uma janela de 9 pontos	51
Figura 15. Espectro Raman pré-processado de uma mancha de sangue usando o laser de 785 nm.....	52
Figura 16. Diagrama representativo da oxidação e redução do ferro presente na hemoglobina	54
Figura 17. Espectros de manchas de sangue, A (3 horas) e B (0 horas) usando o laser de 785 nm	55
Figura 18. Espectro recuperado das manchas de sangue - A (3 horas) e B (0 horas) e espectro obtido usando o laser 785 nm diretamente na mancha C (0 horas).....	56
Figura 19. Número de componentes principais (PCs) em função da variância explicada	57
Figura 20. Gráfico de escores de PC1xPC2 (A), PC1xPC3 (B) e PC2xPC3 (C) além dos escores de PC3 (D) em função das amostras	58
Figura 21. Gráfico de variáveis latentes contra valores de RMSEC e RMSECV	61
Figura 22. Gráfico de T^2 de Hotelling contra os Resíduos Q e gráfico de leverage contra Resíduos em Y	62
Figura 23. Gráfico de Y medido contra Y previsto para o modelo PLS	63

Figura 24. Gráficas de VIP escores para o modelo de regressão	64
Figura 25. Gráfico de Amostras contra os resíduos em Y	64
Figura 26. Espectro Raman da mancha de sangue exposta em condições ambientes e a baixa umidade com 24 horas de exposição.....	65
Figura 27. Espectro da mancha de sangue exposta as condições ambientes e a alta umidade...	66
Figura 28. Representação geométrica do sistema RGB de acordo com ref. 107	69
Figura 29. Representação do sistema de cor HSV de acordo com ref. 108	70
Figura 30. Representação dos sistemas de cores CMYK, HSL, YCbCr, XYZ, L*a*b, YUV, YIQ aseado na ref. 108	70
Figura 31. Produção bibliográfica da área no decorrer dos anos	72
Figura 32. Esquema da quantidade de smatphones, notebook e tablets em milhões no Brasil em 2016 de acordo com ref. 109.....	72
Figura 33. Imagem de satélite da cidade de Brasília no sistema RGB de acordo com ref. 117.	73
Figura 34. Imagem de satélite da cidade de Brasília no sistema CMYK de acordo com ref. 117	73
Figura 35. Imagem de satélite da cidade de Brasília no sistema HSV de acordo com ref. 117.	74
Figura 36. Diferentes substratos usados na construção dos modelos de regressão	77
Figura 37. Esquema da metodologia da coleta da gota de sangue até a aquisição da imagens digitais	77
Figura 38. Sistema de aquisição de imagens digitais	78
Figura 39. Metodologia para obter a região de interesse (ROI) da mancha de sangue no sistema RGB.....	79
Figura 40. Esquema da metodologia para o tratamento das imagens no sistema RGB	81
Figura 41. Imagem RGB mostrando mudanças na coloração de uma mancha de sangue de 0, 1, 2 e 4 horas (esq-dir).....	82
Figura 42. Imagens de uma mancha de sangue com 0, 1, 2 e 4 horas nos sistemas de cores RGB, HSV, CMYK, HSL, L*a*b, YCbCr, YIQ, YUV e XYZ.....	83
Figura 43. Mancha de sangue com 0 horas com (esq.) e sem fundo (dir.).....	84
Figura 44. Histograma do recorte e da mancha completa no sistema RGB	85
Figura 45. Histograma do recorte e da mancha completa no sistema HSV	85
Figura 46. Histograma do recorte e da mancha completa no sistema CMYK	85
Figura 47. Histograma do recorte da imagem no sistema CMYK para 0, 1, 4 horas.....	86
Figura 48. Histograma do recorte da imagem no sistema HSV para 0, 1, 4 horas.....	87
Figura 49. Histograma do recorte da imagem no sistema RGB para 0, 1, 4 hora.....	87
Figura 50. Valores médios de luminosidade para o sistema RGB em ordem crescente de horas	88
Figura 51. Valores médios de luminosidade para o sistema CMYK em ordem crescente de horas	88
Figura 52. Valores médios de luminosidade para o sistema HSV em ordem crescente de horas	89
Figura 53. Número de componentes principais (PCs) em função da variância explicada	91
Figura 54. Escores de PC1 em função das amostras (esq.) e Gráfico de escores PC1 contra PC2 (dir.).....	91
Figura 55. Gráfico de variáveis latentes contra valores de RMSEC e RMSECV	93
Figura 56. Gráfico de T ² de Hotelling contra os Resíduos Q e gráfico de leverage contra Resíduos em Y nas 1 ^o e 2 ^o retirada de outlier	94

Figura 57. Gráfico de T^2 de Hotelling contra os Resíduos Q e gráfico de leverage contra Resíduos em Y para o modelo final	95
Figura 58. Gráfico dos valores y medidos contra os y preditos	95
Figura 59. Resíduos em Y para as amostras do conjunto de calibração e de validação	96
Figura 60. Gráfico de VIP escores para o modelo PLS construído com 4 VLs	96
Figura 61. Variação do canal H em função do tempo de exposição de 0, 1 e 4 horas	98
Figura 62. Mancha de sangue com cerca de 24 horas nos nove sistemas de cores para um doador em duas condições ambientes diferentes: alta e baixa umidade e baixa luminosidade... ..	99
Figura 63. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato Ardósia	100
Figura 64. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato Madeira	100
Figura 65. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato piso	100
Figura 66. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 1	100
Figura 67. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 2	100
Figura 68. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 3	100
Figura 69. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 4	101
Figura 70. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 1	101
Figura 71. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 1	101
Figura 72. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato piso 2	101
Figura 73. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato piso 3	101
Figura 74. Gráfico de LVs contra valores de RMSEC e RMSECV	102
Figura 75. Gráfico de T^2 de Hotelling contra os Resíduos Q e gráfico de leverage contra Resíduos em Y nas 1° e 2° retirada de outlier	103
Figura 76. Gráfico de valores medidos contra preditos	104
Figura 77. Resíduos em Y para as amostras do conjunto de calibração e de validação empregando diferentes substratos	104
Figura 78. Gráfico de VIP escores para o novo modelo PLS incorporando diferentes substratos	105
Figura 79. Manchas de sangue em diferentes substratos com seus respectivos erros relativos	105
Figura 80. Manchas de sangue em diferentes substratos com seus respectivos erros relativos preditos pelo modelo	105
Figura 81. Histograma da ROI e da mancha sem recortes no sistema YUV	126
Figura 82. Histograma da ROI e da mancha sem recortes no sistema YIQ	127
Figura 83. Histograma da ROI e da mancha sem recortes no sistema YCbCr	128
Figura 84. Histograma da ROI e da mancha sem recortes no sistema L^*a^*b	129
Figura 85. Histograma da ROI e da mancha sem recortes no sistema HSL	130
Figura 86. Histograma da ROI e da mancha sem recortes no sistema XYZ	131

Figura 87. Histograma da ROI no sistema YCbCr para 0, 1, 4 horas	132
Figura 88. Histograma da ROI no sistema YIQ para 0, 1, 4 horas.....	133
Figura 89. Histograma da ROI no sistema YUV para 0, 1, 4 horas	134
Figura 90. Histograma da ROI no sistema L*a*b para 0, 1, 4 horas	135
Figura 91. Histograma da ROI no sistema XYZ para 0, 1, 4 horas	136
Figura 92. Histograma da ROI no sistema HSL para 0, 1, 4 horas	137
Figura 93. Valores médios de luminosidade para o sistema L*a*b em ordem crescente de horas	138
Figura 94. Valores médios de luminosidade para o sistema HSL em ordem crescente de horas	139
Figura 95. Valores médios de luminosidade para o sistema XYZ em ordem crescente de horas	140
Figura 96. Valores médios de luminosidade para o sistema YCbCr em ordem crescente de horas	141
Figura 97. Valores médios de luminosidade para o sistema YIQ em ordem crescente de horas	142
Figura 98. Valores médios de luminosidade para o sistema YUV em ordem crescente de horas	143

LISTA DE TABELAS

Tabela 1. Componentes majoritários e suas respectivas modos vibracionais de acordo com ref. 28,91,92,95-103	53
Tabela 2. Modelos multivariados obtidos usando os valores do vetor y em horas e $\log [(horas)]$	60
Tabela 3. Valores para variância explicada na matriz X , vetor y e os parâmetros do modelo PLS	63
Tabela 4. Valores máximos e mínimos de luminosidade dos canais para os seguintes sistemas de cores: RGB, CMYK, HSV, HSL, L^*a^*b e XYZ.....	90
Tabela 5. Parâmetros dos modelos multivariados calculados usando o espectro bruto, o espectro processado e os valores em Y , em horas e $\log(horas)$	92
Tabela 6. Parâmetros dos modelos multivariado usando os valores médios de luminosidade na matriz X e no vetor y , os valores em horas e $\log(horas)$, minutos e $\log(minutos)$	102

LISTA DE ABREVIATURAS

HPLC	Cromatografia Líquida de Alta Eficiência (do inglês, <i>High Performance Liquid Chromatography</i>)
RNA	Ácido Ribonucleico (do inglês, <i>Ribonucleic Acid</i>)
oxi-Hb	Hemoglobina oxigenada (do inglês, <i>Oxyhemoglobin</i>)
meta-Hb	Hemoglobina desoxigenada (do inglês, <i>Desoxyhemoglobin</i>)
Hc	Hemicromo (do inglês, <i>Hemichrome</i>)
NIR	Espectros copia na região do infravermelho próximo (do inglês, <i>Near Infrared Spectroscopy</i>)
LDA	Análise discriminante linear (do inglês, <i>Linear Discriminant Analysis</i>)
SVM-DA	Análise discriminante por máquina de vetores de suporte, (do inglês, <i>Discriminant Analysis Support vector machines</i>)
SIMCA	Modelagem independente de classe (do inglês, <i>Soft independente Modeling of Class Analogy</i>)
PCA	Análise de Componentes Principais (do inglês, <i>Principal Componente Analysis</i>)
PLS-DA	Análise Discriminante pelo método dos mínimos quadrados parciais (do inglês, <i>Discriminant Analysis - Partial Least Squares</i>)
PLS	Regressão pelo método dos mínimos quadrados parciais (do inglês, <i>Partial Least Squares</i>)
PCR	Regressão pelo método das componentes principais (do inglês, <i>Principal Component Regression</i>)
PCs	Componentes principais (do inglês, <i>Principal Component</i>)
VIP	Importância das variáveis na projeção (do inglês, <i>Variable Importance in Projection</i>)
MLR	Regressão linear múltipla (do inglês, <i>Multiple Linear Regression</i>)
RMSEC	Raiz quadrada do erro médio quadrático da calibração (do inglês, <i>Root mean square error of calibration</i>)
RMSECV	Raiz quadrada do erro médio quadrático da validação cruzada (do inglês, <i>Root mean square error of cross-validation</i>)
RMSEP	Raiz quadrada do erro médio quadrático da predição (do inglês, <i>Root mean square error of prediction</i>)
LVs	Variáveis Latentes (do inglês, <i>Latent variable</i>)
CCD	Dispositivo de carga acoplada (do inglês, <i>Charge coupled deviser</i>)
SERS	Efeito Raman intensificado por superfície (do inglês, <i>Surface-enhanced Raman spectroscopy</i>)
COEP	Comitê de Ética em Pesquisa

OSC	Correção de sinal ortogonal (do inglês, <i>Ortogonal sign correction</i>)
AFM	Microscopia de força atômica (do inglês, <i>Atômic force microscopy</i>)
ROI	Região de interesse (do inglês, <i>Region Of Interest</i>)
WLS	Mínimos quadrados ponderados automáticos (do inglês, <i>Automatic Weighted Least Squares</i>)
MSC	Correção do espalhamento multiplicativo (do inglês, <i>Multiplicative Scatter Correction</i>)
SNV	Varição normal padrão (do inglês, <i>Standard Normal Variate</i>)
SVM	Máquinas de vetores de suporte (do inglês, <i>Support vector machines</i>)
LED	Diodo emissor de luz (do inglês, <i>Light Emitting Diode</i>)

SUMÁRIO

CAPÍTULO 1.....	19
1. INTRODUÇÃO	20
1.1 Química forense	21
1.1.1 Manchas de sangue em locais de crime.....	23
1.2 Quimiometria	28
1.2.1 Análise de componentes principais (PCA).....	30
1.2.2 Regressão por mínimos quadrados parciais (PLS).....	32
1.2.3 Pré-processamento dos dados.....	37
CAPÍTULO 2.....	40
2. ESPECTROSCOPIA RAMAN E FERRAMENTAS QUIMIOMÉTRICAS PARA ESTIMAR O TEMPO DE EXPOSIÇÃO DE MANCHAS DE SANGUE	40
2.1 Espectroscopia Raman	41
2.2 Objetivos	44
2.2.1 Objetivos gerais.....	44
2.2.2 Objetivos específicos.....	44
2.3 Parte Experimental	45
2.3.1 Obtenção das amostras	45
2.3.2 Espectros Raman	46
2.3.3 Tratamento dos dados.....	47
2.4 Resultados e Discussão	49
2.4.1 Pré-processamento dos espectros Raman.....	49
2.4.2 Caracterização do espectro Raman das manchas de sangue.....	52
2.4.3 Construção do modelo de calibração.....	56
2.5 Conclusão.....	67
CAPÍTULO 3.....	68
3. ESTIMATIVA DO TEMPO DE EXPOSIÇÃO DE MANCHAS DE SANGUE EMPREGANDO IMAGENS DIGITAIS: UMA ABORDAGEM MULTIVARIADA ...	68
3.1 Introdução	69
3.2 Objetivos	75
3.2.1 Objetivo geral.....	75
3.2.2 Objetivos específicos.....	76
3.3 Parte Experimental	76

3.3.1 Obtenção das amostras	76
3.3.2 Obtenção das imagens digitais	78
3.3.3 Tratamento dos dados.....	79
3.3.3.1 Determinação da região de interesse (ROI) nas imagens digitais	79
3.3.3.2 Construção do modelo de regressão.....	81
3.4 Resultados e Discussão	82
3.5 Conclusão.....	107
CAPÍTULO 4.....	108
4. CONCLUSÃO GERAL	108
REFERÊNCIAS BIBLIOGRÁFICAS	109
APÊNDICE.....	119
A.1 Linhas de comando para os 9 sistemas de cores usados para o tratamento das imagens digitais: ROI e imagem sem recortes	119
A.2 Comparação entre os histogramas da ROI com os da imagem da mancha sem recortes para os sistemas de cores YUV, YIQ, YCbCr, HSL, L*a*b e XYZ.....	126
A.3 Histogramas de uma mancha de sangue para os tempos de exposição de 0, 1 e 4 horas para os sistemas de cores YUV, YIQ, YCbCr, HSL, L*a*b e XYZ.....	132
A.4 Valores médios de luminosidade em função do tempo de exposição para os sistemas de cores YUV, YIQ, YCbCr, HSL, L*a*b e XYZ	138
A.5 Termo de Consentimento Livre e Esclarecido apresentado para os participantes da pesquisa	144

CAPÍTULO 1

1. INTRODUÇÃO

Os vestígios encontrados em locais de um crime constituem uma importante ferramenta para a elucidação do fato investigado. São considerados vestígios quaisquer marca, objeto, ou sinal que possa ter relação com o ocorrido.¹ Dentre os vestígios biológicos que podem ser encontrados em locais de crime, os mais comuns são sangue, cabelo, pelos, saliva, urina e sêmen.^{1,2} A coleta e o armazenamento adequado dessas provas têm grande importância para o decorrer das investigações, pois são elas que irão fornecer as informações sobre o delito por meio de uma análise, seja ela química ou física.^{1,3}

As manchas de sangue são as evidências biológicas mais recorrentes em um local de crime. Por isso, sua identificação e a posterior previsão do seu tempo de deposição tem grande valor no processo investigativo^{2,3}, pois com essas informações, é possível estimar o momento da ocorrência do delito, associar ou não as manchas com o evento criminoso e também correlacionar a origem das provas, se são de um único evento ou de eventos múltiplos.⁴

Atualmente, a identificação das manchas de sangue no local do delito é feita utilizando reagentes como a fenoltaleína e a tetrametilbenzidina que mudam de cor em contato com a peroxidase ou hemoglobina do sangue.^{5,6} Para a identificação das manchas invisíveis é utilizado o reagente luminol. O uso dessas metodologias pode levar a perda da amostra inviabilizando análises futuras, e não traz nenhuma informação sobre o tempo em que o fluido está no local. Além disso, não existe na literatura nenhum método de referência para estimar a idade das manchas de sangue.

A partir disso, é interessante o desenvolvimento de metodologias analíticas que minimizem a perda de amostra e que forneçam respostas rápidas. No caso de manchas de sangue, várias metodologias têm sido propostas na literatura para a determinação da sua idade. Dentre as técnicas, pode-se citar a cromatografia líquida de alta eficiência⁷ (HPLC), microscopia de força atômica (AFM)⁸, análise de RNA⁹, medidas de reflectância na região do visível e infravermelho próximo.^{10,11}

No entanto, a maioria dessas metodologias não são aplicadas na rotina dos laboratórios forenses, principalmente no Brasil. Os motivos são falta de validação dos métodos propostos e as laboriosas etapas de preparo de amostra durante a análise e que dificultam a implementação de tal metodologia em análises de rotina.⁴

A espectroscopia Raman vem ganhando destaque na área forense, pois atende as exigências de ser minimamente destrutível, envolve pouca ou nenhuma etapa de preparo de amostra e permite a máxima obtenção de informações com a realização de uma única análise.⁴ De acordo com o potencial apresentado pela técnica, estudos tem sido feitos aplicando a espectroscopia Raman em conjunto com ferramentas quimiométricas na área forense.^{4,12}

Vários trabalhos vêm sendo relatados na literatura empregando espectroscopia Raman e imagens digitais na análise de manchas de sangue.¹²⁻¹⁹ O uso em conjunto da espectroscopia Raman e da Quimiometria permitiu o estudo de uma grande quantidade/variabilidade de dados, sendo possível correlacioná-los a fatores como idade da mancha e condições ambientais de exposição através do agrupamento das amostras com mesmas características.

Recentemente, o uso de imagens digitais como método analítico têm se destacado no meio científico, uma vez que os métodos de captura de imagem se tornaram uma alternativa viável frente a outras técnicas analíticas já estabelecidas no cenário científico.⁷⁻¹¹ No entanto, no caso de imagens digitais, observa-se que o tratamento dos dados é feito de forma univariada.^{18,19} Dentre as vantagens desta metodologia, pode-se destacar o baixo custo, a simplicidade, a rapidez e a possibilidade de realização de uma análise *in situ*.^{18,19} Além disso, o uso de ferramentas quimiométricas permitiria uma melhor exploração das imagens para a construção de modelos de regressão (análise quantitativa) e modelos de classificação (análise qualitativa).

A proposta desse estudo foi a construção de modelos de regressão para estimar o tempo de exposição da mancha de sangue para fins forense. Para isso a espectroscopia Raman e imagens digitais foram empregadas juntamente com ferramentas quimiométricas para a obtenção dos dados e construção dos modelos. A influência das condições ambientes (umidade e luminosidade) na previsão da idade das manchas de sangue também foi avaliada.

1.1 Química forense

Há muita controvérsia na literatura sobre o início da investigação forense. De maneira geral, pode-se dizer que a investigação científica aplicada a questões legais

começou por meio de tópicos relacionados a medicina, uma vez que existe pouca informação disponível sobre o emprego da química como prova pericial.²⁰

Na Grécia antiga, cerca de 500 a.C., já haviam estudos envolvendo extratos de plantas venenosas e sua posterior utilização como veneno, por exemplo, no caso da cicuta, conhecida como veneno de Sócrates. Além disso, existem escritos egípcios e indianos que também fazem referência ao uso de venenos e antídotos. Na literatura romana também há relatos de veneno de origem vegetal para fins suicidas e homicidas.^{20,21,22}

Em 1248, o livro chinês, *Hsi Yüan Lu (Instructions to Coroners)*²³ foi publicado, e o mesmo descrevia um método para distinguir a morte de uma pessoa por estrangulamento da morte por afogamento. O livro em questão é considerado como o primeiro documento em que um princípio químico foi usado.

Em Tulle, na França, no ano de 1039 foi feito o primeiro relato de um perito, o toxicologista Mathieu Orfila. O testemunho estava associado a um caso de envenenamento, onde a principal suspeita era a esposa da vítima, Marie LaFrange, que foi acusada de matar o marido usando um bolo envenenado com arsênio.

Orfila, por meio do teste de Marsh, avaliou os restos mortais da vítima e também amostras de solo próximas e distantes do túmulo. Não foram detectadas quantidades significativas de arsênio no solo. Logo, os resultados indicaram que o arsênio encontrado provinha do corpo da vítima, corroborando com a hipótese de envenenamento, e assim a suspeita foi condenada.^{20,24}

Segundo Zarzuela (1995)²⁵, a Química Forense pode ser definida como o ramo da química que se ocupa da investigação forense no campo da química especializada, a fim de atender aspectos de interesse judiciário. Tal ramo atende basicamente as áreas de estudos da criminalística e da medicina forense. Em termos penais, é a área da criminalística que se encarrega da análise, classificação e identificação dos elementos ou substâncias encontradas nos locais de ocorrências de um delito ou que podem estar relacionadas a este.

A Química Forense, como área da perícia criminal, se encarrega da análise, classificação e identificação de elementos ou substâncias encontradas nos locais de ocorrência de um delito ou que podem estar relacionadas a este. Com base nisso, as principais aplicações da química forense podem ser sintetizadas em: análise de drogas de abuso, análise de adulteração em numerações de veículos e armas, análise de resíduos de disparos de arma de fogo, investigação de fraude em medicamentos, bebidas e combustíveis e análise de vestígios em cenas de crime.^{20,25,26}

Atualmente, os estudos na área forense se baseiam principalmente no desenvolvimento de metodologias analíticas e na otimização das já existentes. A tendência dos métodos analíticos é que possam ser realizados *in-situ* e forneçam respostas rapidamente.^{6-12,18,19}

1.1.1 Manchas de sangue em locais de crime

O estudo de manchas de sangue em locais de crime pode trazer muitas informações para a investigação criminal, por exemplo, a origem do delito, como e quando o crime ocorreu.² De forma geral, o perfil da mancha de sangue pode ser avaliado para identificar e interpretar como a mancha de sangue foi gerada. No entanto, neste trabalho estamos interessados em determinar o tempo de exposição da mancha de sangue em contato com o ambiente.

O sangue é um fluido corporal que circula no aparelho cardiovascular e é encontrado em todo o organismo, com exceção das unhas e da epiderme. Sua cor varia do vermelho vivo (sangue arterial) ao vermelho escuro (sangue venoso). O pH do sangue é em torno de 7,5 e é mantido por um sistema tampão que inclui principalmente o íon hidrogenocarbonato (HCO_3^-). Ao ser observado no microscópio, mostra ser uma mistura heterogênea, composta de componentes sólidos, como glóbulos brancos, vermelhos e plaquetas e um componente líquido, o plasma.^{2,3}

A hemoglobina, principal cromóforo do sangue, é composta por quatro cadeias polipeptídicas organizadas em dois pares idênticos de subunidades, α e β , que podem ser diferenciadas, uma vez que cada uma contém 141 e 146 resíduos de aminoácidos, respectivamente (Figura 1). Dentro de cada uma destas cadeias está localizada uma molécula heme, composta por uma protoporfirina e um átomo de ferro, que pode apresentar dois diferentes estados de oxidação, Fe^{+2} e Fe^{+3} .^{27,28}

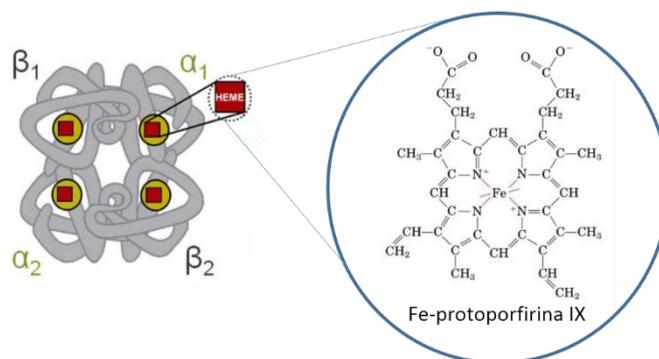


Figura 1. Esquema da estrutura da hemoglobina de acordo com a ref. 28

Em contato com o ambiente, as manchas de sangue podem sofrer alterações tanto físicas quanto químicas. No aspecto físico, ocorre uma alteração na coloração da mancha, variando de um vermelho intenso para um marrom escuro. Do ponto de vista químico, as mudanças observadas podem ser relacionadas com as transformações sofridas pela hemoglobina.^{2,28}

Inicialmente, o ferro (Fe) na hemoglobina possui dois sítios de coordenação livres onde pequenas moléculas podem se ligar, e isto pode dar origem a diferentes moléculas derivadas da hemoglobina.²⁹ No início do envelhecimento da mancha, um dos dois sítios livres pode ser ocupado por resíduos de histidina proximal (His) na posição F8 da cadeia polipeptídica.

Após a metaloproteína estar na presença de oxigênio, forma-se a hemoglobina oxigenada com a molécula de oxigênio ligada reversivelmente. No caso da hemoglobina desoxigenada, o ferro com carga +2 (Fe^{+2}) é colocado aproximadamente a $0,4\text{\AA}$ fora do plano da porfirina, enquanto que na forma oxigenada, a densidade eletrônica é parcialmente transferida do Fe^{+2} para o oxigênio.^{28,29}

A partir disso, ocorrem rearranjos das cadeias polipeptídicas nas reações mencionadas ocasionando mudanças conformacionais da hemoglobina. A consequência direta dessas alterações é que na sua forma oxidada (Fe^{+3}), a mesma é incapaz de se ligar a uma molécula de oxigênio.^{2,3} No entanto, as reações químicas envolvendo a hemoglobina seguem um caminho diferente dentro do organismo, uma vez que existe a presença de uma enzima cuja função é reduzir a forma oxidada, a *citocromo-B5 redutase*.^{2,3} Como a enzima não existe no meio exterior, após a perda da água, o processo de desnaturação segue com a formação da espécie oxidada (meta-Hb) em detrimento da espécie reduzida (oxi-Hb).

Na Figura 2 é apresentado um esquema que exemplifica as mudanças sofridas pela hemoglobina no meio exterior e as espécies formadas em função do tempo de exposição.

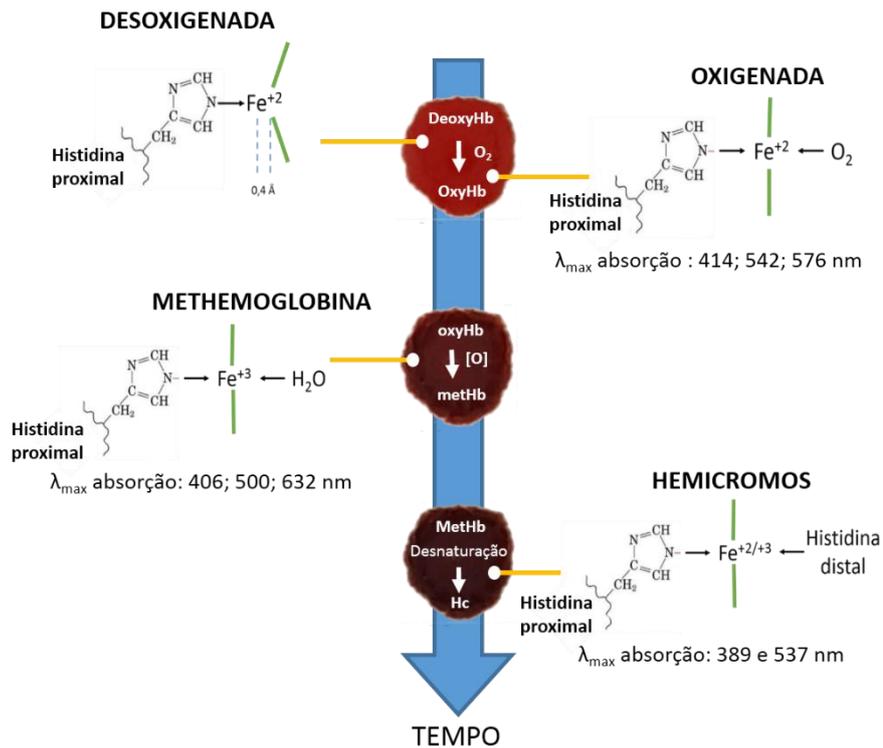


Figura 2. Alterações químicas da mancha em função do tempo de exposição baseada na ref. 28

Devido as espécies formadas apresentarem estruturas ligeiramente diferentes, as técnicas de espectroscopia vibracional podem ser empregadas para a caracterização e previsão da idade das manchas de sangue ao longo do tempo de exposição.^{6,10-12}

Dentre os vários parâmetros que podem interferir no processo de envelhecimento da mancha, pode-se destacar a temperatura, a umidade e a exposição da mancha a diferentes condições de luminosidade (Figura 3).

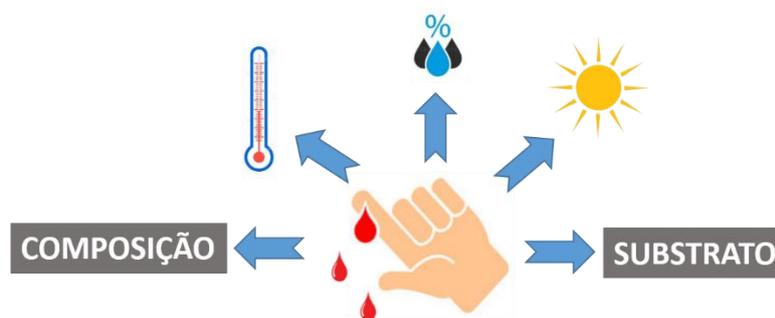


Figura 3. Interferentes na análise de manchas de sangue

Com relação a temperatura, estudos realizados demonstraram que o seu aumento acarreta no aumento da velocidade de envelhecimento da mancha de sangue²⁹⁻³⁶. De acordo com Bremmer et al.³⁵, isso é devido a forte dependência das constantes de velocidade das reações envolvendo a oxi-Hb para a meta-Hb e também da meta-Hb para Hc, com a temperatura. É importante notar que tanto Hanson et al.³⁰ quanto Thanakiatkrai

et al.¹⁹ também demonstraram que existe uma redução na velocidade das reações quando a mancha está armazenada a temperaturas baixas (-20°C), uma vez que não foi notada mudanças significativas nas propriedades em função do tempo.

No contexto forense, esta informação pode ser útil, uma vez que o armazenamento das manchas em baixas temperaturas pode ser uma ferramenta valiosa na preservação de amostras antes da análise.³⁷ No caso do desenvolvimento de metodologias analíticas para determinação do tempo, a informação da temperatura média do local encontrada a mancha é muito importante para a construção de modelos de regressão com veracidade.

Além da temperatura, o processo de envelhecimento da mancha também pode sofrer alterações em função da umidade relativa.^{19,30,35} De acordo com os estudos de Bremmer et. al.³⁵, o emprego de medidas de espectroscopia de reflectância revelou que o processo de conversão da meta-Hb em Hc está correlacionado com um aumento na taxa de umidade. Essas observações estão em acordo com os resultados obtidos pelos trabalhos de Tsuruga et al.³⁸ e Colombo et. al.³⁹

Em relação a presença de luz, nos estudos de Miki et. al.³¹, Fujita et al.³³, Thanakiatkrai et al.¹⁹ e Hanson et. al.³⁴ ficou demonstrado que a taxa de decomposição da mancha de sangue aumentou quando a mancha foi submetida à luz solar. Nestes experimentos, as manchas de sangue foram expostas a diferentes fontes de radiação - luz do sol ou luz fluorescente - e comparadas com aquelas armazenadas no escuro. No entanto, os experimentos de Bauer et al.⁴⁰ obtiveram resultados completamente diferentes, já que os traços de sangue expostos à luz solar e aqueles protegidos das fontes de luz natural não apresentaram praticamente nenhuma diferença significativa na taxa de envelhecimento.

Além disso, existem outros fatores que podem influenciar no envelhecimento das manchas de sangue, como as diferenças na composição, incluindo variação intra e interpessoal e a interação entre a superfície do substrato e a mancha de sangue. No caso do substrato, o mesmo também influencia o tamanho, a forma e, mais importante, a espessura da mancha resultante^{41,42}, o que é um fator determinante para o tempo de secagem. Outro fator preponderante, por exemplo, seria deposição em superfícies porosas, o que permitiria a imersão do sangue no material e neste caso, causando mudanças no tempo de secagem da mancha, fato este que influenciaria na veracidade da predição da idade da mancha.^{30,42-45}

Para evitar a introdução de muitas variáveis, a maioria dos estudos relatados na literatura foi realizada com amostras de sangue coletadas de um único doador ou

pequenos grupos com características semelhantes, por exemplo, idade, sexo, raça e condições de saúde. No entanto, mesmo que o conjunto amostral incluísse indivíduos diferentes,^{11,12,18,19,45-49} o número de amostras analisadas é insuficiente para tirar conclusões confiáveis.

Apesar de não existir um método de referência para estimar a idade de uma mancha de sangue, estudos mostram mudanças de compostos químicos das manchas com o tempo de exposição.^{6,12,16,50} Nos estudos de Botonjic-Sehic *et al* (2009)⁶ foi comprovada a existência de uma relação entre a oxi-Hb e a meta-Hb ao longo do tempo. Ao analisar os espectros obtidos na região do infravermelho próximo (NIR), concluiu-se que a quantidade de oxi-Hb em uma mancha de sangue diminuiu com o passar do tempo, enquanto a de meta-Hb aumentou. Uma possível explicação para esses resultados é que em contato com o meio externo o ferro oxidado prefere se ligar a uma molécula de água (H₂O) do que a de oxigênio (O₂).

De acordo com Li *et. al.* (2011)⁵⁰ um método interessante para a determinação da idade de uma mancha de sangue é a análise da hemoglobina utilizando microespectrofotometria na região do visível e do infravermelho próximo (NIR). A parte experimental foi feita utilizando sangue equino e as mudanças observadas nos espectros em relação ao tempo foram claras, embora os espectros passaram por etapas de pré-processamento para reduzir desvios na linha base e também reduzir o ruído experimental. Os autores construíram um modelo de classificação empregando análise discriminante linear (LDA) para classificar o tempo da mancha de sangue e obtiveram uma seletividade maior que 90%.

Outro trabalho relacionado ao tema e envolvendo espectroscopia vibracional é o de Lednev *et al.*¹². Após a análise dos espectros Raman de manchas de sangue, os autores atribuíram as bandas de 377 e 420 cm⁻¹ como os marcadores da meta-Hb e oxi-Hb, respectivamente. Durante o estudo, os espectros com um tempo de exposição menor do que uma hora já apresentavam a banda da forma oxidada (meta-Hb) mais intensa do que da forma reduzida (oxi-Hb), corroborando com a hipótese de que processo de oxidação é rápido.

Outro estudo relevante de manchas de sangue é o de Mistek *et al* (2016)¹⁶ que buscou discriminar duas raças humanas utilizando amostras de sangue seco por espectroscopia Raman. A pesquisa englobou amostras de vinte indivíduos considerando duas raças, caucasianas e afro-americana, com pessoas de sexo e idades diversificadas. O modelo de classificação adotado foi a análise discriminante por máquina de vetores de

suporte (SVM-DA), que mostrou 83% de probabilidade de classificação correta para as duas raças e uma especificidade e sensibilidade de 80%. Segundo os autores, o método apresentou potencial para ser usado em uma cena real de crime alcançando resultados rápidos e confiáveis com a mínima destruição da amostra.

No entanto, os métodos analíticos propostos na literatura não fazem parte da rotina dos laboratórios forenses, principalmente no Brasil. Essa situação é devida, entre outros fatores, a falta de validação dos métodos propostos e/ou as laboriosas etapas de preparo de amostra durante a análise, exceto em situações envolvendo espectroscopia Raman, já que a técnica demanda poucas ou nenhuma etapa de preparo de amostra.

Os estudos de datação de manchas de sangue encontrados na literatura apresentam uma lacuna quanto ao número de amostras, aos efeitos das condições ambientais de exposição e a validação dos métodos propostos.^{6,12,16,50} Além disso, existem algumas doenças que podem interferir na previsão da idade das manchas de sangue, por exemplo, a hemofilia e a metemoglobinemia. A primeira é associada a falta de fibrina no sangue, responsável pela coagulação. A segunda, existe um aumento do conteúdo de meta-Hb graças a um anomalia genética no metabolismo das células vermelhas do sangue, bem como a exposição a vários fármacos oxidantes ou toxinas.^{51,52}

Esse trabalho almeja contribuir com informações sobre a influência das condições ambientais e do tipo de substrato na datação de manchas de sangue para fins de aplicação no campo forense.

1.2 Quimiometria

Com o grande avanço da tecnologia empregada nas técnicas instrumentais e também dos processadores na década de 70, o volume de dados obtidos a partir de uma simples medida aumentou consideravelmente. Diante da necessidade de extrair, processar e interpretar uma grande quantidade de informações, obtidas cada vez mais rápido e de maneira eficiente, foi necessário desenvolver uma nova forma de interpretar e tratar os dados, o que resultou no desenvolvimento de uma disciplina, a Quimiometria.⁵³

Desde o seu surgimento, essa disciplina vem se desenvolvendo rapidamente e seu uso tem se disseminado a muitas outras áreas do conhecimento, tais como a engenharia, biologia, física, ciências ômicas dentre outras.

A Quimiometria pode ser definida como uma disciplina da química voltada para o tratamento, a interpretação e a previsão de dados químicos usando ferramentas matemáticas, estatísticas e computacionais. Além disso, também é usada no planejamento de experimentos e otimização de condições de processos.⁵⁴⁻⁵⁹

Entre os métodos quimiométricos, os mais empregados no tratamento de dados analíticos são os métodos de reconhecimento de padrões e calibração multivariada. Dos métodos de reconhecimento de padrões não supervisionados, a análise por componentes principais (PCA, do inglês *principal component analysis*) é o método mais usado e importante, uma vez que é a base para o entendimento dos outros métodos quimiométricos. O objetivo da PCA é verificar a existência ou não de algum agrupamento natural dentro do conjunto de dados analisados sem utilizar informações previamente conhecidas.^{54,60,61}

Os métodos de reconhecimento de padrões supervisionados, como o SIMCA (do inglês, *soft independent modeling of class analogy*) e o PLS-DA (do inglês, *partial least square discriminant analysis*) são empregados com o objetivo de atribuir amostras desconhecidas a classes previamente definidas. Nestes métodos, informações sobre as classes das amostras são fornecidas ao modelo para treinamento fazendo com que a classificação supervisionada seja mais objetiva, proporcionando uma maneira sistemática de classificar novas amostras.⁵³

Na calibração multivariada, destaca-se a regressão por mínimos quadrados parciais (PLS, *partial least square*).^{62,63} Os métodos de calibração multivariada apresentam algumas vantagens em relação a univariada, como a possibilidade de determinações diretas na ausência de resolução do sinal analítico, na presença de interferentes desde que presentes no conjunto de calibração, e também a possibilidade de quantificações simultâneas a partir do mesmo conjunto de dados, quando houver mais de um analito ou propriedade a serem previstos.⁵³

Os métodos empregados nesse trabalho foram a análise de componentes principais (PCA), com o objetivo de análise exploratória, e a regressão por mínimos quadrados parciais (PLS) para a construção de modelos de calibração multivariada. Esses métodos serão discutidos com mais detalhes nas próximas seções.

1.2.1 Análise de componentes principais (PCA)

A PCA é definida como um método exploratório, não supervisionado, isto é, apenas a natureza dos dados químicos é utilizada para se obter informações sobre o comportamento das amostras. Foi proposta em 1901 por Karl Pearson, mas sua aplicação começou a partir da década de 1930, após a publicação do trabalho de *Hottelling* que formalizou o método e o aplicou para análises multivariadas na área de psicologia.^{61,64-66}

A aplicação desse método permite descobrir e interpretar mais facilmente os padrões existentes nos dados, além de possibilitar a detecção de amostras com comportamento atípico. Neste método, os dados originais são decompostos em outro sistema de coordenadas não correlacionadas, ou seja, ortogonais, através de operações de autovalores (matriz de correlação) e autovetores (matriz de pesos) com finalidade de extrair informações fazendo uma projeção dos dados em um espaço menor, reduzindo assim a dimensionalidade do espaço original, sem afetar as relações entre as amostras e preservando a maior parte da variância original.⁵³

A partir disso, a combinação linear das variáveis originais altamente correlacionadas é realizada para formar novas variáveis que são chamadas de componentes principais (PCs).^{53,58} A Figura 4 apresenta um esquema da representação geométrica de um sistema com três componentes principais (PCs) de um modelo PCA hipotético. Através de uma operação de mudança de bases nesse sistema de coordenadas, é possível projetar os dados das amostras em um plano formado pelas componentes principais (PC1 e PC2). A projeção das amostras nos novos eixos são chamadas de escores e permitem visualizar a relação entre as amostras sem a perda de informações do sistema de coordenadas original.

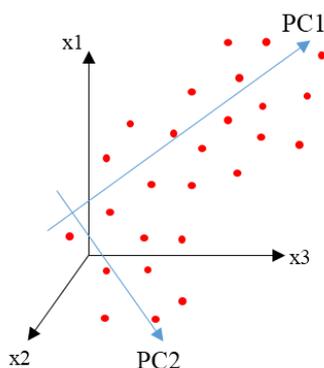


Figura 4. Interpretação geométrica da PCA. (dir.)- Amostras em função das variáveis e (esq.) – Escores de PC1 contra PC2 baseada na ref. 53

A interpretação de como as variáveis originais influenciam na discriminação dos escores, ou seja, das amostras originais, ocorrerá pela análise da influência dos pesos (*loadings*) de cada uma delas no novo sistema de coordenadas, no eixo das componentes principais, e a partir disso é possível identificar os fenômenos ou características responsáveis pelo agrupamento das amostras.

Usando termos matemáticos, a PCA pode ser explicada da seguinte forma: a matriz de dados $\mathbf{X}_{(n,m)}$ formada por n linhas (amostras) e m colunas (variáveis medidas) é decomposta numa matriz de escores $\mathbf{T}_{(n,A)}$, que são as projeções das amostras no novo espaço das componentes principais e em uma matriz de pesos $\mathbf{P}_{(A,m)}$, os pesos são determinados pelo cosseno dos ângulos formados entre as variáveis originais e as novas componentes principais.⁵³

A matriz $\mathbf{E}_{(n,m)}$ é composta pelos resíduos que não foram explicados em nenhuma das PCs selecionadas e irá descrever informações aleatórias (ruídos instrumentais ou irrelevantes), mas que apresenta importância prática para a detecção de amostras anômalas. Essa decomposição matricial é representada pela Figura 5 e pelas equações (1) e (2) apresentadas abaixo:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (1)$$

$$\mathbf{X} = \mathbf{T}_1\mathbf{p}'_1 + \mathbf{T}_2\mathbf{p}'_2 + \dots + \mathbf{T}_A\mathbf{p}'_A + \mathbf{E} \quad (2)$$

sendo A o número de PCs selecionado para o modelo PCA, \mathbf{t}_1 e \mathbf{p}_1 são os vetores de escores e pesos da primeira PC, respectivamente, e \mathbf{E} a matriz de resíduos contendo a variância não descrita pelas PCs incluídas no modelo.

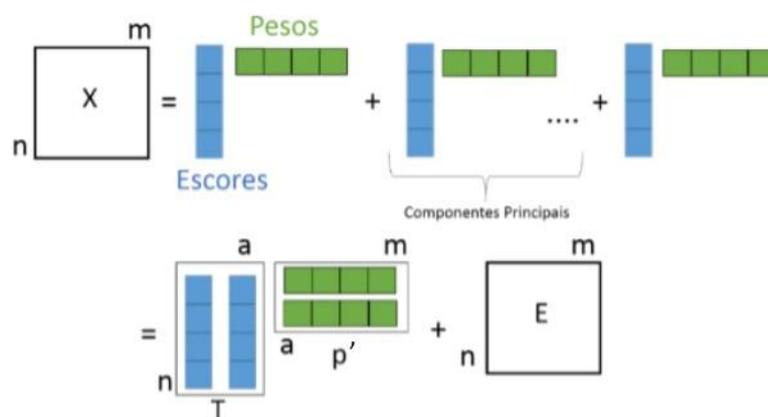


Figura 5. Representação da decomposição da matriz de dados de acordo com o PCA, baseada na referência 53

As PCs são traçadas na direção de maior variabilidade dos dados e selecionar o número adequado de PCs é uma das etapas mais importantes na PCA. O número de PCs significativas deve ser, idealmente, igual ao número de variações linearmente

independentes presentes nos dados. Uma das formas empregadas para a escolha adequada do número de PCs é baseada na quantidade de variância explicada pelo modelo.

Devido a ortogonalidade entre as PCs, é possível interpretar os resultados de forma mais simples, uma vez que as PCs são independentes entre si e a análise dos pesos permite extrair informações a respeito dos fenômenos ou características independentes, que são responsáveis pelo agrupamento das amostras.

Outro ponto importante nos modelos da PCA é a identificação de amostra anômalas (*outlier*) no conjunto de dados. A detecção desse tipo de amostra é feita por meio da avaliação do gráfico de Resíduos Q e os valores de T^2 de *Hotelling*. Os resíduos Q estão associados a variância não modelada pelo modelo e por isso, valores altos indicam que existe muita informação no conjunto amostral que não está sendo levada em consideração pelo modelo. Já os valores de T^2 de *Hotteling* é uma medida da influência de cada amostra no modelo construído. Portanto, as amostras que apresentam um alto valor de resíduo Q e alto valor de T^2 são consideradas anômalas e devem ser retiradas do modelo.

1.2.2 Regressão por mínimos quadrados parciais (PLS)

Os métodos de calibração visam desenvolver expressões matemáticas que relacionem a concentração de um componente ou de outra propriedade de interesse com medidas instrumentais de um sistema químico.⁵³ Na calibração multivariada muitas variáveis são empregadas para quantificar alguma(s) propriedade(s) de interesse. Seu uso é justificado quando não existe uma variável seletiva que permita a aplicação da calibração univariada.

Os métodos de regressão multivariados mais aplicados são a regressão linear múltipla (MLR, *multiple linear regression*), a regressão por componentes principais (PCR, *principal component regression*) e a regressão por mínimo quadrados parciais (PLS, *partial least squares*).⁵³

O método de regressão por mínimos quadrados parciais (PLS) é o método mais utilizado e difundido entre as técnicas de calibração multivariada. O PLS é baseado em uma relação linear entre as variáveis independentes (respostas instrumentais) que formam a matriz \mathbf{X} e a variável dependente ou de interesse, contida no vetor \mathbf{y} .

As linhas da matriz \mathbf{X} correspondem as amostras e as colunas são formadas pelas variáveis, que podem ser, por exemplo, intensidades de um espectro Raman, concentrações de vários compostos ou elementos, e valores de absorvância. No vetor \mathbf{y} , as linhas correspondem as mesmas amostras da matriz \mathbf{X} e a coluna é formada pelos valores da propriedade de interesse. O vetor \mathbf{y} também pode ser uma matriz, embora neste caso ocorra a determinação de mais de uma propriedade de interesse.

No PLS, a PCA é aplicada tanto na matriz \mathbf{X} (resposta instrumental) quanto no vetor \mathbf{y} (propriedade de interesse) com o objetivo de reduzir a dimensão dos dados, de acordo com as equações abaixo:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} \quad (3)$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F} \quad (4)$$

onde \mathbf{T} e \mathbf{U} representam as matrizes de escores de \mathbf{X} e \mathbf{y} , respectivamente; \mathbf{P} e \mathbf{Q} são as matrizes que representam os pesos das variáveis em \mathbf{X} e \mathbf{Y} e \mathbf{E} e \mathbf{F} são os resíduos.

Os escores \mathbf{T} são bons preditores de \mathbf{Y} e assume-se que \mathbf{Y} e \mathbf{X} são modelados pelas mesmas componentes.

$$\mathbf{Y} = \mathbf{TQ}' + \mathbf{F} \quad (5)$$

Os resíduos de \mathbf{Y} , \mathbf{F} , expressam os desvios entre a resposta observada e a modelada. A concentração de novas amostras é estimada a partir dos seus escores, \mathbf{T}^* obtidos pela projeção em \mathbf{P} e dos pesos do modelo, \mathbf{Q} , de acordo com a equação abaixo:

$$\mathbf{Y}_{pred} = \mathbf{TQ}' + \mathbf{F} \quad (6)$$

Dentre as vantagens do método de regressão PLS, pode-se destacar que não é necessário conhecer todos os perfis individuais dos componentes presentes nas amostras e o modelo pode ser construído mesmo na presença de interferentes, embora eles devem estar presentes no conjunto de calibração. A Figura 6 mostra a representação geométrica do PLS, onde uma relação linear é estabelecida entre os escores de \mathbf{X} e \mathbf{y} . Ocorrem mudanças nos escores até que se encontre a melhor relação linear entre eles. Dessa forma, ocorrem leves rotações das componentes principais da matriz \mathbf{X} para aumentar a correlação com \mathbf{y} . As componentes principais são chamadas de variáveis latentes (VLs) no PLS não estão mais na direção de maior variância explicada.

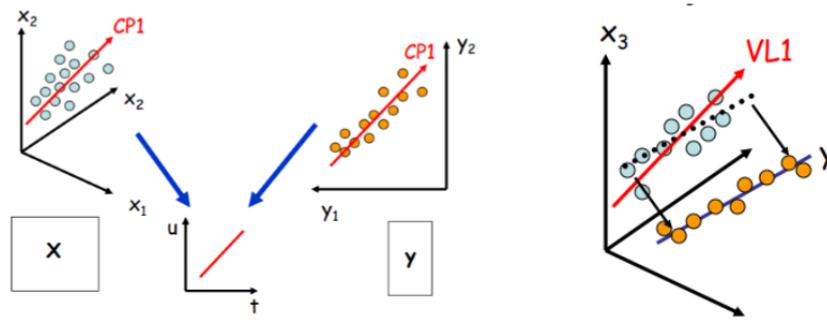


Figura 6. Representação geométrica do PLS, baseada na ref. 53

Duas etapas são importantes durante a construção de um modelo de regressão multivariado: a primeira etapa é a construção do modelo de calibração, ou seja, a determinação dos coeficientes de regressão e a segunda etapa envolve a validação, ou seja, o teste para observar se o modelo construído consegue prever corretamente a propriedade de interesse de amostras que não contidas no conjunto de calibração.⁵³

Para isso, o conjunto de calibração deve conter as amostras mais representativas do conjunto amostral, e isso é obtido fazendo uma seleção adequada das amostras para compor o conjunto de calibração. Geralmente, são empregados algoritmos para selecionar as amostras mais representativas a partir do conjunto total de amostras disponíveis.

O algoritmo Kennard-Stone foi apresentado em 1969 e vem sendo utilizado como a principal abordagem para a escolha das amostras que são utilizadas na construção de modelos de regressão multivariados.^{53,67}

Esse algoritmo utiliza um critério sistemático baseado no cálculo das distâncias euclidianas entre as amostras considerando o hiperespaço definido pelas variáveis originais do conjunto de dados. As duas primeiras amostras selecionadas para o conjunto de calibração serão aquelas que apresentam a maior e a menor distância em relação à origem multivariada dos dados (o ponto médio central). Em seguida, a terceira amostra selecionada será aquela que apresentar a maior distância em relação às amostras já selecionadas.⁵³ Esse processo se repete até que o número de amostras definido, 2/3 do total de amostras, seja selecionado.

Após a seleção das amostras do conjunto de calibração, outro passo importante na construção de um modelo PLS é a determinação correta do número de variáveis latentes. Se for escolhido um número menor que o ideal, informação relevante deixa de ser modelada e o modelo fornece resultados inexatos, nesse caso tem-se um modelo subajustado. Por outro lado, se um número maior que o ideal de VLs é selecionado,

informações redundantes serão incluídas no modelo e este apresentará sobreajuste, sendo adequado apenas para as previsões das amostras usadas na sua construção.⁵³

A escolha do número de variáveis latentes é comumente feita através de validação cruzada, um tipo de validação interna realizada com as amostras do conjunto de calibração. Nessa etapa separa-se uma parte (ou apenas uma) das amostras de calibração e constrói-se o modelo com as demais.^{53,68} Os erros de previsão para as amostras que ficaram de fora da construção do modelo de regressão são estimados empregando-se diferentes números de variáveis latentes e repete-se o processo até que todas as amostras tenham sido previstas.⁵³ O número de variáveis latentes escolhido é o que fornece um menor erro de validação cruzada, RMSECV (do inglês *root mean square error of cross validation*). Este parâmetro é calculado de acordo com a equação abaixo:

$$RMSECV = \sqrt{\frac{\sum(y_r - y_p)^2}{n_c}} \quad (7)$$

onde n_c representa o número de amostras do conjunto de calibração, y_r é o valor de referência contido no vetor y e y_p é o valor previsto pelo modelo.

Os principais métodos de validação cruzada são *leave-one-out*, que remove uma amostra de cada vez, sendo indicado apenas para modelos com pequeno número de amostras (aproximadamente 20), blocos contíguos, que separa amostras em blocos de amostras sequenciais; janelas venezianas (do inglês, *venetian blinds*), que separa amostras sistematicamente espaçadas e subconjuntos aleatórios (do inglês, *Random Subsets*) que separa, aleatoriamente os conjuntos de amostras.⁵³

Após a construção do modelo de regressão, pode-se utilizar uma série de amostras externas (conjunto de validação) para verificar sua capacidade de predição. O desempenho do modelo pode ser verificado pelos parâmetros RMSEP (do inglês, *root mean square error of prediction*) para o conjunto de validação e RMSEC (do inglês, *root mean square error of calibration*) para o conjunto de calibração. Estes parâmetros podem ser calculados de acordo com as equações:

$$RMSEP = \sqrt{\frac{\sum(y_p - y_r)^2}{n_v}} \quad (8)$$

$$RMSEC = \sqrt{\frac{\sum(y_p - y_r)^2}{n_c - v}} \quad (9)$$

onde n_v representa o número de amostras do conjunto de validação, y_r é o valor de referência e y_p é o valor previsto pelo modelo, e v é o número de graus de liberdade perdidos, sendo este número igual ao número de VLs mais um para dados centrados na média. Uma recomendação é que a relação entre RMSEC e RMSEP seja menor que 2,5, uma vez que valores muito maiores de RMSEP indicam sobreajuste do modelo, ou seja, uso excessivo de variáveis latentes.⁵³

A caracterização do modelo também é uma etapa importante, sendo que a análise dos vetores informativos deve ser feita com cuidado, uma vez que eles expressam a importância das variáveis originais no modelo de regressão. Os mais importantes são os vetores dos coeficientes de regressão e os VIP escores (do inglês, *variable importance in projection*).

O cálculo do valor de VIP escores de uma variável J é dado pela equação a seguir^{69,70}

$$VIP_j = \sqrt{d \frac{\sum_{k=1}^h v_k (w_{jk})^2}{\sum_{k=1}^h v_k}} \quad (10)$$

onde d é o número de variáveis do conjunto \mathbf{X} , h é o número de variáveis latentes do modelo, \mathbf{w} é o vetor fator-peso (*weight*) entre o vetor-coluna de \mathbf{X}_j e o vetor \mathbf{y} , e v_k representa a variância de \mathbf{X} calculada a partir dos escores e dos resultados das previsões \mathbf{y} .

Esse vetor é de grande relevância para a interpretação dos modelos e pode ser usado na seleção de variáveis. A média dos quadrados dos valores de VIP escores de todas as variáveis do modelo é igual a um. Por isso, esse valor é utilizado como critério para a “regra do maior que 1”, que aponta as variáveis mais relevantes para a construção do modelo.

Por fim, a detecção de *outlier* também é uma etapa importante na construção dos modelos de calibração multivariada, uma vez que, a presença de amostras anômalas prejudica a capacidade preditiva do modelo. A detecção de *outlier* é feita observando os valores dos resíduos Q e também pelos valores de *leverage* para a matriz \mathbf{X} e para o vetor \mathbf{y} . O *leverage* é uma medida da importância da amostra dentro do modelo, logo quanto maior o *leverage*, maior será o peso desta amostra para o modelo. Quando uma amostra apresenta um alto valor de resíduo e um alto valor de *leverage*, essa amostra é considerada um *outlier* e deve ser retirada do modelo.⁵³

Um outro parâmetro que pode ser usado para avaliar o modelo é o coeficiente de correlação entre o valor experimental (obtidos pelos métodos de referência) da propriedade de interesse e o valor predito pelo modelo (calculado pelo modelo de regressão). Quanto mais próximo de 1 o valor do coeficiente de correlação das amostras do conjunto de calibração, melhor o ajuste. Caso seja o coeficiente de correlação obtido para o conjunto de validação, quanto mais próximo de 1 indica a habilidade de predição do modelo de regressão construído. A equação 11 mostra o cálculo do coeficiente de correção.

$$r = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \bar{y})}{[\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 (y_i - \bar{y})^2]^{1/2}} \quad (11)$$

1.2.3 Pré-processamento dos dados

Antes da aplicação dos métodos quimiométricos faz-se necessário uma etapa de pré-processamento dos dados. Essa etapa tem o objetivo de eliminar ou reduzir a variância aleatória, além de corrigir as fontes de variação sistemática não desejadas. Dessa forma, a extração das informações é direcionada para a variância que realmente interessa.⁶³

Os pré-processamentos podem ser aplicados nas amostras e nas variáveis, ou seja, nas linhas e nas colunas da matriz, respectivamente. Quando aplicado nas linhas, considera-se todas as variáveis e uma amostra de cada vez. De forma análoga, quando aplicado nas colunas, considera todas as amostras e uma variável de cada vez.⁵³

Para os métodos usados nas colunas, isto é, nas variáveis, o mais comum é centrar os dados na média (equação 12). Esse método é escolhido quando as variáveis possuem a mesma natureza e são obtidas da mesma fonte. Nesse método, a média de cada variável é subtraída de seus respectivos elementos, e isso causa uma translação de eixos das variáveis para o valor médio, embora a estrutura dos dados seja preservada.

$$x_{ij (CM)} = x_{ij} - X_j \quad (12)$$

onde $x_{ij (CM)}$, corresponde ao dado centrado na média para a variável j na amostra i ; x_{ij} é o valor da variável j na amostra i e X_j é a média aritmética dos valores de uma determinada variável j , considerando a presença de n amostras.

Além disso, também pode ser usado o escalamento pela variância, onde cada elemento da coluna é dividido pelo desvio padrão da coluna.⁵³

$$x_{ij(EV)} = \frac{x_{ij}}{s_j} \quad (13)$$

onde, $x_{ij(EV)}$ corresponde ao escalamento pela variância da variável j na amostra i ; x_{ij} é o valor da variável j na amostra i e S_j é o desvio padrão associado à variável j .

O autoescalamento são os dois métodos anteriores usados em sequência e é indicado quando se trata de dados discretos com unidades e escalas diferentes. Esse método não é recomendado para dados espectrais, pois é feito a atribuição dos pesos igualmente entre as variáveis, isto é, o sinal do equipamento, o ruído experimental e outros interferentes, o que não é do interesse durante uma análise.

$$x_{ij(AE)} = \frac{x_{ij(CM)}}{s_j} \quad (14)$$

Na Figura 7 está apresentado um esquema dos métodos anteriormente citados que podem ser aplicados nas colunas, ou seja, nas variáveis do conjunto de dados.

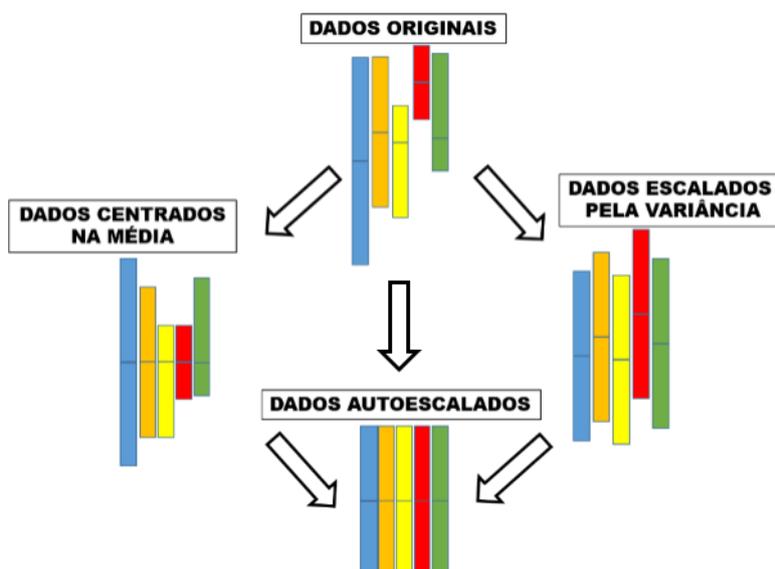


Figura 7. Esquema ilustrativo para o pré-processamento das variáveis baseada na ref. 20

Os métodos geralmente aplicados nas linhas são as derivadas, métodos de normalização, como ‘correção do espalhamento multiplicativo (MSC, do inglês, *multiplicative scatter correction*)⁷¹ e variação normal padrão (SNV, do inglês, *standard normal variate*).⁷² Para reduzir desvios da linha de base, por exemplo, pode ser usado a derivada e também o método dos mínimos quadrados ponderados automáticos^{53,73} (WLS, do inglês, *automatic weighted least squares*). Além disso, a diminuição do ruído experimental pode ser feita empregando o método de alisamento, como *Savitzky-Golay* que usa um polinômio e uma janela com n pontos.

A variância sistemática não desejada de um conjunto de dados pode ser eliminada por métodos de normalização. De forma geral, nesse tipo de pré-processamento os valores

de cada variável de uma determinada amostra i são divididos por um fator de normalização, como por exemplo, a norma da respectiva amostra representada por $\|x_i\|$.

Assim, todas as amostras estarão em uma escala predeterminada, ou seja, com a mesma magnitude. A equação abaixo mostra o cálculo da normalização vetorial aplicado em cada elemento de uma linha da matriz de dados.⁵³

$$x_{ij} (norm) = \frac{x_{ij}}{\|x_i\|}, j = 1, 2, \dots, J. \quad (15)$$

CAPÍTULO 2

2. ESPECTROSCOPIA RAMAN E FERRAMENTAS QUIMIOMÉTRICAS PARA ESTIMAR O TEMPO DE EXPOSIÇÃO DE MANCHAS DE SANGUE

2.1 Espectroscopia Raman

A cena de um crime pode fornecer muitos tipos de evidências e, dentre as amostras biológicas que podem ser coletadas, as manchas de sangue são as mais comuns e a partir delas se pode obter informações importantes, como por exemplo, o momento do delito.^{1,2} As amostras coletadas fazem parte do quebra-cabeças para desvendar o crime e muitas vezes são obtidas em quantidades mínimas e, portanto, é fundamental que a técnica analítica usada pelos profissionais forenses seja minimamente destrutiva, isto é, não altere ou não destrua quantidades significativas das amostras, mas em contrapartida, consiga retirar a maior quantidade de informação possível.⁴

A espectroscopia Raman aplicada a química forense tem se mostrado uma técnica promissora devido sua capacidade na identificação de compostos químicos (inorgânicos e orgânicos), uma vez que fornece informações em nível molecular sobre os compostos presentes em uma amostra e possui pouca ou nenhuma etapa de preparo de amostra. O seu emprego em conjunto com as ferramentas quimiométricas disponíveis tem ampliado o leque de aplicações da técnica, assim como a disponibilidade de equipamentos portáteis, permitindo a realização de medidas diretamente sobre o material ou até mesmo em materiais acondicionados em sacos plásticos ou vidros.^{4,74}

Em um primeiro momento, a técnica de espectroscopia Raman era utilizada apenas para complementar as informações obtidas empregando outras metodologias e técnicas analíticas, como a espectroscopia de absorção na região do infravermelho médio.⁴ No entanto, atualmente essa técnica vêm sendo usada na identificação de substâncias explosivas e pólvora⁷⁵⁻⁸⁰, de corantes e pigmentos para a área de documentoscopia⁸¹⁻⁸³ e obras de arte⁸⁴, de substâncias ilícitas⁸⁵, de fibras têxteis⁸⁶ e para caracterização de fluidos biológicos⁸⁷.

A espectroscopia Raman é uma técnica de espectroscopia vibracional baseada no espalhamento inelástico da radiação eletromagnética após interação com a amostra. Esse fenômeno físico foi descrito pela primeira vez em 1928 pelo físico indiano Sir Chandrasekhara Venkata Raman. No seu experimento, Raman usou a luz do sol como fonte de excitação simulando um laser na região do visível, um espectrômetro de bolso e o olho humano como detector.⁷⁴ O cientista focalizou a luz solar usando várias lentes em um recipiente contendo um líquido purificado e antes da amostra foi colocado um filtro azul para deixar passar apenas a radiação de comprimento de onda mais energética do espectro. Ao observar a amostra em uma direção perpendicular ao feixe de luz incidente,

Raman constatou que ocorreu espalhamento inelástico, pois foi observado um traço luminoso de cor diferente do que foi usado, ou seja, com uma energia diferente da que foi incidida.

Logo, a partir de suas observações e experimentos, Raman comprovou que a luz visível sofria espalhamento inelástico. Com a descoberta do fenômeno, o reconhecimento da comunidade científica foi rápido e por suas contribuições, em 1930, ele foi agraciado com o prêmio Nobel de Física.

A técnica utiliza, como fonte de excitação, uma radiação monocromática na região do visível ou do infravermelho próximo por meio de um laser. A radiação é incidida sobre a amostra, e após interações, fótons são espalhados em todas as direções com energia igual ou diferente da incidida. A maior parte desses fótons são espalhados elasticamente, ou seja, possuem a mesma energia da radiação incidida. Esse tipo de espalhamento não é considerado como fonte de informação a respeito da composição molecular, sendo denominado espalhamento elástico ou Rayleigh.⁷⁴

Em contrapartida, poucos fótons são espalhados com energia diferente da incidida (espalhamento inelástico), e estes são capazes de fornecer informações vibracionais sobre a amostra, pois a diferença de energia entre a radiação incidida e a radiação espalhada é igual a diferença de energia entre dois estados vibracionais da molécula. Dessa forma, os espectros Raman apresentam bandas que são características de uma determinada ligação química e fornecem uma impressão digital/*fingerprint* da amostra permitindo sua identificação.⁴

A obtenção do sinal na espectroscopia Raman está relacionada a variação do momento de dipolo induzido na molécula pelo campo elétrico da radiação⁸⁸, e este campo elétrico pode ser calculado usando a equação abaixo:

$$E = E_0 \cos 2\pi\nu_0 t \quad (16)$$

onde E_0 é a amplitude vibracional e ν_0 é a frequência da radiação incidente

A oscilação do campo elétrico causa na molécula um momento de dipolo oscilante P , cuja frequência é igual ao do campo elétrico externo aplicado.

$$P = \alpha E = \alpha E_0 \cos 2\pi\nu_0 t \quad (17)$$

sendo que α é a constante de proporcionalidade chamada de polarizabilidade, que informa a facilidade de deformação da nuvem eletrônica da ligação química na presença de um campo elétrico.

A ocorrência de vibrações moleculares e movimentos nucleares leva a mudanças ocasionais na polarizabilidade, cujas frequências correspondem aos vários modos

vibracionais de uma molécula. Para uma frequência ν_m , o deslocamento do núcleo para uma molécula diatômica (q) pode ser escrito de acordo com a equação:

$$q = q_0 \cos 2\pi \nu_m t \quad (18)$$

onde ν_m é a frequência vibracional da coordenada interna da ligação, q . A polarizabilidade pode ser considerada como uma função linear de q e que pode ser expandida na seguinte série:

$$\alpha = \alpha_0 + (\delta\alpha/\delta q) * q + \dots \quad (19)$$

onde o primeiro termo é a polarizabilidade na posição de equilíbrio e o segundo termo corresponde a variação na coordenada q em relação a posição de equilíbrio.

O espalhamento elástico ou Rayleigh ocorre quando o fóton emitido interage com a molécula e é espalhado com a mesma energia inicial, ou seja, sem modificação da sua frequência. No espalhamento anti-Stokes, ocorre o espalhamento do fóton com uma energia maior que a incidida e neste caso, o fóton encontra a molécula já num estado excitado e após a interação ocorre o decaimento da molécula para o estado fundamental e a diferença de energia é cedida ao fóton.

Já no espalhamento Stokes, cuja frequência é $\nu_0 - \nu_v$, a molécula no estado fundamental sofre colisão com o fóton de energia $h\nu_0$, passa para um estado intermediário (virtual), pois não corresponde a nenhum estado estacionário da molécula. Em seguida a molécula decai para um estado vibracional excitado de energia e_v e assim o fóton espalhado $h\nu_0 - e_v$, possui energia menor do que o fóton incidente.⁸⁸ Os espalhamentos elástico e inelástico são mostrados esquematicamente na Figura 8.

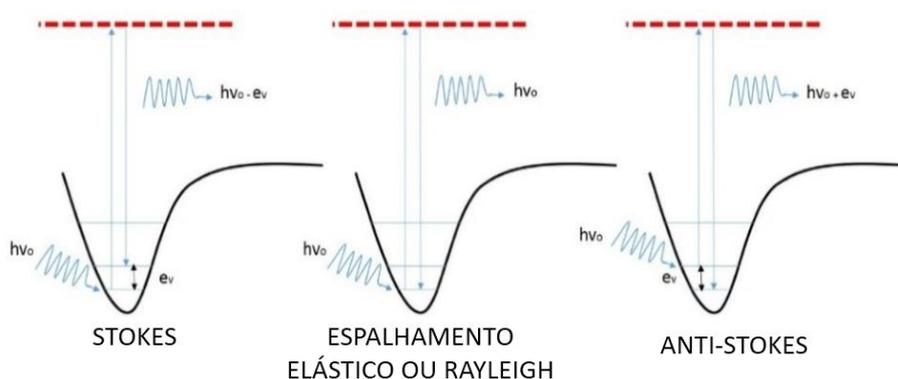


Figura 8. Diagramas de energia envolvidos nos processos de espalhamento Raman e Rayleigh. ($h\nu_0$ energia incidente, e_v é a diferença de energia entre os dois estados vibracionais) baseada na ref. 88

A partir da figura, observa-se que as informações fornecidas pelas bandas Stokes e anti-Stokes são as mesmas, embora suas intensidades sejam diferentes e sua relação

pode ser mostrada pela função de distribuição de Boltzmann, de acordo com a equação a seguir:

$$I_A/I_S = (\nu_o + \nu_v / \nu_o - \nu_v) \exp (-e_v/kT) \quad (20)$$

, onde I_A , intensidade anti-Stokes, I_S , intensidade Stokes, $\nu_o - \nu_v$, frequência do espalhamento anti-Stokes, $\nu_o - \nu_v$, frequência Stokes, T em Kelvin e e_v , energia do estado excitado.

A partir da Figura 8, é possível compreender que as bandas anti-Stokes são oriundas da população do estado excitado e essas estão em um número muito menor que a população do estado fundamental, e é por esse motivo que as bandas anti-Stokes são menos intensas que as bandas Stokes. No entanto, em certos casos, a energia do fóton de excitação se assemelha com a energia de transição eletrônica permitida fazendo com que ao invés de se atingir um estado virtual, ocorra uma transição eletrônica produzindo uma intensificação do sinal Raman.

Esse efeito causa um aumento na sensibilidade do sinal Raman e é chamado de Raman ressonante. Ainda nesse contexto, outro efeito que se observa de intensificação do sinal Raman é o efeito SERS (do inglês, *surface enhanced Raman scattering*).⁸⁹⁻⁹² Nesse caso, são empregadas superfícies metálicas nanoestruturadas para o aumento da sensibilidade dos espectros Raman, que pode chegar a um fator de intensificação de 10^6 vezes.

2.2 Objetivos

2.2.1 Objetivos gerais

O objetivo do estudo foi construir um modelo matemático para estimar o tempo de exposição de manchas de sangue para fins forenses. Para isso, a espectroscopia Raman foi empregada como técnica analítica juntamente com a calibração multivariada (PLS).

2.2.2 Objetivos específicos

- Otimizar o procedimento de recuperação da mancha de sangue;
- Avaliar o melhor pré-processamento dos espectros para a correção da linha de base e na redução do ruído experimental;
- Caracterizar os espectros das manchas de sangue obtidos em diferentes tempos, identificando os principais componentes e modos vibracionais presentes;
- Construção e validação do modelo de regressão por PLS;
- Avaliar a influência das condições ambientais nas principais bandas do espectro Raman relacionadas ao processo de envelhecimento do sangue;
- Avaliar a influência das condições ambientais no modelo PLS.

2.3 Parte Experimental

2.3.1 Obtenção das amostras

O projeto de pesquisa foi submetido ao Comitê de Ética em Pesquisa (COEP) da Universidade Federal de Minas Gerais (UFMG) e aprovado com o seguinte identificador: CAAE – 6691817.5.0000.5149

A parte experimental teve início com a coleta de amostras de sangue de dez doadores, homens e mulheres, com idade média de 29 ± 3 anos. As doações do material biológico foram autorizadas por meio do termo de consentimento livre e esclarecido que está apresentado no apêndice, na seção A.5.

As amostras de sangue foram obtidas utilizando lancetas descartáveis para punção digital e após a coleta, o fluido foi depositado em uma lâmina de microscópio. No momento da obtenção das amostras, foram anotados a data e o horário da coleta, e além disso, as mesmas foram nomeadas usando códigos sem qualquer tipo de relação com o doador. As amostras foram armazenadas dentro de sacos plásticos em condições ambientes até a obtenção dos espectros.

Na aquisição dos espectros, foi usado o procedimento de recuperação, feita pressionando *swabs* umedecidos com água (1 gota) contra a amostra. Um esquema da metodologia, da coleta até a recuperação, está apresentado na Figura 9.

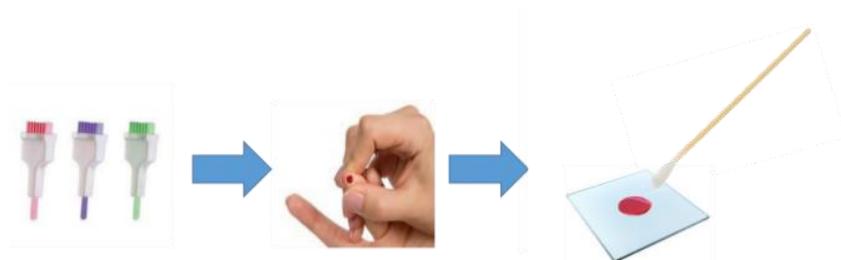


Figura 9. Metodologia de coleta até a recuperação das manchas de sangue

Com base nas informações da coleta (data e hora), foi atribuída a idade das manchas no momento da obtenção dos espectros usando o procedimento de recuperação. A faixa de horas analisada compreendeu de 0 a 369 horas.

Para avaliar os parâmetros de umidade e de luminosidade, algumas manchas de sangue ficaram expostas a condições ambientes controladas por 24 horas com o objetivo de avaliar a influência desses fatores no espectro Raman em função do tempo de exposição. Para os ensaios de umidade, foram empregadas soluções saturadas de cloreto de potássio (86%) e hidróxido de sódio (7%). Nesse estudo, foi coletado sangue de 1 doador, sendo que foram obtidas 4 gotas de sangue. As duas primeiras foram armazenadas em diferentes dessecadores, que continham as soluções para o controle da umidade, 7% (baixa umidade) e 86% (alta umidade).

A terceira mancha foi depositada numa lâmina e esta ficou armazenada dentro de uma caixa sem a presença de luz, embora em contato com o ar. Por fim, a última gota foi depositada numa lâmina que ficou armazenada dentro do saco plástico e este ficou exposto as condições ambientes, temperatura média de 25°C com uma umidade em torno de 40%.

2.3.2 Espectros Raman

Um espectrômetro Senterra Raman (*Bruker Optics*) com um detector CCD, como ilustrado na Figura 10, foi utilizado para todas as medidas. Este equipamento é acoplado a um microscópio óptico Olympus com uma objetiva de 20x. Os espectros foram coletados usando um laser de diodo em 785 nm, com uma potência (nominal) de 10 mW com 30 exposições de 10 s na faixa espectral de 80-2650 cm^{-1} e resolução espectral de 4 cm^{-1} . Esses parâmetros foram otimizados de forma a obter espectros com melhor relação sinal-ruído. O software OPUS (v. 7.2) foi utilizado para obter os espectros.

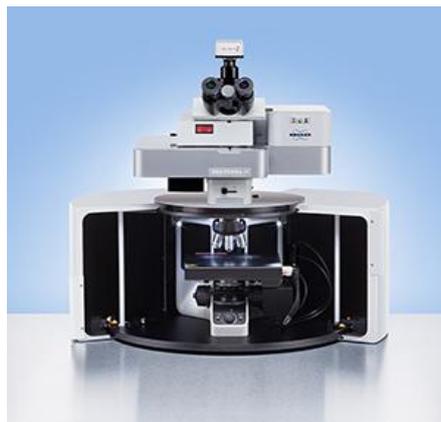


Figura 10. Espectrômetro Senterra Raman (Bruker), detector CCD e microscópio Olympus

2.3.3 Tratamento dos dados

Os espectros obtidos foram exportados do software do equipamento para o ambiente de trabalho *MATLAB* (Mathworks, Natick, Massachusetts, EUA, v. 7.9.0.529). Em seguida, os dados foram tratados usando o pacote *PLS-Toolbox* (*Eigenvector Research*, Manson, Washington, EUA, v. 5.2.2).

A matriz \mathbf{X} foi construída usando as intensidades Raman de 71 espectros na faixa espectral de 180-1800 cm^{-1} . O vetor \mathbf{y} foi construído usando a idade das manchas (tempo após a coleta) em horas e $\log(\text{horas})$. A primeira etapa do tratamento de dados envolveu o pré-processamento dos espectros, os métodos empregados na matriz \mathbf{X} estão listados abaixo:

- (1) Correção da linha de base pelo método WLS (do inglês, *weighted least squares*) usando um polinômio de segunda ordem;
- (2) Normalização por vetor de comprimento unitário;
- (3) Alisamento usando filtro de *Savitzky-Golay*, com polinômio de primeira ordem e janela de nove pontos;
- (4) Centrar os dados na média;

Esses métodos são aplicados com o intuito de reduzir as variações indesejáveis adquiridas na coleta dos dados, como ruído experimental e minimizar os fenômenos que não são do interesse, como por exemplo, a fluorescência. Na Figura 11 estão apresentados

os possíveis interferentes que podem estar presentes em um espectro Raman de uma amostra biológica adquirido em um equipamento dispersivo.

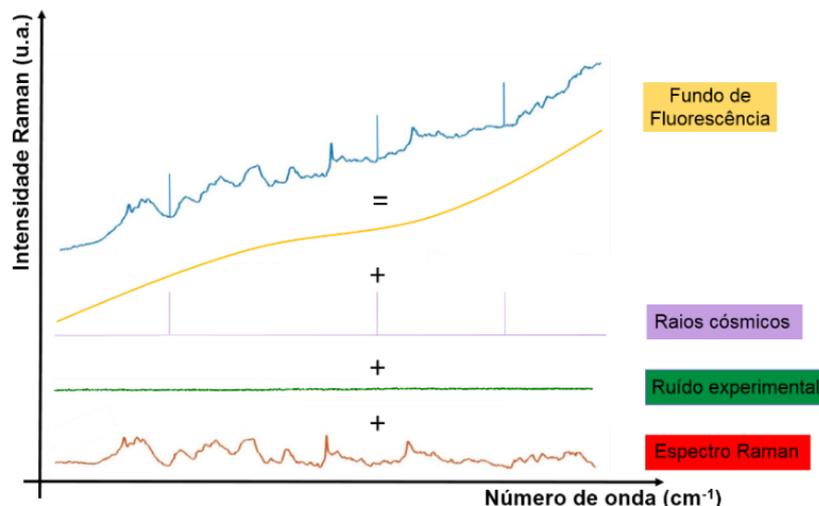


Figura 11. Composição de um espectro Raman de uma mancha de sangue e os seus possíveis interferentes

Além disso, foi avaliado o emprego da correção de sinal ortogonal (OSC do inglês, *orthogonal signal correction*)^{93,94} como pré-processamento dos espectros Raman antes da construção do modelo de calibração. No OSC a variação em \mathbf{X} que não esteja relacionada com o vetor \mathbf{y} é reduzida, dessa forma, o modelo é simplificado. Seu uso é comum quando se quer reduzir o número de VLs no modelo PLS.

O pré-processamento aplicado no vetor \mathbf{y} para os modelos construídos foi centrar os dados na média.

Após a etapa de pré-processamento, a análise exploratória empregando PCA foi realizada com o intuito de observar alguma tendência natural e em seguida, foi usada a regressão pelo método dos mínimos quadrados parciais (PLS) para construir o modelo de regressão para estimar o tempo de exposição das manchas de sangue.

Para construção do modelo de calibração, a matriz \mathbf{X} e vetor \mathbf{y} foram divididos em dois conjuntos usando o algoritmo de Kennard-Stone, conjunto de calibração (70% das amostras) e o conjunto de validação (30% das amostras).

Para a escolha do número de variáveis latentes foi empregada a validação cruzada usando o método *random subsets*. Após a escolha do número de variáveis latentes, o modelo foi validado empregando o conjunto de amostras externas, o conjunto de validação. O desempenho do modelo foi avaliado por meio dos parâmetros: RMSEC, RMSECV, RMSEP e coeficiente de correlação para a validação e para a calibração. A detecção de amostras anômalas também foi realizada.

2.4 Resultados e Discussão

2.4.1 Pré-processamento dos espectros Raman

Após organizar as informações obtidas numa matriz de dados, foi necessário usar métodos de pré-processamento nos espectros para reduzir o ruído experimental e a fluorescência.

Para a correção da fluorescência, o emprego do método WLS (do inglês, *weighted least squares*) é comumente utilizado.^{71,73} O algoritmo usado no WLS ajusta uma linha de base para cada espectro e determina quais variáveis estão acima ou abaixo dela. Após isso, os pontos mais prováveis são selecionados devido apenas à linha de base, e aqueles que estiverem acima são considerados mais significativos na adaptação da linha de base.^{53,71,73}

Normalmente, a linha de base é aproximada por algum polinômio de ordem n ou pode-se fornecer uma linha de base específica ou um conjunto delas. Ao usar um polinômio, deve-se ter cuidado ao selecionar a sua ordem, pois existe o risco de adicionar variância indesejada no conjunto de dados ao selecionar polinômios com um grau mais elevado. Nos espectros obtidos foi usado um polinômio de grau 2. Na Figura 12 é apresentado um espectro de uma mancha de sangue sem nenhum pré-processamento e após fazer a correção da linha de base e também usar a normalização pelo vetor de comprimento unitário.

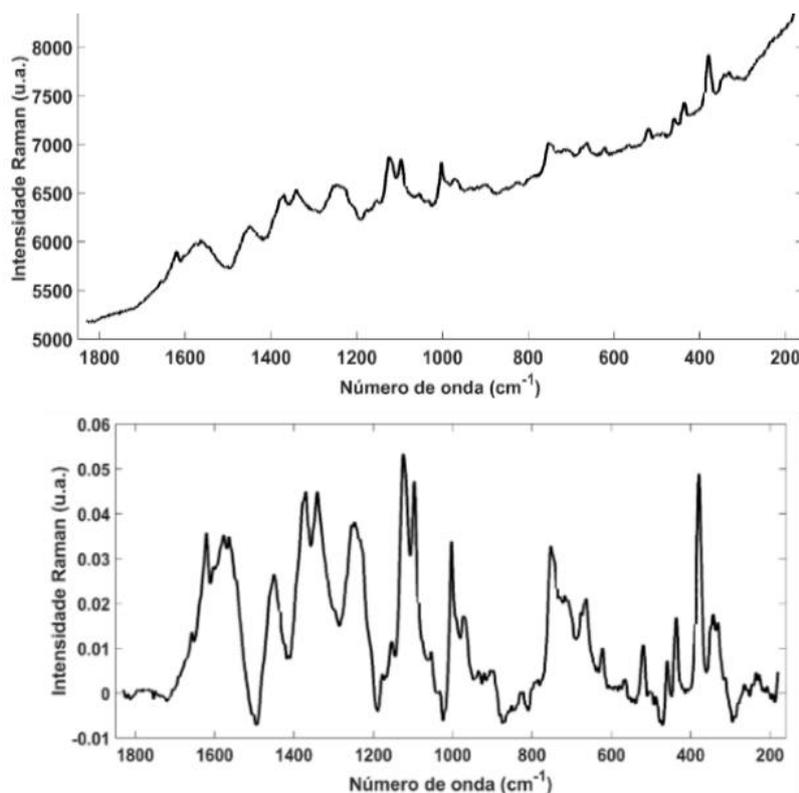


Figura 12. Espectro da mancha de sangue sem (acima) e após (abaixo) o pré-processamento usando mínimos quadrados ponderados – WLS (2º grau) e normalização vetorial

Os métodos de alisamento são empregados em dados espectroscópicos com o objetivo de reduzir o ruído experimental. No entanto, deve-se tomar muito cuidado com o uso de algoritmos de alisamento, em relação a ordem do polinômio e com o tamanho da janela. Na Figura 13 e na Figura 14 é possível observar a influência do tamanho da janela e do grau do polinômio nos espectros obtidos ao incidir diretamente o laser na mancha de sangue e usando o método recuperação, respectivamente.

O cuidado no uso de filtros de alisamento em dados espectrais são de suma importância, uma vez que usar janelas grandes ou polinômios de baixo grau podem causar distorção dos espectros levando a perda de informação, como mostrado nos espectros I e II, nas Figura 13 e Figura 14. Uma boa prática é sempre comparar visualmente os espectro antes e após o alisamento e observar se houve perda de informação.

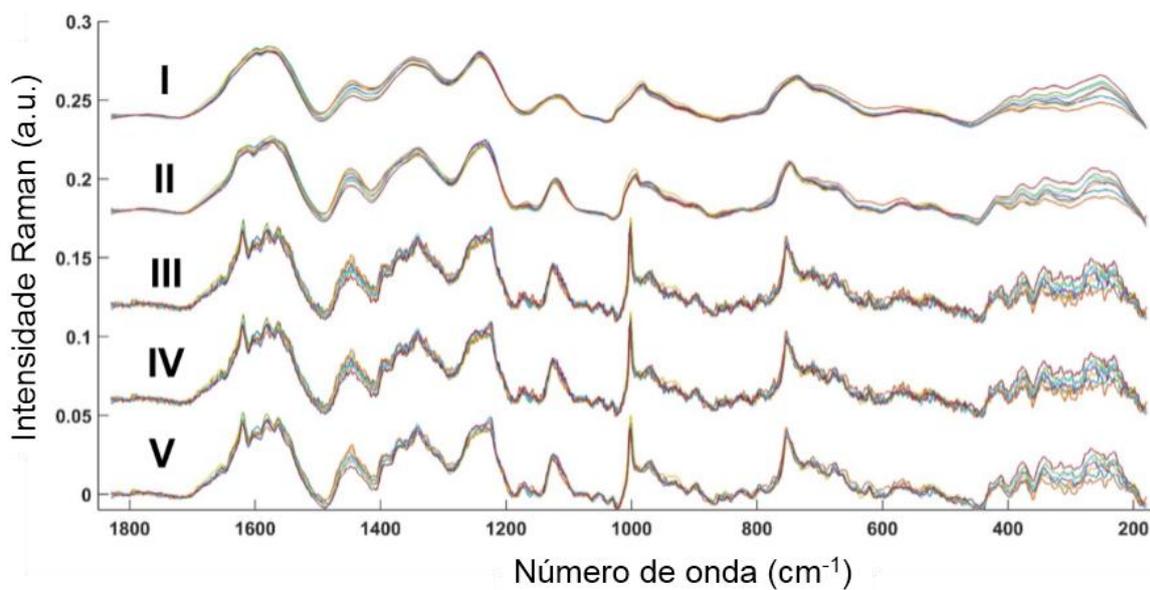


Figura 13. Espectros obtidos incidindo diretamente o laser na mancha. I-Polinômio de 1° ordem com uma janela de 101 pontos; II – Polinômio de 1° ordem com uma janela de 51 pontos; III- Espectro sem alisamento; IV – Polinômio de 6° ordem com uma janela de 9 pontos; V – Polinômio de 1° ordem com uma janela de 9 pontos

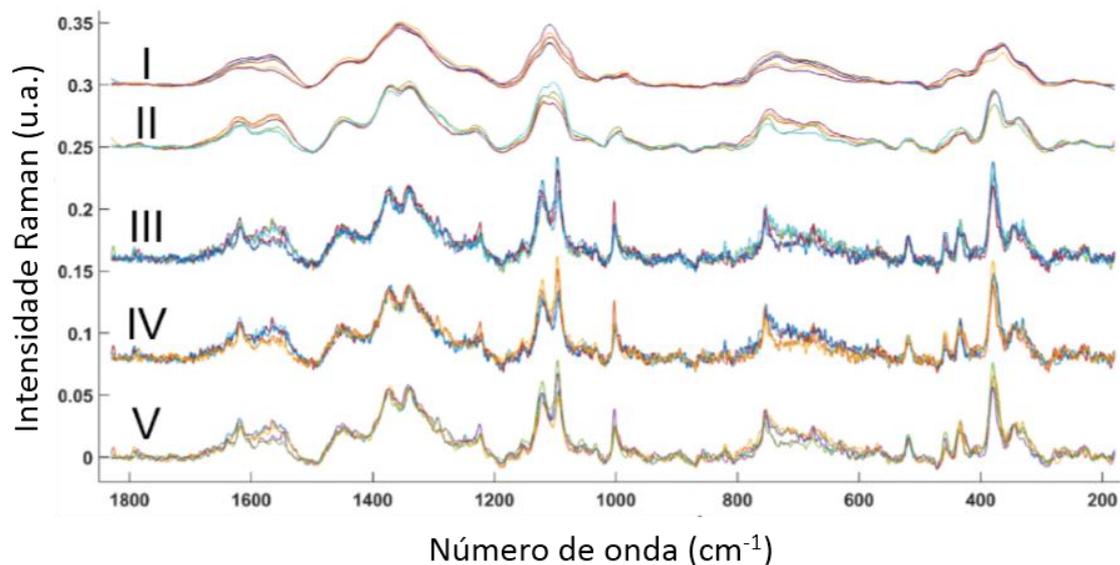


Figura 14. Espectros de uma mancha de sangue usando o método de recuperação. I-Polinômio de 1° ordem com uma janela de 101 pontos; II – Polinômio de 1° ordem com uma janela de 51 pontos; III- Espectro sem alisamento; IV – Polinômio de 6° ordem com uma janela de 9 pontos; V – Polinômio de 1° ordem com uma janela de 9 pontos

Portanto, a partir de uma comparação dos espectros antes e depois do pré-processamento, foi adotado o polinômio de 1° grau com uma janela de 9 pontos (espectro V da Figura 14), pois esta combinação gerou um espectro sem perda de informação química e em contrapartida, com o mínimo de ruído experimental.

2.4.2 Caracterização do espectro Raman das manchas de sangue

Devido ao potencial demonstrado pela espectroscopia Raman para discriminar e identificar amostras biológicas, orgânicas e inorgânicas, seu uso tem sido recorrente na área forense.⁴

A identificação e diferenciação de amostras é feita utilizando as bandas consideradas características de cada composto. A principal vantagem de caracterizar o espectro de um composto é a possibilidade de compará-lo com o espectro de uma amostra desconhecida e assim relacioná-lo com os principais modos vibracionais presentes no espectro de uma amostra de referência e também com a de uma evidência coletada no local do crime.

Os espectros adquiridos usando espectroscopia Raman são como uma impressão digital do composto. A Figura 15 mostra o espectro Raman pré-processado obtido diretamente em uma mancha de sangue fresca. A Tabela 1 mostra a atribuição, baseada na literatura, das principais bandas observadas no espectro aos seus respectivos modos vibracionais.

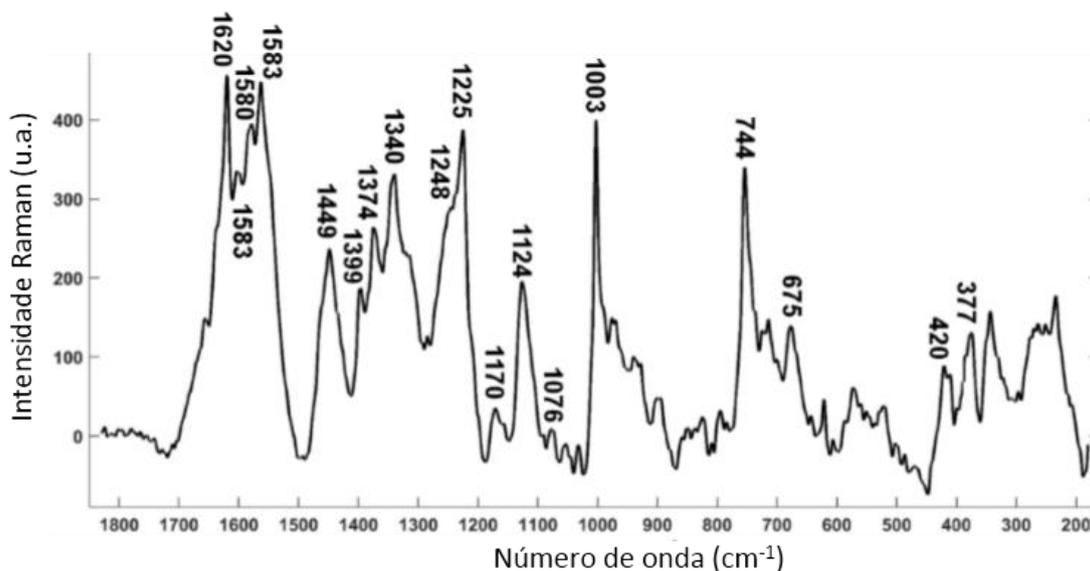


Figura 15. Espectro Raman pré-processado de uma mancha de sangue usando o laser de 785 nm

Tabela 1. Componentes majoritários e suas respectivas modos vibracionais de acordo com ref. 28,91,92,95-103

Número de onda	Modo vibracional	Componente
377	Estiramento $C_{\beta}C_{\alpha}C_{\delta}$	Desoxihemoglobina, meta-Hb
420	Deformação fora do plano Fe-O-O	Oxihemoglobina, oxi-Hb
675	ligação (C-C-N)	Hemoglobina
744	Vibração do anel, ligação (CH_3 , CH_2)	Triptofano
1003	respiração do anel aromático	Fenilalanina
1076	vibração (C – O)	Ácido láctico
1124	CH_3 rocking, vibração (C-O), estiramento assimétrico (C-O-C)	Lactato, polissacarídeos, grupo heme
1170	estiramento (C-C)	Hemoglobina
1225	rotação de spin (Fe), β -folha, vibração da ligação no plano C_m -H	Hemoglobina
1248	amida III	Fibrina, Guanina, Citosina
1340	ligação (C-H)	Triptofano
1374	estiramento simétrico (CH_3)	Grupo heme
1399	estiramento assimétrico (C=N)	Grupo heme
1449	ligação (CH_2, CH_3)	Triptofano
1563	estiramento e vibração do anel pirrólico, vibração ($C_{\beta}C_{\alpha}$)	Hemoglobina
1580	estiramento (C=C)	Grupo heme
1583	ligação (C-H)	Hemoglobina
1620	vibração (C=C)	Grupo heme
1654	Amida I	Proteínas

, onde ν , estiramento; δ , deformação no plano; γ , deformação fora do plano; pyr, anel pirrólico; p, proteína; prop, propionato; subscritos α , β , representam os átomos de carbono na posição alfa, beta da porfirina; a e b corresponde a grupos de carbono vinílicos; c e d aos carbonos de propionato.

Como pode ser observado na Figura 15, o espectro Raman da mancha de sangue é rico em informação vibracional. A caracterização do espectro permitiu identificar os componentes presentes no sangue, como a hemoglobina (675, 1170, 1225, 1563 e 1583 cm^{-1}), o aminoácido triptofano (744, 1340 e 1449 cm^{-1}), o modo vibracional relacionado a presença de proteínas, amida III (fibrina, guanina, citosina em 1248 cm^{-1}), o aminoácido fenilalanina (1003 cm^{-1}), o ácido láctico e lactato (1076 e 1124 cm^{-1}), o grupo heme (1374, 1399, 1580 e 1620 cm^{-1}) e por fim, o modo amida I (1654 cm^{-1}). As bandas que estão relacionadas com o tempo, ou seja, que representam as mudanças químicas em função do tempo de exposição estão localizadas em 377 cm^{-1} associada a meta-Hb e em 420 cm^{-1} , a oxi-Hb.

As intensidades relativas dessas duas espécies sofrem alterações em função do tempo, sendo que as mesmas podem estar relacionadas com o processo de oxidação do ferro (Fe^{+2}) presente na oxi-Hb para a meta-Hb (Fe^{+3}) em contato com o ambiente externo.^{4,5} No entanto, essa mudança também acontece dentro do organismo embora não

seja muito significativa, já que existe uma enzima cuja função é reduzir a forma meta-Hb para a oxi-Hb, uma vez que a desoxihemoglobina é incapaz de transportar a molécula de oxigênio (O_2), a *citocromo- β 5-redutase*.

A Figura 16 mostra um esquema do ciclo que ocorre dentro do organismo envolvendo a oxidação e redução do ferro presente na hemoglobina pela enzima. Como essa espécie não existe fora do corpo humano, a tendência é que o processo de oxidação seja contínuo e que a forma meta-Hb esteja presente em maior quantidade na mancha em função do tempo de exposição do que a forma oxi-Hb. A consequência disso é a perda da capacidade de transporte de oxigênio pela hemoglobina, seguida pelos processos de desnaturação e agregação e finalmente a formação de hemicromos (Hc), Figura 2, que pode ser definida como a forma desnaturada da hemoglobina.²⁸

As espécies mencionadas acima possuem uma geometria ligeiramente diferente e consequentemente irão apresentar diferentes modos vibracionais, como mostrados na Tabela 1. Por isso, as técnicas de espectroscopia vibracional, entre elas, a espectroscopia Raman, podem ser empregadas para monitorar as mudanças químicas em função do tempo de exposição.

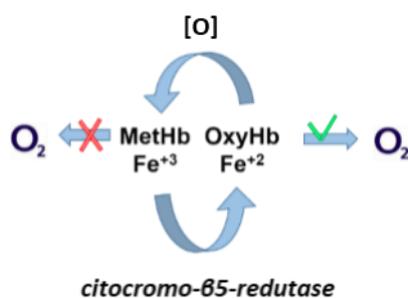


Figura 16. Diagrama representativo da oxidação e redução do ferro presente na hemoglobina

A Figura 17 compara dois espectros obtidos com a idade de 0h e após 3h de exposição ao ambiente. A partir dos espectros adquiridos, as alterações previstas foram confirmadas, pois houve um aumento na intensidade do sinal atribuído a meta-Hb (377 cm^{-1}) em função do tempo se comparado ao sinal da oxi-Hb (420 cm^{-1}), embora os demais sinais, de ambos os espectros, aparentemente não sofreram alterações significativas.

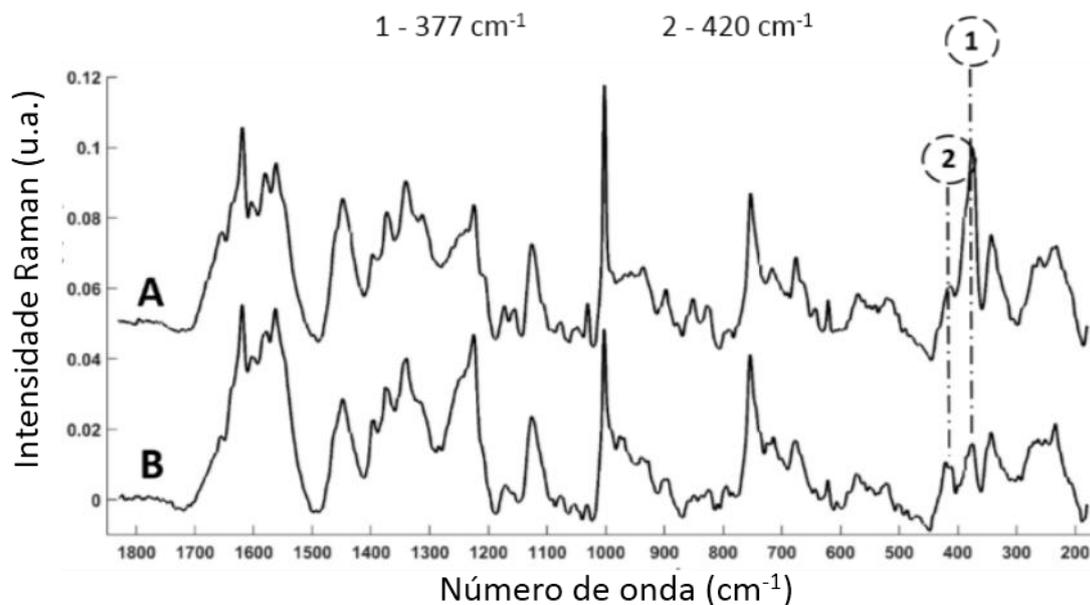


Figura 17. Espectros de manchas de sangue, A (3 horas) e B (0 horas) usando o laser de 785 nm

A Figura 18 mostra os espectros Raman obtidos a partir dos *swabs* que foram aplicados nas manchas de sangue no procedimento de recuperação e também do espectro obtido ao incidir diretamente o laser na mancha de sangue. Após fazer a comparação dos espectros de sangue e do sangue recuperado, foi notada o mesmo perfil entre as intensidades dos sinais de meta-Hb/oxi-Hb. Logo, o procedimento de recuperação na mancha de sangue se mostrou viável, uma vez que as bandas foram observadas tanto na amostra recuperada quanto na amostra sem recuperação.

A vantagem do uso de *swabs* é a facilidade na coleta das amostras, mesmo em superfícies irregulares, no entanto, devido ao tempo transcorrido entre a coleta da amostra, a recuperação e a obtenção do espectro, a banda em 420 cm^{-1} não foi observada com uma intensidade significativa, já que o processo de oxidação (oxi-Hb para meta-Hb) é rápido. Além disso, dois sinais intensos são observados no espectro do sangue recuperado, em 1095 e 1124 cm^{-1} , e estes são característicos do algodão presente nos *swabs*.

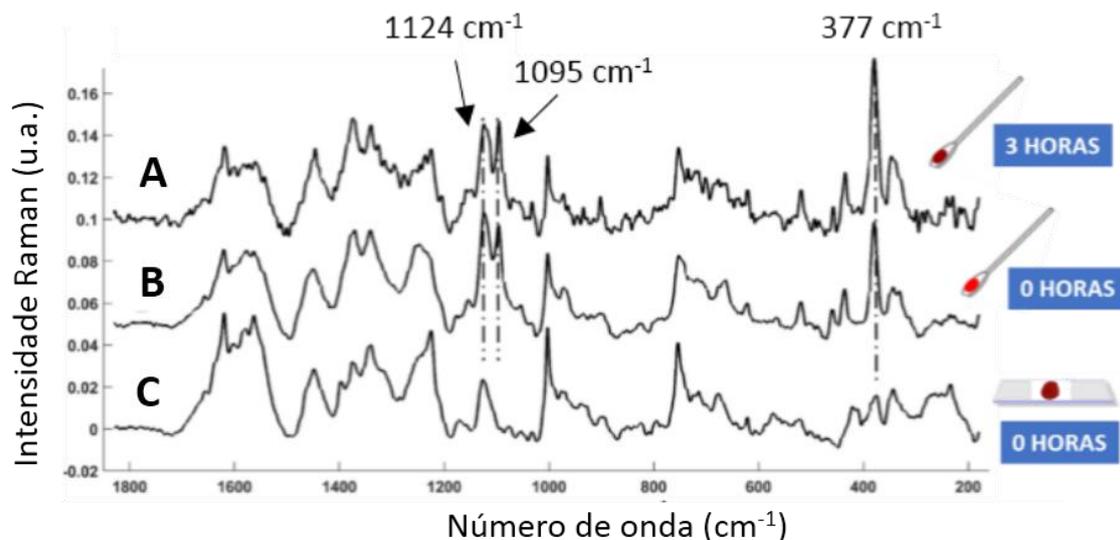


Figura 18. Espectro recuperado das manchas de sangue - A (3 horas) e B (0 horas) e espectro obtido usando o laser 785 nm diretamente na mancha C (0 horas)

2.4.3 Construção do modelo de calibração

Antes da construção do modelo de regressão, foi feita uma análise exploratória, empregando a PCA. A matriz de dados, \mathbf{X} , foi construída organizando as amostras na ordem crescente de tempo, de forma proposital, já que se houvesse alguma tendência natural nos dados, a mesma poderia ser observada nos escores de alguma das componentes principais relevantes (PCs).

O modelo foi construído com 4 PCs com uma variância explicada acumulada de 75,73%: PC1 (31,93%), PC2 (19,66%), PC3 (17,06%) e PC4 (7,08%). A Figura 19 mostra o gráfico usado para a escolha do número de PCs.

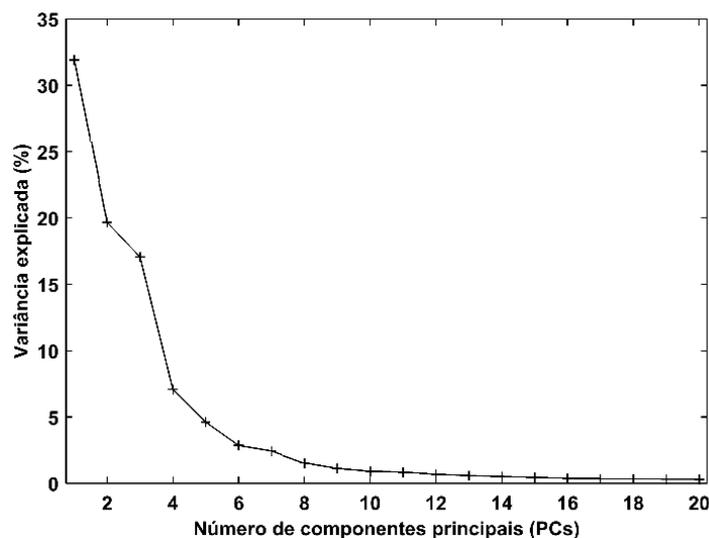


Figura 19. Número de componentes principais (PCs) em função da variância explicada

Os gráficos de escores das PCs que foram consideradas mais relevantes, PC1xPC2, PC1xPC3 e PC2xPC3 são apresentados na Figura 20A, B e C, respectivamente. Com base nas figuras, é possível observar que não houve distinção entre os sexos. Nos escores da PC3 com 17,06%, Figura 20D, foi possível observar uma tendência relacionada com o tempo de exposição da mancha de sangue, já que as amostras foram organizadas de maneira crescente de horas.

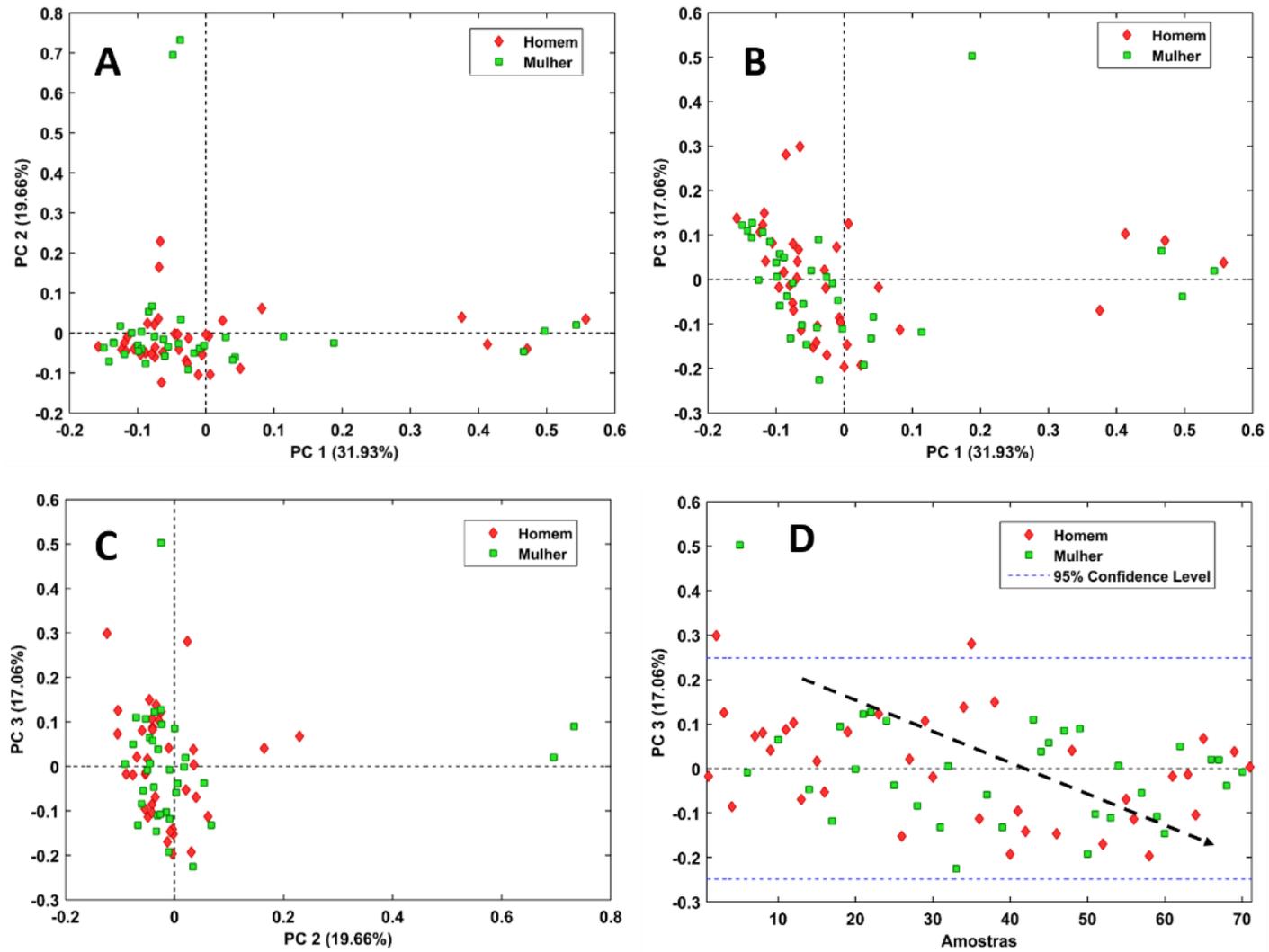


Figura 20. Gráfico de escores de PC1xPC2 (A), PC1xPC3 (B) e PC2xPC3 (C) além dos escores de PC3 (D) em função das amostras

Após a análise exploratória, partiu-se para a construção do modelo de regressão empregando o PLS. Ao todo, 71 espectros foram obtidos em um intervalo de tempo de 0 a 369 horas.

O pré-processamento OSC foi empregado como uma alternativa, já que o mesmo é recomendado quando a matriz \mathbf{X} (resposta instrumental) possui uma variabilidade muito grande em relação a propriedade que se deseja medir, no caso, o vetor \mathbf{y} (variável de interesse).^{93,94}

Considerando as metodologias propostas de pré-processamento para a matriz \mathbf{X} e também para o vetor \mathbf{y} , foram construídos vários modelos, sendo que os mesmos foram comparados usando os seguintes parâmetros: RMSEC, RMSECV, RMSEP, número de LVs e coeficiente de correlação tanto da calibração quanto da validação. O número de variáveis latentes foi escolhido para cada modelo analisando o gráfico dos valores de RMSECV versus o número de variáveis latentes. Foram selecionados um número de variáveis latentes adequado que apresentasse o menor erro de validação cruzada (valor de RMSECV) e uma porcentagem da variância explicada adequada. A validação cruzada foi feita usando amostras retiradas do conjunto de calibração de maneira aleatória (*random subsets*).

Após isso, os coeficientes dos modelos foram calculados e os parâmetros de desempenho RMSEC e RMSEP foram avaliados. Os resultados obtidos com os modelos empregando diferentes pré-processamentos são apresentados na Tabela 2.

Tabela 2. Modelos multivariados obtidos usando os valores do vetor y em horas e log [(horas)]

	y (horas)			y [log(horas)]		
	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5	Modelo 6
Pré- processamentos	1.WLS (2° grau) 2.Normalização (vetor de comprimento unitário) 3. Alisamento Savitzky-Golay (1° grau + 9 pts) 4.Centrar na média	1.OSC (correção do sinal ortogonal)	1.WLS (2° grau) 2.Normalização (vetor de comprimento unitário) 3. Alisamento Savitzky-Golay (1° grau + 9 pts) 4. OSC (correção do sinal ortogonal) 5. Centrar na média	1.WLS (2° grau) 2.Normalização (vetor de comprimento unitário) 3. Alisamento Savitzky-Golay (1° grau + 9 pts) 4.Centrar na média	1.OSC (correção do sinal ortogonal)	1.WLS (2° grau) 2.Normalização (vetor de comprimento unitário) 3. Alisamento Savitzky-Golay (1° grau + 9 pts) 4. OSC (correção do sinal ortogonal) 5.Centrar na média
X %	82,71	93,07	33,55	82,61	92,82	33,2
Y %	97,69	28,90	79,69	93,06	20,33	68,59
RMSEC	12	69	35	0,09	0,31	0,2
RMSECV	44	82	57	0,28	0,4	0,31
RMSEP	16	50	36	0,17	0,2	0,14
LVs	7	2	2	7	2	2
R²_{cal}	0,98	0,29	0,82	0,93	0,2	0,68
R²_{val}	0,93	0,46	0,66	0,43	0,1	0,57

A partir dos resultados obtidos, o modelo 1 apresentou os menores valores tanto para RMSEC quanto para RMSEP, e além disso, foi obtido o maior valor para o coeficiente de correlação (R^2). A partir disso, a detecção de amostras anômalas (*outlier*), determinação do número adequado de variáveis latentes, gráficos de y medido contra y predito dentre outras informações serão apresentados somente para este modelo.

Na Figura 21 é apresentado o gráfico do número de variáveis latentes contra os valores do RMSECV e o RMSEC. A partir dele, foi feita a escolha de 7 LVs para construir o modelo de regressão.

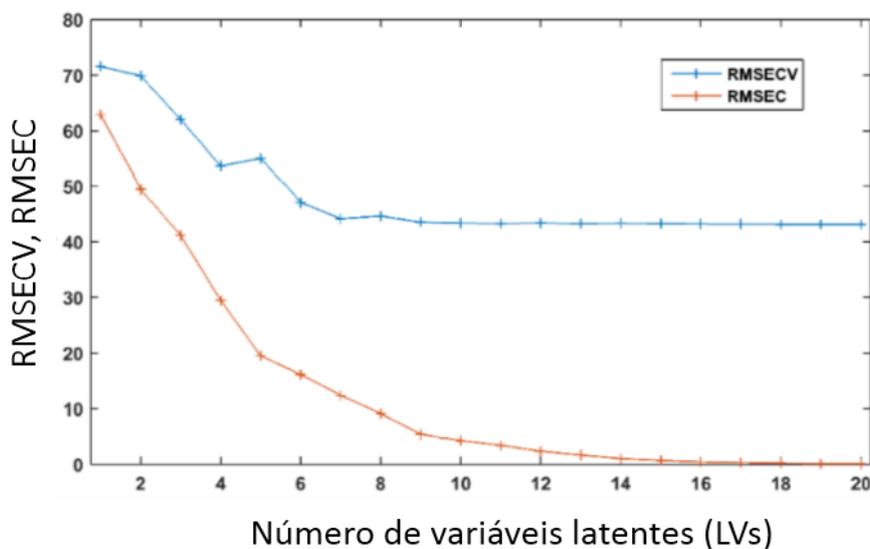


Figura 21. Gráfico de variáveis latentes contra valores de RMSEC e RMSECV

As amostras anômalas foram identificadas por meio da análise do gráfico de Resíduos em Y em relação aos valores de *leverage*. O *leverage* é um parâmetro da influência da variável para o modelo. Os resíduos em Y são os erros associados a cada amostra prevista pelo modelo, tanto na calibração quanto na validação. As amostras com alto valor de resíduos em Y e *leverage* são consideradas anômalas.

Os gráficos de *leverage* contra Resíduos em Y e também do gráfico de T^2 de *Hotteling* contra os Resíduos Q estão apresentados na Figura 22. Os pontos observados nos gráficos representam as amostras do conjunto de calibração, que foram obtidos utilizando o procedimento de recuperação.

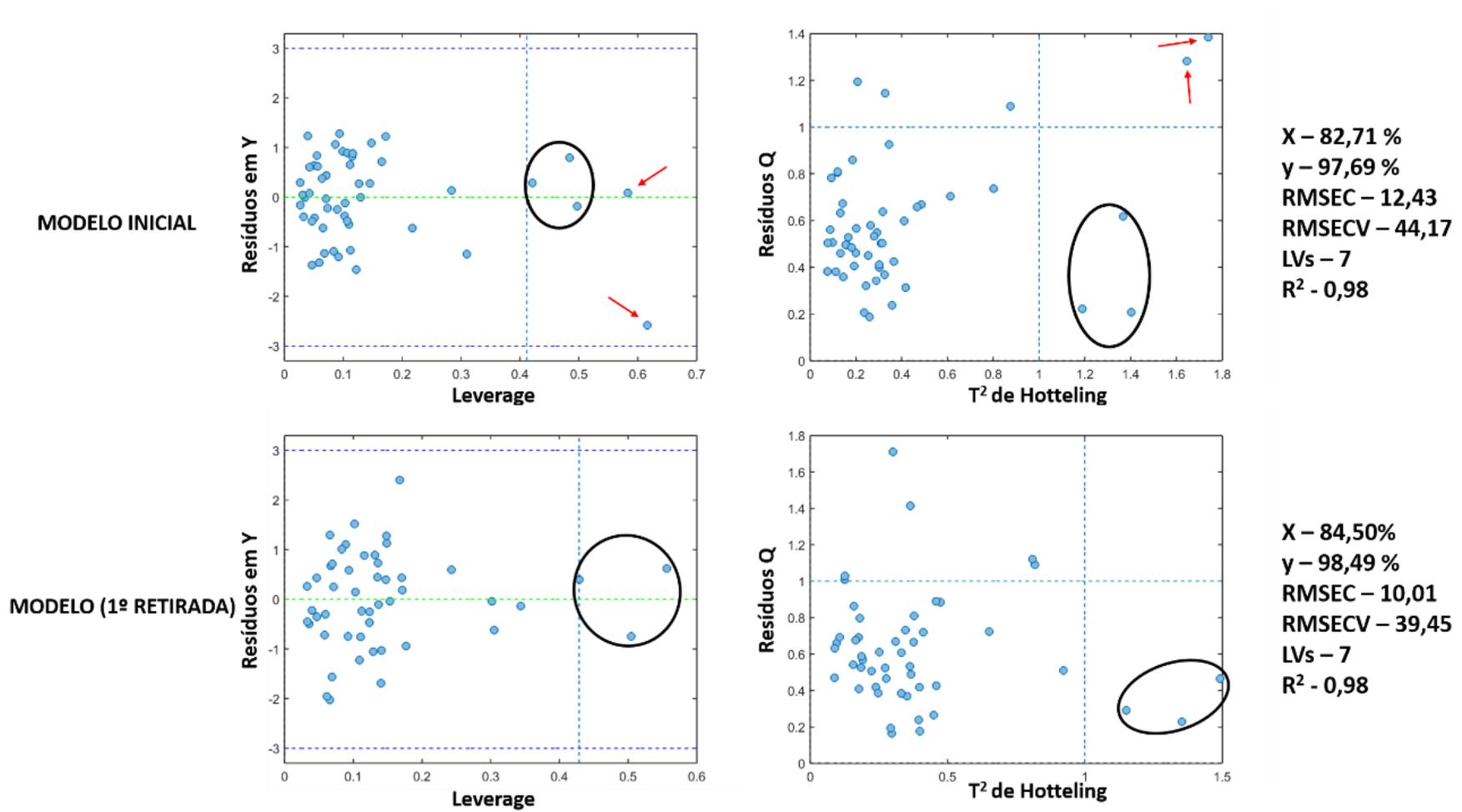


Figura 22. Gráfico de T² de Hotelling contra os Resíduos Q e gráfico de leverage contra Resíduos em Y

Com base no gráfico é possível constatar a presença de duas amostras (modelo inicial-setas vermelhas) que são consideradas importantes para o modelo, ou seja, possuem um alto valor para *leverage*, no entanto, também apresentam um alto valor de resíduos Q, ou seja, a sua variância não está sendo bem modelada o que resulta em altos valores para os resíduos em **Y** (modelo inicial). A partir dessas observações, ambas podem ser consideradas anômalas.

Na Figura 22, também é possível inferir quais são as amostras (espectros) mais importantes para a construção do modelo (círculo preto). Mesmo com as amostras tendo valores consideráveis de resíduos em **Y**, as mesmas apresentam um alto valor de *leverage* e também de T^2 de *Hottelling*, sinalizando sua grande importância para o modelo.

Após a primeira rodada de retirada de amostras anômalas, o modelo foi reconstruído e o mesmo não apresentou *outlier* (Figura 22), embora tenha a presença de duas amostras com um valor considerável de Resíduos em **Y**, as mesmas estão sendo bem modeladas pelo modelo já que possuem um valor de resíduos Q baixo. Após realizar todo esse processo, as amostras do conjunto de validação foram previstas obtendo os seguintes valores para o modelo final apresentadas na Tabela 3.

Tabela 3. Valores para variância explicada na matriz **X**, vetor **y** e os parâmetros do modelo PLS

X %	y %	RMSEC	RMSECV	RMSEP	LVs	R²_{cal}	R²_{val}
84,5	98,49	10	39	18	7	0,98	0,91

Por fim, foi gerado o gráfico de valores de **y** medido contra **y** previsto para o modelo final construído e está apresentado na Figura 23.

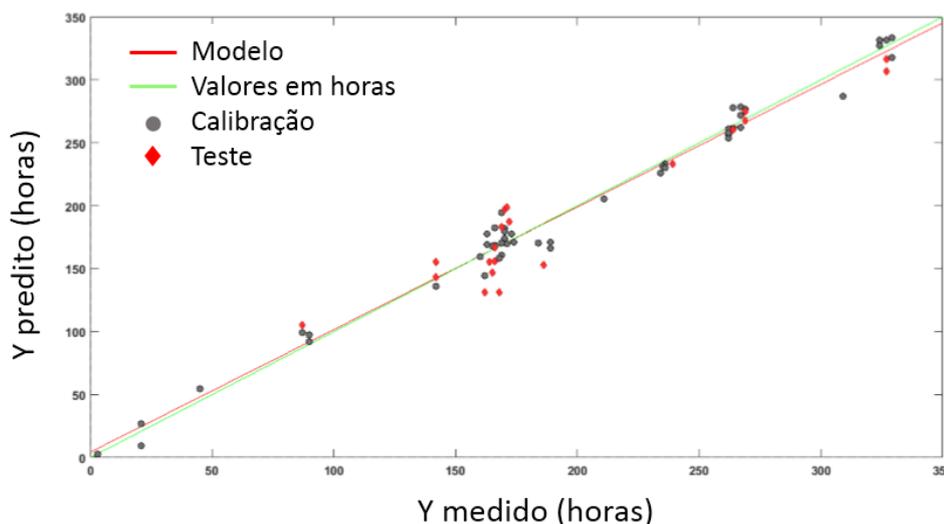


Figura 23. Gráfico de **Y** medido contra **Y** previsto para o modelo PLS

Uma informação importante obtida do modelo PLS é a interpretação das variáveis, no caso, números de onda (cm^{-1}) que foram considerados mais relevantes para a construção do modelo. Essa informação é obtida após gerar o gráfico de VIP scores.

Além disso, também é possível plotar o gráfico das amostras contra os valores de Resíduos em Y (em horas), e a partir dele obter os erros de cada amostras e não apenas os valores para os erros médios do modelo (RMSEC, RMSECV e RMSEP). Os gráficos de VIP scores e também das amostras contra resíduos em Y estão apresentados na Figura 24 e na Figura 25.

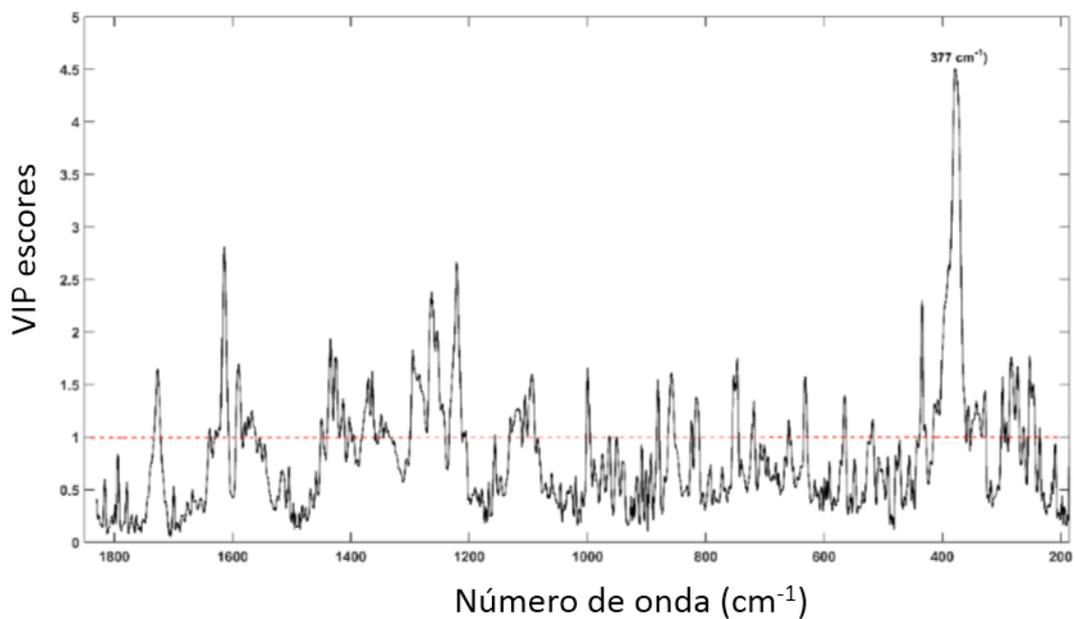


Figura 24. Gráficas de VIP scores para o modelo de regressão

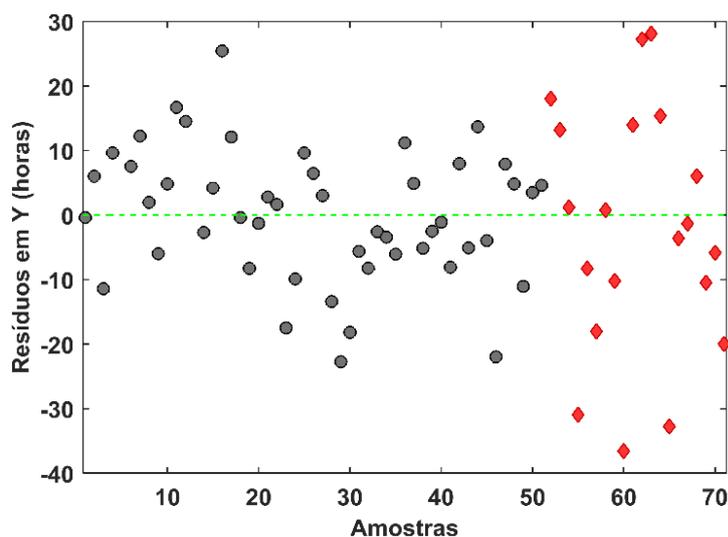


Figura 25. Gráfico de Amostras contra os resíduos em Y

Com base no primeiro gráfico, é possível compreender a importância da banda 377 cm^{-1} para o modelo, uma vez que a mesma está associada a espécie que é produzida em função do tempo, a meta-Hb. Outras regiões importantes para a construção do modelo de regressão estão relacionadas aos modos vibracionais do grupo heme, como na região de $1620, 1580, 1370\text{ cm}^{-1}$. Com relação ao segundo gráfico, é possível observar a distribuição aleatória dos resíduos e se ter uma perspectiva dos erros absolutos de cada amostra no modelo, tanto para o conjunto de calibração quanto para o conjunto de validação. Com base nele, infere-se que o erro máximo esteve na faixa de 0 a 40 horas considerando todas as amostras.

No estudo também foi feita a comparação do espectro Raman de uma mancha de sangue que ficou exposta a baixa umidade (7%) com o espectro de uma outra mancha exposta em condições ambiente, ambas com o tempo de 24 horas. Os experimentos foram feitos em uma temperatura ambiente média de 25°C com uma umidade em torno de 40%. Os resultados obtidos estão apresentados na Figura 26. Os espectros passaram por etapas de pré-processamento, no caso, a correção da linha de base por WLS (2° grau) e normalização pelo vetor de comprimento unitário.

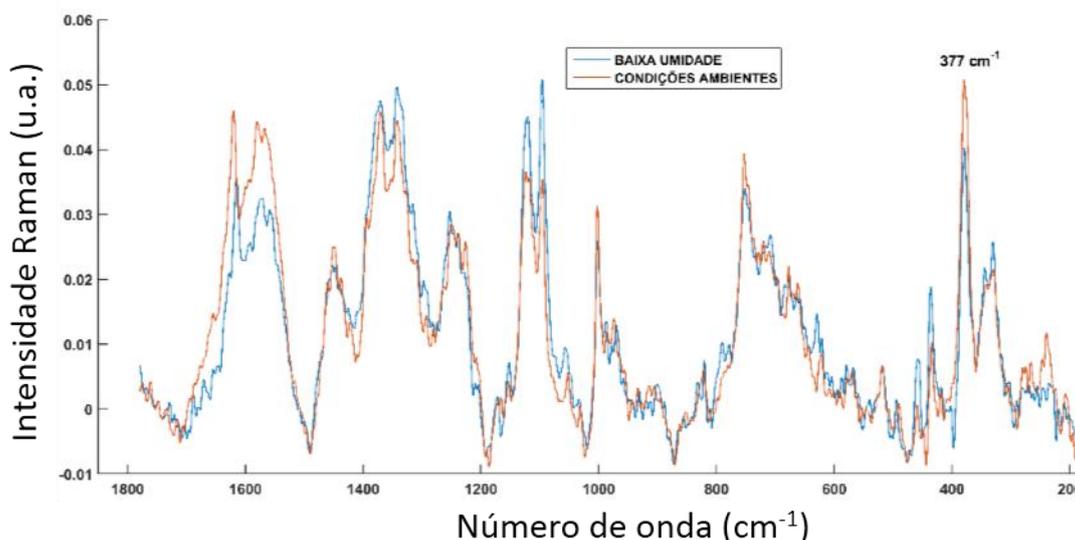


Figura 26. Espectro Raman da mancha de sangue exposta em condições ambientes e a baixa umidade com 24 horas de exposição

A partir da comparação dos espectros é possível notar que a intensidade relativa da banda referente a meta-Hb (377 cm^{-1}) diminuiu em baixa umidade (7%) em relação a mesma banda no espectro da mancha que ficou exposta em condições ambientes. A partir disso, é possível inferir que em baixa umidade as reações químicas envolvidas no

processo de envelhecimento da mancha, oxidação e desnaturação, sofrem um decréscimo em sua velocidade.

No entanto, o resultado contrário é observado no espectro de uma mancha exposta a alta umidade (86%), Figura 27. Com base nos espectros obtidos, é possível inferir que de fato a umidade têm influência direta no processo de envelhecimento da mancha, sendo que quanto maior a umidade maior será a velocidade das reações químicas envolvidas.

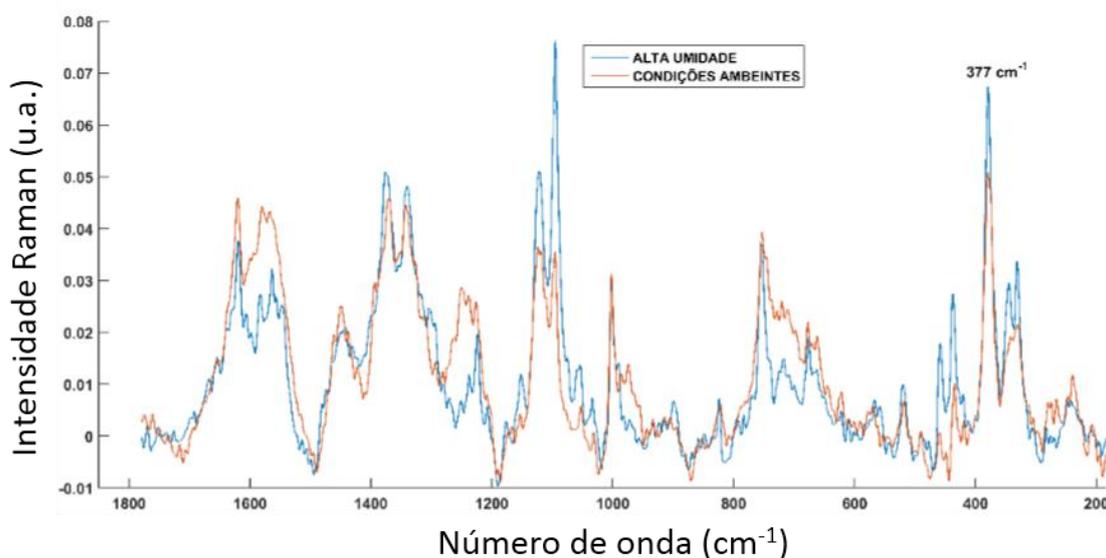


Figura 27. Espectro da mancha de sangue exposta as condições ambientes e a alta umidade

Esses resultados estão coerentes com os estudos presentes na literatura, embora neste caso, é interessante ressaltar a importância do procedimento de recuperação, pois a partir dele, também foi observado as mesmas tendências já apresentadas em outros trabalhos com o objetivo de prever a idade de manchas de sangue.^{11,19}

2.5 Conclusão

A partir dos espectros Raman foi possível observar as mudanças químicas que ocorrem na mancha de sangue com o tempo de exposição. A recuperação da mancha empregando *swabs* apresentou potencial para ser empregado nos laboratórios forenses para estimar a idade de uma mancha de sangue, uma vez que a coleta pode ser feita independente da superfície em que a mancha esteja depositada.

O modelo PLS após a retirada das amostras anômalas foi construído usando 7 LVs com uma variância explicada na matriz \mathbf{X} de 84,5% e no vetor \mathbf{y} de 98,49%. Os valores para os parâmetros RMSEC, RMSECV e RMSEP foram 10, 39 e 19, respectivamente. O coeficiente de correlação calculado entre os valores de \mathbf{y} medido versus o \mathbf{y} predito para a calibração e para a validação foram de 0,98 e 0,91, respectivamente.

Por fim, para a construção de modelos que possam estimar a idade da mancha de sangue com menor erro de previsão e maior variabilidade é necessário aumentar o número de doadores e fazer medidas em diversas condições ambientes. A construção de modelos de classificação empregando faixas de horas também pode ser uma alternativa com os dados obtidos por espectroscopia Raman.

CAPÍTULO 3

3. ESTIMATIVA DO TEMPO DE EXPOSIÇÃO DE MANCHAS DE SANGUE EMPREGANDO IMAGENS DIGITAIS: UMA ABORDAGEM MULTIVARIADA

3.1 Introdução

Uma imagem digital é composta pela união de vários *pixels*, do inglês, *picture element*¹⁰⁴ e a quantidade deles está associada com a resolução de uma imagem digital, uma vez que, quanto maior a quantidade maior será a resolução.^{105,106}

Do ponto de vista matemático, uma imagem pode ser representada por um arranjo bidimensional (x, y) , sendo que para cada *pixel* é associado um valor de luminosidade dentro de uma escala que varia de canal para canal e que juntos formam os sistemas de cores.

A maioria dos dispositivos de captura de imagem, digitais e analógicos, como por exemplo, *smartphones*, *scanners* e câmeras fotográficas utilizam o sistema de cores chamado RGB, composto pelas cores vermelho (R, *red*), verde (G, *green*) e azul (B, *blue*).

Esse sistema de cor pode ser representado como um arranjo tridimensional, onde cada eixo coordenado varia numa escala de 0 a 255, sendo que o 0 é a ausência da cor e o 255 é a cor pura, isto é, sem nenhuma mistura.¹⁰⁷ A combinação entre os canais R, G e B no arranjo que representa o sistema de cores RGB pode gerar mais de 16 milhões de tons diferentes (256^3).

A Figura 28 mostra a representação do sistema RGB, que pode ser visualizado como sendo um cubo em que os vértices localizados nos eixos coordenados estão relacionados com as cores azul, vermelho e verde e os outros vértices com as cores ciano, magenta e amarelo. Na escala de cinza, o ponto na origem representa o preto (0,0,0) e o mais afastado a cor branca (255, 255, 255), e a mesma pode ser representada como sendo uma linha que se estende da origem até o vértice mais distante, ou seja, pode ser representada como sendo a diagonal do cubo.

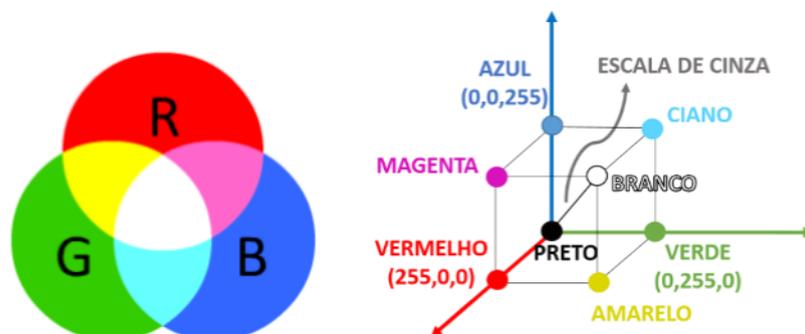


Figura 28. Representação geométrica do sistema RGB de acordo com ref. 107

O RGB é o sistema de cores mais empregado, no entanto, existem outros sistemas de cores, como por exemplo o HSV (**H**ue, **S**aturation, **V**alue) e o CMYK (**C**iano,

Magenta, Yellow, black). Geometricamente, o modelo HSV é representado como sendo um triângulo de cores, Figura 29, e pode ser obtido fazendo um corte transversal do cubo que representa o sistema RGB.

Neste triângulo, os vértices são as cores aditivas e o centro do triângulo está associado a combinação dessas cores, em iguais proporções. Este sistema pode ser de grande ajuda, uma vez que está associado a cor observada pelos indivíduos, pois o matiz (H) é um atributo ligado com o comprimento de onda dominante em uma mistura de ondas de luz. A saturação (S) está relacionada com o brilho, e pode ser definida como sendo a pureza relativa ou a quantidade de luz branca misturada com o matiz.¹⁰⁴⁻¹⁰⁸

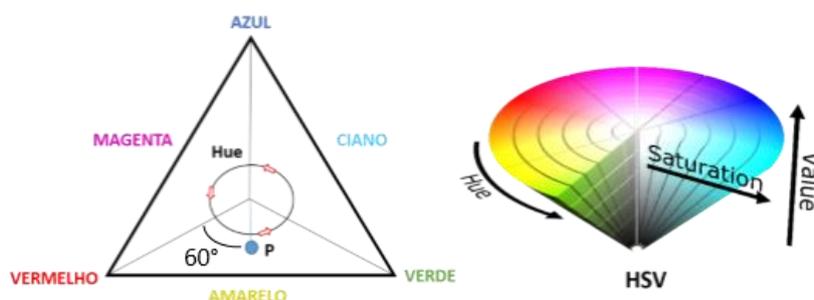


Figura 29. Representação do sistema de cor HSV de acordo com ref. 108

Na figura acima, o valor de H, no ponto P, é devido ao ângulo do vetor em relação a cor vermelha. Assim, para a cor vermelha é atribuído o valor de $H=0^\circ$, e para a cor amarela, $H=60^\circ$, e assim por diante. Os valores de S e de V são atribuídos dentro de uma escala que varia do centro do triângulo ao ponto P.

Existem outros sistemas de cores além daqueles que já foram mencionados, como por exemplo, o HSL, o L^*a^*b , o YCbCr, o XYZ, o YUV e o YIQ. As representações geométricas para esses sistemas estão apresentadas na Figura 30.

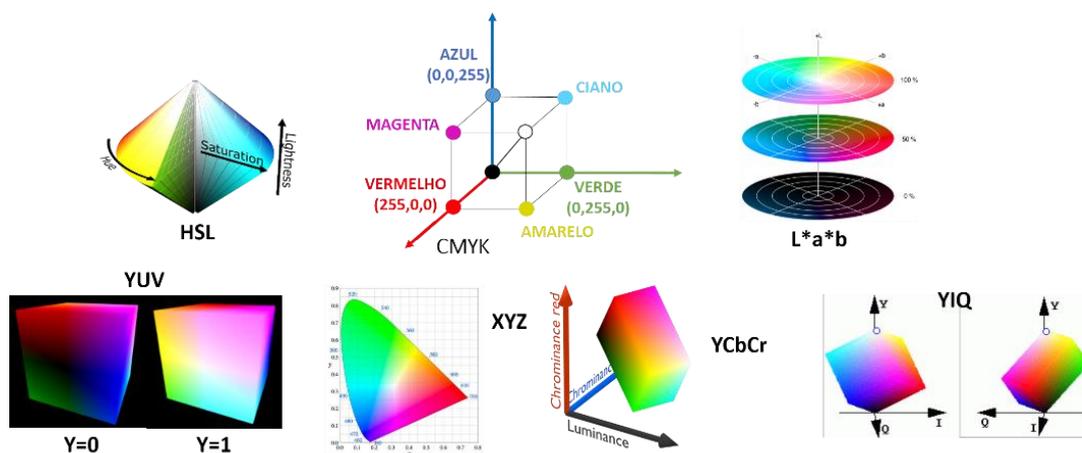


Figura 30. Representação dos sistemas de cores CMYK, HSL, YCbCr, XYZ, L^*a^*b , YUV, YIQ baseado na ref. 108

Todos os sistemas de cores apresentados podem ser obtidos por meio de transformações matemáticas a partir do sistema de cores RGB. Um exemplo disso, é a conversão do sistema RGB para HSV empregando as equações mostradas a seguir, onde *min.* e *máx.* são o menor e o maior valor de luminosidade associado aos canais R, G e B, respectivamente.¹⁰⁶

$$V = \max(R, G, B) \quad (23)$$

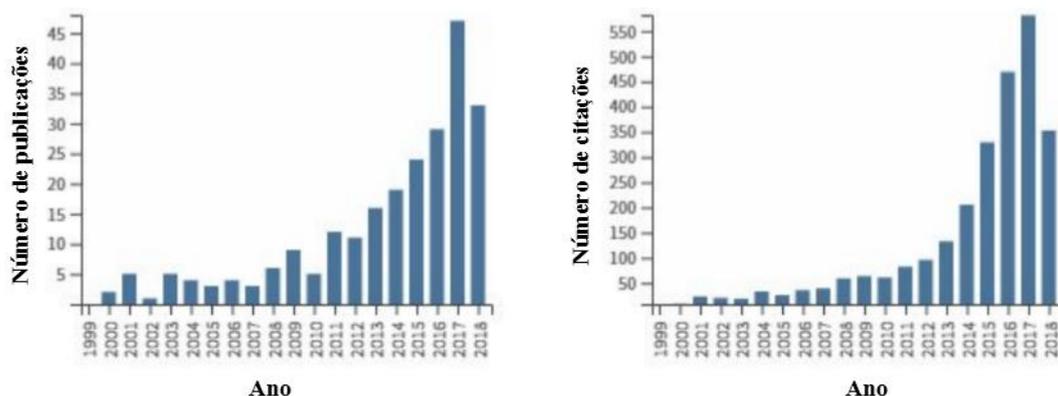
$$S = 1 - \frac{3}{R + G + B} [\min(R, G, B)] \quad (24)$$

$$H = \cos^{-1} \left\{ \frac{1/2[(R-G)+(R-G)]}{2\sqrt{[(R-G)^2+(R-G)(G-B)]}} \right\} \quad (25)$$

O uso de imagens digitais como técnica analítica tem se destacado no meio científico nas últimas duas décadas, uma vez que se tornou uma alternativa viável frente a outras metodologias analíticas. As principais vantagens das metodologias baseadas no emprego de imagens digitais são o baixo custo, simplicidade, rapidez e a possibilidade de realização de uma análise *in situ*.

Embora existam técnicas analíticas que possuem equipamentos portáteis, por exemplo, a espectroscopia Raman e a espectroscopia na região do infravermelho próximo (NIR), os aspectos custo e simplicidade se destacam, já que o valor dos equipamentos usados nessas técnicas é muito maior se comparados ao preço médio de um *smartphone* ou de *scanner* de mão que estão disponíveis comercialmente.

Uma pesquisa feita no site *Web of Science* com as palavras chave, **phone* e **analytical method*, gerou dois gráficos que estão apresentados na Figura 31. O primeiro contém o número de publicações e o segundo o número de citações, ambos em função do ano. A partir dos gráficos é possível notar que de fato, o uso de imagens digitais em pesquisa científica é recente, menos de duas décadas e o número de publicações vem aumentando nos últimos anos. O número de citações acompanha a tendência de aumento do número de publicações.



Fonte: Web of Science. Palavras chave: *phone, analytical method, em 09/07/2018

Figura 31. Produção bibliográfica da área no decorrer dos anos

No Brasil, o número de dispositivos conectados à internet incluindo *smartphones*, *desktops* e *notebooks* tem tido um crescimento elevado. Um dado relevante e expressivo é que no ano de 2016, no Brasil, o número de aparelhos conectados na internet chegou a 208 milhões de dispositivos (Figura 32), ou seja, estatisticamente é um aparelho para cada habitante num país.¹⁰⁹ No entanto, no contexto mundial, o Brasil é o 4º colocado no mercado comercial, atrás de Índia, China e Estados Unidos (EUA) na venda e no consumo de dispositivos eletrônicos incluindo *smartphones*, *notebooks* e *tablets*.

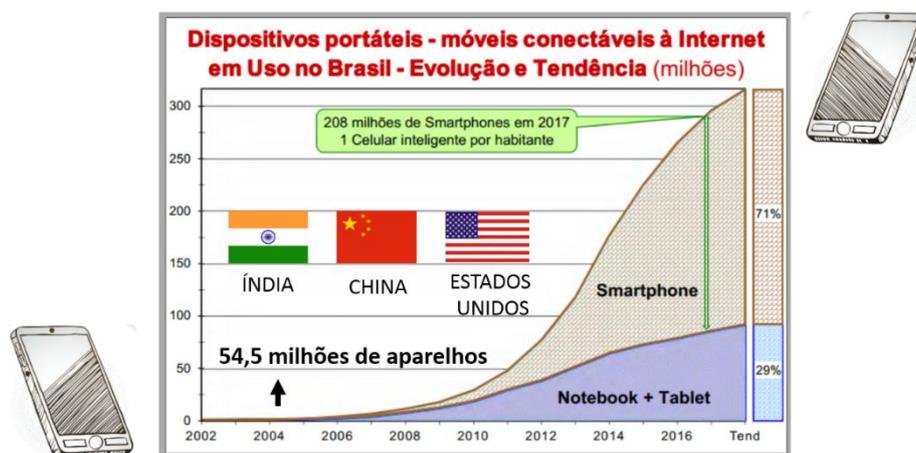


Figura 32. Esquema da quantidade de smartphones, *notebook* e *tablets* em milhões no Brasil em 2016 de acordo com ref. 109

O interesse pelo uso de imagens digitais tem crescido e conseqüentemente novas aplicações têm sido desenvolvidas e utilizadas em diversas áreas. Uma aplicação promissora tem sido monitorar reações químicas onde espécies coloridas são produzidas ou consumidas em função do tempo.

Dos trabalhos encontrados na literatura, destacam-se aplicações na área de controle de qualidade de alimentos, como por exemplo, o monitoramento do processo de

produção de pães¹¹⁰, qualidade de presunto¹¹¹, análise de vinhos¹¹², quantificação de lactobacilos em leite fermentado¹¹³, monitoramento da fermentação de chás¹¹⁴, detecção de doenças em vegetais¹¹⁵ e até na determinação do tamanho de grãos de arroz.¹¹⁶

O sistema RGB é o mais amplamente utilizado e encontrado na literatura, no entanto, o uso de mais de um sistema de cor oferece vantagens, como a possibilidade de extrair mais informações da imagem digital obtida. Para ilustrar a veracidade do argumento, a Figura 33, Figura 34 e Figura 35 mostram uma imagem de satélite da cidade de Brasília nos sistemas de cores, RGB, CMYK e HSV, respectivamente.¹¹⁷



Figura 33. Imagem de satélite da cidade de Brasília no sistema RGB de acordo com ref. 117

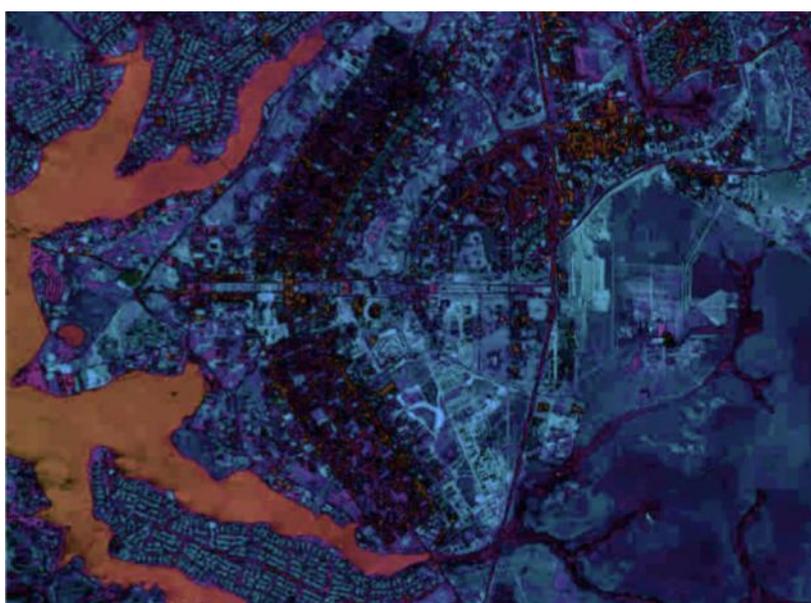


Figura 34. Imagem de satélite da cidade de Brasília no sistema CMYK de acordo com ref. 117

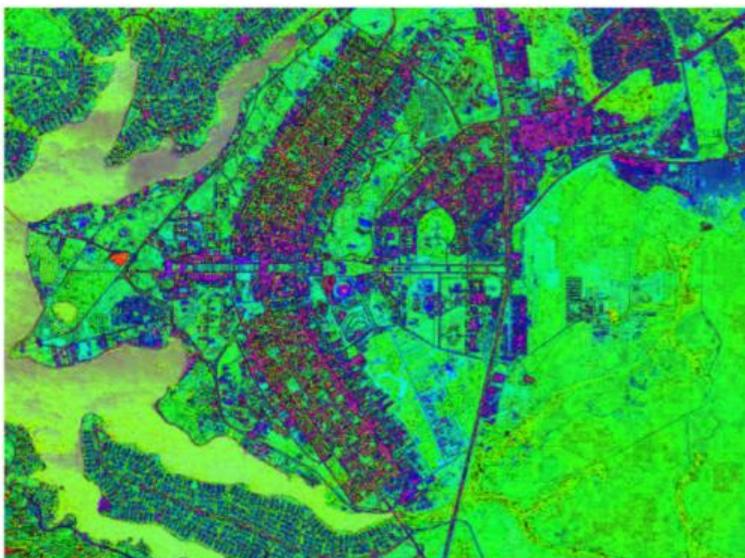


Figura 35. Imagem de satélite da cidade de Brasília no sistema HSV de acordo com ref. 117

Neste caso, usando o sistema HSV é possível fazer a diferenciação visual de uma área urbanizada de uma não urbanizada. Um ponto negativo é que tanto a área de vegetação quanto a região do lago apresentaram o mesmo padrão de cor e se o objetivo da análise fosse identificar pontos onde há presença de água, a escolha desse sistema em questão seria inadequada. Para contornar o problema deve-se tentar buscar outros sistemas de cores, como por exemplo, o CMYK. Usando esse sistema de cor, a diferença entre a região do lago e o restante da imagem fica evidente.

Logo, usando os três sistemas juntos, RGB, HSV e o CMYK é possível discriminar e destacar três regiões importantes da imagem: a área urbanizada, a área não urbanizada e a região do lago. Logo, usar mais de um canal traz benefícios com relação as informações que podem ser extraídas, já que foi possível destacar três regiões presentes na imagem com mais clareza do que na imagem original (RGB).

O emprego de imagens digitais como metodologia analítica para obter informações relacionadas a mudanças químicas tem sido de grande interesse na área forense, pois a mesma mostra ser uma alternativa inovadora, simples e rápida em relação a outras já conhecidas e bem fundamentadas.^{7,9,10,11,18,19} Na literatura, existem somente dois trabalhos que utilizaram imagens digitais para estimar a idade de manchas de sangue.^{18,19}

O estudo de Thanakiatkrai *et al.*¹⁹ que empregou o uso de diferentes substratos (papel de filtro, vidro e algodão) além de ter avaliado os efeitos da variação na câmera empregando diferentes *smartphones* (*Galaxy S*, *iPhone 4*, *iPad 2*) e também usando diferentes condições ambientes. O estudo considerou 4 doadores com uma idade média

de $23,5 \pm 2,4$ anos. O modelo de regressão para determinar o tempo de exposição da mancha foi construído dentro de um intervalo de 42 dias com um erro relativo de 12%.

A partir dos resultados, os autores puderam retirar algumas observações importantes acerca das condições ambientes. Com base na rápida mudança de coloração em temperaturas mais altas, foi proposto que em temperaturas elevadas, a velocidade do processo de envelhecimento da mancha aumenta, principalmente na primeira hora, devido as alterações na reação de oxidação, da oxi-Hb para a meta-Hb, e a posterior desnaturação, de meta-Hb para a formação dos Hc.^{19,118}

Para um ambiente com umidade elevada, a velocidade da mudança de cor sofreu um decréscimo, sendo observado um envelhecimento mais lento. No que se refere ao tipo de luz, as manchas expostas a luz solar envelheceram mais rapidamente do que aquelas expostas a luz ambiente e também as expostas a uma lâmpada fluorescente. No entanto, mesmo analisando três sistemas de cores diferentes (RGB, CMYK e HSL), o estudo construiu o modelo de regressão de maneira univariada, ou seja, empregando somente um canal, justamente aquele em que foi obtido o maior coeficiente de determinação (R^2), no caso, o canal magenta (M).¹⁹

Outro estudo de manchas de sangue com imagem digital foi o trabalho de Shin *et al.*¹⁸. Esse estudo foi mais limitado, uma vez que foram avaliados 2 sistemas de cores (RGB e HSV) e os efeitos das condições ambientes não foram explorados. O modelo de regressão foi construído de forma univariada com um faixa de previsão de 42 horas empregando somente o canal Valor (V) do sistema HSV.

3.2 Objetivos

3.2.1 Objetivo geral

Construção de um modelo matemático para estimar a idade de manchas de sangue em diferentes substratos e em diferentes condições ambientes empregando imagens digitais e ferramentas quimiométricas.

3.2.2 Objetivos específicos

- Obter imagens digitais em diferentes substratos;
- Empregar diferentes sistemas de cores para construir um modelo de regressão multivariado;
- Comparar as informações obtidas usando dois modos diferentes: a imagem sem recortes e selecionar uma região de interesse;
- Comparar os modelos construídos usando dois conjuntos de dados diferentes: os valores médios de luminosidade e a frequência relativa;
- Discriminar os canais mais importantes para o modelo de regressão;
- Avaliar a influência das condições de armazenamento da mancha de sangue no modelo de regressão;

3.3 Parte Experimental

3.3.1 Obtenção das amostras

Após aprovação do projeto junto ao COEP, foram coletadas amostras de seis voluntários, homens e mulheres, com idade média de 26 ± 4 anos. A identificação das amostras foi feita usando códigos que não tinham nenhuma relação com o doador. Antes da coleta, tanto a metodologia quanto os objetivos do estudo foram esclarecidos, sendo que os participantes estavam cientes de que poderiam interromper o procedimento a qualquer momento. A autorização para a doação do material biológico foi feita mediante a assinatura do termo de consentimento livre e esclarecido apresentado no apêndice, seção A.5

As gotas de sangue foram obtidas por meio de uma punção digital usando lancetas descartáveis e a deposição foi feita sobre uma folha de papel A4 fixada em uma lâmina de microscópio. A data e o horário no momento da coleta foram anotados e usados posteriormente no momento da aquisição dos dados para atribuir a imagem obtida sua respectiva idade.

Após a coleta e a deposição da mancha no substrato, a obtenção das imagens teve início. A coleta das imagens não seguiu um planejamento específico, embora foram obtidas imagens para todas as manchas nas horas iniciais, da mancha fresca (0 h) até 5 horas.

Além do papel A4, outros substratos também foram estudados: 5 tipos de tecidos, 4 tipos de pisos (3 pisos cerâmicos e 1 tipo ardósia) e madeira. As imagens dos substratos estão apresentadas na Figura 36 e o esquema da metodologia empregada está apresentado na Figura 37, apenas para a folha A4.

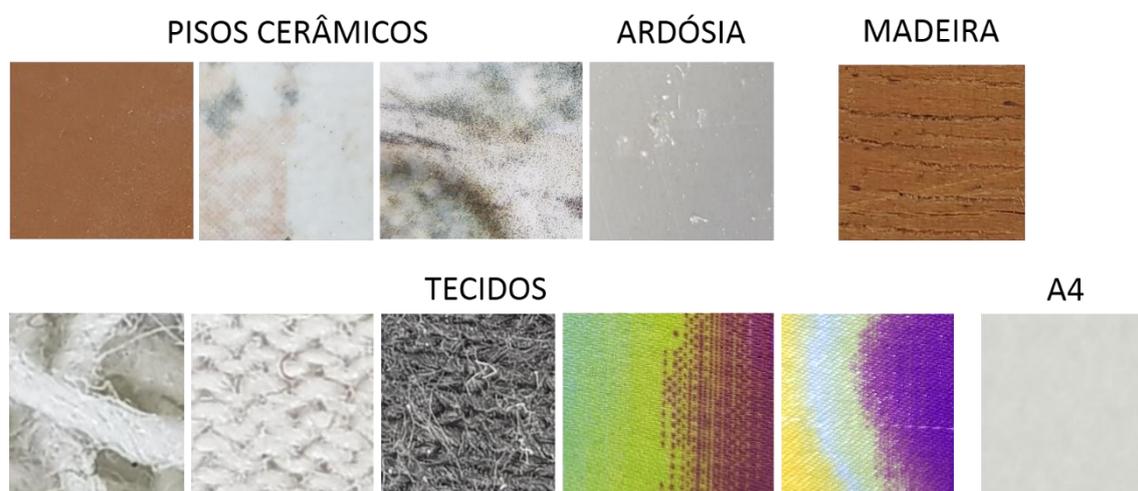


Figura 36. Diferentes substratos usados na construção dos modelos de regressão



Figura 37. Esquema da metodologia da coleta da gota de sangue até a aquisição da imagens digitais

As amostras foram armazenadas em condições ambientes, com uma temperatura média de 25°C e 40% de umidade. Nesse estudo também foi avaliada a influência das condições ambientes (luminosidade e umidade). Para os ensaios de umidade, foram empregadas soluções saturadas de cloreto de potássio (86%) e hidróxido de sódio (7%). As manchas de sangue coletadas foram armazenadas dentro de dois dessecadores diferentes com as soluções saturadas. Já para os testes na ausência de luz, as lâminas com as manchas de sangue foram mantidas dentro de uma caixa fechada, mas em contato com

o oxigênio, sendo que as mesmas foram retiradas somente no momento da coleta da imagem.

3.3.2 Obtenção das imagens digitais

Para a aquisição das imagens digitais foi construído um sistema de baixo custo composto por uma caixa de isopor, uma capa de celular, três pilhas de 1,5 V e um LED (do inglês, *light emitting diode*) de alto brilho na cor branca com uma tensão de 3,5 V. Foram usados elásticos e fita isolante para fixar as pilhas na caixa e após a construção do aparato, o mesmo foi revestido com papel alumínio. Por fim, o suporte para celular foi fixado na parte superior da caixa de isopor. Uma imagem do sistema em funcionamento está apresentada na Figura 38.



Figura 38. Sistema de aquisição de imagens digitais

As imagens foram obtidas empregando um *smartphone Samsung Galaxy S8* modelo SM-G950F, com abertura focal de 1,7, distância focal de 4,20 mm, balanço de branco automático, ISO 160, tempo de exposição 1/40s e câmera traseira de 12 MP. A distância entre a câmera do aparelho até as amostras foi de 5 cm.

A partir das manchas de sangue obtidas foram coletadas 126 imagens referentes as idades de 0 até 3967 horas. Para as amostras expostas a diferentes condições ambientes, foram coletadas imagens em dois tempos específicos, 0 e 24 horas.

3.3.3 Tratamento dos dados

As rotinas utilizadas durante a etapa de tratamento dos dados foram obtidas no *website* da *MathWorks - Makers of MATLAB and Simulink* disponível no endereço eletrônico <https://www.mathworks.com/> e estão apresentadas no apêndice, seção A.1.

Além disso, foram usados o pacote de análise de imagens *Image Processing Toolbox* do *MATLAB* (Mathworks, Natick, Massachusetts, EUA, v. 7.9.0.529). Em seguida, os dados foram tratados usando o pacote *PLS-Toolbox* (*Eigenvector Research*, Manson, Washington, EUA, v. 5.2.2).

As imagens foram importadas para o espaço de trabalho do *software* utilizando o comando ‘*imread*’ e obteve-se um arranjo tridimensional (4000x3000x3), que representa uma imagem de 4.000 por 3.000 *pixel* nos três canais do sistema RGB.

O estudo considerou 9 sistemas de cores: RGB, HSV, CYMK, HSL, L*a*b, YCbCr, YUV, YIQ e XYZ. As linhas de comandos usados na área de trabalho do *Matlab* para obtenção dos diferentes sistemas de cores são apresentados no apêndice, seção A.1, dessa dissertação.

3.3.3.1 Determinação da região de interesse (ROI) nas imagens digitais

Inicialmente, foi feito um recorte nas imagens, sendo que o mesmo foi orientado ao redor da região central da mancha de sangue com dimensões de 501x501x3. Em seguida, o novo arranjo que representa o recorte da imagem, no sistema RGB, foi separado em três matrizes, com as mesmas dimensões, representando os canais: vermelho (R, *red*), verde (G, *green*) e azul (B, *blue*). O esquema dessa metodologia está apresentado na Figura 39.

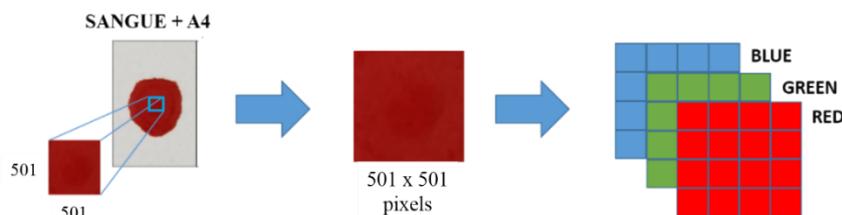


Figura 39. Metodologia para obter a região de interesse (ROI) da mancha de sangue no sistema RGB

Após obter as matrizes dos três canais, dois vetores com dimensões de 1×256 colunas foram criados para cada canal, empregando o comando *'imhist'*. O primeiro vetor contém os valores crescentes de luminosidade para o sistema RGB, que estão associados a escala que varia de 0 a 255. O segundo, possui os valores das frequências absolutas associados aos valores de luminosidade do primeiro vetor e que estão presentes na matriz de dados (*pixel x pixel*) de cada canal.

Em seguida, os vetores com as frequências absolutas de cada canal foram agrupados em uma nova matriz de dados (256×3). Após isso, foi aplicada a função *'sum'* que soma todos os elementos presentes em cada coluna e gera um vetor linha (1×3) com o somatório da frequência de cada canal, isto é, a frequência total.

O somatório da frequência absoluta está relacionado com a quantidade de *pixel* da imagem, uma vez que para cada *pixel* é atribuído um valor para R, G e B. Logo, para o recorte em questão, foi obtido nas três colunas o número 251.001, que está relacionado com a quantidade total de *pixel* na imagem ($501 * 501 = 251.001$). Por fim, o valor de cada elemento na matriz com as frequências absolutas (256×3) foi dividido pelo somatório do respectivo canal e obteve-se assim a matriz com as frequências relativas (256×3).

Para facilitar a construção dos modelos de previsão, os vetores de cada canal do sistema RGB foram reorganizados e agrupados de maneira que, para cada amostra, foi gerado um vetor linha com 768 colunas, ou seja, com cada canal contribuindo com 256 valores.

Após essa etapa, foi obtido a média do valor de luminosidade, para cada canal, tomando como base a matriz de dados que representa o recorte no sistema RGB. Para isso foi usado o comando *'reshape'* que reorganiza a matriz de cada canal (*pixel x pixel*) em uma nova matriz cujas dimensões foram previamente escolhidas com o intuito de facilitar o tratamento dos dados.

Neste caso, foram obtidos vetores linhas de dimensões 251001×1 para cada canal, no sistema RGB, e estes foram agrupados em uma nova matriz de dimensão 251001×3 . Em seguida, foi calculado o valor médio para cada coluna usando a função *'mean'*. O esquema das etapas mencionadas anteriormente estão apresentadas na Figura 40.

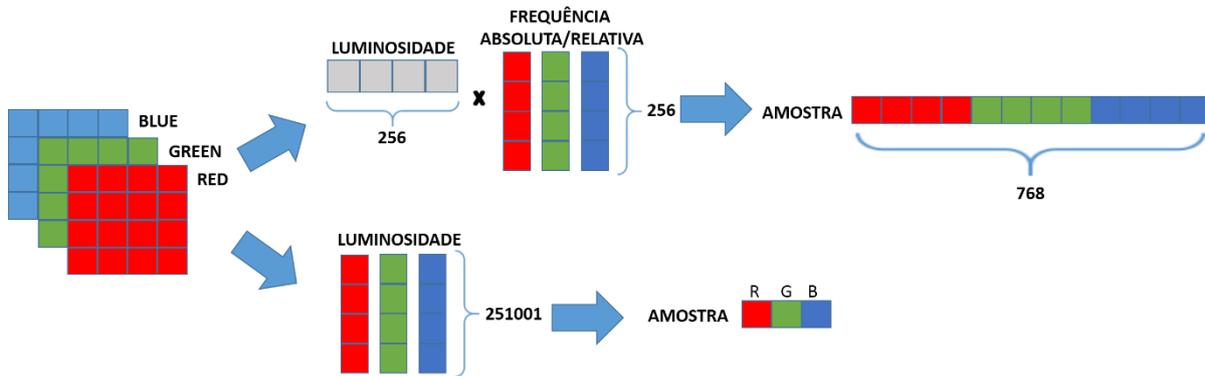


Figura 40. Esquema da metodologia para o tratamento das imagens no sistema RGB

Esse mesmo procedimento foi aplicado para os oito sistemas de cores estudados e para as 159 imagens obtidas, sendo que no fim, foram obtidos dois conjuntos de dados.

O primeiro contém uma matriz (159x28) com os valores médios, relacionados ao recorte e aos 28 canais oriundos dos 9 sistemas de cores. O segundo é dividido em nove matrizes (159x768), para cada sistema de cor, relacionadas ao recorte, exceto o CMYK, que foi obtido uma matriz com 1024 colunas, já que o sistema possui 4 canais ($256 \cdot 4 = 1024$).

O uso da região de interesse da imagem foi comparado com o uso da imagem inteira, sem recorte. Para esse fim, os histogramas da ROI e da imagem inteira das manchas de sangue foram obtidos e comparados. A influência do substrato nas imagens sem o recorte foi corrigida empregando a ferramenta de edição de imagens, *PIXLR EDITOR*, disponível em <https://pixlr.com/editor/> para retirar a cor do fundo. Após retirar a cor de fundo, o tratamento das imagens foi o mesmo descrito anteriormente, embora somente o segundo grupo de matrizes foi obtido.

3.3.3.2 Construção do modelo de regressão

A matriz \mathbf{X} foi construída como descrito anteriormente, com dimensões de 159x28, sendo 159 imagens obtidas das manchas coletadas e 28 valores médios de luminosidade para os 9 sistemas de cores.

O vetor \mathbf{y} foi construído usando métodos diferentes: as idades das manchas em minutos e em horas e também fazendo o logaritmo da base 10 tanto na idade em horas quanto em minutos. Após isso, foram feitos modelos usando a matriz \mathbf{X} e os diferentes

vetores \mathbf{y} construídos, sendo que os mesmos foram comparados usando os seguintes parâmetros: RMSEC, RMSECV, RMSEP e coeficiente de correlação.

Para a construção do modelo de regressão empregando o PLS, as amostras foram separadas em dois grupos, conjunto de calibração (70%) e validação (30%) usando o algoritmo de Kennard-Stone. O pré-processamento adotado tanto na matriz \mathbf{X} quanto no vetor \mathbf{y} foi o autoescalamento.

A escolha do número de variáveis latentes (LVs) foi baseada no menor valor de RMSECV. O erro da validação cruzada foi calculado empregando o método aleatório de seleção de amostras *random subsets*. Por fim, o modelo foi validado usando um novo conjunto de amostras, o conjunto de validação e seu desempenho foi avaliado pelo parâmetro RMSEP e o coeficiente de correlação entre as variáveis dependentes $y_{\text{observado}}$ e y_{predito} das amostra do conjunto de calibração e de validação. O modelo foi caracterizado pelo gráfico de VIP scores, que mostra a importância de cada canal dos sistemas de cores para a construção do modelo de regressão.

3.4 Resultados e Discussão

As imagens digitais obtidas, no sistema RGB, foram convertidas para os 8 sistemas de cores: HSV, CYMK, HSL, L^*a^*b , YCbCr, YUV, YIQ e XYZ. A Figura 41 apresenta as imagens digitais, no sistema RGB, para uma mancha de sangue no intervalo de 0 a 4 horas. A partir da figura é possível inferir que, no aspecto visual, as mudanças mais drásticas na coloração ocorreram na primeira hora de exposição.

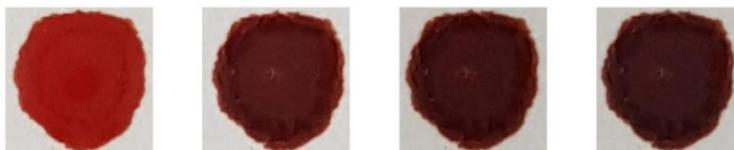


Figura 41. Imagem RGB mostrando mudanças na coloração de uma mancha de sangue de 0, 1, 2 e 4 horas (esq-dir)

Após observar as mudanças na coloração em função do tempo para o sistema RGB, foram geradas imagens para os outros oito sistemas de cores, considerando a mesma mancha de sangue e também o mesmo intervalo de tempo. As imagens geradas foram agrupadas e estão apresentadas na Figura 42.

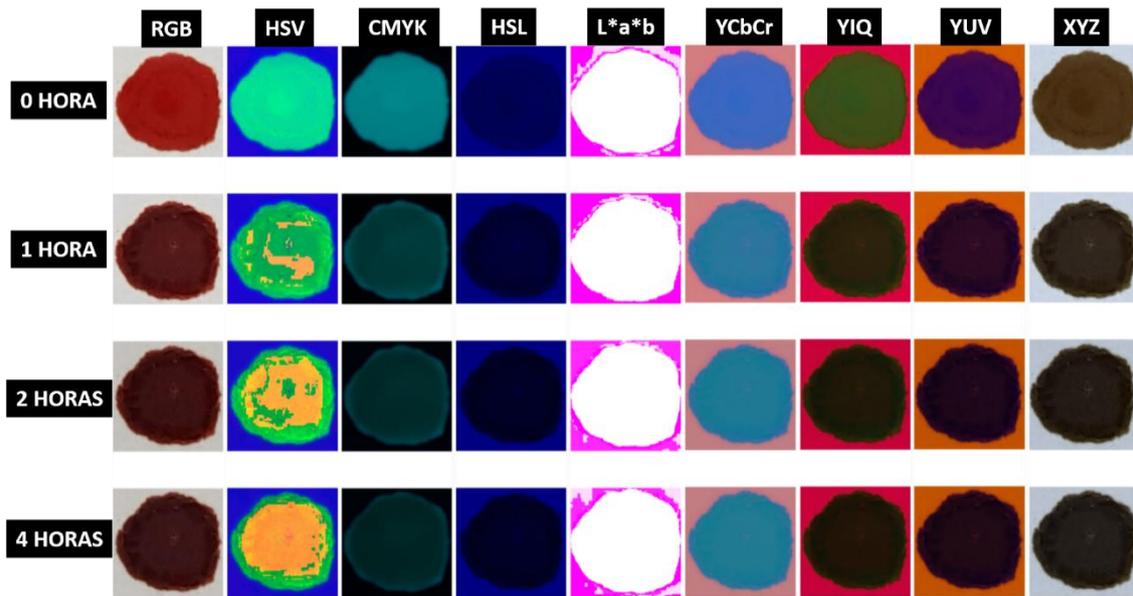


Figura 42. Imagens de uma mancha de sangue com 0, 1, 2 e 4 horas nos sistemas de cores RGB, HSV, CMYK, HSL, L*a*b, YCbCr, YIQ, YUV e XYZ

A partir da figura, nota-se que visualmente, alguns sistemas de cores evidenciaram melhor as alterações sofridas com o tempo de exposição. Neste caso, dentre os sistemas, o RGB, o HSV, o HSL e o CMYK têm destaque, uma vez que são os sistemas com as mudanças visuais mais claras.

Além disso, a partir da imagem é possível compreender que cada sistema de cor destacou um aspecto diferente da mancha em questão, corroborando com a hipótese de que os sistemas de cores e consequentemente, os canais se complementam. A partir desses resultados, pode-se entender o motivo de abordar o problema de maneira multivariada, visto que as características da imagem são descritas de uma maneira mais ampla e mais rica em informações se for empregado mais de um sistema de cor.

O uso da ROI ou da imagem sem o recorte para a construção do modelo de regressão foi avaliado fazendo a comparação dos histogramas gerados para as duas metodologias. A princípio, o método mais recomendado para construir o modelo de regressão seria usar a imagem digital da mancha sem o recorte, ou seja, utilizar todos os *pixels* presentes na imagem.

Uma das vantagens dessa abordagem seria incluir no modelo todas as alterações observadas na mancha de sangue em função do tempo de exposição, uma vez que essas mudanças não são homogêneas se considerar o sistema de cores HSV apresentado na Figura 42. Alguns pontos negativos dessa proposta é o tamanho da mancha, a inclusão da

cor do substrato e conseqüentemente, o aumento no tempo de tratamento das imagens. Para contornar o problema da cor do fundo da imagem, foi empregado um *software* de edição de imagem. A Figura 43 mostra uma das manchas de sangue com e sem fundo, após o procedimento.

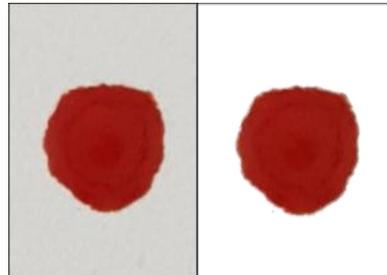


Figura 43. Mancha de sangue com 0 horas com (esq.) e sem fundo (dir.)

Para a ROI, o tamanho do corte foi escolhido de maneira que todas as imagens das manchas pudessem ser recortadas igualmente, obtendo assim recortes nas mesmas dimensões. A vantagem dessa abordagem é a simplificação no tratamento dos dados e o decréscimo no tempo de análise e além da redução de possíveis interferentes na construção do modelo de previsão, como por exemplo, a cor do substrato e o tamanho da mancha de sangue.

A comparação entre as abordagens (imagem com e sem recorte) foi feita usando os histogramas da ROI e da imagem sem o recorte para a mesma mancha no tempo de 1 hora. Os histogramas gerados para o sistema de cor RGB, HSV e CYMK são apresentados na Figura 44 até a Figura 46, de maneira que o primeiro de cada figura é relacionado ao recorte da imagem, com 251.001 *pixels* (501x501) enquanto o segundo usa a imagem sem o recorte, totalizando 12.000.000 *pixels* (3000x4000). Os demais histogramas para os outros sistemas de cores estão apresentados no apêndice seção A.2, Figura 81 a Figura 86.

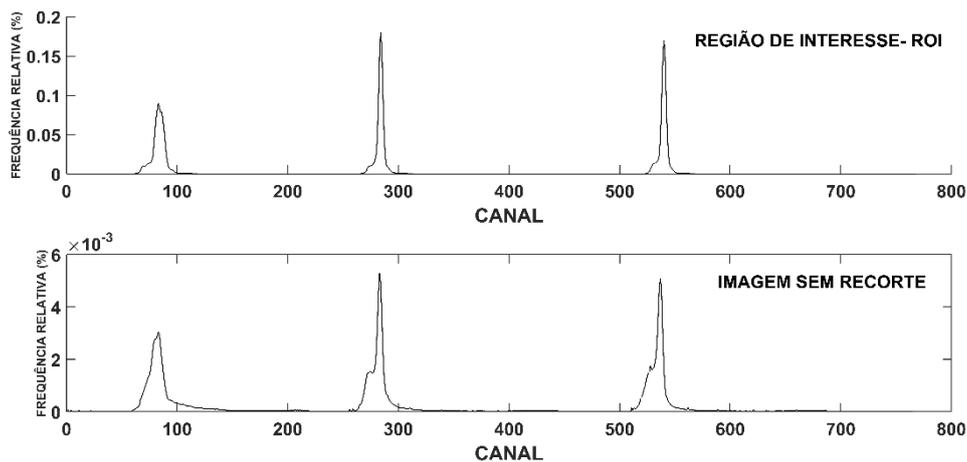


Figura 44. Histograma do recorte e da mancha completa no sistema RGB

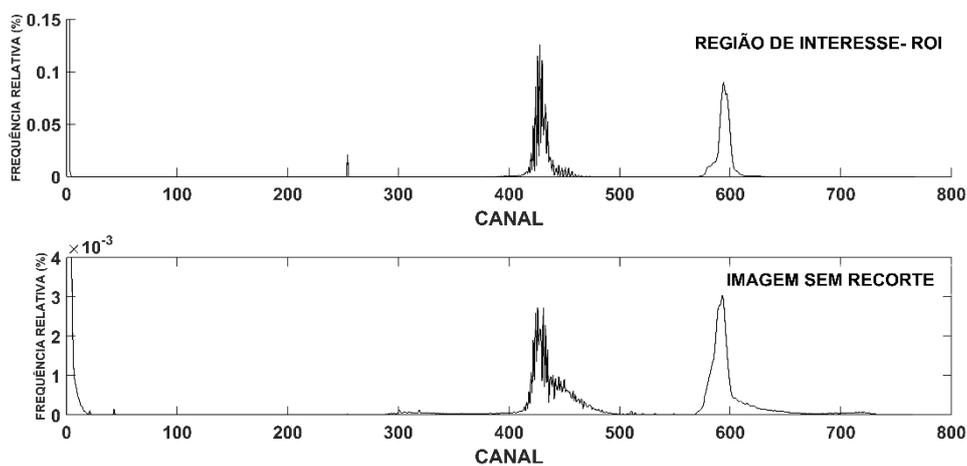


Figura 45. Histograma do recorte e da mancha completa no sistema HSV

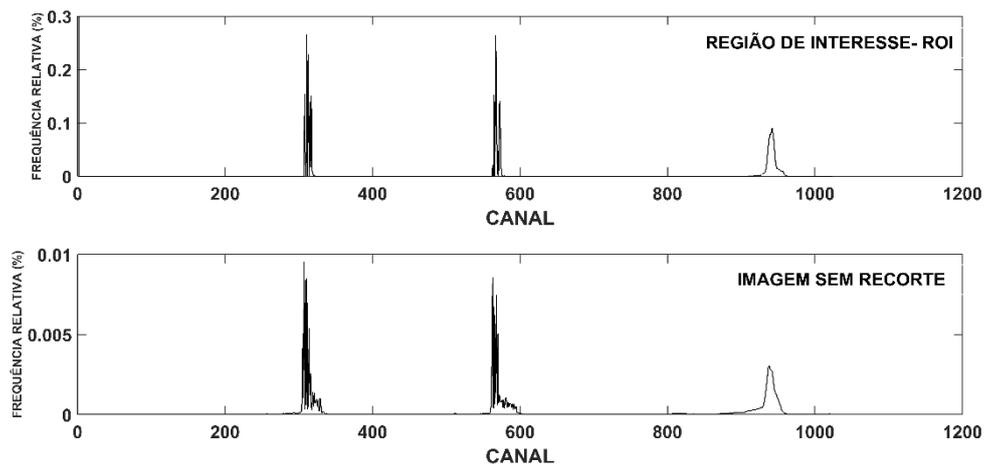


Figura 46. Histograma do recorte e da mancha completa no sistema CMYK

A partir dos resultados obtidos, é possível observar que o recorte da imagem, neste caso, se mostrou representativo, uma vez que, os histogramas em todos os sistemas

de cores estudados apresentaram perfis semelhantes aos obtidos empregando todos os *pixels* da imagem. Este resultado apresenta uma aplicação prática importante, uma vez que no local de um crime serão encontradas manchas de sangue de vários tamanhos, e a princípio, nesta situação, não seria necessário obter a imagem da mancha inteira, facilitando assim o processo de coleta das evidências e além disso, torna o tratamento dos dados mais simples e rápido.

A partir dos resultados obtidos com os histogramas, a construção do modelo de regressão foi feita com a matriz de dados da ROI, uma vez que o recorte se mostrou representativo. Após isso, todas as análises subsequentes foram realizadas empregando somente a região de interesse (ROI). Outra informação relevante que pode ser retirada a partir dos histogramas é a identificação de quais canais variam de forma mais evidente em função do tempo de exposição e a partir disso, obter informações sobre quais poderiam contribuir com um peso maior ou menor no modelo de previsão da idade das manchas de sangue.

A partir disso, usando a ROI foram construídos histogramas para cada sistema de cor considerando uma mesma mancha de sangue em três tempos de exposição diferentes: 0, 1 e 4 horas. Os gráficos para os sistemas de cores CMYK, HSV e RGB estão apresentados na Figura 47 até a Figura 49. As demais figuras para os outros sistemas de cores estão organizadas no apêndice, seção A.3, Figura 87 a Figura 92.

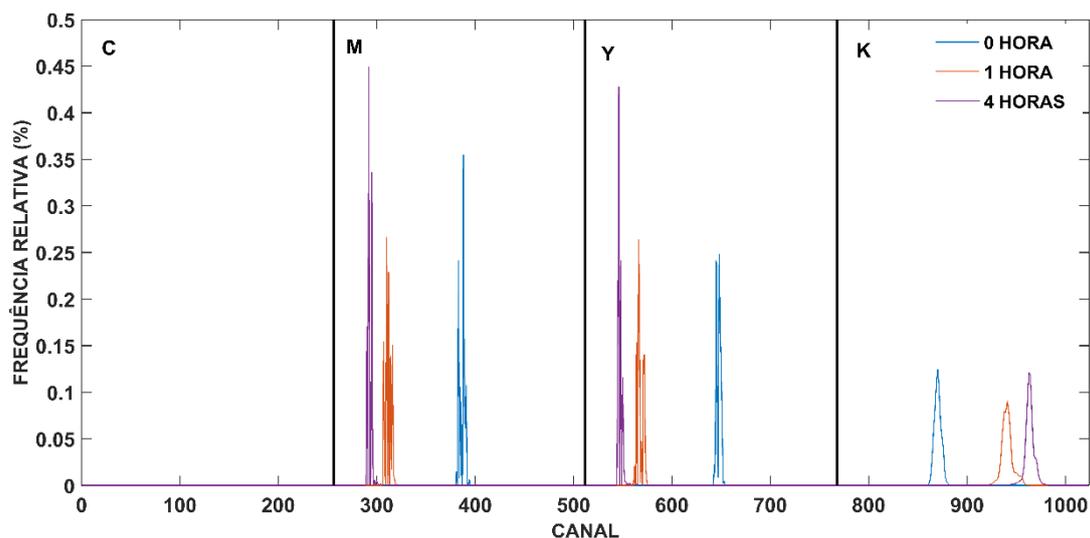


Figura 47. Histograma do recorte da imagem no sistema CMYK para 0, 1, 4 horas

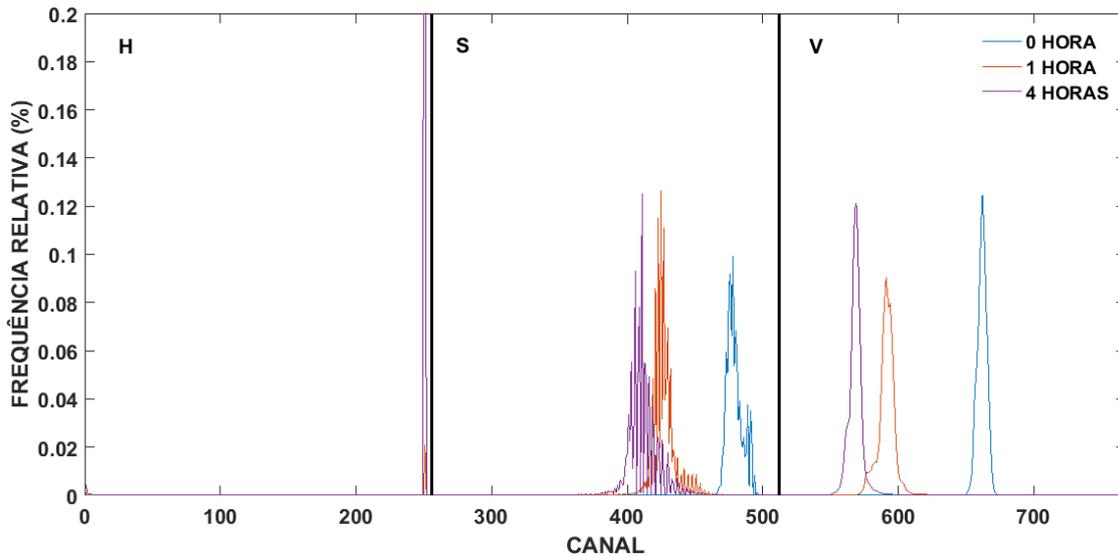


Figura 48. Histograma do recorte da imagem no sistema HSV para 0, 1, 4 horas

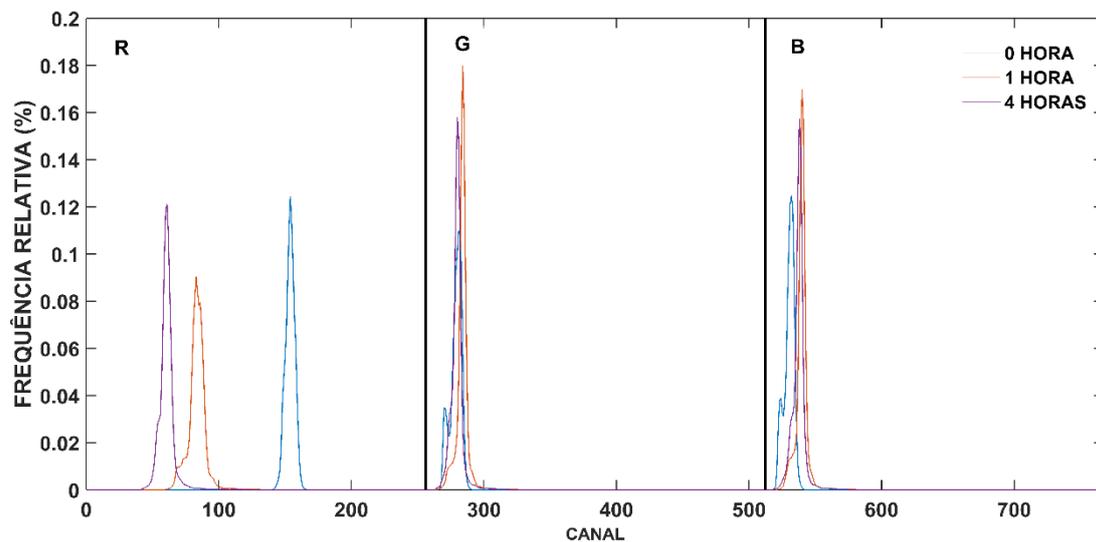


Figura 49. Histograma do recorte da imagem no sistema RGB para 0, 1, 4 hora

Após fazer uma inspeção visual nos histogramas nos intervalos de tempo considerados, fica claro que alguns canais são mais importantes do que outros, uma vez que é possível observar as mudanças nos valores de luminosidade mais frequentes em função do tempo. Essa observação fica evidente para os canais R, X, Y (YCbCr e do YUV), Y (YIQ), Y (XYZ), V (HSV), V (YUV), Cr, M, K, S (HSL e do HSV), Cb, I, Q, L (L^*a^*b), L (HSL), a, b e Z, embora para outros canais como o C, H, G, B e U não foram observadas mudanças relevantes. No entanto, analisando os sistemas de cores como um conjunto, todos eles mostraram possuir relevância na previsão da idade das manchas.

Outro aspecto que deve ser analisado antes de construir o modelo de regressão é o emprego ou não dos valores médios de luminosidade de cada canal e após isso, ser capaz de identificar se existe ou não alguma tendência entre esses valores.

da previsão do modelo. Essa observação também pode ser aplicada para as imagens obtidas usando os diferentes tipos de pisos. Para os outros substratos e também para as diferentes condições ambientes, os erros obtidos foram consideráveis, uma vez que foram obtidos erros maiores que 100 %.

Por fim, para o conjunto de amostras cujo substrato foi o papel A4 é possível notar a interferência das condições ambientes na previsão da idade das manchas de sangue. De acordo com a literatura, as condições ambientes afetam o processo de envelhecimento da mancha, uma vez que conseguem interferir nos processos químicos envolvidos. A falta de luz e o ambiente com baixa umidade retardam o envelhecimento da mancha, e partir disso, é possível compreender o motivo das manchas, nestes dois casos, apresentarem um erro de mais de 100 % na previsão se comparado as manchas que ficaram expostas em condições ambientes.^{19,30,31,33,35,37-41}

Após obter os valores médios dos recortes para todos os canais nos sistemas de cores, estes foram organizados em ordem crescente de idade e a partir disso foram gerados gráficos considerando cada sistema de cor. Os gráficos obtidos para os sistemas RGB, HSV e CMYK organizados na Figura 50 até a Figura 52. Os demais gráficos estão apresentados no apêndice , seção A.4, Figura 93 até a Figura 98.

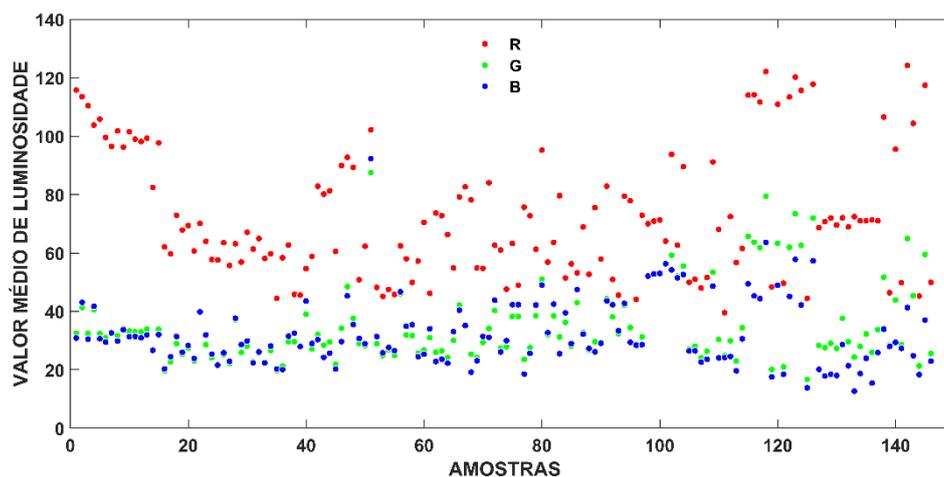


Figura 50. Valores médios de luminosidade para o sistema RGB em ordem crescente de horas

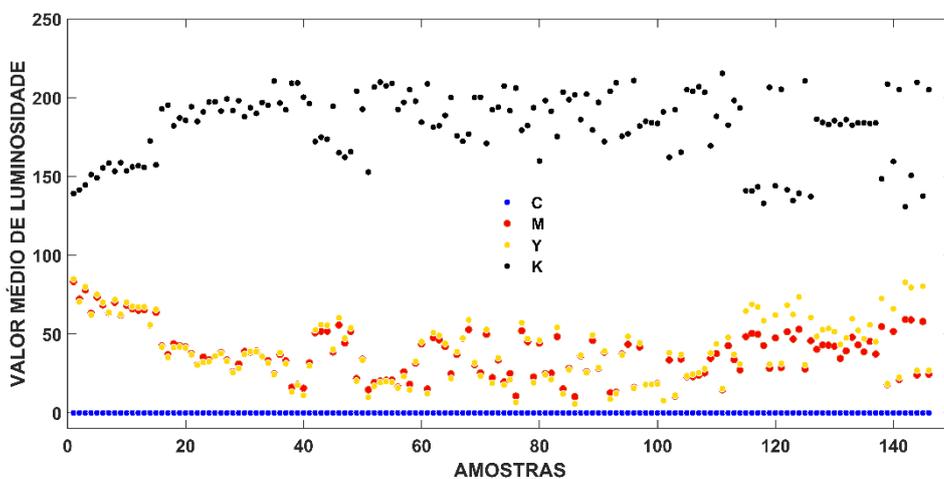


Figura 51. Valores médios de luminosidade para o sistema CMYK em ordem crescente de horas

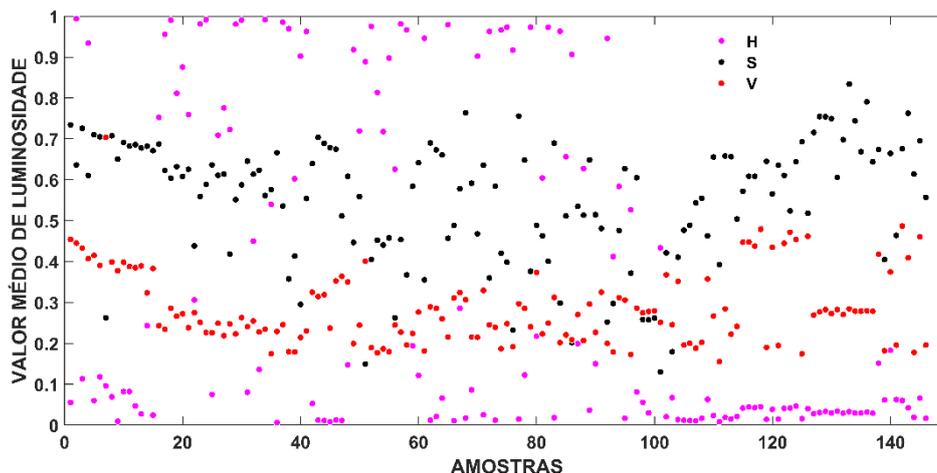


Figura 52. Valores médios de luminosidade para o sistema HSV em ordem crescente de horas

Após observar os gráficos, fica claro que existe uma relação entre os valores médios e o tempo da mancha para alguns canais, no entanto, essa tendência varia de canal para canal. Os canais K e Cb exibem uma tendência logarítmica conforme a equação 26, enquanto os canais L (L^*a^*b e HSL), a, b, Cr, M, Y (CMYK, YCbCr, YIQ, YUV, XYZ), V (HSV e do YUV), R, X, I e Q apresentam uma relação do tipo exponencial, de acordo com a equação 27.

$$y = \log_a x, \text{ para } a > 1 \quad (26)$$

$$y = a^x, \text{ para } 0 < a < 1 \quad (27)$$

Os canais restantes, neste caso o C, H, S, G, B, Z e U, aparentemente variam de forma aleatória ou não variam em função do tempo e por isso não trazem à primeira vista, informações significativas para estimar o tempo das manchas se a abordagem utilizada fosse apenas univariada.

Os dois trabalhos existentes na literatura sobre a predição da idade das manchas usando imagens digitais^{18,19} empregam apenas um canal, enquanto o diferencial desde estudo é fazer uma abordagem multivariada. De acordo com a literatura, já é de conhecimento que a diminuição da temperatura e também da umidade retardam as reações químicas, já que essas condições afetam as quantidades das espécies envolvidas no envelhecimento da mancha (oxi-Hb, meta-Hb e Hc) e conseqüentemente as mudanças na coloração em função do tempo de exposição¹⁹.

No aspecto da exposição a luz, as informações são conflitantes, embora os resultados de Thanakiatkrai *et. al.*¹⁹ sugerem que as manchas de sangue mantidas sob luz solar natural mudaram de cor mais rápido do que as armazenadas no escuro e também aquelas que foram expostas a uma lâmpada fluorescente.

Após a análise preliminar dos dados, os mesmos tiveram que passar por uma etapa de pré-processamento, pois alguns canais possuem valores de luminosidades em escalas diferentes. Com isso, o método escolhido foi o autoescalamamento, uma vez que o mesmo atribui pesos iguais para todas as variáveis. A escala dos valores máximo e mínimo para alguns canais estão apresentada na Tabela 4.

Tabela 4. Valores máximos e mínimos de luminosidade dos canais para os seguintes sistemas de cores: RGB, CMYK, HSV, HSL, L*a*b e XYZ

Sistema de cor	Canal	Escala		Sistema de cor	Canal	Escala	
		Min.	Máx.			Min.	Máx.
RGB	R	0	255	HSL	H	0	360
	G	0	255		S	0	100
	B	0	255		L	0	100
CMYK	C	0	100	L*a*b	L	0	100
	M	0	100		a	-128	128
	Y	0	100		b	-128	128
	K	0	100				
HSV	H	0	360	XYZ	X	0	100
	S	0	100		Y	0	100
	V	0	100		Z	0	100

Fonte: Colorizer.org, disponível em <http://colorizer.org/>

Antes da construção do modelo de regressão, foi feita uma análise exploratória, empregando a PCA. A matriz de dados, \mathbf{X} , foi construída organizando as amostras na ordem crescente de tempo, de forma proposital, já que se houvesse alguma tendência natural nos dados, a mesma poderia ser observada nos escores de alguma das componentes principais relevantes (PCs), ou seja, naquelas que apresentassem uma quantidade de variância explicada significativa.

O modelo foi construído com 3 PCs com uma variância explicada acumulada de 94,91%: PC1 (58,04%), PC2 (27,62%), PC3 (6,1%). A Figura 53 mostra o gráfico usado para a escolha do número de PCs.

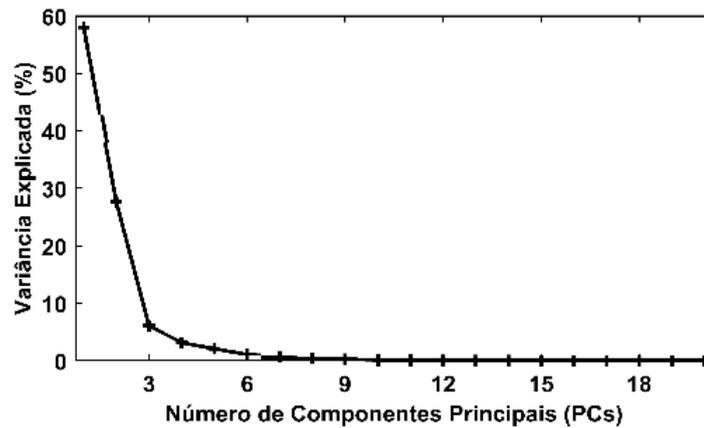


Figura 53. Número de componentes principais (PCs) em função da variância explicada

O gráfico de escores de PC1 contra PC2 é apresentado na Figura 54A e a partir dele é possível observar que não houve distinção entre os sexos. Já o gráfico de escores da PC1 com 58,04% (Figura 54B) apresentou uma tendência relacionada com o tempo de exposição da mancha de sangue.

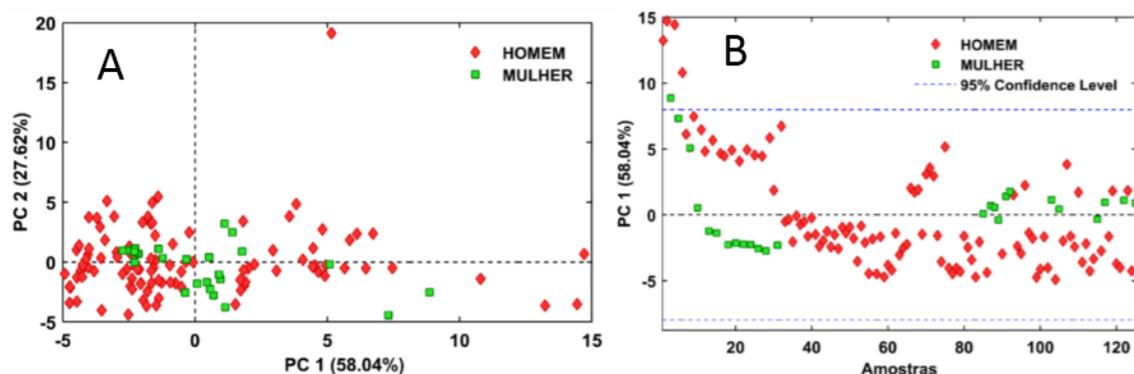


Figura 54. Escores de PC1 em função das amostras (esq.) e Gráfico de escores PC1 contra PC2 (dir.)

Após a análise exploratória, partiu-se para a construção do modelo de regressão empregando o PLS. Para isso foi considerado o valor médio da luminosidade baseado na ROI da imagem da mancha de sangue. Para obter o conjunto de calibração e de validação, foi empregado na matriz \mathbf{X} e no vetor \mathbf{y} o algoritmo de Kennard-Stone.

No entanto, na matriz \mathbf{X} foram excluídas três colunas, isto é, três variáveis que se tratavam de canais repetidos, ou seja, eles apresentaram exatamente o mesmo valor médio de luminosidade. Os canais que apresentaram os mesmos valores foram:

- 1) **H** do **HSL** e o **H** do **HSV**
- 2) **S** do **HSL** e o **S** do **HSL**
- 3) **Y** do **YCbCr** e o **Y** do **YUV**

A partir das diferentes propostas usadas para fazer o vetor y , foram construídos vários modelos, sendo que os mesmos foram comparados usando os seguintes parâmetros: RMSEC, RMSECV, RMSEP, número de LVs e coeficiente de correlação para a calibração e para a validação. O número de variáveis latentes foi escolhido para cada modelo analisando o gráfico dos valores de RMSECV contra o número de variáveis latentes (LVs), com o intuito de selecionar o número que apresentasse o menor erro de validação cruzada (valor de RMSECV) e uma porcentagem da variância explicada adequada.

A validação cruzada foi feita usando amostras retiradas do conjunto de calibração de maneira aleatória (*random subsets*). Os resultados obtidos são apresentados na Tabela 5.

Tabela 5. Parâmetros dos modelos multivariados calculados usando o espectro bruto, o espectro processado e os valores em Y, em horas e log(horas)

Parâmetros	y (horas)	y [log(horas)]	y (minutos)	y [log(minutos)]
X %	95,38	95,6	95,38	93,06
Y %	72,58	88,85	72,58	89,14
RMSEC	821,44	0,39	49286	0,39
RMSECV	898,33	0,43	54154	0,43
RMSEP	806,99	0,42	48419	0,41
LVs	4	4	4	4
R²_{cal}	0,73	0,89	0,73	0,89
R²_{val}	0,67	0,85	0,67	0,86

A partir dos resultados obtidos, os modelos construídos usando o vetor y em logaritmo de minutos e em logaritmo de horas apresentaram os melhores resultados, no entanto, visualmente o uso dos valores em logaritmo de minutos foi escolhido como o mais adequado, pois o gráfico de y predito como y medido teve início no 0 (log 1=0, 1 minuto) enquanto usando valores em horas, o mesmo tinha início em valores negativos.

Portanto, o modelo foi construído usando 4 LVs com uma variância explicada na matriz X de 93,06% e no vetor y de 89,14%. O número de variáveis latentes foi escolhido baseando-se nas informações do gráfico de variáveis latentes contra o RMSEC e o RMSECV apresentado na Figura 55.

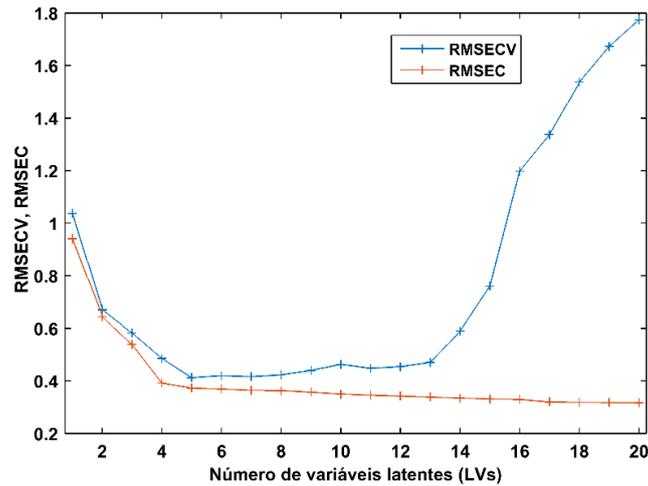


Figura 55. Gráfico de variáveis latentes contra valores de RMSEC e RMSECV

Após a escolha do número ideal de variáveis latentes (LVs), foi feita uma busca por amostras anômalas dentro do modelo, isto é, amostras que tiveram um comportamento diferenciado das demais e que podem causar prejuízos no momento da interpretação dos resultados. A identificação de *outlier* no modelo PLS é feita observando os gráficos de Resíduos Q contra T^2 de *Hotteling* além do gráfico de Resíduos em Y contra o *leverage* para o modelo construído.

Os gráficos estão apresentados na Figura 56, sendo que foram necessárias três rodadas de retirada das amostras anômalas (setas vermelhas) e os valores dos parâmetros RMSEC, RMSECV e R^2_{cal} , a cada passo, também estão apresentados na Figura 56

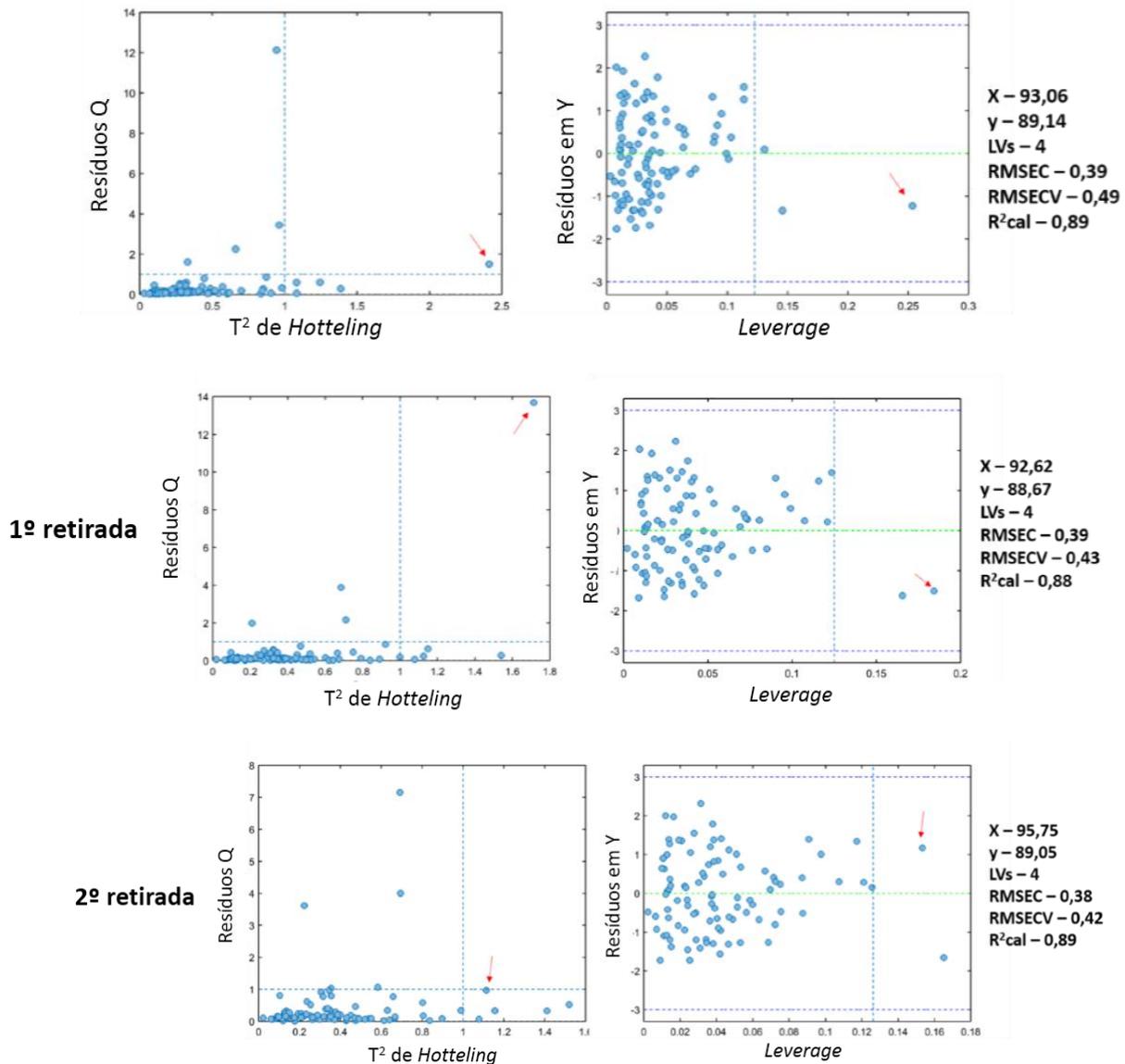


Figura 56. Gráfico de T^2 de Hotelling contra os Resíduos Q e gráfico de *leverage* contra Resíduos em Y nas 1º e 2º retirada de outlier

De acordo com a literatura, é recomendado apenas três rodadas de exclusão de outlier e que o número de amostras não ultrapasse $1/8$ do número total de amostras. É possível notar pelos gráficos a existência de algumas amostras com um alto valor de *leverage* e T^2 de Hotelling, ou seja, amostras importantes para o modelo, no entanto, também é possível observar amostras com valores consideráveis de resíduos Q e de resíduos em Y, fato que indica que a variância da respectiva amostra não está sendo bem modelada.

Outra informação que pode ser retirada do gráfico é o grau de importância das amostras para o modelo, uma vez que sua relevância está relacionada a altos valores de *leverage* e a baixos valores de Resíduos em Y. Após a retirada de amostras anômalas do

conjunto de treinamento, o modelo foi reconstruído sendo que o mesmo não apresentou amostras anômalas (Figura 57).

3ª retirada

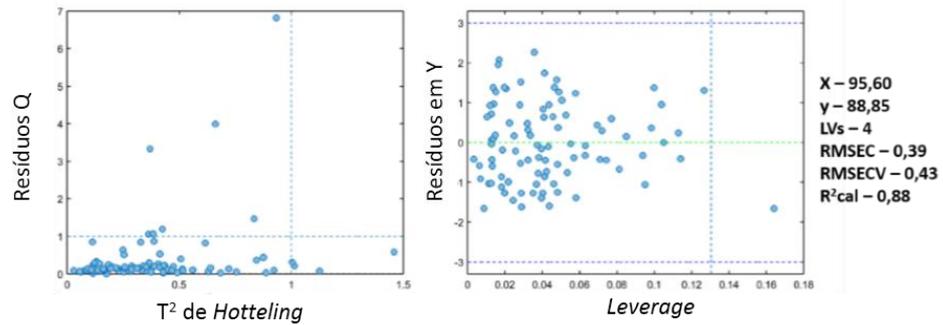


Figura 57. Gráfico de T^2 de Hotelling contra os Resíduos Q e gráfico de *leverage* contra Resíduos em Y para o modelo final

O modelo final foi construído com 4 LVs com uma variância explicada na matriz X de 95,6 % e no vetor y de 88,85 % além disso o modelo apresentou os seguintes valores para RMSEC, RMSECV e RMSEP: 0,39, 0,43 e 0,41, respectivamente. O gráfico de y medido contra y previsto está apresentado na Figura 58. O coeficiente de correlação entre os valores medidos e os valores previstos para o conjunto de calibração foi de 0,89 e para a validação foi de 0,86.

Além disso, na Figura 59 estão apresentados os erros absolutos para todas as amostras presentes no conjunto de calibração e de validação que variaram entre 1 a -1. Observa-se também que não existe nenhuma tendência sistemática na distribuição dos resíduos.

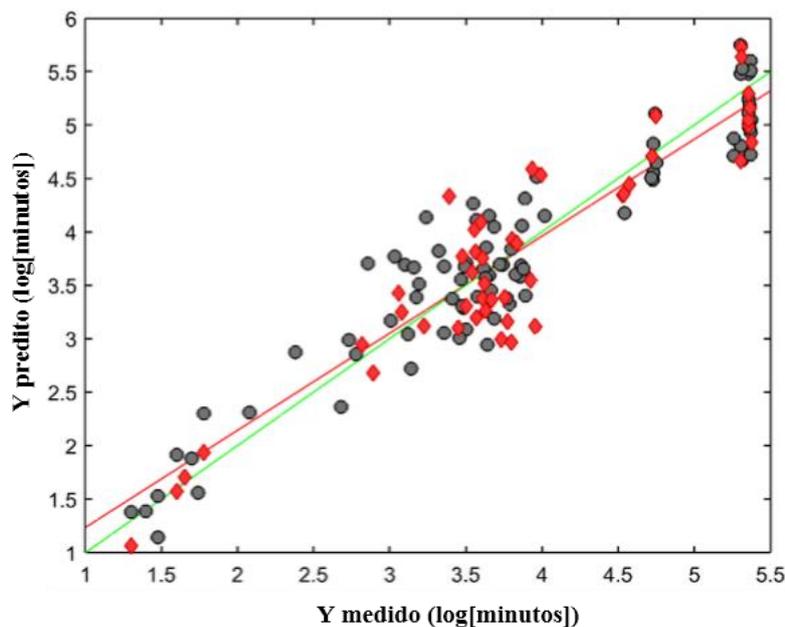


Figura 58. Gráfico dos valores y medidos contra os y preditos

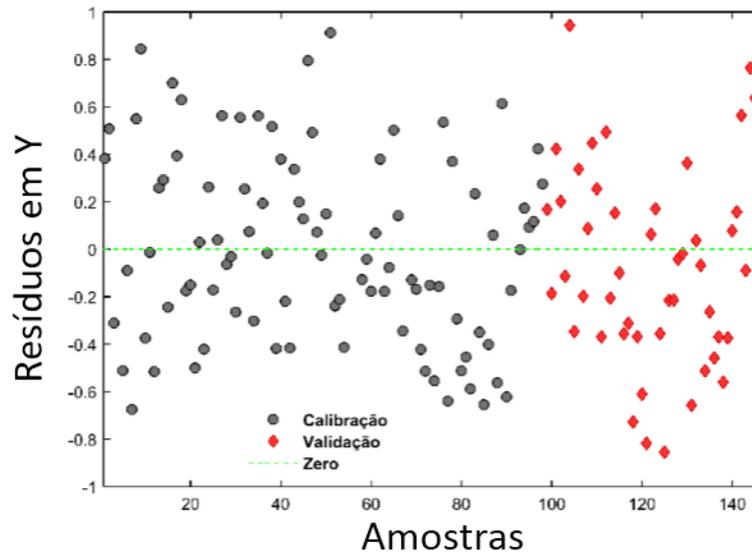


Figura 59. Resíduos em Y para as amostras do conjunto de calibração e de validação

Por fim, o gráfico de VIP scores foi gerado, e a partir dele é possível observar o grau de importância das variáveis, que neste caso, representa a relevância de cada canal dentro do modelo. O gráfico de VIP scores está apresentado na Figura 60 e os canais mais importantes estão destacados em vermelho.

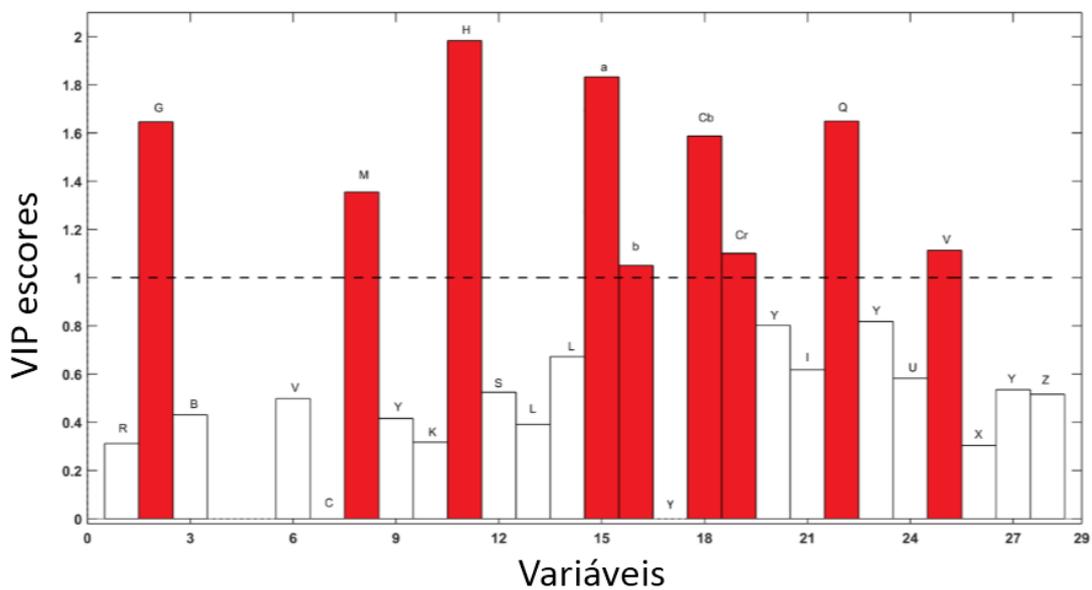


Figura 60. Gráfico de VIP scores para o modelo PLS construído com 4 VLs

A informação obtida do gráfico de VIP scores é diferente da obtida na Figura 50 até a Figura 52, uma vez que esse gráfico é gerado considerando a importância do conjunto de canais dentro do modelo e não considerando a relevância de cada canal separadamente. A partir do gráfico de VIP scores, é possível notar que existem canais mais importantes (>1) e outros que não possuem grande relevância (<1).

Neste caso, na faixa de 0 a 3967 horas, os canais mais importantes são o G, M, H, a, b, Cr, Cb, Q e V. Ao analisar os sistemas de cores como um conjunto, todos podem ser considerados importantes, exceto o XYZ.

Além disso, uma fonte de informação adicional é fazer a comparação entre os resultados obtidos a partir do gráfico de VIP escores com os resultados obtidos considerando os histogramas e também os valores médios para todos os canais em função das horas iniciais apresentados na Figura 47 até a Figura 49 e nas Figura 50 até a Figura 52.

O intuito dessa comparação é perceber a importância de cada canal usando duas abordagens diferentes, uma univariada, isto é, considerando os valores médios de cada canal com as obtidas fazendo uma análise multivariada. Os canais mais importantes com base nos gráficos considerando o VIP escores, os histogramas e valor médio estão listados abaixo:

VIP escores: G, M, H (HSV e HSL), a, b, Cb, Cr, Q e V (YUV).

Histogramas usando frequência relativa: S (HSL e HSV), Y (YUV e YCbCr), R, V (HSV), M, K, Y (CMYK), L (L^*a^*b), L (HSL), a, b, Cb, Cr, Y(YIQ), I, Q, V (YUV), X e Y (XYZ).

Valor médio de luminosidade: Y (YCbCr e YUV), R, V (HSV), M, K, Y (CMYK), L (HSL), L (L^*a^*b), a, b, Cb, Cr, Y (YIQ), I, Q, V (YUV), X e Y (XYZ).

Após fazer esta análise, foi possível identificar os canais que apresentaram alguma tendência natural entre os dados considerando as informações obtidas dos gráficos dos histogramas e dos valores médios. Os canais identificados foram o Y (YUV e YCbCr), R, V(HSV), M, Y (CMYK), K, L (HSL), L (L^*a^*b), Cb, a, b, Cr, Y (YIQ), I, Q, V (YUV), X e Y (XYZ), embora o canal S apresentou uma tendência a partir dos histogramas que não foi visualizada nos gráficos do valor médio.

Após isso, as informações obtidas foram confrontadas com o gráfico de VIP escores obtidos para o modelo. A partir disso, foi possível notar que alguns canais que apresentaram alguma tendência natural em função do tempo também demonstraram possuir importância para o modelo multivariado, como é o caso do M, a, b, Cb, Cr, Q e V (YUV).

Outros no entanto, mesmo apresentando relação com o tempo não foram considerados importantes para o modelo multivariado na faixa de horas usada para construir o modelo, no caso, o Y (YUV e YCbCr), R, V(HSV), Y (CMYK), K, L (HSL), L (L^*a^*b), Y (YIQ), I, X e Y (XYZ).

Além disso, foram identificados canais que mesmo não apresentando relação com o tempo nas horas iniciais foram considerados importantes para o modelo, no caso, foram o G e o H (HSV e HSL).

No caso do canal H, mesmo não demonstrando possuir importância para estimar o tempo de exposição das manchas de acordo com a Figura 52 e a Figura 94 foi considerado importante no gráfico de VIP escores. Essa importância está relacionada com a variação no número de vezes que o valor de luminosidade mais frequente aparece na imagem ($pixel \times pixel$) em função do tempo de exposição para o canal H.

De fato, essa informação é importante, pois para alguns canais o valor de luminosidade mais frequente muda em função do tempo de exposição (Figura 47 a Figura 49), no entanto, para o canal H, ele se mantém o mesmo, considerando o mesmo intervalo de tempo.

A diferença, neste caso, é que a quantidade de vezes que o valor aparece na matriz original ($pixel \times pixel$) aumenta com o tempo, Figura 61, e isso impacta diretamente no cálculo da frequência relativa e também no cálculo do valor médio.

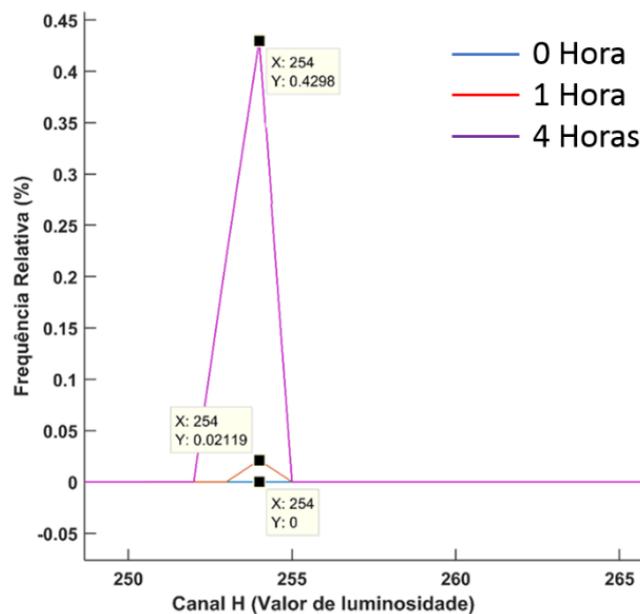


Figura 61. Variação do canal H em função do tempo de exposição de 0, 1 e 4 horas

Após isso, para avaliar a influência das condições ambientes foi feita uma análise preliminar observando as imagens geradas com alta e baixa umidade e na ausência de luz.

As imagens obtidas, no sistema RGB, foram convertidas para os oito sistemas estudados e estão apresentadas na Figura 62A. As imagens geradas foram comparadas

com as obtidas para uma mancha que ficou exposta as condições ambientes, Figura 62B, no mesmo intervalo de tempo, 24 horas.

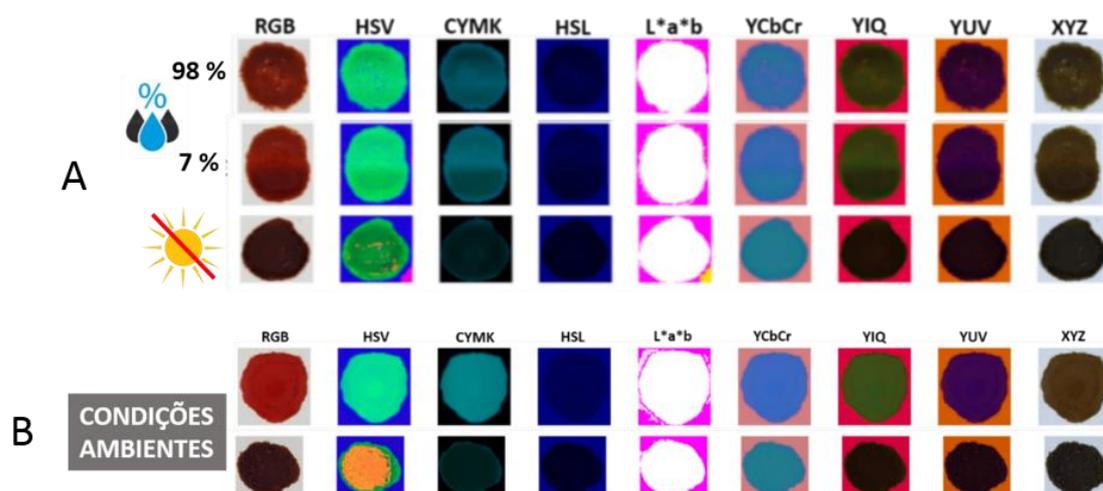


Figura 62. Mancha de sangue com cerca de 24 horas nos nove sistemas de cores para um doador em duas condições ambientes diferentes: alta e baixa umidade e baixa luminosidade

Com base na figura acima, é possível inferir que as condições de armazenamento possuem influência no envelhecimento da mancha, pelo menos no aspecto visual. Isso é verdadeiro, pois as imagens das manchas que ficaram expostas as condições ambientes apresentaram um perfil diferente daquelas expostas as condições controladas, o que pode acarretar erros elevados na previsão do modelo.

Logo, tais informações precisam ser levadas em consideração dentro do modelo de regressão afim de torná-lo mais robusto e menos sujeito a erros de previsão devido a variações na velocidade dos processos químicos envolvidos e consequentemente nas mudanças de coloração da mancha.

A interferência do tipo de substrato também foi estudada, imagens de manchas de sangue obtidas após a deposição em diferentes substratos foram incorporadas no modelo construído usando somente o papel A4. A gota de sangue foi depositada sobre os diferentes substratos e foram obtidas imagens em dois tempos diferentes, fresco (0h) e com 20 minutos. De posse das imagens, no sistema RGB, as mesmas foram convertidas para os outros oito sistemas de cores estudados e as imagens geradas estão apresentadas na Figura 63 a Figura 73, sendo que a primeira linha de cada figura se refere ao tempo 0h e a segunda ao tempo de 20 minutos.

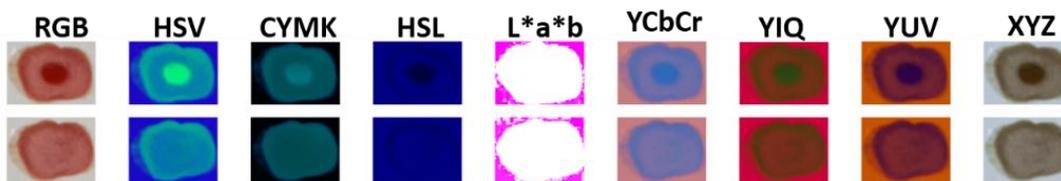


Figura 63. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato Ardósia

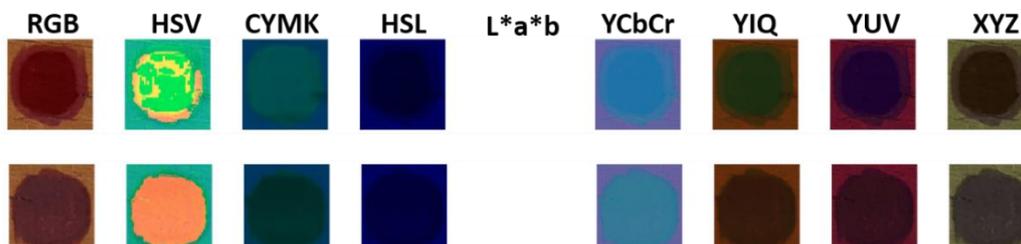


Figura 64. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato Madeira

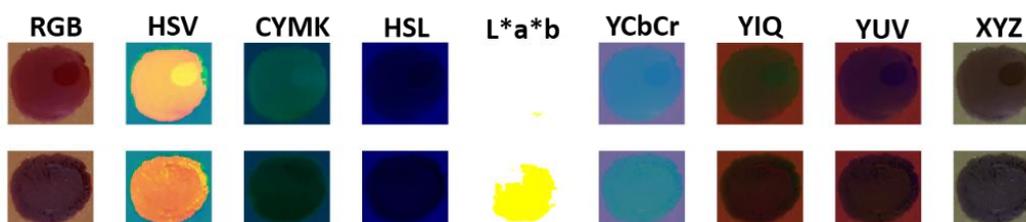


Figura 65. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato piso

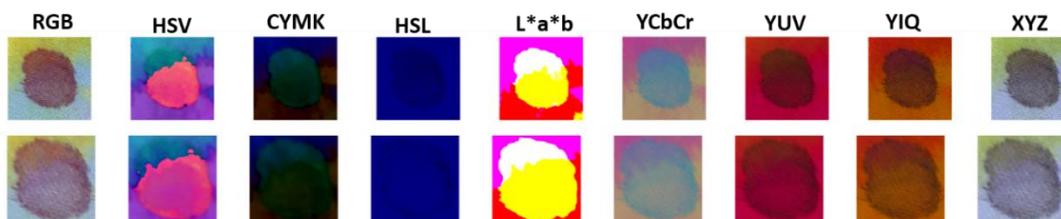


Figura 66. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 1

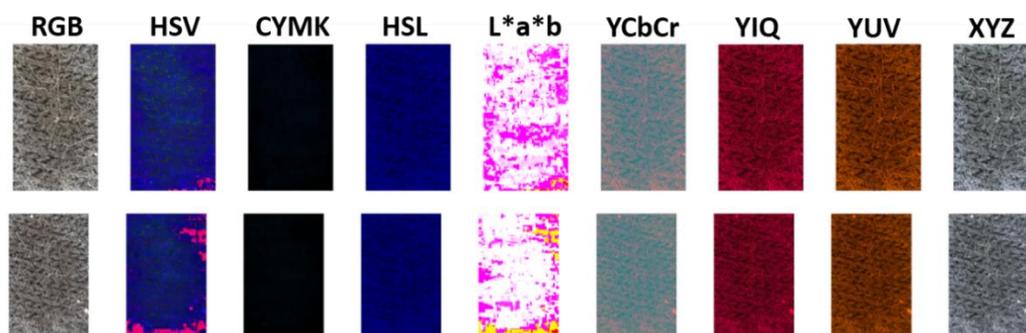


Figura 67. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 2

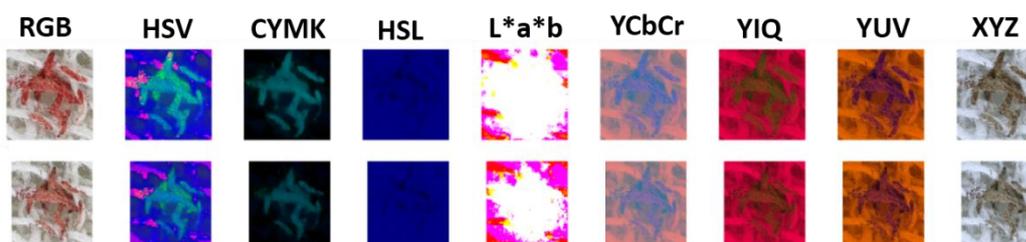


Figura 68. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 3

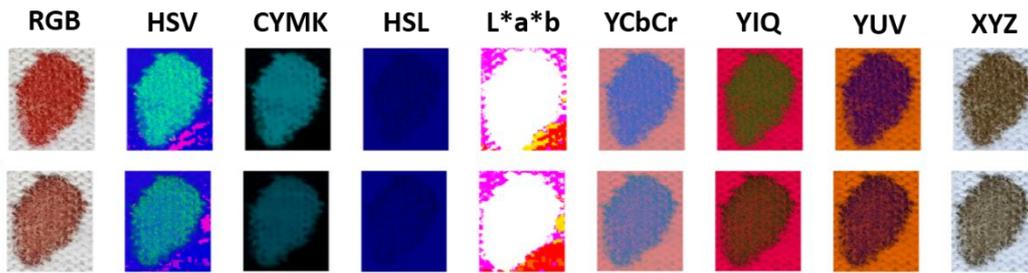


Figura 69. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 4

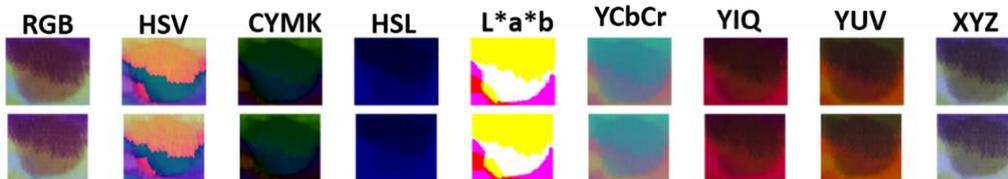


Figura 70. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 1

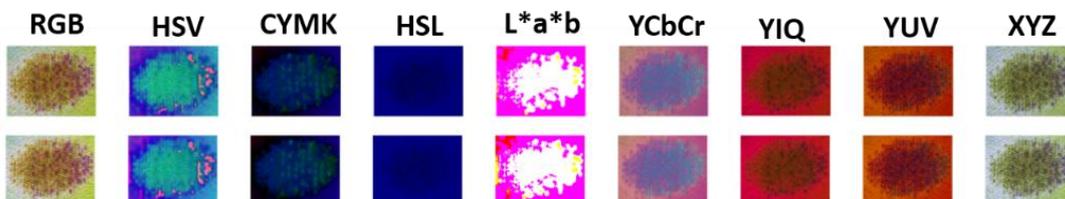


Figura 71. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato tecido 1

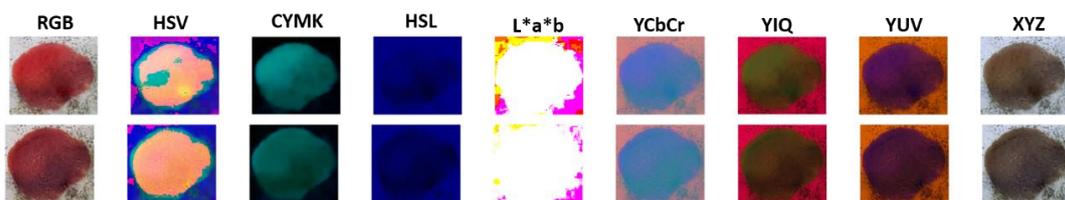


Figura 72. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato piso 2

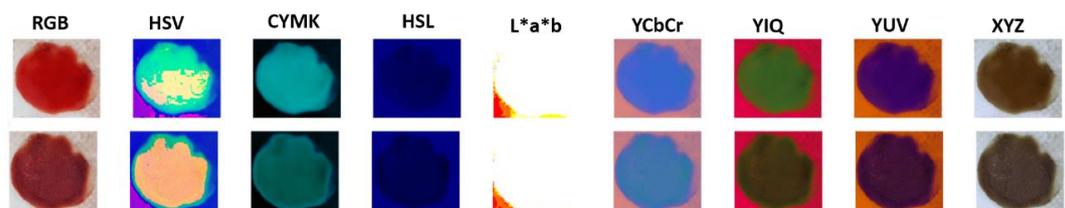


Figura 73. Mancha de sangue no intervalo de tempo de 0 a 20 minutos para o substrato piso 3

De posse das imagens, a matriz \mathbf{X} foi construída usando os valores médios de luminosidade adotando a mesma metodologia já descrita anteriormente. O conjunto de calibração e de validação foram obtidos empregando o algoritmo de Kennard-Stone, sendo que os mesmos foram incorporados aos dois conjuntos já existentes usando somente o papel A4. Nesta etapa, também foi construído modelos usando o vetor \mathbf{y} em horas, $\log(\text{horas})$, minutos e $\log(\text{minutos})$ para avaliar qual seria o melhor modelo. Os resultados obtidos estão descritos na Tabela 6.

Tabela 6. Parâmetros dos modelos multivariado usando os valores médios de luminosidade na matriz X e no vetor y, os valores em horas e log(horas), minutos e log(minutos)

Parâmetros	y (horas)	y (log(horas))	y (minutos)	y (log(minutos))
X %	98,42	98,46	98,42	98,46
Y %	68,7	89,6	68,7	89,6
RMSEC	837	0,41	50240	0,41
RMSECV	3752	2,98	259792	2,21
RMSEP	749	0,43	44962	0,43
LVs	6	6	6	6
R ² _{cal}	0,69	0,89	0,69	0,9
R ² _{val}	0,75	0,88	0,75	0,89

Após analisar os resultados obtidos, o modelo adotado foi aquele construído usando os valores do vetor y em logaritmo de minutos, pois o mesmo apresentou os menores valores tanto para RMSEC quanto para RMSEP, e além disso, foi obtido o maior valor para o coeficiente de correlação. A partir disso, a determinação do número adequado de variáveis latentes, a detecção de amostras anômalas, os gráficos de y medido contra y predito dentre outras informações serão apresentados somente para este modelo.

Logo, o modelo final foi construído usando 6 LVs de acordo com o gráfico de variáveis latentes (LVs) contra RMSEC e RMSECV apresentado na Figura 74.

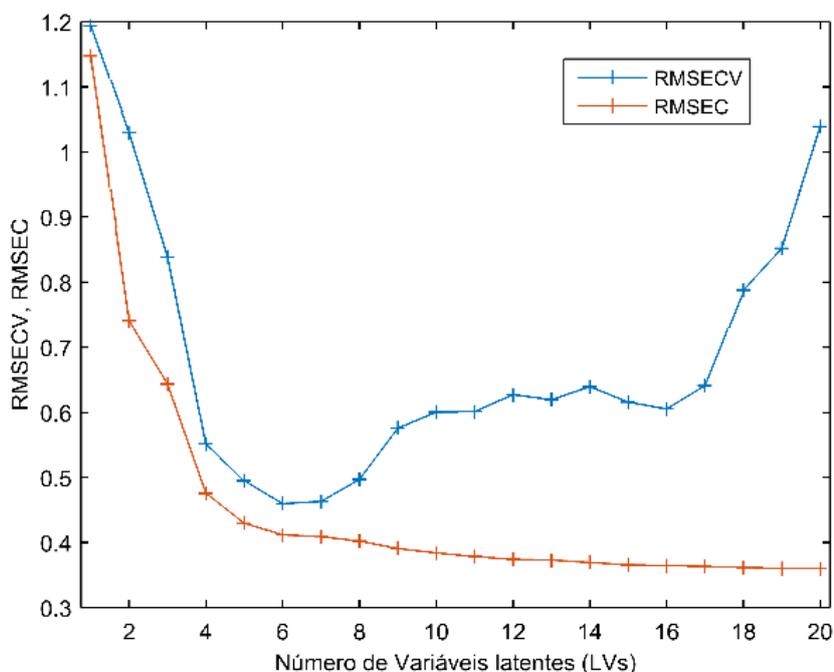


Figura 74. Gráfico de LVs contra valores de RMSEC e RMSECV

De posse do conjunto de calibração, foi feita uma análise para detectar a presença de *outlier* por meio da análise do gráfico de Resíduos em Y contra o *leverage* e também do gráfico de Resíduos Q contra o T^2 de *Hotteling*. As informações dos parâmetros do modelo, no caso, variância explicada na matriz \mathbf{X} e no vetor \mathbf{y} , RMSEC, RMSECV e R^2_{cal} para cada retirada foram anotados e também estão apresentados na Figura 75.

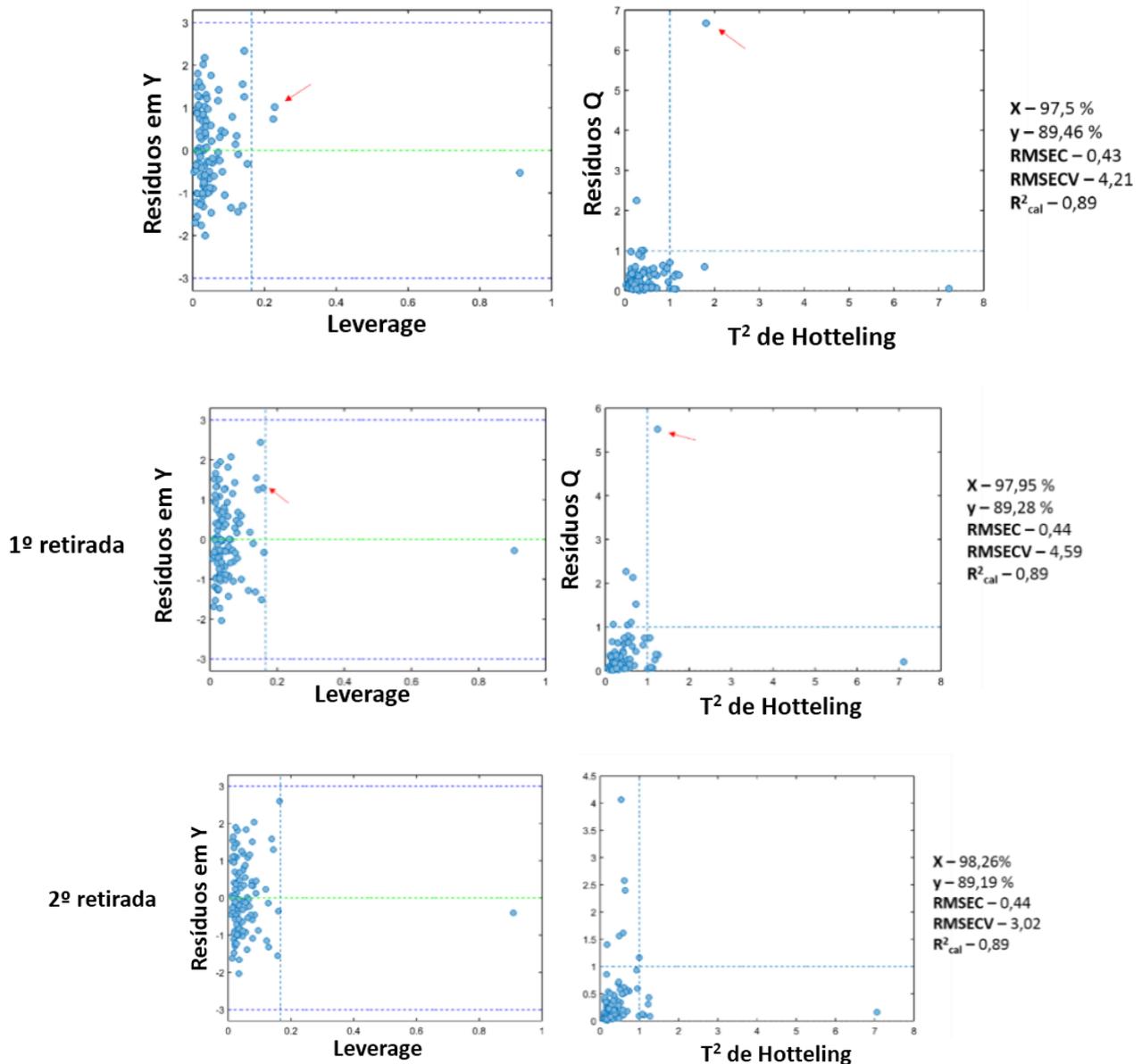


Figura 75. Gráfico de T^2 de *Hotteling* contra os Resíduos Q e gráfico de *leverage* contra Resíduos em Y nas 1ª e 2ª retirada de *outlier*

Após as retiradas das amostras anômalas, o modelo final foi construído usando 6 LVs com uma variância explicada de 98,26 % na matriz \mathbf{X} e de 89,19 % no vetor \mathbf{y} , sendo que os valores para RMSEC, RMSECV e RMSEP foram os seguintes: 0,41, 0,46 e 0,43, respectivamente. O coeficiente de correlação entre os valores preditos e medidos para o conjunto de calibração foi de 0,9 e para a validação foi de 0,89.

O gráfico dos valores medidos contra os valores preditos e os erros para cada amostra tanto do conjunto de calibração quanto o de validação foram gerados e estão apresentados na Figura 76 e Figura 77, respectivamente.

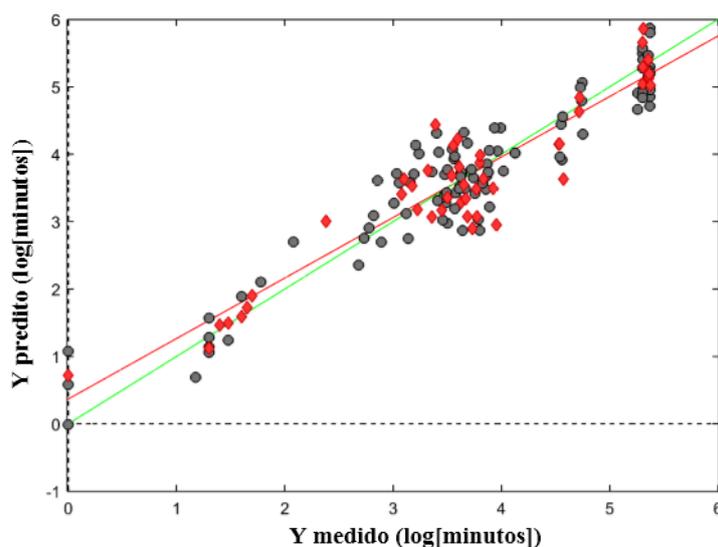


Figura 76. Gráfico de valores medidos contra preditos

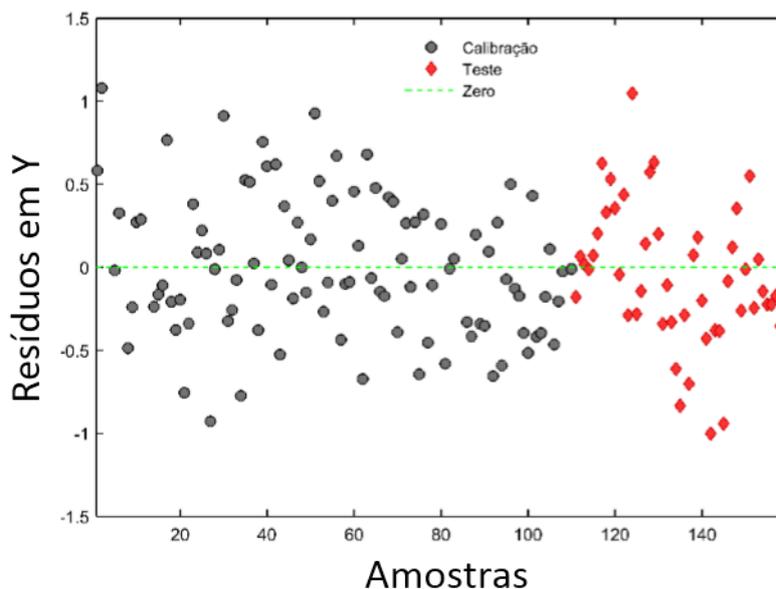


Figura 77. Resíduos em Y para as amostras do conjunto de calibração e de validação empregando diferentes substratos

A partir do gráfico de VIP escores, Figura 78, é possível notar que as variáveis do modelo sofreram modificações, uma vez que algumas variáveis ganharam ou perderam importância se comparado ao do modelo inicial usando somente o papel A4. No caso, os canais Cr e V perderam relevância dentro do modelo multivariado.

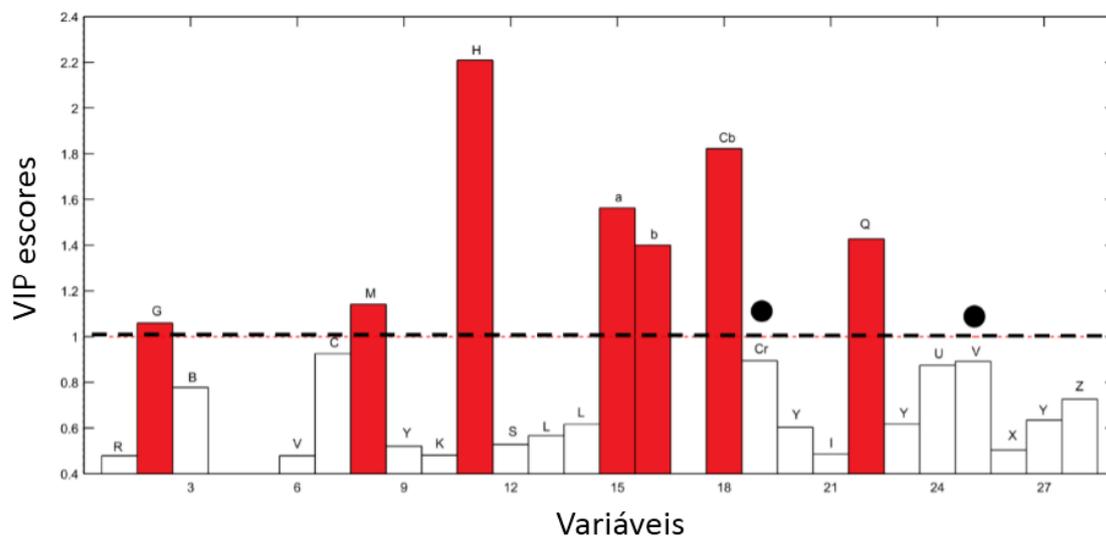


Figura 78. Gráfico de VIP escores para o novo modelo PLS incorporando diferentes substratos

Por fim, as imagens obtidas em diferentes substratos e que foram selecionadas para compor o conjunto de validação tiveram o seu erro relativo calculado. Na Figura 79 e na Figura 80 estão apresentadas as imagens juntamente com os seus erros relativos e o tipo de substrato empregado.

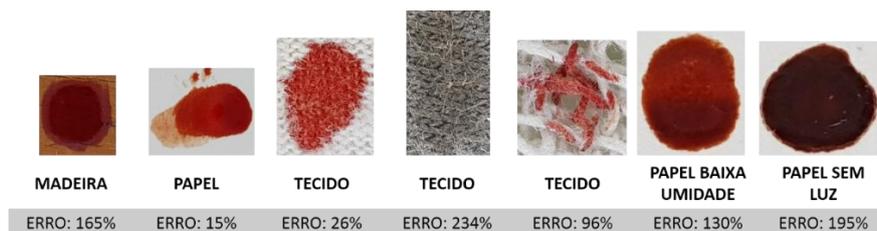


Figura 79. Manchas de sangue em diferentes substratos com seus respectivos erros relativos

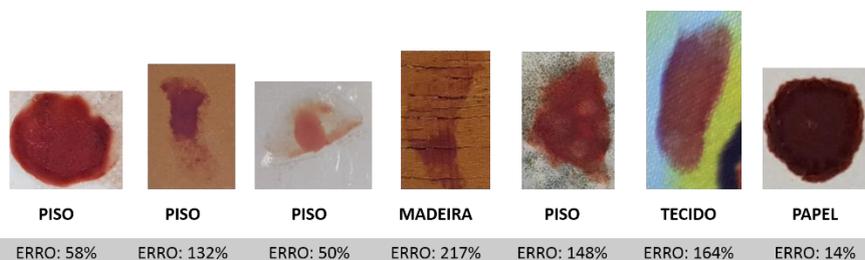


Figura 80. Manchas de sangue em diferentes substratos com seus respectivos erros relativos preditos pelo modelo

A partir das figuras é possível fazer algumas considerações em relação aos erros associados ao tipo de substrato. No caso dos tecidos, quanto mais claro, menor foi o erro

3.5 Conclusão

O uso de imagens digitais em conjunto com análise multivariada apresentou potencial para a estimar o tempo de exposição de manchas de sangue, devido seu baixo custo, simplicidade e a possibilidade de usar o próprio *smartphone* para tratamento de dados por meio de aplicativos.

O modelo de regressão usando somente o papel A4 foi construído usando 4 LVs com uma variância explicada de 95,6 % na matriz \mathbf{X} e de 88,85 % no vetor \mathbf{y} . Os valores para os erros médios do modelo, RMSEC, RMSECV e RMSEP, foram 0,39, 0,43 e 0,41, respectivamente. O coeficiente de correlação para a calibração foi de 0,89 e para a validação foi de 0,86.

Já o modelo usando diferentes substratos, incluindo o papel, foi construído usando 6 LVs com uma variância explicada de 98,26 % na matriz \mathbf{X} e de 89,19 % no vetor \mathbf{y} . Os valores para os erros médios do modelo, RMSEC, RMSECV e RMSEP, foram 0,41, 0,46 e 0,43, respectivamente. O coeficiente de correlação para a calibração foi de 0,9 e para a validação foi de 0,89.

Os resultados obtidos nesse trabalho mostram que as condições ambientais influenciam a velocidade das reações químicas que acontecem na mancha de sangue e causam erros altos de predição. Neste caso, as amostras armazenadas em baixa umidade tiveram um erro de predição de 130% e na ausência de luz, o erro foi de 195%.

Além disso, os menores erros obtidos usando diferentes substratos foram de: 15% para o papel, 50% para o piso, 26% para o tecido de coloração clara, 164% para tecidos coloridos e 165% para madeira. Nesse caso, o modelo carece de informação e mais amostras devem ser adicionadas ao conjunto de calibração para aumentar sua capacidade preditiva.

CAPÍTULO 4

4. CONCLUSÃO GERAL

O uso da espectroscopia Raman e de imagens digitais para acompanhar as mudanças químicas e físicas que ocorrem nas manchas de sangue em função do tempo de exposição se mostrou adequado e com potencial para ser usado em laboratórios de análises forenses

O método de recuperação, ou seja, o uso do *swabs* para coletar amostras de sangue depositadas em diferentes superfícies facilita a coleta e a posterior análise das evidências. Além disso, o uso de imagens digitais se mostrou uma boa alternativa frente a outras técnicas analíticas disponíveis, principalmente devido ao custo. No entanto, é necessário aumentar o conjunto amostral, ou seja, o número de amostras com o objetivo de aumentar a variabilidade do modelo e diminuir o erro de predição.

Como perspectiva futura, é necessário aumentar o número de medidas em diferentes condições ambientes (luz, umidade e temperatura) além de diversificar e aumentar o número de doadores considerando faixas de idades e sexo diferentes. Outra possibilidade seria investigar a influência de diferentes câmeras no momento da aquisição das imagens digitais.

Além disso, uma alternativa frente a metodologia proposta neste estudo seria usar métodos não lineares para construir o modelo de previsão, como por exemplo, SVM (máquinas de vetores de suporte) e também construir modelos de classificação empregando faixas de horas, no caso, um modelo que consiga discriminar se uma mancha tem ou não 24 horas de exposição.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] MALLMITH, D. M. Local de crime. Secretaria da Segurança Pública, Instituto Geral de Perícias, Departamento de Criminalística. Porto Alegre - RS, p. 1-66, 2007
- [2] DOREA, L. E. C. As manchas de sangue como indício em local de crime. Livro técnico sobre procedimentos periciais em locais de crime. Editora Franco Produções. Salvador - BA, p. 1-63, 1989
- [3] MACIEL, D. R. Análise do padrão de manchas de sangue em local de crime: revisão de literatura. Monografia apresentada à Faculdade de Odontologia de Piracicaba (FOB/UNICAMP). Piracicaba - SP, p. 1-164, 2014
- [4] ALMEIDA, M. R. Espectroscopia Raman e quimiometria como ferramentas analíticas para química forense e paleontologia. Tese apresentada ao Instituto de Química da Universidade Estadual de Campinas (UNICAMP). Campinas - SP, p. 1-177, 2015
- [5] PESCHEL, O.; KUNZ, S. N.; ROTHSCCHILD, M. A.; MUTZEL, E. Blood stain pattern analysis. *Forensic Science, Medicine, and Pathology*, N° 7, p. 257-270, 2010
- [6] BOTONJIC-SEHIC, E.; BROWN, C. W.; LAMONTAGNE, M.; TSAPARIKOS, M. Forensic application of near-infrared spectroscopy: aging of bloodstains. *Spectroscopy*, N° 24, p. 42 - 48, 2009
- [7] ANDRASKO, J. The estimation of age of bloodstains by HPLC analysis. *Journal of forensic sciences*, N° 42, p. 601-607, 1997
- [8] STRASSER, S.; ZINK, A.; KADA, G.; HINTERDORFER, P.; PESCHEL, O.; HECKL, W. M. H.; NERLICH, A. G.; THALHAMMER, S. Age determination of blood spots in forensic medicine by force spectroscopy. *Forensic Science International*, N° 170, p. 8-14, 2007
- [9] ANDERSON, S.; HOWARDA, B.; HOBBSB, G. R.; BISHOP, C. P. A method for determining the age of a bloodstain. *Forensic Science International*, N° 148, p. 37-45, 2005
- [10] BREMMER, R. H.; NADORT, A.; LEEUWEN, T. G. V.; GEMERT, M. J. C.; AALDERS, M. C. G. Age estimation of blood stains by hemoglobin derivative determination using reflectance spectroscopy. *Forensic Science International*, N° 206, p. 166-171, 2011
- [11] EDELMAN, G.; MANTI, V.; RUTH, S. M. V.; LEEUWEN, T. V.; AALDERS, M. Identification and age estimation of blood stains on colored backgrounds by near infrared spectroscopy. *Forensic Science International*, N° 220, p. 239 - 244, 2012

- [12] LEDNEV, I. K.; DOTY, K. C.; MCLAUGHLIN, G. A Raman “spectroscopic clock” for bloodstain age determination: the first week after deposition. *Analytical and Bioanalytical Chemistry*, N° 408, p. 3993 - 4001, 2016
- [13] LEDNEV, I. K.; MCLAUGHLIN G. A modified Raman multidimensional spectroscopic signature of blood to account for the effect of laser power. *Forensic Science International*, N° 240, p. 88-94, 2014
- [14] GOMES J. A.; SERCHELI, M. S. Espectroscopia Raman: um novo método analítico para investigação forense em cruzamento de traços. *Revista Brasileira de criminalística*, Vol. 1, N° 1, p. 22-30, 2011
- [15] WEST, M. J.; WENT, M. J. Detection of drugs of abuse by Raman spectroscopy. *Drug Test. Anal.* 3, 532-8 (2011)
- [16] MISTEK, E.; HALÁMKOVÁ, L.; DOTY, K. C.; MURO, C. K.; LEDNEV I. K. Race Differentiation by Raman Spectroscopy of a Bloodstain for Forensic Purposes. *Analytical Chemistry*, N° 88, p. 7453-7456, 2016
- [17] PUGLIERI, T. S.; FARIA, D. L. A. Microscopia Raman aplicada à conservação e autenticação de obras de arte. Resumo expandido publicado no 7º simpósio: Convenção do Patrimônio Imaterial - 10 anos depois (2003-2013). São Paulo – SP, p. 1-11, 2013
- [18] SHIN, J.; CHOI, S.; YANG, J.; SONG, J.; CHOI, J.; JUNG, H. Smart Forensic Phone: Colorimetric analysis of a bloodstain for age estimation using a smartphone. *Sensors and Actuators B: chemical*, vol. 243, p. 221-225, 2017
- [19] THANAKIATKRAI. P.; YAODAM, A.; KITPIPIT, T. Age estimation of bloodstains using smartphones and digital image analysis. *Forensic Science International*, vol. 233, p. 288-297, 2013
- [20] BRUNI, A. T.; VELHO, J. A.; OLIVEIRA, M. F. Fundamentos da química forense: uma análise prática da química que soluciona crimes, editora millennium, 15 de maio de 2017. P.400
- [21] GETTLER, A. O. Historical development of toxicology. *Journal of forensic sciences*, v.1, n° 1, 1958
- [22] RUDIN, N.; INMAN, K.; Forensic Science timeline. Disponível em: www.forensicdna.com/Timeline020702.pdf, acesso 13 ago. 2012
- [23] GILES, H. A.; ABERD, L. D.; OXON, D. L. The ‘HIS YUAN LU’ ou instruction to coroners’, Disponível em: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2201406/pdf/procrsmed01245-0062.pdf>. Acesso em 13 ago. 2012

- [24] BELL, S. Forensic Chemistry: prentice hall, p. 696, 2005
- [25] ZARZUELA, J. L. Química Legal em: Tochetto, D. et al. Tratado de Perícias Criminalísticas. ed. Sagra-DC Luzzatto, Porto Alegre, cap. 8, p. 164, 1995
- [26] OLIVEIRA, M. F. Química forense: a utilização da química na pesquisa de vestígios de crime. química nova na escola, v.24, p. 13, 2006
- [27] BERG, J. M.; TYMOCZKO, J. L. STRYER, L. Hemoglobin: Portrait of a Protein in Action, Biochemistry, New York, p. 191, 2002
- [28] ZADORA, G.; MENŽYK, A. In the pursuit of the holy grail of forensic science - spectroscopic studies on the estimation of time since deposition of bloodstains. Trends in Analytical Chemistry, p. 1, 2018
- [29] STADLER, A. M.; DIGEL, I.; ARTMANN, G. M.; EMBS, J. P.; ZACCAI, G.; BÜLDT, G. Hemoglobin Dynamics in Red Blood Cells: Correlation to Body Temperature, Biophys, vol. 95, p. 5449, 2008
- [30] HANSON, E. K.; BALLANTYNE, J. A blue spectral shift of the hemoglobin Soret band correlates with the age (time since deposition) of dried bloodstains, PLoS ONE 5, p. e12830, 2010
- [31] MIKI, T.; KAI, A.; IKEYA, M. Electron spin resonance of bloodstains and its application to the estimation of time after bleeding, Forensic Sci. Int. vol. 35, p. 149, 1987
- [32] MATSUOKA, T.; TAGUCHI, T.; OKUDA, J. Estimation of bloodstain age by rapid determinations of oxyhemoglobin by use of oxygen electrode and total hemoglobin, Biol. Pharm. Bull. Vol. 18, p. 1031, 1995
- [33] FUJITA, Y.; TSUCHIYA, K.; ABE, S.; TAKIGUCHI, Y.; KUBO, S.; H. SAKURAI, H. Estimation of the age of human bloodstains by electron paramagnetic resonance spectroscopy: long-term controlled experiment on the effects of environmental factors, Forensic Sci. Int. vol. 152, p. 39, 2005
- [34] HANSON, E.; ALBORNOZ, A.; BALLANTYNE, J. Validation of the hemoglobin (Hb) hypsochromic shift assay for determination of the time since deposition (TSD) of dried bloodstains, Forensic Sci. Int. Genet. Suppl. Ser. vol. 3, p. e307, 2011
- [35] BREMMER, R. H.; NADORT, A.; VAN LEEUWEN, T. G.; VAN GEMERT, M. J.; AALDERS, M. C. Age estimation of bloodstains by hemoglobin derivative determination using reflectance spectroscopy, Forensic Sci. Int. vol. 206, p. 166, 2011
- [36] AGUDELO, J.; HUYNH, C.; HALÁMEK, J. Forensic determination of blood sample age using a bioaffinity-based assay, Analyst. Vol. 140, p. 1411, 2015

- [37] EDELMAN, G. Spectral analysis of blood stains at the crime scene (Doctoral dissertation), 2014
- [38] TSURUGA, M.; MATSUOKA, A.; HACHIMORI, A.; SUGAWARA, Y.; SHIKAMA, K. The Molecular Mechanism of Autoxidation for Human Oxyhemoglobin, *J. Biol. Chem.*, vol. 273, p. 8607, 1998
- [39] COLOMBO, M. F.; SANCHES, R. Hydration-dependent conformational states of hemoglobin Equilibrium and kinetic behavior. *Biophysical Chemistry*, vol. 36, p. 33, 1990
- [40] BAUER, M.; POLZIN, S.; PATZELT, D. Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains?. *Forensic Sci. Int.* vol. 138, p. 94, 2003
- [41] LARKIN, B. J. A. Bloodstain pattern analysis: scratching the surface (Doctoral dissertation)
19,30,32,34,35,37,38,39,40,41
- [42] JAMES, S. H.; KISH, P. E.; SUTTON, T. P. Principles of Bloodstain Pattern Analysis, CRC Press, Boca Raton, p. 1, 2005
- [43] LABER, T. L.; EPSTEIN, R. P. Substrate effects on the clotting time of human blood, *Can. Soc. Forens. Sci. J.* vol. 34, p. 209, 2001
- [44] PATTERSON, D. Use of reflectance measurements in assessing the colour changes of ageing bloodstains, *Nature*, vol. 187, p. 688, 1960
- [45] LI, B.; BEVERIDGE, P.; O'HARE, W. T.; ISLAM, M. The age estimation of blood stains up to 30 days old using visible wavelength hyperspectral image analysis and linear discriminant analysis, *Sci. Justice*, vol. 53, p. 270, 2013
- [46] BOTONJIC-SEHIC, E.; BROWN, C. W.; LAMONTAGNE, M.; TSAPARIKOS, M. FORENSIC Application of Near Infrared Spectroscopy: Aging of Bloodstains, *Spectroscopy*, vol. 24, p. 1, 2009
- [47] SUN, H.; DONG, Y.; ZHANG, P.; MENG, Y.; WEN, W.; LI, N.; GUO, Z. Accurate Age Estimation Of Bloodstains Based on Visible Reflectance Spectroscopy and Chemometrics Methods, *IEEE Photon. J.* vol. 9, 2017
- [48] SHINE, S. M.; SUHLING, K.; BEAVIL, A.; DANIEL, B.; FRASCIONE, N. The applicability of Fluorescence lifetime to determine the time since the deposition of biological stains, *Anal. Methods*. Vol. 9, p. 2007, 2017
- [49] DOTY, K. C.; MURO, C. K.; LEDNEV, I. K. Predicting the time of the crime: Bloodstain aging estimation for up to two years, *Forensic Chem.* vol. 5, p. 1, 2017

- [50] LI, B.; BEVERIDGE, P.; O'HARE, W. T.; ISLAM, M. The estimation of the age of a blood stain using reflectance spectroscopy with a microspectrophotometer, spectral pre-processing and linear discriminant analysis. *Forensic Science International*, N° 212, p. 198-204, 2011
- [51] UMBREIT, J.; Methemoglobin-It's Not Just Blue: A Concise Review, *Am. J. Hematol.* vol. 82, p. 134, 2007
- [52] BASKURT, O. K.; HARDEMAN, M. R.; RAMPLING, M. W.; MEISELMAN, H. J. Mechanisms of Blood Rheology Alterations. *Handbook of Hemorheology and Hemodynamics*, IOS Press, Amsterdam, p. 170, 2007
- [63 53] FERREIRA, M. M. C. *Quimiometria: conceitos, métodos e aplicações*. Campinas: UNICAMP, p. 1-493, ISBN-9788526810631, 2015
- [54] KOWALSKY, B.R. "Chemometrics: views and propositions", *J. Chem. Inf. Comp. Sci.* vol. 15, p.201, 1975
- [55] KOWASKI, B.R."Chemometrics, *Chem. Ind.*, vol. 22, p. 882, 1978
- [56] MALINOWSKI, E. R. *Factor analysis in chemistry*, 2° ed., John wiley & Sons : new York, 1991
- [57] MASSART, D. L.; VANDEGINSTE, B. G. M.; BUYDENS, L. M. C.; DE JONG, S.; LEWI, P. J.; VERBEKE, S. J. *Handbook of Chemometrics and Qualimetrics: part A*, Elsevier, Amsterdam, 1997
- [58] WOOD, S. Chemometrics: what do we mean with it, and what do we want from it? *Chemometrics and Intelligent Laboratory Systems*, vol. 30, n° 1, p. 109, 1995
- [59] IUPAC Compendium of chemical terminology-the god bock. Disponível em: <http://goldbook.iupac.org/CT06948.html>, acesso junho de 2018
- [60] BEEB, K. R. *Chemometrics: a practical guide*. Wiley-Interscience, p. 360, 1998
- [61] HOTELLING, H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, vol. 24, p. 417, 1993
- [62] GELADI, P. KOWALSKI, B. R. Partial least-squares regression: a tutorial analytical *chimica acta*, vol. 185, p. 1, 1986
- [63] BEEBE, K. B.; KOWALSKI, B. R. An introduction to multivariate calibration and analysis. *Analytical chemistry*, vol. 59, p. 1007A, 1987
- [64] ALAN, T. M.; ALAN, M. K. "Chemometrics Analysis of NMR spectroscopy data: a reiew. *Annu. Rep. NMR Spectrosc.*, vol. 54, p. 41, 2005
- [65] BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. *Chemometrics: A practical guide*, Wiley, New York, 1998

- [66] SHARAF, M. A.; ILLMAN, D. L.; KOWALSKI, B. R. *Chemometrics*, 1^o ed, John Wiley & Sons, New York, p. 332, 1986
- [67] KENNARD, R. W.; STONE, L. A. *Technometrics*. P. 11, 1969
- [68] THOMAS, E. V. *A primer on multivariate calibration analytical chemistry*, vol. 66, n° 15, p. A795, 1994
- [69] BRERETON, R. G. *Applied Chemometrics for Scientists*. Chichester, Inglaterra: Wiley, 2007
- [70] TRAN, T. N.; AFANADOR, N. L.; BUYDENS, L. M.C.; BLANCHET, L. Interpretation of variable importance in Partial Least Squares with Significance Multivariate 77 Correlation (sMC). *Chemometrics and Intelligent Laboratory Systems*. vol. 138, p.153, 2014
- [71] MARTENS, H.; JENSEN, S.; GELADI, P. *Proc. Nordic. Syrup. On Applied Statistics*. p. 205, 1983
- [72] BARNES, R. J.; DHANOA, M. S.; LISTER, S. J.; *Appl. Spectrosc.* vol. 43, p. 772, 1989
- [73] NIEDZWIECKI, M. On tracking characteristics of weighted least squares estimators applied to nonstationary system identification. *IEEE transactions on automatic control*, vol. 33, 1988
- [74] FARIA, D. L. A.; SANTOS, L. G. C.; GONÇALVES, N. S. Uma demonstraç o sobre o espalhamento inel stico de luz: repetindo o experimento de Raman. *Qu mica Nova*, vol. 20, n° 319, 1997
- [75] BOTTI, S. *et al.* Trace level detection and identification of nitro-based explosives by surface-enhanced Raman spectroscopy. *J. Raman Spectrosc*, vol. 44, p. 463, 2013
- [76] ALI, E. M. A.; EDWARDS, H. G. M.; SCOWEN, I. J. Raman spectroscopy and security applications: the detection of explosives and precursors on clothing. *J. Raman Spectrosc.*, vol. 40, p. 2009, 2009
- [77] ALI, E. M. A.; EDWARDS, H. G. M.; SCOWEN, I. J. In-situ detection of single particles of explosive on clothing with confocal Raman microscopy. *Talanta*, vol. 78, p. 1201, 2009
- [78] L PEZ-L PEZ, M.; FERRANDO, J. L.; GARC A-RUIZ, C. Dynamite analysis by Raman spectroscopy as a unique analytical tool. *Anal. Chem.*, vol. 85, p. 2595, 2013
- [79] L PEZ-L PEZ, M.; GARC A-RUIZ, C. Infrared and Raman spectroscopy techniques applied to identification of explosives. *TrAC Trends Anal. Chem.*, vol. 54, p. 36, 2014

- [80] LÓPEZ-LÓPEZ, M.; FERRANDO, J. L.; GARCÍA-RUIZ, C. Comparative analysis of smokeless gunpowders by Fourier transform infrared and Raman spectroscopy. *Anal. Chim. Acta*, vol. 717, p. 92, 2012
- [81] BRAZ, A., LÓPEZ-LÓPEZ, M., GARCÍA-RUIZ, C. Raman spectroscopy for forensic analysis of inks in questioned documents. *Forensic Sci. Int.*, vol. 232, p. 206, 2013
- [82] GUEDES, A.; ALGARRA, M.; PRIETO, A. C.; VALENTIM, B.; HORTELANO, V.; NETO, S.; ALGARRA, R.; NORONHA, F. Raman Microspectroscopy of Genuine and Fake Euro Banknotes. *Spectrosc. Lett.*, vol. 46, p. 569, 2013
- [83] SKENDEROVIĆ BOŽIČEVIĆ, M.; GAJOVIĆ, A.; ZJAKIĆ, I. Identifying a common origin of toner printed counterfeit banknotes by micro-Raman spectroscopy. *Forensic Sci. Int.* vol. 223, p. 314, 2012
- [84] ALI, E. M. A.; EDWARDS, H. G. M. Analytical Raman spectroscopy in a forensic art context: the non-destructive discrimination of genuine and fake lapis lazuli. *Spectrochim. Acta. A. Mol. Biomol. Spectrosc.* vol. 121, p. 415, 2014
- [85] ROMÃO, W.; SCHWAB, N. V.; BUENO, M. I. M. S.; SPARRAPAN, R.; EBERLIN, M. N.; MARTINY, A.; SABINO, B. D.; MALDANER, A. O. Química forense: perspectiva sobre novos métodos analíticos aplicados à documentoscopia, balística e drogas de abuso. *Química Nova*, vol. 34, nº 10, p. 1717, 2011
- [86] YU, M. M. L.; SANDERCOCK, P. M. L. Principal component analysis and analysis of variance on the effects of Entellan New on the Raman spectra of fibers. *J. Forensic Sci.* vol. 57, p. 70, 2012
- [87] MCLAUGHLIN, G.; DOTY, K. C.; LEDNEV, I. K. Discrimination of human and animal blood traces via Raman spectroscopy. *Forensic Sci. Int.* vol. 238, p. 91, 2014
- [88] SALA, O. Fundamentos da espectroscopia Raman e no infravermelho. São Paulo: UNESP, Ciência & tecnologia, p. 1-223, 1995
- [89] MOSKOVITS, M. Surface roughness and the enhanced intensity of Raman scattering by molecules adsorbed on metals. *J. Chern. Phys.* vol. 69, p. 1, 1978
- [90] CHEN, Y. J.; CHEN, W. P.; BURSTEIN, E. Surface-Electromagnetic-Wave-Enhanced Raman Scattering by Overlayers on Metals. *PHYSICAL REVIEW LETTERS*, vol. 36, nº 20, p. 1207, 1976
- [91] DRESCHER, D.; BUCHNER, T.; DON MCNAUGHTON, D.; KNEIPP, J. SERS reveals the specific interaction of silver and gold nanoparticles with hemoglobin and red blood Cell components. *Phys. Chem.* vol. 15, p. 5364, 2013

- [92] LIU, R.; XIONG, Y.; TANG, W.; GUO, Y.; YAND, X.; SI, M. Near-infrared surface-enhanced Raman spectroscopy (NIR-SERS) studies on oxyhemoglobin (Oxi-Hb) of liver cancer based on PVA-Ag nanofilm. *J. Raman Spectrosc.* vol. 44, p. 362, 2013
- [93] FEARN, T. On orthogonal signal correction. *Chem. Intel. Lab. Syst.* Vol, 50, p. 47, 2000
- [94] SVANTE WOLD, S.; ANTTI, H.; LINDGREN, F.; OHMAN, J. Orthogonal signal correction of near-infrared spectra. *Chem. Intel. Lab.*, vol. 44, p. 175, 1998
- [95] DOTY, K. C.; MCLAUGHLIN, G.; LEDNEV, I. K. A Raman Spectroscopic *clock* for bloodstain age determination: the first week after deposition. *Anal Bioanal Chem.*, vol. 408, p. 3993, 2016
- [96] PREMASIRI, W. R.; LEE, J. C.; ZIEGLER, L. D. Surface-Enhanced Raman Scattering of Whole Human Blood, Blood Plasma, and Red Blood Cells: Cellular Processes and Bioanalytical Sensing. *J. Phys. Chem. B.* vol.116, p. 9379, 2012
- [97] FENG, S.; LIN, D.; LIN, J.; LI, B.; HUANG, Z.; CHEN, G.; ZHANG, W.; WANG, L.; PAN, J.; CHENA, R.; ZENG, H. Blood plasma surface-enhanced Raman spectroscopy for non-invasive optical detection of cervical cancer. *Analyst*, vol. 138, p.3967, 2013
- [98] VIRKLER, K.; LEDNEV, I. K. Raman spectroscopic signature of blood and its potential application to forensic body fluid identification. *Anal Bioanal. Chem.*, vol. 396, p. 525, 2010
- [99] SIKIRZHYTSKI, V.; VIRKLER, K.; LEDNEV, I. K. Discriminant Analysis of Raman Spectra for Body Fluid Identification for Forensic Purposes. *Sensors*, vol. 10, p. 2869, 2010
- [100] LEMLER, P.; PREMASIRI, W.R.; DELMONACO, A.; ZIEGLER, L. D. NIR Raman spectra of whole human blood: effects of laser-induced and in vitro hemoglobin denaturation. *Anal Bioanal. Chem.* vol. 406, p. 193, 2014
- [101] YUN ZOU, Y.; XIA, P.; YANG, F.; CAO, F.; MA, K.; MI, Z.; HUANG, X.; CAI, N.; JIANG, B.; ZHAO, X.; LIU, W.; CHEN, X. Whole blood and semen identification using midinfrared and Raman spectrum analysis for forensic applications. *Anal. Methods*, vol. 8, p. 3763, 2016
- [102] SIKIRZHYTSKAYA, A.; SIKIRZHYTSKI, V.; LEDNEV, I. K. Raman spectroscopy coupled with advanced statistics for differentiating menstrual and peripheral blood. *Joun. Biophotonics*, vol. 7, p. 59, 2014

- [103] LEDNEV, I. K. Raman spectroscopy offers great potential for the nondestructive confirmatory identification of body fluids *Forensic Science International*, vol. 181, p. e1, 2008
- [104] SANTOS, P. M. Aplicação de imagens digitais e técnicas espectroanalíticas combinadas com quimiometria para detecção e quantificação de adulteração em leite bovino, tese, São Carlos, 2013
- [105] PRAT, W. K. *Digital image processing*, 2^o ed. New York, John Wiley, p. 698, 1991,
- [106] GONZALES, R.C; WOODS, R.E.; EDDINS, S. L. *Digital image processing using MATLAB*. 1^o ed. New Jersey, Pearson Prentice Hall, p. 609, 2004
- [107] Website, "A Little About Color: HSV vs. RGB", disponível em https://www.kirupa.com/design/little_about_color_hsv_rgb.htm, acessado em 03/05/2018
- [108] CARVALHO, B.M. Material suplementar sobre: Representações de cores. DIMAp/UFRN, sala 15, ramal 227
- [109] 28^o Pesquisa anual do uso de TI. Centro de tecnologia da informação aplicada, FUNDAÇÃO GETÚLIO VARGAS (EAESP- FGV) IDC: International Data Corporation, Predictions Brazil 2014, TOP Trends. IBGE, Word bank, Teleco, CGI.br, Gartner, ITU
- [110] PAQUET-DURAND, O.; SOLLE D.; SCHIRMER, M.; BECKER, T; HITZMANN, B. Monitoring baking processes of bread rolls by digital image analyses, *J. food. eng.*, p.111, 2012
- [111] JACKMAN, P. SUN, D.W.; ALLEN, P.; VALOUS, V.A.; MENDOZA, F.; WARD, P. Identification of important image features for pork and turkey ham classification using colour and wavelet texture features and genetic selection, *meat Sci.*, p.84, 2010
- [112] TÔRRES, A. R.; LYRA, W.; ANDRADE, S. I. E.; ANDRADE, R. A. N.; SILVA, E. C.; ARAUJO, M. C. U.; GAIÃO, E. N. A digital image-based method for determining of total acidity in red wines using acid-base titration without indicator, *Talanta*, p. 84, 2011
- [113] BORIN, A.; FERRÃO, M.F.; MELLO, C.; CORDI, L.; PATACA, L.C.M.; DURAN,N.; POPPI, R.J. Quantification of lactobacillus in fermented milk by multivariate image analysis with least-square support-vector machines, *Anal. Bioanal. Chem.*, vol. 387, p. 1105, 2007
- [114] BORAH, S.; BHUYAN, M. A computer based system for monitoring colours during the monitoring of tea fermentation, *Int. J. Food. Sci. Tech.*, vol. 40, p. 675, 2005

- [115] KHADABADI, G. C.; RAJPUROHIT V. S.; KUMAR, A.; NARGUND, V.B. Disease Detection in Vegetables Using Image Processing Techniques: A Review. Int. J. Emer. Tech. in Computer Sci. & Electr. , vol. 14, 2015
- [116] VAN-DALEN, G. Determination of the size distribution and percentage of broken kernels of rice flatbed scanning and image analysis. Food, Res. Int., vol. 37, p. 51, 2004
- [117] IMAGEM DE SATÉLITE DA CIDADE DE BRASÍLIA. Fonte: Google Imagens, disponível em: <http://mochileiro.tur.br/brasil2.htm>, acessado em maio de 2018
- [118] BREMMER, R. H.; DE BRUIN, D. M. DE JOODE, M; BUMA, W. J.; VAN LEEUWEN, T. G.; AALDERS, M. C. Biphasic oxidation of oxy hemoglobin in bloodstains, PLOS one, vol. 6, p. e21845, 2011

APÊNDICE

A.1 Linhas de comando para os 9 sistemas de cores usados para o tratamento das imagens digitais: ROI e imagem sem recortes

RGB

```

a=imread('a.jpg');    % Importar a imagem (3000x4000)
rgbimagem1horas=a(1800:2300,1200:1700,:); % Recortar a imagem (501x501)
rgbred1=rgbimagem1horas(:,:,1); % Canal red (501x501)
rgbgreen1=rgbimagem1horas(:,:,2); % Canal green (501x501)
rgbblue1=rgbimagem1horas(:,:,3); % Canal blue (501x501)
[rgby1red,x1r]= imhist(rgbred1); % Frequência absoluta red (256x1)
[rgby1green,x1g]= imhist(rgbgreen1); % Frequência absoluta green (256x1)
[rgby1blue,x1b]= imhist(rgbblue1); % Frequência absoluta blue (256x1)
rgbimagem1=[rgby1red,rgby1green,rgby1blue]; % Frequência absoluta (256x3)
somotorio=sum(rgbimagem1); % Somatório das frequências (1x3)
rgby1red=rgby1red/251001; % Frequência relativa red (1x256)
rgby1green=rgby1green/251001; % Frequência relativa green (1x256)
rgby1blue=rgby1blue/251001; % Frequência relativa blue (1x256)
rgbimagem1=[rgby1red',rgby1green',rgby1blue']; % Matriz (1x768)
I1=reshape(rgbimagem1horas,size(rgbimagem1horas,1)*size(rgbimagem1horas,2),size(
rgbimagem1horas,3)); % Frequência absoluta (256x3)
Media1=mean(I1); % Valores Médios de luminosidade (1x3)

```

HSV

```

a=imread('a.jpg');
rgbimagem1horas=a(1800:2300,1200:1700,:);
hsvimagem1horas= rgb2hsv(rgbimagem1horas); % Converter imagem RGB para HSV
(501x501)
hsv1h=hsvimagem1horas(:,:,1);
hsv1s=hsvimagem1horas(:,:,2);
hsv1v=hsvimagem1horas(:,:,3);
hsvh1 = reshape(hsv1h,[251001,1]); % Valor de Luminosidade, H, (251001x1)
hsvs1 = reshape(hsv1s,[251001,1]); % Valor de Luminosidade, S, (251001x1)
hsvv1 = reshape(hsv1v,[251001,1]); % Valor de Luminosidade, V, (251001x1)
I1=[hsvh1,hsvs1,hsvv1]; % Valor de Luminosidade (251001x3)
Media1=mean(I1); % Valores Médios de luminosidade (1x3)
[hsvy1h,x1h]= imhist(hsvh1);
[hsvy1s,x1s]= imhist(hsvs1);
[hsvy1v,x1v]= imhist(hsvv1);
hsvimagem1=[hsvy1h,hsvy1s,hsvy1v];
somotorio=sum(hsvimagem1);

```

```

hsvy1h=hsvy1h/251001;
hsvy1s=hsvy1s/251001;
hsvy1v=hsvy1v/251001;
hsvimagem1=[hsvy1h',hsvy1s',hsvy1v']; % Matriz (1x768)

```

CMYK

```

a=imread('a.jpg');
c=a(1800:2300,1200:1700,:);
c = rgb2cmk(c); % Converter imagem RGB para CMYK (501x501)
CMYKimagem1horas=c;
CMYKc1=CMYKimagem1horas(:,:,1);
CMYKy1=CMYKimagem1horas(:,:,2);
CMYKm1=CMYKimagem1horas(:,:,3);
CMYKk1=CMYKimagem1horas(:,:,4);
[CMYKy1c,x1c]= imhist(CMYKc1);
[CMYKy1y,x1y]= imhist(CMYKy1);
[CMYKy1m,x1m]= imhist(CMYKm1);
[CMYKy1k,x1k]= imhist(CMYKk1);
CMYKimagem1=[CMYKy1c,CMYKy1y,CMYKy1m,CMYKy1k];
somatorio=sum(CMYKimagem1);
CMYKy1c=CMYKy1c/251001;
CMYKy1y=CMYKy1y/251001;
CMYKy1m=CMYKy1m/251001;
CMYKy1k=CMYKy1k/251001;
CMYKimagem1=[CMYKy1c',CMYKy1y',CMYKy1m',CMYKy1k']; % Matriz (1x768)
I1=reshape(CMYKimagem1horas,size(CMYKimagem1horas,1)*size(CMYKimagem1horas,2),size(CMYKimagem1horas,3));
Media1=mean(I1); % Valores Médios de luminosidade (1x3)

```

HSL

```

a=imread('a.jpg');
hslimagem1horas=a(1800:2300,1200:1700,:);
rgb_in = hslimagem1horas;
rgb=reshape(rgb_in, [], 3);
mx=max(rgb,[],2);
mn=min(rgb,[],2);
lhs1=(mx+mn)/2; % Matriz de dados L
hsl1 = reshape(lhs1,[501,501]);
hslimagem1horas=rgb2hsv(hslimagem1horas); % Converter imagem RGB para HSV(501x501)
hslh1=hslimagem1horas(:,:,1);
hsls1=hslimagem1horas(:,:,2);
hslimagem1horas=cat(3,hslh1,hsls1,hsl1); % Agrupamento das matrizes dos canais separados para formar a imagem do sistema
hsl1 = reshape(hslh1,[251001,1]);

```

```

shsl1 = reshape(hsls1,[251001,1]);
a100 = mean(hhsl1);
a101 = mean(shsl1);
a102 = mean(lhsl1);
Media1=[a100,a101,a102]; % Valores Médios de luminosidade (1x3)
I1=[hhsl1, shsl1, lhsl1];
[hsly1h,x1h]= imhist(hslh1);
[hsly1s,x1s]= imhist(hsls1);
[hsly1l,x1l]= imhist(hsll1);
hslimagem1=[hsly1h,hsly1s,hsly1l];
somatorio=sum(hslimagem1);
hsly1h=hsly1h/251001;
hsly1s=hsly1s/251001;
hsly1l=hsly1l/251001;
hslimagem1=[hsly1h',hsly1s',hsly1l']; % Matriz (1x768)

```

L*a*b

```

a=imread('a.jpg');
rgbimagem1horas=a(1800:2300,1200:1700,:);
labimagem1horas= rgb2lab(rgbimagem1horas); % Converter imagem RGB para LAB
(501x501)
lab1l=labimagem1horas(:,,1);
lab1a=labimagem1horas(:,,2);
lab1b=labimagem1horas(:,,3);
lab1l = reshape(lab1l,[251001,1]);
lab1a = reshape(lab1a,[251001,1]);
lab1b = reshape(lab1b,[251001,1]);
I1=[lab1l,lab1a,lab1b];
Media1=mean(I1); % Valores Médios de luminosidade (1x3)
[laby1l,x1l]= imhist(lab1l);
[laby1a,x1a]= imhist(lab1a);
[laby1b,x1b]= imhist(lab1b);
labimagem1=[laby1l,laby1a,laby1b];
somatorio=sum(labimagem1);
laby1l=laby1l/251001;
laby1a=laby1a/251001;
laby1b=laby1b/251001;
labimagem1=[laby1l',laby1a',laby1b']; % Matriz (1x768)

```

YCbCr

```

a=imread('a.jpg');
rgbimagem1horas=a(1800:2300,1200:1700,:);
rgbred1=rgbimagem1horas(:,,1);
rgbgreen1=rgbimagem1horas(:,,2);
rgbblue1=rgbimagem1horas(:,,3);

```

```

a1=(0.299*rgbred1);
b1=(0.587*rgbgreen1);
c1=(0.114*rgbblue1);
ybcry1=(0+a1+b1+c1); % Converter RGB para Y
a1a=(0.168736*rgbred1);
b1b=(0.331264*rgbgreen1);
c1c=(0.5*rgbblue1);
ybcrcb1=(128-a1a-b1b+c1c); % Converter RGB para Cb
a1aa=(0.5*rgbred1);
b1bb=(0.418688*rgbgreen1);
c1cc=(0.081213*rgbblue1);
ybcrcr1=(128+a1aa-b1bb-c1cc); % Converter RGB para Cr
ybcrimagem1horas=cat(3,ybcry1,ybcrcb1,ybcrcr1);
y1a=reshape(ybcry1,[251001,1]);
cb1a=reshape(ybcrcb1,[251001,1]);
cr1a=reshape(ybcrcr1,[251001,1]);
I1=[y1a,cb1a,cr1a];
Media1=mean(I1); % Valores Médios de luminosidade (1x3)
[ybcry1y,x1y]= imhist(ybcry1);
[ybcry1cb,x1cb]= imhist(ybcrcb1);
[ybcry1cr,x1cr]= imhist(ybcrcr1);
ybcrimagem1=[ybcry1y,ybcry1cb,ybcry1cr];
somatorio=sum(ybcrimagem1);
ybcry1y=ybcry1y/251001;
ybcry1cb=ybcry1cb/251001;
ybcry1cr=ybcry1cr/251001;
ybcrimagem1=[ybcry1y',ybcry1cb',ybcry1cr']; % Matriz (1x768)

```

YIQ

```

a=imread('a.jpg');
rgbimagem1horas=a(1800:2300,1200:1700,:);
rgbred1=rgbimagem1horas(:,:,1);
rgbgreen1=rgbimagem1horas(:,:,2);
rgbblue1=rgbimagem1horas(:,:,3);
a1=(0.299*rgbred1);
b1=(0.587*rgbgreen1);
c1=(0.144*rgbblue1);
yiqy1=(a1+b1+c1); % Converter RGB para Y
a1a=(0.596*rgbred1);
b1b=(0.275*rgbgreen1);
c1c=(0.321*rgbblue1);
yiqi1=(a1a-b1b-c1c); % Converter RGB para I
a1aa=(0.212*rgbred1);
b1bb=(0.523*rgbgreen1);
c1cc=(0.311*rgbblue1);
yiqq1=(a1aa-b1bb+c1cc); % Converter RGB para Q
yiqimagem1horas=cat(3,yiqy1,yiqi1,yiqq1);

```

```

y1a=reshape(yiqy1,[251001,1]);
i1a=reshape(yiqi1,[251001,1]);
q1a=reshape(yiqq1,[251001,1]);
I1=[y1a,i1a,q1a];
Media1=mean(I1); % Valores Médios de luminosidade (1x3)
[yiqy1y,x1y]= imhist(yiqy1);
[yiqy1i,x1i]= imhist(yiqi1);
[yiqy1q,x1q]= imhist(yiqq1);
yiqimagem1=[yiqy1y,yiqy1i,yiqy1q];
somotorio=sum(yiqimagem1);
yiqy1y=yiqy1y/251001;
yiqy1i=yiqy1i/251001;
yiqy1q=yiqy1q/251001;
yiqimagem1=[yiqy1y',yiqy1i',yiqy1q']; % Matriz (1x768)

```

YUV

```

a=imread('a.jpg');
rgbimagem1horas=a(1800:2300,1200:1700,:);
rgbred1=rgbimagem1horas(:,1);
rgbgreen1=rgbimagem1horas(:,2);
rgbblue1=rgbimagem1horas(:,3);
a1=(0.299*rgbred1);
b1=(0.587*rgbgreen1);
c1=(0.114*rgbblue1);
yuvy1=(a1+b1+c1); % Converter RGB para Y
a1a=(0.147*rgbred1);
b1b=(0.289*rgbgreen1);
c1c=(0.436*rgbblue1);
yuvu1=(-a1a-b1b+c1c); % Converter RGB para U
a1aa=(0.615*rgbred1);
b1bb=(0.515*rgbgreen1);
c1cc=(0.1*rgbblue1);
yuvv1=(a1aa-b1bb-c1cc); % Converter RGB para V
yuvimagem1horas=cat(3,yuvy1,yuvu1,yuvv1);
y1a=reshape(yuvy1,[251001,1]);
u1a=reshape(yuvu1,[251001,1]);
v1a=reshape(yuvv1,[251001,1]);
I1=[y1a,u1a,v1a];
Media1=mean(I1); % Valores Médios de luminosidade (1x3)
[yuvy1y,x1y]= imhist(yuvy1);
[yuvy1u,x1u]= imhist(yuvu1);
[yuvy1v,x1v]= imhist(yuvv1);
yuvimagem1=[yuvy1y,yuvy1u,yuvy1v];
somotorio=sum(yuvimagem1);
yuvy1y=yuvy1y/251001;
yuvy1u=yuvy1u/251001;
yuvy1v=yuvy1v/251001;

```

```
yuvimagem1=[yuvy1y',yuvy1u',yuvy1v']; % Matriz (1x768)
```

XYZ

```
a=imread('a.jpg');
rgbimagem1horas=a(1800:2300,1200:1700,:);
rgbred1=rgbimagem1horas(:,1);
rgbgreen1=rgbimagem1horas(:,2);
rgbblue1=rgbimagem1horas(:,3);
a1=(0.431*rgbred1);
b1=(0.342*rgbgreen1);
c1=(0.178*rgbblue1);
xyzy1=(a1+b1+c1); % Converter RGB para x
a1a=(0.222*rgbred1);
b1b=(0.707*rgbgreen1);
c1c=(0.071*rgbblue1);
xyzi1=(a1a+b1b+c1c); % Converter RGB para y
a1aa=(0.02*rgbred1);
b1bb=(0.130*rgbgreen1);
c1cc=(0.939*rgbblue1);
xyzq1=(a1aa+b1bb+c1cc); % Converter RGB para z
xyzimagem1horas=cat(3,xyzy1,xyzi1,xyzq1);
y1a=reshape(xyzy1,[251001,1]);
i1a=reshape(xyzi1,[251001,1]);
q1a=reshape(xyzq1,[251001,1]);
I1=[y1a,i1a,q1a];
Media1=mean(I1); % Valores Médios de luminosidade (1x3)
[xyzy1y,x1y]= imhist(xyzy1);
[xyzy1i,x1i]= imhist(xyzi1);
[xyzy1q,x1q]= imhist(xyzq1);
xyzimagem1=[xyzy1y,xyzy1i,xyzy1q];
somatorio=sum(xyzimagem1);
xyzy1y=xyzy1y/251001;
xyzy1i=xyzy1i/251001;
xyzy1q=xyzy1q/251001;
xyzimagem1=[xyzy1y',xyzy1i',xyzy1q']; % Matriz (1x768)
```

As rotinas empregadas para obter as matrizes relacionadas com as imagens sem recortes foram similares as já apresentadas, com exceção de três mudanças nas linhas de comando. A primeira está associada ao recorte da região de interesse (ROI).

A segunda está ligada a obtenção das matrizes considerando todos os *pixel* de cada canal além do uso da função *reshape* e o cálculo da média.

E por fim, a terceira é na alteração do número de *pixel* da imagem que vai de 251.001 para 12.000.000, sendo que o primeiro valor deve ser substituído pelo segundo

na etapa do cálculo da frequência relativa. As linhas de comando, para o sistema RGB, que sofreram mudanças estão mostradas a seguir:

```
    rgbimagem1horas=a;  
    rgby1red=rgby1red/12000000;  
    rgby1green=rgby1green/12000000;  
    rgby1blue=rgby1blue/12000000;
```

As mesmas alterações foram aplicadas nas rotinas para os outros sistemas de cores.

A.2 Comparação entre os histogramas da ROI com os da imagem da mancha sem recortes para os sistemas de cores YUV, YIQ, YCbCr, HSL, L*a*b e XYZ

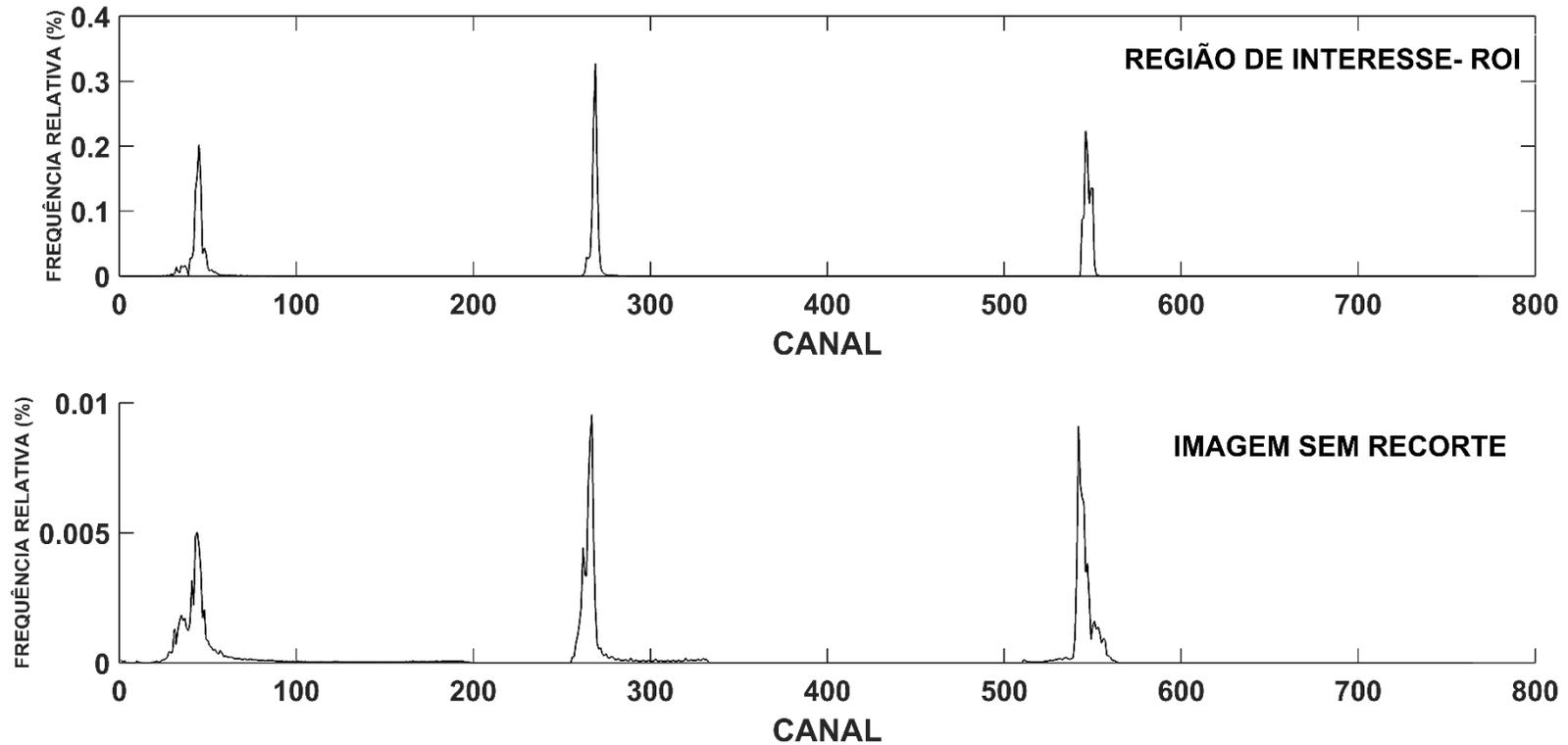


Figura 81. Histograma da ROI e da mancha sem recortes no sistema YUV

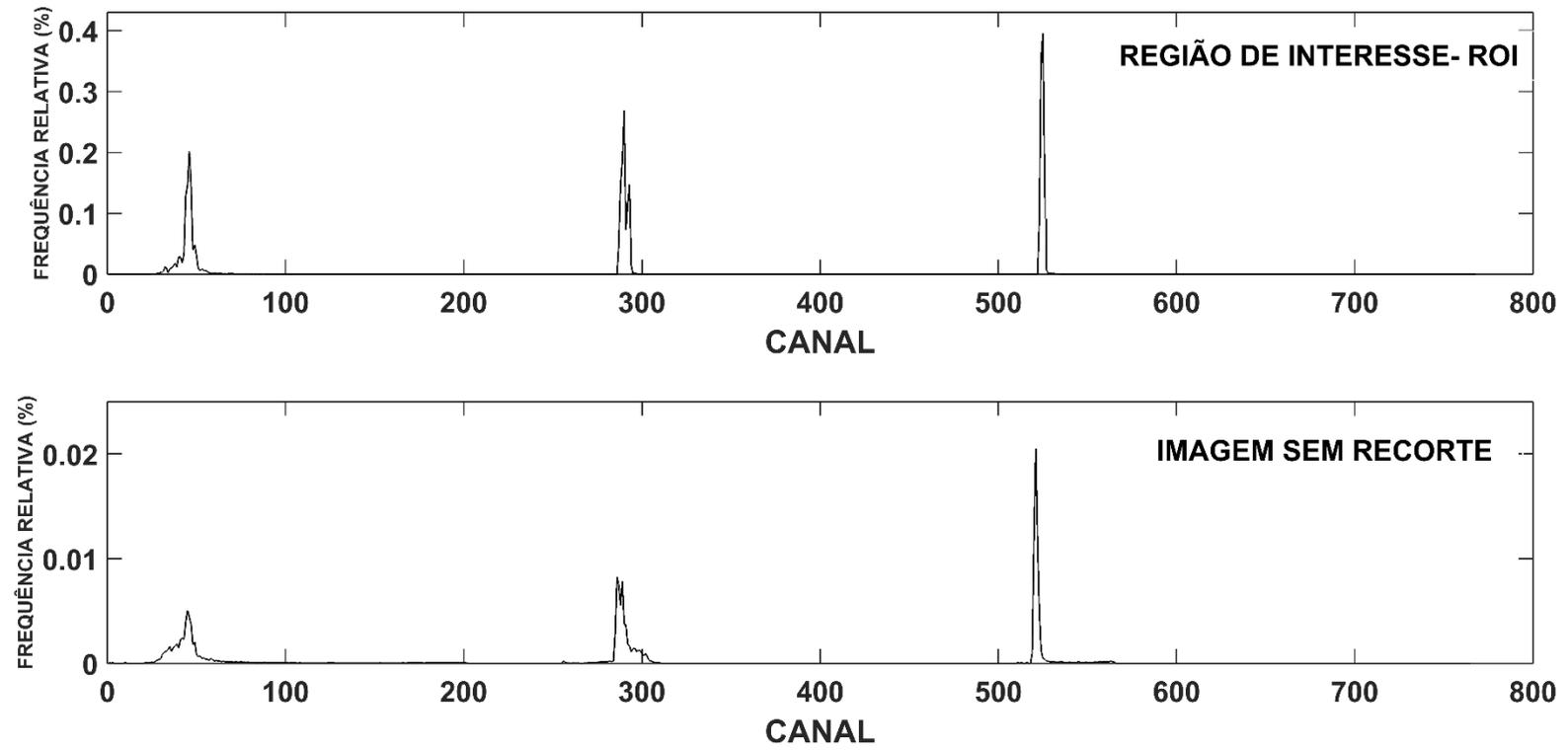


Figura 82. Histograma da ROI e da mancha sem recortes no sistema YIQ

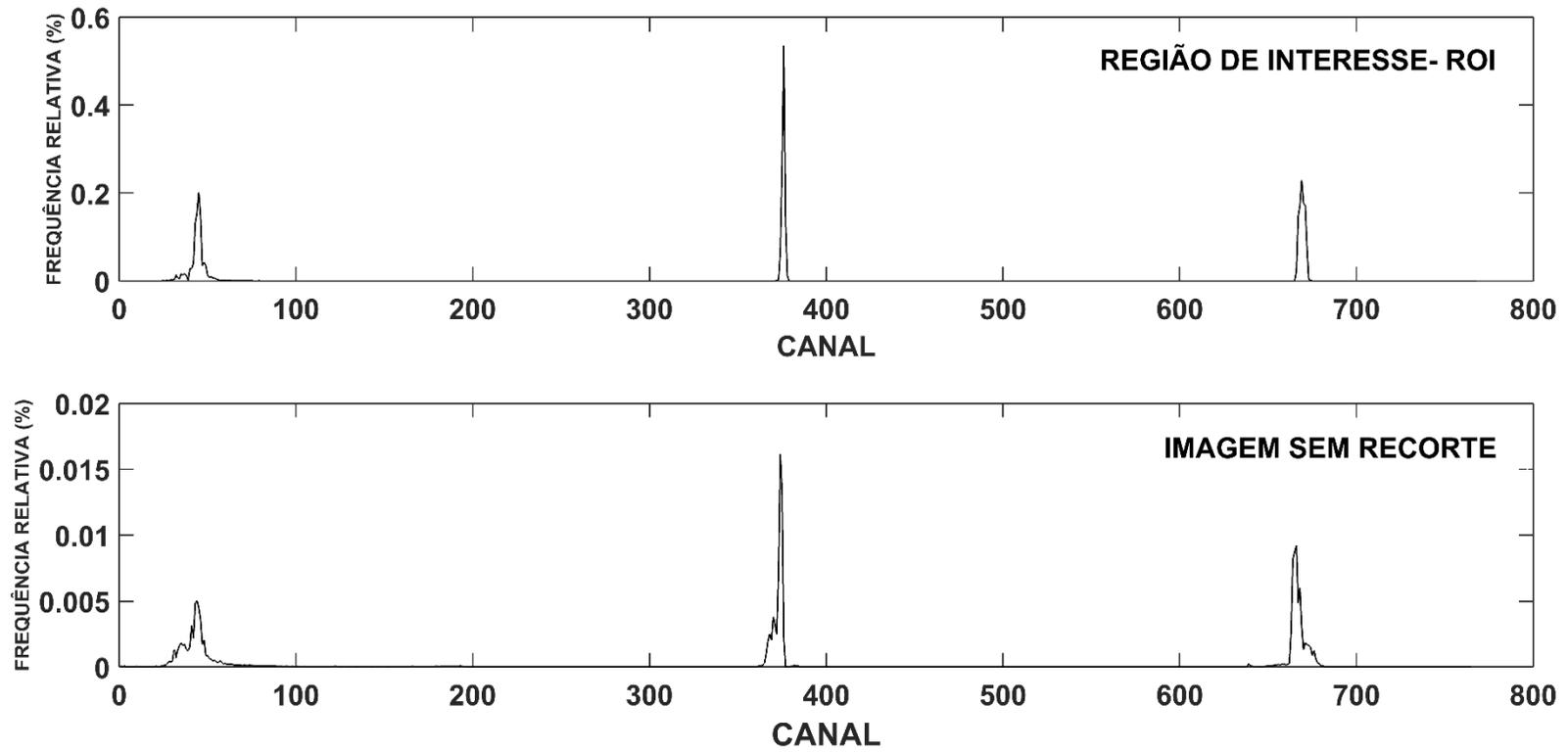


Figura 83. Histograma da ROI e da mancha sem recortes no sistema YCbCr

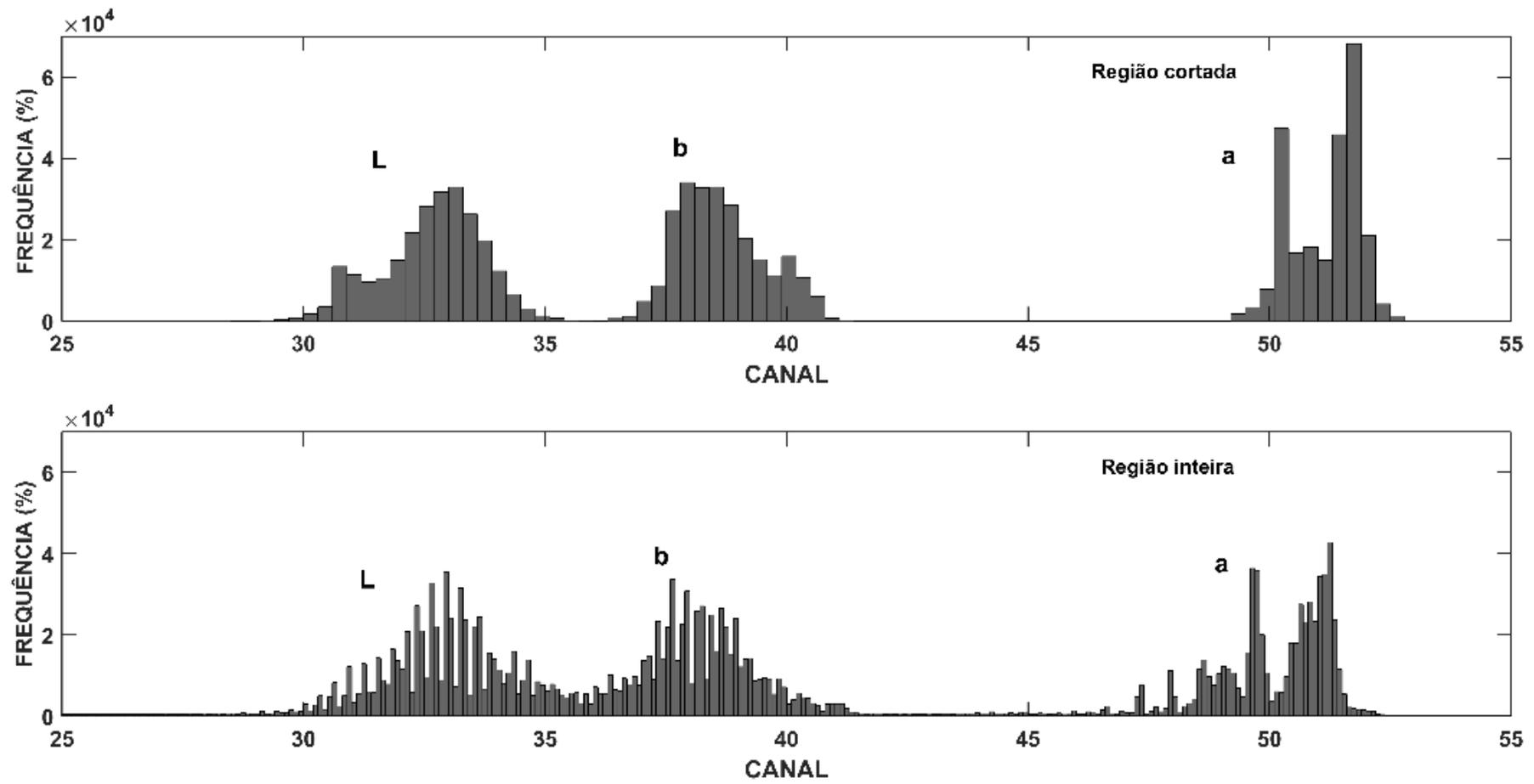


Figura 84. Histograma da ROI e da mancha sem recortes no sistema L*a*b

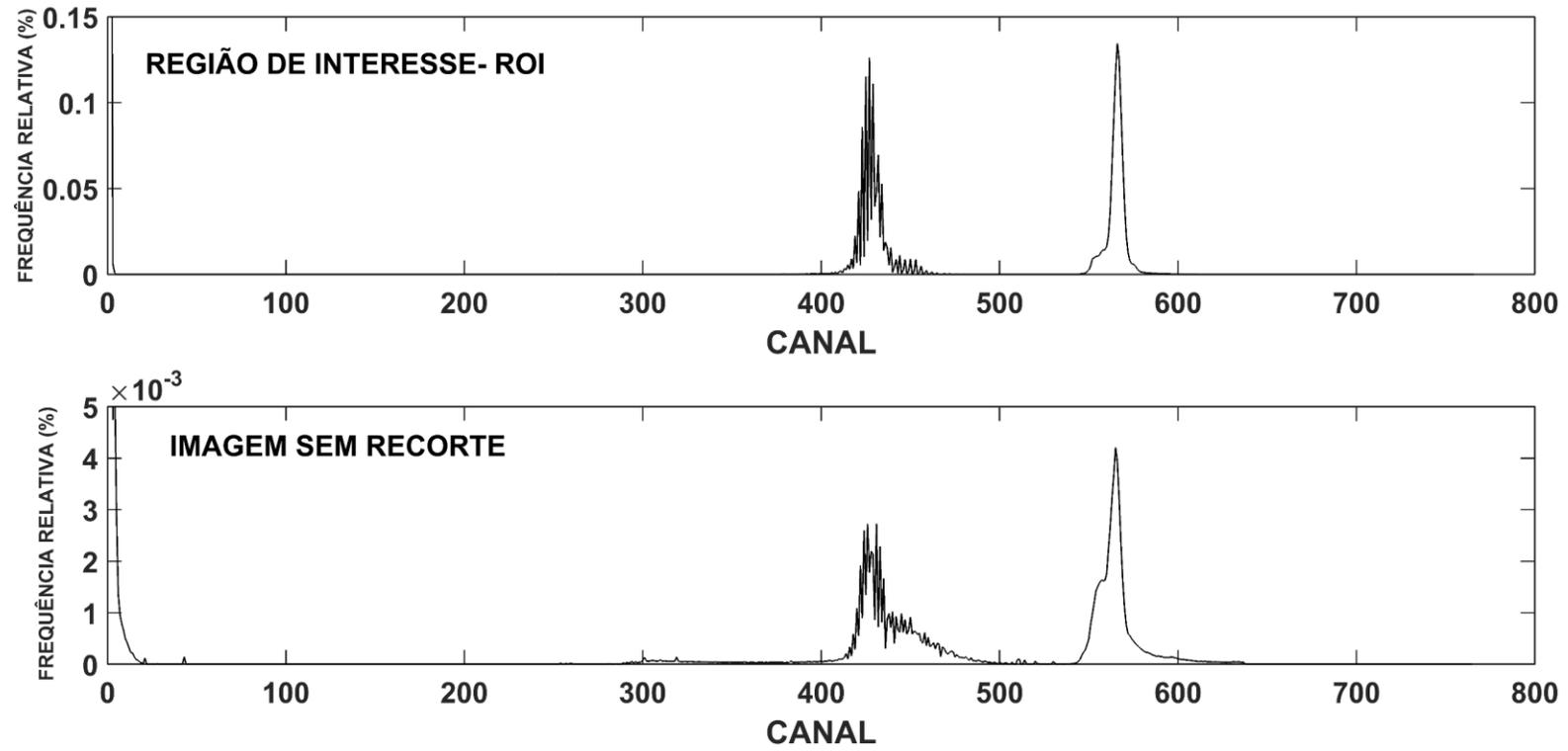


Figura 85. Histograma da ROI e da mancha sem recortes no sistema HSL

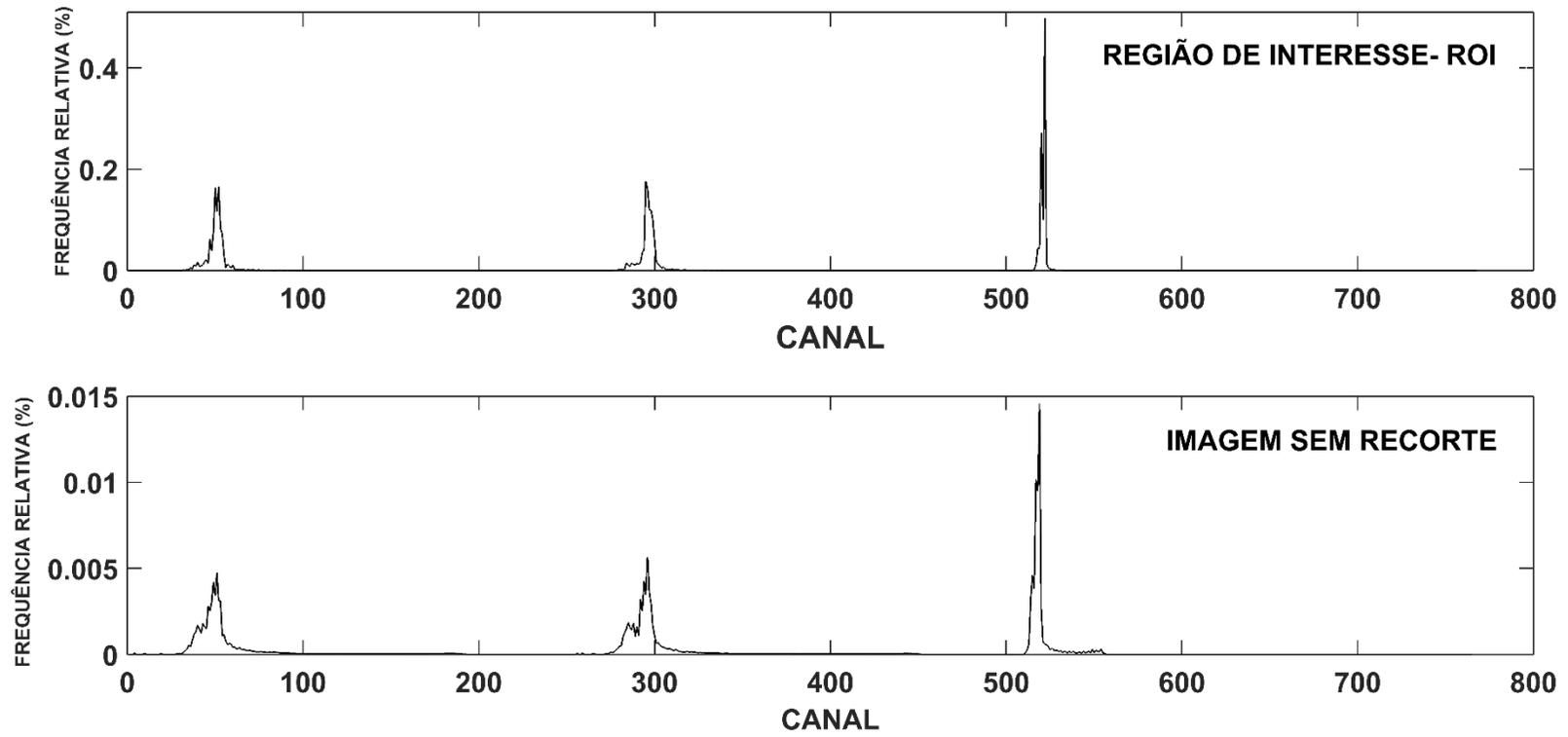


Figura 86. Histograma da ROI e da mancha sem recortes no sistema XYZ

A.3 Histogramas de uma mancha de sangue para os tempos de exposição de 0, 1 e 4 horas para os sistemas de cores YUV, YIQ, YCbCr, HSL, L*a*b e XYZ

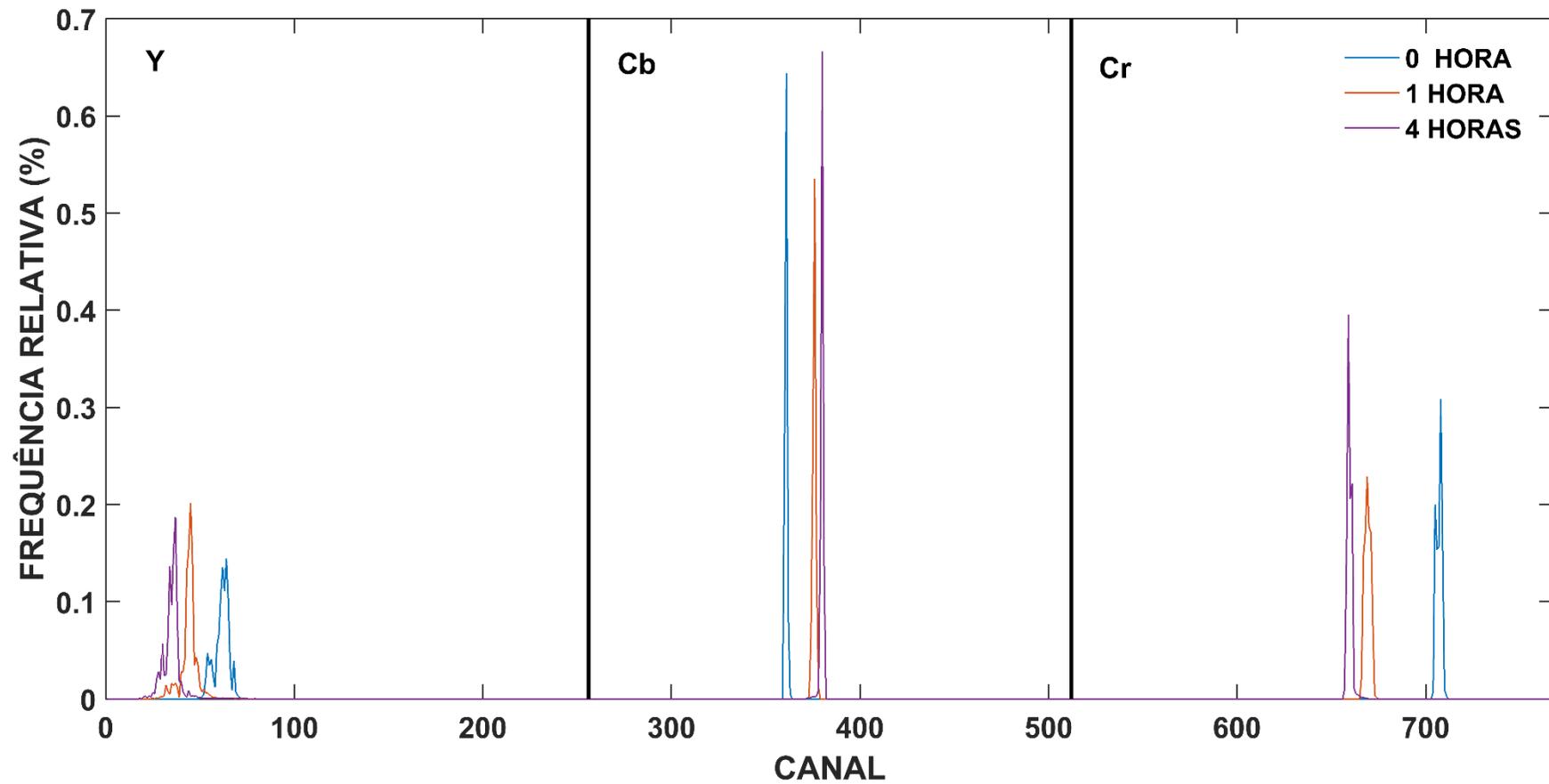


Figura 87. Histograma da ROI no sistema YCbCr para 0, 1, 4 horas

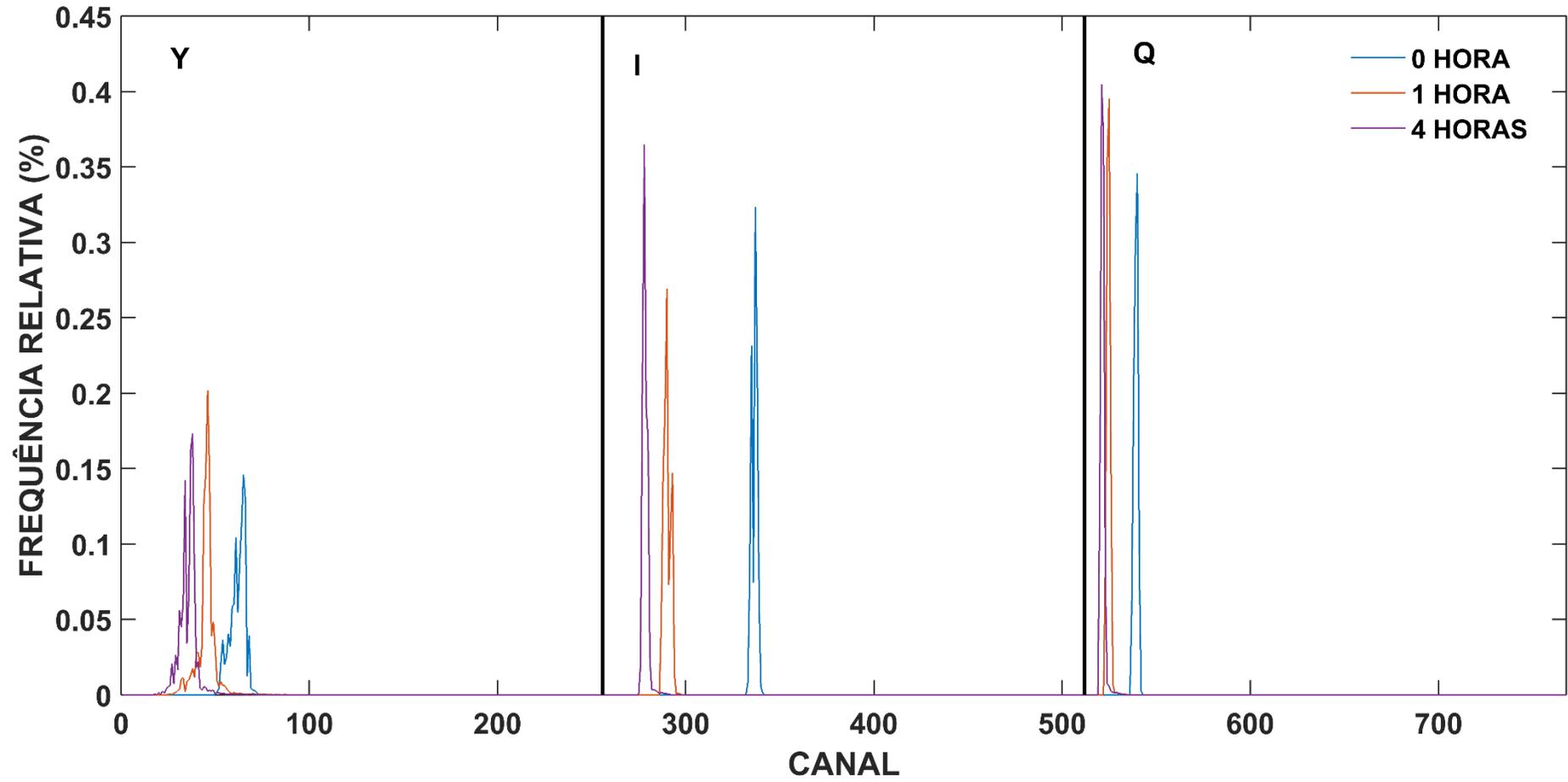


Figura 88. Histograma da ROI no sistema YIQ para 0, 1, 4 horas

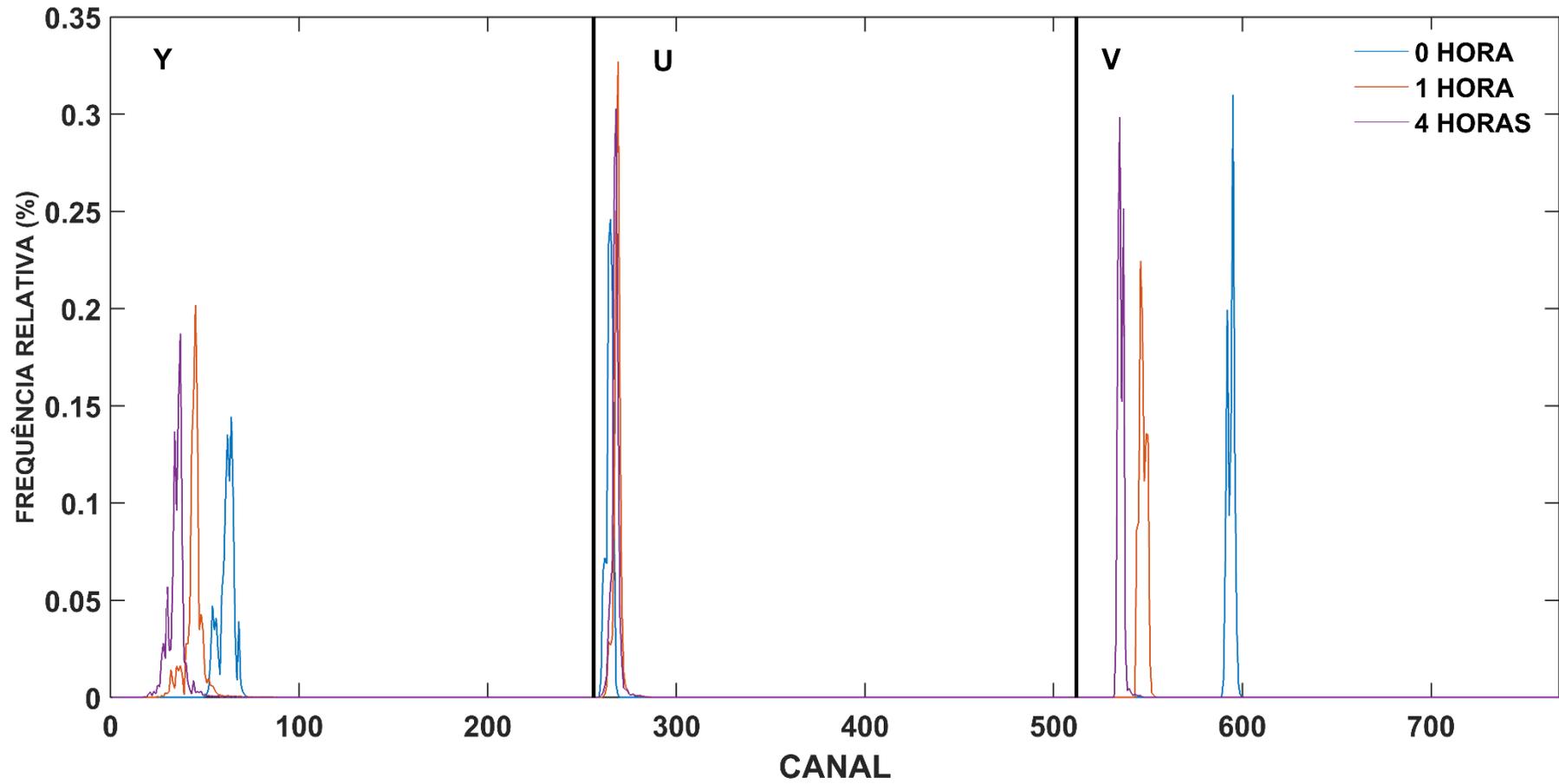


Figura 89. Histograma da ROI no sistema YUV para 0, 1, 4 horas

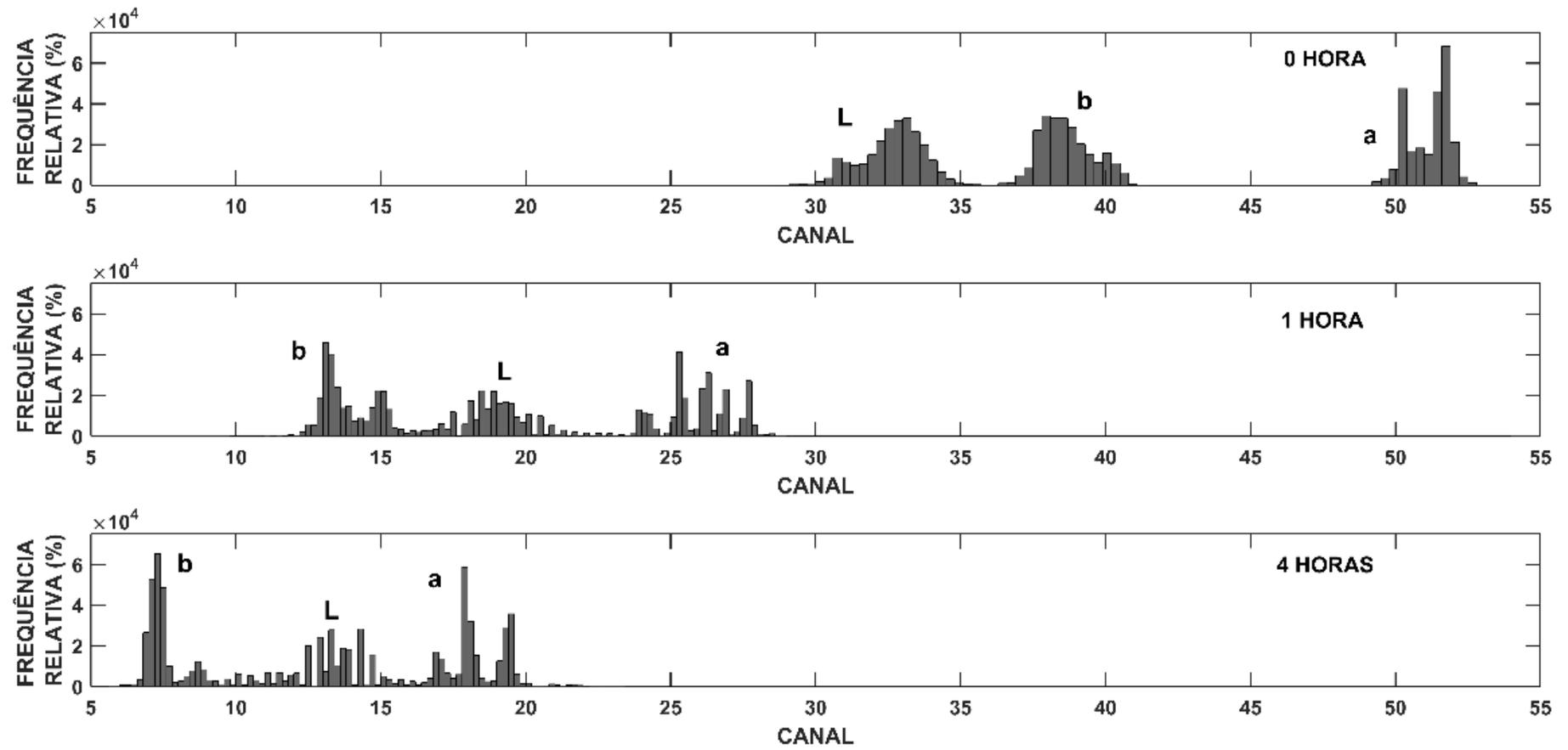


Figura 90. Histograma da ROI no sistema L*a*b para 0, 1, 4 horas

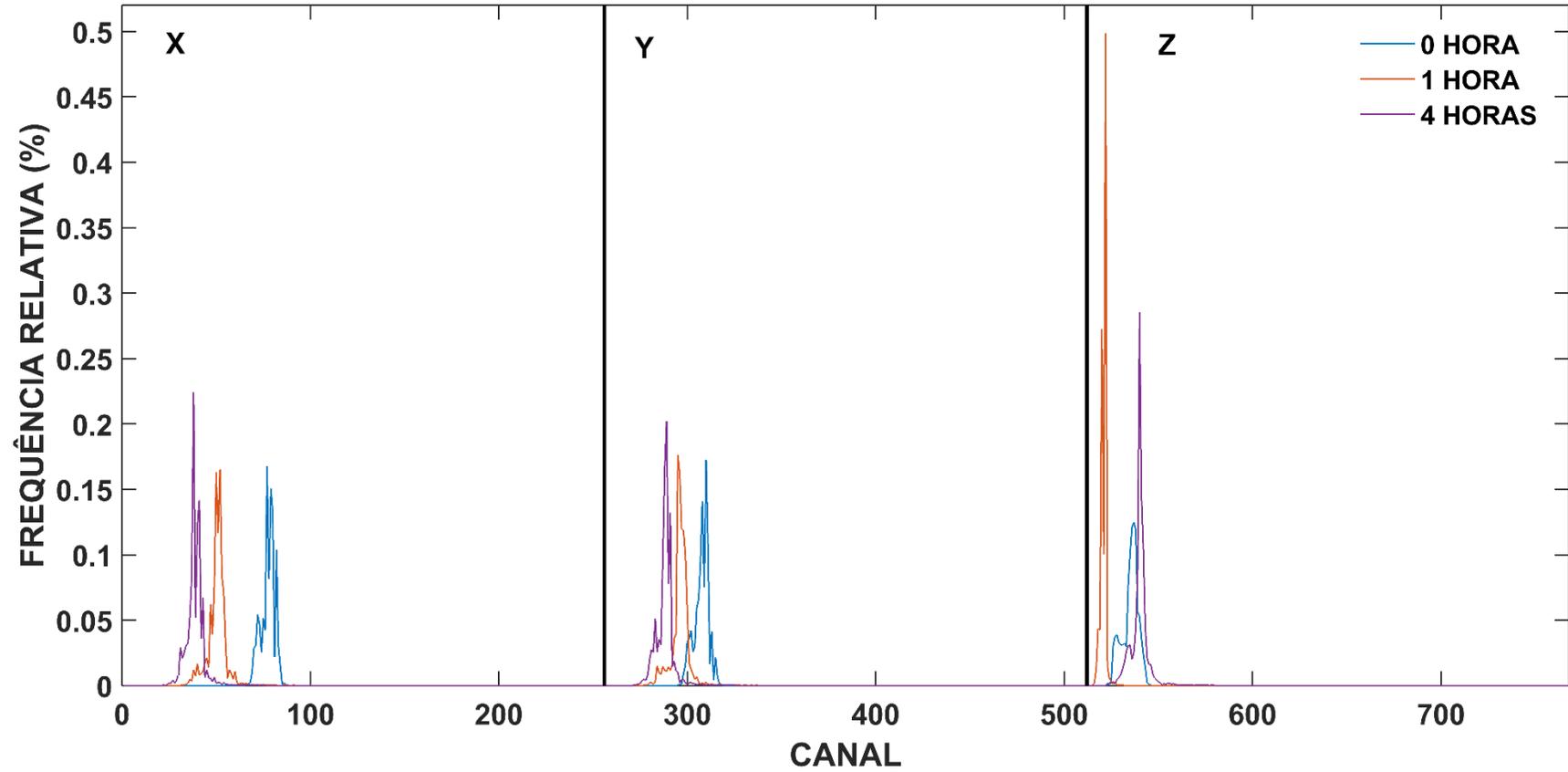


Figura 91. Histograma da ROI no sistema XYZ para 0, 1, 4 horas

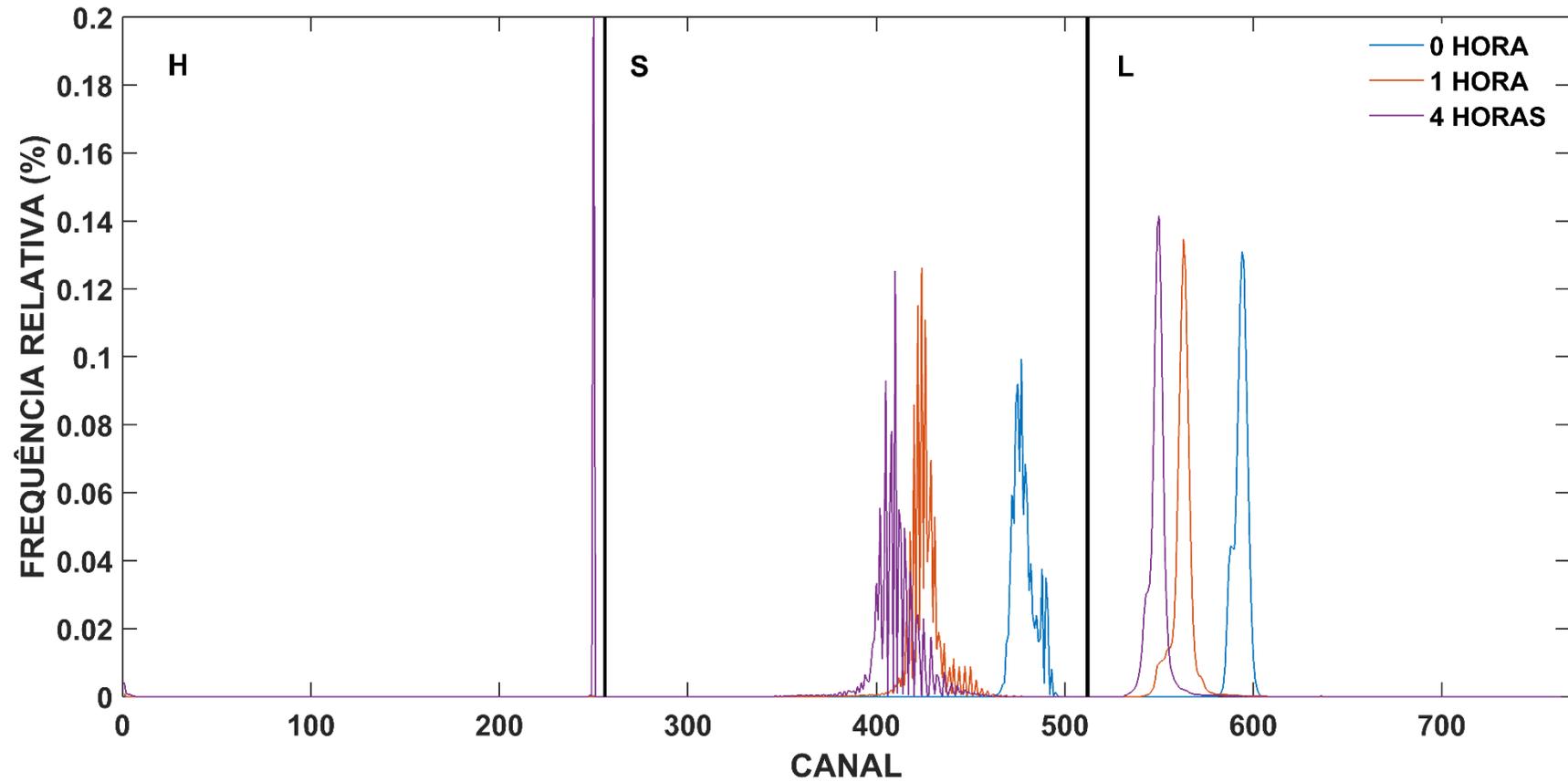


Figura 92. Histograma da ROI no sistema HSL para 0, 1, 4 horas

A.4 Valores médios de luminosidade em função do tempo de exposição para os sistemas de cores YUV, YIQ, YCbCr, HSL, L*a*b e XYZ

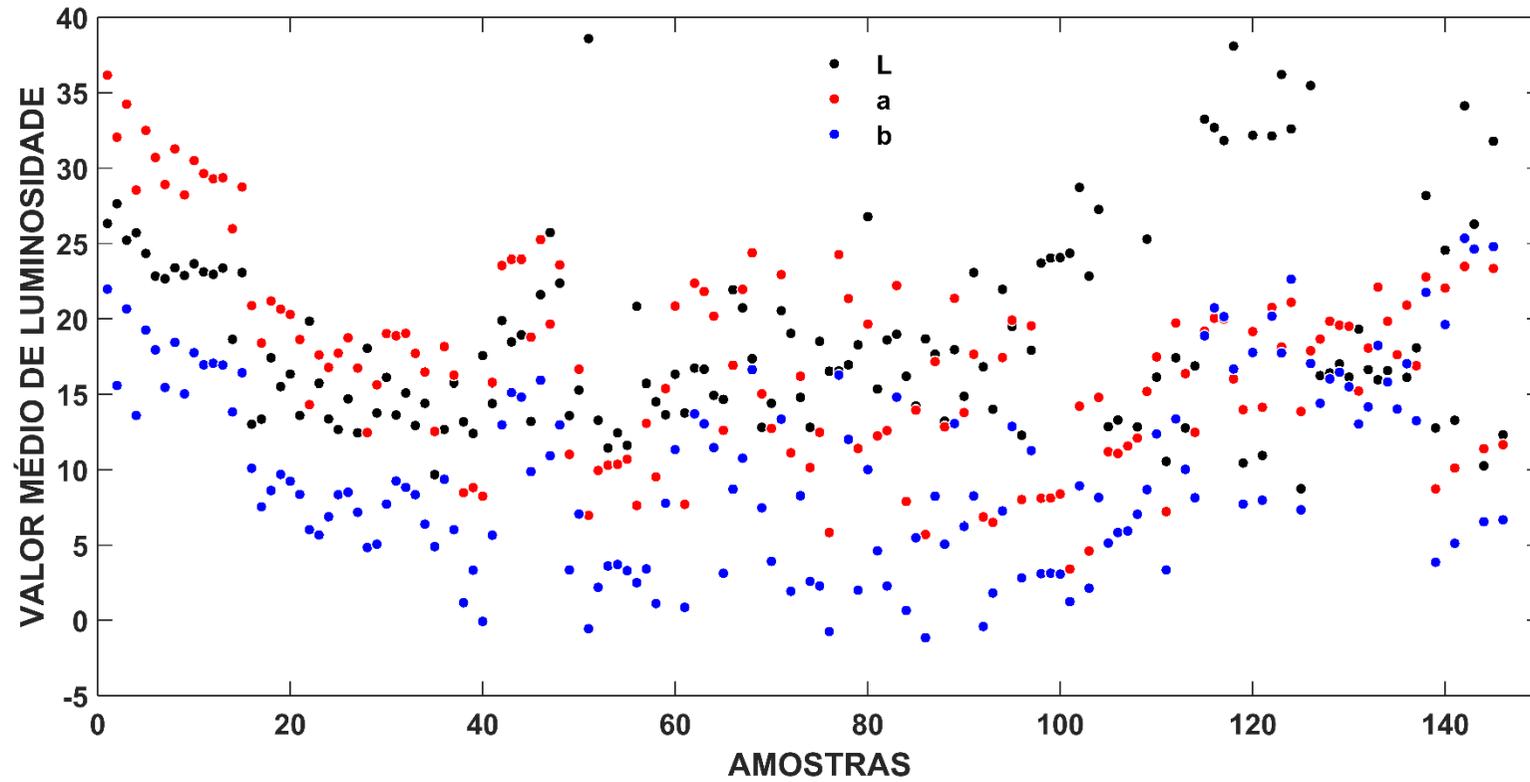


Figura 93. Valores médios de luminosidade para o sistema L*a*b em ordem crescente de horas

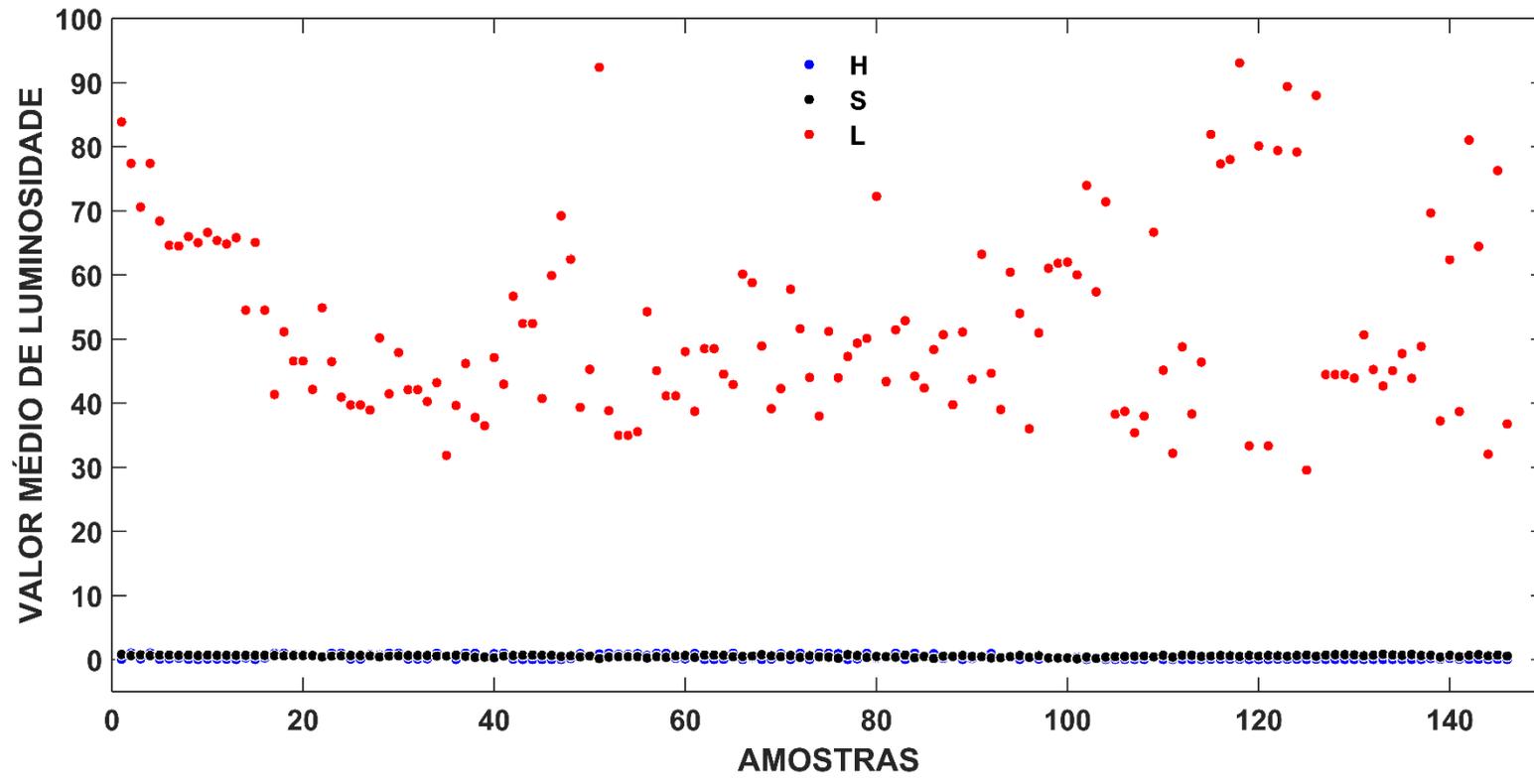


Figura 94. Valores médios de luminosidade para o sistema HSL em ordem crescente de horas

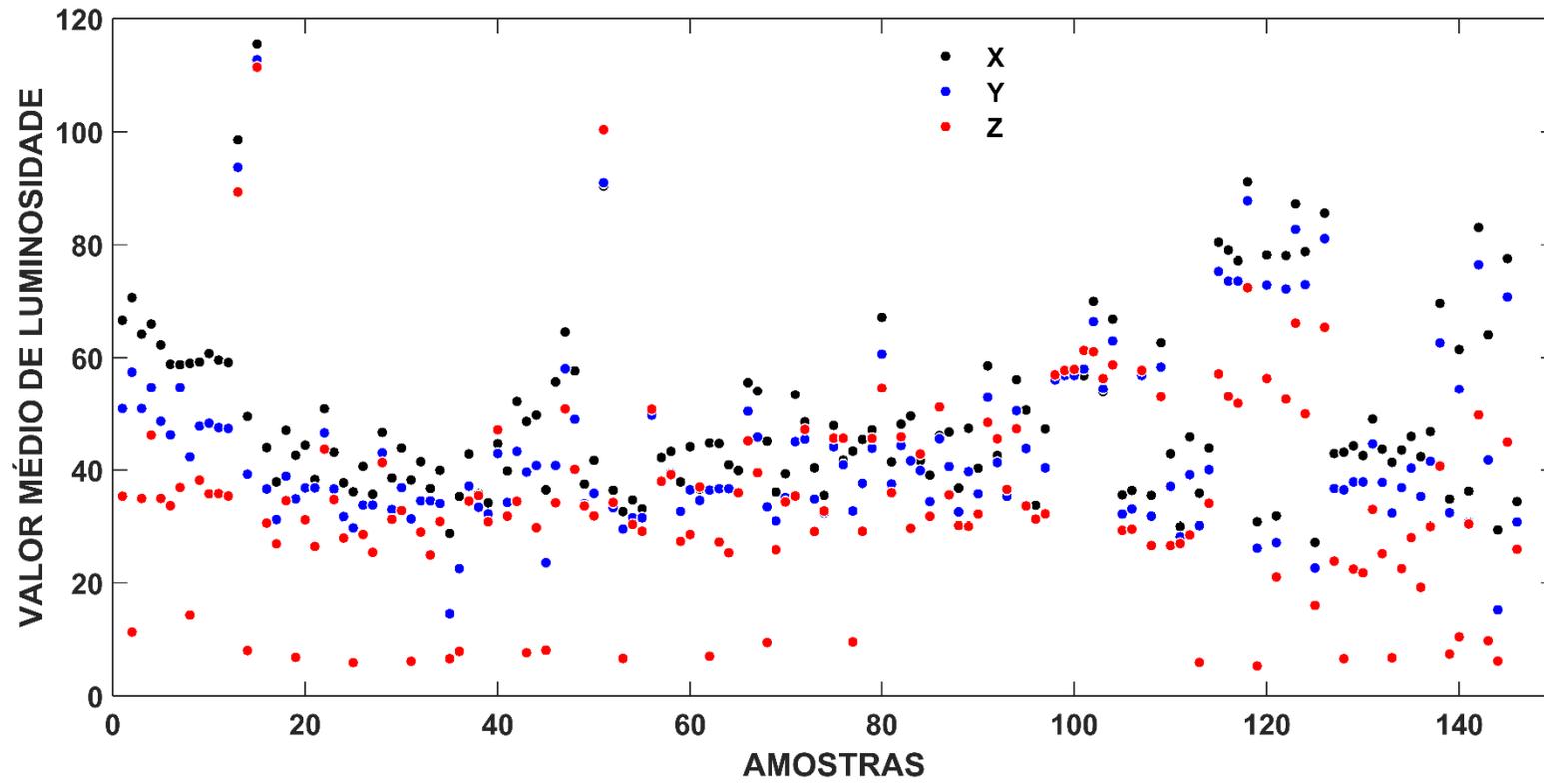


Figura 95. Valores médios de luminosidade para o sistema XYZ em ordem crescente de horas

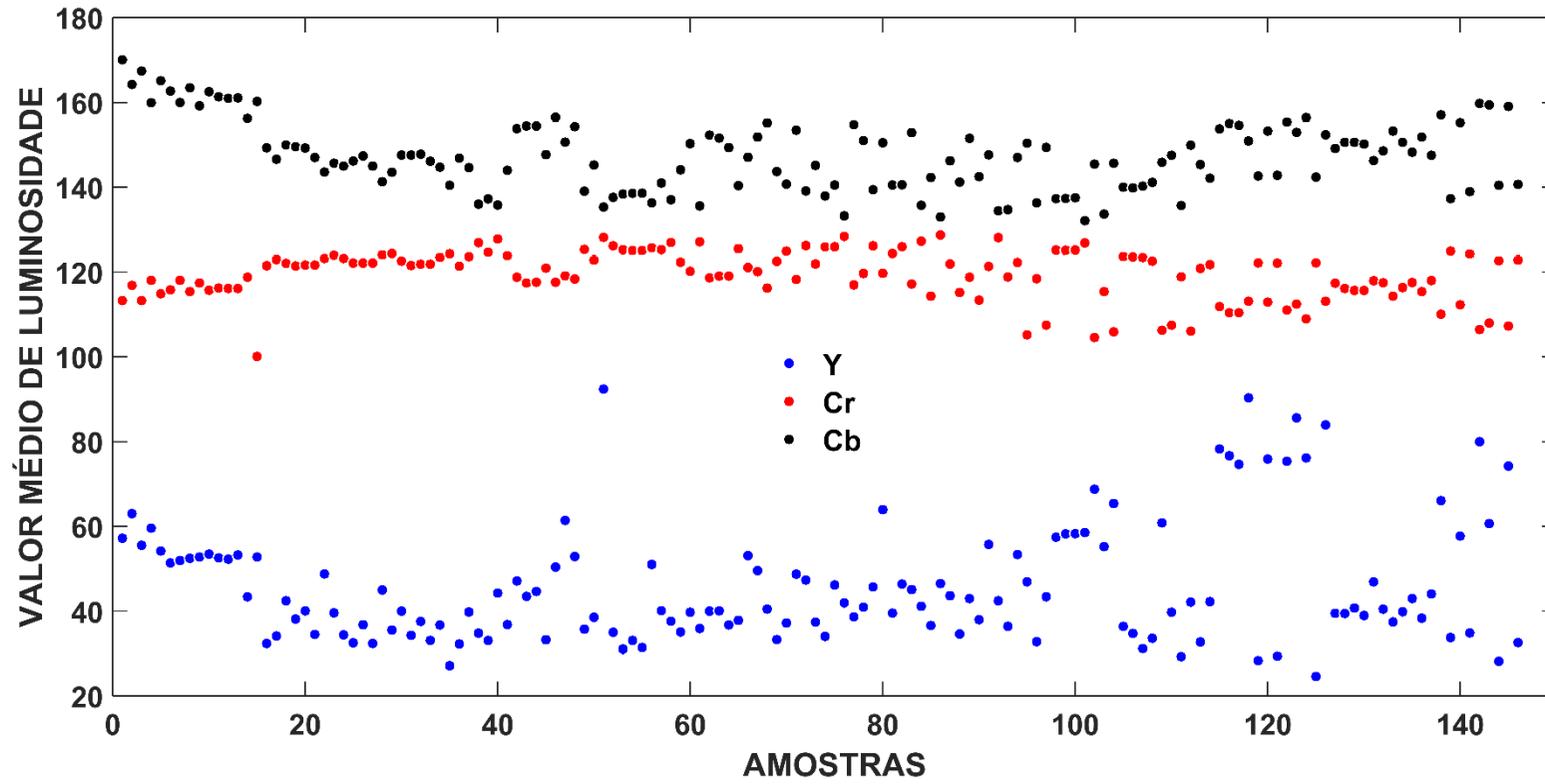


Figura 96. Valores médios de luminosidade para o sistema YCbCr em ordem crescente de horas

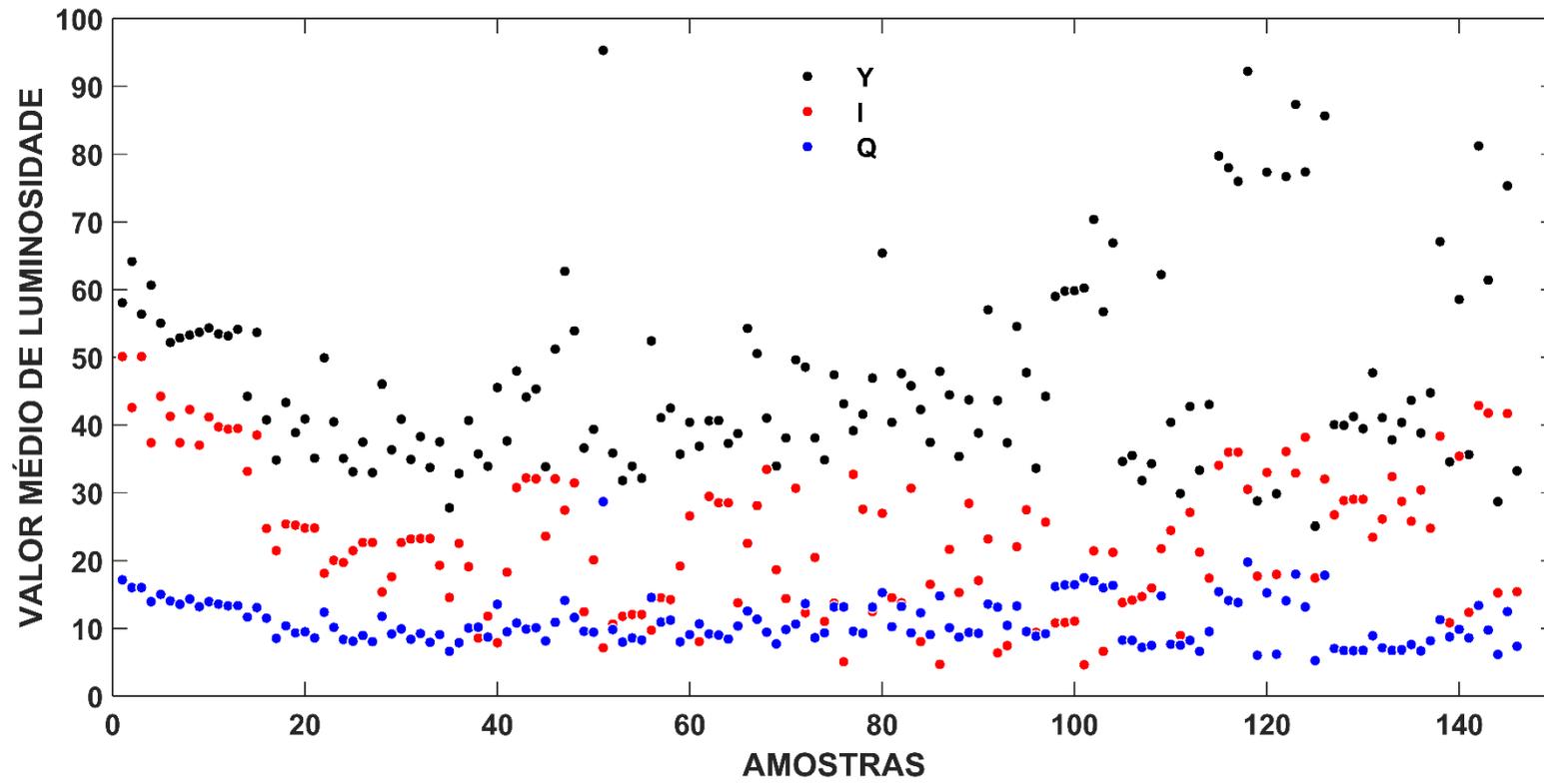


Figura 97. Valores médios de luminosidade para o sistema YIQ em ordem crescente de horas

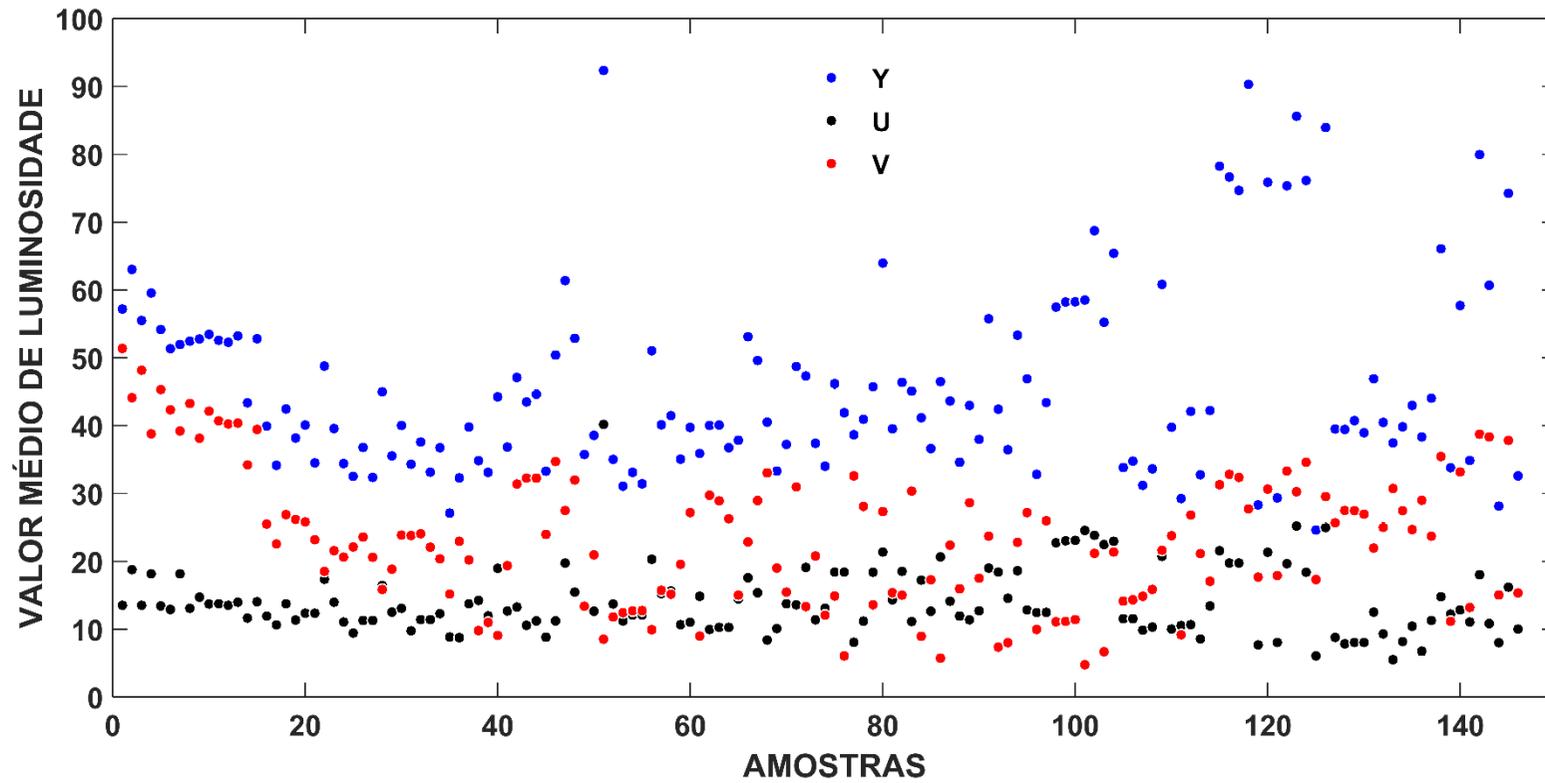


Figura 98. Valores médios de luminosidade para o sistema YUV em ordem crescente de horas

A.5 Termo de Consentimento Livre e Esclarecido apresentado para os participantes da pesquisa

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título da pesquisa: Estudo de manchas de sangue: uma abordagem forense empregando espectroscopia Raman e Quimiometria

Você está sendo convidado(a) a participar de um estudo que tem por objetivo avaliar a degradação do sangue quando exposto ao ambiente e correlacionar as mudanças químicas que ocorrem na mancha de sangue exposta com o tempo. O desenvolvimento de uma metodologia analítica para datação de manchas de sangue em cenas de crime pode ajudar peritos a estabelecer a data em que um crime ocorreu.

A amostra de sangue será obtida por punção digital empregando lanceta descartável com retração automática da agulha. Será realizada uma punção na lateral do dedo e coletadas 2 gotas (40 µL) em lâminas de vidro. A amostra será utilizada apenas para fins da presente pesquisa e posteriormente descartadas em caixas perfuro-cortantes apropriadas. Se você concordar com o uso de suas informações e/ou do material do modo descrito acima, é necessário esclarecer que você não terá quaisquer benefícios ou direitos financeiros sobre eventuais resultados decorrentes desta pesquisa. Em caso de recusa, você não será penalizado(a) de forma alguma.

Os riscos associados com a coleta de sangue incluem: pequeno desconforto no local da coleta, como dor de intensidade leve e hematomas. Os riscos desta coleta serão minimizados por profissional experiente. Em caso de dor, que geralmente é ausente ou mínima, e de curta duração, providenciaremos uma compressa de gelo no local, caso queira. Não estão previstos quaisquer outros tipos de constrangimentos, exceto sua disponibilidade de tempo para leitura desse termo e a coleta do material para a pesquisa. Fica garantido o direito de você retirar o seu consentimento a qualquer tempo. O seu nome não será divulgado em nenhum momento.

Em caso de dúvidas em relação à pesquisa, você poderá solicitar quaisquer informações adicionais e a qualquer tempo aos pesquisadores responsáveis, nos seguintes contatos:

Profa. Dra. Mariana Ramos de Almeida (telefone de contato: (31) 3409-5755, (31) 98257-0549; email: mariramosalmeida@gmail.com)

Prof. Dr. Leonardo de Souza Vasconcellos (telefone de contato: 3409-9774)

Em caso de dúvidas em relação aos aspectos éticos desta pesquisa, você poderá entrar em contato com o Comitê de Ética em Pesquisa da UFMG (COEP):
Telefone: (31) 3409-4592 Endereço: Av. Presidente Antônio Carlos, 6627 – Unidade Administrativa II, Sala 2005 – 2º andar – Pampulha, Belo Horizonte/MG - CEP: 31270-901.

Rúbrica participante

Rúbrica pesquisador

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Título da pesquisa: Estudo de manchas de sangue: uma abordagem forense empregando espectroscopia Raman e Quimiometria

Após ser esclarecido(a) sobre as informações anteriores, no caso de aceitar fazer parte do estudo, assine ao final deste documento, que está em duas vias. Uma delas é sua e a outra é do pesquisador responsável.

Eu, _____, declaro que li as informações do TCLE, esclareci minhas dúvidas, aceitei participar do projeto de pesquisa “Estudo de manchas de sangue: uma abordagem forense empregando espectroscopia Raman e Quimiometria” e o assino livremente. Fui devidamente informado e esclarecido sobre a pesquisa, os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios decorrentes de minha participação. Autorizo a utilização dos dados coletados para os fins científicos relatados. Foi-me garantido o sigilo das informações e sei que posso retirar meu consentimento a qualquer momento, sem que isto leve a qualquer penalidade.