



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I – CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA
CURSO DE MESTRADO EM QUÍMICA**

YURI ANIEL DOS SANTOS VASCONCELOS

**AUTENTICAÇÃO DE ORIGEM GEOGRÁFICA DE CACHAÇAS DA PARAÍBA
UTILIZANDO ESPECTROSCOPIA PORTÁTIL NIR E RAMAN ACOPLADA COM
DD-SIMCA**

**CAMPINA GRANDE-PB
2025**

YURI ANIEL DOS SANTOS VASCONCELOS

AUTENTICAÇÃO DE ORIGEM GEOGRÁFICA DE CACHAÇAS DA PARAÍBA
UTILIZANDO ESPECTROSCOPIA PORTÁTIL NIR E RAMAN ACOPLADA COM
DD-SIMCA

Dissertação apresentada à Coordenação do
Curso de Programa de Pós-Graduação em
Química - Mestrado da Universidade Estadual
da Paraíba, como requisito parcial à obtenção
do título de Mestre em Química.

Linha de pesquisa: Metodologias Analíticas
Avançadas - Campina Grande.

Orientador: Prof. Dr. José Germano Vêras Neto
Coorientador: Prof. Dr. David Douglas de Sousa Fernandes

CAMPINA GRANDE-PB
2025

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

V331a Vasconcelos, Yuri Aniel dos Santos.

Autenticação de origem geográfica de cachaças da Paraíba utilizando espectroscopia portátil NIR e RAMAN acoplada com DD-SIMCA [manuscrito] / Yuri Aniel dos Santos Vasconcelos. - 2025.

58 f.: il. color.

Digitado.

Dissertação (Programa de Pós-Graduação em Química - Mestrado) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2025.

"Orientação: Prof. Dr. Jose Germano Veras Neto, Departamento de Química - CCT".

1. Bebida alcoólica. 2. Análise multivariada. 3. Raman.
4. Infravermelho próximo. I. Título

21. ed. CDD 543

YURI ANIEL DOS SANTOS VASCONCELOS

AUTENTICAÇÃO DE ORIGEM GEOGRÁFICA DE CACHAÇAS DA PARAÍBA
UTILIZANDO ESPECTROSCOPIA PORTÁTIL NIR E RAMAN ACOPLADA COM
DD-SIMCA

Dissertação apresentada à Coordenação do
Curso de Programa de Pós-Graduação em
Química - Mestrado da Universidade Estadual
da Paraíba, como requisito parcial à obtenção
do título de Mestre em Química.

Linha de Pesquisa: Metodologias Analíticas
Avançadas - Campina Grande.

Aprovada em: 25/07/2025.

BANCA EXAMINADORA

Documento assinado eletronicamente por:

- **Mayara Ferreira Barbosa** (***.377.234-**), em **29/08/2025 11:02:32** com chave **ceb8f50884e011f08aea06adb0a3afce**.
- **Jose Germano Veras Neto** (***.206.144-**), em **29/08/2025 10:14:22** com chave **145b83c084da11f0b91606adb0a3afce**.
- **WELMA THAÍSE SILVA VILAR** (***.560.824-**), em **29/08/2025 10:15:57** com chave **4cc156ea84da11f0808906adb0a3afce**.

Documento emitido pelo SUAP. Para comprovar sua autenticidade, faça a leitura do
QrCode ao lado ou acesse https://suap.uepb.edu.br/comum/autenticar_documento/ e informe os dados a seguir.

Tipo de Documento: Folha de Aprovação do Projeto Final

Data da Emissão: 29/08/2025

Código de Autenticação: 5b5472



AGRADECIMENTOS

. A ciência se constrói sobre aprendizado, dedicação e colaboração. Neste percurso, cada experiência foi essencial para o desenvolvimento deste trabalho. Por isso, agradeço a todos que contribuíram, direta ou indiretamente, para sua realização, especialmente aos docentes da UEPB e aos membros do LQAQ.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

A ciência e os números dançam juntos
para revelar padrões invisíveis
(Yuri Aniel)

RESUMO

A cachaça é uma bebida alcoólica destilada tipicamente brasileira, produzida a partir da fermentação e destilação do caldo de cana-de-açúcar. A sua história e influência lhe concedeu o reconhecimento mundial como símbolo da cultura do Brasil. Reconhecida por seu valor histórico e cultural, a cachaça também representa um importante produto de exportação, assim, a autenticação geográfica dessa bebida assume grande relevância econômica e cultural, pois contribui para a valorização do patrimônio, além de proteger consumidores e produtores. Tradicionalmente, a autenticação dessa bebida era realizada por meio de técnicas organolépticas, entretanto, o avanço tecnológico e a necessidade de uma compreensão mais detalhada de seus componentes químicos e sensoriais impulsionaram o desenvolvimento e a aplicação de novas técnicas de análise e autenticação. Portanto, este trabalho tem como objetivo utilizar diferentes técnicas espectroscópicas, como a espectroscopia no infravermelho próximo (NIR) e espectroscopia Raman, empregando equipamentos portáteis, para autenticar amostras de cachaça produzidas no estado paraibano. Essas técnicas de análise, quando combinadas com métodos quimiométricos de classe única, como o DD-SIMCA (*Data-Driven Soft Independent Modeling of Class Analogy*), se mostraram capazes de classificar essas bebidas com precisão, identificando-as como pertencentes a um determinado grupo. Onde foi possível alcançar uma acurácia de 64,80% utilizando os dados NIR, com sensibilidade de 88,90% e especificidade de 60,00%, sendo o melhor pré-processamento a correção de linha de base (offset). A espectroscopia Raman, por outro lado, se destacou ao atingir uma acurácia de 98,10%, com sensibilidade de 100% e especificidade de 97,80%, tendo como melhor pré-processamento a correção de dispersão multiplicativa. Esses resultados apresentam o potencial dessas ferramentas para garantir a autenticidade e a diferenciação geográfica das cachaças, pois são técnicas que oferecem análises rápidas, não destrutivas e confiáveis, promovendo assim uma maior valorização do produto e consolidação da identidade cultural.

Palavras-Chave: bebida alcoólica; autenticação; Raman; infravermelho próximo; análise multivariada

ABSTRACT

Sugar cane brandy is a distilled alcoholic beverage that is typically Brazilian, produced from the fermentation and distillation of sugarcane juice. Its history and cultural influence have granted it worldwide recognition as a symbol of Brazilian culture. Recognized for its historical and cultural value, sugar cane brandy also represents an important export product. Thus, the geographical authentication of this beverage assumes great economic and cultural relevance, as it contributes to the valorization of heritage while protecting both consumers and producers. Traditionally, the authentication of sugar cane brandy was performed through organoleptic techniques, however, technological advances and the need for a more detailed understanding of its chemical and sensory components have driven the development and application of new techniques for analysis and authentication. Therefore, this study aims to apply different spectroscopic techniques, such as near-infrared (NIR) and Raman spectroscopy, using portable equipment, to authenticate sugar cane brandy samples produced in the state of Paraíba. These analytical techniques, when combined with one-class chemometric methods such as DD-SIMCA (Data-Driven Soft Independent Modeling of Class Analogy), proved capable of accurately classifying these beverages, identifying them as belonging to a specific group. It was possible to achieve an accuracy of 64.80% using NIR data, with a sensitivity of 88.90% and specificity of 60.00%, where the best preprocessing method was baseline correction (offset). Raman spectroscopy, on the other hand, stood out by reaching an accuracy of 98,10%, with sensitivity of 100% and specificity of 97.80%, with multiplicative scatter correction being the best preprocessing method. These results highlight the potential of these tools to ensure the authenticity and geographical differentiation of sugar cane brandy, as they provide fast, non-destructive, and reliable analyses, thereby promoting greater product valorization and the consolidation of cultural identity.

Palavras-Chave: alcoholic beverage; authentication; Raman; near infrared; multivariate analysis

SUMÁRIO

1	INTRODUÇÃO	9
2	OBJETIVOS.....	11
2.1	Objetivo Geral	11
2.2	Objetivos Específicos.....	11
3	REFERENCIAL TEÓRICO.....	12
3.1	Cachaça.....	12
3.1.1	<i>Aspectos históricos da cachaça.....</i>	<i>12</i>
3.1.2	<i>Principais constituintes químicos da cachaça</i>	<i>12</i>
3.1.3	<i>Aspectos produtivos e econômicos da cachaça no Brasil.....</i>	<i>13</i>
3.2	Espectroscopia	15
3.2.1	<i>Espectroscopia no infravermelho próximo.....</i>	<i>16</i>
3.2.2	<i>Espectroscopia Raman</i>	<i>17</i>
3.3	Quimiometria	19
3.3.1	<i>Métodos não supervisionados.....</i>	<i>20</i>
3.3.2	<i>Análise por componentes principais (PCA).....</i>	<i>20</i>
3.3.3	<i>Métodos supervisionados.....</i>	<i>22</i>
3.3.4	<i>Modelagem analógica de classes independentes baseada em dados (DD-SIMCA).....</i>	<i>22</i>
4	MATERIAL E MÉTODOS	25
4.1	Amostras	25
4.2	Aquisição dos espectros NIR.....	25
4.3	Aquisição dos espectros Raman	25
4.4	Pré-processamento dos dados	26
4.5	Construção do modelo PCA	26
4.6	Métodos de classificação.....	26
5	RESULTADOS E DISCUSSÃO	28
5.1	Resultados utilizando a espectroscopia NIR.....	28
5.1.1	<i>Pré-tratamento dos dados espectrais NIR</i>	<i>28</i>
5.1.2	<i>Análise exploratória de dados utilizando PCA.....</i>	<i>32</i>
5.1.3	<i>Análise classificatória utilizando DD-SIMCA.....</i>	<i>34</i>
5.2	Resultados utilizando a espectroscopia Raman.....	38
5.2.1	<i>Pré-tratamento dos dados espectrais Raman.....</i>	<i>38</i>
5.2.2	<i>Análise exploratória de dados utilizando PCA.....</i>	<i>43</i>
5.2.3	<i>Análise classificatória utilizando DD-SIMCA.....</i>	<i>45</i>

6	CONCLUSÕES	51
	REFERÊNCIAS	53

1 INTRODUÇÃO

As bebidas alcoólicas são consumidas e apreciadas por pessoas em todo o mundo, sendo o seu consumo e produção uma prática comum em diversas sociedades. Tais bebidas são caracterizadas por fatores como matérias-primas, formas de produção, e condições edafoclimáticas, como a localização geográfica (LIN *et al.*, 2023).

Dentre as diversas bebidas alcoólicas existentes, a cachaça se destaca como sendo uma bebida originada no Brasil, tendo suas raízes na colonização portuguesa no início do século XVI. A cachaça é amplamente reconhecida por seus aspectos históricos e culturais e constitui um importante produto de exportação, que difunde a cultura brasileira pelo mundo. Segundo o Instituto Brasileiro da Cachaça (IBRAC), estima-se que o Brasil possui capacidade de produzir cerca de 1,2 bilhões de litros anuais de cachaça, e que o setor gera mais de 600 mil empregos diretos e indiretos, exercendo assim um grande impacto sobre a economia brasileira e o mercado de trabalho (IBRAC, 2025).

O Estado da Paraíba tem se destacado pela grande produção de cachaça de alambique, denominada de artesanal. Além disso, a cachaça paraibana possui uma qualidade que chama atenção a nível nacional, possuindo características que lhe são únicas, sendo apreciada tanto localmente como também a nível internacional (VILELA *et al.*, 2021).

Devido aos impactos sociais e econômicos desta bebida, é comum a busca por uma cachaça de maior qualidade, que ofereça para o consumidor satisfação em termos sensoriais, assim como oferecer um produto que contenha uma menor exposição a componentes indesejados na bebida. Dessa forma, cada produtor possui meios e formas únicas de produção, o que lhe confere características únicas, construindo assim uma identidade regional (PEREIRA *et al.*, 2024).

De acordo com o Instituto Nacional da Propriedade Industrial (INPI) que é regulamentada pela lei nº 9.279/1996 (lei da propriedade industrial), atualmente o Brasil possui três regiões produtoras de cachaça com identificação geográfica (IG) sendo elas Paraty (Rio de Janeiro), Salinas (Minas Gerais) e a microrregião de Abaíra (Bahia). No presente momento o estado da Paraíba não possui um registro de identificação geográfica (IG) pelo INPI, contudo, devido à tradição e à qualidade na produção de cachaça artesanal, esses fatores representam um forte indicativo do potencial para obtenção futura do registro de identificação geográfica (INPI, 2023).

Diante do aumento do consumo de bebidas e alimentos, a autenticação da origem geográfica torna-se fundamental, garantindo qualidade, segurança e rastreabilidade dos

produtos. A ausência de métodos confiáveis de autenticação pode levar à desvalorização de produtos autênticos, prejuízos econômicos e riscos à saúde do consumidor, decorrentes de adulterações. Assim, o desenvolvimento e o aperfeiçoamento de metodologias de classificação e autenticação são essenciais para a proteção da cadeia produtiva da cachaça. No entanto, essas análises de cachaças e discussões comparativas sistemáticas ainda carecem de trabalhos, necessitando assim de mais estudos que comparem sistematicamente os dados obtidos (LIN *et al.*, 2023).

Portanto, técnicas espectroscópicas, aliadas à quimiometria, desempenham papel essencial na otimização dos processos de produção e no controle de qualidade. Esses recursos permitem identificar contaminantes e adulterantes, assegurando a autenticidade e a origem dos produtos. Nesse contexto, a autenticação torna-se indispensável para garantir qualidade, segurança e confiabilidade. A ausência de autenticação pode desvalorizar produtos autênticos, prejudicar economicamente o setor e expor o consumidor a riscos. Portanto, é fundamental aplicar e aperfeiçoar métodos de autenticação, assegurando maior segurança e confiança no consumo da cachaça (KHARBACH *et al.*, 2023).

Dessa forma, o presente trabalho tem como objetivo contribuir para a autenticação geográfica de cachaças produzidas no estado da Paraíba em relação às de outros estados brasileiros, utilizando espectroscopia portátil NIR e espectroscopia Raman, associadas à análise multivariada por meio do método DD-SIMCA. A proposta busca oferecer uma ferramenta rápida, não destrutiva e eficiente para a diferenciação de origem, promovendo o reconhecimento e a valorização da cachaça paraibana.

2 OBJETIVOS

2.1 Objetivo Geral

- Diferenciar cachaças do estado da Paraíba dos demais estados do Brasil, através de métodos espectroscópicos e modelos de classificação quimiométricos

2.2 Objetivos Específicos

- Investigar o potencial das técnicas espectroscópicas aliadas à quimiometria como ferramentas aplicáveis à autenticação geográfica de cachaças
- Comparar os resultados obtidos por diferentes técnicas espectroscópicas a fim de verificar se estes métodos levam a conclusões semelhantes em relação a autenticação geográfica
- Validar o modelo de classificação para verificar sua eficácia na autenticação regional nesse respectivo contexto

3 REFERENCIAL TEÓRICO

3.1 Cachaça

3.1.1 Aspectos históricos da cachaça

A cachaça é considerada um símbolo da herança cultural brasileira, tendo cerca de 500 anos de existência. Sua história remonta ao período colonial, onde começou a ser produzida quando os portugueses trouxeram a cana-de-açúcar para o Brasil, permitindo assim a produção da primeira bebida destilada das Américas. Sua história é marcada por altos e baixos ao longo dos séculos, pois em muitos contextos já teve uma imagem negativa. Contudo, atualmente é reconhecida como uma bebida típica do Brasil, sendo valorizada tanto nacionalmente como internacionalmente (GUIMARÃES JUNIOR *et al.*, 2024).

A produção da cachaça é datada do início do século XVI, inicialmente esta bebida era consumida principalmente pelos escravizados, devido ao seu custo e facilidade de produção. No entanto, a medida com que a cachaça se popularizava, a bebida passou a ser consumida por diferentes classes sociais, onde os colonizadores começaram a substituir as bebidas típicas de sua região da Europa, pelo consumo da cachaça, contribuindo para a valorização da bebida e consolidação cultural do Brasil (GONÇALVES *et al.*, 2021).

Com o passar do tempo, o setor de cachaça aperfeiçoou-se e vêm melhorando as tecnologias de produção da bebida, desde o cultivo da cana-de-açúcar quanto aos processos de produção, proporcionando assim um aumento da eficiência produtiva e pôr fim a obtenção de um produto de maior qualidade (SILVA *et al.*, 2023).

3.1.2 Principais constituintes químicos da cachaça

O termo cachaça é empregado para designar o destilado da cana-de-açúcar que é produzido exclusivamente no Brasil, cujo teor alcoólico varia entre 38 e 48% de álcool (v/v). A composição química da cachaça é constituída por uma matriz complexa que lhe confere propriedades que lhe são bem características como o aroma e sabor. As principais substâncias presentes nessa bebida são água, etanol, álcoois superiores, ésteres etílicos, aldeídos, cetonas e ácidos carboxílicos (GAMA, 2022; MELO *et al.*, 2021; RATKOVICH *et al.*, 2023).

A qualidade da cachaça está intimamente ligada a fatores como, produção que influencia diretamente a composição química, e por sua vez as qualidades sensoriais em geral, como características de odor, paladar e até aspectos visuais, bem como a qualidade química, ou seja, uma bebida com uma menor concentração de espécies químicas nocivas à saúde e que causam uma diminuição na qualidade sensorial da bebida (BORTOLETTO *et al.*, 2013).

Dentre as principais espécies químicas presentes na bebida, as que desempenham maior impacto sobre suas características sensoriais são os ésteres, álcoois, aldeídos e ácidos carboxílicos, sendo os ésteres o grupo de maior influência nesse aspecto (OLIVEIRA *et al.*, 2020).

O éster predominante nas aguardentes é o acetato de etila, seguido pelo lactato de etila, que juntos correspondem a aproximadamente 95% do total de ésteres presentes na bebida. O acetato de etila confere um aroma frutado característico, embora possa se tornar desagradável quando em concentrações elevadas. Já o lactato de etila também apresenta aroma frutado, contribuindo para o perfil sensorial da cachaça. A destilação realizada em alambiques tende a gerar uma maior concentração de ésteres em comparação à destilação por colunas contínuas. Isso ocorre porque o cobre presente no alambique atua como catalisador das reações de esterificação no ambiente aquecido, intensificando a formação desses compostos. Por esse motivo, aguardentes produzidas em alambiques geralmente exibem aromas frutados mais pronunciados, principalmente devido à presença de ésteres de cadeia curta (NÓBREGA, 2003).

O ácido orgânico mais encontrado em aguardentes de cana é o ácido acético, é o composto orgânico predominante encontrado na cachaça, representando um teor na faixa de 90 a 93% do conteúdo total de ácidos orgânicos da bebida. O controle da concentração de ácido acético é fundamental, pois teores mais baixos estão relacionados a uma acidez mais equilibrada e, conseqüentemente, a uma bebida de melhor qualidade. Outra classe de compostos que têm grande importância na composição de aguardentes são os aldeídos, sendo os principais o acetaldeído, furfural e hidroximetilfurfural, que são caracterizados por apresentarem odores fortes e frequentemente enjoativos. Estas espécies químicas estão relacionadas a diversos sintomas negativos ao organismo humano como náuseas, vômitos, agitação, transpiração, e dores de cabeça resultado em um mal-estar popularmente conhecido como ressaca, dessa forma, baixas concentrações de aldeídos são esperados em uma cachaça de boa qualidade (ALCARDE, 2017; SILVA *et al.*, 2023).

3.1.3 Aspectos produtivos e econômicos da cachaça no Brasil

A cachaça desempenha papel importante na economia brasileira, gerando empregos diretos e indiretos ao longo de toda a cadeia produtiva, movimentando a indústria local e fortalecendo a renda regional. Além disso, constitui um produto com potencial de exportação, contribuindo para a presença do Brasil no mercado internacional de destilados.

De acordo com o anuário de 2024 produzido pela secretaria de defesa agropecuária do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), o estado da Paraíba, sobretudo

a região do brejo paraibano se destaca entre uma das regiões que possui maior concentração de destilarias de cachaça no país, além disso a Paraíba é a terceira maior em densidade de estabelecimentos por unidade da federação tendo uma cachaçaria para cada 79.484 habitantes, indicando assim uma tradição significativa na fabricação da bebida, mostrando com isto uma alta relevância econômica na região e refletindo seu papel importante no cenário nacional (BRASIL, 2024).

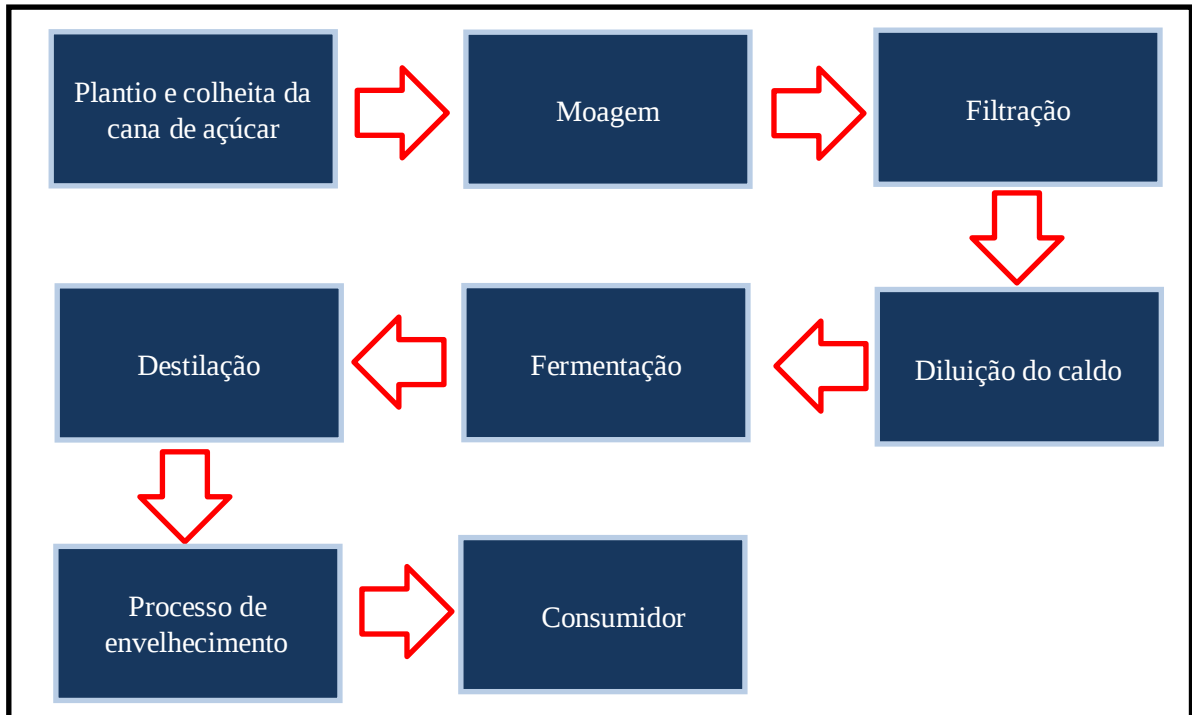
A cachaça é a segunda bebida alcoólica de maior consumo no Brasil, ficando atrás apenas para a cerveja, sendo o Brasil um grande produtor de cachaça, no entanto, apenas cerca de 1 a 2% dessa produção é exportada. A produção de cachaça no Brasil pode ser classificada em dois tipos, a produção artesanal e a industrial, sendo a artesanal o principal meio de produção no estado da Paraíba. O processo de fabricação de cachaça no Brasil é de profundo caráter institucional e legislativo, com destaque no aperfeiçoamento da qualidade da bebida, tendo características únicas, trata-se de uma bebida saborosa, e que com o passar do tempo e com os avanços tecnológicos resultam em uma maior ampliação das exportações e do consumo interno (OLIVEIRA *et al.*, 2022).

Considerando a diversidade de práticas na produção, é natural que cada produtor utilize métodos específicos, o que confere a cada cachaça uma identidade única. Essa bebida resulta da combinação de diversos fatores, incluindo a qualidade da matéria-prima, o tipo de alambique empregado, além dos processos de fermentação e destilação adotados. A combinação desses fatores confere à cachaça características sensoriais únicas de cada produtor.

A qualidade química e sensorial da cachaça está associada às suas etapas de produção demonstrada na **Figura 1**, desde o plantio da cana-de-açúcar até o envelhecimento. A fermentação é um dos processos mais críticos, pois determina a formação do etanol e de diversos outros compostos. Esse processo consiste na adição de leveduras ao caldo de cana, promovendo a conversão dos açúcares em etanol e metabólitos secundários responsáveis pelas características sensoriais da bebida (SCHOENINGER *et al.*, 2024).

Além disso, o tipo de alambique exerce forte influência sobre a qualidade final da cachaça. Os modelos mais utilizados são os de cobre e os de aço inox. O cobre é amplamente empregado por favorecer reações desejáveis na destilação, mas também pode aumentar a concentração de íons cúpricos no produto, exigindo cuidados adicionais por parte dos produtores (SILVA *et al.*, 2023).

Figura 1: Principais etapas de produção de cachaça artesanal



Fonte: adaptado de SCHOENINGER (2024).

Assim, a cachaça, com sua história rica que remonta ao Brasil colonial, tem evoluído não apenas em termos de técnicas de produção, mas também na busca por uma bebida de maior qualidade. Tradicionalmente, o controle da qualidade da cachaça era realizado sem o uso de tecnologias avançadas ou métodos científicos sofisticados, no entanto, com a modernização da indústria da cachaça, surge a necessidade de entender com mais profundidade os componentes químicos e sensoriais da bebida (RATKOVICH *et al.*, 2023). Técnicas espectroscópicas portáteis, como NIR e Raman, surgem como ferramentas promissoras para análise da composição química da bebida, possibilitando maior precisão em aspectos relacionados à sua caracterização e autenticação.

3.2 Espectroscopia

A espectroscopia é a área de estudo da interação da radiação com a matéria, baseado nos comprimentos de onda que são absorvidos, refletidos ou emitidos por uma espécie química. Essa interação acontece por meio da absorção e emissão de fótons, que se manifestam de diferentes maneiras. A dependência dessas interações em relação ao comprimento de onda permite obter informações detalhadas sobre as propriedades químicas de uma substância. Portanto as técnicas espectroscópicas têm promovido um avanço no que se diz respeito ao

entendimento da estrutura da matéria. Os dados espectroscópicos são geralmente representados por um espectro, que é a representação gráfica da resposta de interesse em função da energia. A espectroscopia tem uma aplicação muito vasta, sendo bastante utilizada para determinação, identificação e classificação de muitos compostos orgânicos e inorgânicos (MOURÃ *et al.*, 2023).

Isso se deve ao fato de que as técnicas espectroscópicas apresentam baixo custo em comparação a outras técnicas, como espectrometria de massas, cromatografia líquida e gasosa, entre outras. Além disso, essas técnicas se destacam por fornecerem respostas rápidas e pela praticidade de contar com equipamentos tanto de bancada quanto portáteis (SANTOS *et al.*, 2022).

3.2.1 Espectroscopia no infravermelho próximo

A espectroscopia no infravermelho próximo (NIR) é caracterizada por ser uma técnica vibracional, que corresponde à faixa de comprimento de onda de 800 a 2.500 nm. O NIR é baseado em harmônicos moleculares e vibrações combinadas, onde os espectros são registrados através de modos de reflexão, transmissão ou absorção, fornecendo informações químicas e físicas (BEĆ *et al.*, 2019).

A espectroscopia no infravermelho é uma técnica baseada nas vibrações moleculares, nas quais as ligações químicas estão em constante movimento. A radiação infravermelha interage com essas vibrações, provocando variações no movimento vibracional das ligações, o que pode alterar tanto o ângulo (deformação) quanto o comprimento (estiramento) dessas ligações. As moléculas vibram em frequências específicas, e se a radiação no infravermelho estiver na mesma frequência de vibração, ocorre assim a absorção, resultando no surgimento de bandas características da respectiva ligação (BEĆ *et al.*, 2020).

As diferentes ligações químicas têm sua própria frequência natural de vibração, contudo, devido ao acoplamento vibracional, fenômeno esse que consiste na influência de ligações vizinhas no número de absorção de uma respectiva ligação, os números de onda de uma ligação, apresentam variação de uma molécula para outra. Pode-se pensar que devido às interferências do acoplamento vibracional, a interpretação dos espectros seja dificultada, mas é justamente devido a esse fator interferente que proporciona características únicas a um espectro de uma espécie química, criando assim uma impressão digital, que nos permite discriminar com mais facilidade os espectros de diferentes espécies químicas (BEĆ *et al.*, 2020).

Portanto, a espectroscopia no infravermelho possibilita obter informações qualitativas e quantitativas sobre as espécies químicas presentes em diferentes amostras, por meio da

interação da radiação eletromagnética dessa faixa do espectro com o material analisado. Em especial, a espectroscopia no infravermelho próximo destaca-se por ser uma técnica rápida, não destrutiva e que exige praticamente nenhum preparo de amostras, permitindo a análise de diversos materiais de maneira prática e eficiente (RECH *et al.*, 2020).

3.2.2 Espectroscopia Raman

A espectroscopia Raman assim como a espectroscopia no infravermelho é uma técnica vibracional, no qual a luz interage com as ligações da molécula provocando mudanças em suas vibrações, fornecendo assim informações sobre a estrutura e as condições químicas de uma determinada espécie química. Essas vibrações revelam detalhes importantes sobre os tipos de ligações e a disposição dos átomos, fornecendo assim uma visão detalhada da composição da molécula (DODO *et al.*, 2022).

Como a maioria das moléculas apresenta modos vibracionais e diversos grupos ativos no Raman, essa técnica pode ser utilizada para realizar a análise de diferentes materiais. Devido ao Raman ser caracterizado como uma técnica robusta e versátil entre as técnicas instrumentais, ele é bastante utilizado em diversas áreas. As suas aplicações consistem em grande parte na área da analítica e tendo também suas aplicações em tecidos biológicos, como na pesquisa da pele, medicamentos e utilizações em diagnósticos (ORLANDO *et al.*, 2021; LUNTER *et al.*, 2022).

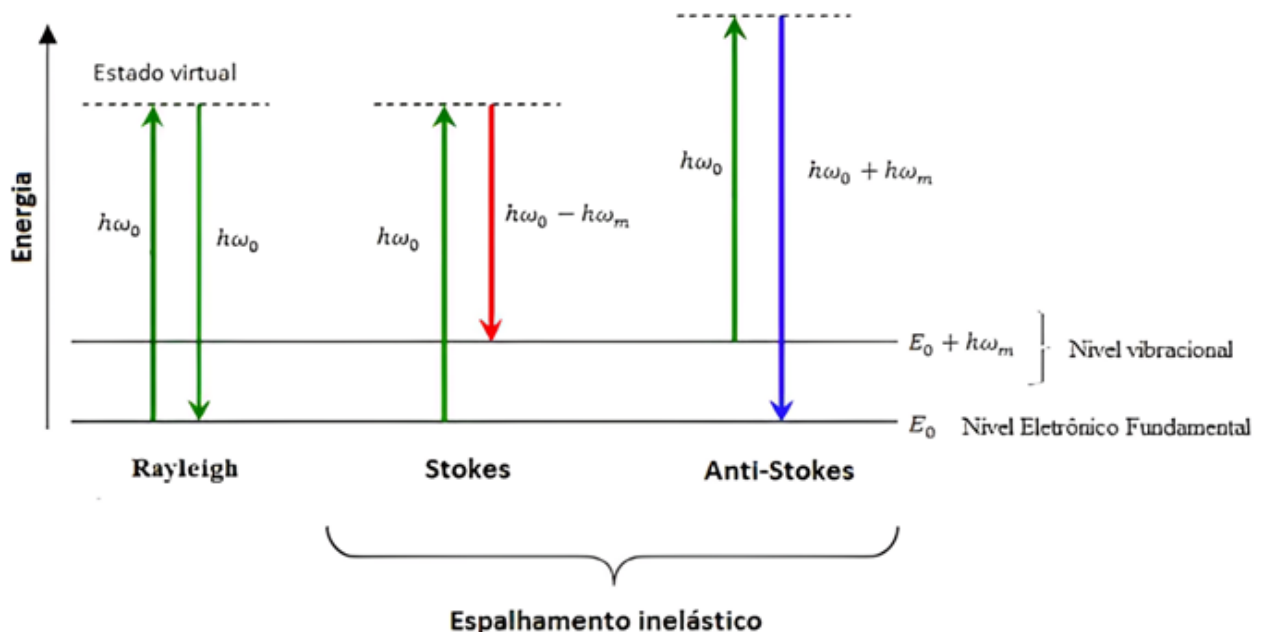
A espectroscopia Raman consiste em sondar os modos vibracionais associados às ligações químicas de uma respectiva espécie química, onde os espectros Raman obtido estão relacionados com as vibrações moleculares de cada molécula, e como cada molécula possui uma dinâmica vibracional única, os espectros Raman terão características únicas. Com isso, os espectros Raman configuram uma impressão digital específica de cada molécula, visto que a menor diferença na estrutura ou na composição da molécula irá resultar em uma dinâmica vibracional diferente, resultando assim em um espectro Raman também com diferenças em relação ao deslocamento Raman (HANNA *et al.*, 2022).

Para obter os espectros Raman, a técnica consiste em utilizar um laser, cuja radiação é monocromática e se encontra geralmente na região do visível (400–700 nm) ou infravermelho próximo (700–1100 nm). Com a irradiação do laser sobre a amostra, ocorre o processo de excitação, onde o fóton incidente interage com a molécula provocando uma vibração nas ligações químicas da amostra. Através desta interação o fóton é disperso (espalhado) com uma mudança na sua energia, desta forma, dois fenômenos de espalhamento podem ocorrer: o espalhamento Rayleigh que consiste em um espalhamento elástico e o espalhamento Raman

que por sua vez é inelástico, fenômenos esses ilustrados na **Figura 2**. O espalhamento Rayleigh é dominante e caracterizado por ter um fóton disperso com a mesma energia do fóton incidente, e como as suas energias são iguais, o espalhamento Rayleigh não produz nenhuma informação específica sobre a molécula, pois resulta em um deslocamento Raman igual a zero. Contudo, o fóton emitido pode apresentar uma variação na sua energia em relação ao fóton incidente, ocorrendo assim o deslocamento Raman. podendo o fóton ter uma energia menor (deslocamento Stokes) ou uma energia maior que o fóton incidente (deslocamento anti-Stokes). As variações de energia no deslocamento Raman estão relacionadas com as vibrações moleculares, e, portanto, o espectro Raman irá apresentar deslocamentos característicos de cada molécula (GABA *et al.*, 2022).

Portanto a espectroscopia Raman consiste em ser uma técnica não destrutiva e não invasiva, possuindo uma rápida velocidade de obtenção dos espectros, permitindo uma varredura e amostragem em tempo real. Dessa forma, a técnica possibilita a análise da amostra em etapas subsequentes, e como os espectros são obtidos em tempo real, isso permite com que o Raman seja utilizado em processos de monitoramento e controle da qualidade do material analisado (ZHANG *et al.*, 2022).

Figura 2: Representação dos espalhamentos Rayleigh, Stokes e Anti-Stokes



Fonte: adaptado de (NASCIMENTO-DIAS 2020)

A fluorescência que pode ser causada pelo laser é um fator interferente na espectroscopia Raman, pois pode obscurecer as bandas do Raman, que são características das

vibrações moleculares de uma espécie química. Portanto, amostras que apresentam altos níveis de fluorescência podem gerar um problema na análise por espectroscopia Raman. Apesar da cachaça possuir espécies químicas fluorescentes como, compostos fenólicos e dentre outras espécies químicas, esse problema pode ser mitigado com o ajuste no comprimento de onda do Raman, que pode intensificar as bandas do Raman, bem como mitigar os efeitos da fluorescência ao selecionar um comprimento de onda distante de bandas de absorção característica do analito analisado (SKOOG *et al.*, 2017).

Apesar de as técnicas espectroscópicas serem eficazes na caracterização de amostras, elas geram uma grande quantidade de dados complexos. Nesse contexto, a quimiometria torna-se essencial, ao organizar, interpretar e extrair informações relevantes, permitindo uma análise mais robusta e confiável.

3.3 Quimiometria

Com o desenvolvimento da química analítica instrumental, impulsionado pela eletrônica de microcomputadores a partir da década de 1970, a quantidade de resultados gerados durante uma análise química aumentou consideravelmente. Isso criou a necessidade de avançar em métodos estatísticos multivariados capazes de interpretar simultaneamente esses dados, extraíndo o máximo de informação com o menor número possível de experimentos. Esse avanço tecnológico e científico possibilitou o surgimento da quimiometria, uma subárea da química que tem como objetivo empregar métodos matemáticos e estatísticos para otimizar a manipulação dos dados, gerando informações químicas úteis por meio de análises multivariadas e da aplicação de ferramentas matemáticas, estatísticas e computacionais (SILVA *et al.*, 2023).

A quimiometria é aplicada em diversas áreas, abrangendo desde o reconhecimento de padrões, utilizado para fornecer uma visão geral das amostras quanto às suas semelhanças e diferenças e para classificá-las em grupos específicos, até a calibração multivariada, voltada à construção e ajuste de modelos matemáticos que relacionam dados experimentais a propriedades químicas. Esses modelos podem posteriormente ser validados para assegurar sua precisão e aplicabilidade. Além disso, a quimiometria permite planejar e otimizar experimentos, buscando a melhor resposta possível com o menor custo experimental.

A técnica de reconhecimento de padrões destaca-se em duas abordagens principais: a exploratória e a classificatória. Entre as metodologias exploratórias, sobressaem a Análise Hierárquica de Clusters (HCA, do inglês *Hierarchical Cluster Analysis*) e a Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*), ambas consideradas métodos não supervisionados. Já na classificação, os métodos mais empregados incluem a

Modelagem Independente Suave de Analogias de Classe (SIMCA, do inglês *Soft Independent Modeling of Class Analogy*) e a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA, do inglês *Partial Least Square Discriminant Analysis*), caracterizados como supervisionados.

3.3.1 Métodos não supervisionados

Os métodos não-supervisionados, como os métodos exploratórios, são técnicas de análise de dados que não necessitam de informação inicial sobre o conjunto de dados, pois o objetivo destes métodos é justamente buscar por padrões ou estruturas no conjunto de dados. Portanto, através desses métodos é possível conhecer no que as amostras se assemelham, assim como identificar variáveis que mais contribuem para semelhanças e diferenças, assim como identificar agrupamentos de amostras e demais conclusões que podem ser obtidas por essas técnicas.

3.3.2 Análise por componentes principais (PCA)

A análise de componentes principais é uma técnica quimiométrica que consiste em realizar uma manipulação matemática dos dados de uma matriz, a fim de obter a variação de muitas variáveis utilizando um pequeno número de fatores ou de componentes principais.

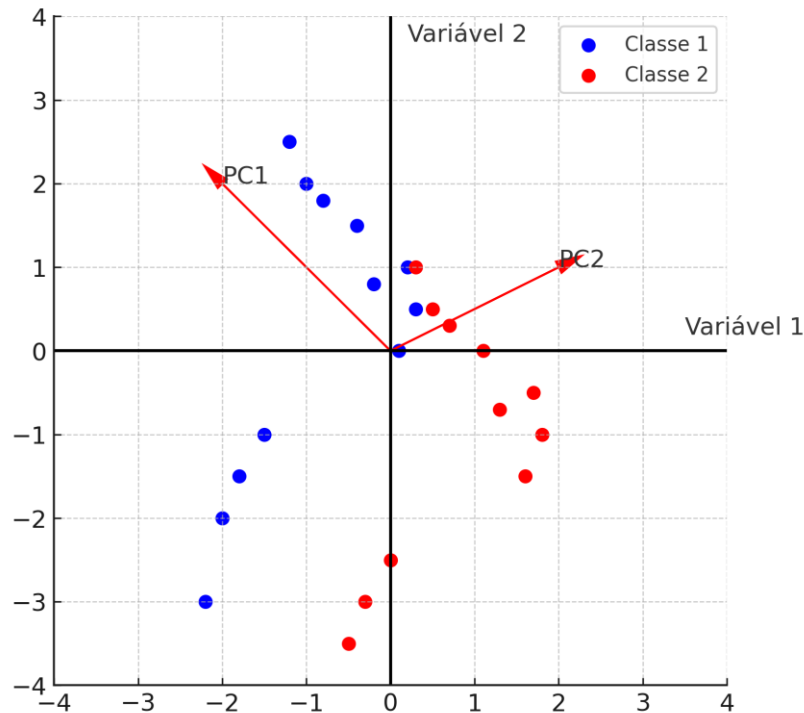
A PCA em termos matriciais consiste em decompor uma matriz (X) em duas matrizes menores, ou seja, uma matriz de escores (T) que contém informações sobre as amostras e uma matriz de pesos (P) que possui informações sobre as variáveis com a presença de resíduos (E), esta relação pode ser visualizada de acordo com a (**Equação 1**).

$$X = p^t T + E \quad (1)$$

Em termos geométricos a PCA projeta a matriz em um subespaço de dimensões reduzidas, onde possui o máximo de informação possível, ou seja, de variância. Neste sentido, o método projeta os dados multivariados em um novo sistema de coordenadas multidimensionais ilustrado pela **Figura 3**, criando assim componentes principais que permitem observar as diferenças entre as variáveis e seu impacto sobre a variação dos dados. A primeira componente principal (PC1) é a combinação linear de máxima variância dos dados originais, assim como a criação de novas componentes principais, a variância explicada por essas novas componentes principais irá explicar uma quantidade menor de variância, levando isto em consideração, em uma análise quimiométrica, geralmente se utiliza as primeiras

componentes principais, pois estas já possuem a maior parte da variância dos dados (SCHNEIDER *et al.*, 2018).

Figura 3: Representação da projeção dos dados sistema de coordenada bidimensional



Fonte: elaborado pelo autor, 2025.

Portanto, através da diminuição da dimensionalidade do conjunto de dados, os dados são plotados em duas ou três dimensões, permitindo assim que se analise os dados de forma mais fácil, visto que isso seria uma tarefa difícil a olho nu, senão impossível em matrizes de alta dimensão.

Dentre essas análises, é possível identificar agrupamentos de amostras, variáveis correlacionadas, variáveis que mais contribuem para as diferenças, assim como diferenciar informação de ruído. Sendo assim a PCA é bastante utilizada em diversas áreas do conhecimento humano, como na medicina em distúrbios cerebrais, ciências ambientais e agronomia como na análise de carbono em solo, análise de bebidas alcoólicas, dentre diversas outras aplicações (SHAFIZADEH-MOGHADAM, 2021; KHERIF *et al.*, 2020; CHU *et al.*, 2021).

Contudo, apesar da eficácia destas técnicas na exploração inicial de padrões, essas técnicas não são meios de classificação, pois não construímos um modelo estatístico de classificação, pois analisamos os dados sem qualquer fornecimento de informação inicial da

amostra, portanto é necessário se utilizar de métodos supervisionados, onde a partir destes, se constroem modelos de classificação, que são capazes de classificar uma nova amostra como sendo pertencente a uma determinada classe ou categoria de amostra.

3.3.3 Métodos supervisionados

Os métodos supervisionados, como os métodos de classificação, são caracterizados pela construção de um modelo quimiométrico capaz de classificar novas amostras em uma determinada categoria de amostra. Portanto o modelo construído é baseado em informações iniciais sobre o conjunto de dados, de forma que o modelo aprende características distintivas de cada classe e desenvolve a habilidade de reconhecer e classificar novos dados com base nesse aprendizado.

Geralmente, os métodos de classificação dividem o conjunto de dados em um conjunto de treinamento e outro de teste, assim como alguns métodos também se utilizam do conjunto de validação dependendo do contexto do conjunto de dados. O conjunto de treinamento é utilizado para construção do modelo e, portanto, é comum que represente maior parte do grupo amostral, pois ele é essencial para a construção e aprendizado do modelo. O conjunto de teste, por sua vez, é constituído por amostras que não fizeram parte do conjunto de treinamento, pois o objetivo é justamente avaliar a capacidade de generalização do modelo diante de dados não vistos anteriormente. Portanto, esses conjuntos são utilizados para criar e avaliar o desempenho dos modelos quimiométricos criados (SALONEN *et al.*, 2019).

3.3.4 Modelagem analógica de classes independentes baseada em dados (DD-SIMCA)

A modelagem analógica de classes independentes baseada em dados, deriva do método SIMCA. Essa é uma técnica de classificação multivariada que cria um modelo estatístico para cada classe ou categoria de amostra. Diferentemente de outros métodos de classificação, como PLS-DA e LDA, que estabelecem uma regra de classificação ou uma fronteira de decisão utilizando informações de todas as classes, o DD-SIMCA trabalha com modelos individuais por classe. Novas amostras são classificadas com base nesses modelos específicos. O SIMCA cria um modelo baseado em análise de componentes principais para cada classe, é importante destacar que esse modelo é criado baseando-se em amostras de classe única, ou seja, amostras que pertencem a um determinado grupo com características específicas. Logo, o SIMCA é um método importante para autenticação, sendo o número de componentes principais e o limite de decisão que determinam o desempenho do modelo (CHEN *et al.*, 2019).

Como o SIMCA cria um modelo estatístico para cada classe, ou seja, uma representação

matemática que descreve a relação entre as variáveis para cada conjunto de dados, logo, em um conjunto de dados onde se têm duas classes, dois modelos serão criados. Assim, em um conjunto de dados onde se têm duas classes de amostras, como amostras de cachaça produzida no estado da Paraíba e amostras de cachaça produzida em demais estados do Brasil, será criado um subconjunto de treinamento e de teste para cada classe, totalizando dois conjuntos de treinamento e dois subconjuntos de teste.

O SIMCA é uma ótima ferramenta para autenticação de amostras, pois permite que amostras sejam classificadas em categorias de amostras, atribuindo novas amostras a uma respectiva categoria. Cada classe de interesse é modelada independentemente, o que permite capturar a variabilidade específica de cada grupo. Isso facilita a identificação de amostras autênticas, já que o método se adapta a características particulares de cada classe (SUHANDY *et al.*, 2021).

O DD-SIMCA é uma técnica quimiométrica de classificação multivariada, que consiste em criar um único modelo estatístico para uma única classe, sendo essa a classe de interesse, portanto o DD-SIMCA é voltado para resolução de problemas onde se têm interesse em apenas uma única classe, logo o DD-SIMCA se mostra uma ferramenta útil para autenticar se uma nova amostra pertence à classe alvo ou não. Portanto, este tipo de técnica se mostra ideal levando em consideração os objetivos do presente trabalho, não havendo assim a necessidade de criar um modelo estatístico para as demais amostras de cachaça produzidas em demais estados brasileiros, mas apenas um único modelo para as cachaças produzida no estado da Paraíba (MAZIVILA *et al.*, 2020).

Assim, o DD-SIMCA cria um modelo estatístico para uma única classe alvo, e seu processo de classificação estabelece a possibilidade de a amostra ser ou não pertencente a classe, demonstrando assim ser uma ferramenta interessante em processos de autenticação. Para que qualquer processo de autenticação seja válido, é essencial que se tenha uma classe genuína previamente identificada, a classe é conhecida como classe-alvo e representa o foco central dos problemas de autenticação. Logo devido a estas características da técnica, qualquer outra categoria de amostra que se encontre fora dos padrões de variabilidade da classe alvo, é considerada não pertencente a classe modelada, portanto dependendo do contexto do problema, essa amostra pode ser considerada adulterada, falsa ou apenas pertencente a uma outra classe (KUCHERYAVSKIY *et al.*, 2024).

Após a criação do modelo estatístico para a classe-alvo, realiza-se a classificação e avaliação para garantir sua confiabilidade. O conjunto de teste é submetido ao modelo treinado, calcula-se a distância da nova amostra em relação à classe-alvo, define-se a zona de aceitação

e, então, as amostras são classificadas como pertencentes ou não à classe. A qualidade da classificação, depende da representatividade das amostras usadas para calibração do modelo, que é um aspecto essencial a ser considerado em estudos de autenticação (ANGELISS *et al.*, 2024; NEVES, *et al.*, 2020).

Dentre as suas aplicações de autenticação, encontram-se trabalhos na literatura sobre diferenciação e identificação geográfica de bebidas alcoólicas, assim como consta no trabalho realizado por (GOMES *et al.*, 2021) onde foi realizado um estudo de autenticação de vinhos Tokaj, devido ao problema de que foram encontrados vinhos Tokaj falsificados, o estudo se utilizou de técnica espectroscópica e modelo DD-SIMCA para resolução do problema. Onde o modelo DD-SIMCA apresentou um desempenho de classificação correta de 97% e apresentou alta especificidade de 100% (GOMES *et al.*, 2021).

Levando em consideração esses fatores, o DD-SIMCA se configura uma técnica adequada em processos de autenticação, devido a sua robustez, capacidade de classificação e flexibilidade, tendo assim uma ampla aplicação em diversos setores e nos mais diferentes tipos de problemas (RODIONOVA *et al.*, 2023).

4 MATERIAL E MÉTODOS

4.1 Amostras

O presente trabalho utilizou um total de 79 amostras de cachaça para análise. Destas, 34 amostras foram produzidas no estado da Paraíba, enquanto as outras 45 foram obtidas através de misturas de amostras de cachaça de diferentes estados brasileiros, abrangendo assim uma variedade de condições geográficas e métodos de produção, permitindo assim uma comparação mais ampla e detalhada.

As amostras de cachaça do estado da Paraíba foram adquiridas em engenhos locais. Após a coleta, algumas delas foram adquiridas em seus recipientes originais e outras foram transferidas para em frascos de vidro âmbar, com o objetivo de proteger o conteúdo da exposição à luz e evitando possíveis alterações químicas, mantendo-se assim armazenadas até a aquisição dos espectros.

As 45 amostras de cachaça que pertencem à classe de amostra produzidas em demais estados brasileiros foram obtidas através de uma mistura de 8 marcas específicas de cachaças de demais estados brasileiros. O objetivo dessas misturas foi obter uma boa representatividade das amostras para construção dos modelos quimiométricos de classificação.

4.2 Aquisição dos espectros NIR

Os espectros NIR foram adquiridos em duplicata a partir de alíquotas utilizando um espectrômetro NIR portátil modelo NIR-S-T2 da marca InnoSpectra operando no modo de absorbância, onde a faixa espectral analisada variou de 900 a 1700 nm. Durante a coleta dos dados, a resolução digital do equipamento foi de 228 pontos, e a largura de banda utilizada corresponde a uma faixa variável de 2,34 a 7,03 nm. A cubeta utilizada foi uma cubeta de quartzo de 1 mm, pois proporcionam maior sensibilidade e precisão nas medições, permitindo a análise de amostras com menor volume.

4.3 Aquisição dos espectros Raman

Os espectros Raman foram obtidos utilizando um Raman portátil mira p da Metrohm cujo laser de excitação tem comprimento de onda específico de 785 nm, comprimento de onda este utilizado para minimização da fluorescência e uma melhor obtenção do sinal Raman. A faixa espectral para o deslocamento Raman obtida foi de 400-2300 cm^{-1} para as amostras de cachaça, faixa esta que abrange regiões importantes para a identificação de vibrações moleculares característica da bebida.

4.4 Pré-processamento dos dados

Inicialmente, os espectros Raman e NIR passaram por um processo de remoção de regiões consideradas ruidosas. Nos espectros Raman, foram eliminadas as 782 variáveis finais do espectro, enquanto nos espectros NIR foram removidas as 10 primeiras e as 32 últimas variáveis, correspondentes às extremidades do espectro. Esse processo é importante, pois essas regiões podem mascarar informações relevantes e afetar negativamente a qualidade dos dados, portanto remover essas regiões elimina contribuições desnecessárias para a análise.

O tratamento inicial dos dados consistiu em realizar diversos pré-processamentos com o intuito de obter a melhor preparação inicial para o processo de classificação, dentre esses processamentos foi realizado o pré-processamento correção multiplicativa de espalhamento (MSC, do inglês *Multiplicative Scatter Correction*), variação normal padrão (SNV, do inglês *Standard Normal Variate*), correção de linha de base (BC, do inglês *Baseline Correction*) variando parâmetros como Offset, Linear e Offset-Linear, e derivação com o filtro Savitzky-Golay (SG), variando os tamanhos de janela, o número de pontos e os graus dos polinômios.

4.5 Construção do modelo PCA

A construção dos modelos PCA para os dados NIR e Raman foi realizada no software MATLAB (versão R2019a), amplamente utilizado para análise e manipulação de dados multidimensionais. Para isso, empregou-se o pacote PCA Toolbox (versão 1.5), ferramenta especializada que oferece recursos para execução de PCA e outras técnicas de análise exploratória de dados não supervisionados. Os modelos foram desenvolvidos a partir de dados centralizados na média, procedimento adequado para este tipo de análise, pois corrige diferenças nas médias das variáveis, que já se encontram na mesma escala, e permite concentrar a interpretação nas variações numéricas relevantes.

A PCA foi aplicada diretamente sobre os dados originais, sem pré-processamento, tanto para o conjunto NIR quanto para o conjunto Raman. Essa abordagem possibilita explorar as características dos espectros em seu estado bruto, identificando padrões, variações principais e potenciais agrupamentos sem influência de transformações adicionais.

4.6 Métodos de classificação

Os modelos de classificação DD-SIMCA para ambos os dados NIR e Raman foram feitos através do software MATLAB, além disso foi utilizado o pacote DD-SIMCA versão 1.2 amplamente utilizado em modelagem e análise de dados em quimiometria (ZONTOV *et al.*, 2017).

O conjunto de treinamento foi construído com 75% das amostras da classe alvo (25 amostras), enquanto o conjunto de teste incluiu os 25% restantes (9 amostras). Essas 9 amostras foram combinadas com outras 45 provenientes de diferentes regiões, totalizando 54 amostras para o teste. A seleção das amostras para o conjunto de treinamento foi realizada utilizando o algoritmo Kennard-Stone, por meio do pacote computacional *Data_Hand_Gui*, desenvolvido pelo professor Adriano de Araújo Gomes. Esse método é amplamente utilizado para escolher amostras de forma sistemática, garantindo a representatividade dos conjuntos e minimizando o viés de uma seleção arbitrária.

A construção dos modelos de predição utilizou diferentes números de componentes principais, conforme a natureza dos dados de cada pré-processamento, com o objetivo de encontrar o número ideal de componentes que proporcionasse os melhores resultados. Foi adotado um nível de significância de 5% (0,05) para a detecção de outliers e para o erro tipo I (falso positivo) nos modelos de predição. Para o conjunto de teste, considerou-se um nível de 5% para o erro tipo II (falso negativo). Assim, para cada classificação individual positiva ou negativa, estabelece-se uma probabilidade máxima de 5% de cometer erro tipo I ou tipo II.

5 RESULTADOS E DISCUSSÃO

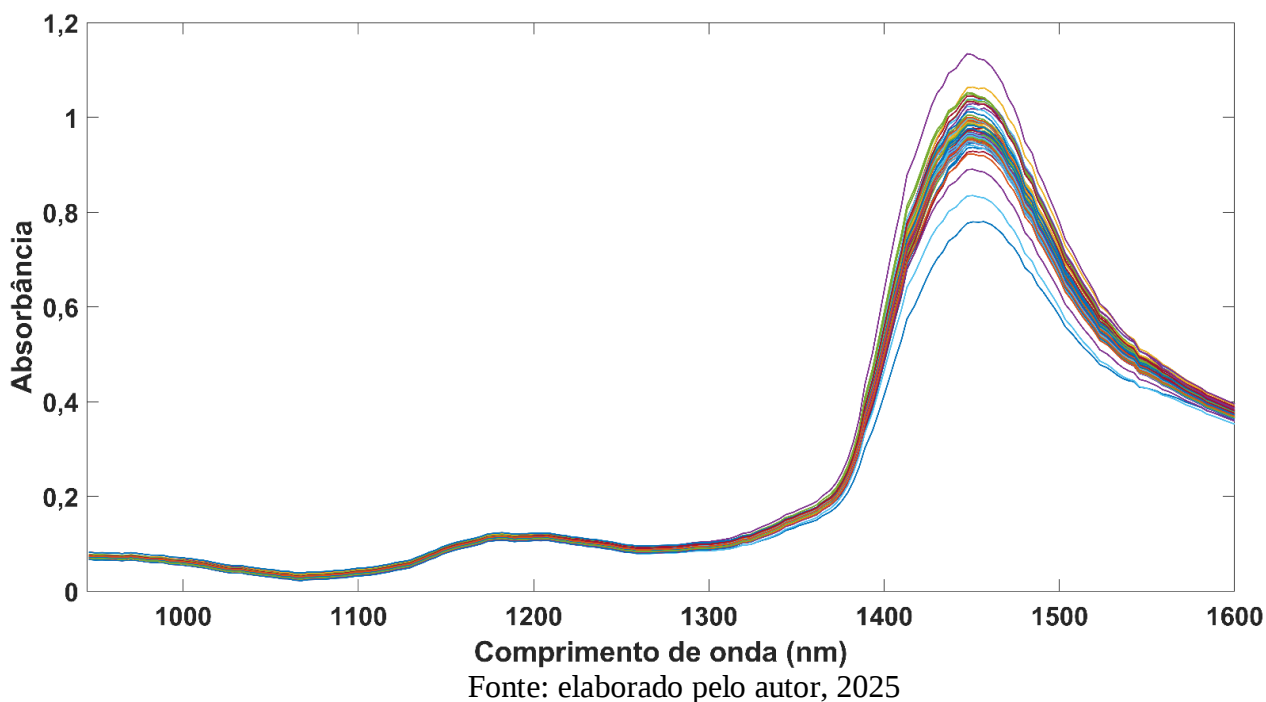
5.1 Resultados utilizando a espectroscopia NIR

5.1.1 Pré-tratamento dos dados espectrais NIR

Os espectros NIR adquiridos com o espectrômetro NIR portátil na região de 945 a 1600 nm de ambas as cachaças pode ser visualizado na **Figura 4**. A região trabalhada é relevante para a caracterização das cachaças, permitindo a identificação de bandas características da bebida, ponto este essencial para a discriminação entre as classes.

Ao trabalhar com a média espectral, obtida a partir de espectros adquiridos em duplicata, triplicata ou mais, reduz-se o impacto de erros aleatórios presentes em espectros individuais, decorrentes de variações instrumentais, condições ambientais e outros fatores. Esse procedimento resulta em um espectro mais confiável e representativo da amostra.

Figura 4: Espectros brutos NIR na faixa de 945 a 1600 nm



O espectro NIR obtido para as amostras de cachaça das duas classes analisadas revela uma série de características importantes que refletem a composição química das bebidas. A partir da análise do espectro, verifica-se que as bandas de absorção estão distribuídas ao longo do espectro, com picos específicos que correspondem às vibrações moleculares de diferentes compostos presentes nas cachaças.

Esses picos são indicativos de grupos funcionais de diversas espécies químicas, os quais

estão presentes em concentrações variáveis nas amostras. A partir da análise das bandas do espectro é possível avançar para uma descrição mais detalhada das bandas características que ajudam a explicar as variações observadas e a relação com as propriedades químicas das cachaças.

De acordo com o espectro NIR obtido, é possível observar uma banda em torno de 1200 nm das amostras de cachaça, onde essa banda está associada a uma combinação de vibrações envolvendo ligações C–H e o segundo sobretom da ligação O–H. Essa região espectral é sensível à presença de grupos metila ($-\text{CH}_3$) e metileno ($-\text{CH}_2-$), comuns em compostos orgânicos como álcoois. A interação entre essas vibrações indica contribuições tanto do etanol quanto de outros componentes menores da matriz alcoólica, refletindo características moleculares importantes para a diferenciação entre amostras (LIU et al., 2004; WEYER et al., 2002).

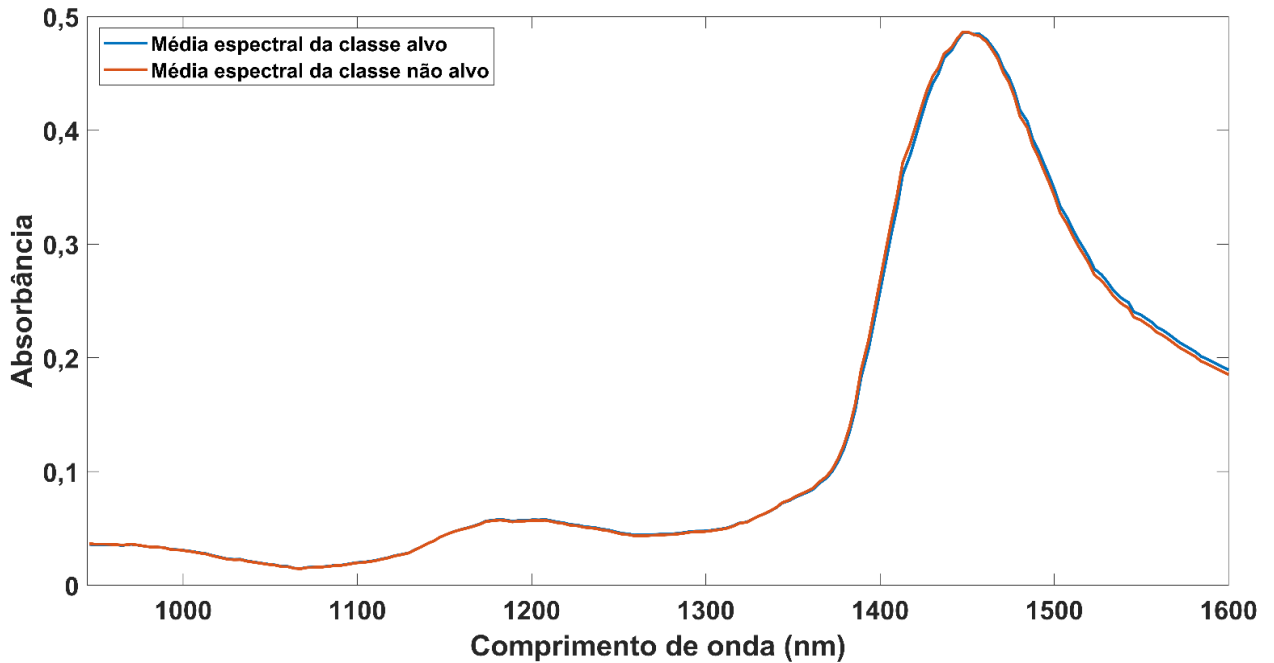
Também é possível observar uma intensa banda registrada na região de aproximadamente 1450 nm, onde essa banda é atribuída ao primeiro sobretom da vibração de estiramento da ligação O–H. Essa absorção é típica de substâncias que contêm grupos hidroxila, como a água e etanol, principais constituintes da cachaça. Devido à sua intensidade e sensibilidade a interações intermoleculares, como ligações de hidrogênio, essa banda fornece informações relevantes sobre a composição e estrutura das amostras, além disso, por estar diretamente relacionada à presença de grupos O–H, essa região espectral tem sido amplamente utilizada na determinação do teor alcoólico em bebidas, já que a intensidade da absorção em torno de 1450 nm varia conforme a concentração de etanol na matriz (CHO et al., 2005; WEYER et al., 2002).

Para facilitar a visualização da sobreposição dos espectros, foi calculada a média dos espectros das cachaças de cada classe, portanto de acordo com a **Figura 5**, a linha azul representa o espectro das cachaças da Paraíba enquanto a linha vermelha representa as cachaças de demais regiões. É possível observar que não existe uma distinção muito clara dos espectros das cachaças da Paraíba e de demais estados brasileiros, pois as suas diferenças são sutis, essa característica impede a identificação visual direta das diferenças entre as classes e destaca a complexidade dos dados obtidos. Esta sobreposição espectral pode ser atribuída à similaridade química das cachaças, pois apesar das cachaças terem sido produzidas em diferentes condições, elas possuem componentes em comum.

Diante disso, a aplicação de técnicas quimiométricas torna-se indispensável, pois essas metodologias permitem explorar a variabilidade presente nos dados espectrais, uma tarefa praticamente impossível de ser realizada a olho nu, identificando assim padrões e diferenças.

Técnicas como PCA e métodos de classificação supervisionados como DD-SIMCA, são ferramentas fundamentais para a análise de espectros que apresentam sobreposição.

Figura 5: Média dos espectros NIR por classe na faixa de 945 a 1600 nm

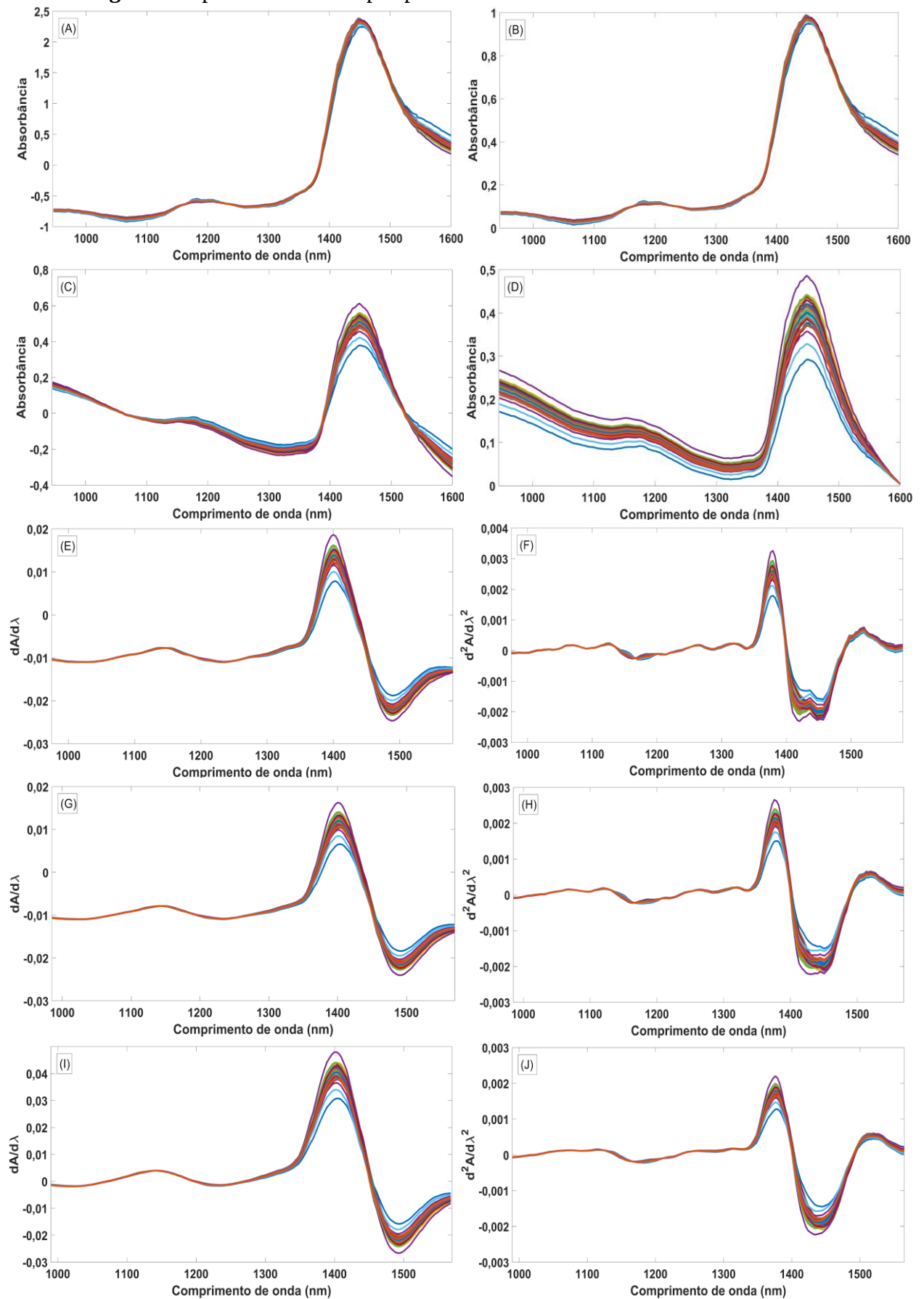


Fonte: elaborado pelo autor, 2025

Dessa forma, o pré-processamento dos dados na região do infravermelho próximo (NIR) é uma etapa fundamental para minimizar variações indesejadas. A **Figura 6** apresenta uma série de espectros submetidos a diferentes tratamentos: (A) SNV, (B) MSC, (C) BC-(Linear), (D) BC-(Offset-Linear), (E) SG-(15,1,1), (F) SG-(15,2,2), (G) SG-(19,1,1), (H) SG-(19,2,2), (I) SG-(23,1,1) e (J) SG-(23,2,2). Os parâmetros utilizados no pré-processamento Savitzky-Golay (SG) são indicados entre parênteses, em ordem: tamanho da janela, ordem da derivada e grau do polinômio.

O pré-processamento BC-(Offset) apresentou os melhores resultados entre os pré-processamentos utilizados para os dados NIR, como pode ser visualizado na **Figura 7**. Esse procedimento corrigiu deslocamentos verticais nos espectros, alinhando-os e permitindo uma comparação mais consistente entre as amostras. A aplicação do BC-(Offset) proporcionou uma visualização mais clara das características espectrais das cachaças, destacando diferenças químicas reais entre as amostras. Como resultado, os espectros corrigidos forneceram uma base mais confiável para análises e classificações mais precisas.

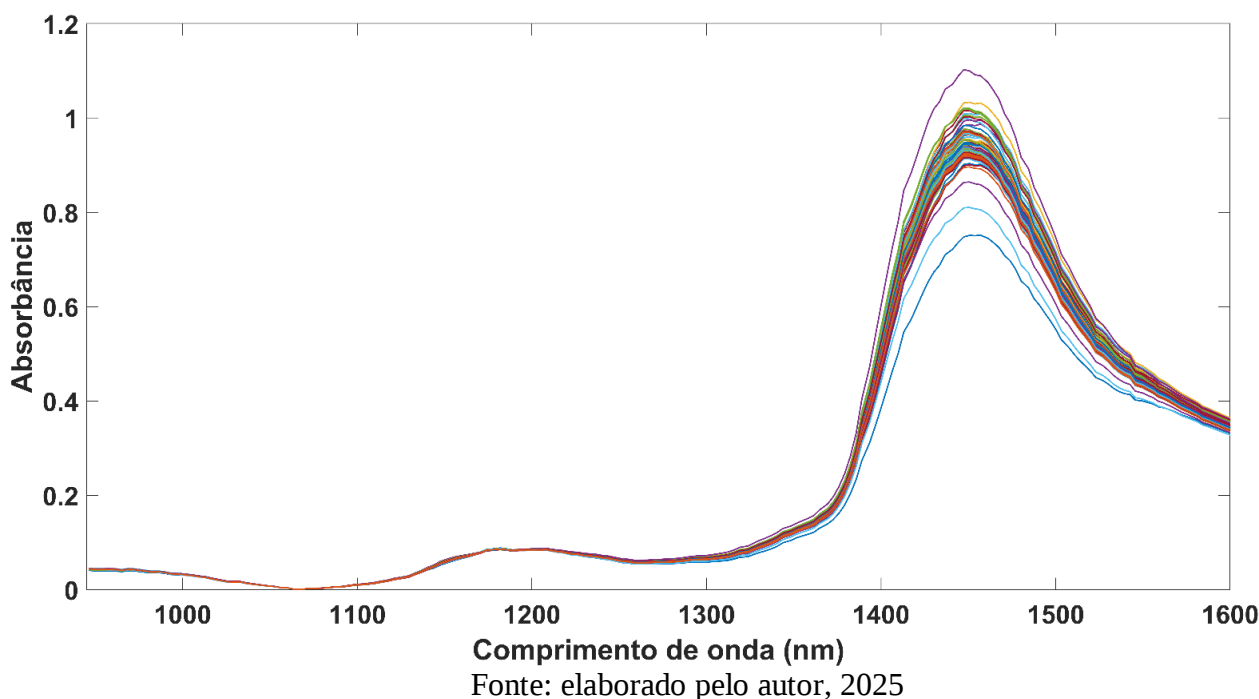
Figura 6 Espectros NIR com pré-processamentos



Fonte: elaborado pelo autor, 2025

Além disso, o pré-processamento de correção de linha de base facilitou a identificação de padrões químicos específicos em cada classe de cachaça, permitindo uma distinção mais nítida entre as amostras. A correção dos deslocamentos verticais melhorou a qualidade dos dados e contribuiu para a construção de modelos analíticos mais robustos, que proporcionam previsões mais confiáveis das características das diferentes classes.

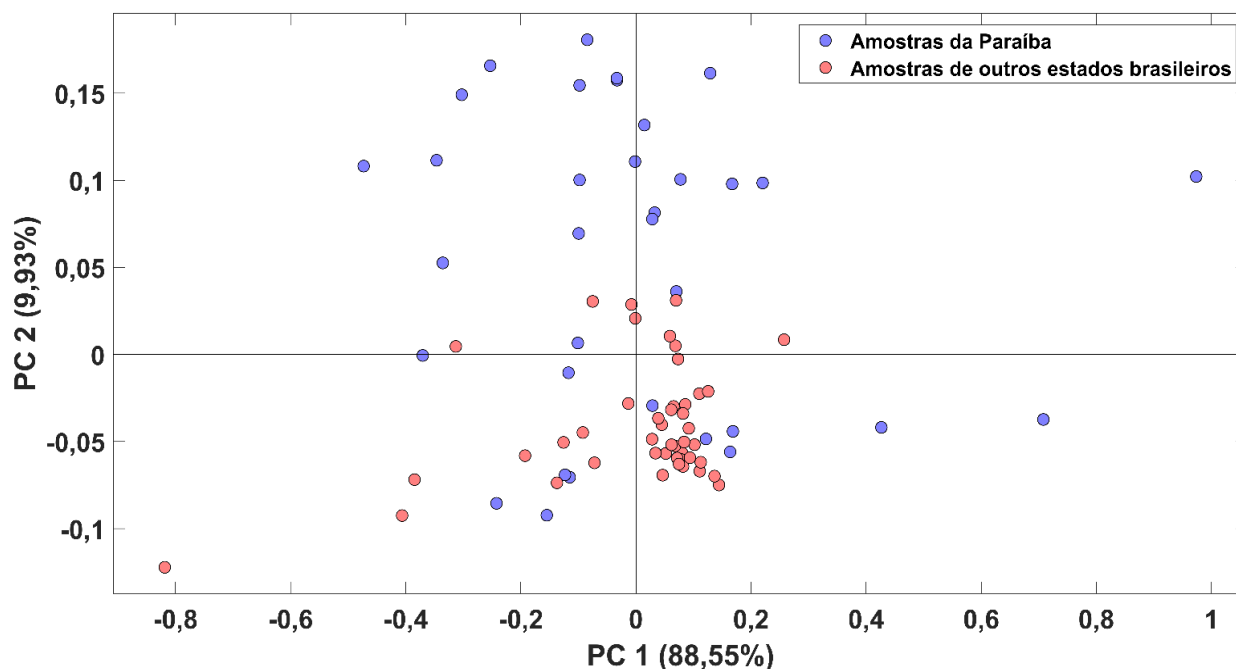
Figura 7: Espectros NIR com pré-processamento BC-(Offset)



5.1.2 Análise exploratória de dados utilizando PCA

A análise de componentes principais apresentada na **Figura 8** evidencia uma pequena tendência de agrupamento entre as classes, mostrando visualmente uma separação parcial no gráfico de score entre as cachaças da Paraíba e as cachaças de outras regiões, permitindo visualizar diferenças e padrões entre as amostras. A componente principal 1 (PC1) explica 88,55% da variância total dos dados, enquanto a componente principal 2 (PC2) explica 9,93%, totalizando 98,48% da variância explicada pelos dois primeiros componentes.

Essa alta variância capturada sugere que as características utilizadas na análise são altamente informativas para descrever a variabilidade das amostras. No entanto, a sobreposição entre os dois grupos de cachaças indica que, embora exista alguma diferenciação entre as amostras geográficas, essa diferenciação não é absoluta.

Figura 8: Gráfico de escores do modelo PCA

Fonte: elaborado pelo autor, 2025

Além disso, observa-se no gráfico de scores que a componente principal PC2 desempenha um papel essencial na distinção parcial das amostras de cachaça provenientes da Paraíba em relação às cachaças provenientes de outras regiões. Essa tendência de agrupamento destaca a importância de PC2 como um fator determinante na distinção entre os dois grupos de amostras. Esse comportamento indica que PC2 captura as variações nos dados que estão fortemente relacionadas às características distintas das cachaças do estado paraibano. Tais características podem incluir fatores ambientais específicos, como condições climáticas e a composição do solo da região, que influenciam diretamente as propriedades sensoriais e químicas da cachaça produzida. Além disso, a técnica de produção, que pode envolver métodos tradicionais ou inovadores, portanto, todos estes fatores podem ser elementos cruciais que contribuem para as variações observadas.

Esse resultado ressalta a complexidade no processo de autenticação geográfica das cachaças e destaca a importância de explorar diferentes técnicas quimiométricas para aprimorar a diferenciação entre os grupos. Para avançar nessa direção, é necessário utilizar um método de classificação supervisionado como DD-SIMCA, permitindo assim uma análise mais detalhada das interações entre as variáveis e potencialmente melhorando a acurácia na distinção das cachaças de diferentes localidades.

5.1.3 Análise classificatória utilizando DD-SIMCA

Para avaliar a eficiência do método DD-SIMCA na análise classificatória, foram testados diferentes pré-processamentos aplicados aos dados espectrais, com o objetivo de otimizar a discriminação entre as classes e reduzir possíveis interferências de ruído, assim como obter o melhor pré-processamento para o conjunto de dados. Os pré-processamentos utilizados estão descritos na **Tabela 1**, que também apresenta os parâmetros de sensibilidade, especificidade e acurácia, além das métricas de classificação (VP – verdadeiro positivo, FN – falso negativo, VN – verdadeiro negativo e FP – falso positivo).

Tabela 1: Desempenho dos diferentes pré-processamentos em termos de métricas de classificação. dos dados NIR.

Pré-processamentos	Sensibilidade (%)	Especificidade (%)	Acurácia (%)	VP	FN	VN	FP
Dados brutos	77,80	62,20	64,80	7	2	28	17
MSC	44,40	71,10	66,70	4	5	35	10
SNV	44,40	68,90	64,80	4	5	31	14
BC-(Offset)	88,90	60,00	64,80	8	1	27	18
BC-(Linear)	88,90	40,00	48,10	8	1	18	27
BC-(Offset-Linear)	44,40	33,30	35,20	4	5	15	30
SG-(15,1,1)	88,90	44,40	51,90	8	1	20	25
SG-(15, 2, 2)	77,80	44,40	50,00	7	2	20	25
SG-(19, 1, 1)	100,00	42,20	51,90	9	0	19	26
SG-(19, 2, 2)	88,90	42,20	50,00	8	1	19	26
SG-(23,1,1)	88,90	33,30	42,60	8	1	15	30
SG-(23,2, 2)	77,80	46,70	51,90	7	2	21	24

Fonte: elaborado pelo autor, 2025

A análise de classificação utilizando o método DD-SIMCA foi realizada com diferentes pré-processamentos (**Tabela 1**), e os resultados obtidos em termos de sensibilidade, especificidade e acurácia revelaram padrões distintos de desempenho.

Entre os pré-processamentos utilizados, a técnica BC-(Offset) destacou-se por apresentar um bom equilíbrio entre os resultados obtidos. Esse método alcançou uma sensibilidade de 88,90%, uma especificidade de 60,00% e uma acurácia de 64,80%, sendo a sensibilidade uma das mais altas entre os pré-processamentos avaliados. Esses valores indicam que o BC-(Offset) apresenta boa capacidade de identificar corretamente os verdadeiros positivos, ao mesmo tempo em que mantém níveis satisfatórios de especificidade e acurácia, configurando-se como uma opção robusta para a discriminação.

Os pré-processamentos MSC e SNV apresentaram resultados semelhantes, com

sensibilidades de 44,40%, especificidades de 71,10 e 68,90%, respectivamente, e acurácia de 66,70 e 64,80%. Esses valores indicam que ambos os métodos tiveram desempenho próximo, destacando-se por uma boa especificidade, embora com sensibilidade relativamente baixa.

Os pré-processamentos baseados em derivações apresentaram um comportamento semelhante, destacando-se a configuração utilizada com uma janela de suavização de 19 pontos, derivada de primeira ordem e polinômio de grau 1 alcançou uma sensibilidade máxima de 100%, indicando que todos os dados da classe foram corretamente identificados. Contudo, sua especificidade foi limitada a 42,20%, o que resultou em uma acurácia de 51,90%, sugerindo maior incidência de falsos positivos.

Por fim, os dados brutos, sem qualquer pré-processamento, apresentaram sensibilidade de 77,80%, especificidade de 62,20% e acurácia de 64,80%. Esse desempenho foi semelhante ao de algumas abordagens testadas, mas ainda com sensibilidade inferior à observada no método BC-(Offset).

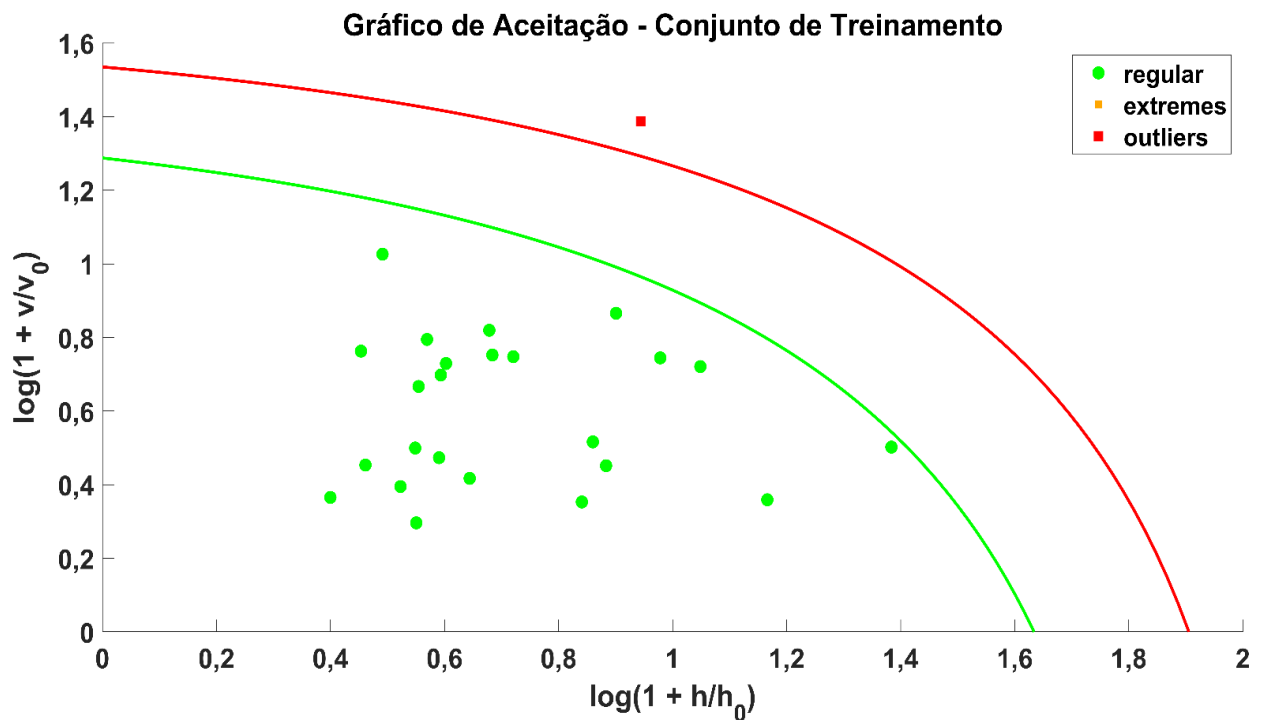
Em suma, o pré-processamento BC-(Offset) demonstrou ser a melhor abordagem geral, equilibrando especificidade, sensibilidade e acurácia, enquanto métodos como MSC e SNV exibiram um bom desempenho em critérios específicos, mas com limitações em outros. Esses resultados ressaltam a importância de selecionar o pré-processamento adequado para otimizar o desempenho do DD-SIMCA em diferentes cenários.

Assim como, é possível observar que, com o pré-processamento BC-(Offset), foram obtidos 8 verdadeiros positivos (VP), o que indica que oito amostras da Paraíba foram corretamente identificadas como pertencentes à classe alvo. No entanto, o modelo apresentou 18 falsos positivos (FP), ou seja, amostras de outras localidades classificadas incorretamente como sendo da Paraíba. Esse resultado evidencia que, embora o método tenha alcançado elevada capacidade de identificação das amostras positivas, apresentou menor especificidade, refletindo dificuldade em distinguir adequadamente as amostras negativas.

Além disso, o modelo classificou 27 amostras como verdadeiros negativos (VN), ou seja, corretamente reconhecidas como pertencentes a outras localidades. Apenas uma amostra real da Paraíba foi classificada incorretamente como negativa, resultando em um único falso negativo (FN). Esse resultado sugere que essa amostra paraibana possui características espectrais semelhantes às de cachaças de outras regiões, o que pode comprometer sua discriminação pelo modelo. A análise da matriz de confusão demonstra que, embora o modelo seja eficiente na identificação da maioria das amostras da classe alvo (alto VP), ainda enfrenta desafios na redução de falsos positivos, reforçando a necessidade de ajustes no modelo ou do uso de técnicas complementares para aprimorar sua capacidade de discriminação.

A **Figura 9** apresenta o gráfico de treinamento utilizando o pré-processamento BC- (Offset), onde é possível evidenciar que de acordo com o gráfico de treinamento, das 25 amostras submetidas a construção do modelo, apenas 1 amostra foi considerada outlier, o que indica uma maior variabilidade ou comportamentos atípicos em relação às demais amostras, mas a maioria das amostras no gráfico está dentro da região verde, indicando que o modelo é representativo para o conjunto de treinamento.

Figura 9: Gráfico da área de aceitação do modelo DD-SIMCA para o conjunto de treino pré-processamento BC-(Offset) dos dados NIR



A presença da amostra que foi considerada como outlier sugere que ela apresenta características espectrais distintas em relação às demais amostras do grupo de treinamento, ou ainda que tenha sido afetada por ruídos experimentais, o que a diferencia do padrão estabelecido pelo modelo.

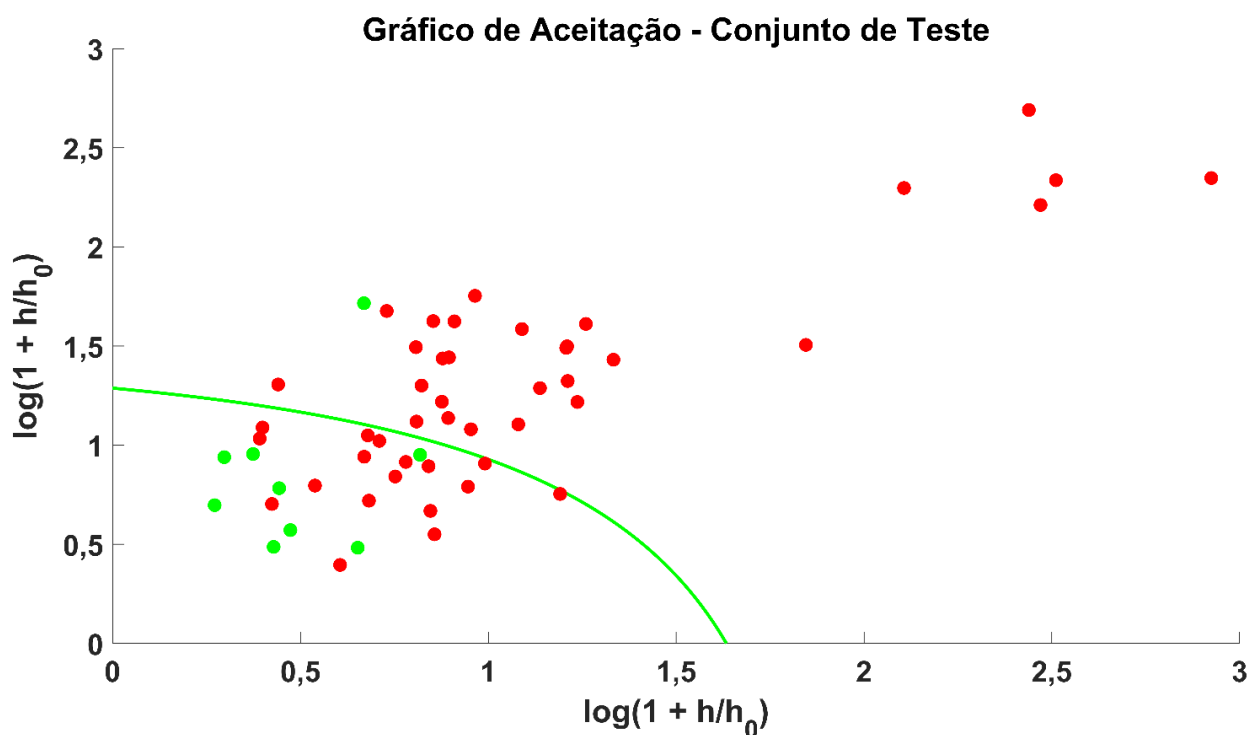
Essa amostra pode representar um caso isolado ou uma situação com características distintas, mas que não compromete de forma significativa a capacidade do modelo em classificar corretamente as classes. A presença de amostras extremas ou atípicas não é incomum em modelos de classificação, sobretudo quando o conjunto de dados apresenta variações naturais ou pequenas dissonâncias entre as características das classes. Nesses contextos, um outlier pode ser definido como um indivíduo cujas características se afastam consideravelmente

das predominantes na maioria dos membros da mesma classe. A ocorrência de outliers em dados experimentais é relativamente comum e pode estar associada a diferentes fatores, como variações naturais, erros de medição, ruídos instrumentais ou preparo inadequado das amostras (BALDOMERO-NARANJO *et al.*, 2021).

Contudo é importante destacar que a maior parte das amostras está posicionada dentro da região verde, indicando que essas amostras foram corretamente classificadas pelo modelo e estão bem representadas na análise.

A **Figura 10** apresenta o gráfico de teste em um espaço de aceitação definido pelo modelo, considerando as variáveis Q-residual (eixo y) e Hotelling T^2 (eixo x) em relação à estrutura modelada. A curva verde representa a região de aceitação para amostras da classe alvo, enquanto aquelas localizadas fora dessa região são classificadas como pertencentes à classe não alvo.

Figura 10: Gráfico da área de aceitação do modelo DD-SIMCA para o conjunto de teste com pré-processamento BC-(Offset) dos dados NIR



A análise dos resultados evidencia de forma clara uma tendência consistente de separação entre as classes, o que indica que o modelo de classificação conseguiu capturar padrões relevantes nos dados espectrais. Observa-se que 60% das amostras pertencentes à classe não alvo foram corretamente classificadas, o que já demonstra uma taxa de

reconhecimento significativa para esse grupo. Por outro lado, as amostras da classe alvo apresentaram um desempenho ainda mais expressivo, uma vez que foram predominantemente posicionadas dentro da região delimitada pela curva de aceitação. Isso resultou em 88,90% das amostras classificadas corretamente, evidenciando a elevada capacidade discriminatória do modelo para essa categoria específica.

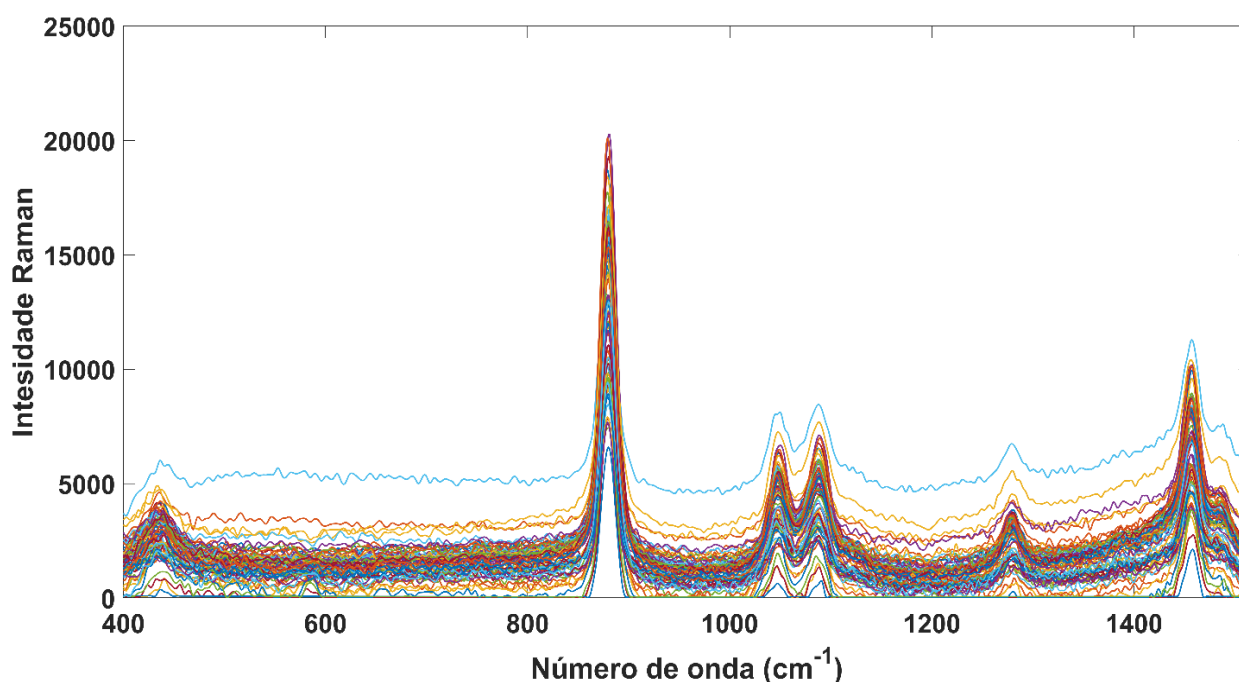
Observa-se uma separação considerável entre as diferentes classes que demonstra a capacidade do modelo em distinguir padrões específicos associados a cada grupo. Esse desempenho reflete a robustez do método na identificação de variações atribuídas tanto às características geográficas das regiões produtoras e dentre outras características. Essa distinção é essencial para garantir a confiabilidade do modelo na classificação das amostras, destacando sua aplicabilidade na autenticação e diferenciação de cachaças com base em seus perfis espectrais.

Portanto, o gráfico de treinamento e de teste obtido a partir do pré-processamento BC- (Offset), aplicado no modelo de classificação DD-SIMCA, mostra-se fundamental, uma vez que fornece uma visão abrangente e detalhada do desempenho do modelo durante o processo de distinção entre as classes presentes no conjunto de dados. Esse recurso gráfico permite observar de forma visual a eficiência da separação, evidenciando quais amostras foram aceitas ou rejeitadas dentro da região de confiança estabelecida pelo modelo. Além disso, sua interpretação auxilia na avaliação da robustez do método, possibilitando verificar se a taxa de acerto é consistente tanto na etapa de calibração quanto na de validação, o que garante maior confiabilidade e aplicabilidade prática dos resultados. Dessa forma, o uso combinado de pré-processamento adequado e métodos de classificação robustos, como o DD-SIMCA, reafirma-se como uma estratégia eficiente para a autenticação geográfica de cachaças com base em seus perfis espectrais.

5.2 Resultados utilizando a espectroscopia Raman

5.2.1 Pré-tratamento dos dados espectrais Raman

A **Figura 11** apresenta a análise dos espectros Raman, onde o espectro fornece uma visão geral das bandas espectrais características das amostras da Paraíba e de outras regiões. As assinaturas espectrais revelam informações importantes sobre os componentes químicos presentes nas cachaças, permitindo uma visão preliminar das variações que podem ser associadas às diferentes origens das amostras. A análise deste espectro abre caminho para discussões sobre a aplicação de técnicas espectroscópicas na diferenciação das cachaças, assim como uma discussão mais aprofundada sobre as características dessas cachaças.

Figura 11: Espectros Raman na faixa de 400 a 1518 cm^{-1} 

Fonte: elaborado pelo autor, 2025

Esses compostos apresentam bandas vibracionais características em regiões específicas do espectro Raman, sendo essas bandas fundamentais para a identificação de componentes-chave nas cachaças. A correta interpretação dessas regiões espectrais, aliada a técnicas avançadas de análise de dados, pode possibilitar uma melhor diferenciação entre as amostras, contribuindo para a caracterização e autenticação da bebida com base em sua assinatura espectral.

Dentre as bandas do espectro Raman, a banda em torno de 880 cm^{-1} é mais intensa no espectro, e esta banda está associada ao modo de estiramento simétrico das ligações C–C e C–O do etanol. Sua presença marcante indica a predominância do etanol como constituinte principal da bebida. Além disso, sua intensidade pode ser usada como parâmetro para monitorar a concentração de etanol, sendo frequentemente utilizada para quantificação em análises vibracionais (MENEVŞEOĞLU *et al.*, 2021).

A banda localizada próxima a 1045 cm^{-1} é atribuída à vibração de estiramento da ligação C–O, que é comum tanto no etanol quanto em álcoois superiores. Esta região é importante para identificar e diferenciar diferentes tipos de álcoois na bebida, sendo útil também para avaliar a pureza do etanol. A intensidade e forma desta banda podem indicar a presença de impurezas orgânicas voláteis, assim como o pico duplo entre 1000 e 1100 cm^{-1} pode ser atribuído aos estiramentos simétricos e assimétricos de CO do etanol com contribuições dos modos de oscilação CH (CORSETTI *et al.*, 2016).

A banda em torno de 1280 cm^{-1} no espectro Raman está associada à deformação angular (wagging) do grupo CH_2 , típica de álcoois e ácidos orgânicos presentes em baixas concentrações na cachaça. Esses compostos contribuem para o perfil sensorial da bebida e são influenciados por fatores como fermentação, tipo de levedura e destilação. Por isso, essa faixa espectral pode ser usada como indicador químico em estudos de autenticação geográfica e controle de qualidade (ULLAL *et al.*, 2024).

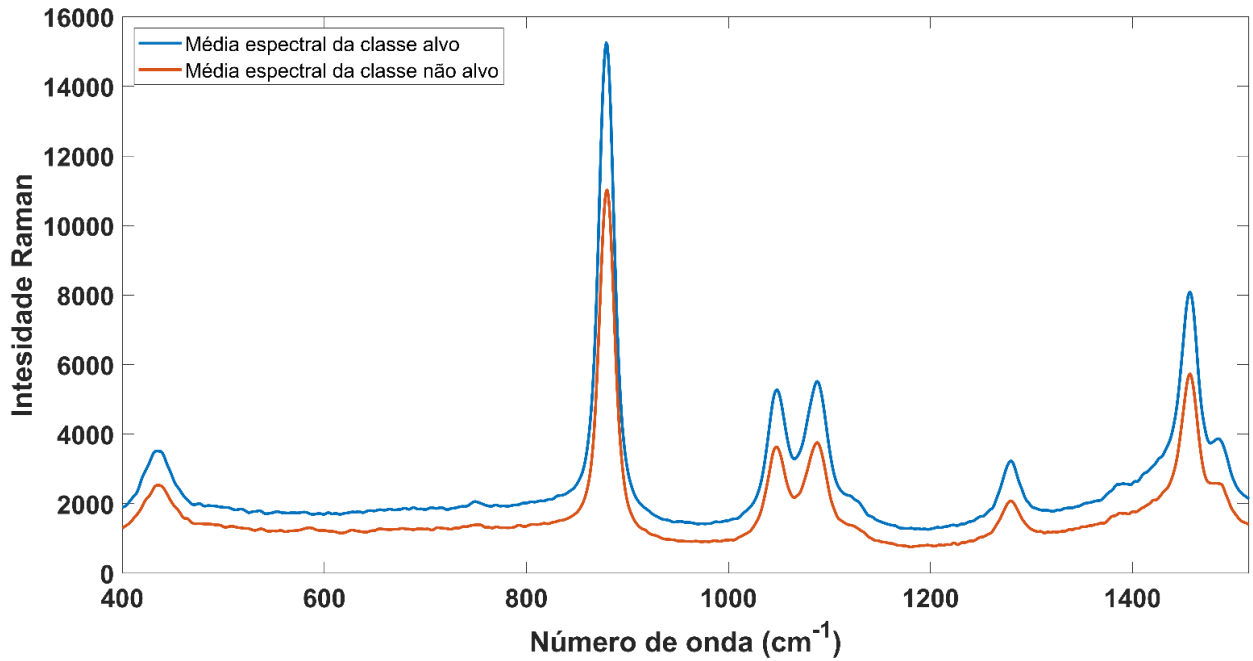
A segunda banda de maior intensidade do espectro corresponde à região em torno de 1450 cm^{-1} . Essa banda está associada à vibração de deformação angular (bending) dos grupos metila (CH_3) e metileno (CH_2), comuns em uma ampla variedade de compostos orgânicos. Sua presença constante e bem definida nos espectros Raman da cachaça indica a contribuição do etanol, principal constituinte da bebida, bem como de compostos secundários como álcoois superiores, ésteres e outros voláteis formados durante a fermentação e a destilação. Por isso, essa faixa pode ser utilizada como um indicador químico complementar na avaliação da qualidade do destilado e na diferenciação de perfis químicos entre amostras (PAPASPYRIDAKOU *et al.*, 2023).

Os espectros Raman das amostras de cachaça apresentaram uma significativa sobreposição entre as classes, o que evidencia a complexidade química da bebida e a dificuldade em distinguir visualmente as classes de cachaças de forma direta. Essa sobreposição é comum pois a composição química das cachaças envolve uma série de compostos que compartilham assinaturas espectrais semelhantes, tornando assim o processo de identificação e discriminação das diferentes classes um processo mais desafiador.

Para facilitar a interpretação dos resultados, foi calculada a média espectral das amostras pertencentes a cada classe. Na **Figura 12**, a linha azul representa a média espectral das cachaças da Paraíba (classe alvo), enquanto a linha vermelha corresponde às cachaças provenientes de outras regiões (classe não alvo). Diferentemente do que foi observado para os espectros no NIR, onde há uma sobreposição mais acentuada entre as classes, no espectro Raman é possível perceber uma diferenciação um pouco mais evidente entre os perfis espectrais, sobretudo na intensidade relativa de alguns picos característicos.

Embora os espectros médios Raman apresentem menor sobreposição entre classes em relação ao NIR, essa diferença não é suficiente para uma distinção clara por inspeção visual. As variações são sutis e distribuídas ao longo do espectro, o que dificulta a interpretação direta. Assim, torna-se essencial o uso de técnicas de classificação multivariada, capazes de explorar simultaneamente múltiplas variáveis e realçar padrões complexos, ampliando o poder discriminatório dos dados Raman.

Figura 12: Média dos espectros Raman por classe na faixa de 400 a 1518 cm^{-1}

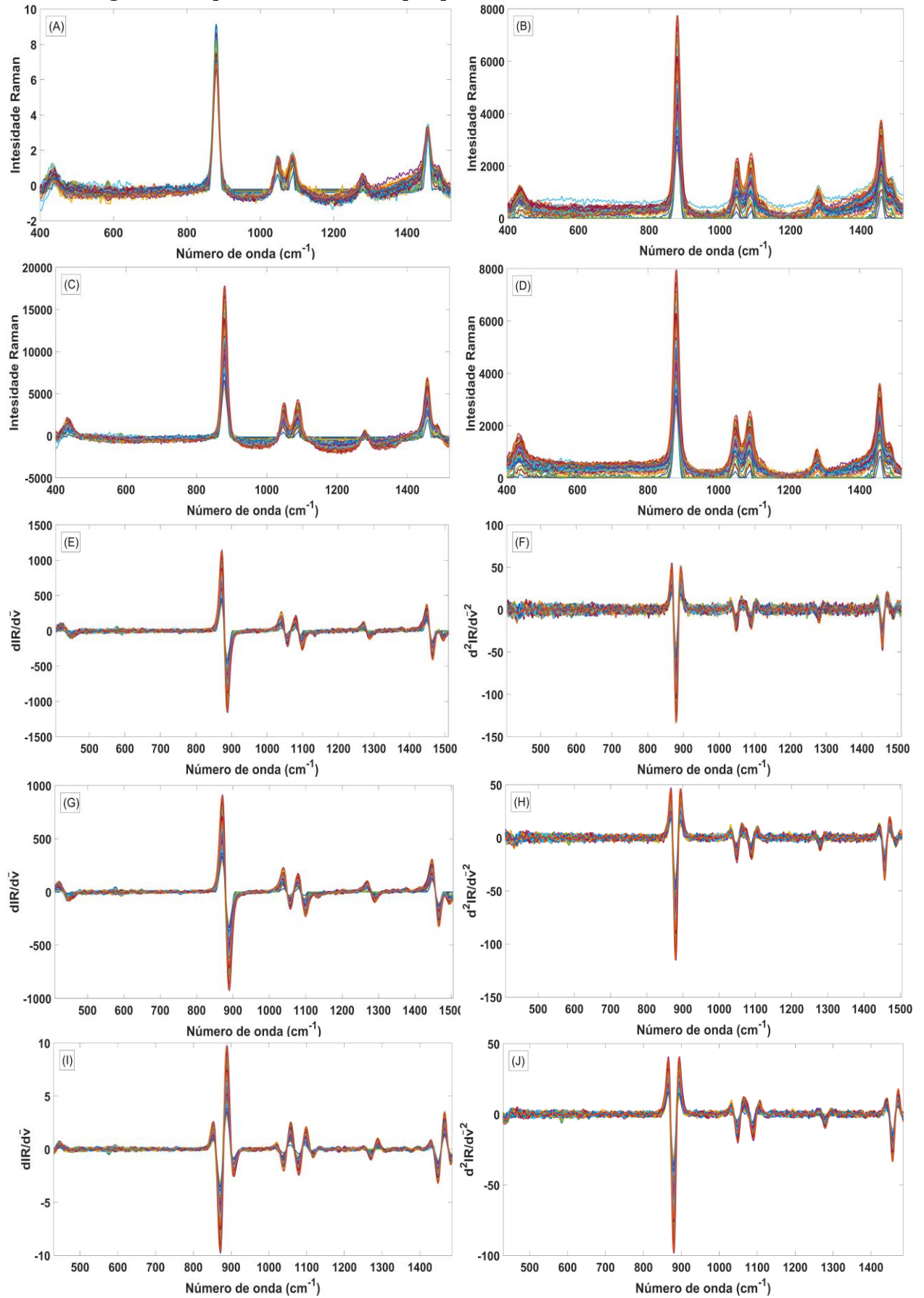


Além disso, o pré-processamento dos dados Raman constitui uma etapa fundamental para tornar o processo de classificação mais eficaz. A **Figura 13** apresenta uma série de espectros submetidos a diferentes tratamentos: (A) SNV, (B) BC–(Offset), (C) BC–(Linear), (D) BC–(Offset-Linear), (E) SG–(15,1,1), (F) SG–(15,2,2), (G) SG–(19,1,1), (H) SG–(19,2,2), (I) SG–(23,1,1) e (J) SG–(23,2,2).

O espectro Raman obtido após a aplicação do pré-processamento MSC ilustrado na **Figura 14** apresenta resultados consistentes evidenciando a eficiência dessa técnica para minimizar efeitos indesejados, como variações de espalhamento e diferenças na intensidade dos sinais. A aplicação do MSC é fundamental para a análise classificatória, visto que promove uma maior uniformidade entre as amostras e reduz o impacto de fatores externos, como variações na concentração ou espessura das amostras analisadas (HEMMATEENEJAD *et al.*, 2024).

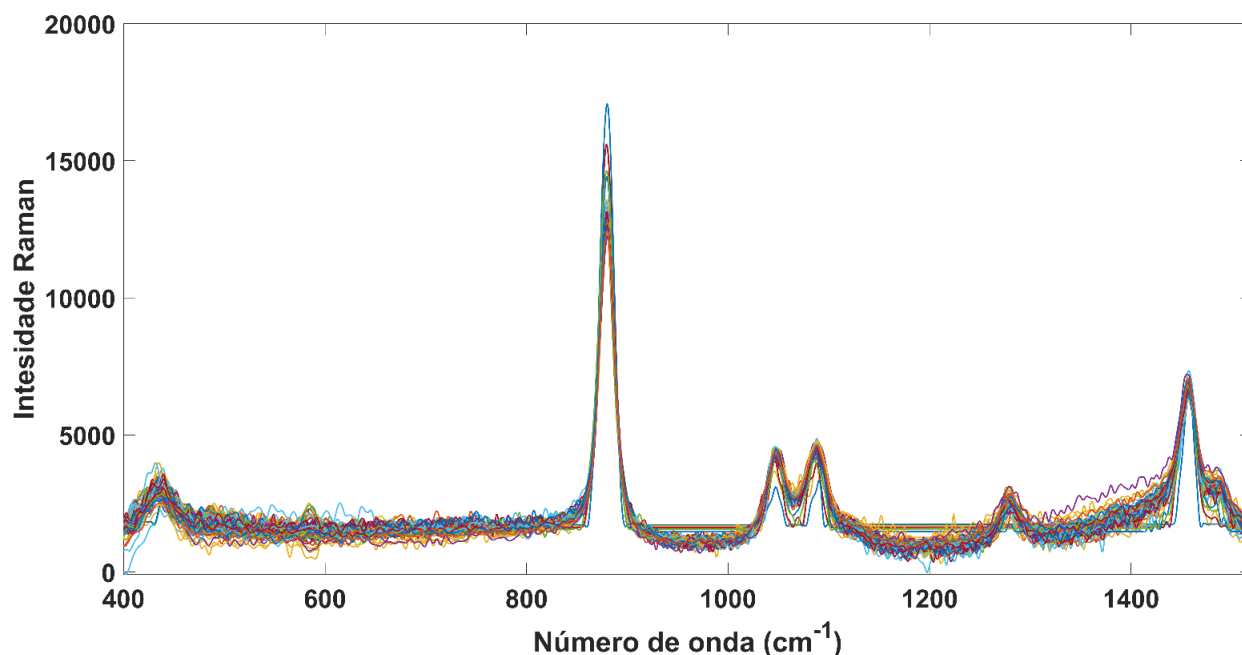
A clara redução do ruído e o realce das variações espectrais sutis são indicativos de que o pré-processamento escolhido foi eficaz em tornar os dados mais interpretáveis. Esses ajustes contribuíram diretamente para melhorar a capacidade de discriminação das cachaças da Paraíba em relação às de outros estados brasileiros, levando em consideração os melhores resultados obtidos com esse tipo de pré-processamento.

Figura 13 Espectros Raman com pré-processamentos



Fonte: elaborado pelo autor, 2025

Figura 14: Espectros Raman pré-processados com MSC na faixa de 400 a 1518 cm^{-1}



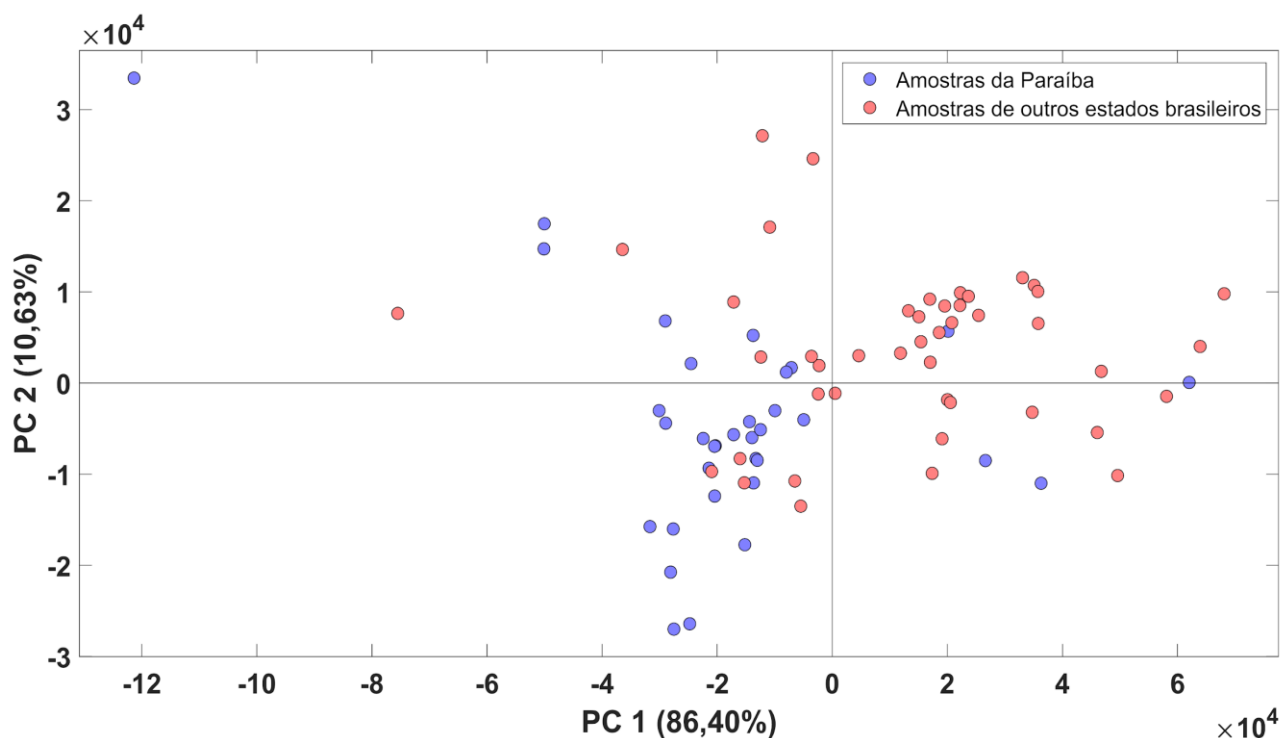
Fonte: elaborado pelo autor, 2025

Portanto, a uniformidade visual observada no espectro sugere que as diferenças características das amostras podem ser exploradas com maior precisão por modelos quimiométricos como o DD-SIMCA, potencializando assim a acurácia na diferenciação das cachaças. Assim, o uso do MSC como pré-processamento destaca-se como uma escolha eficiente, fornecendo uma base sólida para o desenvolvimento de modelos robustos e confiáveis. Essa abordagem reforça o potencial da espectroscopia Raman combinada à quimiometria como ferramenta eficaz para a diferenciação de cachaças com base em suas características regionais.

5.2.2 Análise exploratória de dados utilizando PCA

A **Figura 15** expressa o gráfico da análise de Componentes Principais realizada com os dados brutos Raman, é possível observar que assim como o gráfico de score utilizando os dados NIR, a PCA dos dados Raman também indicam uma separação parcial no gráfico de score entre as amostras do estado paraibano (pontos azuis) e de outras regiões (pontos vermelhos).

A primeira componente principal (PC1) explica 86,40% da variância, enquanto a segunda componente principal (PC2) explica 10,63% da variância, contribuindo para uma menor variação entre os dados. Isso indica que a maior parte da diferenciação entre as amostras é capturada por PC1, sugerindo que as características espectrais mais relevantes para essa separação estão concentradas nessa componente.

Figura 15: Gráfico de scores do modelo PCA

Fonte: elaborado pelo autor, 2025

A alta porcentagem de variância explicada pela primeira componente principal (PC1) indica que as diferenças mais marcantes nas características espectrais das amostras estão concentradas nesse componente, reforçando sua importância na distinção entre as duas classes. Portanto, PC1 desempenha um papel fundamental ao capturar as principais variações químicas ou estruturais que definem a separação observada entre as amostras.

Observando a distribuição dos pontos, observa-se que as amostras do estado paraibano tendem a se agrupar mais à esquerda do gráfico, enquanto as amostras de outros estados se espalham mais à direita. Essa separação parcial e tendência de agrupamento entre as amostras da mesma classe no gráfico de score, sugere que há uma diferença significativa nas características químicas ou estruturais das amostras de cada região, evidenciada pela espectroscopia Raman. Essa análise reforça a capacidade da PCA de reduzir a dimensionalidade dos dados espectrais, permitindo a visualização de padrões e agrupamentos de maneira eficiente.

Portanto a análise exploratória por meio da PCA indica uma separação parcial entre as amostras, sugerindo a existência de padrões diferenciáveis, mas sem uma discriminação definitiva. Devido a isto é necessário analisar os dados obtidos utilizando uma técnica de classificação supervisionada como o método DD-SIMCA que surge como uma abordagem

robusta para a autenticação e diferenciação geográfica das cachaças, permitindo definir um limite de aceitação baseado em um modelo de classe única.

5.2.3 Análise classificatória utilizando DD-SIMCA

Assim como foi feito com os dados NIR, também foram aplicados diferentes pré-processamentos aos dados Raman (**Tabela 2**), com o objetivo de otimizar a discriminação entre as classes e reduzir interferências de ruído. Assim como, foi realizado ajustes nos parâmetros relacionados à correção de linha de base, além da derivação e suavização com o filtro de Savitzky-Golay.

Tabela 2: Desempenho dos diferentes pré-processamentos em termos de métricas de classificação dos dados Raman

Pré-processamentos	Sensibilidade (%)	Especificidade (%)	Acurácia (%)	VP	FN	VN	FP
Dados brutos	33,33	97,80	87,00	3	6	44	1
MSC	100,00	97,80	98,10	9	0	44	1
SNV	100,00	95,60	96,30	9	0	43	2
BC-(Offset)	66,70	97,80	92,60	6	3	44	1
BC-(Linear)	55,60	97,80	90,70	5	4	44	1
BC-(Offset-Linear)	66,70	95,60	90,70	6	3	43	2
SG-(15,1,1)	88,90	84,40	85,20	8	1	38	7
SG-(15, 2, 2)	66,70	88,90	85,20	6	3	40	5
SG-(19, 1, 1)	88,90	80,00	81,50	8	1	36	9
SG-(19, 2, 2)	33,30	95,60	85,20	3	6	43	2
SG-(23,1,1)	88,90	82,20	83,30	8	1	37	8
SG-(23,2, 2)	88,90	91,10	90,70	8	1	41	4

Fonte: elaborado pelo autor, 2025

A **Tabela 2** apresenta a sensibilidade, especificidade e acurácia do modelo de classificação DD-SIMCA utilizando dados espectroscópicos Raman com diferentes pré-processamentos. A escolha do pré-processamento influencia diretamente o desempenho do modelo na distinção entre cachaças da Paraíba, que representa a classe alvo, e cachaças de outras regiões, ou seja, a classe não alvo.

Considerando inicialmente o processo de classificação utilizando apenas os dados brutos, sem qualquer tipo de pré-processamento, observa-se que o modelo apresenta uma baixa sensibilidade de 33,33%, o que indica uma dificuldade significativa em identificar corretamente as amostras da classe alvo. No entanto, a especificidade é alta, alcançando 97,80%, demonstrando assim que as amostras pertencentes à classe não alvo são bem reconhecidas. A

acurácia geral do modelo nesse cenário é de 87%, o que sugere que, mesmo sem correções, o modelo ainda mantém um desempenho relativamente aceitável na classificação das amostras.

Entre os diferentes pré-processamentos avaliados, o MSC apresentou os melhores resultados globais, com uma sensibilidade de 100,00%, especificidade de 97,80% e acurácia de 98,10%. Esse desempenho superior indica que o MSC é altamente eficaz na minimização de variações não relacionadas à composição química das amostras, melhorando a representatividade do modelo. Da mesma forma, o SNV também proporcionou uma sensibilidade de 100,00%, garantindo que todas as amostras pertencentes à classe alvo fossem corretamente identificadas. No entanto, a especificidade do SNV foi menor (95,60%), resultando em uma acurácia global um pouco inferior (96,30%) quando comparado ao MSC.

A correção de linha de base foi utilizada abordando diferentes parâmetros, incluindo Offset, Linear e uma combinação de ambos (Offset-Linear). Em geral, os modelos construídos utilizando esse pré-processamento resultaram em acurácias que variam entre 90,70% e 92,60%, isso evidencia a capacidade da correção de linha de base em minimizar alterações indesejadas na linha de base devido à superposição de efeitos de absorção e dispersão, resultando em deslocamentos variáveis ou oscilações de baixa frequência, melhorando assim a qualidade dos espectros. No entanto, apesar da acurácia semelhante dos pré-processamentos MSC e SNV, a sensibilidade obtida foi inferior, variando entre 55,60% e 66,70% (ANDRIES *et al.*, 2023).

A aplicação de derivações de Savitzky-Golay também foi analisada como estratégia de pré-processamento para realçar variações espectrais sutis e reduzir efeitos de ruído nos espectros Raman. Os melhores desempenhos dentro desse grupo foram observados para as derivações SG-(15,1,1) e SG-(23,2, 2), ambas com sensibilidade de 88,90% e acurácia entre 85,20 e 90,70%. No entanto, outras combinações, como SG-(19, 2, 2), resultaram em uma sensibilidade muito baixa (33,30%), indicando que essa configuração pode estar removendo informações espectrais relevantes. Já a derivação utilizando uma janela de 23 pontos, segunda derivada e um polinômio de grau 2 apresentou um desempenho mais equilibrado, alcançando uma acurácia similar à observada para a correção de linha de base.

Com base nos resultados obtidos, conclui-se que os pré-processamentos MSC e SNV são os mais eficazes para o modelo DD-SIMCA utilizando os dados Raman, com destaque para o MSC, que apresentou uma maior acurácia de 98,10% e uma sensibilidade de 100%, demonstrando sua capacidade de amenizar variações indesejadas nos dados, enquanto preserva as informações essenciais que envolvem às características exclusivas das amostras. Assim como outros métodos de correção de linha de base e derivações podem melhorar a qualidade dos espectros, de tal forma que foi obtido resultados melhores que a utilização dos dados brutos no

processo de classificação, contudo tiveram um desempenho menor que o MSC.

Diante disso, a matriz de confusão descrita na **Tabela 2** reflete o desempenho do modelo DD-SIMCA na classificação de dados Raman pré-processados com MSC. A análise dos valores mostra que o modelo tem uma excelente capacidade de distinguir entre as classes positivas e negativas, cometendo apenas um único erro de classificação.

O modelo identificou corretamente 9 amostras do estado da Paraíba como pertencentes à classe alvo positiva (VP), o que demonstra que ele é altamente eficaz na detecção dos casos de interesse. Além disso, não houve falsos negativos (FN), ou seja, nenhuma amostra positiva foi erroneamente classificada como sendo pertencente a classe não alvo. Isso significa que o modelo tem uma sensibilidade perfeita, conseguindo identificar todos os casos positivos sem falhas.

Na classificação das amostras negativas, o modelo também teve um ótimo desempenho. 44 amostras negativas foram corretamente classificadas como sendo pertencentes a classe não alvo (VN), o que indica uma alta capacidade de reconhecer corretamente os casos que não pertencem à classe alvo. No entanto, houve 1 caso de falso positivo (FP), ou seja, uma amostra que, na realidade, é da classe negativa que foi incorretamente classificada como sendo pertencente a classe alvo, ou seja positiva.

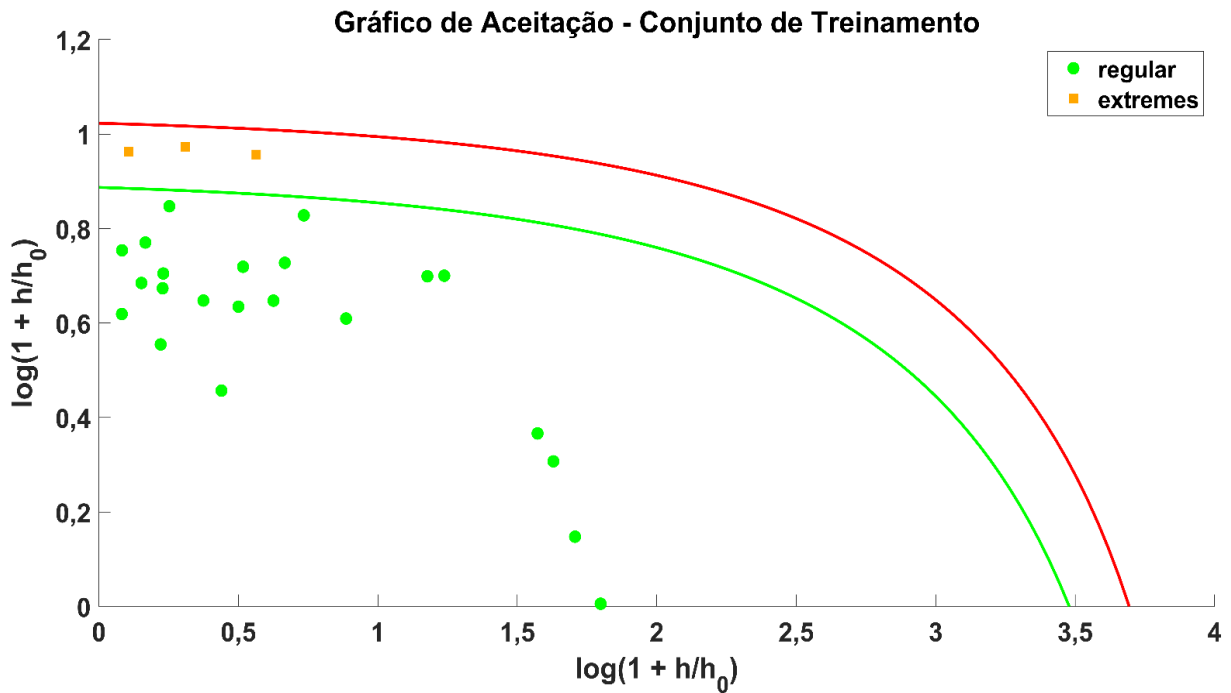
Esse único falso positivo indica que o modelo, embora altamente preciso, pode ocasionalmente classificar erroneamente uma amostra negativa como pertencente à classe positiva. Dependendo da aplicação, isso pode ter um impacto maior ou menor. Por exemplo, em um contexto em que um falso positivo implica em custos elevados ou ações desnecessárias, pode ser interessante ajustar parâmetros do modelo para torná-lo mais rigoroso na aceitação de amostras positivas (MOHAMMADI *et al.*, 2023).

No geral, o modelo apresentou um desempenho excepcional, com uma taxa de erro muito baixa e uma capacidade extremamente confiável de classificar corretamente tanto as amostras positivas quanto as negativas. A ausência de falsos negativos é um ponto extremamente positivo, pois garante que nenhum caso relevante foi perdido. O pequeno número de falsos positivos indica um modelo ligeiramente conservador, mas ainda assim altamente eficaz.

Uma melhor compreensão e visualização do resultado obtido para o MSC pode ser elucidado através do gráfico de treinamento e de teste do modelo DD-SIMCA, aplicado aos dados espectroscópicos Raman para a classificação de cachaças ilustrados na **Figura 16 e 17**. Esses gráficos permitem avaliar a adequação das amostras dentro do modelo construído, identificando aquelas que seguem o comportamento predominante e aquelas que apresentam

características extremas, assim como o resultado obtido do modelo de classificação através do gráfico de teste.

Figura 16: Gráfico da área de aceitação do modelo DD-SIMCA para o conjunto de treino com pré-processamento MSC dos dados Raman



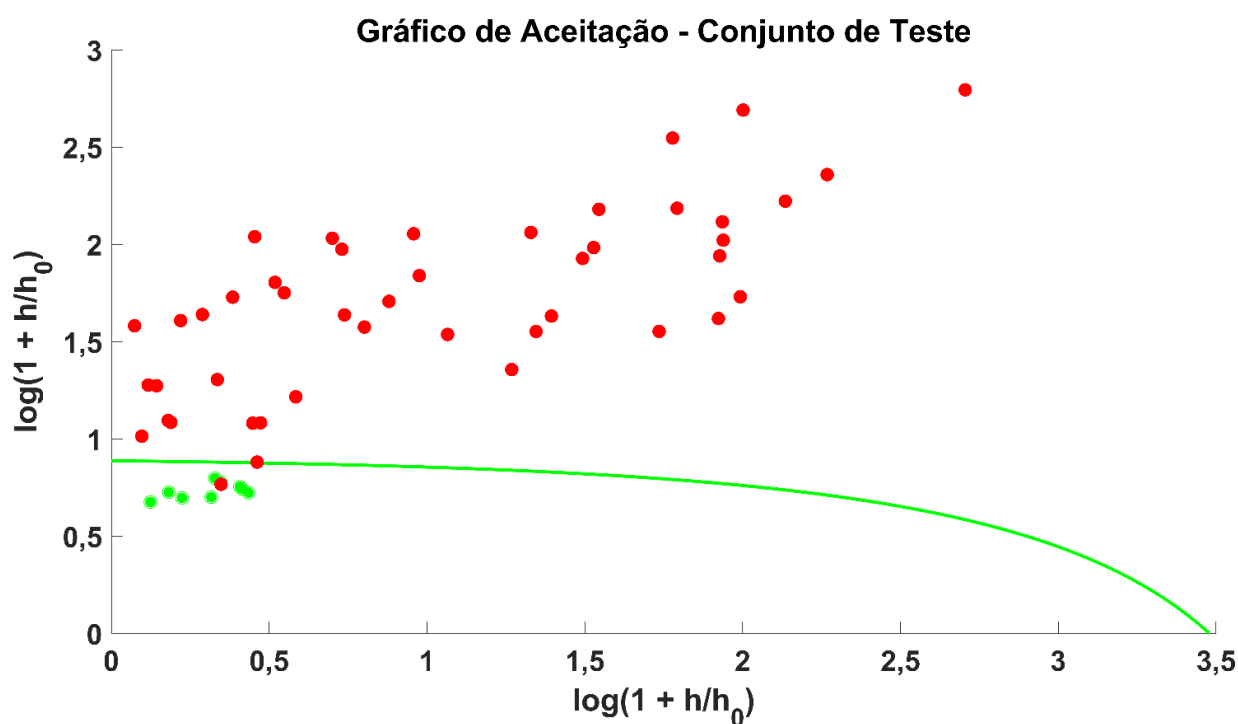
Fonte: elaborado pelo autor, 2025

Observa-se que de acordo com a **Figura 16**, das amostras utilizadas na construção do modelo, a maioria (representada pelos pontos verdes) está localizada dentro da região delimitada pela curva verde, indicando que seguem o comportamento esperado e são bem representadas pelo modelo. No entanto, algumas amostras (destacadas pelos pontos laranjas) foram classificadas como extremas, posicionando-se acima da curva verde e mais próximas da curva vermelha. Essas amostras extremas indicam uma maior variabilidade em relação as demais amostras do conjunto de treinamento, o que pode sugerir que possuem características diferenciadas dentro do contexto analisado. No entanto, apesar dessa maior variabilidade em relação as demais amostras delimitadas pela curva verde, não significa necessariamente que sejam outliers severos, mas sim que apresentam particularidades que as afastam do comportamento predominante da classe modelada. A presença de amostras extremas pode estar relacionada a variações naturais nas composições químicas das cachaças, como diferenças na matéria-prima, no processo de fermentação ou envelhecimento.

De maneira geral, a distribuição das amostras sugere que o modelo de classificação é representativo para o conjunto de treinamento, uma vez que a maioria das amostras está dentro da região de aceitação. Isso demonstra a capacidade do modelo em captar as principais variações presentes nos espectros Raman das cachaças analisadas, garantindo um desempenho confiável na distinção entre diferentes amostras.

A **Figura 17** apresenta o gráfico de teste de aceitação para classificação das amostras da Paraíba e demais regiões. Esse gráfico permite avaliar o desempenho do modelo ao classificar novas amostras com base no padrão aprendido durante o treinamento.

Figura 17: Gráfico da área de aceitação do modelo DD-SIMCA para o conjunto de teste com pré-processamento MSC dos dados Raman



Fonte: elaborado pelo autor, 2025

Portanto de acordo com o gráfico de teste, observa-se que as amostras classificadas como pertencentes à classe alvo representadas por pontos verdes, todas estão localizadas dentro da região delimitada pela curva verde, indicando que essas amostras seguem o comportamento esperado e são aceitas pelo modelo como pertencentes à classe do estado paraibano, isso indica que essas amostras possuem variações que são comuns aos padrões de variabilidade do conjunto de treinamento.

Por outro lado, as amostras de outras regiões, boa parte se encontra acima da curva delimitada para a classe alvo (pontos vermelhos), indicando que o modelo possui uma boa

capacidade de distinção entre a classe alvo e demais amostras não pertencentes a classe alvo, o que deduz que o modelo identificou variações nessas amostras que não são comuns aos padrões de variação do conjunto de treinamento, sendo assim classificadas como não pertencentes à classe.

Portanto, o uso da espectroscopia Raman associado ao modelo DD-SIMCA e ao pré-processamento MSC demonstra um desempenho eficaz na distinção entre cachaças do estado paraibano e de outras regiões. A clara separação entre as amostras reforça a robustez do modelo na identificação de padrões característicos das cachaças da classe alvo, permitindo sua aplicação confiável na autenticação e controle de qualidade dessas bebidas.

O eficiente resultado obtido na diferenciação das cachaças do estado da Paraíba em relação às de outras regiões pode ser explicado pelo fato de que muitas das cachaças pertencentes ao grupo não alvo são produzidas a partir da mistura de destilados provenientes do sul do Brasil. Exemplos disso incluem a cachaça Salinas e a Saliníssima, ambas de Minas Gerais, bem como a Cachaça Sagatiba, que é produzida em São Paulo.

As cachaças produzidas na Paraíba apresentam características sensoriais distintas em relação às cachaças elaboradas em regiões como Minas Gerais e São Paulo que foram utilizadas para construção do conjunto de teste, devido a fatores como, fermentação, destilação e envelhecimento. O clima e o solo desempenham um papel fundamental na qualidade e rendimento da cana-de-açúcar e, conseqüentemente, na produção da cachaça. O clima úmido e quente especificamente do Brejo Paraibano favorece o cultivo de uma cana-de-açúcar com características distintas de outras regiões produtoras. Minas gerais tem condições úmidas em toda sua área no verão e, predominantemente, áridas no inverno, enquanto São Paulo apresenta condições climáticas mais variadas, o que influencia a doçura e a mineralidade das cachaças produzidas nessas regiões, resultando assim em características distintas (CARNEIRO *et al.*, 2021; FLACK-PRAIN *et al.*, 2021; REBOITA *et al.*, 2015).

6 CONCLUSÕES

A autenticação geográfica de cachaças é um aspecto essencial para garantir a valorização dos produtos regionais, assim como assegurar sua autenticidade e fornecer uma base científica para possíveis certificações de origem. Nesse estudo, buscou-se diferenciar cachaças do estado paraibano de cachaças de outras regiões do Brasil por meio da aplicação de técnicas espectroscópicas combinadas com métodos quimiométricos. O uso dessas abordagens se mostrou uma alternativa promissora para a análise da bebida, permitindo um controle da qualidade mais ágil, eficiente e sem a necessidade de preparações complexas de amostras.

As técnicas espectroscópicas utilizadas, NIR e Raman, demonstraram eficácia na distinção das amostras, onde a espectroscopia NIR apresentou 64,80% de acurácia e a espectroscopia Raman, apresentou um desempenho superior, onde através dos dados Raman foi possível obter uma acurácia de 98,10% no processo de classificação, demonstrando assim a sua capacidade de fornecer informações químicas mais detalhadas sobre as amostras. Essa diferença de desempenho pode estar relacionada à maior sensibilidade do Raman a determinadas estruturas moleculares presentes na matriz da cachaça, permitindo uma discriminação mais precisa entre as regiões estudadas.

Além disso, a técnica quimiométrica DD-SIMCA demonstrou eficiência na diferenciação entre as amostras de cachaças da Paraíba e as de outras regiões. O modelo apresentou um desempenho satisfatório, destacando-se pela capacidade de classificar corretamente as amostras e reduzir possíveis erros de atribuição. Esses resultados reforçam a robustez da quimiometria como ferramenta complementar às técnicas espectroscópicas, otimizando a interpretação dos dados e aumentando a confiabilidade dos modelos preditivos.

Os resultados deste estudo ressaltam a importância do uso combinado de espectroscopia e quimiometria na autenticação geográfica de cachaças. A implementação dessas técnicas pode trazer benefícios significativos para produtores, órgãos reguladores e consumidores, ao possibilitar um método analítico rápido, preciso e acessível para a autenticação da bebida. Além disso, a adoção dessas metodologias pode auxiliar no fortalecimento da identidade regional da cachaça do estado paraibano, agregando valor ao produto e fomentando sua competitividade no mercado.

Dessa forma, esse trabalho demonstra a eficácia das metodologias aplicadas, abrindo assim, caminho para futuras pesquisas, podendo ser expandido para a análise de outras bebidas artesanais e produtos agroindustriais que exigem certificação de origem e controle de qualidade.

Além disso, a abordagem apresentada pode ser aprimorada com novas técnicas

espectroscópicas e estratégias quimiométricas, possibilitando uma caracterização ainda mais detalhada.

REFERÊNCIAS

ALCARDE, A. R. *Cachaça: ciência, tecnologia e arte*. 2. ed. São Paulo: Edgard Blücher, 2017.

ANDRIES, E.; NIKZAD-LANGERODI, R. Supervised and Penalized Baseline Correction. *Chemometrics and Intelligent Laboratory Systems*, v. 253, p. 105200, 2024. DOI: <https://doi.org/10.1016/j.chemolab.2024.105200>.

ANGELIS, D.; SUMMO, C.; PASQUALONE, A.; FACCIA, M.; SQUEO, G. Advancements in food authentication using soft independent modelling of class analogy (SIMCA): a review. *Food Quality and Safety*, v. 8, p. fyae032, 2024. DOI: <https://doi.org/10.1093/fqsafe/fyae032>.

BALDOMERO-NARANJO, M.; MARTÍNEZ-MERINO, L. I.; RODRÍGUEZ-CHÍA, A. M. A robust SVM-based approach with feature selection and outliers detection for classification problems. *Expert Systems with Applications*, v. 178, p. 115017, 2021. DOI: <https://doi.org/10.1016/j.eswa.2021.115017>.

BEĆ, K. B.; GRABSKA, J.; HUCK, C. W. Near-infrared spectroscopy in bio-applications. *Molecules*, v. 25, p. 2948, 2020. DOI: <https://doi.org/10.3390/molecules25122948>.

BEĆ, K. B.; GRABSKA, J.; STANZL, M. T.; BRUNNER, M. E.; HUCK, C. W.; BEGANOVIĆ, A. Vibrational coupling to hydration shell – Mechanism to performance enhancement of qualitative analysis in NIR spectroscopy of carbohydrates in aqueous environment. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 237, p. 118359, 2020. DOI: <https://doi.org/10.1016/j.saa.2020.118359>.

BEĆ, K. B.; HUCK, C. W. Breakthrough potential in near-infrared spectroscopy: Spectra simulation. A review of recent developments. *Frontiers in Chemistry*, v. 7, p. 48, 2019. DOI: <https://doi.org/10.3389/fchem.2019.00048>.

BORTOLETTO, A. M.; ALCARDE, A. R. Congeners in sugar cane spirits aged in casks of different woods. *Food Chemistry*, v. 139, p. 695-701, 2013. DOI: <https://doi.org/10.1016/j.foodchem.2012.12.053>.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Anuário da Cachaça 2024: Ano referência 2023. Brasília, 2024. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/anuarios-de-produtos-de-origem-vegetal-pasta/anuario-da-cachaca-2024-ano-referencia-2023.pdf>. Acesso em: 22 nov. 2024.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. *Lista de Indicações Geográficas*. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/sustentabilidade/indicacao-geografica/listaigs>. Acesso em: 22 nov. 2024.

CARNEIRO, K. A. A.; SOUSA, L. R.; FRAGA, V. S.; SILVA, A. P.; VENDRUSCOLO, J. Solos do brejo paraibano: influências do clima, do relevo e do material de origem. *Research, Society and Development*, v. 10, p. e22410716545, 2021. DOI: <https://doi.org/10.33448/rsd-v10i7.16545>.

CHEN, Z.; HARRINGTON, P. de B. Automatic soft independent modeling for class analogies. *Analytica Chimica Acta*, v. 1090, p. 47–56, 2019. DOI: <https://doi.org/10.1016/j.aca.2019.09.035>.

CHO, S.-H.; KIM, H.-J.; PARK, J.-S.; LEE, K.-T. Determination of water content in ethanol by miniaturized near-infrared (NIR) system. *Bulletin of the Korean Chemical Society*, v. 26, p. 115–118, 2005. DOI: <https://doi.org/10.5012/bkcs.2005.26.1.115>.

CHU, B.; JIN, Y.; GAO, A.; WU, S. Optimization of brandy base wine production technique using ginger pomace. *E3S Web of Conferences*, v. 261, p. 02086, 2021. DOI: <https://doi.org/10.1051/e3sconf/202126102086>.

CORSETTI, S.; MCGLOIN, D.; KIEFER, J. Comparison of Raman and IR spectroscopy for quantitative analysis of gasoline/ethanol blends. *Fuel*, v. 166, p. 488–494, 2016. DOI: <https://doi.org/10.1016/j.fuel.2015.11.018>.

DODO, K.; FUJITA, K.; SODEOKA, M. Espectroscopia Raman para pesquisa em biologia química. *Journal of the American Chemical Society*, v. 144, n. 43, p. 19651–19667, 2022. DOI: <https://doi.org/10.1021/jacs.2c05359>

FLACK-PRIAN, S.; SHI, L.; ZHU, P.; DA ROCHA, H. R.; CABRAL, O.; HU, S.; WILLIAMS, M. The impact of climate change and climate extremes on sugarcane production. *GCB Bioenergy*, v. 13, p. 408–424, 2021. DOI: <https://doi.org/10.1111/gcbb.12797>.

GABA, F.; TIPPING, W. J.; SALJI, M.; FAULDS, K.; GRAHAM, D.; LEUNG, H. Y. Raman spectroscopy in prostate cancer: techniques, applications and advancements. *Cancers*, v. 14, p. 1535, 2022. DOI: <https://doi.org/10.3390/cancers14061535>.

GAMA, E. M. Caracterização físico-química da cachaça artesanal comercializada no mercado municipal de Coronel Murta-MG. *RECIMA21 – Revista Científica Multidisciplinar*, v. 3, p. e321168, 2022. DOI: <https://doi.org/10.33448/rsd-v3i2.1168>.

GOMES, Adriano A.; KHVALBOTA, Liudmyla; MACHÝNÁKOVÁ, Andrea; FURDÍKOVÁ, Katarína; ZINI, Claudia A.; ŠPÁNIK, Ivan. Slovak Tokaj wines classification with respect to geographical origin by means of one class approaches. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 257, p. 119770, 2021. DOI: <https://doi.org/10.1016/j.saa.2021.119770>.

GONÇALVES, P. D. A.; CAVICHIOLI, F. A. A evolução do mercado de cachaça artesanal. *Revista Interface Tecnológica*, v. 18, p. 343–354, 2021. DOI: <https://doi.org/10.33448/rsd-v10i7.16545>.

HANNA, K.; KRZOSKA, E.; SPEIRS, V. Raman spectroscopy: current applications in breast cancer diagnosis, challenges and future prospects. *British Journal of Cancer*, v. 126, p. 1125–1139, 2022. DOI: <https://doi.org/10.1038/s41416-021-01659-5>.

HEMMATEENEJAD, B.; MOBARAKI, N.; BAUMANN, K. Robust multiplicative scatter correction using quantile regression. *Journal of Chemometrics*, v. 38, p. e3589, 2024. DOI: <https://doi.org/10.1002/cem.3589>.

INSTITUTO BRASILEIRO DA CACHAÇA (IBRAC). *Mercado Interno*. Disponível em: <https://ibrac.net/servicos/mercado-interno>. Acesso em: 24 ago. 2025.

JUNIOR, E. D. G.; SILVA, L. P.; BAPTISTA, J. A. Diageo e sua importância para a internacionalização da cachaça. *Revista do Encontro de Gestão e Tecnologia*, v. 1, p. 48–55, 2024. DOI: <https://doi.org/10.33448/rsd-v1i3.79>

KHARBACH, M.; MANSOURI, M. A.; TAABOUZ, M.; YU, H. Current application of advancing spectroscopy techniques in food analysis: data handling with chemometric approaches. *Foods*, v. 12, p. 2753, 2023. DOI: <https://doi.org/10.3390/foods12142753>.

KHERIF, Ferath; LATYPOVA, Adeliya. Principal component analysis. In: MECHELLI, Alessandro; VIEIRA, Sergio (Eds.). *Machine Learning*. Academic Press, 2020. p. 209–225. DOI: <https://doi.org/10.1016/B978-0-12-815739-8.00012-2>.

KUCHERYAVSKIY, S.; RODIONOVA, O.; POMERANTSEV, A. A comprehensive tutorial on Data-Driven SIMCA: Theory and implementation in web. *Journal of Chemometrics*, v. 38, p. e3556, 2024. DOI: <https://doi.org/10.1002/cem.3556>

LIN, M.; YANG, B.; DAI, M.; XU, Y.; LI, X.; SUN, B. East meets west in alcoholic beverages: Flavor comparison, microbial metabolism and health effects. *Food Bioscience*, v. 56, p. 103385, 2023. DOI: <https://doi.org/10.1016/j.fbio.2023.103385>.

LIU, Y.; WANG, X.; ZHANG, H.; LI, J.; CHEN, Q. Two-dimensional correlation analysis of visible/near-infrared spectral intensity variations of chicken breasts with various chilled and frozen storages. *Journal of Agricultural and Food Chemistry*, v. 52, p. 505–510, 2004. DOI: <https://doi.org/10.1021/jf034655n>.

LUNTER, D.; KLANG, V.; KOCSIS, D.; VARGA-MEDVECKY, Z.; BERKÓ, S.; ERDŐ, F. Novel aspects of Raman spectroscopy in skin research. *Experimental Dermatology*, v. 31, p. 1311–1329, 2022. DOI: <https://doi.org/10.1111/exd.14645>

MAZIVILA, Sarmiento J.; PÁSCOA, Ricardo N. M. J.; CASTRO, Rafael C.; RIBEIRO, David S. M.; SANTOS, João L. M. Detection of melamine and sucrose as adulterants in milk powder using near-infrared spectroscopy with DD-SIMCA as one-class classifier and MCR-ALS as a means to provide pure profiles of milk and of both adulterants with forensic evidence: A short communication. *Talanta*, v. 216, p. 120937, 2020. DOI: <https://doi.org/10.1016/j.talanta.2020.120937>.

MELO, T. S.; MAGALHÃES, A. E. M.; MACÊDO, L. S.; SIQUEIRA, L. C. S.; LIMA, L. E. B.; SOUZA, T. H. S.; MELO, T. S.; OLIVEIRA, M. M. B. Aguardente e cachaça brasileira, da história ao processamento moderno: a evolução da bebida artesanal a um produto de qualidade. *Brazilian Journal of Development*, v. 7, p. 95093–95111, 2021. DOI: <https://doi.org/10.34117/bjdv7n10-020>

MENEVŞEOĞLU, A.; KAYALAR, Ö.; YILMAZ, E.; DEMİR, F.; KOC, M. Non-invasive monitoring of ethanol and methanol levels in grape-derived pisco distillate by vibrational spectroscopy. *Sensors*, v. 21, p. 6278, 2021. DOI: <https://doi.org/10.3390/s21186278>.

MOHAMMADI, K.; ZHAO, H.; ZHAI, M.; TUNG, F. Ranking regularization for critical rare classes: minimizing false positives at a high true positive rate. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, p. 15783-15792, 2023. DOI: <https://doi.org/10.1109/CVPR52729.2023.01515>.

MOURÃO, R. S.; VALE, B. R. C.; FONSECA, A. F. V.; CARVALHO, T. A. S.; SCHIAVON, M. A. Espectroscopia de fluorescência: dos fundamentos à influência dos parâmetros instrumentais para análises de corantes orgânicos e nanopartículas inorgânicas. *Química Nova*, v. 47, p. e-20230095, 2024. DOI: <https://doi.org/10.21577/0100-4042.20230095>.

NASCIMENTO-DIAS, B. Raman spectroscopy and its peculiarities. *Journal of Experimental Techniques and Instrumentation*, v. 4, p. 15–25, 2020. DOI: <https://doi.org/10.30609/jeti.v4i01.10329>.

NEVES, M. D. G.; POPPI, R. J. Authentication and identification of adulterants in virgin coconut oil using ATR/FTIR in tandem with DD-SIMCA one class modeling. *Talanta*, v. 219, p. 121338, 2020. DOI: <https://doi.org/10.1016/j.talanta.2020.121338>.

NÓBREGA, I. C. C. Análise dos compostos voláteis da aguardente de cana por concentração dinâmica do “headspace” e cromatografia gasosa-espectrometria de massas. *Ciência e Tecnologia de Alimentos*, v. 23, p. 210–216, 2003. DOI: <https://doi.org/10.1590/S0101-20612003000200019>.

OLIVEIRA, L. O.; JUNIOR, E. F. Produção de cachaça artesanal. *Revista Interface Tecnológica*, v. 19, p. 810-818, 2022. DOI: <https://doi.org/10.31510/inf.v19i2.1542>.

OLIVEIRA, R. E. S.; CARDOSO, M. G.; SANTIAGO, W. D.; BARBOSA, R. B.; ALVARENGA, G. F.; NELSON, D. L. Physicochemical parameters and volatile composition of cachaça produced in the state of Paraíba, Brasil. *Research, Society and Development*, v. 9, p. 504974409, 2020. DOI: <https://doi.org/10.33448/rsd-v9i7.4409>.

ORLANDO, A.; FRANCESCHINI, F.; MUSCAS, C.; PIDKOVA, S.; BARTOLI, M.; ROVERE, M.; TAGLIAFERRO, A. A comprehensive review on Raman spectroscopy applications. *Chemosensors*, v. 9, p. 262, 2021. DOI: <https://doi.org/10.3390/chemosensors9090262>.

PAPASPYRIDAKOU, P.; GIANNOUTSOU, P.; ORKOULA, M. G. Non-destructive and non-invasive measurement of ethanol and toxic alcohol strengths in beverages and spirits using portable Raman spectroscopy. *Biosensors*, v. 13, p. 135, 2023. DOI: <https://doi.org/10.3390/bios13010135>.

PEREIRA, Amanda de Andrade Marcondes; ACEVEDO, Maria Soledad M. S. F.; ALCARDE, André Ricardo. Improvement of the Chemical Quality of Cachaça. *Beverages*, v. 10, p. 79, 2024. DOI: <https://doi.org/10.3390/beverages10030079>.

RATKOVICH, N.; ESSER, C.; MACHADO, A. M. R.; MENDES, B. A.; CARDOSO, M. G. The spirit of cachaça production: an umbrella review of processes, flavour, contaminants and quality improvement. *Foods*, v. 12, p. 3325, 2023. DOI: <https://doi.org/10.3390/foods12173325>.

REBOITA, M. S.; RODRIGUES, M.; SILVA, L. F.; ALVES, M. A. Aspectos climáticos do estado de Minas Gerais. *Revista Brasileira de Climatologia*, v. 17, p. 209–229, 2015. DOI: <https://doi.org/10.5380/abclima.v17i0.38903>.

RECH, Â.; WERNER, S. S. Utilização da tecnologia NIRS para predição dos valores nutricionais de forrageiras. *Agropecuária Catarinense*, v. 33, p. 11–14, 2020. DOI: <https://doi.org/10.52945/rac.v33i1.370>.

RODIONOVA, O.; TITOVA, A. V.; POMERANTSEV, A. L. Expanding the DD-SIMCA concept: a case study of precision medicine. *Analytica Chimica Acta*, v. 1250, p. 340958, 2023. DOI: <https://doi.org/10.1016/j.aca.2023.340958>.

SALONEN, T.; AHRANS, B.; BOVENS, M.; ELIAERTS, J.; HUHTALA, S.; NORDGAARD, A.; ALBERINK, I. Chemometrics in forensic chemistry—Part II: Standardized applications—Three examples involving illicit drugs. *Forensic Science International*, v. 307, p. 110138, 2020. DOI: <https://doi.org/10.1016/j.forsciint.2019.110138>.

SANTOS, G. R.; PAULINO, G. S. P.; BORGES, G. P. I.; SANTIAGO, A. F.; DA SILVA, G. A. Avanços analíticos baseados em modelos de calibração de primeira ordem e espectroscopia UV-Vis para avaliação da qualidade da água: uma revisão—parte 1. *Química Nova*, v. 45, p. 314–323, 2022. DOI: <https://doi.org/10.21577/0100-4042.20170830>.

SCHNEIDER, M. S.; VACARO, B. B.; ALEXANDRE, M. C.; FERRÃO, M. F. Métodos de análise exploratória aplicados ao ensino de espectrometria no infravermelho. *Revista Virtual de Química*, v. 10, p. 229–243, 2018. DOI: <https://doi.org/10.21577/1984-6835.20180019>.

SCHOENINGER, V.; COELHO, S. R. M.; SILOCHI, R. M. H. Cadeia produtiva da cachaça. *Energia na Agricultura*, v. 29, p. 292–300, 2014. DOI: <https://doi.org/10.17224/EnergAgric.2014v29n4p292-300>.

SHAFIZADEH-MOGHADAM, H. Fully component selection: an efficient combination of feature selection and principal component analysis to increase model performance. *Expert Systems with Applications*, v. 186, p. 115678, 2021. DOI: <https://doi.org/10.1016/j.eswa.2021.115678>.

SILVA, A. P.; SILVELLO, G. C.; BORTOLETTO, A. M.; ALCARDE, A. R. Chemical composition of sugar cane spirit produced from different distillation methods. *Brazilian Journal of Food Technology*, v. 23, p. e2018308, 2020. DOI: <https://doi.org/10.1590/1981-6723.30818>.

SILVA, K. C.; SOUZA, L. R.; OLIVEIRA, M. A.; PEREIRA, F. R.; ALMEIDA, J. P. Café com quimiometria: uma aplicação do planejamento fatorial fracionário 2^{5-1} . *Química Nova*, v. 46, p. 98–107, 2023. DOI: <https://doi.org/10.21577/0100-4042.20220123>.

SILVA, V. P.; OLIVEIRA, R. M.; SANTOS, F. L.; CARDOSO, M. G.; ALMEIDA, T. R. Cachaça production: from sugar cane to spirit. *Journal of the Institute of Brewing*, v. 129, p. 259–275, 2023. DOI: <https://doi.org/10.1002/jib.753>.

SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. Principles of Instrumental Analysis. 7. ed.

New York: Sunder College Publisher, 2017.

SUHANDY, D.; YULIA, M. The use of UV spectroscopy and SIMCA for the authentication of Indonesian honeys according to botanical, entomological and geographical origins. *Molecules*, v. 26, p. 915, 2021. DOI: <https://doi.org/10.3390/molecules26040915>.

ULLAL, N.; SAHOO, B.; SUNIL, D.; KULKARNI, S. D.; BHAT, U. B. K.; ANAND, P. J. Pyrene carbaldehyde derived carbon dots for detecting water in alcohol and security printing. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, v. 703, p. 135362, 2024. DOI: <https://doi.org/10.1016/j.colsurfa.2024.135362>.

VILELA, A. F.; SANTOS, F. L.; OLIVEIRA, R. M.; CARDOSO, M. G.; ALMEIDA, T. R. Assessment of sensory and physical-chemical quality, and potential for certification of cachaças from the state of Paraíba, Brazil. *Food Science and Technology*, v. 41, p. 661–668, 2021. DOI: <https://doi.org/10.1590/fst.20320>.

WEYER, L. G.; LO, S. C. Spectra-structure correlations in the near-infrared. *Handbook of Vibrational Spectroscopy*, v. 3, p. 1817–1837, 2002. DOI: <https://doi.org/10.1002/0470027320.s4103>.

ZHANG, F.; LI, H.; WANG, Y.; CHEN, X.; ZHOU, J. Application and progress of Raman spectroscopy in male reproductive system. *Frontiers in Cell and Developmental Biology*, v. 9, p. 823546, 2022. DOI: <https://doi.org/10.3389/fcell.2021.823546>.

ZONTOV, Y. V.; IVANOV, P. A.; PETROV, S. M.; KUZNETSOV, A. N. DD-SIMCA—a MATLAB GUI tool for data driven SIMCA approach. *Chemometrics and Intelligent Laboratory Systems*, v. 167, p. 23–28, 2017. DOI: <https://doi.org/10.1016/j.chemolab.2017.03.002>.