



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

BRENO HENRIQUE DANTAS DOS SANTOS

**AVALIAÇÃO DE EQUIPAMENTOS NIR E RAMAN PORTÁTEIS E TÉCNICAS DE  
APRENDIZAGEM DE MÁQUINA PARA ESTIMAR O TEMPO DE DEPOSIÇÃO DE  
SANGUE EM PISOS PARA FINS FORENSES**

Recife

2020

BRENO HENRIQUE DANTAS DOS SANTOS

**AVALIAÇÃO DE EQUIPAMENTOS NIR E RAMAN PORTÁTEIS E TÉCNICAS DE APRENDIZAGEM DE MÁQUINA PARA ESTIMAR O TEMPO DE DEPOSIÇÃO DE SANGUE EM PISOS PARA FINS FORENSES**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito à obtenção do título de Mestre em Engenharia Química.

**Área de concentração:** Engenharia de Processos Químicos e Bioquímicos;

**Orientadora:** Prof. Dra. Fernanda A. Honorato

**Coorientadora:** Prof. Dra. Maria Fernanda Pimentel

Recife  
2020

Catálogo na fonte  
Bibliotecário Gabriel Luz, CRB-4 / 2222

- S237a Santos, Breno Henrique Dantas dos.  
Avaliação de equipamentos NIR e raman portáteis e técnicas de aprendizagem de máquina para estimar o tempo de deposição de sangue em pisos para fins forenses / Breno Henrique Dantas dos Santos – Recife, 2020.  
108 f.: fig., tabs., abrev. e siglas, simbol.
- Orientadora: Profa. Dra. Fernanda Araújo Honorato.  
Coorientadora: Prof. Dra. Maria Fernanda Pimentel.  
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG.  
Programa de Pós-Graduação em Engenharia Química, 2020.  
Inclui referências e anexo.
1. Engenharia Química. 2. Ciências forenses. 3. Manchas de sangue. 4. Equipamentos portáteis. 5. SVM. 6. I. Honorato, Fernanda Araújo (Orientadora). II. Pimentel, Maria Fernanda (Coorientadora). III. Título.

UFPE

660.2 CDD (22. ed.)

BCTG / 2020-238

BRENO HENRIQUE DANTAS DOS SANTOS

**AVALIAÇÃO DE EQUIPAMENTOS NIR E RAMAN PORTÁTEIS E TÉCNICAS DE APRENDIZAGEM DE MÁQUINA PARA ESTIMAR O TEMPO DE DEPOSIÇÃO DE SANGUE EM PISOS PARA FINS FORENSES**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Engenharia Química.

Aprovada em: 15/ 06/ 2020.

**BANCA EXAMINADORA**

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Fernanda Araújo Honorato(Orientadora)  
Universidade Federal de Pernambuco

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Maria Fernanda Pimentel(Coorientadora)  
Universidade Federal dePernambuco

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Daniella Carla Napoleão (ExaminadoraInterna)  
Universidade Federal de Pernambuco

---

Prof. Dr. José Geraldo de Andrade Pacheco Filho(ExaminadorInterno)  
Universidade Federal de Pernambuco

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Maria José de Filgueiras Gomes (Examinadora Externa)  
Universidade Federal Rural de Pernambuco

## **AGRADECIMENTOS**

A dádiva da vida, da minha saúde e das oportunidades que surgiram permitindo o estudo e o conhecimento, me fazem agradecer primeiramente a Deus!

Agradeço a minha esposa Tatiane por me apoiar durante toda a graduação e desde a inscrição no Mestrado, me ajudando a organizar os documentos necessários, até a finalização, sempre me apoiando!

A minha família que me mostra desde criança que os bons caminhos da vida surgem com o estudo, que nunca é fácil, mas é sempre gratificante!

A todos que me apoiaram, que se importaram com meu estudo, amigos, colegas de turma, amigos de perícia e a todos do Laboratório de Combustíveis (LAC).

A Francielson por me mostrar que é ajudando que se aprende e que se conquista amizades duradouras, agradeço a dedicação em me ajudar sempre!

As Professoras Maria Fernanda Pimentel e Fernanda Honorato, por serem exemplo de dedicação e carinho com os alunos, por disponibilizar o pouco tempo livre para me orientar! Tomo-as como exemplo de profissionais para o resto da vida!

A Coordenação de Aperfeiçoamento de Ensino Superior (CAPES) pela bolsa concedida durante todo o Mestrado.

Como também ao suporte financeiro promovido pelos projetos NUQAAPE/FACEPE, NEQUIFOR/PROFORENSE/CAPES, INCTAA e CNPq.

## RESUMO

Ciência forense é a ciência que auxilia a justiça na resolução de delitos. Os vestígios produzidos durante a prática da ação delituosa são de extrema importância para relacionar uma autoria e motivação. O sangue é um vestígio muito comum, sendo um importante elemento para o entendimento dos fatos. Através da extração de DNA e da análise do padrão de distribuição das manchas de sangue, por exemplo, permite-se elucidar a dinâmica do crime e desvendar a sequência de eventos. Além disso, a partir do tempo de deposição da mancha de sangue, permite-se indicar o tempo transcorrido após um determinado crime ocorrer. As espectroscopias Raman e na região do infravermelho próximo (NIR) associadas a equipamentos portáteis são de grande interesse para analisar o sangue como vestígio diretamente no local do crime. Desta forma, este estudo buscou avaliar a possibilidade de datar o tempo de deposição do sangue considerando a sua degradação, utilizando a análise espectral por Raman e NIR. Para tanto, o sangue coletado de quatro diferentes doadores do sexo masculino foi depositado nas superfícies cerâmica e porcelanato, em seguida as amostras foram submetidas a diferentes condições de temperatura (de 9, 20, 27 e 45°C) durante 45 dias. Os espectros foram analisados por componentes principais (PCA), regressão por mínimos quadrados parciais (PLS) e máquinas de vetores de suporte (SVM), aplicadas após correção dos espectros por diferentes pré-processamentos. Para os espectros NIR, pôde ser observado a intensa influência da água principalmente até 24 horas após a deposição da amostra. A PCA evidenciou uma separação entre amostras recentes (com menos de 24 horas de deposição) e amostras antigas (após 24 horas). Os modelos PLS construídos para datação não se mostraram satisfatórios, obtendo-se altos erros de predição devido à complexidade bioquímica da degradação do sangue. Modelos SVM, na faixa de 1 a 49 dias, tiveram valores de RMSEP de cerca de 3 dias para ambos os substratos. Esta técnica aliada a outros exames do local podem indicar um tempo estimado da prática do crime. Já para os dados empregando a espectroscopia Raman, a influência do substrato nos espectros aliada à baixa intensidade do sinal Raman para amostras biológicas não permitiu observar os componentes do sangue. Apesar de ter utilizado uma metodologia testada previamente tendo o alumínio como substrato.

**Palavras-chave:** Ciências forenses. Manchas de sangue. Equipamentos portáteis. SVM.

## ABSTRACT

Forensic science is the science that assists justice in the resolution of crimes. The traces produced during the practice of criminal action are extremely important to relate an authorship and motivation. Blood is a very common vestige, being an important element for understanding the facts. Through DNA extraction and analysis of the distribution pattern of bloodstains, for example, it is possible to elucidate the dynamics of crime and unravel the sequence of events. In addition, from the time of deposition of the bloodstain, it is possible to indicate the time elapsed after a certain crime occurs. Raman and near infrared (NIR) spectroscopies associated with portable equipment are of great interest to analyze blood as a trace directly at the crime scene. Thus, this study aimed to evaluate the possibility of dating the time of blood deposition considering its degradation, using spectral analysis by Raman and NIR. Therefore, the blood collected from four different male donors was deposited on the ceramic and porcelain surfaces, then the samples were submitted to different temperature conditions (9, 20, 27 and 45°C) for 45 days. The spectra were analyzed by principal component analysis (PCA), partial least squares regression (PLS) and support vector machine (SVM), applied after correction of the spectra by different pre-processing. For NIR spectra, the intense influence of water could be observed mainly up to 24 hours after sample deposition. PCA showed a separation between recent samples (with less than 24 hours of deposition) and old samples (after 24 hours). The PLS models constructed for dating were not satisfactory, obtaining high prediction errors due to the complexity of blood degradation biochemistry. SVM models, in the range of 1 to 49 days, had RMSEP values of about 3 days for both substrates. This technique combined with other examinations of the site may indicate an estimated time of the practice of the crime. For the data using Raman spectroscopy, the influence of the substrate on the spectra combined with the low intensity of the Raman signal for biological samples did not allow observing the blood components. Although it used a methodology previously tested with aluminum as substrate.

**Keywords:** Forensic science. Blood. Portable equipment. SVM.

## LISTA DE FIGURAS

Figura 1 - Estrutura da molécula da hemoglobina com os grupamentos Heme .....	20
Figura 2 - Desnaturação da hemoglobina .....	20
Figura 3 - Principais parâmetros definidos para o SVMR.....	38
Figura 4 - Esquema de deposição das manchas de sangue nas placas .....	50
Figura 5 - Espectros NIR das manchas de sangue na cerâmica nas temperatura de 9°C (CE09), 20°C (CE20), 27°C (CE27) e 43°C(CE43).....	54
Figura 6 - Espectros brutos das manchas de sangue sobre a cerâmica para diferentes temperaturas: em azul os espectros nos tempos iniciais (10 horas para 9°C; 2 horas para 20 e 27°C; 20 minutos para 43°C); em vermelho os espectros para os demais tempos .....	55
Figura 7 - Comparação entre a média dos espectros brutos, nos tempos de 0 minutos a 24 horas (em azul) e dos espectros dos tempos de 24 horas a 45 dias (em vermelho) e do espectro de referência do substrato cerâmico (em preto), para a temperatura de 27°C .....	56
Figura 8 - Comparação entre a média dos espectros brutos, nos tempos de 0 minutos a 24 horas (em azul) e dos espectros dos tempos de 24 horas a 45 dias (em vermelho) após a remoção do espectro de referência do substrato para a temperatura de 27°C .....	57
Figura 9 - Dados truncados em 1100-1650 nm e pré-processados aplicando a derivada, normalização pela faixa e centralização na média.....	58
Figura 10 - Gráfico dos Escores para os dados pré-processados por derivada, considerando todos os tempos.....	59
Figura 11 - Gráfico dos pesos para os dados pré-processados por derivada, considerando todos os tempos.....	60
Figura 12 - a) Gráfico dos escores para amostras frescas (< 24 horas); b) Gráfico dos escores para amostras antigas (≥24 horas). .....	62
Figura 13 - a) Gráfico dos pesos para amostras frescas (< 24 horas); b) Gráfico dos pesos para amostras antigas (≥ 24 horas). .....	63

Figura 14 - Espectros da mancha de sangue na cerâmica a 27°C após o pré-processamento SNV .....	64
Figura 15 - Gráfico dos escores para os dados pré-processados por SNV, considerando todos os tempos. ....	65
Figura 16 - Gráfico dos pesos para os dados pré-processados por SNV, considerando todos os tempos .....	65
Figura 17 - a) Gráfico dos escores para amostras frescas pré-processados por SNV (< 24 horas); b) Gráfico dos escores para amostras antigas pré-processados por SNV (≥ 24 horas) .....	66
Figura 18 - a) Gráfico dos pesos para amostras frescas pré-processados por SNV (< 24 horas); b) Gráfico dos pesos para amostras antigas pré-processados por SNV (≥ 24 horas).....	67
Figura 19 - a) Y CV Predito <i>versus</i> Y Medido para amostras frescas pré-processados por SNV (< 24 horas); b) Y CV Predito <i>versus</i> Y Medido para amostras antigas pré-processados por SNV (de 24 horas até 15 dias). ....	69
Figura 20 - a) Y CV Predito (tempo em horas) <i>versus</i> Y Medido (tempo em horas) para amostras frescas (< 24 horas); b) Y CV Predito (tempo em horas) <i>versus</i> Y Medido (tempo em horas) para amostras antigas (de 24 horas até 15 dias) .....	70
Figura 21 - a) Dados ampliados brutos, acima os dados introduzidos, abaixo os dados coletados para esta pesquisa; b)Dados ampliados após aplicação do pré-processamento SNV.....	73
Figura 22 - Valores preditos <i>versus</i> referência do modelo SVM para os Dados 4 na cerâmica. Dados de calibração em cinza e os dados de validação externa em vermelho .....	77
Figura 23 - Espectros no infravermelho próximo obtidos em equipamento portátil para a placa de porcelanato nas temperaturas de 9°C (PO09), 20°C (PO20), 27°C (PO27) e 43°C (PO43).....	79
Figura 24 - Comparação entre a média dos espectros nos tempos iniciais (em azul), dos espectros restantes (em vermelho) e do espectro de referência do substrato porcelanato (em preto).....	80

Figura 25 - Média dos espectros da mancha de sangue na cerâmica nos tempos iniciais (em azul) e nos tempos seguintes (em vermelho) após a remoção do espectro de referência do substrato.....	80
Figura 26 - Valores preditos versus referência do modelo SVM para os Dados 3 no porcelanato. Dados de calibração em cinza e os dados de validação externa em vermelho. ....	84
Figura 27 - Espectro Raman de sangue depositado no substrato de alumínio, obtido pela subtração direta do espectro de referência do espectro da amostra .....	86
Figura 28 - Espectros Raman das manchas de sangue depositadas na cerâmica a 9°C (CE09), 20°C (CE20), 27°C (CE27) e 43°C (CE43) em todos os tempos.....	87
Figura 29 - Média dos tempos dos espectros Raman na cerâmica. ....	88
Figura 30 - Espectro de referência da placa de cerâmica .....	88
Figura 31 - Espectros das amostras (em preto) na comparação com o espectro de referência (em vermelho).....	89
Figura 32 - Espectro resultante após a subtração do espectro de referência do substrato ao espectro da amostra a 27°C.....	90
Figura 33 - Média do espectro resultante da subtração dos dados brutos com o espectro de referência para 27°C .....	91
Figura 34 - Comparação entre os espectros das amostras (em preto) e do espectro de referência (em vermelho) após a normalização pela faixa. ....	91
Figura 35 - Espectros resultantes da subtração dos espectros de referência com espectros das amostras previamente normalizados.....	92
Figura 36 - Média dos espectros resultantes da subtração dos espectros de referência com espectros das amostras previamente normalizados.....	93
Figura 37 - Média dos espectros Raman na cerâmica com aplicação do AsLS para diferentes tempos.....	93
Figura 38 - Espectros da amostra a 27°C após aplicação do AsLS da normalização e após a subtração dos dados com o espectro de referência que teve correção de linha de base e normalização.....	94
Figura 39 - Média do espectro resultante, com os picos característicos de ligações moleculares correspondentes a componentes do sangue .....	95

Figura 40 - Amostras de sangue depositadas em porcelanato a 9°C (PO09), 20°C (PO20), 27°C (PO27) e 430C (PO43) para todos os tempos .....	96
Figura 41 - Média dos tempos dos espectros Raman no porcelanato.....	97
Figura 42 - Média do espectro resultante após a subtração dos espectros pré-processados por AsLS .....	98

## LISTA DE TABELAS

Tabela 1 - Modos vibracionais de absorção dos componentes do sangue no NIR.....	41
Tabela 2 - Modos vibracionais de absorção dos componentes do sangue no Raman .....	45
Tabela 3 - Valores obtidos da regressão PLS após pré-processamento por derivada para diferentes intervalos de tempo.....	70
Tabela 4 - Valores obtidos da regressão PLS após pré-processamento SNV .....	71
Tabela 5 - Valores obtidos da regressão PLS das médias das triplicatas após pré-processamento derivada e SNV .....	72
Tabela 6 - Valores obtidos com o PLS para os dados ampliados.....	75
Tabela 7 - Valores obtidos para os modelos PLS para as médias das replicatas dos dados ampliados.....	76
Tabela 8 - Conjunto de dados selecionados para calibração e para validação externa.....	77
Tabela 9 - Parâmetros dos modelos SVM com diferentes combinações de dados para calibrar e validar externamente .....	77
Tabela 10 - Valores obtidos da Regressão PLS das médias das triplicatas após pré- processamento derivada e SNV .....	81
Tabela 11 - Valores obtidos com o PLS para as médias das triplicatas .....	82
Tabela 12 - Parâmetros do PLS para intervalos de tempo da média das replicatas dos dados ampliados das amostras. ....	83
Tabela 13 - Parâmetros dos modelos SVM com diferentes combinações de dados para calibrar e validar externamente as amostras depositadas em porcelanato .....	83

## LISTA DE ABREVIATURAS E SIGLAS

CEP-UFPE	Comitê de Ética em Pesquisa envolvendo Seres Humanos da Universidade Federal de Pernambuco
CCR	Taxa de correta classificação ( <i>corrected classification rate</i> )
desoxiHb	Desoxi-hemoglobina
EMSC	Extensão de Correção de Sinal Multivariado ( <i>Extended Multivariate Signal Correction</i> )
FIR	Infravermelho distante ( <i>far infrared</i> )
FT-NIR	Infravermelho Próximo por Transformada de Fourier
Hb	Hemoglobina
HbO <sub>2</sub>	Oxi-Hemoglobina
Hc	Hemicromo
IV	Infravermelho
InGaAs	Arseneto de gálio e Índio
LDA	Análise Discriminante Linear ( <i>Linear Discriminant Analysis</i> )
LVF	Filtro linear variável ( <i>Linear Variable Filter</i> )
MSC	Correção de Espalhamento Multiplicativa ( <i>Multiplicative Scatter Correction</i> )
MIR	Infravermelho Médio ( <i>middle infrared</i> )
metHb	Meta-hemoglobina
NIR	Infravermelho Próximo ( <i>near Infrared</i> )
PLS-DA	Análise discriminante por Mínimos Quadrados Parciais ( <i>Partial Least Square Discriminant Analysis</i> )
PLS	Mínimos Quadrados Parciais ( <i>Partial Least Square</i> )
PC	Componente Principal
PCA	Análise de Componentes Principais
RMSEP	Raiz Quadrada do Erro Médio Quadrático de Previsão
RMSEC	Raiz Quadrada do Erro Médio Quadrático de Calibração
RMSECV	Raiz Quadrada do Erro Médio Quadrático de Validação Cruzada ( <i>Cross-Validation</i> )

SG	Savitzky-Golay
SNV	Varição Normal Padrão ( <i>Standard Normal Variate</i> )
SVM	Máquina de vetores de suporte ( <i>Support Vector Machines</i> )
SVMR	Regressão por máquina de vetores de suporte ( <i>Support Vector Machines Regression</i> )
TCLE	Termo de Consentimento de Livre Esclarecimento
UV-Vis	Ultravioleta-Visível
VL	Variável latente
VS	Vetor de suporte
WLS	Correção de linha de base com mínimos quadrados ponderados ( <i>weighted least squares</i> )

## LISTA DE SÍMBOLOS

$\alpha_1$ e $\alpha_2$	Cadeias polipeptídicas $\alpha$
$\beta_1$ e $\beta_2$	Cadeias polipeptídicas $\beta$
$R^2$	Coefficiente de determinação
$\beta_0$ e $\beta_1$	Coefficientes mínimos quadrados
$\hat{B}$	Coefficiente de regressão do modelo
<b>T</b>	Escores
$X_i$	Espectro
$X_i^*$	Espectro corrigido por MSC
$X_r$	Espectro de referência
$N$	Frequência do laser
$I$	Fótons
$^{\circ}\text{C}$	Graus Celsius
$H$	Hora
<b>X</b>	Matriz original
<b>E</b>	Matriz residual
$\nu$	Modo vibracional normal
nm	Nanômetro
<b>P</b>	Pesos
$t$	Tempo
$X_{ij}$	Valores individuais de uma coluna
$\bar{X}_j$	Valores médios de uma coluna
$\xi$	Variável de folga

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>17</b>
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA .....</b>	<b>19</b>
2.1	BIOQUÍMICA DO SANGUE E DA MANCHA DE SANGUE .....	19
2.2	ESPECTROSCOPIA POR ABSORÇÃO NO INFRAVERMELHO .....	22
2.3	ESPECTROSCOPIA POR ESPALHAMENTO RAMAN .....	25
2.4	QUIMIOMETRIA.....	30
<b>2.4.1</b>	<b>Pré-processamento .....</b>	<b>31</b>
2.4.1.1	Centragem na Média .....	31
2.4.1.2	Normalização.....	32
2.4.1.3	Suavização (Filtros Savitzky Golay).....	32
2.4.1.4	Derivadas.....	33
2.4.1.5	Correção de Espalhamento Multiplicativa e Variação Normal Padrão.....	33
2.4.1.6	Suavização por Mínimos Quadrados Assimétricos.....	34
<b>2.4.2</b>	<b>Análise Exploratória e Calibração Multivariada .....</b>	<b>35</b>
2.4.2.1	Análise Exploratória: Análise por Componentes Principais .....	35
2.4.2.2	Calibração Multivariada: Regressão por Mínimos Quadrados Parciais .....	36
2.4.2.3	Máquinas de Vetores de Suporte .....	37
<b>2.4.3</b>	<b>Parâmetros para avaliar a qualidade do modelo .....</b>	<b>38</b>
2.5	ESPECTROSCOPIA E QUIMIOMETRIA NAS MANCHAS DE SANGUE.....	40
<b>3</b>	<b>METODOLOGIA .....</b>	<b>49</b>
3.1	OBTENÇÃO DAS MANCHAS DE SANGUE.....	49
3.2	ANÁLISE COM INFRAVERMELHO E RAMAN PORTÁTEIS.....	50
<b>3.2.1</b>	<b>Espectros Totais no IV .....</b>	<b>52</b>
3.3	TRATAMENTO DE DADOS .....	52
<b>3.3.1</b>	<b>Valores esperados dos parâmetros de validação dos modelos de calibração.....</b>	<b>53</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>54</b>
4.1	ESPECTROSCOPIA NIR.....	54
<b>4.1.1</b>	<b>Espectros NIR no substrato de cerâmica .....</b>	<b>54</b>
4.1.1.1	Análise Exploratória.....	54

4.1.1.2	Modelos de calibração para datação das manchas .....	68
<b>4.1.2</b>	<b>Espectros NIR substrato de porcelanato.....</b>	<b>79</b>
4.2	ESPECTROSCOPIA RAMAN .....	85
<b>4.2.1</b>	<b>Espectros Raman no substrato de cerâmica .....</b>	<b>87</b>
<b>4.2.2</b>	<b>Espectros Raman no Substrato de porcelanato.....</b>	<b>96</b>
<b>5</b>	<b>CONCLUSÃO .....</b>	<b>100</b>
<b>6</b>	<b>PERSPECTIVAS PARA TRABALHOS FUTUROS .....</b>	<b>101</b>
	<b>REFERÊNCIAS .....</b>	<b>102</b>
	<b>ANEXO A – TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO</b>	
	<b>(TCLE) .....</b>	<b>107</b>

## 1 INTRODUÇÃO

A ciência forense é formada pela junção de conhecimentos científicos com o objetivo de auxiliar a justiça na resolução de delitos. A perícia criminal se utiliza desta ciência para materializar provas por análise de vestígios e assim obter informações sobre o fato criminoso.

O sangue é um dos vestígios mais encontrados em locais de crime, por ser bioquimicamente rico ele fornece informações particulares sobre os possíveis agentes do fato delituoso. Uma das formas de analisar o sangue é a partir do perfil genético, sendo possível determinar o DNA de um indivíduo e comparar se o mesmo é compatível com o encontrado na cena de crime (MURO *et al.*, 2016). Além disso, análises do padrão de distribuição da mancha de sangue podem mostrar onde e com qual instrumento uma vítima foi atingida (BEVEL *et al.*, 2002; MACIEL, 2014), estimar o tempo de deposição do fluido em uma superfície específica, podendo assim determinar quando o crime ocorreu (DOTY *et al.*, 2016; DOTY *et al.*, 2017), distinguir a raça entre afro-americano e caucasiano (MISTEK *et al.*, 2016), determinar a idade (AGUDELO *et al.*, 2016; DOTY *et al.*, 2018; FUJIHARA *et al.*, 2019) e o gênero (SIKIRZHYTSKAYA *et al.*, 2017) do depositador da mancha, que podem ser de suspeitos, vítimas, curiosos ou até mesmo de animais.

O primeiro a enfatizar a mancha de sangue como vestígio foi Edward Piotrowski: “É de grande importância para o campo da medicina legal dar maior atenção às manchas de sangue encontradas na cena de um crime, porque elas podem lançar luz sobre um assassinato e fornecer uma explicação para os momentos essenciais do incidente” (PIOTROWSKI, 1895). Assim que o sangue entra em contato com o ambiente externo e se deposita em uma superfície ocorrem diversos processos físico-químicos irreversíveis, como a perda de água do plasma ocasionando a formação da mancha (ZADORA *et al.*, 2018). Além do processo de degradação ao longo do tempo, dependente das condições externas (BREMNER *et al.*, 2011). Devido a essas características têm-se buscado tecnologias que possam auxiliar as análises periciais de forma mais rápida, prática e menos oneroso.

As espectroscopias na região do infravermelho e Raman têm se mostrado promissoras para avaliação do fluido biológico, podendo identificar os componentes da amostra de forma não destrutiva e com confiabilidade (EDELMAN *et al.*, 2012). Assim, técnicas espectroscópicas vibracionais têm sido muito aplicadas nas ciências forenses devido às suas vantagens, como a capacidade de determinação de amostras em diferentes estados físicos, facilidade de

implementação, preservação dos componentes das amostras e rapidez na aquisição de espectros (SKOOG *et al.*, 2014).

Além do fato destas técnicas permitirem análises *in loco* devido à portabilidade dos equipamentos, uma vez que vestígios biológicos como o sangue são encontrados em locais diversos, contornando a dificuldade de coletar e transportar amostras depositadas nos substratos. Estes equipamentos portáteis, que cada vez mais apresentam tecnologia de baixo custo, alta reprodutibilidade, alta sensibilidade, sem consumo de reagentes e sem preparo das amostras, tem facilitado as perícias em locais ermos e com muitos vestígios. Aliando as técnicas espectroscópicas a ferramentas quimiométricas adequadas, pode-se obter informações confiáveis provenientes dos espectros como a construção de modelos de regressão (análise quantitativa) e modelos de classificação (análise qualitativa) de forma a elucidar questões forenses (JACKSON *et al.*, 2001).

Este trabalho tem o diferencial por buscar análises espectroscópicas para datação de manchas de sangue de forma mais simples e objetiva, isto é, sem preparo e coleta da amostra, sem alterar o substrato ao qual a mesma está depositada, utilizando equipamentos portáteis para analisar diretamente na cena do crime e considerando as variações de temperatura. Além do mais, há o comparativo de duas técnicas espectroscópicas promissoras para a área forense e combinações de pré-processamentos para determinar qual método melhor se aplica à amostra em questão.

Com isso, esse trabalho tem como objetivo geral avaliar a possibilidade de datar manchas de sangue humano depositadas em pisos, empregando espectroscopia na região do infravermelho próximo (NIR) e espalhamento Raman associadas às técnicas quimiométricas para tratamento dos dados. Como objetivos específicos pretende-se: avaliar a eficiência de equipamentos NIR e Raman portáteis, que podem ser facilmente levados ao local do crime; avaliar os efeitos da temperatura (de 9°C a 43°C) nos espectros obtidos ao longo do tempo (até 45 dias); avaliar o efeito dos tipos de substratos (porcelanato e cerâmica) ao construir modelos empregando as técnicas de regressão por mínimos quadrados parciais (PLS) e por máquinas de vetores de suporte (SVM), na intenção de fazer um estudo temporal e poder estimar de forma mais precisa a datação da deposição do sangue.

## 2 REVISÃO BIBLIOGRÁFICA

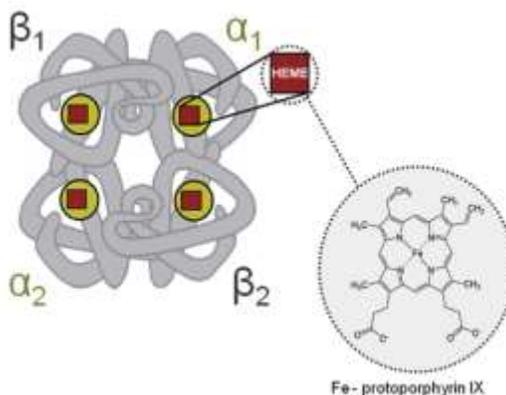
Os tópicos a seguir constituem uma fundamentação teórica sobre a composição do sangue e como sua bioquímica pode ser de interesse forense; as técnicas espectroscópicas que se aplicam na análise do sangue como vestígio, isto é, a técnica de absorção por infravermelho próximo (NIR) e por espalhamento do sinal Raman; e também pela conceituação de técnicas quimiométricas para analisar de forma multivariada os dados espectroscópicos. Esta revisão bibliográfica traz também a aplicação das técnicas espectroscópicas aliadas às técnicas quimiométricas para analisar o sangue por análise multivariada de classificação e calibração, previamente aplicada em outros estudos.

### 2.1 BIOQUÍMICA DO SANGUE E DA MANCHA DE SANGUE

O sangue é composto basicamente por plasma que se encontra em maior quantidade, sendo esta a parte líquida contendo cerca de 90% de água, além das hemácias, que são os glóbulos vermelhos, leucócitos, que são os glóbulos brancos, e plaquetas que são os fragmentos celulares (ZADORA *et al.*, 2018).

Quando externo ao corpo, o sangue passa por processos físico-químicos ocasionando a sua desidratação formando a mancha, ficando com composição de cerca de 90% de hemoglobina (Hb) (proteína das hemácias). Este é o principal cromóforo no sangue e se apresenta de várias formas, chamadas de derivativos da hemoglobina. Dentro do corpo humano a hemoglobina se apresenta de duas formas: sem o oxigênio, chamada de desoxi-hemoglobina (desoxiHb) e saturado de oxigênio, conhecida como a oxi-hemoglobina (HbO<sub>2</sub>) (SHIKAMA *et al.*, 1998). A Hb faz parte das proteínas com grupo heme, tendo a composição de ferro na sua molécula que se liga ao O<sub>2</sub> fazendo o transporte deste, oxigenando todo o organismo. Cada molécula de hemoglobina humana é composta de quatro hemes contendo cadeias polipeptídicas, sendo duas  $\alpha$  idênticas ( $\alpha 1$  e  $\alpha 2$ ) e duas  $\beta$  idênticas ( $\beta 1$  e  $\beta 2$ ) (Figura 1) (BERG *et al.*, 2002). O grupo heme é uma molécula formada por quatro anéis pirrólicos ligados a um átomo central de ferro. Além destas ligações, o ferro forma duas ligações adicionais, uma em cada lado do plano do grupo heme, podendo ser de moléculas de oxigênio apenas se o ferro (Fe) estiver no estado de oxidação de Fe<sup>+2</sup>, fazendo com que o grupo heme da hemoglobina seja o responsável pelo transporte do oxigênio (ZADORA *et al.*, 2018).

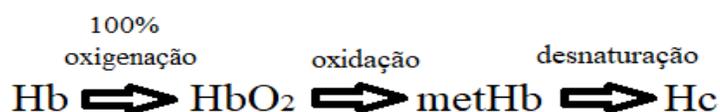
Figura 1 – Estrutura da molécula da hemoglobina com os grupamentos Heme.



Fonte: Adaptado de ZADORA *et al.* (2018)

Fora do corpo a hemoglobina é totalmente oxidada devido à presença do  $O_2$  do ambiente, formando  $HbO_2$ . Como não há mais enzimas atuando, a reação da  $HbO_2$  para meta-hemoglobina (metHb) se torna irreversível (MARRONE *et al.*, 2009). Ocorre por meio da auto-oxidação do  $Fe^{+2}$  a  $Fe^{+3}$  a uma taxa aproximada de 3% ao dia, formando a meta-hemoglobina (metHb) e um radical superóxido (UMBREIT *et al.*, 2007). A meta-hemoglobina, então sofre desnaturação irreversível se transformando em hemicromo (Hc) (Figura 2) (ZADORA *et al.*, 2018). O envelhecimento da mancha de sangue é caracterizado por tais mudanças físico-químicas na proteína, que paralelamente muda sua conformação à medida que ocorre sua degradação.

Figura 2 - Desnaturação da hemoglobina.



Fonte: O autor (2020).

Em termos físicos, assim que deixa o corpo humano, o sangue sofre um processo de desidratação fazendo com que a sua coloração varie de um vermelho intenso para um marrom escuro. Paralelamente, se inicia o processo de coagulação, que irá aumentar a viscosidade, formando um material mucilaginoso. Em seguida se torna gelatinoso, com a água do plasma continuamente evaporando, formando a mancha (RAMSTHALER *et al.*, 2016). Balthazard *et al.* (1939) perceberam que as manchas de sangue começam a se formar pela secagem das periferias do

sangue depositado que é a parte mais espessa, seguindo para o centro que é a parte menos espessa, isto devido ao impacto da deposição. Por se tratar de uma suspensão coloidal, o sangue após longos dias fora do corpo começa a se tornar um sólido, surgindo com o tempo fendas paralelas na mancha, similares a rachaduras conhecidas como *cracks*, definido por Allain *et al.* (1995) como o padrão formado pela secagem direcionada de uma fina camada de uma suspensão coloidal. Os *cracks* ou rachaduras, são frestas que aumentam gradativamente e são dependentes do tipo de sangue. Seu aumento ocorre devido à competição mecânica, proveniente da abertura da rachadura, com o aumento do estresse mecânico devido à evaporação. Logo, quanto maior o diâmetro da mancha, mais as bordas terão espessura menor e assim a abertura de *cracks* será maior e mais rápida (ALLAIN *et al.*, 1995).

Paralelamente a formação da mancha de sangue, há o processo de degradação, ocasionando alteração na sua composição bioquímica, que é influenciada por fatores internos como por exemplo, raça, gênero, idade, nível de hidratação do doador, doenças associadas à hemofilia e a metemoglobinemia, que afetam a coagulação e a concentração de metHb no sangue, entre outros (UMBREIT *et al.*, 2007). Além disso, fatores externos como a umidade, exposição à luz, temperatura, substrato que o sangue foi depositado, entre outros, influenciam nesse processo de degradação. Entre os fatores externos, a temperatura é o maior influenciador da degradação de decomposição do sangue, pois a conversão das diferentes formas de hemoglobina, como a HbO<sub>2</sub> em metHb, seguido da desnaturação da meta-hemoglobina para formar o Hc, é dependente desta. Como exemplo, Hanson *et al.* (2010) demonstraram que a baixa temperatura (20°C) diminui o processo de degradação do sangue. Já a umidade pode ter relativa influência no processo de transferência de massa, interferindo no processo de envelhecimento da mancha (LI *et al.*, 2013). O substrato também é um fator determinante, já que para diferentes substratos (tecido, pisos, entre outros) o sangue se deposita de diferentes formas, influenciando no tempo de degradação (BRODECK, 2012).

Quando a formação da mancha está completa, ou seja, não há mais líquidos presentes, há basicamente células vermelhas na sua composição, sendo possível a partir destas um estudo cinético da degradação para estimar o tempo de deposição (ZADORA *et al.*, 2018). O estudo para determinar a idade da mancha de sangue foi primeiro proposto por Schwarzacher em 1930, que estimou o tempo de deposição avaliando a variação da concentração de água na amostra. Surgiram então os estudos ópticos por espectros de refletância da mancha na região do visível, feitos por

Patterson em 1960, que estudou o tempo de degradação pela mudança de cor do vermelho para marrom. Porém, essa mudança é muito dependente das condições ambientais, logo concluiu que tal metodologia era fonte de diversos erros. Depois de anos de pesquisas, o próprio Patterson recomendou estudos espectrais mais aprofundados para determinação do tempo de deposição (BREMNER *et al.*, 2011). Com isso, vários trabalhos, como exemplo o de Doty *et al.*, em 2018 e Fujihara *et al.*, em 2017 com uso de diferentes técnicas espectroscópicas, aplicando diferentes metodologias e buscando outros resultados têm sido realizados na intenção de coletar informações acerca das transformações do sangue ao longo do tempo.

## 2.2 ESPECTROSCOPIA POR ABSORÇÃO NO INFRAVERMELHO

A região do infravermelho no espectro eletromagnético compreende comprimentos de onda na faixa de 750 a  $10^6$  nm, sendo 750-2500 nm a região do infravermelho próximo (NIR, *near infrared*), 2500-16.000 nm infravermelho médio (MIR, *middle infrared*) e 16.000 a  $10^6$  nm a região do infravermelho distante (FIR, *far infrared*) (PAVIA *et al.*, 2010).

A espectroscopia por absorção no infravermelho (IV) está relacionada com as vibrações correspondentes de cada ligação química, permitindo a identificação de grupos funcionais presentes na molécula. A absorção da radiação IV pelas moléculas ocasiona transições vibracionais que podem ser de estiramento (simétrico ou assimétrico), deformações angulares simétricas no plano (tesoura) e fora do plano, além de deformações angulares assimétricas no plano e fora do plano (torção) das ligações covalentes. São absorvidas apenas energias nas frequências da radiação IV que equivalem às próprias frequências vibracionais que são naturais da molécula, fazendo com que a radiação absorvida aumente a amplitude do movimento vibracional. Contudo, nem todas as ligações, mesmo que possuam igual frequência da radiação são capazes de absorver no IV. A absorção só ocorre para aquelas que possuem momento dipolo variando em função do tempo (PAVIA *et al.*, 2010).

O infravermelho próximo é a região do espectro eletromagnético imediatamente após a região do visível, em termos de comprimento de onda, de onde deriva o nome. Foi primeiro descrito por Herschel em 1800, porém somente foi utilizado como técnica espectroscópica em 1968, por Norris (LIMA *et al.*, 2011). As ocorrências espectrais na região NIR são predominantemente sobretons e combinação de bandas de absorção, relacionados a modos vibracionais que envolvem

átomos leves. Os sobretons ocorrem quando as moléculas sofrem transições do estado fundamental para o segundo ou terceiro estado energético (1° e 2° sobretons). O valor da frequência é correspondente a múltiplos inteiros das frequências de um modo vibracional fundamental ( $\nu$ ), sendo de  $2\nu$ ,  $3\nu$  e assim sucessivamente. Já as bandas de combinação são provenientes da interação entre as frequências de diferentes ligações, surgindo bandas de absorção com frequências no valor aproximado da diferença ou soma de  $\nu$ , sendo  $(\nu_a + \nu_b)$ ,  $(\nu_a - \nu_b)$ . Logo, as bandas de absorção para a combinação, são o resultado das interações da radiação em diferentes ligações químicas. As bandas de absorção no espectro NIR são de baixa intensidade no comparativo com as bandas de absorção da transição entre o estado fundamental e o primeiro estado excitado (transições fundamentais), sendo de até 1000 vezes menor para os sobretons e de até 10 vezes menor para as combinações (HOLLER *et al.*, 2009).

Uma grande variedade de compostos orgânicos e inorgânicos que tenham ligações covalentes absorvem frequências da região eletromagnética no infravermelho. Devido a particularidade de cada ligação, da enorme quantidade de ligações possíveis e de moléculas existentes, um espectro de absorção no IV de uma amostra nunca será igual ao de outra (PAVIA *et al.*, 2010). Tal técnica tem se difundindo em diversas áreas e com isso surge o advento de novos equipamentos e metodologias buscando informações mais confiáveis, rápidas e de baixo custo.

Equipamentos de bancada foram os precursores nas análises espectroscópicas. Para o infravermelho, o início se deu com equipamentos dispersivos. Em seguida surgiram os equipamentos com Transformada de Fourier (FT-IV) que empregam um interferômetro, que gera um sinal complexo mas completo em informações. A Transformada de Fourier (FT) é utilizada para obter um espectro de Intensidade do sinal *versus* Frequência a partir de dados de Intensidade do sinal *versus* Tempo. A vantagem deste sobre os dispersivos é a velocidade de detecção, uma vez que ele produz um interferograma em menos de um segundo, sendo assim mais veloz e de maior sensibilidade. Desta forma os equipamentos dispersivos basicamente foram substituídos pelos equipamentos com transformada de Fourier (PAVIA *et al.*, 2010). Porém, os equipamentos de bancada em geral apesar de terem alta resolução espectral e alta razão sinal ruído eles ocupam grande espaço de laboratório, não podem ser facilmente transportados, além de exigirem alto custo de compra e manutenção. Logo, uma nova tendência surge com o intuito de facilitar ainda mais as análises espectroscópicas com o uso de equipamentos portáteis, que tem tomado cada vez mais

espaço devido a praticidade, baixo custo de compra, de análise, de rápida aplicação, além de ser possível obter resultados com confiabilidade semelhante aos equipamentos de bancada.

Em 2016, Da Silva e colaboradores realizaram uma pesquisa para avaliar a eficiência dos equipamentos portáteis frente aos equipamentos de bancada. No trabalho eles compararam os resultados obtidos com três diferentes instrumentos NIR portáteis com um equipamento de bancada na identificação de polimorfos do mebendazol. Os equipamentos portáteis apresentaram resultados similares e estatisticamente comparáveis ao de bancada (DA SILVA *et al.*, 2016).

Equipamentos portáteis na região NIR vem sendo muito estudados para o uso *in loco*, nos casos onde a análise tem que ser realizada de forma rápida, sem preparo da amostra e no local de coleta, e, portanto, adequados às situações forenses. Um destes usos é nas análises de sangue em locais de crime, já que a hemoglobina na forma de desoxi-hemoglobina (desoxiHb) e oxi-hemoglobina (HbO<sub>2</sub>) possuem pico de absorção com comprimentos de onda iguais a 760 e 920 nm, respectivamente, como foi determinado por Lima *et al.*, em 2011 utilizando equipamento de bancada.

Em 2017, Pereira *et al.* utilizaram o MicroNir 1700 (infravermelho próximo portátil da Viavi Solutions) para analisar diferentes manchas de sangue, em diferentes substratos (porcelana, cerâmica, vidro e metal). Foram depositados sangue humano, sangue animal e substâncias a qual a mancha aparentaria ser sangue (falso-positivos) como batom, vinho, entre outros para avaliar o potencial do equipamento portátil em identificar a classe sangue humano. Após a obtenção dos espectros e o uso de técnicas quimiométricas, os pesquisadores puderam determinar qualitativamente qual a natureza da mancha por meio dos espectros no NIR.

Morillas *et al.*, em 2018, utilizaram o equipamento NIR portátil (SCiO®) que tem faixa espectral de 700 a 1100 nm (NIR), ou seja, a região do terceiro sobretom do espectro NIR, para determinar qualitativamente as manchas de sangue depositadas em lâminas de vidro de laboratório dentre outras manchas de fluidos biológicos. O NIR foi utilizado, uma vez que este consegue determinar classes específicas do sangue em amostras de pequenos volumes 10-20 µL, podendo então ser aplicado em amostras localizadas na cena do crime. Os pesquisadores observaram duas bandas em 760 nm e 970 nm referentes a informação de água, e então selecionaram uma banda entre 800 e 940 nm, no intuito de avaliar apenas a variação da oxi-hemoglobina (que tem pico a 930 nm) sem influência da água. Logo, conseguiram diferenciar sangue de outras manchas com até

94% de confiabilidade e assim sugeriram fazer uma análise temporal com equipamentos portáteis (MORILLAS *et al.*, 2018).

Hespanhol *et al.*, em 2019, utilizaram o equipamento NIR portátil de baixo custo (\$ 1000,00), o NanoNIR (Texas Instruments Inc.), para analisar cocaína apreendida em várias regiões do Brasil, qualitativa e quantitativamente. O intuito era determinar as concentrações, os diluentes, adulterantes, entre outros, além de avaliar a aplicabilidade do equipamento para uso *in loco*, eliminando assim a necessidade de transporte de amostra. Aliando a espectroscopia NIR a técnicas quimiométricas, Hespanhol e colaboradores caracterizaram a cocaína sem preparo e sem destruir a amostra para fazer análise *in loco*, utilizando o NanoNIR. Apesar da metodologia empregada não conseguir determinar todos os diluentes e adulterantes da amostra, devido à alta complexidade da mistura.

### 2.3 ESPECTROSCOPIA POR ESPALHAMENTO RAMAN

A espectroscopia Raman é uma técnica de espalhamento baseada no Efeito Raman que se caracteriza por uma alteração na frequência de uma fração da radiação incidente quando esta é espalhada por moléculas ou estruturas cristalinas. A técnica consegue detectar o espalhamento inelástico dos fótons de uma molécula sendo complementar à técnica espectroscópica de absorção no infravermelho (CIALLA-MAY *et al.*, 2019).

A técnica utiliza uma fonte de luz monocromática (um laser de baixa potência) que ao incidir na amostra tem a radiação espalhada, podendo ser com o mesmo comprimento de onda da radiação incidente (espalhamento elástico - Rayleigh *Scattering*) que representa cerca de 0,1% da radiação incidente; ou com comprimentos de onda diferentes (espalhamento inelástico - Raman *Scattering*), cerca de 1 fóton dentre  $10^6$ - $10^7$  fótons. O espalhamento Raman pode ter baixas frequências (Espalhamento Stokes) ou altas frequências (Espalhamento Anti-Stokes) quando comparado com a luz incidente (IZAKE *et al.*, 2010). O espalhamento inelástico é rico em informações acerca da composição química do analito, uma vez que a diferença de energia entre a radiação incidente e a radiação espalhada é causada pela interação com um estado vibracional da molécula. Portanto, ao se identificar a frequência destas vibrações pode-se determinar quais ligações químicas estão presentes, além de possibilitar a determinação da geometria molecular e a identificação de diferentes fases polimórficas presentes na amostra (FARIA *et al.*, 2002).

A espectroscopia Raman possui como vantagens: cada molécula tem espectro único frente ao espalhamento; pode ser aplicado a uma vasta gama de amostras, contando apenas que tenham interações de espalhamento com a radiação; pode analisar amostras sólidas, líquidas, gasosas, transparentes ou não; capaz de analisar amostras da ordem de  $1 \mu\text{m}^2$  a vários  $\text{dm}^2$ , da ordem de 1 milímetro a vários metros, ou seja, tamanho da amostra não é fator limitante da técnica. Além do mais, a região do *fingerprint* é independente do comprimento de onda da radiação monocromática, as análises não sofrem influência da luz externa e a técnica já conta com equipamentos portáteis (IZAKE *et al.*, 2010), embora o custo seja, em geral, maior que a de um equipamento NIR portátil.

O espectro Raman é uma combinação do sinal da amostra, do sinal do *background* (este que é uma combinação da fluorescência da amostra e do substrato, além dos próprios elementos óticos do equipamento) e de ruído (KERR *et al.*, 2015).

A espectroscopia Raman vem sendo utilizada nas análises bioanalíticas e biomédicas, uma vez que proteínas, lipídios, ácidos nucleicos e carboidratos produzem um significativo espalhamento da radiação. A técnica ainda possui a vantagem de ser não destrutiva, o que é muito importante do ponto de vista forense (CIALLA-MAY *et al.*, 2019).

Algumas dificuldades são encontradas no uso da técnica: a mesma possui baixa sensibilidade, uma vez que geralmente apenas 1 fóton dentre  $10^6$ - $10^7$  sofre espalhamento inelástico (CIALLA-MAY *et al.*, 2019). Outro problema frequente é o da fluorescência indesejável, que somado à baixa intensidade do sinal Raman faz com que haja dificuldades na aquisição de determinados espectros, a depender do equipamento utilizado, do comprimento de onda do laser, do tipo de amostra e do substrato. A escolha precisa do comprimento de onda a ser utilizado é de fundamental importância, já que o número de fótons ( $I$ ) espalhados é diretamente proporcional a quarta potência da frequência do laser ( $\nu$ ). A forma mais simples de lidar com esta fluorescência é evitá-la diretamente, por exemplo, utilizando substratos que não espalham a radiação. Pode-se também selecionar um comprimento de onda na qual a fluorescência não é observada ou tem baixa influência. Técnicas matemáticas e algoritmos foram desenvolvidos na tentativa de remover a fluorescência, contudo, nem sempre é facilmente retirada e acaba influenciando na informação final da amostra (KERR *et al.*, 2016).

O fraco sinal Raman obtido em amostras biológicas é fonte de constantes pesquisas que tentam intensificar este sinal. Em muitos casos ocorre a não detecção da amostra, uma vez que os espectros ficam sobrepostos pela alta fluorescência de determinados substratos e pela dispersão da

luz, causado pelo ‘Mie Scattering’ (KERR *et al.*, 2016). Este último é o espalhamento elástico da radiação que incide em partículas que tem diâmetro igual ao tamanho do comprimento de onda da radiação incidente. Logo, o Mie Scattering é o quadrado do diâmetro da partícula, com intensidade maior que o sinal Raman e o sinal Rayleigh, este que difere do Mie Scattering por ser o espalhamento elástico de partículas de diâmetro menores que o do comprimento de onda da radiação.

Tais problemáticas ocorrem em equipamentos portáteis, que tem maior limitação na obtenção dos espectros, geralmente acoplados com um único tipo de laser (um único comprimento de onda) e trabalha em uma faixa espectral limitada, mas também com os equipamentos de Microscopia Raman e equipamentos de bancada, estes que possuem maior capacidade de detecção, maior faixa espectral de análise, além de ser possível determinar o comprimento de onda da radiação pela escolha do laser. A presença do sinal do substrato pode comprometer a capacidade de extração da informação desejada para esta técnica (KOURKOUMELIS *et al.*, 2012). Para Kerr e colaboradores (2015), os substratos ideais devem produzir baixos sinais Raman, ser biocompatíveis, atóxicos para as células e tecidos analisados, de baixo custo e fácil acesso. Porém, estes substratos ideais além de caros, são disponibilizados em diferentes espessuras e purezas, produzindo diferentes espectros para um mesmo material.

Um dos trabalhos que aborda com detalhes os problemas da técnica Raman foi publicado por Kerr *et al.* (2015), ressaltando que para minimizar a intensidade do sinal do substrato para análises de amostras biológicas, deve-se utilizar um laser com comprimento de onda apropriado para o substrato de interesse. A intenção é diminuir a intensidade do sinal do substrato e aumentar a intensidade do sinal da amostra com a redução do espalhamento Mie (Mie Scattering). Kerr e colaboradores (2015) buscaram otimizar o sinal Raman analisando o impacto que alguns comprimentos de onda do laser (473 nm, 532 nm, 660 nm, 785 nm e 830 nm) têm nas amostras biológicas em determinados substratos (fluoreto de cálcio, fluoreto de magnésio, filme de alumínio de espessura de 100 nm no vidro, filme de alumínio de espessura de 1500 nm no vidro, vidro, sílica fundida, bromito de potássio, cloreto de sódio, seleneto de zinco). Com isso o objetivo era determinar uma correlação ótima entre a escolha do comprimento de onda e do substrato utilizado e assim obter um bom sinal da amostra biológica (células da bochecha humana). Os autores fizeram um comparativo da influência do comprimento de onda entre 830 nm e 473 nm, mostrando que um laser de 473 nm produz linhas Raman que são 9,5 vezes mais intensas que as produzidas para 830

nm, devido a isso os espectros a 830 nm apresentam muito ruído e poucos picos definidos na comparação com 473 nm. Logo, a eficiência do espalhamento Raman é maior para menores comprimentos de onda, baixos tempos de integração e menor potência do laser.

Com relação aos substratos, ainda no trabalho de Kerr e colaboradores (2015), os pesquisadores concluíram que os filmes de alumínio depositados no vidro em ambas as espessuras de 100 nm e 1500 nm obtiveram espectros mais consistentes, além de serem baratos, biocompatíveis, e apresentarem espectro de baixa fluorescência. Nos comprimentos de onda 785 nm e 830 nm, que recaem na região do infravermelho próximo, o filme de alumínio depositado no vidro foi o que apresentou melhor resultado, uma vez que este recobre o vidro fazendo com que apenas o alumínio seja irradiado pelo laser. Isso acontece pelo fato das superfícies metálicas reflectivas aumentarem em até quatro vezes o número de fótons espalhados pelo Raman, no comparativo com substratos transparentes como o vidro puro, intensificando o sinal e obtendo melhor sinal da amostra. Porém, para estes mesmos comprimentos de onda, o vidro, o fluoreto de cálcio e a sílica fundida, apresentaram uma alta intensidade do substrato, fazendo que o sinal da amostra biológica praticamente não seja detectado.

Outra forma de solucionar o problema da grande intensidade do sinal do substrato é usar algoritmos de subtração do sinal de fundo. Kerr e Henelly (2016) analisaram a eficiência de alguns algoritmos para subtrair o sinal Raman (laser de 532 nm) do vidro utilizado como substrato para amostras biológicas. A lâmina de vidro é bastante comum para deposição de amostras para análises clínicas por ser de baixo custo e poder ser utilizado em microscópios. Alguns pesquisadores evitam o uso do vidro nas análises clínicas com espectro Raman devido a sua alta fluorescência (principalmente na região de  $1050-1150\text{ cm}^{-1}$ ), que atenua os sinais Raman das amostras biológicas que possuem características de serem de fraca intensidade (KERR *et al.*, 2015). Logo, buscam utilizar substratos como o fluoreto de cálcio ( $\text{CaF}_2$ ) ou quartzo, mas estes materiais não são viáveis para análises em larga escala devido ao alto custo. Vários algoritmos têm sido desenvolvidos para minimizar o efeito da fluorescência do substrato na amostra, com correção de linha de base e ajuste polinomial. No trabalho citado, foram comparados dois algoritmos que removem a influência do vidro, ambos por subtração baseada num método polinomial iterativo de aproximação. Contudo, mesmo após a aplicação destes algoritmos, a influência do substrato ainda era fator dominante e podia ser observada por meio da análise exploratória de dados. Também foi testado a Correção de Sinal Multivariado Extendido (*Extended Multivariate Signal Correction* - EMSC), este que se

mostrou eficiente na remoção da informação do vidro. O EMSC atua simulando o espectro de referência, podendo subtrair a informação deste dos espectros das amostras. Porém é indicado para experimentos que não tenham como obter espectros de referência do substrato, pois a simulação do espectro é passível de erros (KERR & HENELLY, 2016).

Outros estudos sugerem o uso de substratos alternativos para fazer a análise Raman, evitando assim a fluorescência e a baixa intensidade. Cui *et al.* (2016) também sugeriram usar folha de alumínio como potencial substrato para análises de absorção de infravermelho e espectroscopia Raman, uma vez que substratos que não interagem com a radiação são de alto custo. A folha de alumínio, além de acessível, robusta e de baixo custo, possui inexpressiva intensidade de espalhamento, não interferindo no espectro da amostra de forma significativa, fazendo com que as informações importantes sejam preservadas, mesmo para amostras com espessura da ordem de 2  $\mu\text{m}$ . Como a folha de alumínio é composta basicamente do metal puro, considerando que para obter espalhamento Raman é necessário que ocorra uma mudança de polarização durante a vibração molecular, no metal puro isto não acontece. Para avaliar a eficiência do alumínio como substrato, Cui *et al.* (2016) compararam o uso da folha de alumínio frente a outros substratos, utilizando um Microscópio Raman com laser de 785 nm. Foi observado que o vidro comum e o vidro tipo Low-E apresentaram uma larga e forte banda de fluorescência a  $1382\text{ cm}^{-1}$ , que se sobrepõe a diversos picos da região do *biofingerprint* para amostras biológicas, causando uma grande influência nas análises, inclusive se tornando impossível a obtenção de espectros de amostras biológicas por subtração da banda de vidro. Em contraste, a folha de alumínio e lâmina revestida a ouro apresentaram uma inexpressiva intensidade. Após análises com material biológico depositado nestes diversos substratos, Cui e colaboradores (2016) concluíram que na folha de alumínio o espectro obtido teve alta qualidade, baixa fluorescência, permitindo seu uso em análises com materiais biológicos.

A espectroscopia Raman tem sido muito estudada nos últimos anos, incluindo a área forense devido à flexibilidade que a técnica tem para analisar diversas amostras, identificando drogas (HODGES *et al.*, 1990), fibras (THOMAS *et al.*, 2005), tintas (SUZUKI *et al.*, 2001), fluidos corporais (MURO *et al.*, 2016), entre outros. Virkler e Lednev em 2008, utilizaram a técnica Raman para identificar os diferentes fluidos corporais como saliva, sêmen, sangue, suor e sangue vaginal em locais de crime, de forma não destrutiva (VIRKLER & LEDNEV, 2008).

Apesar de utilizarem equipamentos e fundamentos diferentes, nas quais o Raman utiliza informação da radiação espalhada e o infravermelho a quantidade de radiação absorvida, os espectros trazem informações sobre transições vibracionais que muitas vezes são complementares. Como por exemplo, o fato da espectroscopia Raman sofrer pouca influência da água nas análises, já que há um fraco espalhamento por parte da ligação H-O-H, diferentemente da espectroscopia por absorção no infravermelho, uma vez que há um forte momento dipolo da molécula de água (CIALLA-MAY, 2019). Por outro lado, o infravermelho não tem tanta influência da fluorescência indesejada na obtenção do espectro comparado com o Raman (PEREIRA *et al.*, 2016).

Análise espectroscópica gera resultados com muitas variáveis de várias amostras, cuja observação direta de tais dados promove conclusões limitadas, sem prever tendências, padrões, erros associados, entre outras informações que são importantes para avaliar as amostras. Logo, é necessário utilizar ferramentas de análise multivariada que permitam obter dos dados informações relevantes.

## 2.4 QUIMIOMETRIA

As técnicas quimiométricas representam uma abordagem estatística e matemática para problemas químicos que tratam com grande quantidade de dados, como os produzidos por técnicas instrumentais incluindo as técnicas espectroscópicas (DEFERNEZ *et al.*, 1997). A Quimiometria é útil na interpretação das respostas analíticas, uma vez que essas são muitas vezes complexas e difíceis de interpretar visualmente ou de forma univariada. Utilizando tais ferramentas é possível verificar correlações estatísticas presente nos dados obtidos para uma dada questão química (KUMAR *et al.*, 2014), analisando as informações primárias (dados espectrais, cromatográficos, etc.) e as transformando em informações secundárias (modelos de classificação, modelos de calibração) permitindo interpretar os dados de forma mais abrangente a partir do modelo criado (GASTEIGER & ENGEL, 2003).

A análise multivariada contempla: a análise exploratória de dados, denominada técnicas não supervisionadas que permitem identificar padrões, agrupamentos e tendências nos dados; bem como técnicas de calibração, que são técnicas supervisionadas, onde se correlaciona uma matriz de dados ( $\mathbf{X}$ ) contendo espectros, por exemplo, com um vetor de características ( $\mathbf{y}$ ) que apresenta correlação química com a matriz original (BRERETON, 2003) obtendo-se modelos de calibração.

Porém, antes de aplicar essas técnicas, é importante remover as informações indesejadas e otimizar a análise estatística através do pré-processamento dos dados (BEEBE *et al.*, 1998).

### 2.4.1 Pré-processamento

As técnicas de pré-processamento são geralmente utilizadas devido à presença de informações indesejadas obtidas junto com as informações do analito, tais como ruído instrumental, efeitos de espalhamento (*scattering*), fluorescência, diferenças de intensidades entre espectros, efeito aditivo, entre outros (GAUTAM *et al.*, 2015). A escolha do pré-processamento é dependente do tipo de informação obtida com as análises e com a finalidade desejada (BRERETON, 2003). Na maioria das vezes, a escolha do pré-processamento adequado é realizada por tentativa e erro, avaliando se após o tratamento de dados a informação indesejada foi removida. Porém, conhecer a composição dos dados e o comportamento dos espectros facilita a escolha das ferramentas a serem utilizadas, que podem ser aplicadas às amostras e às variáveis.

Nas seções seguintes, alguns pré-processamentos de maior relevância para os espectros NIR e Raman foram conceituados. Assim, será possível entender como os espectros terão efeitos indesejados removidos ou atenuados.

#### 2.4.1.1 Centragem na Média

Consiste em um pré-processamento aplicado nas variáveis, que subtrai os valores individuais de uma coluna ( $X_{ij}$ ) pelos valores de sua média ( $\bar{X}_j$ ), sendo aplicado a todos os valores individualmente obtendo os valores centralizados ( $X_{ij}^{cen}$ ) de acordo com a Equação 1.

$$X_{ij}^{cen} = X_{ij} - \bar{X}_j \quad (1)$$

Centragem na média (*Mean Centering*) faz com que os dados tenham suas coordenadas mudadas para o centro do conjunto de dados, permitindo perceber diferenças nas intensidades relativas das variáveis facilitando sua interpretação (BRERETON, 2003). Sua aplicação faz com que as variações das variáveis partam todas da origem (nível zero), com isso pode-se melhorar o ajuste do modelo criado a partir destes dados e assim evitar erros numéricos, uma vez que a

arbitrariedade dos dados brutos faz com que haja diversos pontos de partida dos dados (BRO & SMILDE, 2003).

#### 2.4.1.2 Normalização

Devido à variação de intensidades que ocorrem em análises espectroscópicas ocasionada pela variabilidade na radiação durante análises, além de outros efeitos físicos, um conjunto de espectros para uma mesma amostra pode apresentar diferenças significativas de intensidades entre si. Este pré-processamento tem por objetivo remover ou minimizar variações sistemáticas das amostras (BEEBE *et al.*, 1998).

A normalização pode ser obtida dividindo-se cada variável por uma constante, assim as variáveis ficam numa mesma escala. Esta constante pode ser a soma dos valores absolutos de todas as variáveis para uma determinada amostra, sendo denominada de normalização pela área total. Também pode ser aplicada fazendo uma transformação linear, onde todos os valores referentes a resposta Y fiquem no intervalo de mínimo 0 e máximo 1, chamada de normalização pelo mínimo e máximo, sendo a forma mais simples de normalizar (GEMPERLINE, 2006). Como também pode ser normalizada pelo comprimento unitário, na qual divide-se cada elemento pela constante C, definida na Equação 2.

$$c = \sqrt{\sum_{j=1}^{max} x_j^2} \quad (2)$$

Na qual, j são as variáveis do vetor dos dados (x). A escolha de qual constante vai ser utilizada para normalizar depende de qual variação deseja-se remover (BEEBE *et al.*, 1998).

#### 2.4.1.3 Suavização (Filtros Savitzky Golay)

A suavização atua nas amostras ao longo das variáveis, corrigindo ruído espectral. A suavização pelo algoritmo proposto por Savitzky e Golay (SG) funciona ajustando-se um polinômio numa região simétrica ao redor do ponto de interesse. Em seguida substitui o valor deste ponto pelo valor previsto dado pela função ajustada. O tamanho da janela deste polinômio deve ser

escolhido de forma a atenuar o ruído, sem encobrir informações importantes. Aplicando-se o filtro sem optar por derivar o polinômio, tem-se apenas a suavização dos dados. Com menos ruído tem-se respostas analíticas com menor quantidade de influência física e que podem gerar modelos com melhor ajuste (RINNAN *et al.*, 2009).

#### 2.4.1.4 Derivadas

As derivadas conseguem remover os efeitos aditivos (1ª derivada) e multiplicativos (2ª derivada) nos espectros. Há alguns algoritmos aptos a aplicar as derivadas nos espectros, são as derivadas por diferenças finitas. A primeira derivada aplica a diferença entre dois pontos, já a segunda derivada considera a mesma diferença, porém entre pontos de espectros que já foram derivados; além do Norris-Williams e Savtizky-Golay (RINNAN *et al.*, 2009).

O método derivativo por filtros Savtizky-Golay é o mais utilizado, uma vez que além de derivar, o algoritmo também aplica uma suavização nos espectros, melhorando a razão sinal/ruído. Seu cálculo é realizado derivando um polinômio ajustado em torno de uma janela simétrica para um ponto no espectro. Essa operação é aplicada ponto a ponto ao longo do espectro. Tanto o número de pontos da janela quanto o grau do polinômio devem ser avaliados, pois janelas muito grandes removem ruído, mas podem remover picos informativos e causar distorção dos picos remanescentes (RINNAN *et al.*, 2009).

#### 2.4.1.5 Correção de Espalhamento Multiplicativa e Variação Normal Padrão

Tanto Correção de Espalhamento Multiplicativa (*Multiplicative Scatter Correction*-MSC) quanto Variação Normal Padrão (*Standard Normal Variate*-SNV) destinam-se a corrigir espalhamento da radiação apresentando resultados semelhantes apesar de operarem de forma diferente (GEMPERLINE, 2006).

MSC tenta compensar efeitos da dispersão da radiação aplicando uma regressão linear simples para cada espectro ( $\mathbf{x}_i$ ) contra um espectro de referência ( $\mathbf{x}_r$ ), que geralmente é o espectro médio do conjunto de dados, como pode ser visto na Equação 3.

$$\mathbf{x}_r \approx \beta_0 + \beta_1 \mathbf{x}_i \quad (3)$$

Onde,  $\beta_0$  e  $\beta_1$  são os coeficientes da regressão obtidos por mínimos quadrados, estimados previamente e então aplicados para obter os espectros corrigidos por MSC ( $X_i^*$ ), como pode ser visto na Equação 4.

$$X_i^* \approx \beta_0 + \beta_1 X_i \dots \dots \dots (4)$$

Como exemplo, tem-se a aplicação do MSC realizado nos espectros de sangue detectados pelo Infravermelho Médio, no trabalho de Lin *et al.*, em 2017 com a intenção de reduzir os efeitos de dispersão da luz.

Já o SNV atua centralizando na média o conjunto de dados, ou seja, os valores de cada espectro ( $x_i$ ) são subtraídos pelo valor médio ( $\bar{x}_i$ ), em seguida são divididos pelo desvio padrão dos valores de seu respectivo espectro. Após aplicar o SNV os dados ficam centrados no eixo zero, uma vez que ocorre a subtração pela média (GEMPERLINE, 2006).

No trabalho de Sun *et al.*, em 2017, por exemplo, para analisar manchas de sangue por meio de espectro visível foi relatado dificuldade em utilizar o MSC devido ao fato de ser necessário selecionar um espectro de referência. Assim, aplicaram o SNV para remover os efeitos de dispersão.

#### 2.4.1.6 Suavização por Mínimos Quadrados Assimétricos

A Suavização por mínimos quadrados assimétricos (*Asymmetric Least Square Smoothing-AsLS*) é um corretor de linha de base, este que tem a função de colocar a linha de base dos espectros partindo de um mesmo ponto, geralmente no ponto zero do eixo (PENG *et al.*, 2010). O AsLS é muito utilizado nos espectros Raman, sendo aplicado inicialmente para suavizá-los. Em seguida parte desta suavização para estimar a correção de linha de base, utilizando a tendência assimétrica dos pesos dos desvios da suavização aplicada.

Considerando um vetor ( $\mathbf{y}$ ) com amostras de igual intervalo, com uma série suavizada e fiel ao vetor  $\mathbf{y}$ , chamado de  $\mathbf{z}$ , com  $\lambda$  para balancear os dois termos, tem-se as Equações 5 e 6.

$$S = \sum_i (\mathbf{y}_i - \mathbf{z}_i)^2 + \lambda \sum_i (\Delta^2 \mathbf{z}_i)^2 \quad (5)$$

Sendo:

$$\Delta^2 \mathbf{z}_i = (\mathbf{z}_i - \mathbf{z}_{i-1}) - (\mathbf{z}_{i-1} - \mathbf{z}_{i-2}) = \mathbf{z}_i - 2\mathbf{z}_{i-1} + \mathbf{z}_{i-2} \quad (6)$$

A aplicação do AsLS faz diminuir a quantidade de ruídos, cria picos definidos enfatizando o sinal Raman, assim facilitando a identificação da interação da radiação com a molécula (EILERS & BOELEN, 2005).

## 2.4.2 Análise Exploratória e Calibração Multivariada

Os tópicos seguintes introduzem e conceituam a Análise Exploratória de dados e a Calibração Multivariada.

### 2.4.2.1 Análise Exploratória: Análise por Componentes Principais

A Análise por Componentes Principais (*Principal Componente Analysis* - PCA) é a técnica mais utilizada de análise exploratória. A técnica permite redução da dimensionalidade dos dados (geralmente em duas ou três) representando-os em um novo sistema de eixos, chamados de componentes principais (PC). O novo sistema de eixos, representados pelas PC's, são obtidos a partir da direção de máxima variância dos dados e permitem identificar similaridades entre amostras, além de detectar amostras anômalas (*outliers*) (SANTANA, 2013). Na PCA, a matriz de dados original ( $\mathbf{X}$ ) é decomposta no produto de duas matrizes: Escores ( $\mathbf{T}$ ), que são as coordenadas das amostras no novo sistema de eixos; e os Pesos ( $\mathbf{P}$ ), que são os cossenos dos ângulos formados entre as variáveis originais e as componentes principais. A PCA pode ser matematicamente representada pela Equação 7.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (7)$$

Na qual  $\mathbf{E}$  é a matriz residual e  $\mathbf{P}^T$  a transposta da matriz de Pesos (BRERETON, 2003).

Sendo a direção da PC's baseada na máxima variância, PC1 terá a máxima variância entre os dados e a PC2 que é ortogonal a PC1, tendo a segunda máxima variância, e assim sucessivamente.

#### 2.4.2.2 Calibração Multivariada: Regressão por Mínimos Quadrados Parciais

A regressão por mínimos quadrados parciais (*Partial Least Square Regression* - PLSR) é considerado o método de regressão mais importante, isso se deve à sua robustez frente a outros métodos. É útil para prever propriedades por meio dos modelos criados, para prever similaridades entre conjuntos de amostras diferentes, a partir do modelo obtido pelos dados de calibração (BRERETON, 2003).

O PLSR é um método de regressão multivariada que encontra uma relação linear entre a matriz de variáveis dependentes ( $\mathbf{X}$ ) e a matriz de variáveis independentes ( $\mathbf{Y}$ ), utilizando variáveis latentes (LV's), que são combinações lineares das variáveis originais, para que se tenha a máxima covariação entre  $\mathbf{X}$  e  $\mathbf{Y}$  durante a regressão (LIN *et al.*, 2017). A matriz resposta  $\mathbf{X}$  é decomposta em componentes de forma semelhante a PCA, gerando a matriz de escores ( $\mathbf{T}$ ) e a matriz de pesos ( $\mathbf{P}$ ) como pode ser visto na Equação 8, bem como para  $\mathbf{Y}$ , com escores ( $\mathbf{U}$ ) e pesos ( $\mathbf{Q}$ ), na Equação 9. As matrizes  $\mathbf{E}$  e  $\mathbf{F}$  são chamados de resíduos (GEMPERLINE, 2006).

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (8)$$

$$\mathbf{X} = \mathbf{UQ}^T + \mathbf{F} \quad (9)$$

Um cálculo iterativo faz com que  $\mathbf{T}$  e  $\mathbf{U}$  se ajustem de forma a maximizar a correlação entre a matriz dependente e a matriz independente. Como não há necessariamente ortogonalidade entre  $\mathbf{T}$  e  $\mathbf{U}$ , a correlação tenta aproximar a zero as matrizes residuais, como pode ser visto na Equação 10.

$$\mathbf{U} = \mathbf{TW} \quad (10)$$

Onde,  $\mathbf{W}$  é uma matriz de pesos.

Combinando algebricamente as equações 8 e 9 descritas acima, obtém-se uma matriz para se estimar os coeficientes de regressão do modelo ( $\hat{\mathbf{B}}$ ). Assim, para prever uma propriedade  $\mathbf{y}$  de uma amostra quando seu espectro  $\mathbf{x}$  é conhecido, utiliza-se a matriz ( $\hat{\mathbf{B}}$ ) pelas Equações 11 e 12.

$$\hat{\mathbf{B}} = \mathbf{P}(\mathbf{P}^t\mathbf{P})^{-1}\mathbf{W}\mathbf{L} \quad (11)$$

$$\hat{y} = X\hat{B} \quad (12)$$

Onde,  $L$  é a matriz peso.

É importante notar que, em geral, quanto maior for o número de variáveis latentes para obter um modelo, maior é o desvio da linearidade entre os dados utilizados para calibração. (SANTANA, 2013).

#### 2.4.2.3 Máquina de Vetores de Suporte

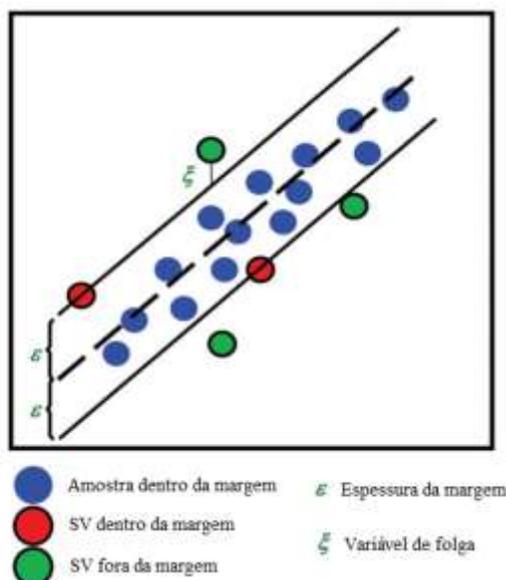
Máquina de vetores de suporte (*Support Vector Machine - SVM*) é uma ferramenta de aprendizado de máquinas supervisionada utilizada para previsão e classificação de propriedades de interesse. Esta ferramenta trabalha mapeando um conjunto de dados, geralmente não linear, para projetá-lo em um novo espaço de características multidimensionais (*kernelling*), onde o limite de decisão no espaço de entrada é representado por um hiperplano (HAMDI *et al.*, 2017).

O hiperplano é calculado por meio dos vetores de suporte (VS), que são os dados das margens de cada conjunto de dados (amostras), em torno de 2 ou 4 vetores. Assim que se define os vetores de suporte, nenhuma outra amostra tem influência na margem, sendo uma das vantagens do SVM frente a outros métodos de classificação, que usam todas as variáveis para determinar as fronteiras entre classes. Muitas vezes os dados não são tão fáceis de ter conjuntos de dados separados linearmente, logo o cálculo da função Kernel faz com que haja a análise dos VS's em mais de um plano, assim o hiperplano pode ser definido de forma a haver uma separação entre as classes. O Kernel pode ser uma função polinomial, gaussiana, sigmoidal, etc; a depender dos dados utilizados. Por dividir os conjuntos de dados, o SVM consegue determinar *outliers* e assim desconsiderá-los, por isso é um *machine learning*. Logo, se enquadra bem ao ser utilizado para classificação e calibração, podendo ser aplicado a dados químicos não lineares (BRERETON & LLOYOD, 2010).

Em termos de calibração, a regressão por SVM (SVMR) utiliza o parâmetro  $\epsilon$ , que define a largura da margem (definida pelos VS's), uma vez que o objetivo da calibração é incluir todas as amostras dentro desta margem. À medida que a linha da regressão se ajusta aos dados, a largura da margem diminui, conseqüentemente diminui  $\epsilon$ . Assim, este parâmetro pode ser considerado como fator de tolerância para a modelagem, limitando os erros e desvios do modelo. Contudo, em casos

práticos nem sempre todas as amostras ficam dentro das margens, muitas vezes ao forçar a inclusão destes vetores pode-se obter um modelo não representativo. Assim, SVMR permite que amostras fiquem de fora da margem, definindo-se a variável de folga  $\xi$ , que é a distância entre a amostra e a margem (Figura 3). Logo, o modelo desconsidera *outliers*, sendo esta outra vantagem do SVMR na comparação com outros métodos de calibração.

Figura 3 - Principais parâmetros definidos para o SVMR.



Fonte: Adaptado de Brereton & Lloyd (2010).

Assim como na classificação, para a calibração quando se tem amostras não lineares utiliza-se uma função Kernel para poder incluir o máximo de amostras possíveis dentro das margens, uma vez que aumenta a dimensão para ajustar linearmente os dados. Assim como um método de calibração pode-se determinar a qualidade do modelo SVMR com parâmetros de desempenho já conhecidos, como descrito a seguir (BRERETON & LLOYD, 2010).

### 2.4.3 Parâmetros para avaliar a qualidade do modelo

Diferentes parâmetros são utilizados para determinar a qualidade do modelo. No caso dos modelos de regressão multivariados, tem-se como um dos principais parâmetros as raízes dos erros quadráticos médios (*Root Mean Square Error* - RMSE), que podem ser de calibração (RMSEC);

de validação cruzada (*cross-validation* - RMSECV) e de predição (RMSEP) (SHARMA & KUMAR, 2018).

O RMSEC é o erro que representa o grau de concordância entre os valores de  $y$  estimados pelo modelo de calibração e os valores de  $y$  de referência usados para criar o modelo. O RMSEC fornece estimativas da capacidade que um modelo tem para prever dados corretamente e pode ser determinado pela Equação 13.

$$RMSEC = \left[ \frac{1}{n-m-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right]^{1/2} \quad (13)$$

Onde,  $n-m-1$  representa os graus de liberdade;  $y_i$  é o valor medido e  $\hat{y}_i$  é o valor predito.

O RMSECV é o erro obtido quando um modelo é avaliado por uma validação interna, por meio de uma validação cruzada (*cross validation* – CV). A CV atua separando um ou alguns elementos do conjunto de calibração. Calcula-se o modelo com os demais elementos, e em seguida efetua a previsão da propriedade de interesse dos elementos separados. Há várias formas de separar os elementos do conjunto de dados para a validação interna, como por exemplo a validação cruzada *leave-one-out* (LOOCV) que separa um elemento do conjunto de calibração por vez, procede a modelagem e a previsão até que todos os dados de calibração sejam utilizados para a validação.

Os erros de validação (RMSEP e RMSECV) representam de forma mais correta a representatividade do modelo. Para obter o RMSEP deve ser realizada uma validação externa, com dados obtidos de forma independente dos dados de calibração, como o nome já sugere são dados externos. Serve para determinar o quanto o modelo é abrangente, uma vez que se analisa o quanto os dados da validação externa se ajustam ao modelo. O número de amostras do conjunto de amostra de validação externa ou de predição ( $p$ ) deve ser grande o suficiente, de forma que o RMSEP leve em consideração todas as possíveis fontes de variabilidade possíveis e é calculado pela Equação 14 (GEMPERLINE, 2006).

$$RMSEP = \left[ \frac{1}{p} \sum_{i=1}^p (y_i - \hat{y}_i)^2 \right]^{1/2} \quad (14)$$

Outro parâmetro utilizado para avaliar a qualidade de um modelo é o coeficiente de determinação ( $R^2$ ). Seu valor é definido no intervalo de 0 a 1, quanto mais próximo de 1, mais

próximo de zero são os erros residuais. Contudo, quanto mais próximo de zero, indica que os coeficientes de regressão não conseguem prever  $y$  com precisão, de acordo com a Equação 15.

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (15)$$

Onde,  $\hat{y}_i$  é o valor predito;  $y_i$  é o valor medido;  $\bar{y}_i$  é a média do valor de referência dos valores de calibração e predição.

Outro parâmetro de desempenho que determina a tendência ou o viés (do inglês *bias*) do modelo é o valor de *bias*. Este que permite avaliar erros sistemáticos no conjunto de dados. Um valor de *bias* próximo de zero indica ausência de erros sistemáticos (NAES *et al.*, 2002).

## 2.5 ESPECTROSCOPIA E QUIMIOMETRIA NAS MANCHAS DE SANGUE

As técnicas espectroscópicas têm mostrado uma larga aplicação devido à grande abrangência de moléculas que interagem com a radiação, além de serem não destrutivas e permitirem a portabilidade de instrumentos. Assim, suas aplicações vêm sendo muito difundido para fins forenses, principalmente para análises de mancha de sangue. Dentre as técnicas destacam-se o uso da radiação no infravermelho próximo por absorção (NIR) e espalhamento (Raman) que aliados a um tratamento multivariado de dados tem-se meios para auxiliar na resolução de delitos.

Edelman *et al.*, em 2012, utilizaram a espectroscopia na região do infravermelho próximo com um equipamento de bancada, para identificar mancha de sangue dentre outras manchas similares (manchas de vinho, ketchup, batom e alvejante) e para determinar o tempo de deposição do sangue durante 77 dias. Tanto a identificação como a datação foram realizados em substratos coloridos de algodão (cinco cores diferentes), utilizando a ferramenta quimiométrica PLS para calibrar. No trabalho são descritas as principais bandas de absorção referentes aos componentes do sangue, conforme mostra a Tabela 1, para avaliar os espectros de sangue ao longo do tempo. Os espectros foram pré-processados removendo a influência do substrato por meio da subtração dos espectros de referência, que são os espectros apenas dos algodões coloridos. Ao obter os espectros foram observados picos relevantes em 970 nm, 1454 nm e 1940 nm para o sangue imediatamente depositado (0 minutos), indicando que a água estava presente em alta concentração. Após 10

minutos, foram observados picos de hemoglobina, albumina e globulina, porém nem todas as bandas de absorção relatadas na Tabela 1 foram observadas.

Tabela 1 – Modos vibracionais de absorção dos componentes do sangue no Infravermelho Próximo.

<b>Comprimento de onda (nm)</b>	<b>Componente</b>	<b>Modo Vibracional</b>
930	Oxi-hemoglobina	Terceiro sobretom do -CH e -CH <sub>2</sub> vibração de estiramento
970	Água	Combinação de H-O-H vibração de estiramento simétrico e assimétrico
1454	Água	Combinação de H-O-H vibração de estiramento simétrico e assimétrica
1690	Hemoglobina, Albumina, Globulina	Primeiro sobretom do -CH vibração de estiramento
1740	Hemoglobina, Albumina, Globulina	Primeiro sobretom da banda a 3477 nm
1940	Água	Combinação do H-O-H deformação angular e estiramento simétrico e assimétrico
2056	Hemoglobina, Albumina, Globulina	Combinação do amido A e amido II ou outra combinação
2170	Hemoglobina, Albumina, Globulina	Combinação do amido B e amido II ou sobretom do amido II
2290	Hemoglobina, Albumina, Globulina	-CH vibração de estiramento e combinação de deformação angular
2350	Hemoglobina, Albumina, Globulina	-CH vibração de estiramento e combinação de deformação angular

Fonte: Adaptado de Edelman *et al.* (2012)

Os resultados de Edelman e colaboradores em 2012 mostraram que esses substratos coloridos tiveram uma maior influência e significância na região do infravermelho próximo, porém a partir de 1150 nm, os espectros dos substratos se sobrepuseram (*overlapping*). Com isso, a região de 1150-2500 nm foi selecionada para classificar e datar as manchas de sangue, já que não tem influência do substrato de algodão. Assim, os pesquisadores conseguiram identificar sangue com 100% de sensibilidade e 100% de especificidade dentre outras manchas similares. Já na datação da

mancha eles observaram que a confiabilidade do modelo PLS decai ao longo dos dias e por isso criaram o modelo com dados de deposição até o 30º dia. Assim, obtiveram um valor de  $R^2$  médio de 0,9 e RMSEP médio de 9%, ou seja, um erro de aproximadamente 3 dias. Justificaram o alto erro devido à alta influência dos fatores externos como temperatura, umidade e luminosidade na reação bioquímica de degradação do sangue.

Em 2016, Edelman *et al.* aplicaram seus conhecimentos de datação de manchas de sangue por meio de técnicas espectroscópicas em situação real de crime. Este que é passível de muitas influências externas, como condições ambientais (variações de temperaturas, umidade, etc.), presença de diversos substratos onde o sangue pode ter se depositado, além de diferentes tamanhos e espessuras das manchas. Assim, para poder determinar o tempo de deposição da mancha os pesquisadores criaram equações para dados obtidos na região do visível, que tem dependência com a espessura da mancha, uma vez que influencia na resposta espectral devido a possibilidade da radiação atingir o substrato. As equações conseguiram determinar as frações de hemoglobina, meta-hemoglobina e hemicromo nas amostras. Com isso, os pesquisadores coletaram espectros de mancha de sangue utilizando equipamento portátil (com faixa espectral de 450-800 nm), proveniente de 40 doadores para criar uma base de dados que relaciona as frações de derivados da hemoglobina, metahemoglobina e hemicromo em função do tempo.

Na cena de crime, Edelman e colaboradores (2016) obtiveram espectros aplicando o mesmo equipamento portátil, analisando as manchas em substratos de algodões coloridos (branco, rosa, azul, verde e amarelo) proveniente de tecidos encontrados no local, no tempo de 2 horas a 19 dias. Os espectros foram normalizados aplicando o SNV e apenas consideraram analisar os derivados da hemoglobina, uma vez que a metahemoglobina e o hemicromo são muito dependentes da temperatura e umidade podendo causar erro na determinação das suas frações ao longo do tempo. Com um *threshold* de  $R^2=0,9$ , apenas as manchas de sangue analisadas em algodão branco, rosa e amarelo coincidiram com o sangue de referência em algodão branco. Assim, apenas estes tiveram a fração de hemoglobina determinada com o tempo, evitando-se erros devido a influência do substrato na datação da mancha. No local do crime os tecidos estavam dispostos de forma aleatória no térreo e no primeiro andar da casa, e foram comparadas as frações de hemoglobina nestes substratos proveniente de sangue de diferentes pessoas (triplo homicídio). À medida que a fração de hemoglobina se apresentava em menor ou maior quantidade, os pesquisadores conseguiam

determinar se uma mancha era mais antiga que a outra. Aplicando a base de dados criada por eles, os pesquisadores determinaram com 95% de confiança o tempo de deposição daquelas manchas.

Em 2017, Lin *et al.* realizaram um estudo da datação da mancha de sangue para condições internas e externas, simulando condições reais de crime, nos tempos de 0,25 a 107 dias, utilizando a espectroscopia na região do infravermelho médio. Os espectros foram obtidos por reflexão total atenuada (*attenuated total reflection-fourier transform infrared* - ATR-FTIR) e as técnicas quimiométricas PLS e análise discriminante por PLS (PLS-DA) foram empregadas. Com 556 amostras para construir o modelo (dados de treinamento - 228 em ambiente interno e 228 em ambiente externo) e 228 amostras para validar o modelo (dados de validação - 114 para ambiente interno e 114 para ambiente externo), os sangues dos doadores (dois masculinos e dois femininos) foram depositados imediatamente em lâminas de vidro após a coleta. Contudo, para analisar a mancha formada, a mesma foi coletada em tubo Eppendorf e misturada com uma solução salina. Um microlitro da mistura foi então depositado no cristal ATR na faixa espectral de  $900-1800\text{ cm}^{-1}$  (*biofingerprint region*).

Considerando os espectros obtidos no trabalho de Lin *et al.* (2017), o *background* foi automaticamente removido, além de ter sido aplicado como pré-processamento a correção de linha de base, a normalização e MSC, para reduzir os efeitos físicos nos espectros. Os autores observaram que a estrutura secundária da hemoglobina muda constantemente com o passar dos dias, devido à reação cinética da autoxidação, desnaturação e degradação da hemoglobina (Hb->HbO<sub>2</sub>->met-Hb->Hc), provando que a degradação da mancha começa imediatamente na deposição e pode ser detectada desde os tempos iniciais até longos períodos de tempo. Porém, perceberam que é impraticável a datação da mancha de sangue analisando isoladamente as intensidades de um ou vários picos variando com o tempo, devido a sobreposição dos espectros da mancha, por isso foram utilizadas as técnicas PLSR e PLS-DA.

Ainda no trabalho de Lin *et al.* (2017) a análise quimiométrica foi realizada por meio da validação interna aplicando a validação cruzada (*cross-validation* - CV), além da validação externa. Foi observado que tanto para condições internas e externas, os tempos iniciais de 0,25 a 7 dias (sangue considerado fresco) apresentaram alta discrepância na regressão, tornando esta ainda maior no tempo de 0,25 a 1 dia. Além disso, ao se aproximar do 107º dia, os erros de previsão foram ainda maiores, ou seja, o modelo não se ajusta adequadamente até aquele determinado tempo, e assim o 107º dia foi removido da análise de regressão, ficando o 85º dia como o último dia do

modelo. Como um único modelo englobando todos os dias não se mostrou adequado, o conjunto de dados foi separado em dois conjuntos, de acordo com o tempo: 0,25-7 dias e 7-85 dias, obtendo-se um modelo melhor para 7-85 dias, principalmente em condições internas. Já para o intervalo de tempo de 0,25-7 dias, o modelo não produziu resultados satisfatórios para ambas as condições, por apresentar RMSEP de 1,1 dias e  $R^2=0,8$  para ambiente externo e RMSEP de 1,2 dias e  $R^2=0,7$  para ambiente interno.

Lin e colaboradores (2017) relatam que não houve ajuste adequado na validação externa devido à reação cinética da hemoglobina apresentar duas etapas, uma inicial de rápido decaimento (durando algumas horas) e outra final de lento decaimento (que perdura até o décimo dia). Além disso, a oxidação da hemoglobina é dependente da temperatura que varia fortemente neste experimento devido às variações naturais de temperatura ao longo do dia. Já a transição de meta-hemoglobina a hemicromo é fortemente dependente da umidade, ocasionando uma alta variação entre as amostras ao longo do tempo, tanto as mantidas em condições externas quanto as internas. Com isso houve uma grande influência na obtenção da reação cinética. O teste de expansão indicou que o modelo criado para a condição experimental não abrange outras condições de tempo e temperatura, devido a um baixo  $R^2$  e alto RMSEP. Logo, os pesquisadores concluíram que ocorreu uma alta influência do ambiente na cinética da mancha e com isso o modelo não apresenta um alto grau de linearidade nos tempos iniciais.

Partindo do conhecimento obtido, Lin et al., em 2017, utilizaram o PLS-DA para discriminar entre manchas frescas ( $\leq 24$  horas) e manchas antigas ( $> 24$  horas), já que o modelo PLS não se mostrou adequado devido à alta variabilidade dos dados para as manchas de 0,25 a 7 dias. O uso da ferramenta quimiométrica de classificação conseguiu diferenciar, com 99% de acurácia para condições internas e 98% acurácia para condições externas, entre as manchas recentes e antigas.

Já para o Raman, também se encontra trabalhos correlacionando a técnica com a análise de sangue para situações forenses. O sangue que tem os picos de absorção definidos em diferentes trabalhos, mas os picos são praticamente os mesmos. Um destes trabalhos é o de Virkler e Lednev (2008), que teve os picos do espectro Raman para o sangue definidos na Tabela 2.

Tabela 2 – Modos vibracionais de absorção dos componentes do sangue no Raman.

<b>Comprimento de onda (cm<sup>-1</sup>)</b>	<b>Componente</b>	<b>Modo Vibracional</b>
742	Triptofano	Vibração do anel
1001	Anel aromático da Fenilalanina	Vibrações de alongamento ( <i>breathing</i> )
1123	Glicose	Estiramento assimétrico do C-O-C
1247	Proteína	Vibração do amido III
1341	Triptofano	Deformação simétrica no plano do C-H
1368	Grupo Heme	Estiramento simétrico CH <sub>3</sub>
1446	Triptofano	Deformação simétrica no plano do CH <sub>2</sub> e do CH <sub>3</sub>
1576	Grupo Heme	Vibração C=C
1619	Grupo Heme	Vibração C=N

Fonte: Adaptado de Virkler & Lednev (2008);

Fujihara e colaboradores fizeram trabalhos interessantes aplicando a espectroscopia Raman para análises forenses de sangue. Em 2017, Fujihara *et al.* utilizaram um equipamento Raman Portátil (com laser de 532 nm e faixa espectral de 100-2000 cm<sup>-1</sup>) para identificar sangue humano dentre outros sangues (animais diversos). A análise foi realizada em condições laboratoriais depositando as amostras em gaze para formar a mancha de sangue, sendo analisadas após a secagem de 48 horas a temperatura ambiente da sala (22°C). Empregando PCA, os pesquisadores conseguiram distinguir os diferentes sangues analisados, mesmo após 3 meses de deposição. Concluíram então que o equipamento pode ser utilizado em condições reais de perícia.

Um trabalho similar foi realizado por Doty e Lednev em 2018, onde aplicaram a espectroscopia Raman aliada a técnica quimiométrica de análise discriminante PLS-DA para identificar sangue humano dentre sangue de dezesseis espécies de animais, na intenção de auxiliar a justiça na resolução de crimes ambientais. Num total de 84 doadores humanos e 129 doadores animais, com idades, raças e gêneros diversificados, os espectros foram coletados por microscopia Raman (com faixa espectral de trabalho 300 - 1700 cm<sup>-1</sup>). Os espectros foram pré-processados por correção de linha de base com mínimos quadrados ponderados (*weighted least squares* -WLS), normalizados pela área total e centrados na média. Com 135 dados para treinamento (35 de

humanos e 100 de animais) o modelo PLS-DA foi criado com 6 variáveis latentes, validação interna por CV e validação externa com dados dos 78 doadores que não entraram na calibração. O modelo demonstrou 99% de acertos ao fazer a discriminação dos doadores, uma vez que dos 14 doadores humanos da validação externa, o modelo determinou todos como humanos e dos 29 doadores animais da validação externa o modelo determinou 28 destes como animais. O animal que foi determinado erroneamente como humano foi o chimpanzé, sua classificação apresentou erro devido à similaridade da hemoglobina com o dos humanos.

Doty e colaboradores, em 2016, conseguiram desenvolver um método utilizando um microscópio Raman (com laser de 785 nm e faixa espectral de 300-1800  $\text{cm}^{-1}$ ) capaz de discriminar manchas de sangue (obtidos de dois doadores em substrato de folha de alumínio) em escala temporal de horas (de 1 hora a 168 horas, ou seja, uma semana), de forma não destrutiva e sem aplicação de anticoagulantes. Os pesquisadores primeiramente observaram um aumento da linha de base ao longo do tempo devido ao aumento da concentração de metHb e Hc, que contribuem para uma alta fluorescência. Logo, tentaram determinar o tempo de deposição analisando a variação de intensidade ao longo do tempo em comprimento de onda específico (400,5  $\text{cm}^{-1}$ ), porém verificaram que o *background* é causado por diversos fatores, incluindo o tipo de substrato, não sendo possível determinar o tempo de deposição da mancha de sangue apenas analisando um comprimento de onda. Partindo para a análise estatística multivariada, PLSR com validação interna (CV) e externa (dados de outros doadores) foi utilizado para obter um modelo preditivo, resultando em valores de RMSEC de 1 hora e  $R^2=0,9$ . Na validação interna obteve-se RMSECV de 1,3 horas e para validação externa RMSEP de 2,2 horas.

Na conclusão dos resultados obtidos, Doty *et al.* (2016) consideraram que os resultados para validação externa não foram coerentes devido a um alto RMSEP, porém consideraram que o modelo se ajustou bem aos dados da validação externa, com  $R^2$  de 0,9. Com isso consideraram que modelo poderia ser aplicado a uma situação forense. Sugeriram fazer análises utilizando a mesma metodologia, porém considerando diferentes substratos e variações na temperatura.

Já em 2017, Doty e colaboradores fizeram outro estudo aplicando a técnica Raman para identificar e datar a mancha de sangue com tempo de deposição de 2 anos. Para isso utilizaram um microscópio Raman (laser de 785 nm com faixa espectral de 300-1800  $\text{cm}^{-1}$ ). O sangue a ser analisado foi obtido de dois doadores (masculino e feminino) depositado em folhas de alumínio, sob condições ambientais não controladas, com temperatura média em torno de 22°C e umidade

relativa de 50%. Os pesquisadores perceberam que mesmo após um longo tempo o espectro da mancha de sangue apresenta mudanças significativas. Conseguiram classificar a mancha como sendo de sangue humano (dentre manchas de saliva, sêmen, suor, entre outras) até 1 mês de deposição na calibração, os modelos PLS apresentaram baixos valores de RMSE's, com RMSECV de 1,5 horas com  $R^2$  de 0,9 e RMSEP de 1,9 horas com  $R^2$  de 0,9, considerando o tempo ao longo de 2 anos. Os pesquisadores concluíram que a espectroscopia Raman tem grande potencial na determinação do tempo de deposição da mancha de sangue, porém o conceito deve ser estendido considerando variação nas condições ambientais (temperatura e umidade).

Em 2018, Doty e Lednev por meio da espectroscopia Raman com comprimento de onda a 785 nm, operando com faixa espectral de  $250-1800\text{ cm}^{-1}$ , aliada à classificação por máquina de vetores de suporte (SVM-DA), tentaram discriminar o sangue baseado na idade cronológica do doador. Os doadores eram recém-nascidos (< 1 ano de vida), adolescentes (11-13 anos de vida) e adultos (43-68 anos de vida). A diferença entre os sangues destes indivíduos, a diferentes idades, pode ser observada na estrutura da hemoglobina, que possui características distintas para cada idade da vida humana. Para isso utilizaram 45 doadores, destes 10 recém-nascidos, 15 adolescentes e 20 adultos, com sangue depositado em alumínio e seco durante pernoite. Utilizaram um corretor de linha de base (*automatic weighted least squares* - AWLS) e a normalização pela área como pré-processamento. O modelo SVM-DA foi obtido com 1080 espectros de calibração de um total de 1350. A classificação apresentou bom ajuste dos dados na validação externa: todos os 57 espectros para validação externa dos recém-nascidos foram corretamente classificados; para os adolescentes, dos 90 espectros apenas 3 foram classificados erroneamente; já para os adultos, dos 118 espectros, 18 foram classificados de forma errada (2 como recém-nascidos e 16 como adolescentes). Logo, o modelo apresentou-se satisfatório, porém os pesquisadores recomendaram considerar mais variáveis, como a dieta, saúde, e hábitos dos doadores, para tornar o modelo mais robusto.

Já em 2019, Fujihara *et al.* utilizaram a técnica Raman, por meio de um microscópio Raman (com laser de 632,8 nm e faixa espectral de  $500-1800\text{ cm}^{-1}$ ) para discriminar entre sangue infantil e adulto. Os pesquisadores conseguiram determinar a histidina a  $1105\text{ cm}^{-1}$ , que só é encontrado em sangue adulto e assim puderam diferenciar entre os dois grupos em estudo

Alguns trabalhos utilizam análises espectrais aliadas às técnicas multivariadas para avaliar o comportamento cinético dos componentes do sangue. Como por exemplo, o trabalho de Bremmer *et al.* (2011), que mostra um estudo cinético utilizando espectros de absorção na região visível

focando no comportamento da hemoglobina. Foi analisada a formação de meta-hemoglobina e do hemicromo nos tempos de 1 hora, 1 dia, 7 dias e 63 dias. Ao longo do tempo os espectros mostraram diferentes comportamentos para Hb, metHb e Hc. Logo após a deposição, começa a ocorrer a oxidação e se formar a meta-hemoglobina e em seguida o hemicromo, fazendo decair a concentração de HbO<sub>2</sub>. Os autores observaram que a reação bioquímica é rápida, de acordo com as observações, a meta-hemoglobina atinge seu máximo até aproximadamente 15 dias e depois há o decaimento da sua concentração. Foi observada somente a formação de meta-hemoglobina e hemicromo, nenhum outro composto foi originado. Assim, essa reação bioquímica foi utilizada para determinar a idade da mancha de sangue. Foi determinado a curva cinética da mancha de sangue e em seguida construiu-se um gráfico comparando a estimativa da idade da mancha de sangue e a real idade, de forma a analisar o quanto a estimativa por meio do modelo proposto varia com a idade real da mancha. Concluíram que o modelo proposto foi passivo de muitos erros, chegando a uma margem de erros em dias, sendo necessário maiores investigações sobre os efeitos do ambiente na amostra, como temperatura, umidade, exposição a luz solar, entre outros.

Um outro estudo cinético para manchas de sangue depositadas em papel foi realizado por Li *et al.* (2013), que obtiveram imagens hiperespectrais (de 400 nm a 720 nm) por meio de equipamento portátil. Pelo fato da reação bioquímica de degradação do sangue ser dependente da variação de temperatura e umidade, a temperatura foi controlada a 30°C e umidade a 40%. As análises foram realizadas durante 30 dias e os espectros foram pré-processados aplicando o SNV. Analisando o comportamento do logaritmo natural da refletância pelo tempo em dias, pode-se observar uma curva exponencial de decaimento, porém com uma etapa lenta e outra rápida, sugerindo haver dois processos cinéticos distintos. O processo lento é relativo à oxidação da oxihemoglobina a meta-hemoglobina e o processo rápido é a degradação de meta-hemoglobina a hemicromo. Um modelo de classificação foi obtido por meio da Análise Discriminante Linear (*Linear Discriminant Analysis - LDA*) para datar a mancha de sangue dentre o período de 30 dias. O modelo LDA apresentou melhores resultados para até 7 dias, taxa de classificação correta (*corrected classification rate-CCR*) de 83,9% (uma média de erro de 1,2 dias). Porém, para os dias restantes o modelo teve decréscimo na confiabilidade, principalmente após o 15º dia, com CCR de 53,5% (uma média de erro de 1,6 dias). Os pesquisadores concluíram que a técnica, após alguns aperfeiçoamentos pode prover informações de grande interesse forense.

### 3 METODOLOGIA

A coleta das amostras, suas respectivas análises com os equipamentos portáteis, assim como os pré-processamentos, técnicas quimiométricas aplicadas nos dados, além dos *softwares* utilizados estão descritos nas seções seguintes.

#### 3.1 OBTENÇÃO DAS MANCHAS DE SANGUE

O sangue empregado para produzir as manchas foi obtido de quatro diferentes doadores do sexo masculino (denominados 1, 2, 3 e 4). Os voluntários se apresentavam com idade média de 28 anos ao tempo da coleta, mas com diferentes portes físicos e hábitos. Questões de variação sanguínea por variação de gênero, idade, raça, tipagem sanguínea não foram levadas em consideração, uma vez que o foco de estudo era o tempo de deposição da mancha do sangue. Logo, levar em ponderação cada fator presente no sangue humano acarretaria em análises de outras variáveis e acabaria fugindo do escopo do trabalho. Cada doador depositou duas gotas em cada placa e para cada gota foram obtidos espectros em triplicata. Logo, as manchas obtidas dos quatro doadores conseguiram ser representativas para o estudo.

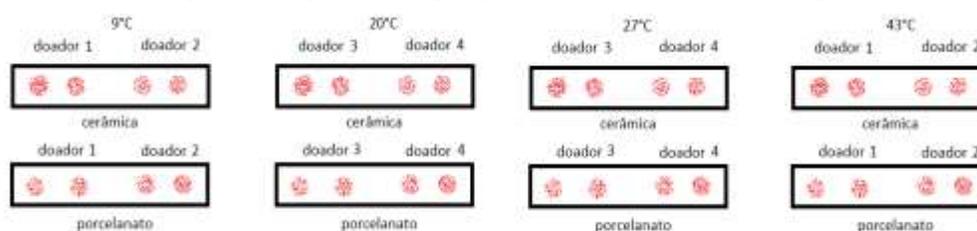
A amostra foi obtida utilizando agulha para punção dos dedos (a mesma utilizada para análise de glicose). Buscou-se manter um volume médio da amostra, de cerca de uma gota de sangue, gerando manchas de espessuras e tamanhos semelhantes. Devido à natureza das amostras, o projeto foi submetido para avaliação ao Comitê de Ética em Pesquisa envolvendo Seres Humanos da Universidade Federal de Pernambuco (CEP-UFPE), sendo aprovado e recebendo o parecer nº1.059.225. Todos os doadores voluntários para o estudo assinaram o Termo de Consentimento de Livre Esclarecimento (TCLE), este que está no Anexo A.

O material foi depositado em dois tipos de substratos, cerâmica e porcelanato, uma vez que estes são os tipos de pisos mais encontrados em locais de crimes em ambientes internos e externos. As amostras foram submetidas a diferentes temperaturas para simular condições ambientais nas quais um crime poderia ocorrer. As temperaturas utilizadas foram: 9°C (temperatura obtida por meio de uma geladeira), 20°C (utilizando uma sala isolada com ar condicionado), 27°C (utilizando uma sala isolada, sendo esta considerada a temperatura ambiente) e 43°C (por meio de uma estufa). As amostras a 9°C ficaram acondicionadas em um compartimento plástico, já para as demais

temperaturas as amostras ficaram expostas ao ambiente. A umidade não foi controlada e a exposição a luz artificial acontecia apenas nos momentos da análise das amostras, não sendo fator de fundamental influência na degradação, uma vez que o maior influenciador era de fato a temperatura que estava sendo avaliada.

O sangue dos doadores 1 e 2 foram depositadas nas placas de cerâmica e porcelanato nas temperaturas de 20 e 27°C e dos doadores 3 e 4 para as placas de cerâmica e porcelanato nas temperaturas de 9°C e 43°C (Figura 4). As manchas foram assim depositadas sem critério de definição.

Figura 4 - Esquema de deposição das manchas de sangue nas placas.



Fonte: O autor (2020)

Os espectros foram coletados em intervalos de tempo de folga de 10 minutos para cada tempo pontual, ou seja, a 0 minutos (imediatamente após a deposição) a análise foi realizada em um intervalo de 0 a 10 minutos. Assim, ficou definido: 10 minutos ficou sendo o período entre 10 a 20 minutos; e assim sucessivamente. Os outros intervalos de tempo foram 20 minutos, 1, 2, 10, 24, 48, 72, 96 horas, 5, 6, 7, 8, 9, 10, 12, 14, 15, 30 e 45 dias, para cada temperatura e substrato, totalizando 252 espectros para cada placa em cada temperatura. Tal forma de determinação dos tempos de coleta foram assim definidos por questões de análise, uma vez que foram utilizados dois equipamentos e os espectros foram coletados em triplicatas havendo dificuldades em analisar em tempos pontuais.

### 3.2 ANÁLISE COM INFRAVERMELHO E RAMAN PORTÁTEIS

Com os equipamentos portáteis MicroNir (para absorção no infravermelho próximo) e Mira DS (para espalhamento Raman no infravermelho próximo) foram obtidos os espectros do estudo.

O equipamento portátil MicroNir 1700, da empresa Viavi Solutions tem um filtro de variação linear (LVF), com detector de arseneto de gálio e índio (InGaAs). Com dimensões compactas de 45 mm de diâmetro e 42 mm de altura, pesando apenas 60 gramas, sendo assim atrativo para uso em campo. A faixa espectral varia de 908 a 1676 nm, com resolução espectral de 12,5 nm.

O equipamento Raman portátil Mira DS da empresa Metrohm tem faixa espectral de 400-2300  $\text{cm}^{-1}$ , pesando em torno de 705 gramas, com 45 mm de altura e 88 mm de largura na parte externa. O laser utilizado tem potência de 5 a 500 mW, sendo classificado como laser de classe 3B, com resolução espectral de 8 - 10  $\text{cm}^{-1}$  e comprimento de onda único de 785 nm. Os espectros foram obtidos por meio do software do próprio fabricante do equipamento Raman, chamado de Mira Cal DS Software.

Para definir uma condição ótima de análise das amostras de sangue por espectroscopia Raman, de forma rápida e com alta razão sinal/ruído, foram depositadas gotas de sangue em uma superfície feita de folhas de alumínio. Os metais foram definidos como intensificadores de sinal Raman, sendo as folhas de alumínio uma opção prática e de baixo custo de acordo com a literatura (KERR *et al.*, 2015; CUI *et al.*, 2016). Desta forma, chegou-se aos seguintes parâmetros de aquisição: tempo de integração de 0,8 segundos, que é o tempo que leva para o equipamento registrar um espectro; com média de duas aquisições, ou seja, a resposta de uma análise vai ser a média de dois espectros; e a potência do laser foi definida em 1 unidade, dentre uma escala de 1-5 unidades de potência, sendo a 5 a potência mais alta. Foi definido 1 unidade no intuito de não penetrar muito na amostra e não degradar por uma alta intensidade do laser, já tentando minimizar a fluorescência do substrato.

Para este o equipamento Raman portátil em questão é possível acoplar na abertura do laser diferentes acessórios, a depender da finalidade do uso do equipamento. Foi utilizado neste trabalho o acessório universal, que é empregado para diversos tipos de análise, incluindo análise direta nas amostras. Tal acessório tem 3 diferentes distâncias de amostragem: a posição 1 é utilizada para focar através de superfícies de vidro, como por exemplo em garrafas de vidro, com ponto focal de aproximadamente 5 mm do fim do acessório. A posição 2 é utilizada para plásticos finos, com ponto focal de aproximadamente 3 mm. A posição 3 é utilizada para contato direto com a amostra, com ponto focal < 1 mm, logo esta posição do acessório foi a utilizada para obter os espectros, uma

vez que as manchas estavam depositadas na superfície das placas (Manual do Mira Cal DS-Methrom).

### 3.2.1 Espectros Totais no IV

Buscando ampliar a quantidade de espectros de sangue obtidos pela técnica do NIR, para dar maior variabilidade (em termos de dias e doadores) aos modelos criados e realizar validação externa, dados de outro estudo foram utilizados. Tais espectros foram denominados de espectros adicionados. Estes foram obtidos do estudo de Pereira *et al.* (2017) utilizando as mesmas placas de piso (cerâmica e porcelanato), empregando dois doadores masculinos diferentes H1 e H2, além de duas doadoras femininas M3 e M4. As análises foram realizadas a temperatura ambiente (27°C) sob os tempos de deposição de 2 horas, 24 horas, 5 dias, 8 dias, 18 dias, 28 dias e 49 dias. Assim, pôde-se incrementar o número de doadores para classes de tempos que já haviam sido determinadas nesse estudo, como de 2 e 24 horas, 5 e 8 dias. Como também adicionar mais dias ao conjunto de dados, como é o caso de 18, 28 e 49 dias. Observando que os dados obtidos para este estudo foram considerando o tempo final de 45 dias, mas os espectros adicionados foram analisados até 49 dias após a deposição.

## 3.3 TRATAMENTO DE DADOS

Os dados foram organizados em matrizes para melhor aplicação das ferramentas estatísticas computacionais. A matriz de dados (**X**) com 252 variáveis de intensidades para o Raman e de 252 absorbâncias para o infravermelho próximo (NIR). As variáveis foram organizadas em classes de tempo (em horas) para facilitar o seu entendimento nos resultados quimiométricos. Os pré-processamentos e demais análises quimiométricas (PCA, PLS, SVM) foram realizadas em rotinas do MATLAB® R2010a 7.10.0.49 9, Mathworks e o PLS ToolBox da Eigvector.

Os pré-processamentos foram realizados para minimizar os efeitos indesejados, já citados anteriormente, uma vez que apenas dados relacionados à informação química da amostra é de interesse. Nos espectros NIR e Raman foram avaliados os seguintes pré-processamentos: remoção do substrato por subtração direta do espectro de referência do substrato, suavização (filtros Savitzky Golay) com polinômio de 2° ordem e janela de 11 pontos, derivadas de 1° grau (com

janela de 11 pontos; e polinômio de 2° ordem), SNV, normalização, centralização na média. Mais especificamente nos espectros Raman, onde se observou mais informação indesejada, foi aplicado também o corretor de linha de base, o AsLS.

Após a etapa de pré-processamento, a análise exploratória empregando PCA foi realizada com o intuito de obter informações acerca da similaridade entre as amostras e as respectivas variáveis responsáveis por isso. Para a regressão pelo método dos mínimos quadrados parciais (PLS), empregado para construir o modelo de regressão para estimar o tempo de deposição das manchas de sangue, foi obtido por meio da projeção em uma reta linear do  $Y_{\text{predito}}$  (tempo em horas) com o  $Y_{\text{medido}}$  (tempo em horas). No PLS o  $Y$  predito é obtido por meio da validação cruzada (CV), por isso é dito  $Y_{\text{CV Predito}}$ . A determinação do número de variáveis latentes e os parâmetros da validação interna foram obtidos por meio da validação cruzada do tipo *Venetian blind*, com 5 amostras por *blind*. A validação cruzada também determinou o número de variáveis latentes (VL). O desempenho do modelo foi avaliado por meio dos parâmetros: RMSEC, RMSECV, coeficiente de correlação para calibração e o *bias*. *Outliers* foram removidos sem critérios matemáticos de determinação, apenas por observações.

Os modelos de regressão por SVM foram construídos utilizando valores de calibração. Para este foi utilizado uma parte de 70% ou 50% dos dados totais, separados de forma aleatória. O restante 30% ou 50% foram utilizados como dados de validação externa. Com estes últimos obteve-se a análise linear com os dados do  $Y$  medido (calibração) obtendo-se parâmetros de RMSEC, RMSECV, RMSEP, o coeficiente de correlação para validação e o *bias*, para poder avaliar o ajuste do modelo.

### 3.3.1 Valores esperados dos parâmetros de validação dos modelos de calibração

Para valores de  $R^2$  próximos a 1 e do *bias* próximo a 0, assim como baixos valores de RMSE's indicam que o modelo de regressão possui um bom ajuste linear para os dados. Baseado em estudos anteriores da literatura, como de Edelman *et al.* (2012), as regressões para datação da mancha de sangue apresentaram valores de  $R^2$  maiores que igual a 0,8, e RMSEP menores ou iguais a 3 dias. Desta forma, tais valores serão considerados satisfatórios para o estudo em questão.

## 4 RESULTADOS E DISCUSSÃO

Nas seções seguintes serão apresentados os resultados obtidos utilizando a metodologia empregada. Além de discutir o comportamento dos dados brutos e mediante aplicação de pré-processamentos, assim como das técnicas quimiométricas. Para então compará-los com estudos correlatos.

### 4.1 ESPECTROSCOPIA NIR

Primeiramente foram discutidos os espectros obtidos com a técnica do NIR.

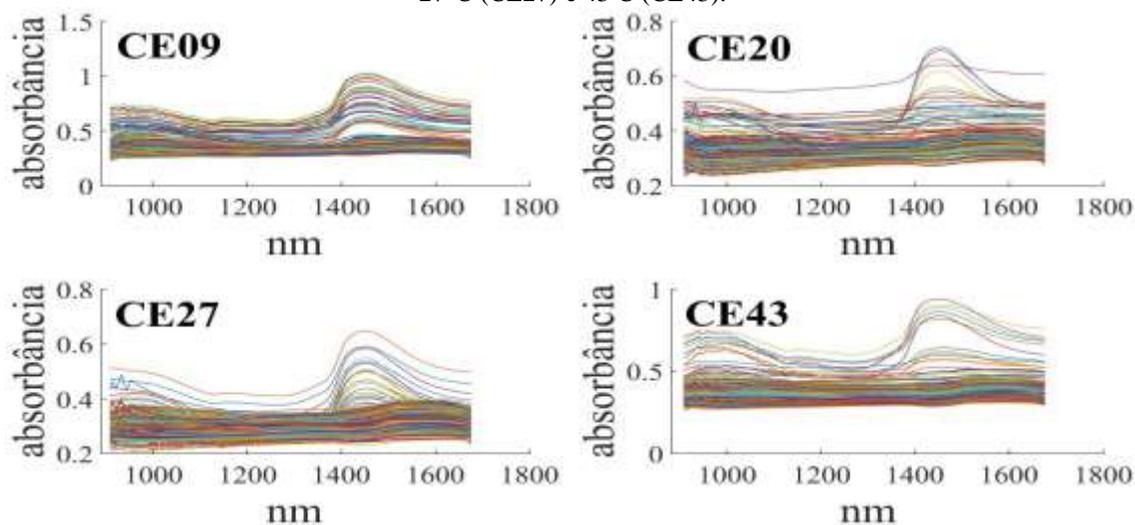
#### 4.1.1 Espectros NIR no substrato de cerâmica

Os resultados obtidos nas placas de cerâmica foram discutidos inicialmente.

##### 4.1.1.1 Análise Exploratória

Os espectros das manchas de sangue depositadas em cerâmica apresentaram um comportamento semelhante para as diferentes temperaturas (Figura 5).

Figura 5 - Espectros brutos NIR do sangue na placa de cerâmica nas temperaturas de 9°C (CE09), 20°C (CE20), 27°C (CE27) e 43°C (CE43).

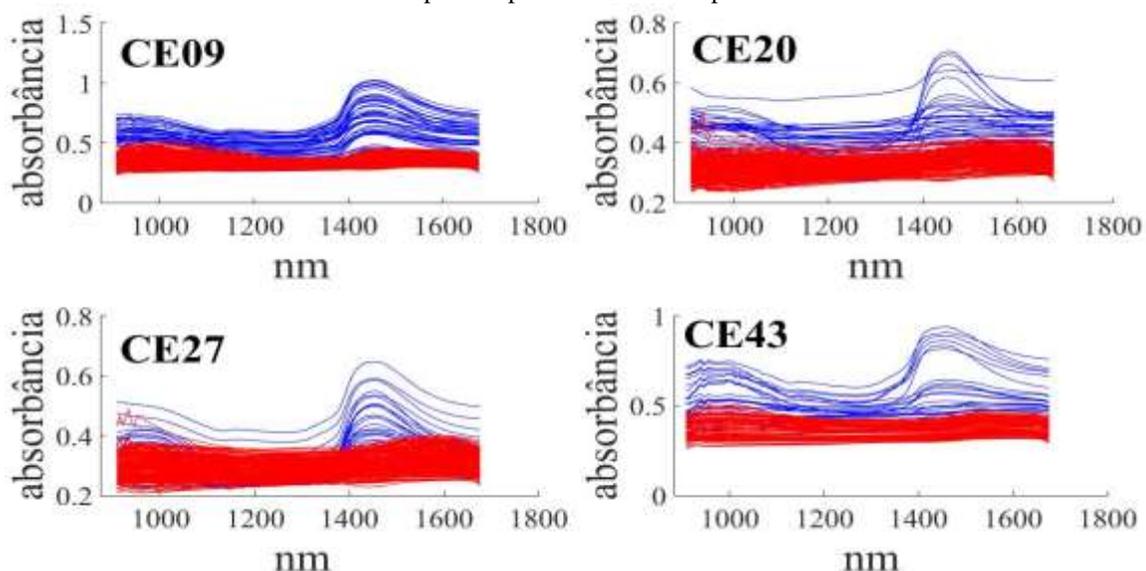


Fonte: O autor (2020)

Da figura anterior observa-se que o efeito de espalhamento estava presente nos espectros, além de ruído elevado, mais especificamente na faixa de comprimentos de onda entre 908 nm a 1100 nm. Isso pode ser devido à proximidade desta região com a região espectral do visível (300-800 nm). Os espectros para os tempos iniciais apresentaram duas bandas bem definidas nas regiões de 908-1100 nm e de 1350-1650 nm. Isso é referente às bandas de absorção da molécula de água em 970 nm e 1454 nm que predominam nesses espectros (entre 0 minutos e 10 horas após deposição do sangue) devido à grande quantidade de plasma, a depender da temperatura.

No intuito de melhor avaliar até que classe de tempo as bandas de informação da molécula de água estava presente nos dados brutos, uma separação em cores (azul para os espectros com bandas de água presente; vermelho para os espectros sem a informação da água) foi realizada, como pode ser visto na Figura 6.

Figura 6 - Espectros brutos das manchas de sangue sobre a cerâmica para diferentes temperaturas: em azul os espectros nos tempos iniciais (10 horas para 9°C; 2 horas para 20 e 27°C; 20 minutos para 43°C); em vermelho os espectros para os demais tempos.



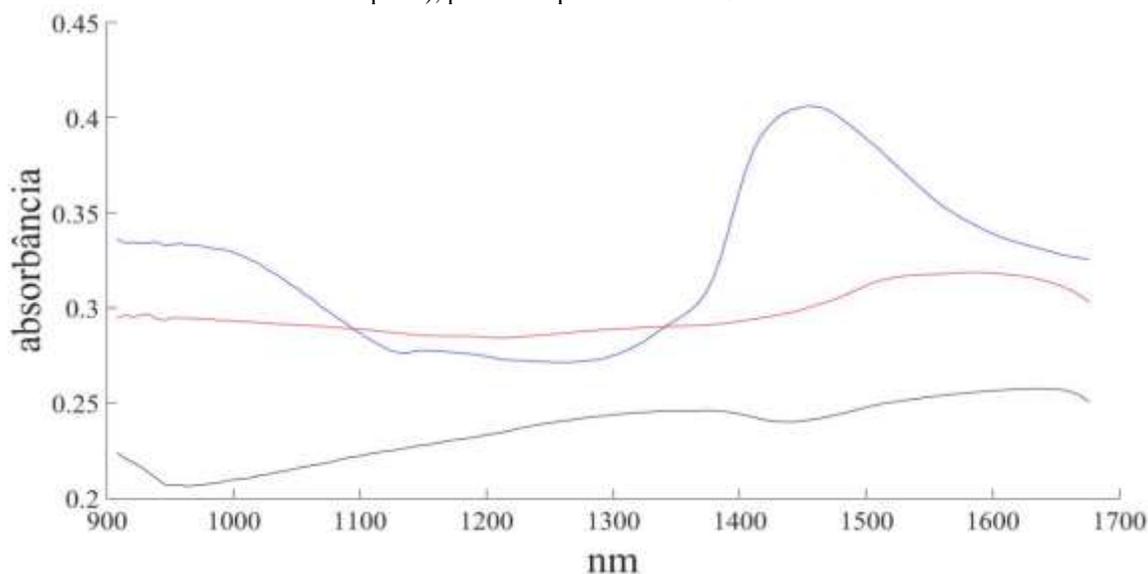
Fonte: O autor (2020)

Da Figura 6 foi observado que para a temperatura de 9°C as bandas características de água são observadas até cerca de 10 horas. Isso ocorre devido à baixa quebra das pontes de hidrogênio em baixas temperaturas, fazendo com que a evaporação da água seja mais lenta. Para 20°C e 27°C bandas de água são observadas até 02 horas após a deposição. Contudo, para a temperatura mais alta de 43°C a água evapora mais rapidamente e em 20 minutos a informação da água praticamente

não se encontra nos espectros. Após a secagem, basicamente 90% da composição da mancha é hemoglobina. A banda de absorção da oxi-hemoglobina a 930 nm encontra-se dentro da faixa espectral deste equipamento (908-1676 nm). Entretanto, apenas parte inicial da banda de absorção da hemoglobina, com máximo em 1690 nm, pôde ser observada (EDELMAN *et al.*, 2012). Logo, os espectros da mancha seca possuem baixa variação da absorbância quando comparados aos espectros das amostras frescas.

Com relação à influência do substrato, seu espectro de referência (média de uma triplicata coletada diretamente na placa de cerâmica) a Figura 7 compara as médias dos espectros nos tempos iniciais com os espectros dos tempos seguintes e o espectro de referência da cerâmica. Para avaliar o comportamento do conjunto de dados, foi selecionada a temperatura ambiente (27°C) como temperatura representativa para as demais, uma vez que o comportamento espectral foi semelhante para todas, com exceção no tempo de evaporação da água.

Figura 7 - Comparação entre a média dos espectros brutos, nos tempos de 0 minutos a 24 horas (em azul) e dos espectros dos tempos de 24 horas a 45 dias (em vermelho) e do espectro de referência do substrato cerâmico (em preto), para a temperatura de 27°C.

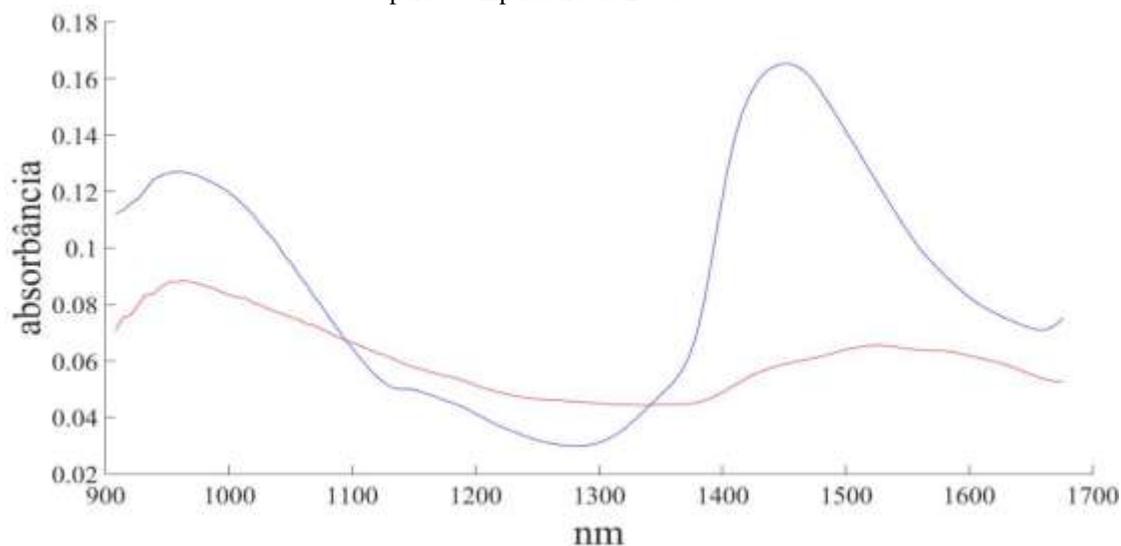


Fonte: O autor (2020)

Da figura anterior percebe-se que o espectro da placa teve baixa absorbância no comparativo com a da amostra, indicando uma inexpressiva influência nos espectros das amostras. Contudo, a informação da placa pode causar influência nos dados de interesse. Com isso o espectro do substrato foi removido do conjunto de dados para que se analise apenas o sangue e seus

componentes. A exclusão foi feita subtraindo o espectro de referência do substrato de todos os espectros obtidos do sangue neste mesmo substrato (Figura 8).

Figura 8 - Comparação entre a média dos espectros brutos, nos tempos de 0 minutos a 24 horas (em azul) e dos espectros dos tempos de 24 horas a 45 dias (em vermelho) após a remoção do espectro de referência do substrato para a temperatura de 27°C.



Fonte: O autor (2020).

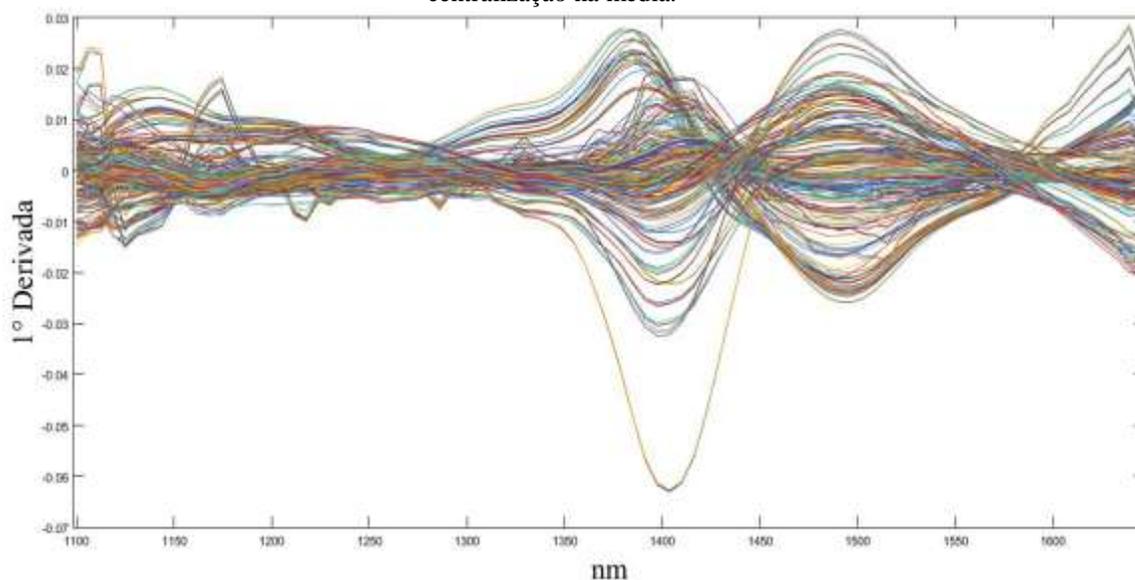
Da figura anterior, pôde-se observar que sem a influência do substrato, uma comparação entre a média dos espectros resultantes iniciais (0 minutos a 10 horas) e a média dos espectros resultantes dos tempos seguintes (24 horas a 45 dias) pode ser melhor analisado. Avaliando-se a média do tempo inicial (em azul) destaca-se as bandas de absorção bem definidas referentes a água (970 nm e 1454 nm) como já previsto, além da oxi-hemoglobina a 930 nm. Porém, esta última se encontrou em uma região muito ruidosa devido aos efeitos físicos e proximidade com a região do visível. Por outro lado, avaliando os tempos a partir de 24 horas (em vermelho) pode-se inferir que o perfil espectral não obteve mudanças significativas após a remoção da informação do substrato, em termos de bandas de absorção. Bandas de água a 970 e 1454 nm foram observadas, porém com bem menor intensidade do que para os tempos iniciais. Entretanto, a banda que se inicia em torno de 1400 nm é mais alongada possivelmente devido à sobreposição da banda da água com as bandas da hemoglobina, albumina e globulina que teriam máximo em 1690 nm (EDELMAN *et al.*, 2012).

Para melhor analisar o comportamento dos dados, foi realizado uma análise exploratória empregando PCA. Contudo, antes da análise multivariada foi de interesse aplicar as ferramentas de pré-processamento. Baseados em estudos prévios foram selecionadas duas combinações diferentes de pré-processamentos para avaliar qual melhor se aplicaria aos dados obtidos. Um

destes foi a primeira derivada, seguida de normalização pela faixa e centralização na média. Um outro grupo de pré-processamentos foi composto de SNV, com suavização Savitzky-Golay, seguido de centralização na média. Devido à similaridade entre os resultados obtidos aplicando-se o SNV e o MSC, este trabalho se ateve a usar apenas o SNV por verificar previamente que este se adequa melhor aos dados.

Primeiramente foi aplicada a combinação envolvendo a 1° derivada (Figura 9).

Figura 9 - Dados truncados em 1100-1650 nm e pré-processados aplicando a derivada, normalização pela faixa e centralização na média.



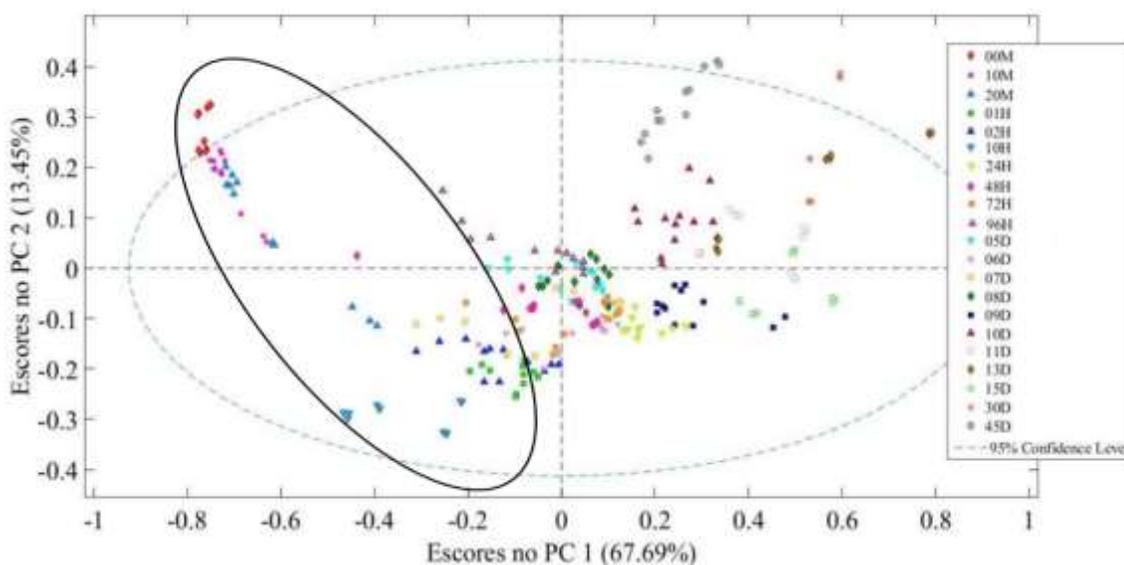
Da Figura 9 percebe-se o efeito da derivada, uma vez que há pontos de inflexão na região onde havia maior absorbância nos dados brutos, assim como a suavização dos espectros, que se apresentam menos ruidosos.

Evidenciou-se que as extremidades dos espectros (908-1100 nm e 1640-1676 nm) apresentaram forte influência no conjunto de dados devido a picos de alta intensidade. Por meio de uma PCA ficou claro no gráfico de pesos que esta região de extremidade influenciou de forma significativa na análise exploratória de dados, com picos elevados na PC1 e PC2. Além de que não foi possível observar no gráfico dos escores uma consistência nos dados com relação ao tempo, podendo ser devido a influência desta região da extremidade, mais ruidosa e menos confiável. Apesar de haver uma banda de oxi-hemoglobina a 930 nm, esta região foi removida devido ao alto ruído observado. A região final dos espectros também apresentou uma alta variabilidade a partir

de 1650 nm, com isso a região entre 1650 a 1676 nm também foi removido para evitar erros. Assim, os espectros analisados foram truncados na faixa espectral de 1100 nm a 1650 nm, como pôde ser observado nos dados pré-processados pela 1ª derivada ainda na figura anterior.

Após a remoção das regiões ruidosas e pré-processamento dos dados, a PCA foi aplicada (Figura 10).

Figura 10 - Gráfico dos escores para os dados pré-processados por derivada, considerando todos os tempos.

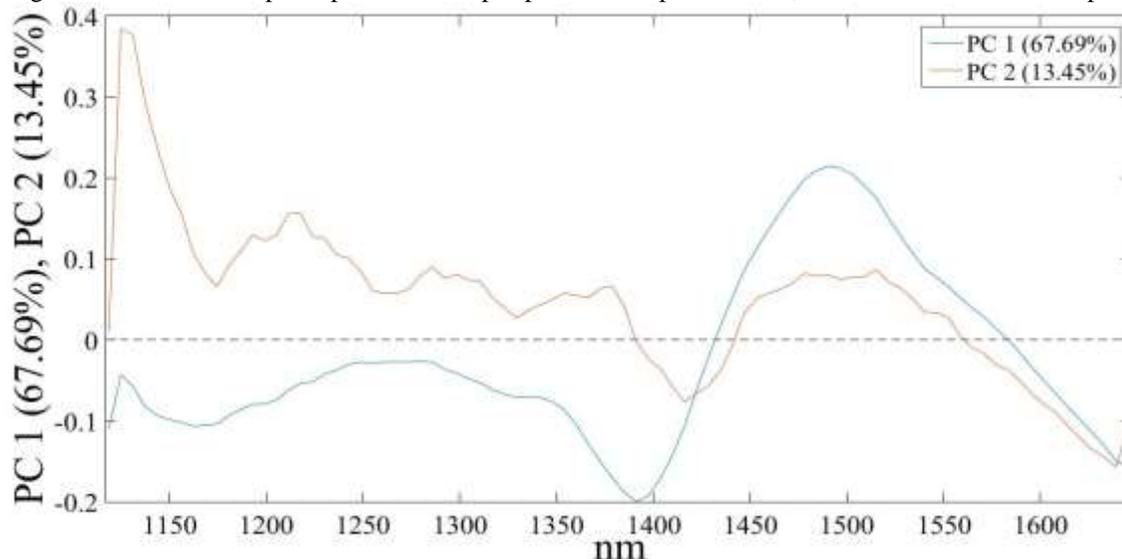


Fonte: O autor (2020).

Foi possível perceber uma separação considerável entre os dados iniciais (0 minutos a 10 horas), estes que foram circulado para melhor visualização, e os dados a partir de 24 horas, como já era esperado, devido a influência da água. Infere-se que as classes de 0 minutos, 10 minutos e 20 minutos ficaram bem próximos, já as classes para 1 hora e 2 horas ficaram mais distantes dos minutos iniciais, porém próximas entre si. A classe para 10 horas ficou próxima das classes a partir de 24 horas, evidenciando que às 10 horas o sangue já se encontra basicamente seco. Uma tendência temporal ao longo da PC1 foi observada nos dados a partir de 24 horas, porém não tão claro quanto para os tempos iniciais. Percebe-se que de 24 horas ao 8º dia as classes se encontraram bem próximas entre si e em torno da origem das PC's, havendo uma maior separação para os dados a partir do 9º dia até o 45º dia, que variam ao longo da PC1 com 67,7% de variância. Já ao longo da PC2, com variância de 13,5%, havendo uma tendência temporal apenas para os dados iniciais, principalmente para os minutos (de 0 a 20 minutos) e para os tempos finais de 9 a 45 dias.

O gráfico de pesos (Figura 11) também possui informações relevantes.

Figura 11 - Gráfico dos pesos para os dados pré-processados por derivada, considerando todos os tempos.



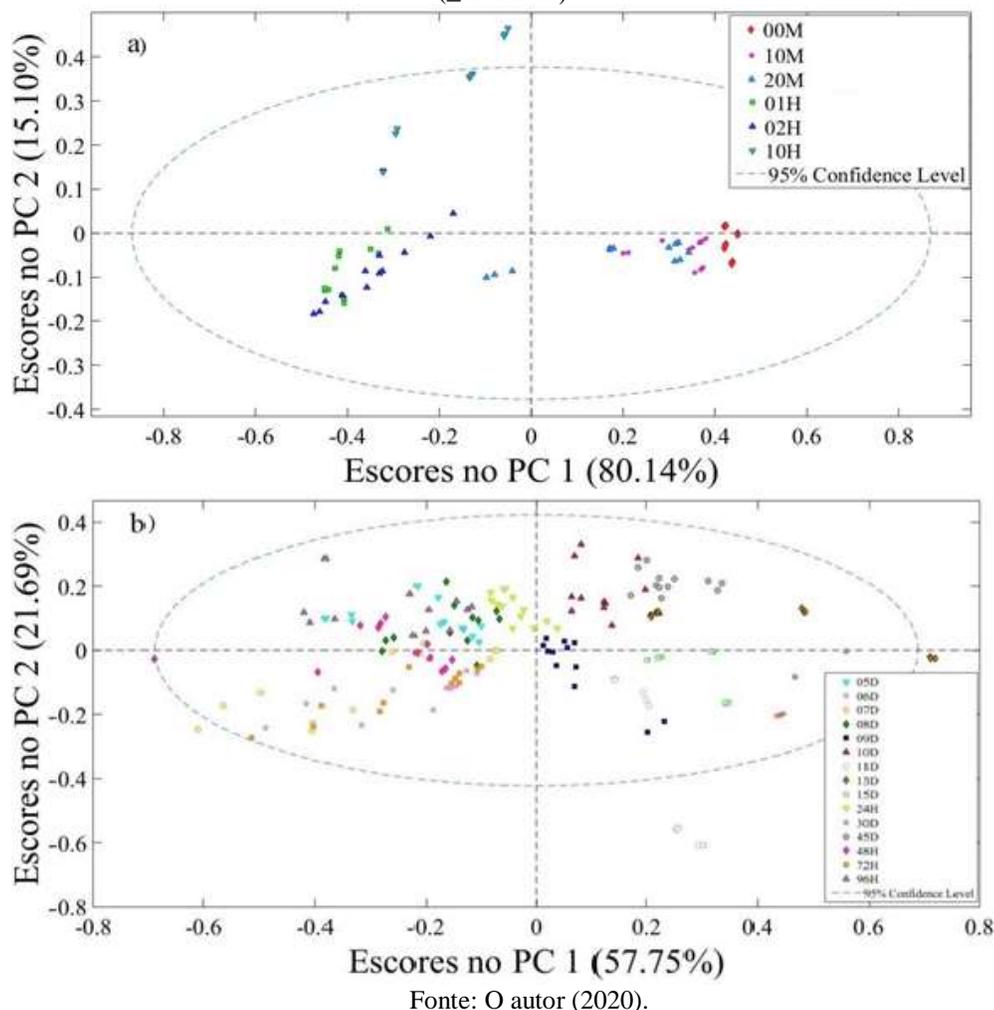
Fonte: O autor (2020).

A PC1 com 67,7% de variância mostrou a influência da água variando ao longo desta componente, com a derivada no ponto do pico da água em torno de 1425 nm, ficando este ponto na origem ( $X=0$ ) mostrando a alta influência desta banda de absorção nos dados obtidos. No gráfico dos pesos a PC2 apresenta banda na região de 1400 a 1600 nm, além de se observar um alto pico em torno de 1100 nm podendo ser proveniente ainda dos ruídos iniciais ou das caudas finais das bandas com picos em 930 nm e 970 nm (hemoglobina e água).

Há uma grande diferença nos escores das amostras nos tempos iniciais, ou seja, os tempos menores de 24 horas (para sangue fresco) e os tempos a partir de 24 horas (sangue antigo). Logo,

optou-se por fazer duas PCA: uma para as amostras frescas (Figura 12-a) e uma para amostras com maior tempo de deposição (Figura 12-b).

Figura 12 - a) Gráfico dos escores para amostras frescas (< 24 horas); b) Gráfico dos escores para amostras antigas ( $\geq 24$  horas).

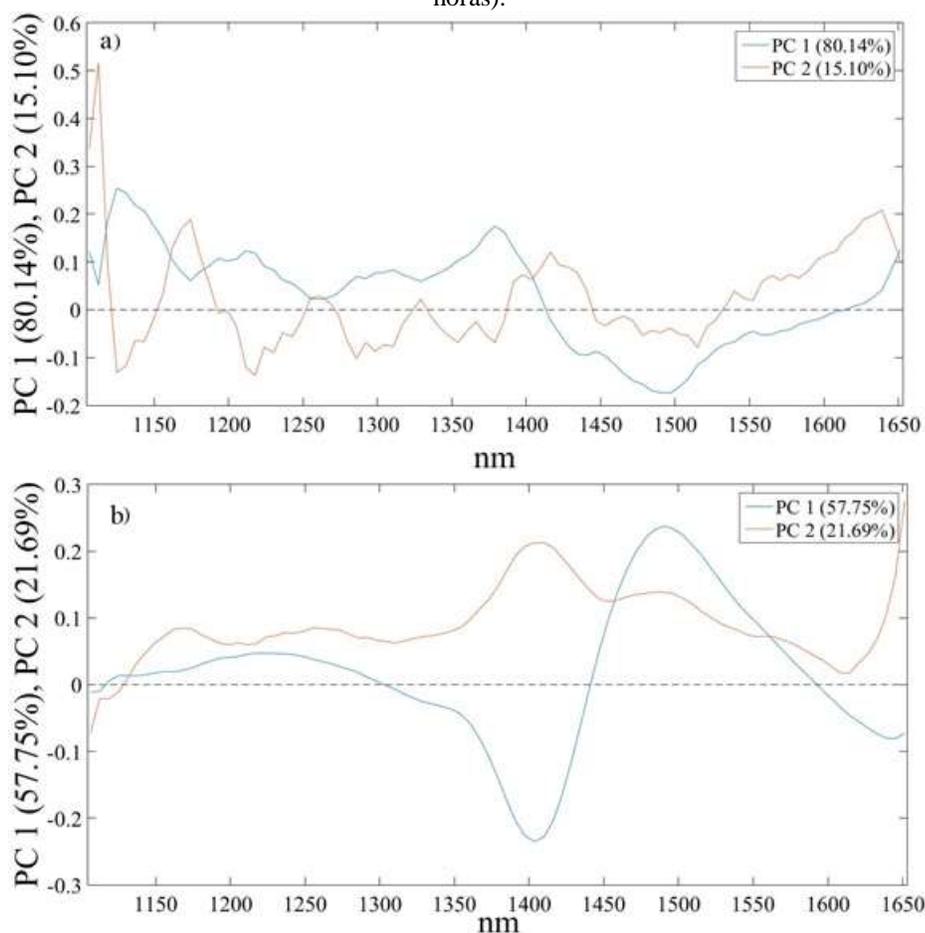


Da Figura 12-a, observa-se uma correlação temporal ao longo da PC1 para os tempos iniciais, principalmente para os minutos (0, 10 e 20 minutos), porém no sentido negativo da PC1. Foi perceptível que os tempos coletados ao longo dos minutos se agruparam, porém, ficaram distantes dos tempos medidos em horas (1 hora, 2 horas e 10 horas), indicando que o decaimento de água ocorreu rapidamente. Na PCA para as amostras antigas (Figura 12-b), não ficou evidente um comportamento temporal. Apenas foi observado que os dados com tempo menor ou igual a 8 dias permaneceram do lado esquerdo do eixo da PC1, enquanto os maiores ou iguais a 9 dias permaneceram do lado direito da PC1, porém de forma bastante aleatória, podendo sugerir que esta separação tenha algum significado químico. Isso pode ser devido a reação bioquímica de degradação do sangue, mais especificamente no processo de autoxidação da oxi-hemoglobina que

apresenta duas etapas de decaimento, como citado no trabalho de Bremmer *et al.* (2011). O rápido decaimento tem a duração de até 10 dias, em seguida passa-se a uma etapa de decaimento lento, tal efeito também foi identificado por Lin *et al.* (2017).

O gráfico de pesos (Figura 13-a, para amostras < 24 horas; e 13-b, para amostras  $\geq$  24 horas) apresentaram comportamentos bem distintos.

Figura 13 - a) Gráfico dos pesos para amostras frescas (< 24 horas); b) Gráfico dos pesos para amostras antigas ( $\geq$  24 horas).

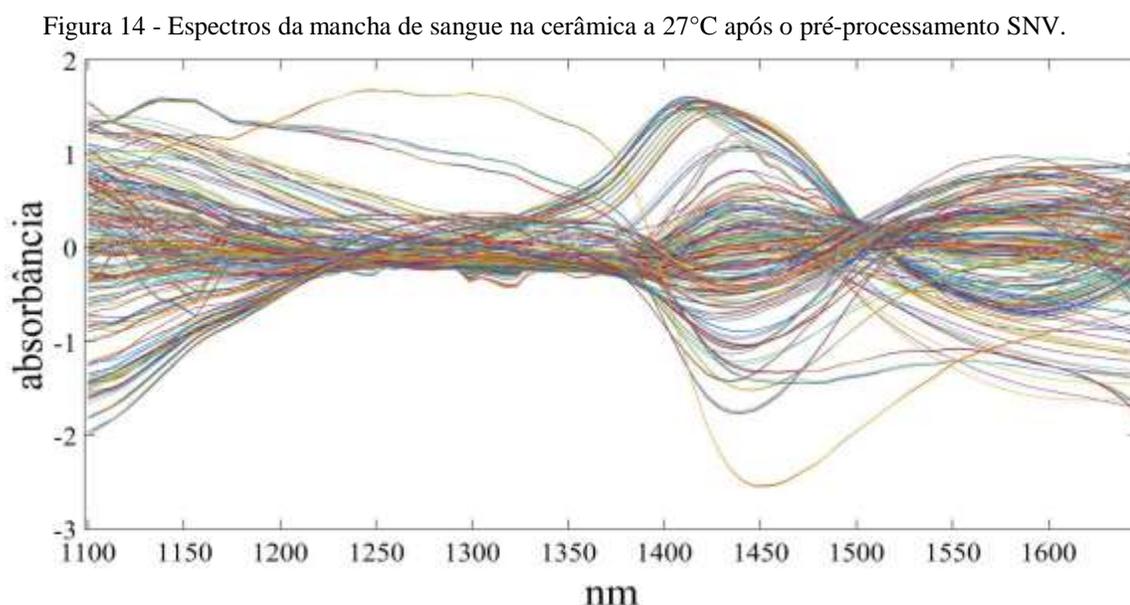


Fonte: O autor (2020).

Para as amostras recentes (Figura 13-a) a PC1 e PC2 apresentaram o ponto 1454 nm na origem do eixo das PC's, indicando que tal ponto era o valor máximo de absorvância ao ser derivado, ou seja, a maior influência foi da banda de água a 1454 nm. O mesmo comportamento é observado em torno de 1630 nm, podendo ser referente a influência do início da banda de absorção da hemoglobina a 1690 nm. Na PC2, observou-se um pico intenso a 1100 nm, podendo ser devido

a influência das bandas de absorção a 970 nm/930 nm, como já mencionado. No gráfico para amostras com tempo  $\geq 24$  horas (Figura 13-b), percebeu-se melhor o ponto de inflexão da PC1 a 1454 nm, indicando maior influência da banda de água a 1454 nm.

Diante do observado na análise exploratória de dados após realizar o conjunto de 1º derivada como pré-processamento, buscou-se a utilização de outro tratamento de dados na tentativa de obter a melhor informação possível dos espectros em questão. Aplicou-se aos dados a suavização (SG), seguido do SNV e centralização na média, previamente excluindo o ruído espectral das extremidades e a informação do substrato, obtendo-se os seguintes perfis espectrais mostrados na Figura 14.



Fonte: O autor (2020).

Observou-se que os espectros apresentaram um fim de banda de absorção de 1100 a 1250 nm, proveniente da absorção da hemoglobina e água em 930 e 970 nm, respectivamente. Além de uma banda completa de 1330 a 1500 nm, proveniente da absorção da água a 1454 nm, além de outra banda de 1500 a 1650 nm, proveniente da influência da banda de água com início da banda de absorção da hemoglobina, albumina e globulina a 1690 nm.

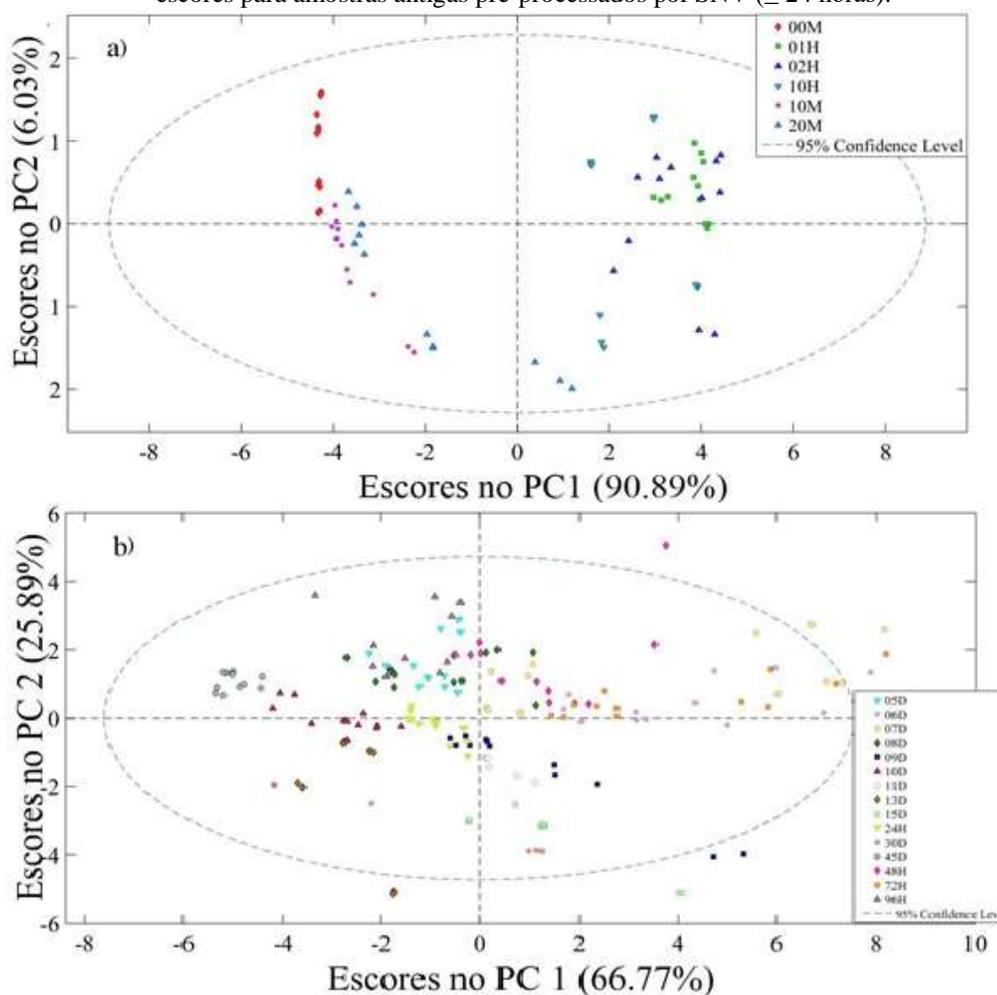
Para melhor avaliar o comportamento das classes de tempo com a variabilidade de doadores para as amostras pós tratados por SNV, foi realizado uma PCA (Figura 15).



Da figura anterior observa-se que a PC1, devido a banda de absorção com máximo em 1454 nm, com 56,7% de variância. Na PC2, com 34,9% de variância, um comportamento oposto ao eixo da PC1 foi observado, com uma banda de máximo em 1600 nm, sendo proveniente do início da banda de absorção da hemoglobina, albumina e globulina com máximo em 1690 nm. Como a PC1 e PC2 apresentaram comportamentos opostos, principalmente após 1300 nm, indica que os dados após aplicação do SNV e centralização na média ficaram normalizados de forma simétrica em torno do eixo zero, dando essa variância à PC1 e PC2 de comportamentos simetricamente opostos. Contudo, a variância da PC1 é influenciada principalmente pela informação da água a 1454 nm.

Da mesma forma realizada para a 1ª derivada, os dados foram explorados pela PCA nos tempos iniciais e finais separadamente (Figura 17).

Figura 17 - a) Gráfico dos escores para amostras frescas pré-processados por SNV (< 24 horas); b) Gráfico dos escores para amostras antigas pré-processados por SNV ( $\geq 24$  horas).

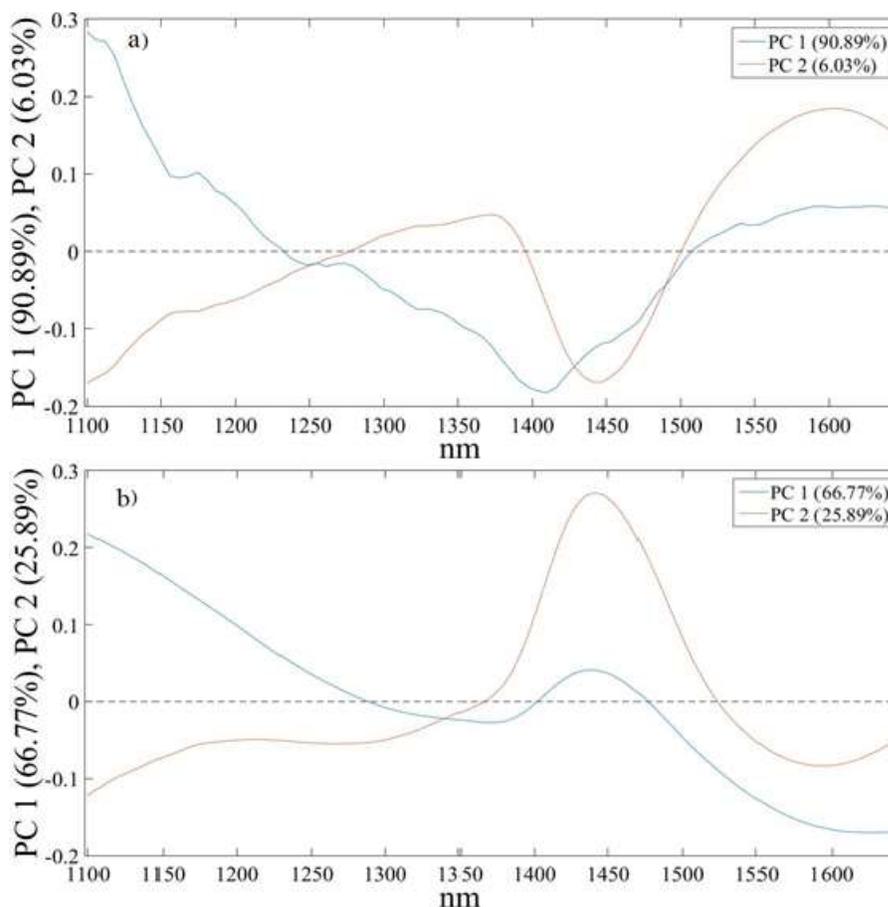


Fonte: O autor (2020).

Observa-se que para as classes de tempos iniciais na PC1 (Figura 17-a) ocorre uma tendência temporal, assim como foi observado utilizando dados derivados, porém agora no sentido positivo da PC1. Para os tempos finais (Figura 17-b), as classes se apresentam totalmente aleatórias, com nenhuma relação temporal. O comportamento observado após aplicação da 1ª derivada: uma separação de 24 horas a 8 dias e de 9 dias a 45 dias nos escores, não foi observada para o SNV.

Os gráficos de pesos para os dados separados após o SNV também apresentam informações relevantes (Figura 18).

Figura 18 - a) Gráfico dos pesos para amostras frescas pré-processados por SNV (< 24 horas); b) Gráfico dos pesos para amostras antigas pré-processados por SNV ( $\geq 24$  horas).



Fonte: O autor (2020).

Para os tempos iniciais (Figura 18-a) na PC1 é observado um decaimento na faixa de 1100 nm a 1200 nm, sendo causado pela influência da banda de água que absorve a 970 nm. Mesmo

tendo removido a faixa espectral inicial de 908 nm a 1100 nm para não haver influência do ruído espectral, esse decaimento é proveniente do final da banda de absorção. Com isso, devido a concentração de água a PC1 obteve 90,9% de variância. Para o gráfico de pesos a partir de 24 horas (Figura 18-b), a influência da água a 1454 nm é observado na PC2, com 25,9% de variância, porém, a PC1 com 66,8% tem maior influência da banda inicial de água (1100-1200 nm), assim como para os tempos iniciais. Analisando a PCA para todos os tempos após aplicação do SNV (Figura 18-b), o gráfico de pesos não demonstrou ter influência da variação de água na faixa de 1100-1200 nm. Porém, na análise dos tempos iniciais e dos tempos finais separadamente, os gráficos de pesos indicaram forte influência da água a 1454 nm.

Da mesma forma que na 1° derivada, para o SNV a partir de 24 horas foi possível perceber alta variabilidade entre os dados de mesmas classes e nenhuma relação temporal ao longo das PC's. Para a 1° derivada foi possível observar uma separação entre grupos de tempo menores ou iguais a oito dias e maiores ou iguais a nove dias. Porém, para o SNV essa tendência de separação não ficou evidente, com as classes dispersas aleatoriamente. Suspeita-se que os tempos finais de 30 e 45 dias estejam influenciando fortemente a PCA, podendo encobrir alguma informação até o 15° dia. Logo, outra análise exploratória de dados foi realizada apenas para o intervalo de tempo de 24 horas a 15 dias, tentando buscar uma lógica temporal para estes espectros, utilizando os dois conjuntos de pré-processamentos, mas também não foram observadas tendências temporais.

O mesmo comportamento geral observado para as manchas de sangue na cerâmica submetido a temperatura de 27°C foi observado nas PCA's realizadas com os espectros das amostras obtidas nas outras temperaturas investigadas (9°C, 20°C e 43°C), sob os mesmos pré-processamentos. Com isso, a análise exploratória de dados mostrou que após a secagem da amostra o complexo processo bioquímico de degradação da mancha não pode ser eficientemente detectado pelo equipamento. Porém, ficou evidente a diferença dos dados coletados nos tempos iniciais (até 24 horas), devido à alta quantidade de plasma (90% de água), com os dados para os tempos seguintes.

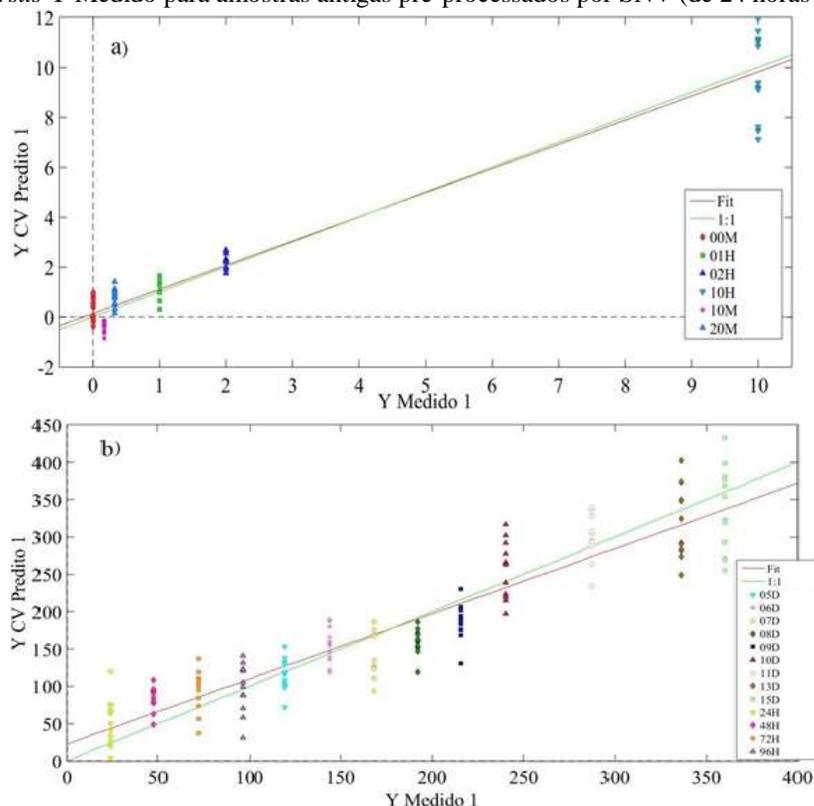
#### 4.1.1.2 Modelos de calibração para datação das manchas

No intuito de obter um modelo que possa datar o tempo desde a deposição do sangue, utilizou-se inicialmente regressão por PLS. Os modelos foram construídos utilizando-se os mesmos

pré-processamentos das PCA's e ainda considerando diferentes intervalos de tempo: 0 minutos a 10 horas; de 24 horas a 14 dias; de 24 horas a 45 dias; e de 0 minutos a 45 dias.

Primeiramente, a regressão foi realizada utilizando a combinação de pré-processamento da 1º derivada cujos gráficos de Y CV predito x Y medido estão na Figura 19.

Figura 19 - a) Y CV Predito *versus* Y Medido para amostras frescas pré-processados por SNV (< 24 horas); b) Y CV Predito *versus* Y Medido para amostras antigas pré-processados por SNV (de 24 horas até 15 dias).



Fonte: O autor (2020).

Das figuras anteriores observa-se que o ajuste foi melhor para os dados iniciais (Figura 19-a). Contudo, há de se considerar que o modelo é tendencioso, já que existe um lapso temporal grande entre as classes de 2 a 10 horas. Já para o modelo da Figura 19-b, há muitos pontos entre 24 horas a 15 dias, fazendo com que se perceba um comportamento entre 24 horas a 8 dias e um outro comportamento de 9 a 15 dias, que foi observado na PCA. Já as classes após 15 dias, possuem também um grande lapso de tempo (de 15 a 45 dias), fazendo com o modelo seja tendencioso. Logo, regressões foram feitas considerando diferentes intervalos de tempo, para contornar a falta

de dados entre classes de tempo. Os parâmetros que determinam a qualidade do modelo estão dispostos na Tabela 3 e organizados pelos intervalos de tempo para facilitar o entendimento.

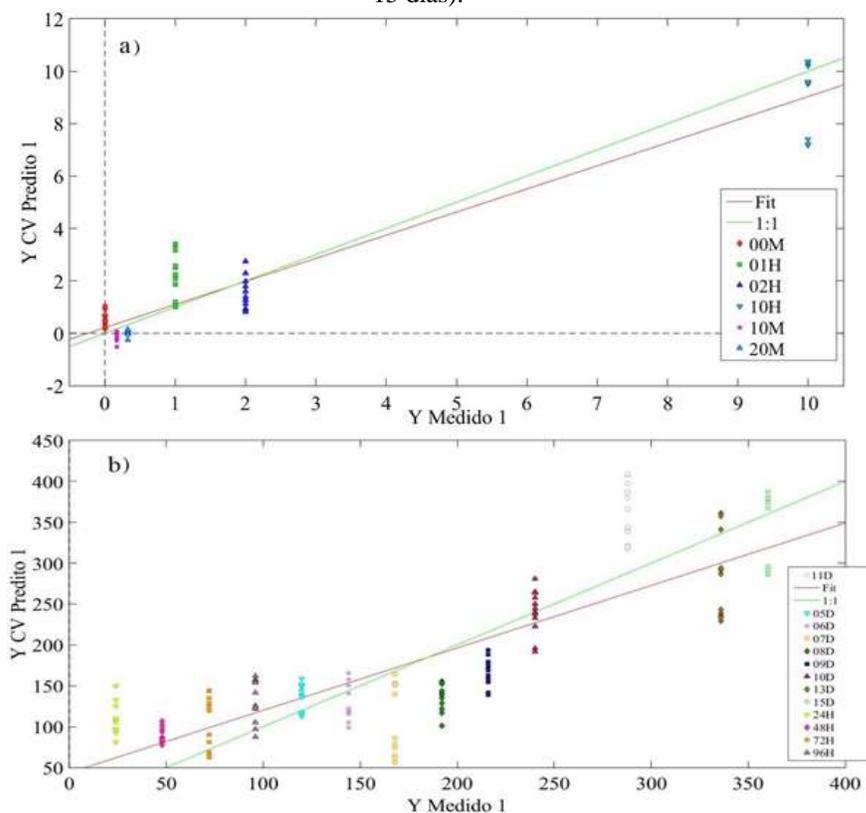
Tabela 3 - Valores obtidos da regressão PLS após pré-processamento por derivada para diferentes intervalos de tempo.

<b>Intervalo de tempo</b>	<b>VL</b>	<b>RMSEC (horas)</b>	<b>RMSECV (horas)</b>	<b>Bias</b>	<b>R<sup>2</sup></b>
0 minutos a 10 horas	6	0,5	0,8	0,1	0,9
24 horas a 15 dias	5	47,7	54,2	0,8	0,7
24 horas a 45 dias	5	87,6	101,9	-0,1	0,9
0 minutos a 45 dias	4	90,6	106,3	-3,7	0,8

Fonte: O autor (2020).

A tabela anterior mostra que o modelo de 0 a 10 horas teve um R<sup>2</sup> de 0,9, porém altos erros RMSE's. Logo, o modelo foi tendencioso a ser linear por haver poucos pontos entre as classes de 2 a 10 horas. Para os demais intervalos de tempo, também foram observados altos RMSE's. Logo, não se obteve um bom ajuste com a derivada como pré-processamento dos dados. Também foi utilizado o pré-processamento por SNV para tentar obter um bom modelo de regressão (Figura 20).

Figura 20 - a) Y CV Predito (tempo em horas) *versus* Y Medido (tempo em horas) para amostras frescas (< 24 horas); b) Y CV Predito (tempo em horas) *versus* Y Medido (tempo em horas) para amostras antigas (de 24 horas até 15 dias).



Fonte: Autor (2020).

Os parâmetros para avaliar a qualidade do modelo foram tabelados para facilitar o entendimento (Tabela 4).

Tabela 4 - Valores obtidos da regressão PLS após pré-processamento SNV.

Intervalo de tempo	VL	RMSEC (horas)	RMSECV (horas)	CV Bias	R <sup>2</sup>
0 minutos a 10 horas	6	0,6	1,1	0,1	0,9
24 horas a 15 dias	5	50,9	61,4	1,9	0,6
24 horas a 45 dias	6	93,5	113,6	0	0,8
0 minutos a 45 dias	5	101,8	117,9	-0,4	0,8

Fonte: O autor (2020).

De acordo com a tabela anterior o modelo aplicando o SNV para o intervalo de 0 minutos a 10 horas, obteve basicamente os mesmos parâmetros no comparativo com o pré-processamento da derivada. O modelo de regressão considerando todas as amostras para todas as classes de tempo

(de 0 minutos a 45 dias) obteve um erro RMSEC e RMSECV com altos valores, indicando que não houve um ajuste adequado. Da mesma forma, para modelos envolvendo intervalos de tempo (de 24 horas a 45 dias; 24 horas a 15 dias; 0 minutos a 10 horas) os RMSE's se apresentaram com altos valores, logo os modelos não se mostraram adequados. Foram observados valores altos de RMSEC e RMSECV (em horas), para ambos os pré-processamentos. Para o SNV os erros obtidos foram ainda maiores do que os para a 1° derivada (Tabela 3).

Observa-se uma grande dispersão entre os dados para uma mesma classe, devido as triplicadas de cada doador. A regressão para os dados de 0 minutos a 45 dias, faz com que os tempos em minutos e horas fiquem aglomerados no início da regressão, devido a grande diferença de tempo, sendo fonte de erros. Também ficou evidente na PCA que o comportamento difere entre as amostras consideradas frescas (< 24 horas) e antigas ( $\geq$  24 horas). Para o intervalo de tempo de 24 horas a 15 dias, é perceptível uma tendência curva quadrática de 24 horas a 8 dias e de outra curva quadrática de 9 dias a 14 dias, esse comportamento foi observado nos escores da PCA dos dados após a derivada, sugerindo que os dados não possuem tendência linear.

Devido à alta dispersão dos dados fez-se a média das triplicatas para avaliar o comportamento dos doadores, consequentemente com ambos pré-processamentos e aplicação do PLS (Tabela 5).

Tabela 5 - Valores obtidos da regressão PLS das médias das triplicatas após pré-processamento derivada e SNV.

<b>Pré-processamento</b>	<b>Intervalo de tempo</b>	<b>VL</b>	<b>RMSEC (horas)</b>	<b>RMSECV (horas)</b>	<b>CV Bias</b>	<b>R<sup>2</sup></b>
Derivada	0 minutos a 10 horas	5	0,7	1,6	0,2	0,8
	24 horas a 15 dias	5	50,2	80,9	6,6	0,5
SNV	0 minutos a 10 horas	4	1,1	1,5	0,1	0,8
	24 horas a 15 dias	3	60,5	69,9	2,2	0,5

Fonte: O autor (2020).

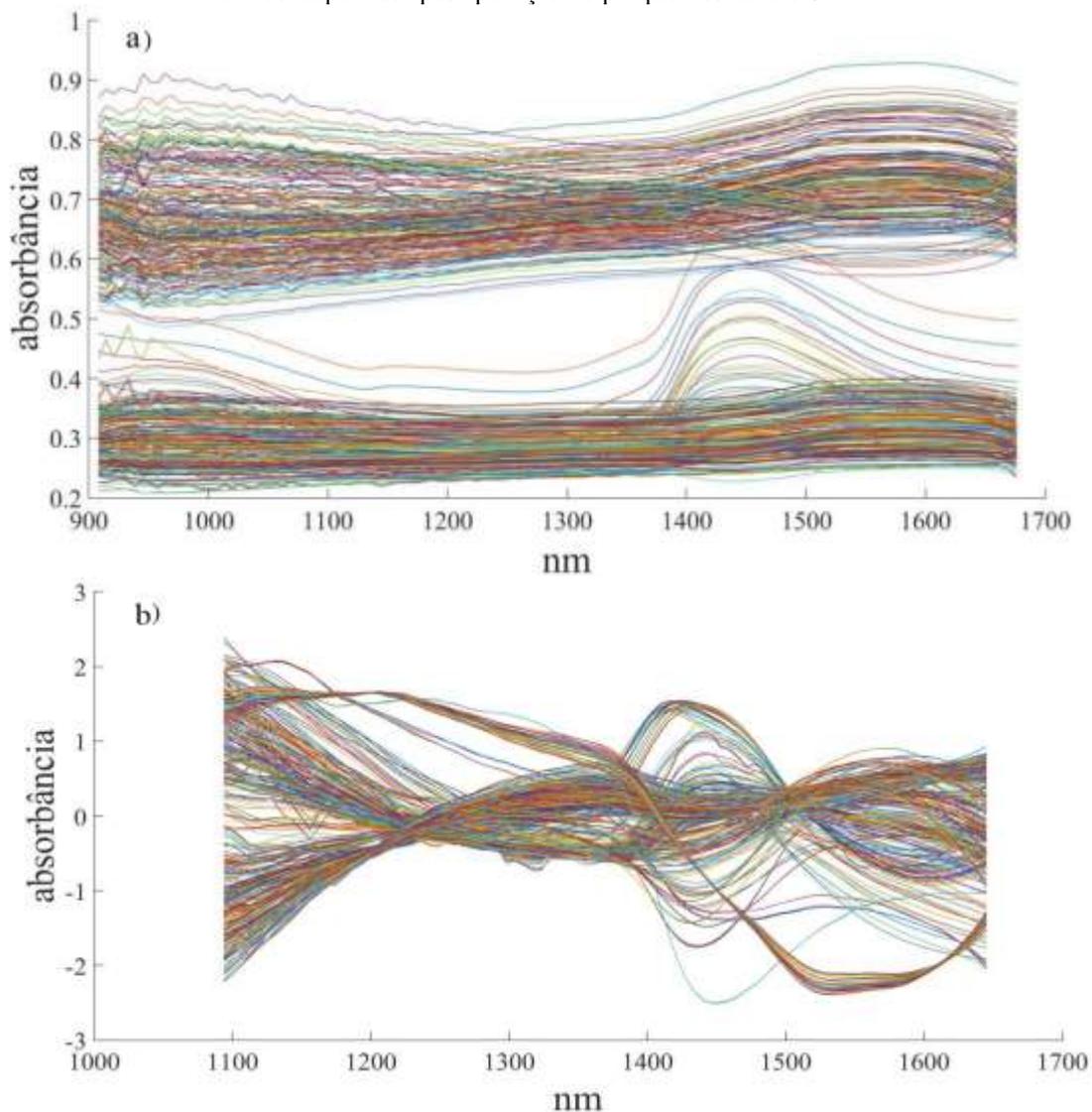
Na aplicação do PLS no conjunto de dados (Tabela 5), ficou evidente a sensibilidade da técnica frente a diferentes doadores, uma vez que baixos R<sup>2</sup> foram obtidos e altos erros RMSE's foram observados para ambos pré-processamentos. Apesar de obter triplicatas, as amostras (duas de cada doador para uma placa) apresentavam dispersão entre si, além da dispersão entre os

próprios doadores para uma mesma classe de tempo, fazendo com que se confirmasse a influência na diferença do sangue, que é particular para cada indivíduo, como já foi comentando em estudos anteriores, como o de Virkler & Lednev em 2010.

Na intenção de tornar o modelo mais robusto e adicionar maior variabilidade nos dados, foi adicionado ao conjunto de dados, outros doadores (dois homens e duas mulheres: H1, H2, M3, M4) para tempos já existentes (2 horas, 24 horas, 05 dias, 08 dias) e adição de outros tempos (18, 28 e 49 dias), provenientes de outros dados cedidos por outro pesquisador do grupo. Estes dados que foram cedidos por um pesquisador do grupo, que os obteve na realização do trabalho Pereira *et al*, (2017). Com a inserção dessas novas amostras a faixa de tempo foi coberta mais uniformemente e com um número mais elevado de amostras foi possível selecionar diferentes conjuntos para validar externamente os modelos de regressão.

Os espectros brutos mostram uma grande diferença de linha de base (Figura 21-a), devido às condições do equipamento, uma vez que os dados introduzidos foram coletados com uma diferença de 2 anos para os dados coletados para esta pesquisa. Entretanto, essa variação na linha de base foi reduzida a partir do pré-processamento SNV (Figura 21-b). Porém, foi observado que ambos os dados contêm o ruído espectral nas extremidades (908-1100 nm e 1650-1676 nm), assim essa faixa foi excluída, como já feito anteriormente.

Figura 21 - a) Dados ampliados brutos, acima os dados introduzidos, abaixo os dados coletados para esta pesquisa; b) Dados ampliados após aplicação do pré-processamento SNV.



Fonte: O autor (2020)

Das figuras anteriores observa-se que o SNV foi efetivo ao corrigir a diferença de linhas de base entre os espectros de dois estudos diferentes, porém que empregaram a mesma metodologia. Com isso, Modelos PLS foram obtidos para os mesmos intervalos de tempo já comentados: de 0 minutos a 49 dias; 24 horas a 49 dias; 24 horas a 15 dias; e de 0 minutos a 10 horas (Tabela 6).

Tabela 6 - Valores obtidos com o PLS para os dados ampliados.

<b>Intervalo de tempo</b>	<b>VL</b>	<b>RMSEC (horas)</b>	<b>RMSECV (horas)</b>	<b>CV Bias</b>	<b>R<sup>2</sup></b>
0 minutos a 10 horas	4	1,0	1,2	0,1	0,8
24 horas a 15 dias	7	64,1	72,1	0,6	0,7
24 horas a 49 dias	7	123,9	136,9	-0,4	0,8
0 minutos a 49 dias	3	218,9	223,8	0,9	0,6

Fonte: O autor (2020).

De 0 minutos a 49 dias, o valor de RMSECV foi de 223,8 horas, cerca de 10 dias, que é bastante alto, mas o R<sup>2</sup> de 0,8 mostra que os dados têm comportamento linear. Já nos modelos construídos na faixa de 24 horas a 49 dias o RMSECV foi de 136,9 horas, cerca de 5,7 dias, um valor também muito alto para casos forenses que requerem boa exatidão, mantendo o R<sup>2</sup> de 0,8. De 24 horas a 15 dias, obteve-se um RMSECV de 72,1 horas, cerca de 3 dias, sendo um resultado basicamente igual ao dos dados antes da ampliação, contudo o R<sup>2</sup> de 0,7 já mostra que os valores medidos e preditos não apresentam tanta relação linear como esperado. De 0 minutos a 10 horas, tem-se o RMSECV de 1,2, cerca de 1 hora. Pelo fato de haver poucos níveis de tempo e com uma distribuição não uniforme ao longo da faixa, o modelo PLS para as amostras frescas não foram adequados. Considerando também que inicialmente há apenas a evaporação da água nas amostras sendo detectado pelo equipamento, isso torna difícil a reprodução, não sendo totalmente representativo um modelo para o estudo de datação da mancha, como também citado no estudo de Lin *et al.* (2017).

Visando diminuir a dispersão e buscando dirimir a diferença da grande quantidade de dados que foram adicionados para certas classes, foi realizada uma regressão PLS das médias das triplicatas de cada doador (Tabela 7).

Tabela 7 -Valores obtidos para os modelos PLS para as médias das replicatas dos dados ampliados.

<b>Intervalo de tempo</b>	<b>VL</b>	<b>RMSEC (horas)</b>	<b>RMSECV (horas)</b>	<b>CV Bias</b>	<b>R<sup>2</sup></b>
0 minutos a 10 horas	4	1,1	1	0,3	0,7
24 horas a 15 dias	9	35	71	2,3	0,5
24 horas a 49 dias	8	107,6	149,5	2,9	0,8
0 minutos a 49 dias	6	170,2	220,1	-18,8	0,6

Fonte: O autor (2020).

Os valores de RMSECV obtidos para as distintas faixas de tempo foram relativamente similares aos dos dados sem a obtenção das médias, apesar de que os R<sup>2</sup> ficaram bem abaixo no comparativo com os dados sem a média. Além de que se observou um aumento no número de variáveis latentes, indicando que o modelo necessitou de mais variáveis para poder se ajustar aos dados de calibração, fazendo com que os erros RMSEC diminuíssem no comparativo com os dados sem a média, porém na validação cruzada o RMSECV foi basicamente o mesmo.

Na intenção de aplicar um método que lida melhor com não linearidades, foi utilizada a técnica de regressão SVM. Os espectros foram pré-processados utilizando SNV, suavização, faixa reduzida e remoção do substrato. As médias das replicatas foram obtidas objetivando diminuir a dispersão.

Como a técnica SVM pode levar a modelos que funcionam bem na calibração, mas não são adequados para previsão, fez-se a divisão dos dados em calibração e validação externa. Quatro divisões diferentes no conjunto de dados foram realizadas (Tabela 8). Os Dados 1 e 2 foram obtidos utilizando 50% para calibração e o restante para validação externa, já nos Dados 3 e 4 foram obtidos utilizando 70% para criar o modelo e 30% para a validação externa. Os doadores utilizados para calibração nos Dados 1 foram utilizados para validação externa nos Dados 2, e assim foi feito com os outros dados desses dois conjuntos. Já a separação para os Dados 3 e 4 ocorreram de forma aleatória. Além disso, os dados usados para construir os modelos foram a partir de 24 horas até 49 dias.

Tabela 8 – Conjunto de dados selecionados para calibração e para validação externa.

<b>Cerâmica a 27°C (doadores 1, 2, H1, H2, M3, M4)</b>							
Dados 1		Dados 2		Dados 3		Dados 4	
Calibração	Validação externa	Calibração	Validação externa	Calibração	Validação externa	Calibração	Validação externa
Doadores 2/H2/M4	Doadores 1/H1/M3	Doadores 1/H1/M3	Doadores 2/H2/M4	Doadores 1/H1/H2/M4	Doadores 2/M3	Doadores 2/H2/M3/M4	Doadores 1/H1

Fonte: O autor (2020).

Com isso os parâmetros de regressão obtidos com aplicação do SVM estão apresentados na Tabela 9. Para analisar como os dados da validação externa se comportaram na regressão.

Tabela 9 -Parâmetros dos modelos SVM com diferentes combinações de dados para calibrar e validar externamente.

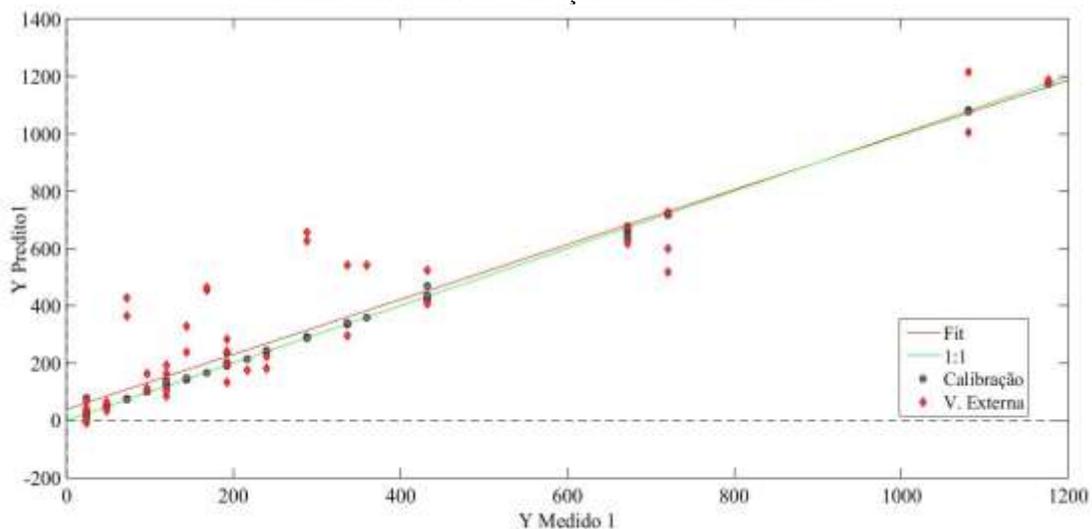
Dados	Intervalo de tempo	RMSEC (horas)	RMSECV (horas)	REMSEP (horas)	CV Bias	R <sup>2</sup>
1	24 horas a 49 dias	32,3	77,4	80,8	0,5	0,9
2	24 horas a 49 dias	3,5	114,4	64,9	3,5	0,9
3	24 horas a 49 dias	3,5	85,3	109,7	8,8	0,9
4	24 horas a 49 dias	12,2	65,7	79,6	0,2	0,9

Fonte: O autor (2020).

Da tabela anterior observa-se que os R<sup>2</sup> de 0,9 indicam que os ajustes entre os valores de calibração com os valores de validação externa apresentam ajuste com tendência linear, fazendo com que o SVM tenha maior influência na obtenção de melhores modelos de calibração.

O modelo SVM dos Dados 4 foi utilizado para a representação gráfica (Figura 22), uma vez que apresentou menor erro de predição (REMSEP).

Figura 22 – Valores preditos *versus* referência do modelo SVM para os Dados 4 na cerâmica. Dados de calibração em cinza e os dados de validação externa em vermelho.



Fonte: O autor (2020).

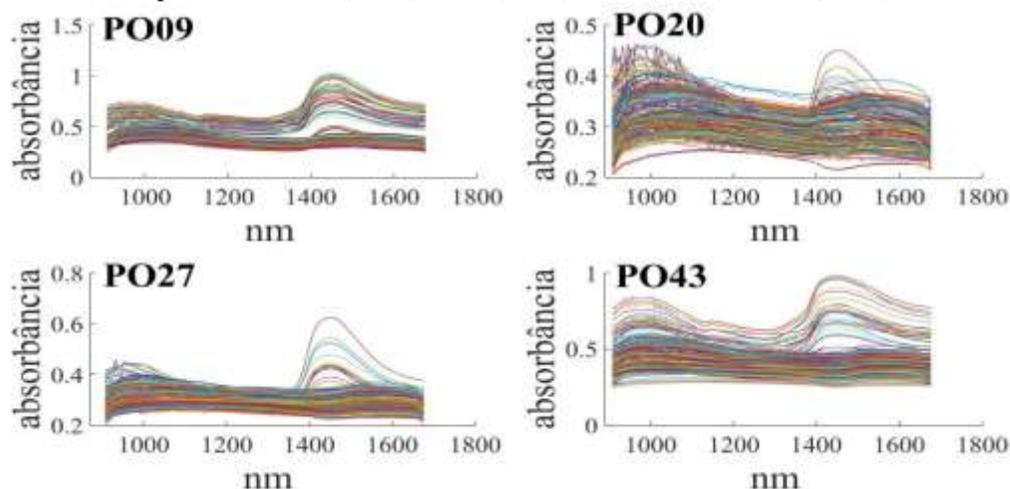
No comparativo com a regressão PLS das médias dos dados ampliados, infere-se que para o período de 24 horas a 49 dias o modelo SVM se apresentou com parâmetros mais ajustados, com erros RMSE's mais baixos e  $R^2$  mais altos. Porém, considerando a finalidade do modelo, os ajustes ainda não se mostraram adequados para prever com confiabilidade o tempo de deposição do sangue. Para os Dados 2 e 3, o modelo apresentou um RMSEC baixo de 3,5 horas para ambos os conjuntos de dados no comparativo com REMSECV e REMSEP elevados. Isto sugere que este modelo está sobreajustado (*overfitting*), ou seja, os dados utilizados na etapa de calibração se ajustaram adequadamente, mas não foi capaz de generalizar os dados na etapa de validação e previsão. Para os Dados 1, observa-se que o RMSEP foi de 80,8 horas, cerca de 3,4 dias de erro, já para os Dados 4 o RMSEP foi de 79,8, cerca de 3,3 dias sendo um valor relativamente alto para fins forenses. Contudo, levando em consideração a limitação do equipamento (faixa espectral limitada) e a quantidade de dias avaliados pelo modelo: de 24 horas (1 dia) a 49 dias; infere-se que 3,4 e 3,3 dias de erros poderiam, baseado em resultados da literatura, serem considerados um resultado satisfatório. Os modelos SVM poderiam ser empregados para obter uma estimativa do tempo de deposição, porém os associando a outros fatores: outros vestígios do local do crime, testemunhas, filmagens, tempo de desaparecimento relatado por parentes. Com isso, obtendo um contexto que possa auxiliar na resolução de delitos.

Os RMSE's relativamente elevados são característicos da alta complexidade e influência do ambiente, mesmo que controlando as temperaturas, na reação bioquímica de degradação do sangue, como já mencionado por outros autores. Ressalta-se também a faixa espectral limitada monitorada pelo equipamento portátil utilizado. Contudo, é possível determinar qualitativamente se as amostras são frescas ( $< 24$  horas) ou amostras antigas ( $\geq 24$  horas) por meio de análise exploratória de dados. Na determinação de manchas antigas, pode-se utilizar o modelo SVM obtido para estimar o tempo de deposição do sangue, porém com um erro em torno de 3 dias para uma faixa de análise de 1 a 49 dias, ou seja, 6,25% de erro.

#### 4.1.2 Espectros NIR no substrato de porcelanato

Os espectros obtidos apresentaram um comportamento semelhante para todas as temperaturas, com efeito de espalhamento espectral e com ruído de 908 a 1100 nm, com alguns espectros com bandas de absorção a 970 nm e 1454 nm e outros com pouca variabilidade de absorbância (Figura 23).

Figura 23 - Espectros no infravermelho próximo obtidos em equipamento portátil para a placa de porcelanato nas temperaturas de 9°C (PO09), 20°C (PO20), 27°C (PO27) e 43°C (PO43).



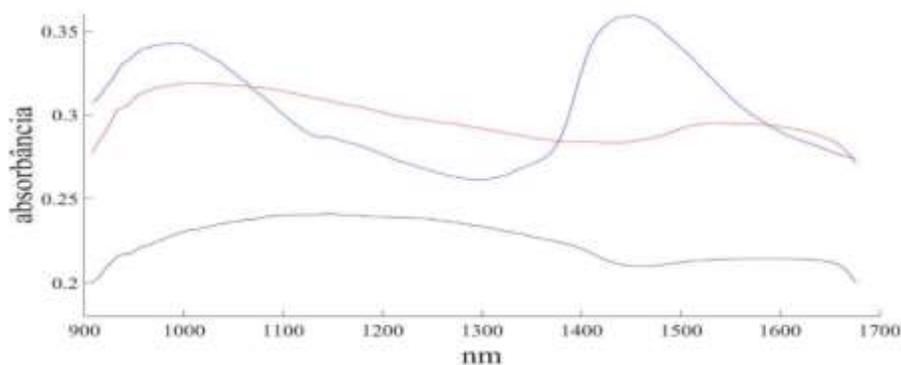
Fonte: O autor (2020).

Os espectros brutos obtidos para o sangue depositado nas placas de porcelanato foram semelhantes aos espectros obtidos na cerâmica, inclusive no tempo de evaporação da água nas

amostras para cada temperatura. Utilizou-se os dados a 27°C, aplicou-se os mesmos pré-processamentos e os espectros foram analisados por PCA, PLS e SVM.

O espectro de referência do porcelanato (obtido da média de uma triplicata dos espectros coletados da placa a 27°C) no comparativo com a média dos espectros nos tempos iniciais e com a média dos espectros nos tempos restantes (Figura 24), apresentou baixa significância e pouca influência sobre os dados.

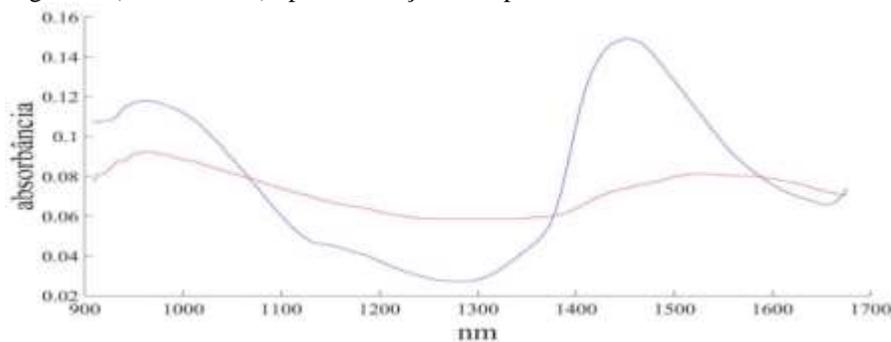
Figura 24 - Comparação entre a média dos espectros nos tempos iniciais (em azul), dos espectros restantes (em vermelho) e do espectro de referência do substrato porcelanato (em preto).



Fonte: O autor (2020).

Especificamente para os espectros dos tempos iniciais (em azul) foi obtido um comportamento parecido com o da cerâmica. Observou-se bandas de oxi-hemoglobina a 930 nm e de água a 970 nm e 1454 nm, mostrando a alta influência da absorção das moléculas de água devido ao seu forte momento dipolo. De qualquer forma foi considerável fazer a remoção dos efeitos da absorção do substrato (Figura 25).

Figura 25 - Média dos espectros da mancha de sangue na cerâmica nos tempos iniciais (em azul) e nos tempos seguintes (em vermelho) após a remoção do espectro de referência do substrato.



Fonte: O autor (2020).

Após a remoção do substrato os picos de absorção se intensificaram. Para o porcelanato os espectros também apresentaram ruído espectral na extremidade (908-1100 nm) devido a aproximação com a região do visível. Logo, estes foram removidos do conjunto de dados, ficando a análise espectral na região de 1100 nm a 1650 nm.

PCA's foram realizadas com as combinações de pré-processamentos já mencionadas. Independente do pré-processamento, os mesmos comportamentos foram observados em comparação com os dados na cerâmica, ou seja, uma separação entre amostras frescas (< 24 horas) devido a uma alta concentração de água e entre amostras mais antigas ( $\geq$  24 horas) devido à presença apenas de hemoglobina, nestas últimas.

Utilizando os mesmos pré-processamentos, modelos PLS foram obtidos. Analisando as amostras frescas, pelo fato de haver um alto lapso de tempo (de 2 horas a 10 horas e de 10 horas a 24 horas) e pela limitada região espectral do equipamento, os modelos PLS não se mostraram adequados para detectar a reação bioquímica de degradação do sangue. A metodologia foi eficiente ao mostrar a variação de água ao longo do tempo. Porém, já foi ressaltado que datar a mancha de sangue a partir da concentração de água é muito suscetível a erros. Logo, o estudo da regressão foi realizado com foco nas amostras antigas ( $\geq$  24 horas), como pode ser visto na Tabela 10.

Tabela 10 - Valores obtidos da Regressão PLS das médias das triplicatas após pré-processamento derivada e SNV.

<b>Pré-processamento</b>	<b>Intervalo de tempo</b>	<b>VL</b>	<b>RMSEC (horas)</b>	<b>RMSECV (horas)</b>	<b>CV Bias</b>	<b>R<sup>2</sup></b>
Derivada	0 minutos a 10 horas	2	1,6	1,9	-0	0,7
	24 horas a 15 dias	5	54,4	61,9	-1,2	0,6
	24 horas a 45 dias	3	127,2	147,7	-5,6	0,7
	0 minutos a 45 dias	4	94,1	110,1	-2,8	0,8
SNV	0 minutos a 10 horas	5	1,2	1,3	0	0,9
	24 horas a 15 dias	6	43,8	45,4	0,2	0,8
	24 horas a 45 dias	5	123,2	127,4	1,2	0,8
	0 minutos a 45 dias	5	119,1	121,9	-0,2	0,8

Fonte: O autor (2020).

A regressão apresentou alto erro devido à alta dispersão dos dados entre as próprias triplicatas dos doadores, ao lapso temporal de análises que há entre 15-30 dias e de 30-45 dias. Devido a isso, foi considerado que os dois dias finais (30° e 45° dia) podem ser fontes de erros consideráveis na regressão por apresentarem tempos muito esparsos entre o outro, fazendo com que a regressão tenha caráter linear de forma tendenciosa. Porém, mesmo fazendo regressão de 24 horas a 15 dias, observou-se para o porcelanato que o modelo PLS não foi adequado devido aos altos erros RMSE's encontrados. Para os intervalos entre 24 horas a 8 dias e de 9 dias a 15 dias, foram observadas duas tendências não lineares, com características de curvas quadráticas, indicando haver mudanças bioquímicas nesses intervalos de tempo.

No intuito de dirimir o erro da dispersão dos dados foi realizado a média das triplicatas, porém a regressão também não obteve resultados satisfatórios (Tabela 11). Isso se deu por ter poucos pontos após obter as médias, o modelo não apresentou robustez e os mesmos erros foram observados.

Tabela 11 - Valores obtidos com o PLS para as médias das triplicatas.

<b>Pré-processamento</b>	<b>Intervalo de tempo</b>	<b>VL</b>	<b>RMSEC (horas)</b>	<b>RMSECV (horas)</b>	<b>CV Bias</b>	<b>R<sup>2</sup></b>
Derivada	0 minutos a 10 horas	4	0,7	1,6	0,4	0,8
	24 horas a 15 dias	6	44,2	70,6	2,4	0,6
SNV	0 minutos a 10 horas	3	1,2	1,7	0,2	0,8
	24 horas a 15 dias	3	66,1	80,9	0,9	0,4

Fonte: O autor (2020).

Para dirimir tais problemáticas, novos dados foram inseridos no conjunto de dados, cobrindo as mesmas classes de tempo (2 horas, 24 horas, 5 dias, 8 dias, 18 dias, 28 dias e 49 dias) e os mesmos doadores (H1, H2, M3 e M4) adicionados na cerâmica, com exceção que esses dados foram obtidos no porcelanato. Assim, uma nova regressão PLS foi realizada, porém os RMSE's continuaram altos, não obtendo modelos adequados para a aplicação esperada. Tentou-se ainda, fazer regressão PLS com a média das replicatas das classes de tempo, prevendo diminuir as diferenças entre as quantidades de amostras para cada classe. Já que após os dados terem sido

adicionados, algumas classes ficaram com muitas amostras fazendo com que a reta linear tivesse maior intervalo de pontos, sendo fonte de erros na regressão (Tabela 12).

Tabela 12 - Parâmetros do PLS para intervalos de tempo da média das replicatas dos dados ampliados das amostras.

<b>Intervalo de tempo</b>	<b>VL</b>	<b>RMSEC (horas)</b>	<b>RMSECV (horas)</b>	<b>CV Bias</b>	<b>R<sup>2</sup></b>
0 minutos a 10 horas	3	1,8	2,1	-0	0,5
24 horas a 15 dias	7	29,1	33,6	-0,8	0,9
24 horas a 49 dias	6	116,6	124,5	-1,1	0,9
0 minutos a 49 dias	4	164,3	171,1	0,9	0,7

Fonte: O autor (2020).

Porém, mesmo com a média, os erros RMSE's também foram elevados.

A técnica SVM foi então aplicada, uma vez que os dados possuem tendência não linear, apenas para o intervalo de 24 a 49 dias. Foram selecionados os mesmos grupos de doadores para a calibração e para a validação externa utilizados para cerâmica indicados na Tabela 7. Com isso, foram obtidos os seguintes resultados para os modelos SVM (Tabela 13).

Tabela 13- Parâmetros dos modelos SVM com diferentes combinações de dados para calibrar e validar externamente as amostras depositadas em porcelanato.

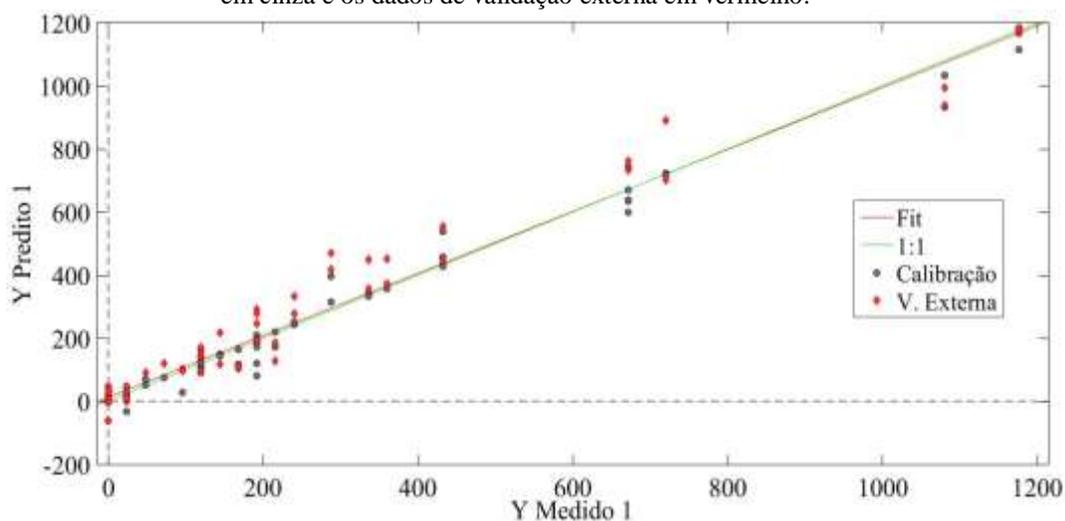
<b>Dados</b>	<b>Intervalo de tempo</b>	<b>RMSEC (horas)</b>	<b>RMSECV (horas)</b>	<b>REMSEP (horas)</b>	<b>CV Bias</b>	<b>R<sup>2</sup></b>
1	24 horas a 49 dias	28,1	59,5	93,8	-0,5	0,9
2	24 horas a 49 dias	36,7	58,2	75,1	-9,2	0,9
3	24 horas a 49 dias	39,7	58,8	65,6	-6,5	0,9
4	24 horas a 49 dias	27,8	47,6	74,3	-3,9	0,9

Fonte: O autor (2020).

No comparativo com os valores PLS para as médias das replicatas dos dados ampliados no porcelanato, no intervalo de 24 horas a 49 dias, os modelos SVM se apresentaram com menores

erros quadráticos (RMSE's) e maiores valores do coeficiente de regressão ( $R^2$ ), principalmente para os Dados 3 e Dados 4. Para avaliar o modelo SVM com os dados da validação externa por meio do gráfico de  $Y_{\text{predito}}$  vs  $Y_{\text{medido}}$ , a Figura 26 da regressão SVM para os Dados 3 foi plotada abaixo. Estes que foram analisados com 70 e 30% dos dados para calibração e validação, respectivamente.

Figura 26 - Valores preditos *versus* referência do modelo SVM para os Dados 3 no porcelanato. Dados de calibração em cinza e os dados de validação externa em vermelho.



Fonte: O autor (2020).

No comparativo com o modelo SVM para a cerâmica, os dados no porcelanato se apresentaram um pouco mais ajustados, uma vez que não houve sobreajuste na calibração. Assim como as diferenças entre os erros RMSECV e RMSEP foram menores, indicando que a validação interna e a validação externa apresentaram coerência avaliando o ajuste do modelo.

Para os Dados 3 com RMSEP de 65,6 horas, cerca de 2,7 dias e para os Dados 4 com um RMSEP de 74,3 horas, cerca de 3,1 dias, as mesmas considerações feitas para o modelo SVM na cerâmica podem ser feitas para estes modelos SVM no porcelanato. Apesar do alto erro, levando em consideração as limitações do equipamento portátil frente a este tipo de amostra, a quantidade de dias avaliadas na calibração, de 1 a 49 dias, um modelo com erro de 2,7 e 3,1 dias, pode ser levado em consideração para dar uma ideia inicial com relação ao tempo de deposição a serem elucidados em um crime, desde que seja aliado a outras análises e ao contexto da situação.

Foi observado que os dados se comportaram de forma semelhante para todas as temperaturas e para ambos os pisos (cerâmica e porcelanato). Isso se deve possivelmente à faixa

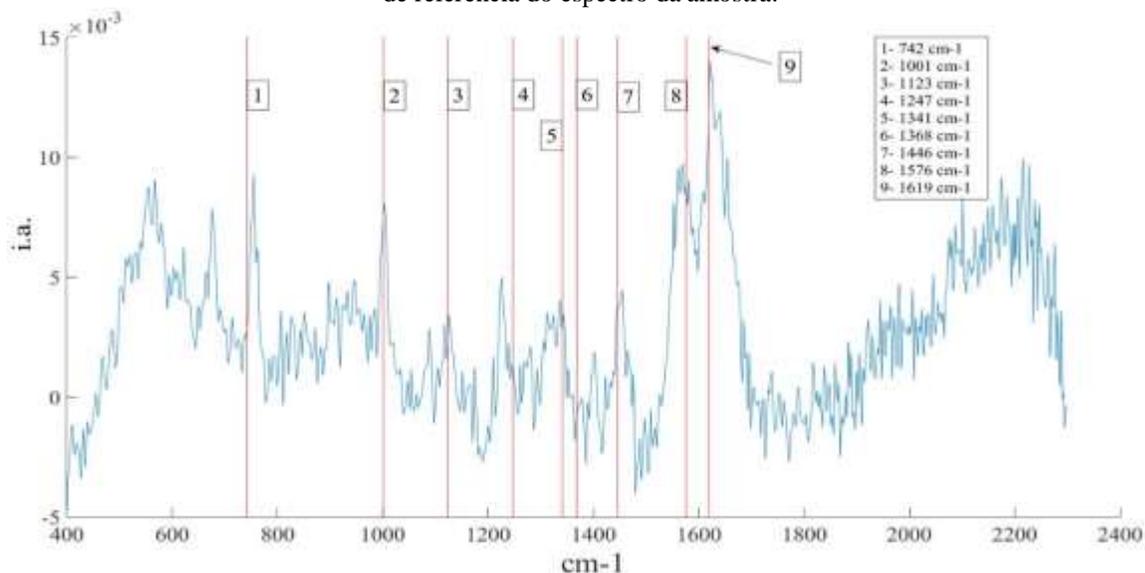
de detecção do equipamento portátil utilizado, que não permitiu analisar todas as bandas de absorção possíveis para o sangue na região do infravermelho próximo. Mesmo com 90% de hemoglobina após a secagem da mancha, a banda completa não foi detectada, deixando informações relevantes fora da análise. Apenas o início da banda foi observado, uma vez que o seu valor máximo absorve a 1690 nm, ou seja, fora da capacidade de detecção do equipamento. Apesar de ser possível observar a banda de absorção da oxi-hemoglobina a 930 nm, o equipamento entre 908-1100 nm é muito susceptível a ruído devido à proximidade com a região do visível, impedindo uma melhor análise da variação da oxi-hemoglobina com o tempo. Apesar disso, técnicas de regressão foram realizadas e foram obtidos modelos com alto RMSE's em horas, tanto para o PLS quanto para o SVM, com erros de previsão em torno de 3 dias.

Ficou evidente a influência da água nos espectros NIR e o quanto está presente no sangue inicialmente e ausente nos tempos seguintes. Com isso é possível distinguir amostras frescas (< 24 horas) e amostras antigas ( $\geq$  24 horas) pela presença de bandas de absorção da água.

#### 4.2 ESPECTROSCOPIA RAMAN

Afim de otimizar as condições do equipamento com o intuito de aplicar a melhor metodologia para obter espectros de sangue de forma clara e objetiva, foi realizado um conjunto de análises prévias com o Raman portátil em sangue diretamente depositado em substrato de alumínio, conforme os trabalhos de Kerr *et al.*, 2015 e Cui *et al.*, 2016. O espectro obtido do sangue em alumínio (média de uma triplicata), sem nenhum preparo da amostra, foi subtraído do espectro de referência do alumínio (média de uma triplicata do espectro coletado diretamente no alumínio). O espectro resultante teve então sua linha de base corrigida (Figura 27) obtendo-se picos definidos com razão sinal/ruído característico da própria técnica Raman para este tipo de amostra, como é possível observar em muitos trabalhos que usam esta técnica, como o de Boyd *et al.*, em 2011. A análise do espectro resultante foi realizada comparando-se os picos de espalhamento Raman do sangue a partir de dados da literatura de Lednev *et al.*, 2008 e Fujihara *et al.*, 2017.

Figura 27- Espectro Raman de sangue depositado no substrato de alumínio, obtido pela subtração direta do espectro de referência do espectro da amostra.



Fonte: O autor (2020).

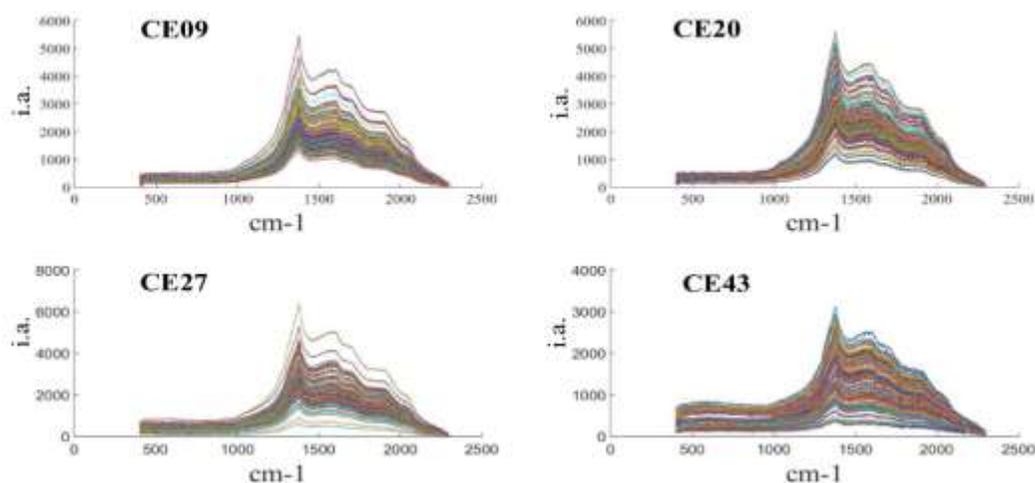
De acordo com Lednev *et al.*, 2008 e Fujihara *et al.*, 2017 os picos correspondentes a  $742\text{ cm}^{-1}$  são provenientes das bandas relativas à vibração do anel de triptofano, este que é um aminoácido obtido a partir dos alimentos, apesar do pico ter ficado um pouco deslocado a  $750\text{ cm}^{-1}$ . Já em  $1001\text{ cm}^{-1}$  há vibrações de alongamento (*breathing*) do anel aromático da fenilalanina presente na albumina; em  $1123\text{ cm}^{-1}$  o pico se apresenta de forma mais discreta, mas é proveniente do estiramento assimétrico do C-O-C da glicose. Já em  $1247\text{ cm}^{-1}$  há um pico referente ao amido III de proteína, porém este não foi observado, apesar de haver um pico a  $1238\text{ cm}^{-1}$ . Em  $1341\text{ cm}^{-1}$  há a deformação simétrica no plano do C-H do triptofano; em  $1368\text{ cm}^{-1}$  tem-se o pico referente ao estiramento simétrico  $\text{CH}_3$  do grupo Heme, porém não foi observado; a  $1446\text{ cm}^{-1}$  tem-se a deformação simétrica no plano do  $\text{CH}_2$  e do  $\text{CH}_3$  proveniente do triptofano. Os picos mais significativos a  $1576\text{ cm}^{-1}$  se referem à vibração C=C do grupo Heme e a  $1619\text{ cm}^{-1}$  à vibração C=N do grupo Heme. Desta forma, pode-se verificar que o espectro resultante é de sangue, confirmando que a técnica e os parâmetros operacionais adotados são eficientes na detecção deste tipo de amostra, assumindo assim tal espectro como o de referência para o sangue.

#### 4.2.1 Espectros Raman no substrato de cerâmica

O equipamento portátil Raman demonstrou ser de fácil manipulação e prático, na qual sua capacidade de armazenamento de espectros se mostrou vantajosa com relação ao instrumento NIR portátil, que necessita de um computador para armazenar os espectros. Contudo, a análise Raman das amostras se mostrou mais lenta, ou seja, mais demorada do que a análise por NIR.

Os espectros Raman das manchas de sangue na cerâmica apresentaram um perfil semelhante para todos os tempos e temperaturas, variando nas intensidades arbitrárias (i.a.) e nas linhas de base (Figura 28).

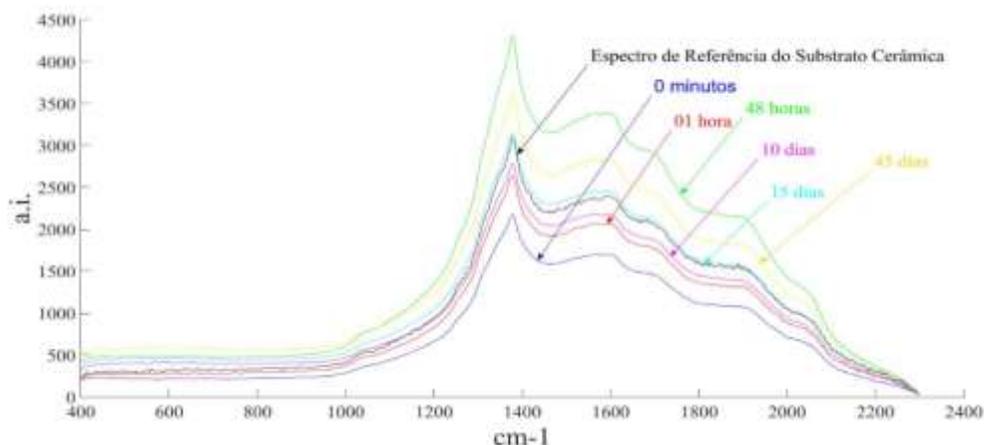
Figura 28 - Espectros Raman das manchas de sangue depositadas na cerâmica a 9°C (CE09), 20°C (CE20), 27°C (CE27) e 43°C (CE43) em todos os tempos.



Fonte: O autor (2020).

Para entender como essa variação de linha de base e intensidades se comportam com relação ao tempo, uma análise entre as médias das amostras (triplicatas) para cada tempo foi realizada (Figura 29).

Figura 29 - Média dos tempos dos espectros Raman na cerâmica.

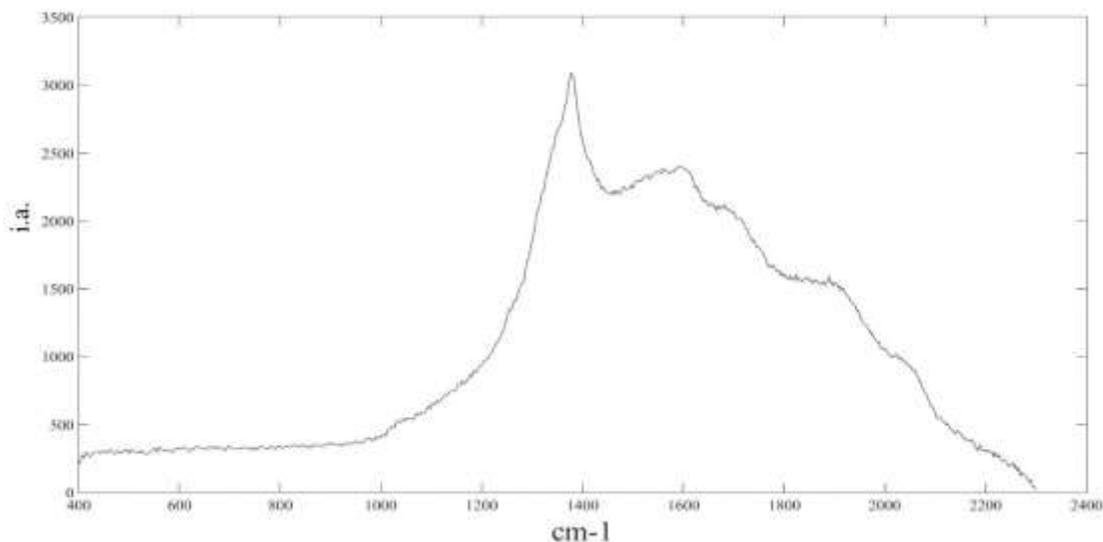


Fonte: O autor (2020).

Como o mesmo comportamento foi observado para todas as temperaturas, apenas a temperatura a  $27^{\circ}\text{C}$  foi considerada para esta discussão, aplicando os resultados obtidos às demais temperaturas. Pode-se observar, nos espectros da Figura 29, que após a deposição (0 minutos), a linha de base tende a aumentar com o tempo, atingindo seu máximo 48 horas após a deposição. A média dos espectros a 15 dias ficou próximo do espectro de referência. Já para os tempos seguintes como de 45 dias, por exemplo, as intensidades foram maiores que o do espectro de referência, porém menores que 48 horas. Logo, observa-se uma relação temporal aleatória, sem uma tendência linear ao longo do tempo. Uma vez que a maior intensidade é atingida a 48 horas e a 15 dias as amostras ficam com intensidade próximas a dos espectros da placa, ou seja, não há uma relação linear da intensidade com o tempo.

Ainda pela Figura 29, é presente em todos os espectros um pico de alta intensidade a  $1379\text{ cm}^{-1}$ , além de picos a  $1610\text{ cm}^{-1}$ ,  $1705\text{ cm}^{-1}$ ,  $1950\text{ cm}^{-1}$  e  $2050\text{ cm}^{-1}$ , além de baixa variação de intensidade de  $400$  a  $1000\text{ cm}^{-1}$  e convergência dos espectros de  $2200\text{ cm}^{-1}$  a  $2300\text{ cm}^{-1}$ . Com isso, suspeita-se que os perfis dos espectros das manchas de sangue na cerâmica obtiveram o mesmo perfil espectral da própria cerâmica de referência (Figura 30).

Figura 30 - Espectro de referência da placa de cerâmica.

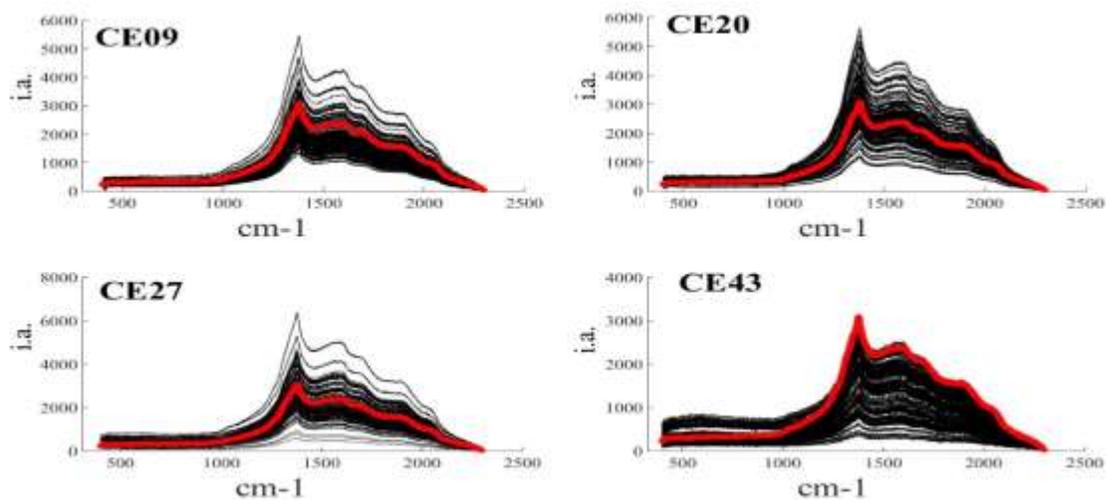


Fonte: Autor (2020)

Este espectro apresentou um pico característico a  $1379\text{ cm}^{-1}$ , sendo bem próximo do pico para o vidro a  $1382\text{ cm}^{-1}$ , relatado por Cui *et al.* em 2016 para análise Raman sob laser de 785 nm. Indicando que o material cerâmico tenha forte interação com a radiação Raman, assim como os materiais vítreos.

Para comparar melhor o espectro de referência com os espectros das amostras de sangue, um gráfico foi proposto sobrepondo o espectro de referência aos espectros dos dados (Figura 31).

Figura 31 - Espectros das amostras (em preto) na comparação com o espectro de referência (em vermelho).

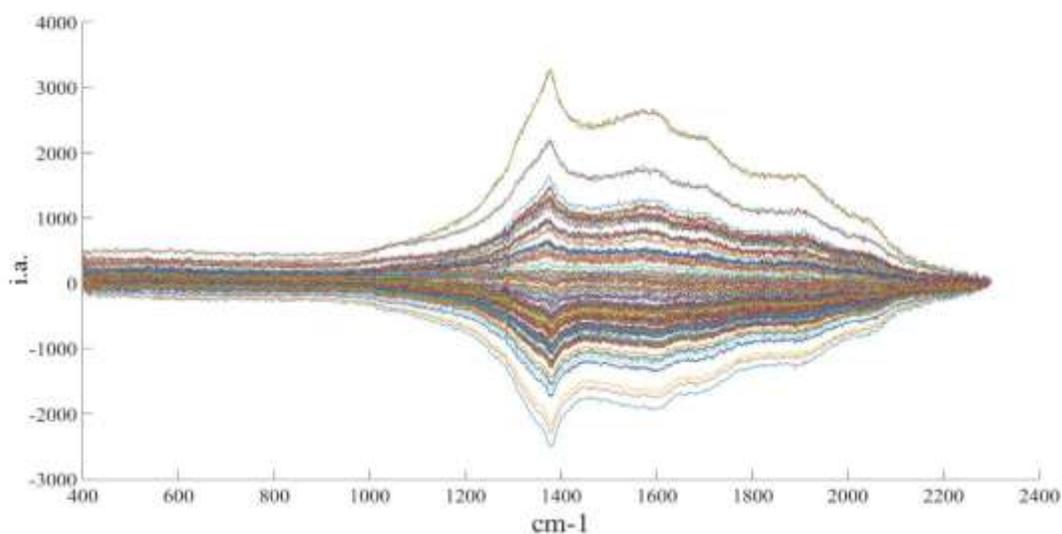


Fonte: O autor (2020)

Infere-se do gráfico anterior que no comparativo com o espectro de referência da cerâmica, pode-se observar que os espectros das amostras têm o mesmo perfil do espectro da cerâmica, confirmando que a fluorescência da placa está influenciando de forma significativa nos espectros das amostras. Logo, os espectros brutos demonstram que apenas o sinal da placa de cerâmica foi detectado pelo equipamento, assim como ocorre com o vidro em estudos citados. Justifica-se a variação da linha de base devido a variação com a qual a radiação monocromática atinge a placa, uma vez que a amostra de sangue, que não está sendo detectado ou que está tendo seu espectro sobreposto pelo da placa, apenas influencia no quanto a radiação atinge a placa, ou seja, funciona apenas como uma barreira física para a radiação. Assim, considerando as várias alterações físicas da mancha ao longo do tempo, obtém-se as variações de intensidade e linha de base ao longo do tempo observada para os espectros.

Apesar dessas considerações, tentou-se contornar a forte influência da fluorescência nas análises fazendo a remoção do espectro do substrato para obter espectros puros de sangue. Desse modo utilizou-se a mesma metodologia que foi determinada com o sangue depositado no alumínio, ou seja, subtração direta do espectro de referência do substrato, sendo este com intensidade de 3.000 i.a. Logo, obteve-se espectros resultantes que apresentaram valores positivos, centrados no zero e negativos, como pode-se observar no perfil do espectro a 27°C, na figura abaixo (Figura 32).

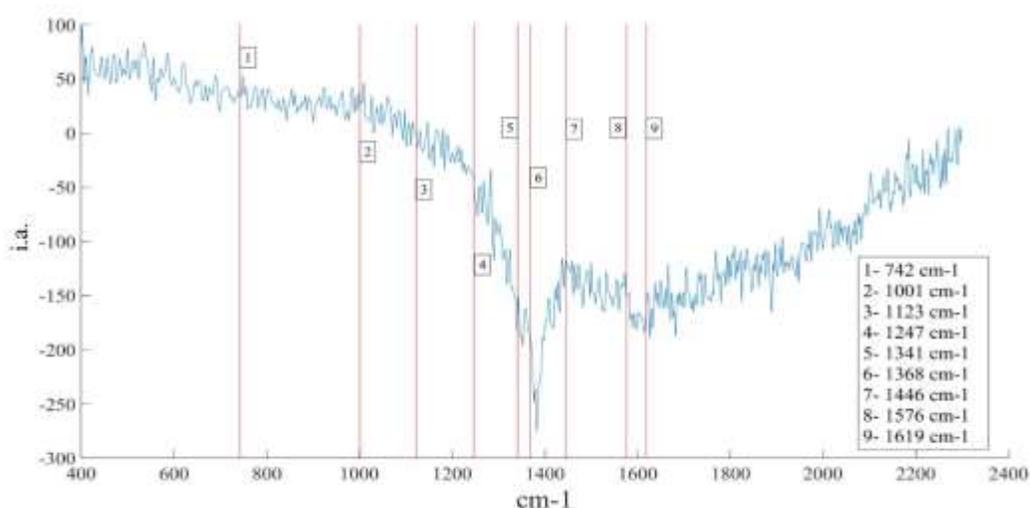
Figura 32 - Espectro resultante após a subtração do espectro de referência do substrato ao espectro da amostra a 27°C.



Fonte: O autor (2020)

Na Figura 32 pode-se observar que em torno de  $1379\text{ cm}^{-1}$  há um pico de alta intensidade característico da cerâmica que se apresenta em todos os espectros obtidos. Além de ficar claro que não há interação da radiação com a amostra de  $400\text{ a }1000\text{ cm}^{-1}$ . No intuito de tentar determinar picos de sangue no espectro resultante, foi obtido a média deste espectro resultante para melhor identificar os picos (Figura 33).

Figura 33 - Média do espectro resultante da subtração dos dados brutos com o espectro de referência para  $27^{\circ}\text{C}$ .

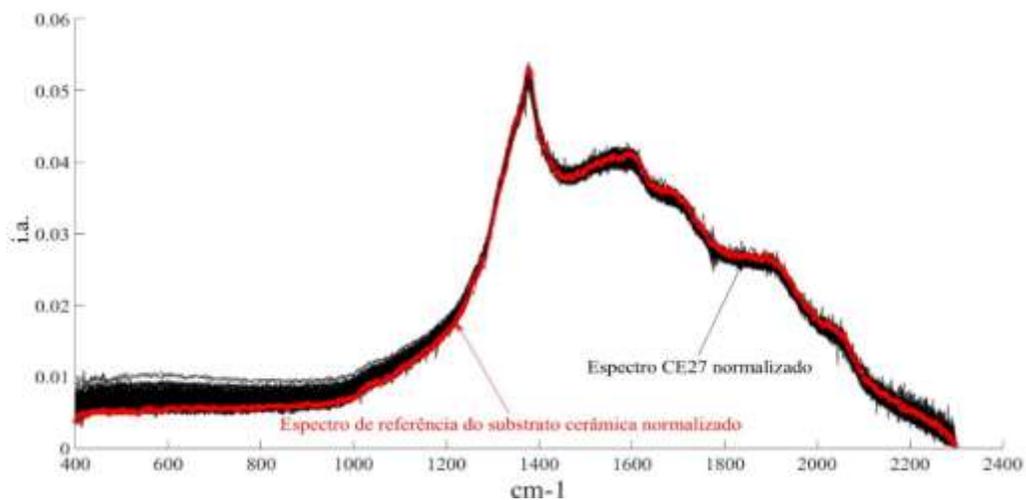


Fonte: O autor (2020).

Apesar de facilitar a identificação de picos na média do espectro resultante, não foi possível determinar a presença de sangue após a remoção da informação do espectro da cerâmica. Apenas foi verificada a presença de ruídos espectrais e nenhum pico significativo, no comparativo com os picos obtidos da literatura.

Para dirimir os efeitos das diferenças de intensidades após a subtração, os espectros foram previamente normalizados pela faixa, tanto das amostras (em preto) como do espectro de referência (em vermelho), conforme na Figura 34.

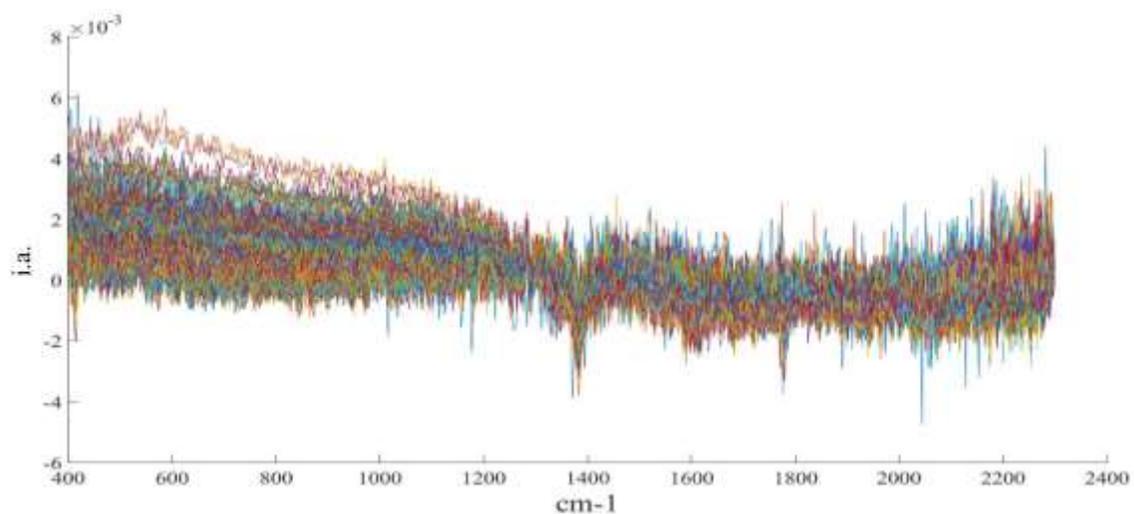
Figura 34 - Comparação entre os espectros das amostras (em preto) e do espectro de referência (em vermelho) após a normalização pela faixa.



Fonte: O autor (2020).

Da Figura 34, infere-se que as diferenças de intensidade foram dirimidas com a aplicação da normalização. Há praticamente a sobreposição dos espectros normalizados da cerâmica (em vermelho) com os espectros do conjunto de dados. Desta forma fez-se a subtração direta (Figura 35).

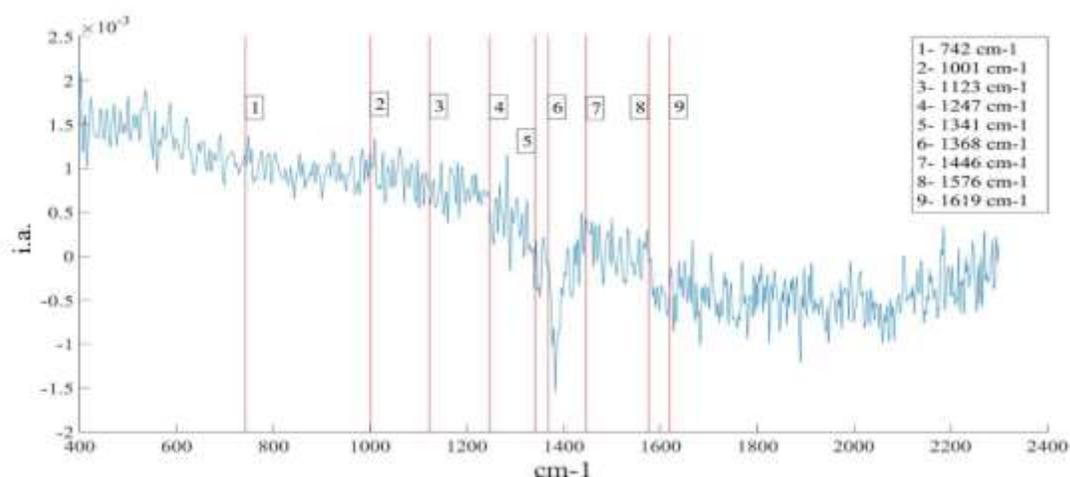
Figura 35 - Espectros resultantes da subtração dos espectros de referência com espectros das amostras previamente normalizados.



Fonte: O autor (2020).

Os espectros resultantes da subtração (Figura 35) apresentaram baixa intensidade, em torno de 0,001 i.a., devido à grande semelhança de intensidade após normalização dos espectros. Além de um sinal com bastante ruído, não sendo possível identificar os picos das interações moleculares que caracterizam o sangue. Obtendo-se o valor médio desses espectros (Figura 36) pode-se perceber ruídos espectrais originários da própria técnica e da subtração de espectros com intensidades muito próximas.

Figura 36 - Média dos espectros resultantes da subtração dos espectros de referência com espectros das amostras previamente normalizados.

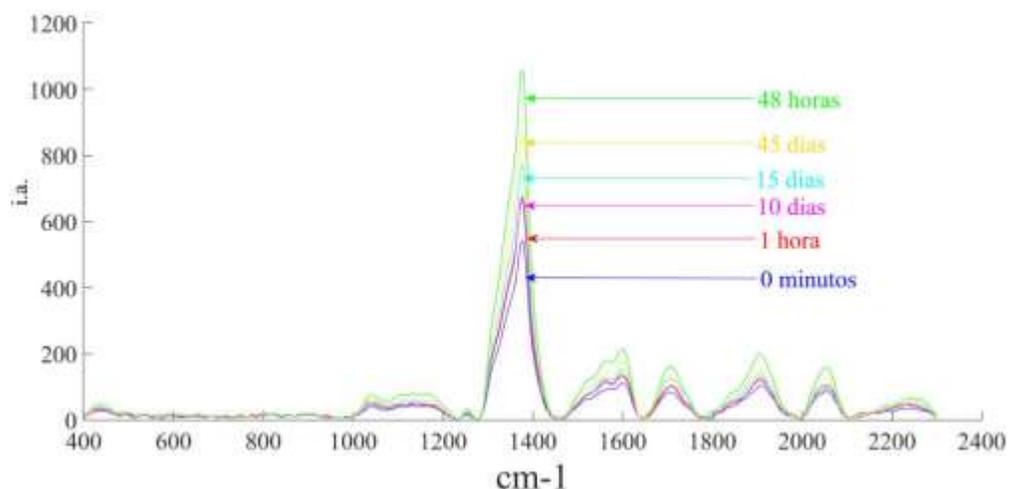


Fonte: O autor (2020).

O espectro resultante na Figura 36 também apresenta baixa intensidade e um pico negativo em torno de  $1379\text{ cm}^{-1}$  referente a subtração do espectro de referência da cerâmica que tem o pico positivo característico a  $1379\text{ cm}^{-1}$  inviabilizando a identificação de sangue, uma vez que este praticamente não foi detectado. Apesar de sobrepor as linhas referentes as posições dos nove picos característicos do sangue relatados na literatura, observa-se que o ruído é mais presente do que picos característicos.

Um outro pré-processamento foi adotado para tentar dirimir os efeitos da fluorescência das placas de cerâmica e na diferença de intensidades entres os espectros, fazendo correção de linha de base a fim de identificar os picos característicos do espalhamento das ligações moleculares dos componentes do sangue. O AsLS foi aplicado separadamente nos espectros das amostras e no espectro de referência da placa antes da subtração. Uma análise temporal foi realizada comparando-se as médias dos espectros obtidos após aplicação da correção de linha de base (Figura 37).

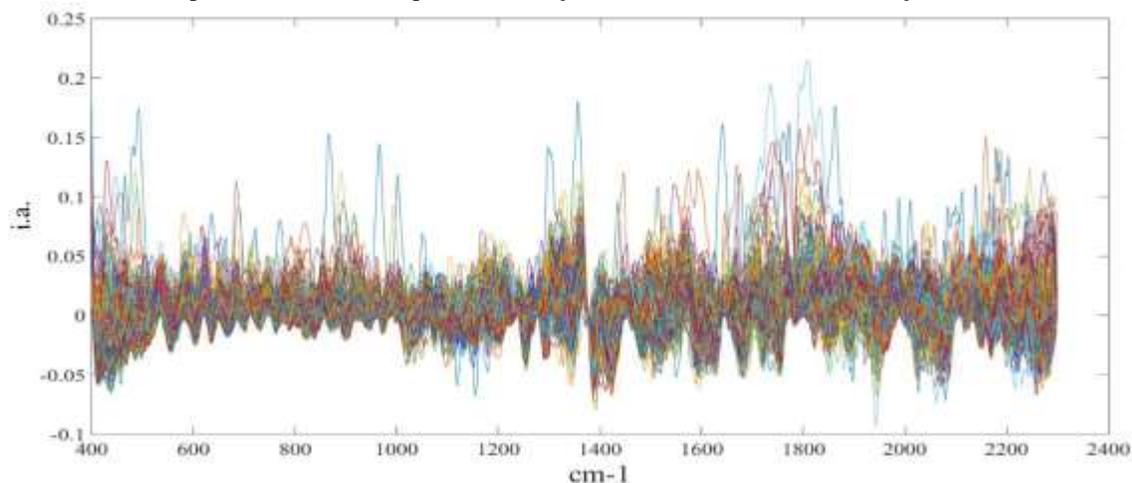
Figura 37 - Média dos espectros Raman na cerâmica com aplicação do AsLS para diferentes tempos.



Fonte: O autor (2020).

Após aplicação do AsLS, observou-se que houve uma intensificação nos picos observados nos espectros brutos, tornando a melhor visualização destes. Porém, com relação a variação de intensidade dos espectros médios para cada tempo, estes mantiveram a mesma relação de intensidades arbitrarias com relação ao tempo no comparativo com os espectros brutos, ou seja, não houve alteração de intensidade entre os tempos. Como há uma alta variabilidade de intensidades nos espectros brutos foi realizada uma normalização para que as intensidades fiquem na mesma dimensão, tornando possível uma subtração do espectro de referência, obtendo-se então o espectro resultante (Figura 38).

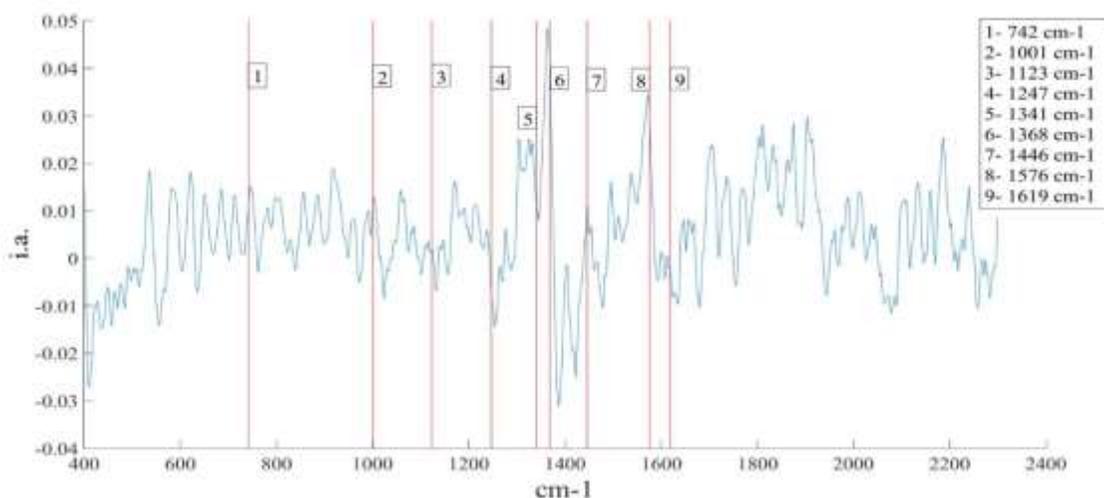
Figura 38 - Espectros da amostra a 27°C após aplicação do AsLS da normalização e após a subtração dos dados com o espectro de referência que teve correção de linha de base e normalização.



Fonte: O autor (2020).

Diretamente observando o espectro resultante da Figura 38, percebe-se é inconclusivo obter alguma informação relevante a partir desta. Devido a isso, foi obtido o espectro resultante médio (Figura 39) onde pode-se observar mais facilmente os picos, apesar do ruído e da baixa intensidade. Os picos de espalhamento identificados por Lednev *et al.* (2008) com suas respectivas ligações moleculares foram comparados ao espectro resultante para analisar se havia a presença dos picos característicos do sangue.

Figura 39 - Média do espectro resultante, com os picos característicos de ligações moleculares correspondentes a componentes do sangue.



Fonte: O autor (2020)

Alguns picos do sangue determinados pela literatura se encontram presentes na média do espectro resultante, estes que podem ser vistos na figura anterior sobrepostos pelas linhas em vermelho. Apesar de haver uma correspondência com muitos picos, a maioria se confunde com ruído, sendo de difícil identificação. Apenas os picos 6 e 8 apresentaram intensidades relativamente altas quando comparadas com o ruído, porém estes são muito sutis (baixa i.a.), não sendo útil para afirmar que o sangue fora detectado.

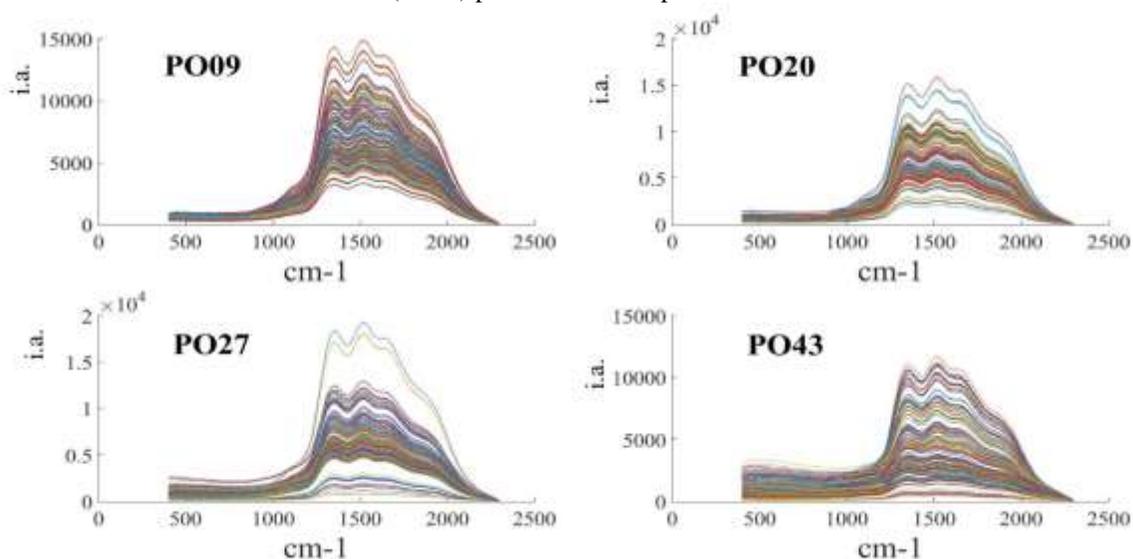
Parar tentar avaliar como este espectro resultante se comporta numa análise exploratória de dados, foi realizado uma PCA com os espectros pré-processados pelo AsLS truncados de 1300 a 1600  $\text{cm}^{-1}$ , ou seja, englobando apenas os picos 6 e 8 que tiveram maior destaque. Contudo, nenhuma relação temporal foi observada nos gráficos dos escores e dos pesos. Uma calibração também foi realizada e foram obtidos erros RMSE's altos e valores de  $R^2$  praticamente zero. Logo,

mesmo que seja sinal de sangue, não se pode fazer uma análise temporal, pois o espectro resultante é ruidoso e de baixa intensidade.

#### 4.2.2 Espectros Raman no substrato de porcelanato

Um comportamento parecido com os espectros obtidos nas placas de cerâmica foi observado nas placas de porcelanato. Um perfil similar em todos os tempos e em diferentes temperaturas, apresentando bandas características em torno de  $1352\text{ cm}^{-1}$ ,  $1520\text{ cm}^{-1}$  e  $1670\text{ cm}^{-1}$ , além do efeito de espalhamento e alta variação de linha de base. Os espectros apresentaram altas intensidades arbitrárias no comparativo com as placas de cerâmica, na faixa de 160.000 i.a. para  $9^{\circ}\text{C}$ ,  $20^{\circ}\text{C}$  e  $27^{\circ}\text{C}$  e 120.000 para  $43^{\circ}\text{C}$  (Figura 40).

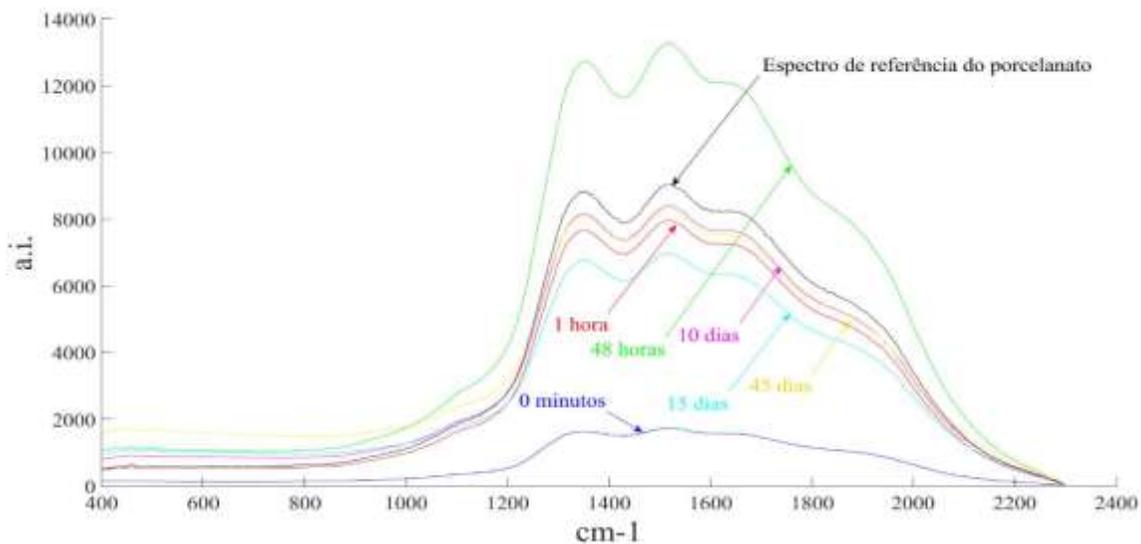
Figura 40 - Amostras de sangue depositadas em porcelanato a  $9^{\circ}\text{C}$  (PO09),  $20^{\circ}\text{C}$  (PO20),  $27^{\circ}\text{C}$  (PO27) e  $43^{\circ}\text{C}$  (PO43) para todos os tempos.



Fonte: O autor (2020).

Da mesma forma realizada para a cerâmica, uma análise das médias para cada tempo foi realizada, considerando apenas a temperatura ambiente ( $27^{\circ}\text{C}$ ) (Figura 41).

Figura 41 - Média dos tempos dos espectros Raman no porcelanato.



Fonte: O autor (2020).

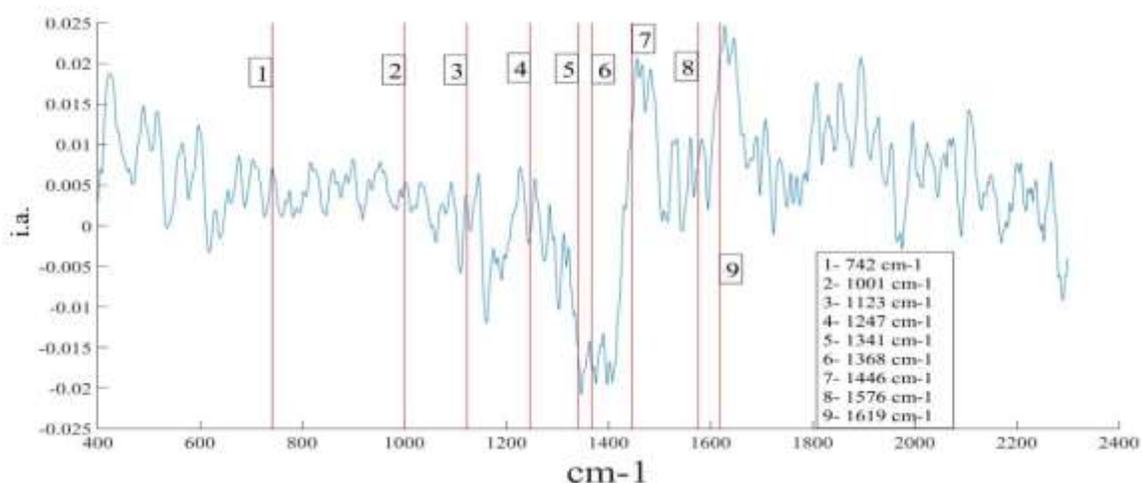
Uma ordem diferente para as intensidades das médias foi observada no comparativo com o da cerâmica. Algumas intensidades se apresentaram maiores ou menores, por exemplo, a deposição a 10 dias possui maior intensidade do que a de 15 e 45 dias no porcelanato. Em ambas as placas (cerâmica e porcelanato) foi possível perceber que a intensidade para a média dos espectros em 0 minutos após a deposição é a de menor i.a., indicando que assim que há a deposição, o sangue por ser uma suspensão coloidal atua como barreira física na interação do laser com a placa. Já a 48 horas após a deposição, pode-se perceber que ocorre a maior i.a. entre as médias dos espectros. Indicando que após 48 horas o laser tem maior interação com a placa, podendo haver alguma intensificação por parte da amostra, uma vez que após 48 horas a i.a. tende a decair.

Assim como na cerâmica, todos os espectros obtidos da amostra de sangue possuem o mesmo perfil do espectro de referência da placa de porcelanato, indicando que este substrato também apresenta alta fluorescência do material frente ao espalhamento Raman. Com isso, influenciando de forma significativa a análise da mancha em si, uma vez que os espectros obtidos basicamente são os da placa de porcelanato. Logo, a amostra apenas atua como barreira física, tanto atenuando como intensificando a interação do laser com o piso. Alguns trabalhos como de Kerr *et al.* (2015), relataram a alta influência nos espectros de materiais vítreos comuns, ou seja, aqueles de laminas de vidros para laboratório. Contudo, nenhum trabalho até o momento relacionou pisos com a espectroscopia Raman.

Na tentativa de remover a influência do substrato nas amostras, a mesma metodologia aplicada aos dados na cerâmica foi feita para o porcelanato (subtração direta e subtração após a normalização). Os mesmos resultados obtidos para as médias dos espectros resultantes na cerâmica foram observados para o porcelanato: sem picos definidos; alto ruído espectral; e baixa intensidade arbitrária.

Considerou-se aplicar o AsLS, tendo o espectro resultante analisado obtendo-se a média e o comparando com os comprimentos de onda referentes as bandas moleculares provenientes da literatura (Figura 42).

Figura 42 - Média do espectro resultante após a subtração dos espectros pré-processados por AsLS.



Fonte: O autor (2020).

Alguns picos se apresentaram no espectro médio. Os mais significativos foram a  $1446\text{ cm}^{-1}$  (7), referente a deformação simétrica no plano do  $\text{CH}_2$  e do  $\text{CH}_3$  proveniente do triptofano e a  $1619\text{ cm}^{-1}$  (9), referente à vibração  $\text{C}=\text{N}$  do grupo Heme. Sugerindo que pode ter detectado sangue, porém com baixa intensidade.

Assim como para a cerâmica, foi realizado uma PCA e uma calibração com os espectros pré-processados pelo AsLS truncados de  $1400$  a  $1700\text{ cm}^{-1}$ , ou seja, englobando os picos 7 a 9. Contudo, nenhuma relação temporal foi observada nos gráficos dos escores e dos pesos, ou seja, as mesmas observações foram obtidas. Para a calibração foram obtidos  $R^2$  próximos a zero. Logo, mesmo que haja a presença de algum sinal de sangue, não se pode fazer uma análise temporal, pois o espectro resultante é ruidoso e de baixa intensidade.

Tanto para a cerâmica como para o porcelanato os espectros de sangue ficaram sobrepostos aos espectros dos substratos devido à alta fluorescência. Após tentativas de obter o espectro do sangue em si, por subtração direta, subtração dos dados normalizados e subtração por AsLS, apenas neste último conseguiu-se perceber alguns picos referentes ao sangue, porém bastante sutis. Logo, foi observado o que fora relatado em estudos anteriores, como de Kerr *et al.*, em 2016, que tiveram uma grande dificuldade em remover a informação do substrato. Além desse fato comprometer o uso de pré-processamentos para extrair as informações do analito. Logo, nem todos os picos foram observados para o espectro médio resultante do AsLS (nem para o sangue em alumínio isto ocorreu), mas a baixa intensidade e o alto ruído espectral impedem que o estudo prossiga em termos de datação de mancha de sangue para este conjunto de dados. Considera-se que o tipo de equipamento, mais especificamente o laser no comprimento de onda de 785 nm; o tipo de amostra, pelo fato de ser amostra biológica que tem baixa intensidade frente a espectroscopia Raman; e o substrato, com alto espalhamento da radiação Raman, dificultaram sobremaneira a análise.

Grande parte dos trabalhos que utilizaram espectroscopia Raman em amostras de sangue utilizaram equipamentos robustos de bancada, com amostras preparadas em laboratório, laser de menor comprimento de onda (em torno de 534 nm) para evitar fluorescência. Além disso, muitos fizeram datação de mancha de sangue sob placas de vidro com sobreposição de alumínio, que intensifica o sinal da amostra na superfície. Outros estudos sugerem que a espectroscopia Raman em substratos fluorescentes deve ser utilizada em últimas circunstâncias. Ainda sugerem técnicas para contornar a fluorescência do substrato como coletar o sangue com *swab* para analisar diretamente sob o algodão, até mesmo coletar e analisar o sangue em suportes vítreos de alto custo, feitos de compostos que não interagem com a radiação monocromática (KERR *et al.*, 2016). Considerando que um dos objetivos deste trabalho era avaliar a capacidade de analisar de forma direta, sem preparo de amostras, com rapidez e baixo custo diretamente no piso, usar estratégias para driblar as condições diretas de análise que não seja por pré-processamentos estão fora de escopo.

## 5 CONCLUSÃO

Para as duas técnicas espectroscópicas utilizadas, os espectros, assim como os resultados apresentaram comportamentos bem distintos. Com relação a espectroscopia NIR o substrato não influenciou de forma significativa os espectros das amostras, por outro lado a informação da absorção da água presente na amostra foi dominante. O equipamento conseguiu detectar bandas de absorção da água presente no plasma e partes das bandas atribuídas a moléculas de hemoglobina. Devido a isso ficou evidente as amostras com altas e baixas concentrações de água ao longo do tempo por meio da análise exploratória de dados. Logo, a técnica NIR se mostrou adequada para avaliar qualitativamente a discriminação entre manchas recentes e antigas, sendo dependente da temperatura. Os modelos PLS não se mostraram satisfatórios para o NIR devido à alta complexidade bioquímica da degradação do sangue, não tendo comportamento linear e assim ficando a regressão com altos RMSE's. Já os modelos SVM para datação das manchas apresentaram RMSEP de cerca de 3 dias, valores muito próximos aos encontrados por trabalhos relatados na literatura utilizando equipamentos de bancada, preparo da amostra e substratos não interferentes. Com isso, devido as circunstâncias e a quantidade de dias ao qual o modelo foi obtido (de 1 a 49 dias) aliado a outras evidências relacionadas ao delito, pode auxiliar em uma estimativa do tempo de deposição da mancha.

Na espectroscopia Raman, como o sangue nos pisos não foi detectado, um estudo com relação ao tempo e analisando as diferentes temperaturas não puderam ser realizadas. Observou-se que a variação nas linhas de bases era proveniente, principalmente, da interação da radiação com os substratos, com o sangue funcionando como barreira física para a radiação. Assim, com as mudanças físico-químicas do sangue a variação da linha de base ocorria, porém sem relação lógica com o tempo. Contudo, o sangue em alumínio foi facilmente detectado, pois o equipamento se mostrou bastante versátil, com uma faixa espectral que detecta informação do sangue. Logo, o equipamento empregando a técnica Raman se mostrou apto à detecção de sangue, porém utilizando substrato que não tenha contribuição na análise. Devido a isso, em uma situação real com sangue depositado em substrato de alumínio pode-se inferir que a confirmação de que aquela mancha é sangue humano ocorreria de forma rápida, sem consumo e preparo da amostra aplicando a metodologia aqui utilizada.

## **6 PERSPECTIVAS PARA TRABALHOS FUTUROS**

Para a espectroscopia Raman, fazer estudos aplicando a mesma metodologia, mesmo equipamento, porém utilizando um substrato que não cause tanta influência nos espectros das amostras de sangue, para que a análise temporal possa ser realizada com a variação da temperatura visando novamente fins forenses de datação da mancha de sangue.

Para a espectroscopia de absorção do infravermelho, fazer um estudo temporal utilizando um equipamento com uma faixa espectral maior, para que informações na região das combinações (acima de 2.000 nm) possam ser detectadas e assim conseguir fazer um estudo do tempo de deposição da mancha de sangue variando a temperatura.

Para ambas as técnicas introduzir mais amostras em tempos diferentes, com diferentes doadores em termos de sexo, idade, raça, hábitos, entre outros para dar mais robustez aos modelos, seguindo a tendência das pesquisas atuais.

## REFERÊNCIAS

- AGUDELO, J.; HALÁMKOVÁ, L.; BRUNELLE, E.; RODRIGUES, R.; HUYNH, C.; HALÁMEK, J.; Ages at a Crime Scene: Simultaneous Estimation of the Time since Deposition and Age of Its Originator. **Anal. Chem.** **88**, 6479-6484, 2016.
- ALLAIN, C.; LIMAT, L.; Regular patterns of cracks formed by directional drying of a colloidal suspension, **Phys. Rev. Lett.** **74**, 2981-2984, 1995.
- BALTHAZARD, V.; PIEDELIEVRE, R.; DESOILLE, H.; DEROBERT, L.; Etude des gouttes de sang projete, **XXIIe congresso de medicina legal da língua francesa, Paris**, 1939.
- BEEBE, K.; PELL, R.; SEASHOLTZ, M. B.; **Chemometrics: A Pratical Guide**. 1998.
- BERG, J. M.; TYMOCZKO, J. L.; STRYER, L.; **Biochemistry. 5º Edição**, W. H. Freeman and company. 2002.
- BEVEL, T.; GARDNER, R. M. Bloodstain Pattern Analysis: With an Introduction to Crime Scene Reconstruction, **CRC Series in Practical Aspects of Criminal and Forensic Investigations**, 2002.
- BOYD, S.; BERTINO, M.; SEASHOLS, S.; Raman spectroscopy of blood samples for forensic applications, **Forensic Science International** **208**, 124-128, 2011.
- BREMMER, R. H.; BRUIN, K. G.; GEMERT, M. J. C.; LEEUWEN, T. G.; AALDERS, M. C. G.; Forensic quest for age determination of bloodstains. **Forensic Science International** **216**, 1–11, 2012.
- BREMMER, R. H.; NADORT, A.; LEEUWEN, T. G.; MARTIN, J.C.; AALDERS, M.C.G.; Age estimation of blood stains by hemoglobin derivative determination using reflectance spectroscopy. **Forensic Science International** **206**, 166–171, 2011.
- BREMMER, R. H.; DE BRUIN, D. M.; DE JOODE, M.; BUMA, W. J.; VAN LEEUWEN, T. G., ET AL. Biphasic Oxidation of Oxy-Hemoglobin in Bloodstains. **PLOS ONE**, 2011.
- BRERETON, R. G; LLOYD, G. R.; Support Vector Machines for classification and regression. **The Royal Society of Chemistry**. 2010.
- BRERETON, R. G.; **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. 2003.
- BRO, R.; SMILDE, A. K.; Centering and Scaling in Component Analysis. **Journal of Chemometrics**. 16-33. 2003.

BRODECK, S.; Introduction to bloodstain pattern analysis. **Journal for Police Science and Practice** **2**, 51-57, 2012.

CIALLA-MAY, D.; SCHMITT, M.; POPP, J. Theoretical principles of Raman spectroscopy, **Physical Sciences Reviews**. 2019.

CUI, L.; BUTLER, J. B.; MARTIN-HIRSCHB, P. L ; MARTIN., F. L. Aluminium foil as a potential substrate for ATRFTIR, transfection FTIR or Raman spectrochemical analysis of biological specimens. **Anal. Methods** **8**, 481-487, 2016.

DA SILVA, V. H.; DA SILVA, J. J.; PEREIRA, C. F; Portable near-infrared instruments: Application for quality control of polymorphs in pharmaceutical raw materials and calibration transfer. **Journal of Pharmaceutical and Biomedical Analysis** **134**, 287-294, 2016.

DOTY, K. C.; MURO, K. C.; LEDNEV, I. K.; Predicting the time of the crime: Bloodstain aging estimation for up to two years. **Forensic Chemistry** **5**, 1–7, 2017.

DOTY, K. C.; MCLAUGHLIN, G.; LEDNEV, I. K.; A Raman “Spectroscopic clock” for bloodstain age determination: the first week after deposition. **Anal Bioanal Chem** **408**, 3993-4001, 2016.

DOTY, K. C.; LEDNEV, I. K.; Differentiation of human blood from animal blood using Raman spectroscopy: A survey of forensically relevant species, **Forensic Science International** **282**, 204-210, 2018.

DEFERNEZ, M.; KEMSLEY, E. K.; The use and misuse of chemometrics for treating classification problems. **Trends in analytical chemistry** **16**, 216-221, 1997.

EDELMAN, G.; MANTI, V.; RUTH, S. M.; LEEUWEN, T.; AALDERS, M.; Identification and age estimation of bloodstains on colored backgrounds by near infrared spectroscopy. **Forensic Science International** **220**, 239–244, 2012.

EDELMAN, G. J.; ROOS, M.; BOLCK, A.; AALDERS, M. C.; **Practical Implementation of Blood Stain Age Estimation Using Spectroscopy**, 2016.

EILERS, P. H. C.; BOELEN, H. F. M. Baseline Correction with Asymmetric Least Squares Smoothing, 2005.

FARIA, D. L. A.; AFONSO, M. C.; EDWARDS, H. G. M. Espectroscopia Raman: uma nova luz no estudo de bens culturais. **Rev. do Museu de Arqueologia e Etnologia**, 12: 249-267, 2002.

FUJIHARA, J.; FUJITA, Y.; YAMAMOTO, T.; NISHIMOTO, N.; KIMURA-KATAOKA, K.; KURATA, S.; TAKINAMI, Y.; T. YASUDA, T.; TAKESHITA, H. Blood identification and discrimination between human and nonhuman blood using portable Raman spectroscopy, **Int J Legal Med** **131**, 319-322, 2017.

FUJIHARA, J.; NISHIMOTO, N.; YASUDA, T.; TAKESHITA, H.; Discrimination Between Infant and Adult Bloodstains Using Micro-Raman Spectroscopy: A Preliminary Study. **J Forensic Sci** **64**, 698-701, 2019.

GASTEIGER, J.; ENGEL, T. Chemoinformatics. **Weinheim: WILEY-VCH Verlag GmbH & Co**, 2003.

GAUTAM, R.; VANGA, S.; ARIESE, F.; UMAPATHY, S.; Review of multidimensional data processing approaches for Raman and infrared spectroscopy. **EPJ Techniques and Instrumentation**, 2:8, 2015.

GEMPERLINE, P.; Practical Guide to Chemometrics. **CRC Press Taylor & Francis Group**, 2006.

HAMDI, T.; ALI, J. A.; CONSTANZO, V.; FNAIECH, F.; MOREAU, E.; GINOUX, J. Accurate prediction of continuous blood glucose based on support vector regression and differential evolution algorithm. 2017.

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**, e12830, 2010.

HESPANHOL, M. C.; PASQUINI, C.; MALDANER, A. O. Evaluation of a low-cost portable near-infrared spectrophotometer for in situ cocaine profiling, **Talanta** **200**, 2019.

HODGES, C. M.; AKHAVAN, J.; The use of Fourier transform Raman spectroscopy in the forensic identification of illicit drugs and explosives, **J. Mol. Spectrosc.** **46**, 303–307, 1990.

HOLLER, F. J.; SKOOG, D. A.; CROUCH, S. R. Princípios de Análise Instrumental. Tradução de Celio Pasquini, Jarbas José Rodrigues Rohwedder, *et al.* 6. ed. **Porto Alegre: Bookman**, 2009.

IZAKE, E. L. Forensic and homeland security applications of modern portable Raman spectroscopy, **Forensic Science International**, 2010.

JACKSON, M.; MANTSCH, H.H.; Pathology by infrared and Raman spectroscopy. **Handbook of Vibrational Spectroscopy**, vol. 4, pp. 5-6, 2001.

KERR, L. T.; BYRNE, H. J.; HENNELLY, B. M. Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen. **Anal. Methods**, 2015.

KERR, L. T.; HENNELLY, B. M. **An investigation of background subtraction algorithms for Raman spectra of cytology samples recorded on glass slides**, 2016.

KUMAR, N.; BANSAL, A.; SARMA, G.S.; RAWAL, R. K.; Chemometrics tools used in analytical chemistry: An overview. **Talanta** **123**, 186–199, 2014.

KOURKOUMELIS, N. A. Polymeros and M. Tzaphlidou, **Spectrosc. Int. J.**, 2012.

- 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. **Science and Justice** **53**, 270–277, 2013.
- LIN, H.; ZHANG, Y.; WANG, Q.; LI, B.; HUANG, P.; WANG, Z.; Estimation of the age of human bloodstains under the simulated indoor and outdoor crime scene conditions by ATR-FTIR spectroscopy. **Nature, Scientific Reports**, 2017.
- LIMA, A.; BAKKER, J.; Espectroscopia no infravermelho próximo para a monitorização da perfusão tecidual. **Revista Brasileira Terapia Intensiva**, **23(3)**:341-351, 2011.
- MARRONE, A.; BALLANTYNE, J.; Changes in Dry State Hemoglobin over Time Do Not Increase the Potential for Oxidative DNA Damage in Dried Blood. **PLoS ONE** **4**, 2009.
- Manual do Produto. Metrohm Instant Raman Analyzers (Mira). 2018.
- MACIEL, D. R. **Análise do padrão de manchas de sangue em local de crime: revisão de literatura**. 2014
- MISTEK, E.; HALAMKOVA, L.; DOTY, K. C.; MURO, K.; LEDNEV, I. K.; Race Differentiation by Raman Spectroscopy of a Bloodstain for Forensic Purposes. **Anal. Chem.** **88**, 7453-7456, 2016.
- MORILLAS, A. V.; GOOCH, J.; FRASCIONE, N.; Feasibility of a handheld near infrared device for the qualitative analysis of bloodstains. **Talanta** **184**, 1–6, 2018.
- MURO, K. C.; DOTY, C. K.; SOUZA L.; K. LEDNEV, I.; Forensic body fluid identification and differentiation by Raman spectroscopy. **Forensic Chemistry** **1**, 31–38, 2016.
- NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. A user-friendly guide to multivariate calibration and classification, **Chichester: NIR Publications**, 2002.
- PAVIA, D. L.; LAMPMAN, G. M.; KRIZ, G. S.; VYVYAN, J. R.; Introdução à espectroscopia. **Tradução da 4ª edição norte-americana**, 2010.
- PEREIRA, J. F. Q.; SILVA, C. S.; VIEIRA, M. J. L.; PIMENTEL, M. F.; BRAZ, A.; HONORATO, R. S.; Evaluation and identification of blood stains with handheld NIR spectrometer. **Microchemical Journal** **133**, 561–566, 2017.
- PENG, J.; PENG, S.; JIANGA, A.; WEI, J.; LI, C.; TANA, J. Asymmetric least squares for multiple spectra baseline correction. **Analytica Chimica Acta**, 2010.
- PIOTROWSKI, E.; **Concerning the Origin, Shape, Direction and Distribution of the Bloodstains Following Head Wounds Caused by Blows**, 1895.

RAMSTHALER, F.; SCHLOTE, J.; WAGNER, C.; FISCINA, J.; KETTNER, M.; The ring phenomenon of diluted blood droplets. **Int J Legal Med**, **130**:731–736, 2016.

RINNAN, A.; VAN DEN BERG, F.; ENGELSEN, S. B.; Review of the most common pre-processing techniques for near-infrared spectra, **Trends in Analytical Chemistry**, **Vol. 28**, No. 10, 2009.

SANTANA; **Quimiometria com R**, 2013.

SHARMA, V.; KUMAR, R.; Trends of chemometrics in bloodstain investigations. **Trends in Analytical Chemistry** **107**, 181e195, 2018.

SHIKAMA, K. The Molecular Mechanism of Autoxidation for Myoglobin and Hemoglobin: A Venerable Puzzle. **Chemical Reviews**, 1998.

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. **Photonics Journal**, 2017.

SUZUKI, E. M.; CARRABBA, M.; In situ identification and analysis of automotive paint pigments using line segment excitation Raman spectroscopy: I. Inorganic topcoat pigments, **J. Forensic Sci.** **46**, 1053–1069, 2001.

SKOOG, D.; WEST, D.; HOLLER, J.; CROUCH, S.; Fundamentos de Química Analítica. **Cengage Learning; Edição: 1ª**, 2014.

SCHWARZACHER; Determination of the age of bloodstains. **American Journal of Police Sciences** **1**, 1930.

SIKIRZHYTSKAYA, A.; SIKIRZHYTSKI, V.; LEDNEV, I. K. Determining Gender by Raman Spectroscopy of a Bloodstain, **Anal. Chem.**, 2017.

THOMAS, J.; BUZZINI, P.; MASSONNET, G.; REEDY, B.; ROUX, C.; Raman spectroscopy and the forensic analysis of black/grey and blue cotton fibres part 1: investigation of the effects of varying laser wavelength, **Forensic Sci. Int.** **152**, 189–197, 2005.

UMBREIT, J. Methemoglobin - It's Not Just Blue: A Concise Review. **American Journal of Hematology** **82**:134–144, 2007.

VIRKLER, K.; LEDNEV, I.K.; Raman spectroscopy offers great potential for the nondestructive confirmatory identification of body fluids. **Forensic Science International** **181**, e1–e5, 2008.

ZADORA, G.; MENZYK, A.; In the pursuit of the holy grail of forensic science e Spectroscopic studies on the estimation of time since deposition of bloodstains. **Trends in Analytical Chemistry** **105**, 137-165, 2018.

## ANEXO A-TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO (TCLE)



**UNIVERSIDADE FEDERAL DE PERNAMBUCO**  
**CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA**  
**DEPARTAMENTO DE QUÍMICA FUNDAMENTAL**

### **TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO (PARA MAIORES DE 18 ANOS OU EMANCIPADOS - Resolução 466/12)**

Convidamos o (a) Sr. (a) para participar como voluntário (a) da pesquisa **“Desenvolvimento de métodos analíticos baseados em imagens hiperespectrais para fins forenses”**, que está sob a responsabilidade da pesquisadora **Carolina Santos Silva** (Rua Dr. Genaro Guimarães, 12, Casa Amarela, Recife-PE, CEP: 52070-040; tel: (81)92456456, (81)32680145, carolinasantosilva@gmail.com) e está sob a orientação de: M<sup>a</sup> Fernanda Pimentel, Telefones para contato: (81)21267233; (81)21267235, e-mail (mfp@ufpe.br). Também participam também desta pesquisa:(Ricardo S. Honorato) Telefones: (21373969).

Este Termo de Consentimento pode conter informações que o/a senhor/a não entenda. Caso haja alguma dúvida, pergunte à pessoa que está lhe entrevistando para que o/a senhor/a esteja bem esclarecido (a) sobre sua participação na pesquisa. Após ser esclarecido (a) sobre as informações a seguir, caso aceite em fazer parte do estudo, rubriche as folhas e assine ao final deste documento, que está em duas vias. Uma delas é sua e a outra é do pesquisador responsável. Em caso de recusa o (a) Sr. (a) não será penalizado (a) de forma alguma. Também garantimos que o (a) Senhor (a) tem o direito de retirar o consentimento da sua participação em qualquer fase da pesquisa, sem qualquer penalidade.

#### INFORMAÇÕES SOBRE A PESQUISA:

Essa pesquisa tem como objetivo desenvolver métodos confiáveis, rápidos e não destrutivos para identificação e datação de vestígios de fluidos corporais em cenas de crimes utilizando imagens químicas das manchas dos fluidos doados em diferentes tecidos. Amostras de material biológico (sangue e/ou sêmen) serão depositadas em diferentes superfícies para identificação. A coleta dos materiais será realizada:

1. Sêmen:
  - a. A coleta do material será realizada no centro de Pesquisa Vidas.
  - b. O procedimento de coleta consiste na masturbação/autoestimulação voluntária do doador, não podendo fazer uso de água, lubrificante, saliva, sabonete ou qualquer outro produto químico durante o procedimento. O doador deve depositar o material no recipiente fornecido pela equipe.
  - c. O período de participação do voluntário na pesquisa consiste em apenas um dia de visita para a doação do material.
  - d. Serão colhidos aproximadamente 2,0 ml de sêmen por doador.
  - e. Para os doadores voluntários de sêmen, não há riscos associados ao procedimento de doação.
2. Sangue:
  - a. A coleta será realizada utilizando material descartável em ambiente limpo e apropriado. A coleta será realizada pelo farmacêutico participante da equipe.
  - b. O período de participação do voluntário na pesquisa consiste em apenas um dia de visita para a doação do material.
  - c. Serão colhidos aproximadamente 0,05 ml de sangue (equivalente a uma gota) por punção na falange distal do 4º dedo da mão direita.
  - d. Para os doadores de sangue, o risco associado ao procedimento de coleta é um leve desconforto no local puncionado.

Não há benefícios diretos à saúde dos voluntários. Os benefícios dessa pesquisa estão associados a questões de segurança pública.

As informações desta pesquisa serão confidenciais e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos voluntários, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre a sua participação. Os dados coletados nesta pesquisa (material biológico), ficarão armazenados em computador pessoal, sob a responsabilidade do pesquisador, no endereço acima informado, pelo período de mínimo 5 anos.

O (a) senhor (a) não pagará nada para participar desta pesquisa. Se houver necessidade, as despesas para a sua participação serão assumidas pelos pesquisadores (ressarcimento de transporte e alimentação). Fica também garantida indenização em casos de danos, comprovadamente decorrentes da participação na pesquisa, conforme decisão judicial ou extra-judicial.

Em caso de dúvidas relacionadas aos aspectos éticos deste estudo, você poderá consultar o Comitê de Ética em Pesquisa Envolvendo Seres Humanos da UFPE no endereço: (Avenida da Engenharia s/n – 1º Andar, sala 4 - Cidade Universitária, Recife-PE, CEP: 50740-600, Tel.: (81) 2126.8588 – e-mail: cepccs@ufpe.br).

\_\_\_\_\_  
(assinatura do pesquisador)

#### CONSENTIMENTO DA PARTICIPAÇÃO DA PESSOA COMO VOLUNTÁRIO (A)

Eu, \_\_\_\_\_, CPF \_\_\_\_\_, abaixo assinado, após a leitura (ou a escuta da leitura) deste documento e de ter tido a oportunidade de conversar e ter esclarecido as minhas dúvidas com o pesquisador responsável, concordo em participar do estudo **“Desenvolvimento de métodos analíticos baseados em imagens hiperespectrais para fins forenses”**, como voluntário(a).

Fui devidamente informado (a) e esclarecido (a) pelo(a) pesquisador (a) sobre a pesquisa, os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios decorrentes de minha participação. Foi-me garantido que posso retirar o meu consentimento a qualquer momento, sem que isto leve a qualquer penalidade (ou interrupção de meu acompanhamento/assistência/tratamento).

Local e data: \_\_\_\_\_

Assinatura do participante: \_\_\_\_\_

Impressão Digital  
(Opcional)

**Presenciamos a solicitação de consentimento, esclarecimentos sobre a pesquisa e o aceite do voluntário em participar.** (02 testemunhas não ligadas à equipe de pesquisadores):

Nome:	Nome:
Assinatura:	Assinatura: