



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE QUÍMICA

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

MATEUS PIRES SCHNEIDER

Origem geográfica e perfil varietal de vinhos tintos produzidos no Rio Grande do Sul via técnicas espectroscópicas assistidas por ferramentas quimiométricas

DISSERTAÇÃO DE MESTRADO

Porto Alegre, Janeiro de 2021



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE QUÍMICA

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

MATEUS PIRES SCHNEIDER

Origem geográfica e perfil varietal de vinhos tintos produzidos no Rio Grande do Sul via técnicas espectroscópicas assistidas por ferramentas quimiométricas

Dissertação submetida ao Programa de Pós-Graduação em Química da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do grau de Mestre em Química.

Prof. Dr. Adriano de Araújo Gomes
Orientador

Porto Alegre, Janeiro de 2021

A presente dissertação foi realizada inteiramente pelo autor, exceto as colaborações as quais serão devidamente citadas nos agradecimentos, no período entre junho de 2018 e janeiro de 2021, no Instituto de Química da Universidade Federal do Rio Grande do Sul sob Orientação do Professor Doutor Adriano de Araújo Gomes. A dissertação foi julgada adequada para a obtenção do título de Mestre em Química pela seguinte banca examinadora:

Comissão Examinadora:

| | | |
|---|---------------------|------------------|
| Prof. Dr. José Germano Verás Neto | UEPB | Titular |
| Prof. Dr. Celito Crivellaro Guerra | EMBRAPA Uva e Vinho | Titular |
| Prof. ^a Dr. ^a Carla Sirtori | UFRGS | Titular |
| Prof. Dr. Marco Flôres Ferrão | UFRGS | Titular |
| Prof. Dr. Adriano de Araújo Gomes | UFRGS | Prof. Orientador |

AGRADECIMENTOS

Aos órgãos de fomento CNPq e CAPES, pela manutenção das bolsas e apoio para congressos. Ao Instituto de Química da UFRGS, pelo espaço físico e suporte técnico. Aos professores Fabiano Rodembusch e Diogo Ludtke, em nome do PPGQ - UFRGS, pela excelente administração frente a COVID-19. E à Embrapa Uva e Vinho, bem como ao Concurso de Avaliação Nacional de Vinhos, pelas amostras cedidas.



Agradeço ao Prof. Marco Flôres Ferrão, pelas portas abertas e ouvidos pacientes. Ao meu orientador, Prof. Adriano de Araújo Gomes, por investir, guiar, instigar e inspirar em mim a vida de pesquisador.



Sou grato, em especial, à Fernanda Böck, Danielle Kochenborger e Kelly de Souza, pelo convívio e apoio nas tarefas em laboratório ao longo da pós-graduação.



Ao Josué Pereira Dias, meu namorado, pelas belas contribuições artísticas e pelos chás de camomila.

À Paola Del Vecchio, Estela Kochlrausch e ao Lázaro Aleixo, pela amizade e olhos escrutinadores.



Ao Luís Henrique Schneider e Anelise Pires Schneider, meus pais, pelo amor e suporte (inclusive financeiro), cruciais para que minhas preocupações fossem somente voltadas para o meu desenvolvimento intelectual.

Agradeço também à toda família Pires, pela oportunidade e confiança em poder continuar usufruindo do presente que o vô Lírio nos deixou.



Agradeço a Deus e às variáveis da Natureza que, me conscientizando do quão ínfimo é a meritocracia, convergiram para a conclusão deste sonho.

EPÍGRAFE



FONTE: Inspirado na obra “Última ceia” de Sarah Jenkins e Extrato de Recitativo Mt 26, 23-29, BWV 244 de Johann Sebastian Bach

RESUMO

Por meio do presente trabalho busca-se incrementar o conhecimento acerca do perfil dos vinhos tintos produzidos em território Sul-Rio-Grandense, de modo a possibilitar desenvolvimento de metodologias analíticas que sejam ambiental e economicamente sustentáveis e acessíveis. O entendimento dessas tecnologias pode facilitar o controle de qualidade e a identificação fraude exigido no setor. Quando dados de espectroscopia no infravermelho são aplicados no tratamento para reconhecimento de padrões, a quimiometria contribui na diferenciação entre safras de produção e origens de procedência. No presente trabalho, foram analisadas amostras de vinho tinto cedidas pela Empresa Brasileira de Pesquisa Agropecuária – Embrapa Uva e Vinho, totalizando 92 amostras de produtores locais para o reconhecimento de padrões e 4 amostras de monovarietal (3 Cabernet Sauvignon e 1 Merlot – produzidas em condições controladas) para os estudos de regressão. Os dados espectroscópicos obtidos foram tratados com derivação e suavização, para permitir avaliações concentradas no desempenho estatístico obtido da variação de parâmetros do tratamento de sinal derivada Savitsky Golay. Por meio de análise de componentes principais (PCA), foi possível diferenciar safras de produção entre as amostras, enquanto a diferenciação quanto à origem foi realizada a uma taxa de exatidão de 71% (SPA – LDA) e 81% (GA – LDA e PLS – DA). Estudos de regressão demonstraram que a abordagem quimiométrica permite quantificar variações de concentração de monovarietal dentro das quantidades previstas legalmente e critério para rotulagem. Mediante métodos de seleção de variáveis (iSPA – PLS e iPLS), foi possível encontrar modelos cuja taxa de erro de previsão do grupo de validação se aproxima a 12%. Em ambos estudos, as etapas de seleção de variáveis evidenciaram a relevância da concentração de álcoois e de compostos polifenólicos, o que reflete as diferenças nas condições geográficas e de composição dos monovarietais analisados.

Palavras-chave: Vinhos; Serra e Campanha Gaúcha; Reconhecimento de Padrões; Regressão multivariada; Seleção de variáveis.

ABSTRACT

The goal of the present work is to increase knowledge about red wine profile produced in Rio Grande do Sul, southern state of Brazil, in order to ease further development of analytical methodologies that turn to be environmental friendly and sustainable and affordable economically. Moreover, this methodology may integrate technologies to guarantee quality control against fraud, which are so important to the industrial field of food. After pattern recognition tools applied to infrared spectroscopy data, Chemometrics showed that it is possible to differentiate samples based on their harvest production and geographic origin. All results discussed on the present work became possible thanks to the donation of Embrapa Grape and Wine, which handed us 92 samples of locally produced red wine for pattern recognition and more 4 samples of specific cultivars (3 Cabernet Sauvignon and 1 Merlot – produced under controlled conditions) to regression studies. All performance evaluation focused on statistical effects caused from parameter variation of smoothing Savitsky – Golay derivatives. By PCA it was possible to differentiate harvest production years among samples, meanwhile origin designation has been achieved with an accuracy rate of 71% (SPA – LDA) and 81% (GA – LDA and PLS – DA). The chemometric approach of regression studies showed possibilities to quantify concentration variations of wine mixtures that, although legally provided, has labeling defined limits. Through variable selection algorithm (iSPA – PLS and iPLS), it was possible to quantify validation samples with a prediction rate of error about 12%. In both studies, variable selection showed that alcoholic and polyphenolic concentrations are crucial to the characterization, reflecting differences in terms of geographic conditions of production and grape species composition.

Keywords: Wine; Serra and Campanha Gaúcha; Pattern Recognition; Multivariate Regression; Variable selection.

LISTA DE FIGURAS

| | |
|--|----|
| Figura 1. (a) Catabolismo da glicose na fermentação alcoólica e (b) Conversão do ácido málico em ácido lático..... | 24 |
| Figura 2. Principais famílias de compostos fenólicos no vinho | 25 |
| Figura 3. Taninos no vinho..... | 26 |
| Figura 4. Insolação média ao longo dos meses de maturação vitícola | 28 |
| Figura 5. Unidades Geomorfológicas do Rio Grande do Sul | 29 |
| Figura 6. Precipitação Média Anual no RS | 30 |
| Figura 7. Regiões vitivinícolas de destaque no Rio Grande do Sul | 31 |
| Figura 8. Análises químicas em vinhos para controle de qualidade brasileiro..... | 39 |
| Figura 9. Infravermelho e resposta molecular vibracional | 45 |
| Figura 10. Espectro no infravermelho e regiões do <i>fingerprint</i> | 47 |
| Figura 11. Forma de organização matricial dos dados FTIR | 50 |
| Figura 12. Decomposição da matriz de dados..... | 53 |
| Figura 13. Modos de execução do algoritmo em PLS-DA..... | 55 |
| Figura 14. Probabilidades de classificação e Limite de decisão. | 56 |
| Figura 15. Comparação entre respostas dos métodos PCA e LDA..... | 57 |
| Figura 16. Ilustração da Análise Discriminante Linear (LDA)..... | 58 |
| Figura 17. Direção da máxima da variância em y e covariância entre x e y. | 59 |
| Figura 18. Seleção de variável no espectro infravermelho..... | 60 |
| Figura 19. Seleção de vetores projeção para seleção de variáveis. | 62 |
| Figura 20. Distribuição randômica dos resíduos de regressão. | 64 |
| Figura 21. Erro residual do modelo e a inclusão de fatores que o compõe | 65 |
| Figura 22. Relação de uvas e vinhos tintos produzidos no Rio Grande do Sul..... | 68 |
| Figura 23. Espectrômetro FTIR – ATR, Cary 630..... | 71 |
| Figura 24. Fluxograma do Estudo de Regressão | 72 |
| Figura 25. Esquema de tratamento das amostras de Regressão | 73 |
| Figura 26. Esquema do estudo de Reconhecimento de Padrões..... | 76 |
| Figura 27. Grandes regiões produtoras de uva e vinho | 77 |
| Figura 28. Absorbância com variação do caminho óptico | 81 |
| Figura 29. PCA, Dados Brutos; Origens | 83 |
| Figura 30. PCA, Dados pré-processados; Origens | 83 |
| Figura 31. PCA, Dados Brutos; Safras vs. Origens..... | 84 |

| | |
|---|-----|
| Figura 32. PCA, Dados Brutos; Safras vs. Origens..... | 84 |
| Figura 33. PCA, Dados Brutos; Safras | 85 |
| Figura 34. PCA, Dados pré-processados; Safras | 85 |
| Figura 35. Loadings PCA | 86 |
| Figura 36. Variação do Erro com as Variáveis Latentes | 87 |
| Figura 37. Média Geométrica do conjunto de modelos PLS – DA..... | 88 |
| Figura 38. Precisão do conjunto de modelos PLS – DA | 89 |
| Figura 39. Área abaixo da Curva para os modelos PLS – DA..... | 90 |
| Figura 40. Exatidão dos modelos PLS – DA..... | 91 |
| Figura 41. MCC dos modelos PLS – DA | 91 |
| Figura 42. Dados Brutos..... | 93 |
| Figura 43. Variáveis do espectro significativas para a separação de classes | 94 |
| Figura 44. Distribuição das amostras de treino e teste via PLS – DA..... | 96 |
| Figura 45. Relação Sensibilidade e Especificidade dos modelos SPA – LDA | 98 |
| Figura 46. Precisão do conjunto de modelos SPA – LDA | 99 |
| Figura 47. Média Geométrica do conjunto de modelos SPA – LDA..... | 100 |
| Figura 48. Área abaixo da Curva para os modelos SPA – LDA | 100 |
| Figura 49. MCC para conjunto de modelos SPA – LDA | 101 |
| Figura 50. Taxa de Falso Positivo nos modelos SPA – LDA | 101 |
| Figura 51. Taxa de Falso Positivo nos modelos PLS – DA | 101 |
| Figura 52. Espectro no Infravermelho após SG com 2ª Derivada..... | 103 |
| Figura 53. Distribuição de amostras de treino e teste segundo SPA – LDA..... | 105 |
| Figura 54. Aptidão do Algoritmo Genético aplicado aos dados X31..... | 107 |
| Figura 55. Aptidão do Algoritmo Genético aplicado aos dados X50..... | 107 |
| Figura 56. Frequência de variáveis selecionadas e espectro médio de X31 | 109 |
| Figura 57. Frequência de variáveis selecionadas e espectro médio de X50..... | 110 |
| Figura 58. Relação Sensibilidade e Especificidade dos modelos PLS – DA | 112 |
| Figura 59. Relação Sensibilidade e Especificidade dos modelos SPA – LDA | 112 |
| Figura 60. Média Geométrica do Conjunto Reduzido..... | 113 |
| Figura 61. Coeficiente de Correlação de Matthews do Conjunto amostral reduzido... | 113 |
| Figura 62. Gráfico de Loadings PLS – DA; Conjunto reduzido. | 115 |
| Figura 63. Perfil médio do IR e variáveis selecionadas por SPA – LDA..... | 115 |
| Figura 64. Perfil médio do IR e frequência de seleção de variáveis via GA – LDA.... | 116 |
| Figura 65. Comparação de desempenhos: conjunto integral vs. reduzido | 117 |

| | |
|--|-----|
| Figura 66. Funções Multiobjetivo | 119 |
| Figura 67. EJCR - MCS01..... | 121 |
| Figura 68. EJCR - MCS02..... | 121 |
| Figura 69. EJCR - MCS03..... | 122 |
| Figura 70. EJCR - MCS04..... | 122 |
| Figura 71. EJCR - MCS05..... | 123 |
| Figura 72. EJCR - MCS06..... | 123 |
| Figura 73. EJCR - MCS07..... | 124 |
| Figura 74. Seleção de intervalo em FTIR pré-processado em MCS01 | 126 |
| Figura 75. Seleção de intervalo em FTIR pré-processado em MCS02 | 126 |
| Figura 76. Seleção de intervalo em FTIR pré-processado em MCS03 | 127 |
| Figura 77. Seleção de intervalo em FTIR pré-processado em MCS04 | 127 |
| Figura 78. Seleção de intervalo em FTIR pré-processado em MCS05 | 128 |
| Figura 79. Seleção de intervalo em FTIR pré-processado em MCS06 | 128 |
| Figura 80. Seleção de intervalo em FTIR pré-processado em MCS07 | 129 |
| Figura 81. Seleção de intervalo em FTIR pré-processado em MCS07b | 129 |
| Figura 82. Regiões de calibração, validação e pontos de anomalia | 132 |
| Figura 83. Amostras de Validação MCS01 | 133 |
| Figura 84. Amostras de Validação MCS02 | 134 |
| Figura 85. Amostras de Validação MCS03 | 134 |
| Figura 86. Amostras de Validação MCS04 | 135 |
| Figura 87. Amostras de Validação MCS05 | 135 |
| Figura 88. Amostras de Validação MCS06 | 136 |
| Figura 89. Amostras de Validação MCS07 | 136 |

LISTA DE TABELAS

| | |
|---|-----|
| Tabela 1. Compostos fenólicos flavonoides | 26 |
| Tabela 2. Estrutura das Antocianidinas | 27 |
| Tabela 3. Parâmetros de controle dos vinhos tintos provenientes de IG gaúchas. | 33 |
| Tabela 4. Regras de Indicação de Procedência ou Denominação de Origem | 35 |
| Tabela 5. Parâmetros da composição de vinhos previstos pela legislação brasileira | 37 |
| Tabela 6. Regiões significativas da absorção no infravermelho médio. | 47 |
| Tabela 7. Opções de prismas disponíveis no mercado para FTIR – ATR..... | 49 |
| Tabela 8. Concentrações de monovarietais presentes nas misturas | 69 |
| Tabela 9. Pré-processamento dos dados FTIR – ATR | 75 |
| Tabela 10. Composição do conjunto amostral..... | 78 |
| Tabela 11. Comparação de desempenhos de modelos PLS – DA..... | 92 |
| Tabela 12. Desempenho de modelos selecionados na etapa de predição via PLS – DA | 95 |
| Tabela 13. Desempenho de modelos selecionados na etapa de predição via SPALDA. | 102 |
| Tabela 14. Desempenho de modelos SPA – LDA selecionados para teste. | 104 |
| Tabela 15. Desempenho de Grupos de Teste e variáveis significativas..... | 108 |
| Tabela 16. Conjunto de amostras reduzido | 111 |
| Tabela 17. Desempenho dos algoritmos de Reconhecimento de Padrão | 114 |
| Tabela 18. Algoritmos de Regressão em conjuntos de dados selecionados..... | 130 |

LISTA DE SIGLAS

| Sigla | Português | Inglês |
|-----------------|--|---|
| DO | Denominação de Origem | <i>Designation of Origin</i> |
| EJCR | Região Elíptica de Confiança Conjunta | <i>Elliptical Joint Confidence Region</i> |
| FAAS | Espectrômetro de absorção atômica de chama | <i>Flame Atomic Absorption Spectrometry</i> |
| FTIR– | Infravermelho com transformada de Fourier | <i>Fourier transformed infrared with</i> |
| ATR | por refletância total atenuada | <i>attenuated total reflectance</i> |
| GA–LDA | Análise Discriminante Linear com Algoritmo Genético | <i>Genetic Algorithm based Linear Discriminant Analysis</i> |
| HPLC | Cromatografia líquida de alta performance com detector de: | <i>High performance liquid chromatography with:</i> |
| | DAD Arranjo de diodos | <i>Diode array detector</i> |
| | MS Espectrometria de massas | <i>Mass spectrometry</i> |
| | RID Índice de refração | <i>Refractive index detector</i> |
| | UV/Vis Região do espectro eletromagnético do ultravioleta ao visível | <i>Electromagnetic spectra comprising ultraviolet and visible regions.</i> |
| ICP–OES | Espectrometria de Emissão Óptica com Plasma Indutivamente Acoplado | <i>Inductively Coupled Plasma – Optical Emission Spectrometry</i> |
| IP | Indicação de Procedência | <i>Indication of Origin</i> |
| iPLS | Regressão por Mínimos Quadrados Parciais em intervalos | <i>Interval partial least-squares regression</i> |
| IR | Infravermelho | <i>Infrared</i> |
| | MID Região média do espectro | <i>Mid spectral region</i> |
| iSPA–PLS | Regressão por Mínimos Quadrados Parciais com Algoritmo de Projeções Sucessivas em intervalos | <i>Partial Least Squares regression coupled with the successive projection algorithm for interval selection</i> |
| PCA | Análise de Componentes Principais | <i>Principal Component Analysis</i> |
| PLS(R) | Regressão por Mínimos Quadrados Parciais | <i>Partial Least Squares regression</i> |
| PLS–DA | Análise Discriminante por Mínimos Quadrados Parciais | <i>Discriminant analysis by partial least squares</i> |
| SPA–LDA | Análise Discriminante Linear com Algoritmo de Projeções Sucessivas | <i>Linear discriminant analysis with successive projections algorithm for spectral selection</i> |

Sumário

| | |
|--|----|
| 1. INTRODUÇÃO..... | 18 |
| 1.1 Objetivos..... | 19 |
| 1.1.1 Objetivos Específicos | 19 |
| 2. REVISÃO DA LITERATURA | 22 |
| 2.1 Química do vinho – Composição e fatores de alteração..... | 22 |
| 2.1.2 Composição polifenólica: Antocianinas e Taninos | 25 |
| 2.1.3 Maturação da Uva e os fatores climáticos | 27 |
| 2.2 Geografia do Rio Grande do Sul e a Vitivinicultura | 28 |
| 2.3 Indicações Geográficas e Regiões Produtoras Gaúchas | 30 |
| 2.4 Legislações Brasileiras | 33 |
| 2.5 Análises instrumentais e certificação de origem..... | 36 |
| 2.6 Radiação eletromagnética e o Infravermelho | 44 |
| 2.6.1 Radiação eletromagnética..... | 44 |
| 2.6.2 Infravermelho | 45 |
| 2.6.3 Origem da Informação..... | 46 |
| 2.6.4 Espectroscopia no infravermelho aplicado a vinhos tintos | 46 |
| 2.6.5 Formas de medida no infravermelho médio | 47 |
| 2.7 Quimiometria..... | 49 |
| 2.7.1 Reconhecimento de Padrões..... | 51 |
| 2.7.2 Regressão..... | 58 |
| 2.7.3 Seleção de variáveis..... | 60 |
| 2.7.4 Avaliação de desempenho de modelos..... | 62 |
| 3. METODOLOGIA..... | 67 |
| 3.1 Amostras | 67 |
| 3.1.1 Preparo de amostras..... | 67 |
| 3.2 Registro dos Espectros IR..... | 70 |

| | | |
|-------|---|-----|
| 3.3 | Processamento de Dados | 71 |
| 3.3.1 | Regressão..... | 71 |
| 3.3.2 | Reconhecimento de Padrões..... | 76 |
| 4. | RESULTADOS E DISCUSSÃO | 81 |
| 4.1 | Escolha do Caminho óptico | 81 |
| 4.2 | Reconhecimento de padrão | 82 |
| 4.3 | Estudos extra em Reconhecimento de Padrões..... | 111 |
| 4.3.1 | Efeitos na redução do conjunto amostral..... | 111 |
| 4.4 | Regressão | 118 |
| 5. | CONCLUSÕES | 140 |
| 6. | PERSPECTIVAS FUTURAS | 143 |
| | REFERÊNCIAS | 144 |
| | Apêndice A – Vinhos Embrapa | 161 |
| | Apêndice B – Datas e Horários no FTIR – ATR..... | 164 |
| | Apêndice C – Participações em Produção Científica | 166 |
| | Apêndice D – Participação em eventos | 168 |



Por Anelise Maria Wasem Pires Schneider (2020)

1. INTRODUÇÃO

O vinho tem acompanhado a sociedade ao longo dos tempos. Suas características e métodos de processamento são adaptações e melhoramentos construídos no mesmo passo que o humano valoriza o poder do conhecimento. Cada país possui um particular entendimento e legalidade do que configura a composição. Além disso, é também particular de cada localidade se adaptar às condições geográficas de modo a desenvolver produtos de qualidade única.

Uma das preocupações do setor vitivinícola é a fraude, haja vista o apelo pela qualidade única relacionada às origens. A fraude é um artifício usado para o lucro indevido e cujas técnicas e táticas estão em constante renovação. Desvios no processo industrial, apesar de possivelmente involuntários, também integram os fatores que justificam a necessidade de metodologias para o controle de qualidade produtivo. Contudo, enquanto ações fraudulentas colocam em risco a reputação de uma empresa, o interesse de mercado é mais voltado à procura por técnicas de fiscalização de baixo custo, mas que assegurem a integridade de sua produção.

Frente a todas as técnicas analíticas consolidadas, a espectroscopia no infravermelho se apresenta como alternativa complementar quando aliada às ferramentas matemáticas, estatísticas e computacionais da quimiometria. A sinergia dessas áreas integradas oferece atrativos às instituições, como redução tanto no consumo de reagentes como na manutenção e operação de equipamentos caros e complexos, além de agilizar as tomadas de decisão para aplicações analíticas mais específicas.

Dadas preocupações são pertinentes ao setor da vitivinicultura do Estado do Rio Grande do Sul. Com regiões consolidadas e outras em crescente desenvolvimento na qualificação dessa produção, é de interesse do setor público e privado que o vinho gaúcho seja quimicamente caracterizado, aperfeiçoado segundo as possibilidades impostas pelas condições geográficas da localidade e, não menos importante, que seja fiscalizado, certificando adequação para consumo e origem de procedência.

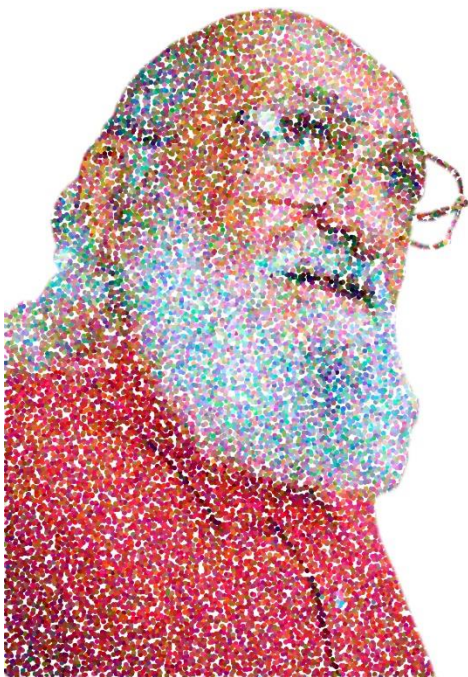
Assim, o conhecimento dos perfis do vinho local se faz necessário para a valorização sócio – cultural e econômica das regiões produtoras do Rio Grande do Sul. Para tanto, o presente trabalho se propõe a identificar a procedência geográfica e quantificar proporções de misturas de monovarietais mediante aplicação de informação obtida por uma técnica analítica secundária, porém simples e de baixo custo, aliada às ferramentas de tratamento de dados multivariados.

1.1 Objetivos

Avaliar a origem geográfica e perfil varietal dos vinhos tintos produzidos em duas grandes regiões Gaúchas, combinando espectros FTIR – ATR e ferramentas da quimiometria.

1.1.1 *Objetivos Específicos*

- ✓ Observar tendências e desempenhos ao aplicar no conjunto de espectros a técnica de valorização da razão sinais ruído do tipo Derivada de Savitzky Golay, alternando valores do grau de polinômio (2 ou 3) e derivada (1^a ou 2^a) a uma gama de valores da janela (partindo de 7 até 33 pontos).
- ✓ Identificar padrões de agrupamento do conjunto amostral via método de classificação não-supervisionado, a análise de componentes principais – PCA.
- ✓ Observar efeito dos diferentes pré-tratamentos espectrais discriminados no objetivo anterior e extrair informação discriminatória de classes.
- ✓ Por métricas da classificação, avaliar o desempenho de métodos de reconhecimento de padrões supervisionados aplicados ao espectro como um todo e por seleção de variáveis; respectivamente PLS – DA e SPA – LDA. Os modelos considerarão, além da informação espectral, a procedência geográfica do vinho.
- ✓ Aplicar conjuntos de dados ao reconhecimento de padrões via Algoritmo Genético. Comparar desempenho dessa classificação frente aos resultados obtidos por meio das abordagens supracitadas.
- ✓ Quantificar variações de concentração em misturas de vinhos de tipos distintos de uvas (monovarietais) por FTIR – ATR e métodos de Regressão que considerem o espectro todo, como o PLS, ou que façam seleção de variáveis (iSPA – PLS e iPLS).
- ✓ Observar a influência do pré-processamento de dados Derivada de Savitzky Golay, via desempenho estatístico na validação cruzada, na eficiência em quantificar corretamente as misturas que compõem o modelo.
- ✓ Avaliar a habilidade de quantificação apresentada pelos modelos de desempenho adequados na validação interna, quando submetidos a amostras de concentração desconhecida.



“
Não basta ler mecanicamente “Eva viu a uva”.
É necessário compreender qual posição
Eva ocupa no seu contexto social,
quem trabalha para produzir uvas
e quem lucra com esse trabalho.”

Paulo Freire

2. REVISÃO DA LITERATURA

2.1 Química do vinho – Composição e fatores de alteração

De acordo com a Organização Internacional da Uva e Vinho e regramento legal Brasileiro, o Vinho é a bebida obtida pela fermentação alcoólica do mosto simples da uva sã, fresca e madura ^{1,2}. Para além das definições técnicas ou legais, o vinho é uma mistura complexa resultante da atividade conjunta entre natureza e humano ³.

A uva responde majoritariamente pela composição do vinho, seja fornecendo componentes da polpa quanto da casca. Frutos pertencentes à família botânica *Vitaceae*, possuem diferentes estruturas, composições e indicações para melhor consumo. De maior importância agrônômica, o gênero *Vitis* engloba a espécie *vitis labrusca* (nativa da América do Norte) e *vitis vinífera* (europeia tradicional). A fabricação da maior parte dos vinhos do mundo moderno resulta dessa última espécie, sob a qual será baseado o presente trabalho ⁴⁻⁶.

Cada espécie e variedade de uva responde diferentemente às múltiplas variáveis, componentes do *terroir*. Ou seja, o clima, solo e a intervenção humana nos processos de vinificação afetam as composições químicas e metabólicas da uva e, por conseguinte, na exigência de controle de qualidade organoléptica precisam ser consideradas para a excelência do produto final ⁷⁻¹¹.

Depois de atingir o estágio de maturação, determinado pelo acompanhamento da concentração de açúcares, acidez e principais polifenóis – antocianinas e taninos, as uvas são colhidas e esmagadas e maceradas. O líquido resultante desse processo leva o nome de mosto que, na produção de vinhos tintos, é fermentado na presença das cascas, para então ser filtrado ¹²⁻¹⁴.

Similar à composição da polpa, o mosto é rico em água, açúcares, ácidos, compostos nitrogenados, vitaminas e minerais. Esse meio é subsídio para o crescimento e desenvolvimento dos micro-organismos envolvidos no processo de fermentação ¹⁵. A suplementação se torna necessária à medida que o mosto é composto por quantidades reduzidas de nitrogenados, assim evitando fermentações lentas ou a formação de ácido sulfídrico e mercaptanas, que conferem odores desagradáveis ^{14,16}.

O desempenho da fermentação alcoólica também é dependente da composição de substâncias minerais do mosto. Extraídos do solo pelas raízes das videiras, esses minerais têm sua concentração reduzida ao decorrer da fermentação, por precipitação e filtração de sólidos e coloides. Todavia, adições intencionais ou oriundas da composição

dos tanques de vinificação podem aumentar a concentração de alguns desses íons, justificando seu controle para que a qualidade do vinho não seja comprometida.

Tanto o mosto quanto o vinho apresentam o grupo de vitaminas hidrossolúveis do complexo B (B1, B2, B3, B5, B6, B7, B8, B9 e B12), vitamina C (ácido ascórbico) e vitamina P (flavonoides). A presença de tais vitaminas é essencial no desenvolvimento das leveduras e bactérias durante a vinificação. Ainda, as vitaminas componentes do vinho configuram o fator que induz a discussões frequentes quanto a considerar a bebida um alimento ou não.

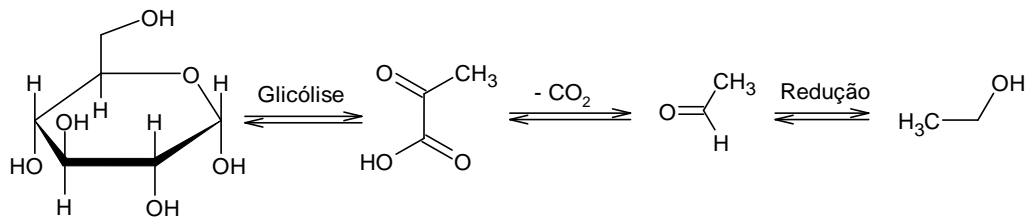
Dentre as famílias químicas que compõem o mosto, os açúcares têm um papel importante na conversão alcoólica. Os mais abundantes desse grupo são os monossacarídeos com 6 carbonos: glicose e frutose. Ambos açúcares são encontrados, praticamente, nas mesmas concentrações. Em níveis de concentração dependentes da espécie de uva, a sacarose é outro açúcar que participa do processo de vinificação, porém, à medida que sua ligação glicosídica é dividida, forma-se frutose e glicose.^{17,18}

A adição exógena de sacarose ao mosto não fermentado é uma prática comum. Conhecida como *chaptalização*, essa adição tem como intuito a elevação da graduação alcoólica do produto final. Outros açúcares que podem ser identificados no mosto e ao longo da fermentação incluem arabinose, xilose, ramnose, maltose e rafinose. Por fim, os açúcares do mosto também são metabolizáveis a álcoois superiores, ésteres de ácidos graxos e aldeídos¹⁸.

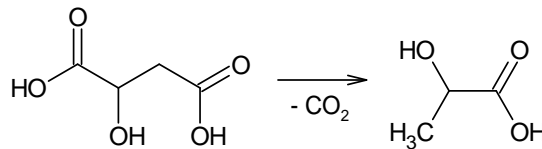
Como esquematizado na Figura 1 (a), a fermentação pode ser entendida como a produção de ácido pirúvico partindo da glicose e a subsequente conversão desse produto em etanal, sendo por fim reduzido a etanol. Na prática, o rendimento da respiração e fermentação das leveduras *Saccharomyces cerevisiae* na produção de etanol é baixo. Reações secundárias formam coprodutos como: glicerol, ácido etanoico, ácido succínico e 2,3-butenodiol, acetona, diacetila, ácido butírico, ácido láctico, ácido aceto acético e acetona¹⁹.

Concomitante ou posteriormente ao processo de fermentação dos açúcares, ocorre a conversão do ácido málico em láctico Figura 1 (b). Essa redução é uma reação enzimática comumente chamada de fermentação malolática e é um recurso para a diminuir a acidez do meio e elevar a complexidade do vinho. Por questões de qualidade organoléptica, essa reação é quase obrigatória na produção de vinhos tintos.

a)



b)



Fonte: Adaptado de *Wine Chemistry and Biochemistry* ¹⁴

Figura 1. (a) Catabolismo da glicose na fermentação alcoólica e (b) Conversão do ácido málico em ácido láctico

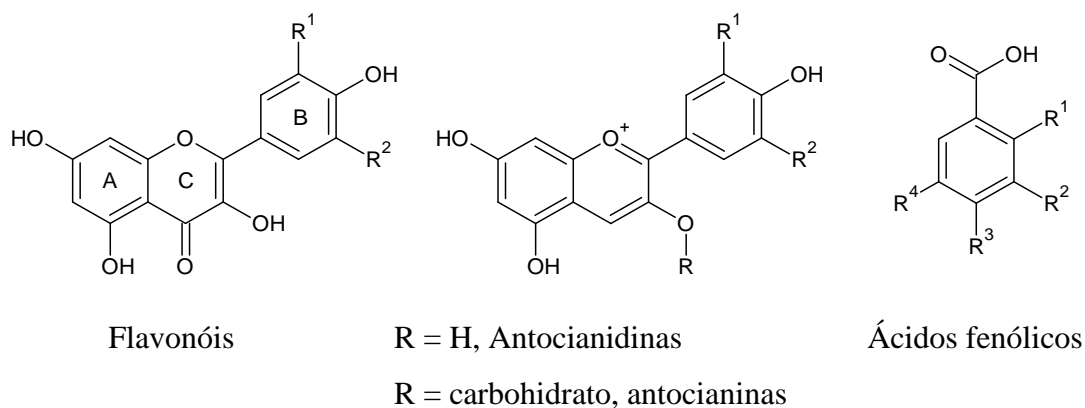
Durante a vinificação, polifenóis e compostos aromáticos são extraídos das cascas da uva e dos barris de envelhecimento, conferindo cor e aroma ao produto final ¹⁹. Isso pode ser atribuído a uma maior afinidade do meio por essas substâncias, visto o aumento da concentração de etanol ao longo da fermentação ²⁰.

A concentração de compostos minoritários no vinho inclui álcoois superiores, compostos carbonílicos e ésteres. Os álcoois superiores são divididos entre álcoois majoritários (1-propanol, isobutanol, álcoois isoamilicos, 2-feniletanol) e demais álcoois (butanol, pentanol, hexanol, etc). Os compostos carbonílicos são representados pelas concentrações de etanal, propanal, hexanal, benzaldeído, 2,3-butanodiona e 3-hidroxi-2-butanona. Por fim, os ésteres presentes são classificados de acordo com os radicais ligantes à estrutura do grupo funcional éster: derivados do ácido acético (os mais voláteis), etil ésteres (tartarato, malato, lactato e monoetil e dietil succinato) e demais ésteres (derivados de outros álcoois, além de etanol).

Além das variações de composição decorrentes da espécie de uva, origem geográfica e etapas da vinificação, outros componentes podem afetar as características do vinho a ponto de torná-lo único. O vinho é transferido para um tanque limpo, onde passa pelo processo de clarificação, eliminando componentes causadores de turvamento. O amadurecimento, quando empregado, pode ser feito em tanques de madeira ou aço inox. Finalmente, há o processo de estabilização, filtração final e engarrafamento ^{21,22}.

2.1.2 Composição polifenólica: Antocianinas e Taninos

A coloração e demais características das uvas e de seus vinhos indica a presença de diferentes compostos, entre eles os polifenóis. Apesar de não participarem dos processos de crescimento e desenvolvimento dos micro-organismos da fermentação, os polifenóis têm um papel essencial na caracterização sensorial do mosto e do vinho ¹⁴.



Fonte: Adaptado de Pisano e colaboradores ¹⁰

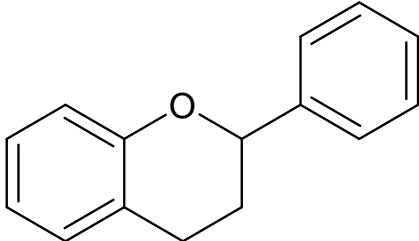
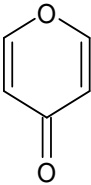
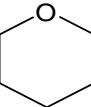
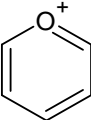
Figura 2. Principais famílias de compostos fenólicos no vinho

Os polifenóis apresentam inúmeras estruturas e funções químicas (representadas simplificada na Figura 2) e, através dessas, são classificados como:

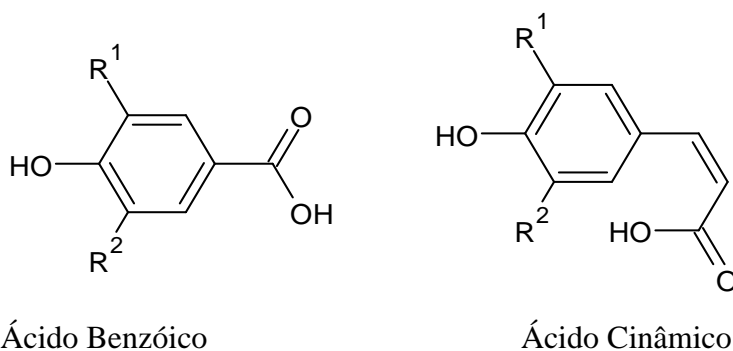
- Flavonoides: englobam os flavanóis, flavonóis, antocianidinas, flavonas e flavanonas. Como demonstrado na Tabela 1, compartilham de uma estrutura similar, que consiste em dois anéis fenólicos (A e B) ligados através de um anel heterocíclico de pirano (C);
- Não-flavonoides: simples fenóis, constituídos de anel benzênico. Apesar de suas cadeias carbônicas relativamente menores, estilbenos também estão inclusos no grupo de não-flavonoides.
- Taninos: polimerizações de moléculas de compostos fenólicos, como representado na **Figura 3**, grandes o suficiente para formar complexos estáveis com proteínas. Conferem amargor e coloração marrom ao vinho;^{10,14}

Os compostos fenólicos que mais contribuem com as tonalidades e nuances presentes no vinho tinto são os flavonóis e as antocianinas. Apesar de contribuir pouco para a coloração do vinho tinto, os flavonóis são pigmentos amarelos provenientes das cascas das uvas ²³. Já as antocianinas são os pigmentos responsáveis pela coloração avermelhada dos vinhos tintos.

Tabela 1. Compostos fenólicos flavonoides

| Estrutura dos Flavonoides | Grupo flavonoide | Heterociclo | Estrutura |
|---|------------------------------|-------------|---|
|  | Flavonóis e dihidroflavonóis | Pirona |  |
| | Flavanas e flavanóis | Pirano |  |
| | Antocianidinas | Pirrol |  |

Fonte: Adaptado de Soleas e Colaboradores (1997) ¹⁸



Fonte: Adaptado de Soleas e Colaboradores (1997) ¹⁸

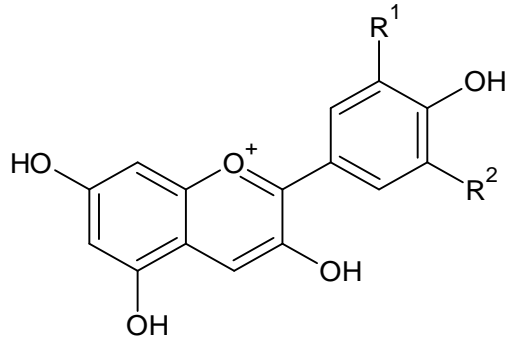
Figura 3. Taninos no vinho

Antocianina é o termo usado para designar substâncias específicas do grupo dos flavonóis: as antocianidinas e as antocianinas intrinsecamente. Enquanto as antocianidinas são agliconas de estrutura definida, as antocianinas são a antocianidinas ligadas a um açúcar e ácidos orgânicos, através de uma ligação glicosídica. Mais especificamente, como demonstrado na Tabela 2, as antocianinas são os derivados polihidroxilados e polimetoxilados do cátion 2-fenilbenzopirílio, o cátion flavílio.

Devido a variação de concentrações das antocianinas minoritárias, aliadas a mais abundante em todas as variedades de uvas (malvidina), podemos observar diferentes nuances na coloração rubra do vinho tinto. Dado fenômeno resulta também do efeito de íons e outras moléculas presentes no vinho, que interagem com as moléculas de antocianinas ^{22,24}.

Tabela 2. Estrutura das Antocianidinas

| Antocianidina | R1 | R2 |
|---------------|------------------|------------------|
| Cianidina | OH | H |
| Peonidina | OCH ₃ | H |
| Delfinidina | OH | OH |
| Petunidina | OCH ₃ | OH |
| Malvidina | OCH ₃ | OCH ₃ |



Fonte: Adaptado de Kähkönen e Colaboradores (2003) ²⁵

Taninos são produtos de uma polimerização de moléculas de compostos fenólicos menores que, como as antocianinas, estão presentes na casca das uvas e são extraídos dessa para o mosto ao longo da vinificação. Uma vez que há concentrações desses compostos nas sementes, o tempo de maceração influencia na concentração de taninos no produto final ^{26,27}.

2.1.3 *Maturação da Uva e os fatores climáticos*

Quando destinada para produção de vinhos, o momento ideal para colheita precisa considerar o estágio de maturação das uvas. Algumas etapas são comuns e guiam para melhor aproveitamento dos frutos da videira, tanto visando melhores desempenhos na fermentação alcoólica quanto no conteúdo de componentes fenólicos ²⁸.

A composição fenólica e de açúcares e ácidos em cada estágio da maturação da uva reflete as condições de microclima da região produtora, ou seja, são consequência, entre outros, da topografia local, vegetação e solo. A maturação da uva é favorecida por uma baixa pluviosidade, principalmente nos subestágios finais de maturação. Isso significa que as uvas não absorverão água a ponto de diluir seus componentes internos o que reduz necessidades de adições exógenas ²⁸⁻³¹.

Outro fator climático de grande importância é a amplitude térmica. Na maturação é ideal uma amplitude térmica grande, ou seja, dias ensolarados e mais quentes que as noites. O desenvolvimento dos frutos se dá, assim, de maneira lenta e controlada, com balanço de açúcares e ácidos e uma maior extratibilidade dos componentes fenólicos ^{29,30,32,33}.

Por tanto, além dos processos de fermentação, o vinho reflete a escolha da uva, o local de cultivo e variáveis relativas ao clima de uma região. As etapas de controle de

qualidade, assim, precisam ser consolidadas a ponto de garantir universalidade de parâmetros, mas principalmente, precisam seguir adaptações e adequações que facilitem a aquisição e comparação dessas métricas.

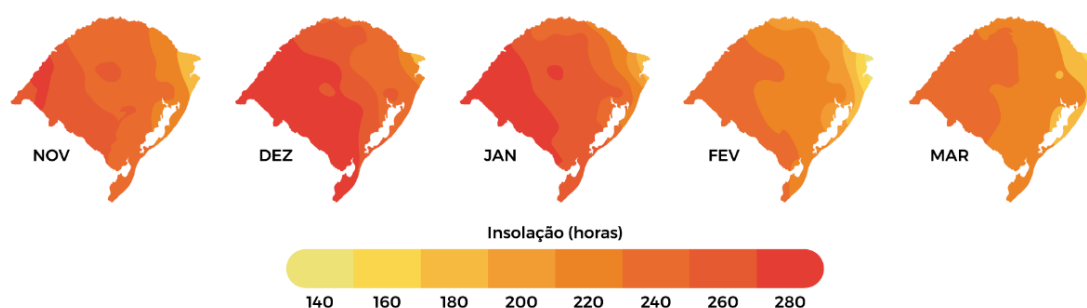
2.2 Geografia do Rio Grande do Sul e a Vitivinicultura

O Estado do Rio Grande do Sul está no extremo sul do Brasil, assim apresenta grandes diferenças geográficas com relação ao restante do país. O clima do Estado é Temperado do tipo Subtropical. A posição geográfica reforça as influências das massas de ar oriundas da região polar e da área tropical continental e Atlântica, influenciando as taxas de precipitação, umidade e variações de temperatura ao longo do ano.

A composição do solo, bem como a altitude dessas regiões, ambas resultam de fenômenos geográficos ocorridos ao longo dos milênios. Além da particularidade climática por reflexo da latitude, as regiões produtoras de uva e vinho no Estado se localizam em zonas geomorfológicas distintas e características, conforme **Figura 5**, o que interfere no nível de industrialização do sistema produtivo e nas potencialidades de desempenho para cada local.

Os níveis pluviométricos anuais e de insolação nos meses de maturação das uvas corroboram na diferenciação geográfica de duas grandes regiões. A precipitação média anual nas regiões da porção Norte e Nordeste são as maiores do território gaúcho, contrapondo o que se observa na região limítrofe ao Uruguai (**Figura 6**). Possivelmente diretamente relacionados com a quantidade de chuva, os índices de insolação são maiores no Sul e Sudoeste, quando em comparação com a Norte e Nordeste (**Fonte:** Adaptado de Atlas Climático do RS ³⁴. Arte de Josué Pereira Dias

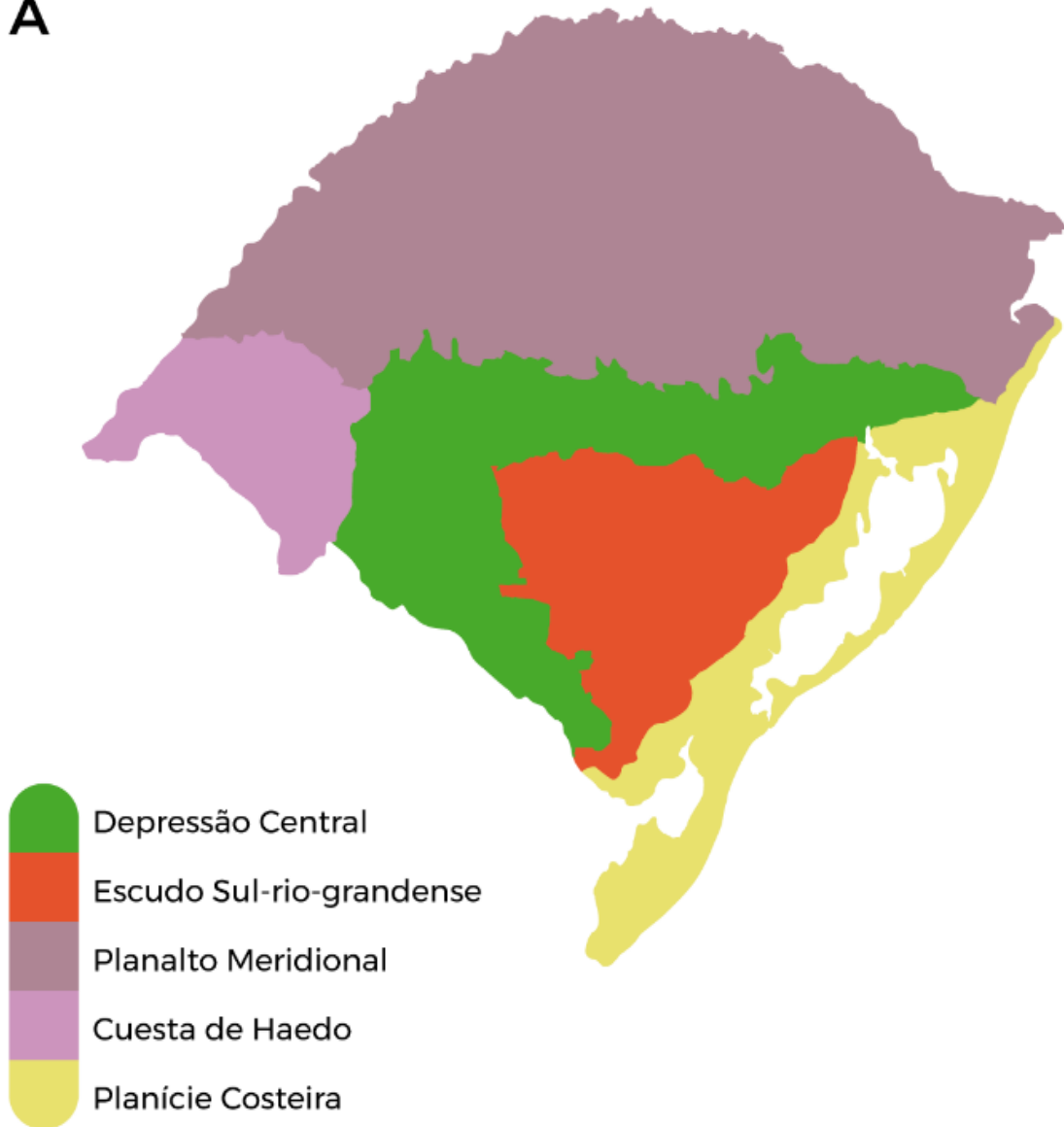
Figura 4).



Fonte: Adaptado de Atlas Climático do RS ³⁴. Arte de Josué Pereira Dias

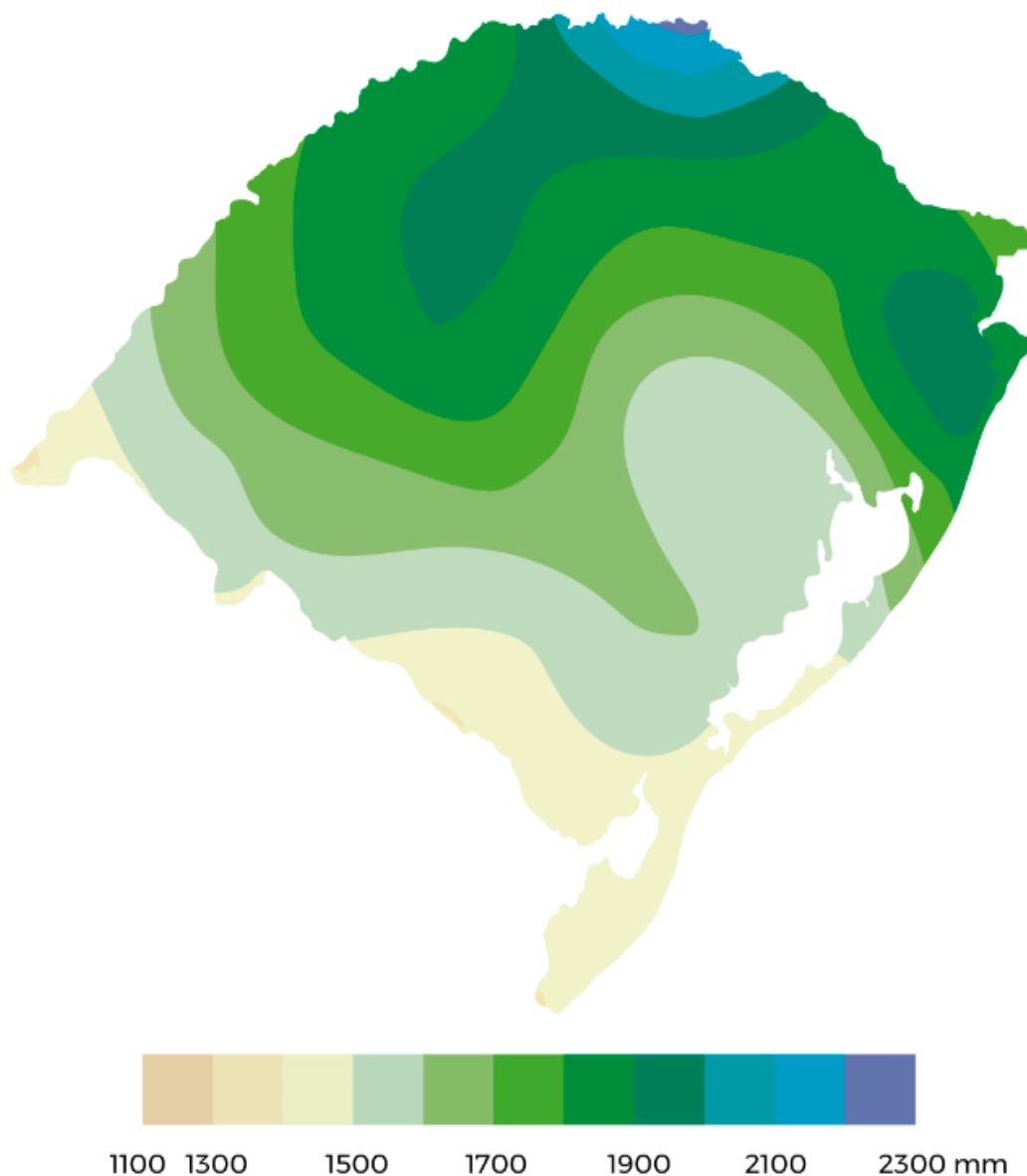
Figura 4. Insolação média ao longo dos meses de maturação vitícola

A



Fonte: Adaptado de Atlas Climático do RS ³⁴. Arte de Josué Pereira Dias
Figura 5. Unidades Geomorfológicas do Rio Grande do Sul

B



Fonte: Adaptado de Atlas Climático do RS ³⁴. Arte de Josué Pereira Dias
Figura 6. Precipitação Média Anual no RS

2.3 Indicações Geográficas e Regiões Produtoras Gaúchas

O conceito de qualidade é relativo e continuamente mutável, inclusive se relacionando com fatores de percepção e julgamento social e cultural. Além do reflexo de um conjunto de atributos físico-químicos inerentes à natureza do produto, a legislação brasileira atua na proteção à propriedade industrial das chamadas indicações geográficas, baseadas na existência de qualificações que diferenciam uma região ou território na elaboração de um dado produto ³⁵.

Para que produções apresentem um signo de indicação geográfica, exige-se que a cadeia produtiva envolvida siga uma regulamentação definida e específica. Depois de um estudo das condições climáticas e territoriais locais, ficam delineadas as áreas produtivas de uva, bem como definidas as melhores espécies de uvas possíveis para a viticultura.

A lei federal n.º 9279 (1996) viabiliza a proteção de produção industrial da viticultura ao definir objetos de indicação de procedência (IP) e de Denominação de Origem (DO) ³⁶. Resumidamente, ambos designam o nome da região geográfica produtora do vinho. Enquanto IP refere-se à localidade reconhecida como centro de extração e produção, considera-se DO o vinho cujas qualidades ou características se devam exclusiva ou essencialmente ao meio geográfico, incluindo fatores naturais e humanos.

Dentre as regiões vitivinícolas brasileiras, o Estado do Rio Grande do Sul é onde cerca de 90% de todo vinho nacional é produzido. De acordo com o Instituto Brasileiro do Vinho, dentro do território gaúcho existem seis regiões produtoras, das quais quatro são as mais renomadas e representadas na Figura 7: Serra do Sudeste, Campanha, Serra Gaúcha e Campos de Cima da Serra ³⁷. Na Figura 7, conjuntamente há uma aproximação dessas localidades considerando as atribuições geográficas.

Para o Rio Grande do Sul, a vitivinicultura tem grande importância social e econômica uma vez que, apesar da bebida não ser de grande preferência brasileira, o setor emprega aproximadamente oitenta mil pessoas e movimentando anualmente cerca de 3 bilhões de Reais. Assim, é de interesse do Estado e do setor privado que a produção local seja, ao menos, protegida de descaminhos e contrabandos, já que decisões políticas correntes tendem a beneficiar a importação de produtos Europeus ³⁸⁻⁴².

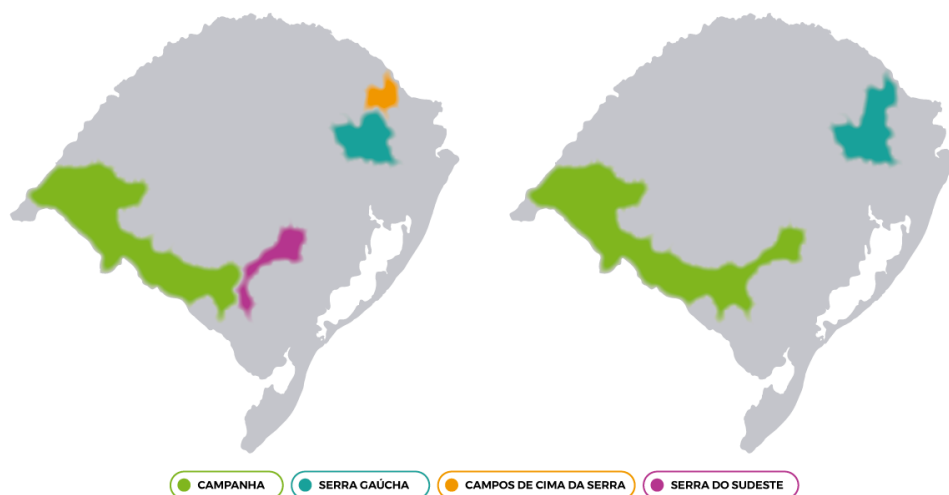


Figura 7. Regiões vitivinícolas de destaque no Rio Grande do Sul

Situada na também conhecida como Escudo Sul-Rio-Grandense, a Serra do Sudeste teve plantações experimentais entre as décadas de 1970 e 1990, que melhor se desenvolveram a partir dos anos 2000. Caracterizada por suas coxilhas e superfícies planas, de altitude não maior que 750 metros acima do nível do mar. Seu solo antigo é rico em calcário e granito. O clima da região é moderado, com verões secos e temperaturas levemente superiores às da Serra Gaúcha, destino de grande parte da produção vitícola local ^{6,37}.

Como região gaúcha mais promissora, a Campanha compreende a extensão territorial do Rio Grande do Sul que faz divisa com o Uruguai e Argentina. Mesmo que historicamente conhecida pelo cultivo de arroz e criação de gado, a vitivinicultura tem se mostrado um atrativo econômico desde meados de 1970, culminando inclusive no registro de Indicação de Geográfica. Essa região apresenta terreno propício para a mecanização dos processos no vinhedo, solos argilosos e ricos em granito, elevada amplitude térmica e clima seco e quente, que se torna mais frio na direção noroeste ^{6,37}.

Os vinhedos de Campos de Cima da Serra se encontram a 975 metros do nível do mar, em uma localidade de clima frio e ventoso. Por muito tempo, essa região esteve às sombras da localidade próxima, a Serra Gaúcha, mas o entendimento das condições ambientais possibilitou a produção de vinhos renomados ³⁷.

Respondendo por 85% da produção nacional de vinhos, a Serra Gaúcha é a maior e mais importante região vinícola do Brasil. Aproximadamente de 580 a 800 metros do nível do mar, seu solo é basáltico e o clima é temperado e úmido, com temperatura média anual de 17°C. A Serra Gaúcha é composta de microrregiões de indicação geográfica, ou seja, possuem reconhecimento de qualidade ^{6,37}.

Até o presente momento, apenas a região sul do Brasil contém indicações geográficas reconhecidas pelo Instituto Nacional da Propriedade Intelectual. Dentre as 7 indicações na produção de vinhos, ao todo 5 são referentes a produtoras gaúchas de vinhos tintos: Altos Montes, Campanha Gaúcha, Monte Belo, Pinto Bandeira e Vale dos Vinhedos ⁴³⁻⁵¹.

Em termos de cultivares (espécies de uvas) permitidas e sua concentração recomendada na produção de vinhos, cada indicação geográfica apresenta particularidades nos registros de especificações técnicas, como pode ser observado nas Tabela 3 e Tabela 4.

Tabela 3. Parâmetros de controle dos vinhos tintos provenientes de IG gaúchas.

| <i>Parâmetros</i> | <i>AM</i> | <i>CA</i> | <i>MB</i> | <i>PB</i> | <i>VV</i> |
|---------------------------------------|-------------------|-----------|-----------|-----------|-----------|
| <i>Acidez volátil (mEq/ℓ)</i> | ≤ 15 | ≤ 15 | ≤ 12 | ≤ 15 | ≤ 14 |
| <i>Anidrido sulfuroso total (g/ℓ)</i> | ≤ 0,15 | ≤ 0,13 | 0,15 | ≤ 0,13 | ≤ 0,15 |
| <i>Polifenóis totais</i> | ≥ 50 | | | | ≥ 50 |
| <i>Intensidade de cor</i> | | | | | ≥ 1,2 |
| <i>Acidez total (mEq/ℓ)</i> | ≤ 90 | | | | |
| ¹³ C/ ¹² C | ✓ | | | ✓ | |
| ✓ | Requerida | | | | |
| <i>AM</i> | Altos Montes | | | | |
| <i>CA</i> | Campanha | | | | |
| <i>MB</i> | Monte belo | | | | |
| <i>PB</i> | Pinto Bandeira | | | | |
| <i>VV</i> | Vale dos Vinhedos | | | | |

Fonte: Adaptada dos regulamentos de uso da Indicação de Procedência ^{45,52-55}

Sendo assim, ainda que dentro de uma região de identificação geográfica, percebe-se que a definição de um padrão para o vinho tinto é complexa e multivariada, mesmo com protocolos de produção e registros de procedência. Por isso, legislações pertinentes a esse assunto regulamentam os requisitos de composição para comercialização segura e isenta de fraudes.

2.4 Legislações Brasileiras

Ao longo dos séculos, o conceito de vinho se moldou conforme as viabilidades de produção e transporte. As preocupações com relação à qualidade e uniformidade produtiva representavam uma consequência da classe social para qual a produção era destinada. Foi somente no século XX que o vinho foi definido como a bebida feita com uva ou suco de uva, especificando a inclusão ou proibição de aditivos e variações no processo produtivo ^{3,12}.

No Brasil, as Leis ordinárias do Vinho (7678/1988) e de Bebidas (8919/1994), respectivamente, regulamentam o setor vitivinícola. A primeira das leis é mais específica sobre o tema de Vinhos no Brasil, enquanto a segunda, apesar de não tratar sobre vinhos especificamente, aborda derivados da uva e do vinho, se tornando por vezes uma auxiliar no tema ⁵⁶.

Regulamentada pelo Decreto 8.198 ¹, a Lei do Vinho passou por algumas reformulações no seu texto original. Cabe ressaltar, também, a Lei 10.970/2004, que foi

criada para relacionar a legislação Brasileira com o Regulamento Vitivinícola do MERCOSUL ^{57,58}.

A legislação brasileira também inclui, através do Decreto n.º 5863, as Resoluções anualmente aprovadas pela Organização Internacional da Uva e do Vinho (OIV) ⁵⁹. Mesmo que no Brasil seu caráter seja não obrigatório, essas Resoluções podem ser usadas como indicativo e seguidas quando as práticas enológicas não forem previstas pela legislação brasileira ⁵⁷.

Tabela 4. Regras de Indicação de Procedência ou Denominação de Origem

| | <i>AM</i> | <i>CA</i> | <i>MB</i> | <i>PB</i> | <i>VV</i> |
|--|---|---|--|--|---|
| <i>Origem da Uva</i> | 85% local | 100% local | 100% local | 85% local | 100% local |
| <i>Varietais* e Demais uvas permitidas</i> | Ancelota; Cabernet Franc; Cabernet Sauvignon; Marselan; Merlot; Pinot Noir; Refosco; Tannat. | Cabernet Franc. Cabernet Sauvignon; Merlot; Pinot Noir; Tannat; | Alicante Bouschet. Cabernet Franc* ; Cabernet Sauvignon* ; Egiodola; Merlot* ; Tannat; | Ancelota; Cabernet Franc* ; Cabernet Sauvignon* ; Merlot* ; Pinot Noir* . Pinotage; Sangiovese* ; Tannat* ; | Cabernet Franc. Cabernet Sauvignon; Merlot* ; Tannat; |
| <i>Proporções previstas?</i> | x | x | 40% Merlot. | x | ≥60% Merlot. |
| <i>Demais uvas?</i> | x | Negociável | x | x | x |
| <i>Chaptalização</i> | Negociável e ≤ 1% | ≤ 2% | ≤ 2% | Não especifica. | Negociável. |

Fonte: Adaptada dos regulamentos de uso da Indicação de Procedência ^{45,52-55}

Através desses protocolos, o vinho brasileiro passou a ser um produto de composição e metodologia produtiva definida, balizando a produção interna e a entrada de importações do setor vitivinícola ⁶⁰. Dentre as oito classes previstas por lei, encontram-se os vinhos de mesa e os vinhos finos. As demais classificações (frisante, gaseificado, leve, licoroso, moscato espumante e composto) dizem respeito a variações de produto, que não serão detalhadas pois não são o foco do presente estudo.

A diferença na classificação quanto ao vinho ser “de mesa” ou “fino” se deve à variedade de uva utilizada na produção. Enquanto os vinhos rotulados como “de mesa” podem ser produzidos com uvas americanas, viníferas ou a mistura destas, os vinhos rotulados como “vinho fino” só apresentam variedades de viníferas nobres, que, no caso de vinhos tintos, compreendem (dentre outras) as variedades Cabernet Franc, Cabernet Sauvignon, Merlot, Pinot Noir, Sémillon, Gamay Beaujolais, Malbec, Petite Syrah ⁶¹.

Os vinhos das classes de mesa ou fino também podem ser caracterizados com relação a sua coloração (branco, rosado ou tinto) e teor de açúcar (seco, *demi-sec* ou suave). Na Tabela 5 encontram-se informações extraídas da legislação e adaptadas para uma melhor demonstração da composição dos vinhos, sejam esses de mesa ou finos ⁶⁰.

A classificação do produto é descrita no rótulo em conjunto com a indicação da variedade da uva utilizada e safra da colheita da uva. Para que o produtor possa rotular o vinho como resultado de uma variedade específica, a composição dessa uva deve ser igual ou superior a 75%. Em relação à indicação de safra, 85% do vinho deve ser composto por uvas de uma mesma safra ⁶².

Pela Instrução Normativa de n.º 24, de 8 de setembro de 2005, foi aprovado o manual operacional de Bebidas e Vinagre, onde consta, entre outros, o padrão de métodos de análise de bebidas fermentadas, se tornando referência oficial para análises físico-químicas de vinhos no país ⁶³. O compêndio, que será discutido brevemente na próxima seção, abrange análises que aplicam metodologias tradicionais na detecção e quantificação de analitos prejudiciais à saúde e marcadores específicos de autenticidade.

2.5 Análises instrumentais e certificação de origem

Um vinho autêntico é aquele que se mantém da maneira como foi produzido e, mais precisamente, da maneira que se enquadra nas legislações vigentes de controle de qualidade. A autenticidade do vinho, evidenciada por análises químicas clássicas e instrumentais, é então certificada pela constatação da ausência de substâncias convencionalmente proibidas. Todavia, as variadas possibilidades de adulteração, além

da complexidade do vinho, exigem o desenvolvimento contínuo de técnicas analíticas cada vez mais sofisticadas ⁶⁴.

As análises químicas descritas na regulamentação brasileira, resumidas no esquema da Figura 8, são majoritariamente focadas na composição química dos mostos e vinhos. Mesmo que de caráter subjetivo, as avaliações sensoriais também são significativas e previstas por lei para atestar a qualidade desse produto ^{12,57}.

Tabela 5. Parâmetros da composição de vinhos previstos pela legislação brasileira

| | | | (Continua) | |
|--|---|------------------------------|------------|--------|
| PARÂMETROS LEGAIS | | | MÍNIMO | MÁXIMO |
| Graduação alcoólica, expressa em %, em v/v, a 20°C | | | ≥ 8,6 | ≤ 14 |
| Pressão gasosa, em atm, a 20°C | | | - | 1 |
| Teor de açúcar, em g/L | <i>Classificação quanto ao Teor de açúcar</i> | Seco | - | ≤ 4 |
| | | <i>Demi-sec</i> ou Meio-seco | > 4 | ≤ 25 |
| | | Vinho de mesa | > 25 | - |
| | | <i>Suave</i> ou <i>Doce</i> | > 25 | - |
| | | Vinho de mesa de americanas | > 25 | - |
| | | Vinho de mesa de viníferas | > 25 | ≤ 80 |
| | | Vinho fino | > 25 | ≤ 80 |
| Acidez total, em mEq/L (pH 8,2) | | | 40 | 130 |
| Acidez volátil, em mEq/L | | | - | 20 |
| Ácido Cítrico, em g/L | | | - | 1 |
| Sulfatos totais, expresso em sulfato de potássio, em g/L | Vinhos sem ou com tempo de envelhecimento menor que 2 anos | | - | 1,2 |
| | Vinhos com tempo de envelhecimento igual, ou maior que 2 anos | | - | 1,5 |

Tabela 5. Parâmetros da composição de vinhos previstos pela legislação brasileira

| PARÂMETROS LEGAIS | (Conclusão) | |
|---|--------------------|---------------|
| | MÍNIMO | MÁXIMO |
| Cloretos Totais, expresso em cloreto de sódio, em g/L | - | 0,2 |
| Cinzas, em g/L | 1,5 | - |
| Extrato seco reduzido, em g/L | 21 | - |
| Água exógena | | Ausência |
| Corante Artificial | | Ausência |
| Edulcorante | | Ausência |
| CONTAMINANTES | MÍNIMO | MÁXIMO |
| Ocratoxina A, em limite máximo tolerado (LMT), em µg/kg | - | 2 |
| Álcool metílico, em g/L | - | 400 |
| Arsênio, em mg/L | - | 0,2 |
| Chumbo, em mg/L | - | 0,15 |
| Cádmio, em mg/L | - | 0,01 |
| Estanho, em mg/kg | - | 150 |

Fonte: Adaptado de Consolidação das Normas de Bebidas (2020) ⁶⁰

Algumas empresas com capacidade de investimento, assim como as entidades de apoio às micro e pequenas empresas se propuseram a desenvolver um sistema interativo que promove a transparência do setor vitivinícola. Mesmo que sua aplicação esteja em estágio inicial, o que se propõe é informar desde a quantidade de uva produzida pela empresa, até as proporções de diferentes variedades utilizadas na composição do vinho ⁶⁵.

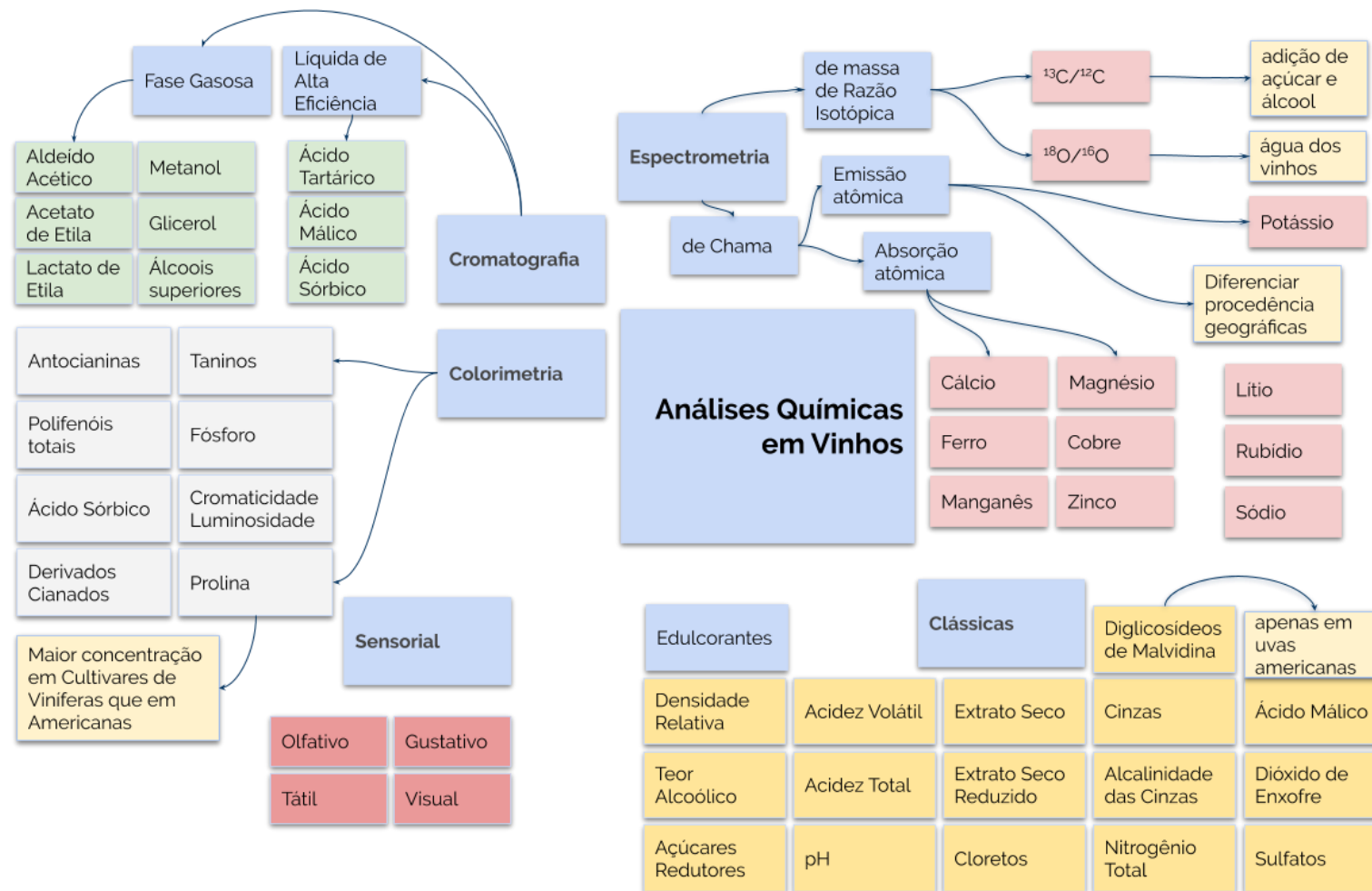


Figura 8. Análises químicas em vinhos para controle de qualidade brasileiro.

Fonte: Adaptado pelo Autor, da Instrução Normativa 24/2005 e Rizzon. ^{63,66}

O desenvolvimento de metodologias de autenticação pode ser feito por uma abordagem que considere na amostra a sua composição e atividade biomolecular, a sua composição isotópica, ou os seus constituintes químicos^{12,67}. Seja por meio de uma abordagem focada ou não em compostos alvo, denominadas respectivamente de *targeted* e *non-targeted*, a literatura das recentes décadas aponta um interesse na extração de informações através de ferramentas de estatística multivariada⁶⁸⁻⁷¹.

A abordagem biomolecular é construída utilizando informações da composição proteica, DNA e afins. Com essa informação, arranjos de sensores sintéticos são estudados para compor línguas e narizes eletrônicos em análises de vinhos^{72,73}. Em conjunto, narizes e línguas eletrônicas se equiparam a técnicas analíticas consolidadas ao classificar corretamente vinhos compostos de diferentes variedades de uvas, ou ainda quantificar proporções de misturas de vinhos, por exemplo, baseando-se na composição de taninos^{70,72}.

A autenticação focada na composição de certos isótopos estáveis é executada através de metodologias oficiais brasileiras e da OIV, que empregam a Espectrometria de Massa de Razão Isotópica (IRMS). Por essa abordagem são analisadas as razões de $^{18}\text{O}/^{16}\text{O}$ de água e $^{13}\text{C}/^{12}\text{C}$ de etanol, respectivamente expressas $\delta^{18}\text{O}$ e $\delta^{13}\text{C}$ ⁷⁴⁻⁷⁷.

Pela abundância relativa do isótopo ^{18}O sugere a presença ou ausência de água exógena, enquanto a razão $\delta^{13}\text{C}$ indica a adição de açúcar durante o processo de vinificação. No primeiro caso, é sabido que a abundância de ^{18}O é menor que ^{16}O , porém de concentração maior em águas botânicas do que em água de torneira⁷⁸. Na avaliação de proporção isotópica de carbono, a razão $\delta^{13}\text{C}$ aponta presença de açúcares de origens biossintéticas distintas, sendo menor na uva do que na cana-de-açúcar, por exemplo⁷⁹.

A literatura tem apresentado, contudo, que a variação dessas razões, $\delta^{18}\text{O}$ e $\delta^{13}\text{C}$, reflete condições climáticas regionais, podendo auxiliar no processo de autenticação de origem^{64,80,81}. No mesmo sentido, tanto a razão isotópica do estrôncio ($^{87}\text{Sr}/^{86}\text{Sr}$)⁸² quanto o perfil elementar aliado à razão isotópica chumbo-estrôncio ($^{207}\text{Pb}/^{206}\text{Pb}$, $^{208}\text{Pb}/^{206}\text{Pb}$, $^{204}\text{Pb}/^{206}\text{Pb}$, $^{87}\text{Sr}/^{86}\text{Sr}$)⁸³ se mostraram, também, alternativas para a investigação de procedência de vinhos.

Outra estratégia para o desenvolvimento de metodologias de autenticação é mediante a determinação da composição química da amostra. Uma das vias estratégicas é a análise direta de um componente químico na amostra. Há também, metodologias cujas respostas analíticas atestam indiretamente a autenticidade da amostra, seja por perfil

químico ou por impressão digital (*fingerprint*). As metodologias indiretas são aquelas que necessitam de uma base de dados ou um método de referência para comparação ^{64,68}.

Cromatografia, espectrometria e espectroscopia são técnicas presentes nas metodologias de autenticação de vinhos e que seguem as estratégias fundamentadas na composição química. Como dificuldades desses métodos, tem-se a indisponibilidade de padrões de referência, a complexidade do vinho e a elevada suscetibilidade de variação dos seus componentes, a necessidade de um preparo de amostra elaborado, além da necessidade de detectores específicos acoplados aos equipamentos, também de alto custo.

Técnicas cromatográficas podem ser empregadas com objetivo da distinção de origem de procedência do vinho, avaliando a composição de antocianinas e outros compostos fenólicos por *HPLC-UV/Vis*. Com ferramentas de estatística multivariada, vinhos do Nordeste e Sul Brasileiros, Chile e Argentina foram diferenciados conforme sua origem geográfica, principalmente devido às concentrações discriminantes de ácido gálico e proporções de 9 antocianinas ^{22,33}.

Coelho e colaboradores (2018 ⁸⁴) avaliaram vinhos e sucos, também do Nordeste Brasileiro, com relação à composição de ácidos orgânicos e açúcares por *HPLC-DAD-RID*. Mediante aplicação de método validado para caracterização dos vinhos e metodologia sugerida pela OIV, os autores concluíram que há adequabilidade regional para produção vitivinícola na localidade, equiparando resultados a outras partes da América do Sul.

Pisano, Silva e Olivieri (2014 ¹⁰) obtiveram importantes resultados através de uma abordagem cromatográfica de *fingerprint*, aliada à estatística multivariada. Sem etapas de preparo da amostra e por *HPLC-DAD-RID*, uma variedade específica de uva foi completamente diferenciada das demais analisadas, enquanto a discriminação de origem geográfica foi parcialmente atingida. Similarmente, porém através de detectores de fluorescência, UV-Vis e posterior referência complementar por *HPLC-MS*, Serrano-Lourido et al. (2012 ⁸⁵) conseguiram discriminar amostras de vinhos de 3 diferentes indicações geográficas espanholas com base na resposta instrumental dos diferentes compostos fenólicos presentes.

A micro-extração em fase sólida acoplada a espectrômetro de massas, com separação cromatográfica a gás ou não (*SPME-MS* e *SPME-GC/MS*, respectivamente) também são abordagens recorrentes na literatura. Pelo *fingerprint* de compostos voláteis via *SPME-MS*, Ziólkowska, Wasowicz e Jelen (2016 ⁸⁶) conseguiram diferenciar vinhos de acordo com o tipo de uva e a origem geográfica. Por meio de *SPME-GC-MS*, em

análise de marcadores de adulteração em vinhos, Sagandykova et al. (2017 ⁸⁷) desenvolveram uma metodologia alternativa para determinação de aditivos flavorizantes ou prejudiciais à saúde.

A espectrometria atômica é outra área consolidada da química analítica instrumental, e sua finalidade é quantificar elementos da composição mineral. Por exemplo, para o controle da concentração de Arsênio, Cádmio, Chumbo e Estanho, a legislação brasileira fixou um limite máximo permitido para cada um desses elementos^{60,88}. Técnicas como AAS, ICP OES e ICP-MS são aplicadas em análises de segurança alimentar e, uma vez que seus analitos refletem condições ambientais específicas de origem e procedência, estão presentes nos variados estudos sobre metodologias de autenticação ^{80,89}.

A terminologia empregada para designar os resultados das análises multielementares via espectrometria atômica é ora tratada como *profile* (perfil), ora *fingerprint* (impressão digital). Enquanto o termo *profile* pode ser entendido como uma forma gráfica, o termo *fingerprint* se mostra mais adequado, visto que os elementos discutidos nas análises não são alvo específico (metodologias *non-target*), e contribuem como variáveis indiretas de autenticação de origem ^{68,90}.

Através do *fingerprint* da composição multielementar, empregando *ICP-OES* e ferramentas de classificação da estatística multivariada, González et al. (2009) identificaram corretamente a procedência de amostras oriundas de diferentes sub-regiões da designação de origem de Valencia foi corretamente classificada ⁹¹. Semelhantemente, agora também demonstrando correspondência de composição atômica do solo e vinho de regiões produtoras argentinas, um estudo com espectrômetro de absorção atômica com chama (*FAAS*) classificou corretamente 92% e 100% as amostras de solo e vinho, respectivamente ⁹². Por fim, através da análise multielementar por *ICP-OES* e *ICP-MS*, Soares et al. (2018 ⁹³) obtiveram 99,9% de exatidão na classificação correta quanto à origem de amostras de vinhos de quatro países sul-americanos.

Com o objetivo de se sobressair à rápida adaptação e diversificação de fraudes em alimentos, o desenvolvimento de testes de adulteração *target* enfrenta crescente dificuldades ⁹⁰. Apesar da robustez e precisão, muitas das metodologias oficiais e alternativas não atendem às demandas da indústria vitivinícola. Tais metodologias necessitam de processos elaborados de preparo de amostra, são de elevado custo, trabalhosas, morosas, destrutivas e geradoras de resíduos tóxicos ¹³. Diante desse

conjunto de inconvenientes, a espectroscopia vibracional surge como uma possível solução, contrapondo abordagens citadas anteriormente.

Com resultados analíticos promissores, a espectroscopia vibracional se mostra vantajosa no controle de parâmetros químicos em cada etapa produtiva da vitivinicultura: do solo, passando pela saúde das videiras e maturação das uvas, até a avaliação do produto final. Nessa perspectiva, a classificação de variedades de uvas e a origem de procedência são estudadas via espectroscopia de fluorescência, ultravioleta, visível e infravermelho com transformada de Fourier nas regiões do espectro próximo e médio ¹³.

A fluorescência é uma técnica que, em poucos segundos, fornece espectros das bandas de excitação e emissão, frequentemente usados como *fingerprint* da amostra em estudo. Ácidos fenólicos, estilbenos, antocianinas, flavonóis e taninos são componentes fluorescentes, cuja proporção indica variedade e origem no vinho ⁹⁴⁻⁹⁶. Com os mesmos objetivos de classificação, também observa-se a sinergia dos dados analíticos de fluorescência com outras técnicas espectroscópicas, como UV-Vis ⁹⁷.

Metodologias UV-Vis têm se mostrado apropriadas para análise da composição fenólica, reflexo da habilidade de componentes botânicos com ação protetiva contra radiação ultravioleta e presença de coloração. Por apresentar respostas de correção considerável com análises via HPLC, a espectroscopia no UV-Vis pode ser uma técnica alternativa no monitoramento da composição fenólica ao longo da fermentação e no produto final ⁹⁸. A fusão de dados UV-vis e IR possibilitou a pesquisadores a classificação de vinhos segundo origem ⁹⁹, o auxílio em avaliação sensorial ¹⁰⁰, a calibração de componentes fenólicos ¹⁰¹ e um correto prognóstico da composição de antocianinas ¹⁰².

Ainda, no controle de processos químicos envolvidos na produção vitivinícola, a espectroscopia na região do infravermelho possui qualidades atrativas, como tempos de aquisição dos dados espectrais na ordem de segundos, equipamentos por vezes até portáteis, baixo custo relativo de operação, cujas análises não-destrutivas e simples possibilitam uma geração mínima de resíduos e o fácil entendimento de execução da rotina ⁹⁰. Estudos mostram sua aplicação na classificação de uvas segundo sua variedade ¹⁰³, identificação e quantificação de compostos voláteis no vinho ¹⁰⁴, além da diferenciação segundo região de origem ¹⁰⁵.

Abordagens via cromatografia, espectrometria ou ainda espectroscopia vibracional necessitam de tratamento computacional das respostas analíticas. Para estudo da autenticidade do vinho, as abordagens sugeridas se mostram convenientes seja pelo

grau de desenvolvimento de equipamentos com melhores resoluções instrumentais, ou para a separação de sinais redundantes ou interferentes ^{13,89,90,106}.

As *fingerprints* da espectroscopia são particularmente muito complexas para serem processadas visualmente. Os espectros brutos, caracterizados por sobreposição de bandas e sinais não muito intensos, refletem a natureza e heterogeneidade da amostra, ruído instrumental, espalhamento da radiação e condições ambientais ⁹⁰. Para a extração de informações úteis, que contribuam para análises qualitativas ou quantitativas, o desenvolvimento de metodologias com análise multivariada se mostra indispensável.

A legislação brasileira vigente regulamenta técnicas de análise química consolidadas e que seguem o tratamento de dados via calibração univariada. Devido a capacidade de geração de dados, a pesquisa em metodologias de calibração multivariada se apresenta como pertinente no sentido de extrair informação útil e valorizar as potencialidades, tanto da instrumentação analítica tradicional quanto de alternativas mais sustentáveis.

Nesse sentido, o estudo no controle de qualidade e origem de procedência de vinhos tem avançado na direção da aplicabilidade de metodologias e ferramentas da quimiometria. Como apresentado ao longo desse capítulo, sobressaindo os reflexos das atividades humanas e agrícolas, a quimiometria possibilitou conclusões importantes com relação a aspectos de qualidade do produto ⁸⁹.

2.6 Radiação eletromagnética e o Infravermelho

2.6.1 Radiação eletromagnética

O termo luz muitas vezes é empregado como alusão ao conhecimento verdadeiro. Como forma de energia, do qual a luz visível é apenas uma estreita faixa, radiações de um amplo espectro eletromagnético são base para métodos de análises de identificação e quantificação de espécies atômicas e moleculares.

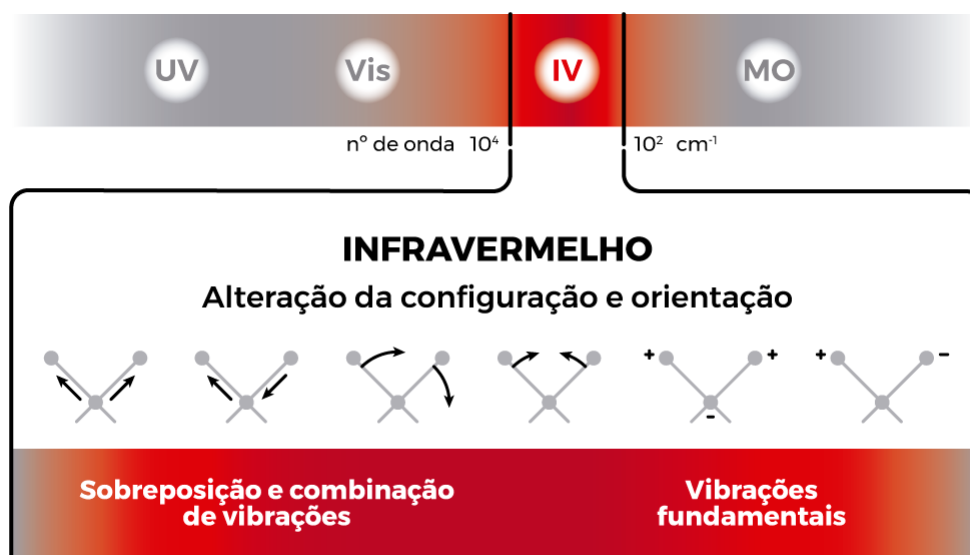
A radiação eletromagnética é uma forma de energia cujas propriedades são explicadas por modelos da física clássica e quântica. De modo complementar, enquanto a primeira considera a natureza macroscópica representada pelo modelo ondulatório de componentes elétrico e magnético, a segunda examina os efeitos e a probabilidade numérica da interação do fóton com a matéria ¹⁰⁷.

A transmissão da radiação eletromagnética através de um meio, que não o vácuo, ocorre a velocidades levemente inferiores à da luz (c). Ondas eletromagnéticas são descritas pela frequência (f), o número de oscilações por unidade de tempo, e pelo

comprimento de onda (λ), cujo seu recíproco é definido como o número de ondas por centímetro (ν). Os efeitos das interações entre radiação e matéria, reflexo da quantização de energia, são proporcionais ao número de onda e frequência, porém inversamente proporcionais ao comprimento de onda ¹⁰⁷.

2.6.2 Infravermelho

Enquanto frequências (f) diferentes de ondas sonoras produzem diferentes tons, diferentes frequências de ondas eletromagnéticas produzem diferentes tipos de luz. Por exemplo, se a luz for do espectro visível, ondas de diferentes frequências produzem diferentes cores. Abaixo da região de frequências do visível encontra-se a faixa espectral do infravermelho que, por sua vez, é convencionalmente subdividida em mais regiões, como representado na Figura 9.



Legenda: **UV** = Ultravioleta; **Vis** = Visível; **IV** = Infravermelho; **MO** = Microondas.

Fonte: adaptado de Encyclopaedia Britannica ¹⁰⁸, por Josué Pereira Dias

Figura 9. Infravermelho e resposta molecular vibracional

Dentro do conjunto de métodos espectroquímicos inclui-se a espectroscopia no infravermelho, que lida com a medição da radiação absorvida em fenômenos de níveis energéticos vibracionais e rotacionais. Suas respostas analíticas apresentam particularidades com a região de trabalho no espectro e a escolha de acessórios da instrumentação, tornando-a uma técnica simples, rápida e de caráter não destrutivo ^{107,109}.

As regiões do espectro infravermelho empregadas em abordagens de interesse analítico compreendem a faixa de números de onda entre 14300 a 200 cm^{-1} : de 12800 e 4000 cm^{-1} (Infravermelho Próximo) e 4000 a 200 cm^{-1} (Infravermelho Médio). Ambas

regiões são utilizadas para a elucidação de compostos orgânicos presentes em amostras, como o vinho ^{103,104,110}.

2.6.3 *Origem da Informação*

Na espectroscopia do infravermelho, a amostra é estimulada aplicando-se energia na forma de luz com números de onda entre 250 a 25000 cm^{-1} . As moléculas da amostra, em seu estado de energia fundamental, absorvem o estímulo da energia IV e entram no estado excitado. A radiação absorvida, proveniente de fonte externa e transformada em fenômenos de transição vibracional ou rotacional, é medida em função do comprimento de onda (λ) ^{109,110}.

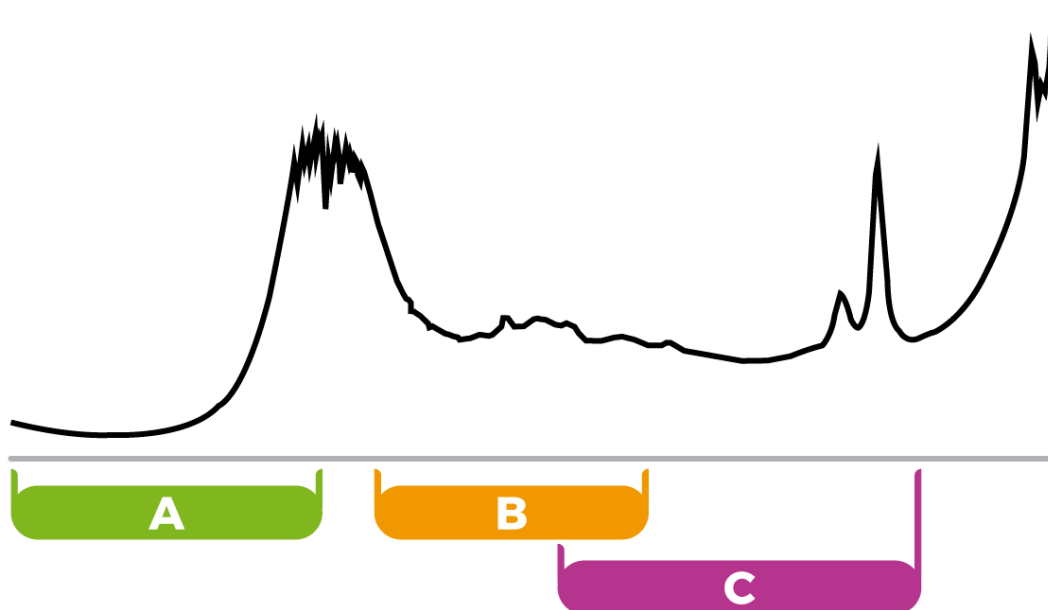
Quando a frequência do campo elétrico alternado – produzido pela fonte externa – coincide exatamente com a frequência necessária para que haja variação do momento dipolar molecular, há uma variação de amplitude da vibração molecular e, eventualmente, uma absorção de radiação. Da mesma forma, a rotação de moléculas assimétricas ao redor do seu centro de gravidade gera uma variação de dipolo que também absorve radiação. Enquanto a radiação absorvida na faixa com números de onda entre 10000 e 100 cm^{-1} é transformada em energia vibracional, a radiação de intensidade inferior é convertida a transições rotacionais¹⁰⁹.

Considerando, então, que as transições vibracionais e rotacionais refletem a composição de espécies atômicas e suas interações, com o espectro de infravermelho é possível fazer inferências das funções orgânicas características das moléculas presentes na amostra. Essa avaliação é feita principalmente nos espectros de infravermelho médio, por apresentar bandas intensas e picos com absorbâncias altas, além da faixa conhecida como impressão digital (1200 a 600 cm^{-1}) ¹¹¹.

Além disso, a intensidade dessa absorção é proporcional à abundância do componente sujeito à interação com a radiação. Todavia, a quantificação é dificultada pela complexidade amostral e ações de interferentes ^{112–114}.

2.6.4 *Espectroscopia no infravermelho aplicado a vinhos tintos*

As amostras de vinho tinto analisadas no presente trabalho apresentam espectro médio no infravermelho como representado na **Figura 10**. O perfil amostral observado é discutido na Tabela 6, e está em concordância com a literatura consultada ^{103,115,116}.



Fonte: O autor; arte de Josué Pereira Dias.

Figura 10. Espectro no infravermelho e regiões do *fingerprint*.

Tabela 6. Regiões significativas da absorção no infravermelho médio.

| REGIÃO | COMENTÁRIO |
|----------|--|
| A | Região de absorção forte, que representa o alongamento de moléculas que contém C=O; éster, aldeído e ácido carboxílico. Todavia, amostras líquidas apresentam essa banda devido a interferência da absorção da água. |
| B | Alongamento de C=C de anéis aromáticos e de C – O da estrutura anelar que forma o pirano, presente em flavonoides. |
| C | Além da influência de moléculas de outras famílias, como ésteres, a região concentra as absorções referentes a alongamentos C–O, presentes em álcoois terciários, secundários e, nesse caso principalmente, primários. |

Fonte: Adaptado de Sigma-Aldrich e Versari *et al.* ^{117,118}

2.6.5 Formas de medida no infravermelho médio

Espectrômetros de infravermelho com Transformada de Fourier são constituídos de um interferômetro de Michelson, dispositivo que possibilita a detecção e registro de todos os comprimentos de onda simultaneamente. A radiação produzida na fonte é dividida em dois feixes, cada qual direcionado ao espelho estacionário ou móvel. Ao refletir na direção da amostra, a variação da posição do espelho móvel faz com que as amostras sejam atingidas por feixes de frequências em fases variantes. Finalmente, no

detector é registrado o interferograma, cujos sinais são convertidos em espectros interpretáveis através das funções matemáticas da transformada de Fourier e apodização^{109,119}.

Um dos fatores relevantes para o desenvolvimento de metodologias analíticas via espectroscopia é a possibilidade de extrair informação de amostras com mínimo, senão nenhum, preparo prévio. Tal característica implica em promover diferentes modos de interação da radiação com a amostra. Os modos mais comuns e presentes no mercado são: transmissão, refletância difusa, refletância especular e refletância total atenuada^{120,121}.

O método de transmissão fornece a porcentagem de radiação que, ao atravessar a amostra transparente, atingiu o detector. Uma variação desse método se chama transletância, onde o feixe de radiação é redirecionado e refletido, de modo a atravessar duas vezes a amostra. Ambos modos seguem a Lei de Beer-Lambert, inclusive nas características limitantes¹⁰⁹.

Outro método é o de Reflexão difusa (DRIFTS – Diffuse Reflectance IR Fourier Transform Spectroscopy). Este é empregado na análise de amostras sólidas, pulverizadas e sem tratamento prévio. A absorvância interna e o espalhamento contribuem com uma variação de intensidade do sinal, implicando em um tratamento rigoroso dos sinais que substitui a abordagem matemática de Lei de Beer-Lambert por variações de Kubelka-Munk. Já no método por reflexão especular (modo de interatância) a radiação incidente atravessa duplamente a amostra, geralmente um filme fino, disposta no suporte¹²¹.

Finalmente, o método via Refletância Total Atenuada (ATR) tem se tornado o mais popular nas análises por espectroscopia no infravermelho. Aplicável em amostras sólidas ou líquidas, um ponto crítico na aquisição de espectros é a necessidade do contato efetivo das superfícies da amostra e do prisma empregado. Tal cuidado se deve ao fato de que a profundidade de penetração das ondas fornecidas pela fonte de infravermelho depende dos índices de refração dos dois meios, o mais e o menos denso, além do comprimento de onda envolvido. O detector registra a atenuação das ondas emitidas pela fonte como resposta às absorções das ondas evanescentes, ondas emitidas através do prisma que, ao penetrarem a amostra, têm a frequência correspondente para gerar estímulo vibracional nas moléculas do meio¹²².

Há diferentes opções de cristais que podem ser usados nos métodos via ATR. Essa escolha depende da faixa espectral de interesse, o quão profundo a onda evanescente penetra a amostra e, por fim, a robustez mecânica do material do cristal às ações do

ambiente. Obviamente, a escolha também é ponderada devido ao custo de cada tipo de material. A Tabela 7 apresenta os diferentes prismas presentes no mercado ^{120,123}.

Tabela 7. Opções de prismas disponíveis no mercado para FTIR – ATR

| <i>material</i> | <i>faixa espectral(cm⁻¹)</i> | <i>Aplicações</i> |
|-----------------|---|--|
| <i>Germânio</i> | 5500 – 675 | Bom para maioria das amostras absorvedoras |
| <i>ZnSe</i> | 15000 – 650 | Uso geral |
| <i>Diamante</i> | 25000 – 100 | Bom para maioria das amostras. Extremamente duras ou cáusticas |

FONTE: adaptado de *ThermoFisher Scientific* e *Shimadzu Corporation* ^{120,123}

O desenvolvimento de novos equipamentos de tamanho reduzido viabilizou a aplicação analítica em bancada ou campo, conferindo popularidade aos espectrômetros do tipo FTIR. Em adição, o avanço tecnológico de computadores e superiores velocidades de processamento de dados promoveram alterações necessárias para tornar mais acessível e atrativo o custo desse instrumento. Ainda, os espectrômetros FTIR são instrumentos de alta resolução, precisão e exatidão, com maior sensibilidade na conversão das frequências de infravermelho ¹⁰⁹.

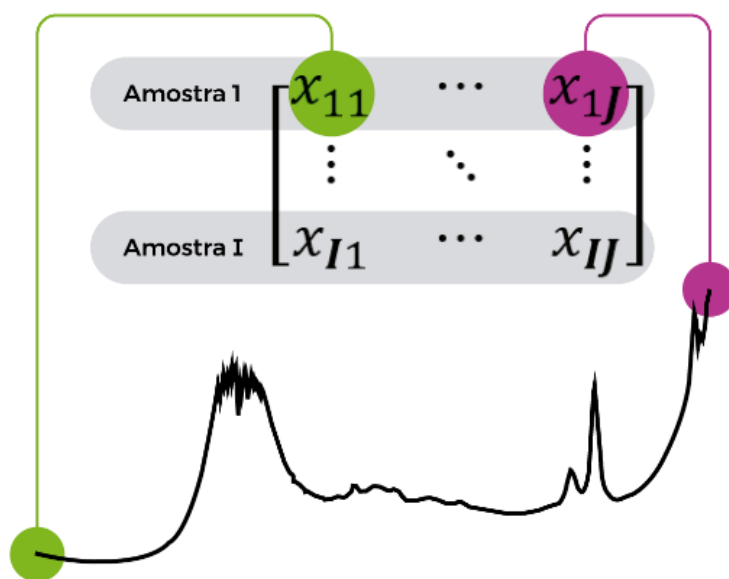
2.7 Quimiometria

A crescente e expressiva quantidade de informação gerada pela instrumentação analítica nas últimas décadas acompanhou e encontrou apoio no desenvolvimento de microprocessadores e computadores. Assim, da década de 1970 em diante, um tratamento estatístico, já obrigatório em metodologia científica, tem se tornado mais complexo nos propósitos analíticos ¹²⁴.

Variadas definições convergem à ideia de que Quimiometria é a região de intercessão da matemática e estatística multivariada, e um braço da química analítica. Seu escopo de trabalho inclui o planejamento ou otimização de procedimentos experimentais e conhecimento sobre sistemas químicos. Além desses, a quimiometria se objetiva à análise de dados multivariados para extração de informação relevante por métodos de reconhecimento de padrões e calibração, com ou sem aplicação de abordagens de seleção de variáveis ¹¹⁹.

Para a modelagem e aplicação das técnicas quimiométricas, os dados instrumentais são organizados em um arranjo matricial, \mathbf{X} ($I \times J$), como ilustrado na Figura

11. Cada uma das linhas i corresponde ao conjunto de informações j de uma amostra. Ou seja, em dados de espectrometria molecular, *e.g.*, cada elemento das J colunas corresponde ao registro da intensidade de absorção em dado número de onda j ¹¹⁹.



Fonte: o autor. Arte de Josué Pereira Dias

Figura 11. Forma de organização matricial dos dados FTIR

Uma vez que os sinais instrumentais têm contribuições determinísticas e estocásticas, é razoável inspecionar o arranjo matricial formado. Mesmo que transformações e pré-processamentos sejam aplicados para a exclusão de variáveis indesejáveis, a falta de parcimônia pode provocar distorções nos modelos quimiométricos¹¹⁹.

Seja por um problema instrumental ou de amostragem, as transformações são usadas para corrigir defeitos ao longo da resposta analítica através de alisamento e correção de linha de base, por exemplo. Enquanto no alisamento há destaque do sinal analítico e amenização do ruído instrumental, nas funções derivadas o deslocamento e a inclinação da linha de base são corrigidos.^{125,126}

Pré-processamentos são os algoritmos empregados em correções no âmbito das variáveis, ou seja, das colunas da matriz de dados¹¹⁹. Sem prejuízo da variância interna dos dados, a centragem de dados na média é o método mais frequente na análise de dados espectroscópicos. Seu atrativo reside na possibilidade de redução de influências não relacionadas ao analito^{115,127}.

Posterior ao tratamento inicial dos dados, outra intervenção quimiométrica é o fracionamento do conjunto amostral. Essa etapa é necessária para a construção de

modelos representativos de calibração e predição (ou treinamento e teste, em casos de classificação), onde a reprodução da variabilidade nesses processos precisa ser considerada. Nesse contexto, uma ágil partição pode ser executada por meio de algoritmos, como Kennard-Stone (KS) e SPXY.

O algoritmo KS (1969) utiliza de distância Euclidiana $d_x(p,q)$ para selecionar as amostras mais distantes entre si. Seguido dessa escolha, uma terceira amostra é rastreada, também distante deste primeiro subconjunto. O fim do ciclo é dado quando o número de amostras selecionadas atinge o número escolhido pelo analista. Em dados espectrais, por exemplo, a informação de distância espacial é extraída da comparação entre intensidades dos números de onda (j) em x_p e x_q ¹²⁸.

$$d_x(p, q) = \sqrt{\sum_{j=1}^J [x_p(j) - x_q(j)]^2}; p, q \in [1, N] \quad (1)$$

De modo similar, o algoritmo de partição baseado em distâncias SPXY considera a informação conjunta das variáveis independente e dependente ($\mathbf{x}-\mathbf{y}$), como expresso na Equação 2. Após a normalização de cada termo, a distância $\mathbf{x}\mathbf{y}$ atribui importâncias equivalentes na distribuição das amostras em ambos espaços considerados¹²⁸. O SPXY é geralmente adotado em métodos de regressão¹²⁹.

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p, q \in [1, N]} d_x(p, q)} + \frac{\sqrt{(y_p - y_q)^2}}{\max_{p, q \in [1, N]} d_y(p, q)}; p, q \in [1, N] \quad (2)$$

2.7.1 Reconhecimento de Padrões

Os métodos quimiométricos usados na diferenciação entre amostras são fundamentados na identificação de tendências e formação de agrupamentos. Semelhanças e diferenças com relação à procedência e composição do vinho, *e.g.*, refletem propriedades físicas e químicas do conjunto de amostras em estudo. Mais ainda, é interessante que tais métodos possam converter a massiva quantidade de dados para a visualização em espaço bi- ou tridimensional^{119,124}.

Algumas funções matemáticas se valem apenas das características químicas instrumentais registradas; outros consideram informações relacionadas à identidade da amostra. Respectivamente, o reconhecimento de padrões é dividido entre os métodos “Não-supervisionados” e “Supervisionados”, ambos com ampla aplicação na análise de bebidas e vinhos¹³⁰⁻¹³².

Métodos não-supervisionados

Também chamadas de análise exploratória, os métodos não-supervisionados visam facilitar a inspeção na formação de agrupamentos caso informações prévias de classificação não estejam disponíveis, ou ainda se há amostras anômalas no conjunto de dados analíticos. Nesse sentido, um dos métodos mais utilizados é a Análise de Componentes Principais, do inglês *PCA* ¹²⁴.

PCA

A análise de componentes principais, ou do inglês *Principal Component Analysis*, é um exemplo clássico de ferramentas matemáticas cuja aplicação, especialmente em química, foi postergada em décadas após sua introdução e desenvolvimento formal. Tal mudança de realidade se deve à integração de computadores nas rotinas de laboratório ¹¹⁹. Com agilidade típica de uma máquina bem programada, foi facilitada a compressão de dados multivariados, como expresso na Equação 3, em

$$\begin{bmatrix} \mathbf{X} \end{bmatrix} = \hat{\mathbf{X}} + \mathbf{E} = \sum_{a=1}^A \begin{bmatrix} \mathbf{t}_a \end{bmatrix} \begin{bmatrix} \mathbf{I}_a^T \end{bmatrix} + \begin{bmatrix} \mathbf{E} \end{bmatrix} \quad (3)$$

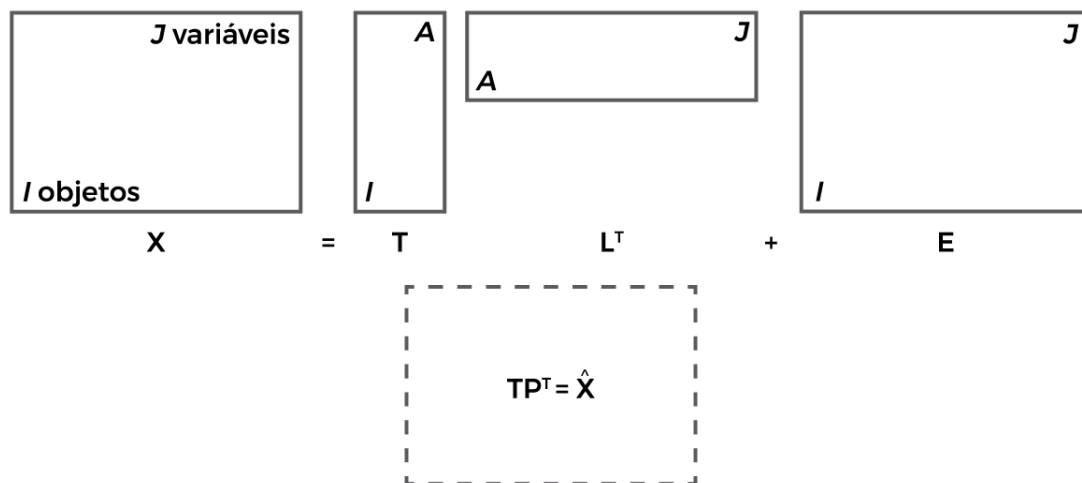
dimensões de inspeção visual de padrões.

A compressão de dados espectrais reduz o espaço multidimensional pela combinação de variáveis correlacionadas. As componentes principais são o conjunto novo de coordenadas formado, de modo combinar variáveis correlacionadas nas dimensões originais. Consecutivamente, cada novo eixo calculado será ortogonal a seu antecessor (bem como aos demais) e, progressivamente, compreenderá toda variação que explique o sistema ¹³³.

Esquemáticamente, a Figura 12 demonstra a decomposição dos dados de \mathbf{X} em vetores das componentes principais (\mathbf{t} , \mathbf{l}) e a matriz de resíduos (\mathbf{E}). Os pesos (\mathbf{l}) são vetores de coeficiente da regressão, ajustados para minimizar a soma dos quadrados das distâncias ponto-plano. Ao passo que os escores (\mathbf{t}) são a combinação linear da contribuição i das variáveis em J , essas ponderadas por (\mathbf{p}).

Atingir dimensões reduzidas significa encontrar os vetores da matriz dados \mathbf{X} que sejam linearmente independentes (autovetores). O posto \mathbf{A} (enquanto conjunto desses vetores) corresponde ao número de PCs suficiente para explicar o sistema. A interpretação do modelo depende da quantidade \mathbf{A} dessas componentes, que explicitam o sentido químico, mas que não causem um sobreajuste. Assim, a variância atribuída aos ruídos e informações sistemáticas irrelevantes é designada à matriz \mathbf{E} ($I \times J$).

Por meio da observação conjunta dos gráficos de escores e pesos (**T** e **L**), o analista pode inferir sobre a composição e a proporção relativa, respectivamente, de constituintes químicos da amostra. Conseqüentemente, agrupamentos formados podem ser relacionados a variáveis mais discriminantes.



Fonte: Adaptado de Bro; Smilde (2014).¹³³

Figura 12. Decomposição da matriz de dados.

Métodos supervisionados

Na construção desses modelos, a informação de classe padrão é preestabelecida nas amostrais integrantes do grupo de treinamento. No caso do estudo envolvendo diferentes regiões de origem de vinhos, por exemplo, o padrão de concentração mineral serve de subsídio do modelo a reconhecer as amostras teste⁹².

Os métodos supervisionados são frequentemente classificados conforme considerações estatísticas assumidas e forma de modelagem das relações entre amostras. Em contraste com os não-paramétricos, que não são baseados em informações sobre a distribuição populacional, chamamos de paramétricos os métodos que pressupõem (e empregam) uma distribuição de probabilidade amostral particular, como a normal.

Também, os modelos podem ser classificados como discriminantes ou modelativos. Os do primeiro tipo trabalham com diferenças entre amostras de diferentes classes, operando através da divisão do hiperespaço de variáveis em regiões, tantas quanto forem as classes pré-determinadas. Assim, as amostras pertencem a uma, e somente uma, classe. Contrapondo esses, as técnicas modelativas constroem limites entre cada classe e o resto do universo, de modo que as amostras em estudo podem ser avaliadas como

pertencentes ou não a cada classe, ou ainda, pertencente a duas classes simultaneamente, ou a nenhuma delas ¹³⁴.

Por fim, podem ser técnicas determinísticas ou probabilísticas. A primeira classifica uma amostra, porém o grau de confiança dessa decisão não é mensurado. No entanto, o método probabilístico provê uma estimativa da confiança da decisão de classificação ^{119,135}.

PLS-DA

O método de análise discriminante por mínimos quadrados parciais (do inglês, *Partial Least Squares Discriminant Analysis*, PLS-DA) é a variação do modelo desenvolvido inicialmente para calibração multivariada, PLS. Seu algoritmo adapta as variáveis originais em combinações lineares (variáveis latentes, VL) com máximo de covariância entre as variáveis dependentes Y (*dummy*). Ou seja, essa relação reflete as fontes de variabilidades amostrais, cujas coordenadas no novo hiperespaço discriminam as classes umas das outras ^{119,136}.

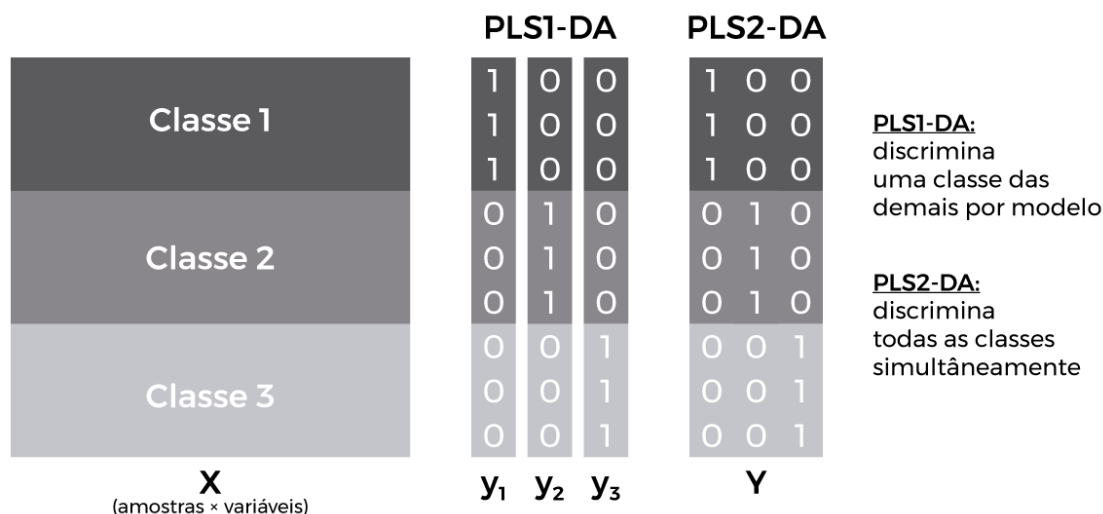
As variações de abordagem via PLS-DA raramente são explicitadas na literatura ¹³⁷. Todavia, dependendo do número de classes dependentes envolvidas simultaneamente, o método de regressão pode ser PLS1 ou PLS2. O algarismo após o acrônimo PLS (1 ou 2) sugere a dimensionalidade do arranjo binário (0 ou 1) das propriedades a serem modeladas.

Como representado na Figura 13, no PLS1 cada coluna y será modelada individualmente, resultando em um conjunto de escores e variáveis latentes específico por variável em inspeção. No método PLS2, todas as colunas Y são modeladas em conjunto ^{119,136}.

A dicotomia referente ao pertencimento de uma amostra a dada classe é estimada por base nas variáveis em Y (ou em cada y , visto PLS1-DA). Como resposta do modelo, o algoritmo retorna valores não necessariamente iguais a, porém entre 0 e 1. Nessa estimativa, basicamente, as amostras com valores próximos a 1 pertencem a dada classe de estudo.

Todavia, a decisão para alocar uma amostra em uma classe particular precisa ser delimitada por um *threshold* (ou limite). Fundamentado na suposição de distribuição Normal, esse Limite de Decisão é comumente estimado pela probabilidade condicional bayesiana. Assim, representado na Figura 14, a região será a interseção entre as

probabilidades de encontrar a posteriori, por exemplo, amostras oriundas da Serra ($P(A/x)$) ou da Campanha Gaúcha ($P(B/x)$)^{134,138,139}.



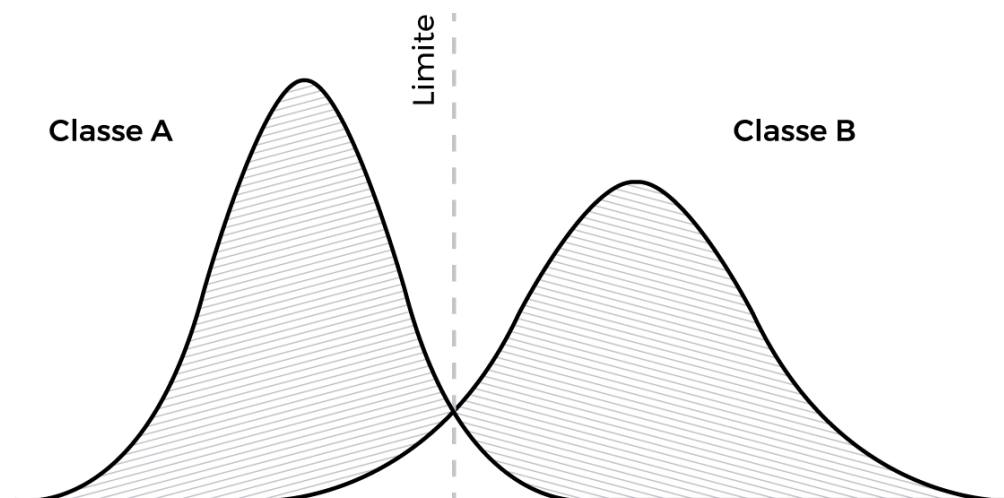
Fonte: Adaptado de Ferreira, 2015, p.461¹¹⁹.

Figura 13. Modos de execução do algoritmo em PLS-DA

A performance de classificação do modelo reflete a habilidade preditiva do modelo sobre si mesmo, a partir da seleção ótima de variáveis latentes com base na validação cruzada. Genericamente, a partir de frações do conjunto amostral, esse processo cíclico consiste em calibrar e estimar os erros de previsão em função do número de VL aplicadas. As abordagens mais recorrentes são: venezianas, blocos contínuos e *leave-one-out* (LOO).

Se tratando das duas primeiras, a disposição do conjunto de dados precisa ser considerada, de modo a não fomentar um modelo sem representatividade das classes envolvidas. No LOO, por sua vez, o ciclo de isolamento, calibração, predição e reintegração chega ao fim quando todas as frações do conjunto amostral tenham sido submetidas a esse processo. Em estudos de classificação envolvendo espectroscopia vibracional, LOO é o mais aplicado, principalmente pela robustez em conjuntos amostrais limitados^{136,140-142}.

$$P(A|x) = \frac{P(x|A) \times P(A)}{P(x|A) \times P(A) + P(x|B) \times P(B)} \quad (4)$$



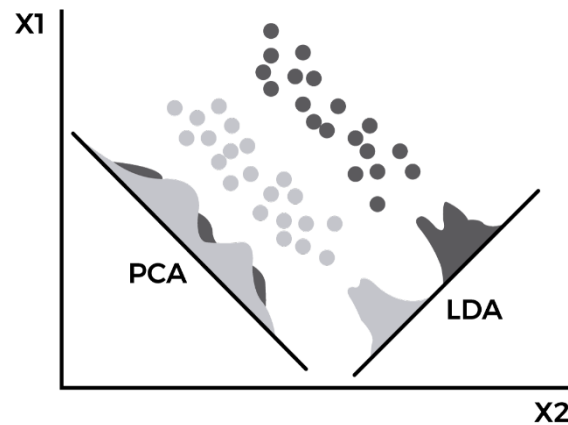
Fonte: Adaptado de MARTINEZ; MARTINEZ,2008, p.365 ¹³⁸

Figura 14. Probabilidades de classificação e Limite de decisão.

LDA

A mais antiga das técnicas publicadas sobre classificação, e ainda uma das mais utilizadas, a análise discriminante linear (LDA) foi introduzida por Fisher em 1936. Sua popularidade se sustenta devido a sua simplicidade. Sua robustez apresenta severas ressalvas na literatura, alterações no seu algoritmo, portanto, visam melhores desempenhos ^{134,143}.

Esse método é baseado na determinação de combinações lineares das variáveis mensuradas, com suposição de não correlação entre as variáveis. A redução de dimensionalidade resultante na PCA também ocorre na LDA. Porém, o algoritmo do segundo método seleciona a direção que resulte na separação maximizada entre classes, mas minimizada entre indivíduos da mesma classe ^{144,145}.



Fonte: Adaptado de Malakar ¹⁴⁵, arte de Josué Pereira Dias.

Figura 15. Comparação entre respostas dos métodos PCA e LDA

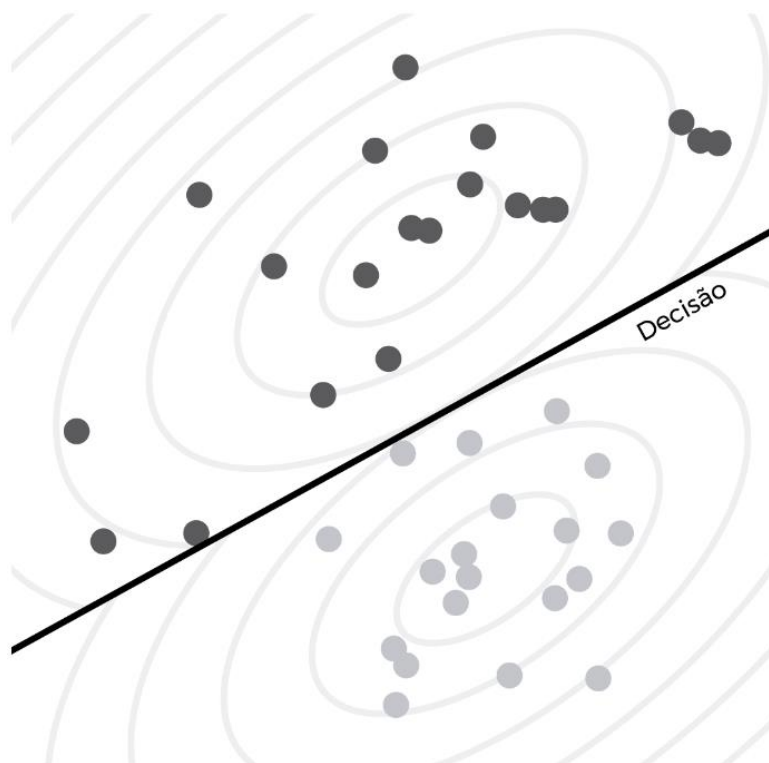
Em um sistema de variáveis bidimensional, como na Figura 15, é possível perceber o processo matemático simplificado, em que a redução dimensional é representada pela equação 5. Inicialmente, para atingir o objetivo expressado anteriormente, o centro da LDA está em encontrar \mathbf{Z} , combinações lineares dos dados \mathbf{X} , cujos coeficientes β são ponderações das características mensuradas.

$$\mathbf{Z} = \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2 \quad (5)$$

Esses coeficientes são função da matriz de covariância conjunta \mathbf{C} e do centroide de cada classe envolvida \bar{x}_1 e \bar{x}_2 . Ainda, expressa matematicamente pela Equação 6, essa relação integra as posições elípticas representados pela Figura 16, conhecidas como distância de Mahalanobis ^{143,146}.

$$D_{1-2} = \sqrt{(\bar{x}_1 - \bar{x}_2)^T \mathbf{C}^{-1} (\bar{x}_1 - \bar{x}_2)} \quad (6)$$

Com a definição de marcadores específicos para o estudo, a LDA tem se mostrado eficaz na discriminação de vinhos de diferentes origens e variedades ¹⁴⁷. Para atingir uma melhor performance da técnica é necessário atentar às suas limitações estatísticas. A quantidade de variáveis precisa ser menor que a quantidade de amostras envolvidas e as variáveis em análise não podem se correlacionar ^{111,134}.



Fonte: Adaptado de Brereton e Lloyd ¹⁴⁶.

Figura 16. Ilustração da Análise Discriminante Linear (LDA)

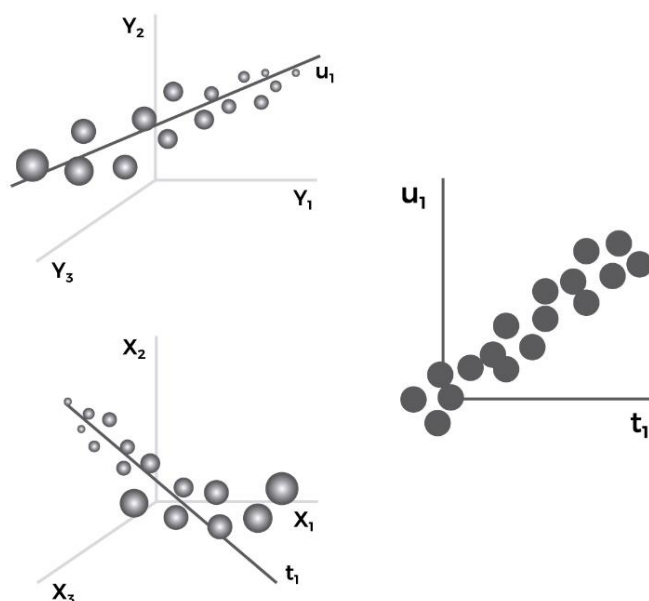
Muito frequentemente, todavia, esse não é o caso ao se avaliar a procedência de vinhos por espectro de infravermelho. Alternativas são propostas na literatura, como a utilização dos escores da PCA ¹⁴⁸ ou a escolha do número de variáveis latentes envolvidas ¹⁴⁹. Em ambos os casos, a redução de características redundantes e não discriminatórias sugere campo de aplicação para algoritmos de seleção de variáveis.

2.7.2 Regressão

PLS

Originalmente desenvolvido para aplicações em Economia, o método chamado Regressão por Mínimos Quadrados Parciais – PLS foi formulado em meados de 1960 e posteriormente adaptado à modelagem de dados químicos. O PLS é fortemente disseminado em quimiometria por sua vantagem de primeira ordem, ou seja, a capacidade de modelar e prever amostras mesmo na presença de interferentes ¹¹⁹.

Por esse método linear, a regressão considera a estrutura latente em ambos conjuntos de dados: nas variáveis explanatórias (**X**) e nas dependentes (**Y**). Para prever a propriedade de interesse, o alvo da redução de dimensionalidade dos dados em **X** tem como critério a preservação da explicação da variância em **X** com a previsão da propriedade de interesse, **y**. A Figura 17 resume a ideia do processo.



Fonte: O autor. Arte de Josué Pereira Dias

Figura 17. Direção da máxima da variância em y e covariância entre x e y .

Genericamente, PLS pode ser comparado à análise de componentes principais, porém uma versão supervisionada, cujos escores em U e T possuem uma relação interna, em outras palavras, máxima covariância um com o outro^{146,150}. Aqui, como já demonstrado factível, as variáveis latentes são combinações lineares dos dados espectroscópicos em X , relacionados com as concentrações provenientes de vinhos produzidos a partir de outros cultivares, em Y ^{13,151}.

O processo é caracterizado pela possibilidade de extração de informação a partir de mais de um componente. Como em PCA, do resíduo não modelado pela primeira componente PLS, uma segunda componente pode ser extraída. Cada repetição desse processo resultará em um componente de regressão, que será utilizado na etapa de predição das amostras desconhecidas.

A decomposição da matriz X em escores e pesos pode ser feita por meio de diversos algoritmos, como NIPALS, utilizado nos estudos de regressão e reconhecimento de padrões do presente trabalho. Dados como os de infravermelho, com um número elevado de variáveis em relação a um pequeno posto químico, o algoritmo NIPALS é o mais recomendado, uma vez que esse calcula as variáveis latentes (autovetores) um a um^{119,152}.

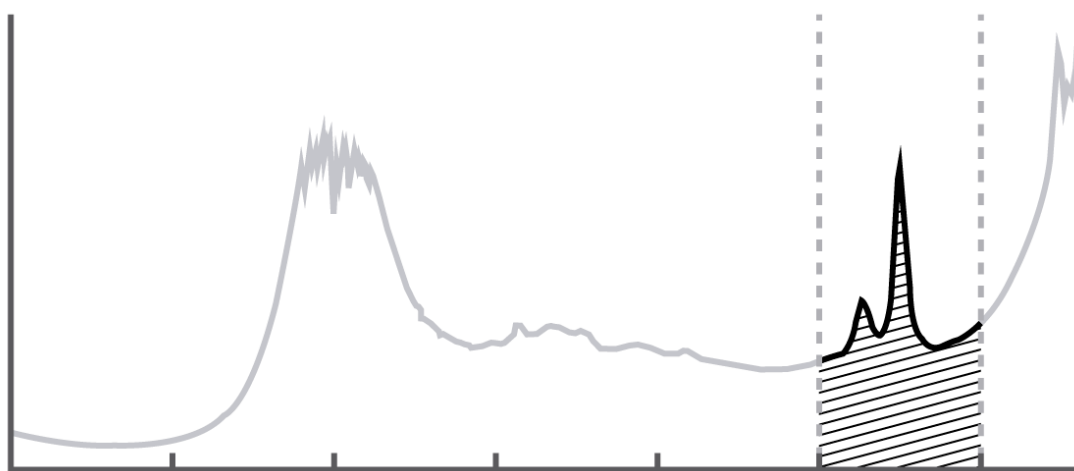
Frente a outras abordagens da quimiometria, a capacidade de agregar variáveis correlacionadas é um atrativo do modelo PLS. Consequentemente, o conjunto de amostras pode ser melhor estudado por meio de características que sobressaiam as

demais, apesar da presença de interferências. Ainda assim, é possível facilitar a interpretação do modelo, bem como a capacidade preditiva do algoritmo por meio da seleção de variáveis ¹¹⁹.

2.7.3 Seleção de variáveis

Grande parte da similaridade constatada entre amostras analisadas via espectroscopia MIR reflete sua composição química frequentemente constante. Na tentativa de sobressair tal situação, como ilustrado pela Figura 18, a seleção de variáveis tem se demonstrado uma etapa crucial no processamento de dados, principalmente por valorizar as influências discriminatórias, apesar de interferentes, comuns em misturas complexas ^{119,153}.

Assim como as diferentes abordagens em calibração e regressão, a seleção de variáveis pode ser estudada e executada por uma variedade de métodos. Os métodos aqui abordados serão os de algoritmos genéticos e de sucessivas projeções (GA e SPA, respectivamente).



Fonte: O autor. Arte de Josué Pereira Dias

Figura 18. Seleção de variável no espectro infravermelho.

Algoritmo Genético

Uma técnica de otimização bem popular empregada para solucionar problemas de otimização combinatória. As etapas que compõem cada ciclo do algoritmo podem ser visualizadas como uma generalização do processo biológico evolutivo, de onde derivou o nome “genético”. Aplicando no contexto do presente trabalho, o objetivo de cada ciclo é modelar a composição ótima de intensidades relativas aos números de onda, para classificar corretamente as amostras do grupo teste.

Dependente de sua forma estocástica e do número de ciclos selecionado por quem analisa, o algoritmo seleciona quais comprimentos de onda podem discriminar amostras. Em cada ciclo, o cruzamento das informações selecionadas gera amostras híbridas. A essas novas amostras é incrementado elementos de mutação, que auxiliarão a exploração do universo de respostas ¹⁵⁴.

O algoritmo tende à formação de modelos superajustados ¹¹⁹, ou com resultados pouco reprodutíveis ¹⁵⁵. Todavia, estudos têm aplicado com sucesso algoritmos genéticos na validação de metodologias que empregam FTIR-ATR na análise de vinhos e bebidas ^{156,157}.

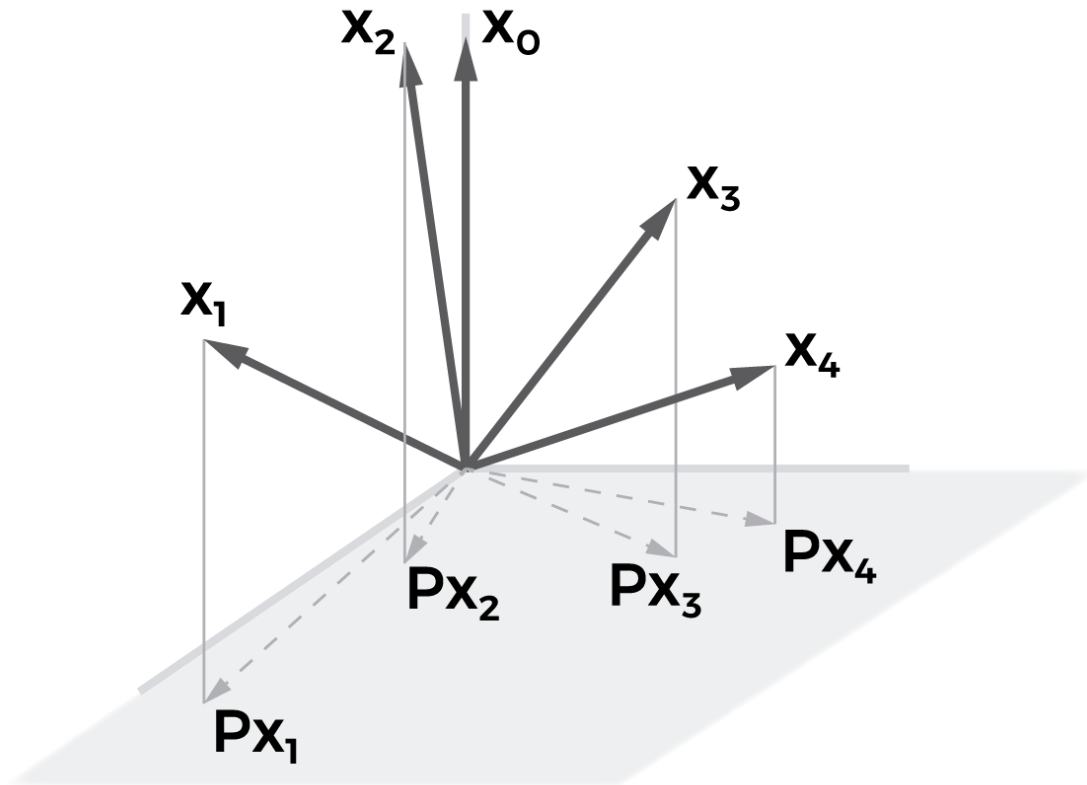
SPA

Ainda que selecionar variáveis a partir de um espectro de infravermelho seja uma etapa importante para a maximização de exatidão, sua seleção criteriosa é desafiadora. Algoritmos genéticos eventualmente apresentam soluções otimizadas, porém, justamente por sua natureza estocástica, a busca exaustiva exige emprego considerável de tempo computacional ^{155,157}.

A redução considerável de tempo de processamento é uma característica do algoritmo de projeções sucessivas (SPA), configurando-o como alternativa para seleção de variáveis. Com isso, seleciona-se números de onda cujas intensidades informadas sejam minimamente redundantes, solucionando assim problemas de colinearidade entre variáveis ¹⁵⁵.

Sem alterações na relação entre as variáveis espectrais e seus respectivos vetores com intensidades, o SPA é organizado para minimizar a colinearidade entre variáveis balizadas pelo erro de predição. As variáveis selecionadas respeitam o compromisso de conter informações úteis ao excluir variáveis ruidosas, com o cuidado de manter um conjunto representativo para aplicar nas predições.

Partindo de um vetor correspondente a um número de onda \mathbf{x}_0 , progressivamente e a cada ciclo, o algoritmo procura, seleciona e incorpora a um novo conjunto de vetores, as intensidades amostrais do número de onda com maior projeção no plano ortogonal do vetor \mathbf{x}_0 , como ilustra a Figura 19. Adaptações no algoritmo foram bem-sucedidas na análise de vinhos e bebidas via espectroscopia no infravermelho e por imagens digitais ^{129,132}.



FONTE: Adaptado de Araújo *et al.* ¹⁵⁵

Figura 19. Seleção de vetores projeção para seleção de variáveis.

2.7.4 Avaliação de desempenho de modelos

Métricas de desempenho analítico da classificação

O desempenho das técnicas de reconhecimento de padrão é avaliado pela alocação das amostras nas classes envolvidas, essas modeladas com o sinal analítico proveniente do infravermelho (*e.g.*) e a definição do limite de decisão. Assim, é avaliado o número de falsos positivos, falsos negativos e suas respectivas taxas de ocorrência.

O número de falsos positivos (FP) é como se nomeia a porção de amostras que, classificadas como pertencentes, de fato não pertencem a dada classe. A taxa de falsos positivos (FPR), por conseguinte, é a razão entre o número de falsos positivos e sua soma com o total de amostras pertencentes a outra classe (TN), como representado na Eq. 7.

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

Analogamente, o número de falsos negativos (FN) se refere à quantidade de amostras que pertencem a uma classe, mas ali não foram alocadas. A taxa de falsos negativos (FNR), representada na Eq. 8, é definida como a razão entre falsos negativos e sua soma com o total de amostras conhecidas dessa classe (TP).

$$FNR = \frac{FN}{FN + TP} \quad (8)$$

As informações citadas acima, juntamente com as quantidades corretamente classificadas (verdadeiros positivos e negativos, VP e VN), são organizadas na forma da matriz de confusão, comumente incluída como resposta dos algoritmos de classificação, de modo a facilitar a inspeção. Não somente, as quantidades e taxas consequentes da classificação correta ou não, suplementam os parâmetros de validação Exatidão (Eq.9), Sensibilidade (Eq.10), Taxa de Falso Alarme (Eq.11) e Seletividade (Eq.12) ^{119,149}.

Outra figura de mérito estatística aplicada nas diversas áreas de análise de dados é o Coeficiente de Correlação de Matthew (CCM). Como as anteriormente citadas, esse coeficiente relaciona as classificações por meio dos agrupamentos VP, VN, FP e FN. Diferentemente da Exatidão, como pode-se perceber pela versão de classificação binária na Equação 13, no CCM considera-se acertos e erros de ambas classes, formando uma possibilidade de análise com menor tendência a eficiências de classificação superestimadas ^{158,159}

$$EXAT = \frac{VP + VN}{VP + FN + FP + VN} \times 100 \quad (9)$$

$$SEN = \frac{VP}{VP + FN} \times 100 \quad (10)$$

$$TFA = FPR \times 100 \quad (11)$$

$$SEL = \frac{VN}{FP + VN} \times 100 \quad (12)$$

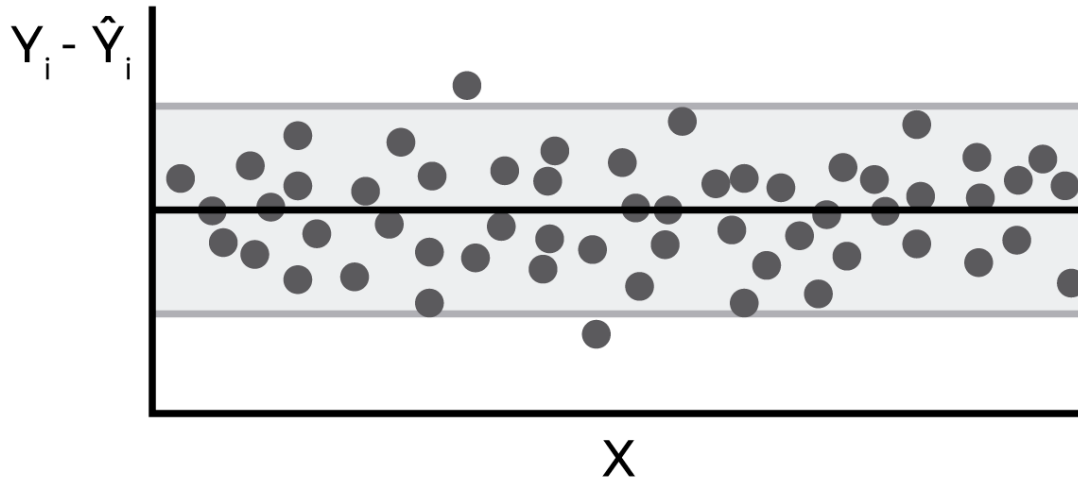
$$CCM = \frac{VP \cdot VN - FP \cdot FN}{[(VP + FN) \cdot (VP + FP) \cdot (VN + FP) \cdot (VN + FN)]^{1/2}} \quad (13)$$

Calibração

O desempenho dos modelos de regressão pode ser avaliado pela magnitude resultante dos desvios nas etapas de calibração, validação cruzada e predição (RMSEC, RMSECV e RMSEP, respectivamente). Cada entidade, matematicamente expressa como a raiz quadrada da média entre desvios, é um indicador da variabilidade randômica na relação entre os dados reais e os previstos mediante o processo de regressão ¹⁶⁰.

A Figura 20 é uma representação genérica de quaisquer das três expressões de desvio randômico do modelo. Marginal à linha preta (resíduo zero), está a região proporcional à magnitude média dos desvios decorrentes da modelagem. Entendida como

desvio do modelo com relação aos dados, a relação resultante de cada elemento e a linha central ($(y_i - \hat{y}_i)$ da equação 14) é chamada de resíduo. Por fim, dependendo da etapa avaliativa, o componente Graus de Liberdade (GL) assume diferentes valores, conforme pré-processamentos e números de amostras (N) nas etapas de calibração (RMSEC e RMSECV) ou em previsão (RMSEP).



Fonte: O autor.

Figura 20. Distribuição randômica dos resíduos de regressão.

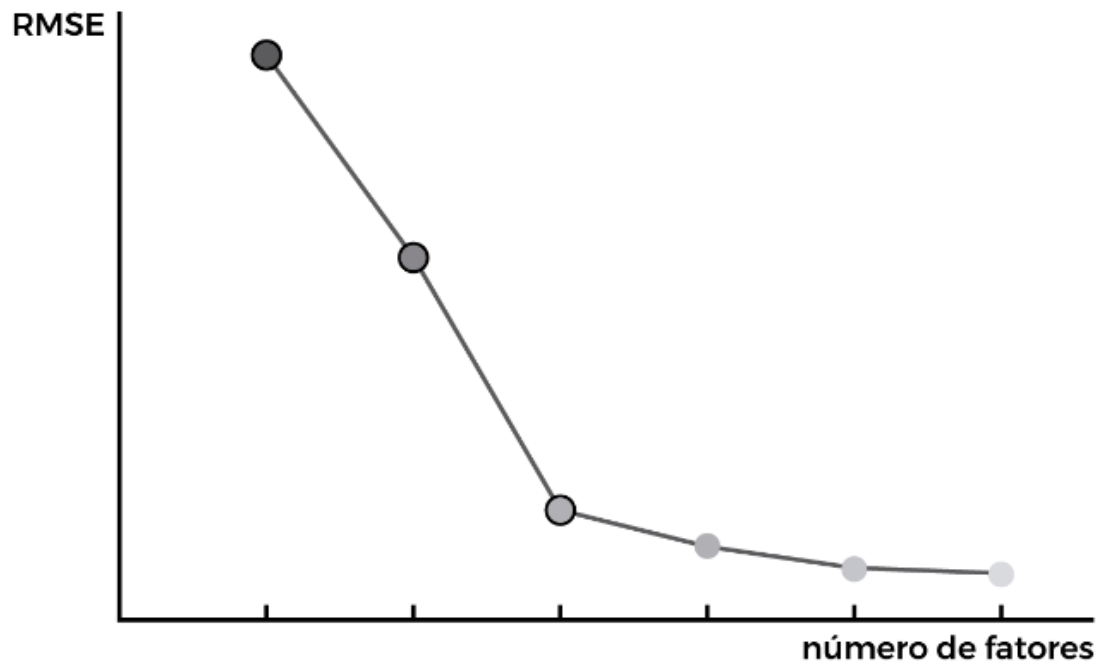
É importante salientar que, a título de simplificação, a aproximação simbolizada pela equação 14 é correta. Ao assumir essa aproximação, assim, ignora-se o rigor estatístico, sem perda numérica relevante por desconsiderar termos referentes aos graus de liberdade ¹¹⁹.

$$RMSE = \sqrt{\frac{1}{GL} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (14)$$

Ambos, RMSEC e RMSECV, informam a performance do modelo ajustado em relação aos dados de calibração, envolvendo todo o conjunto de amostras ou nos processos de validação cruzada, respectivamente. O RMSECV, porém, é empregado na seleção do número de variáveis latentes a se utilizar nas regressões ^{153,156}. De modo a sumarizar os resíduos da predição, o RMSEP reflete a escolha de fatores utilizados na regressão e devida atenção precisa ser dada a essas etapas para evitar sub- ou sobre-ajustes ^{119,161}.

Para fundamentar a decisão segura sobre a composição do modelo, frequentemente, utiliza-se o gráfico da relação componente de variância *versus* número de fatores. Uma vez que a inclusão de fatores aumenta proporcionalmente a taxa de variância explicada pelo modelo, é razoável selecionar a mínima parcela cujos incrementos sejam expressivos ¹¹⁹. Em outras palavras, representando na Figura 21, escolhe-se o número de fatores a partir dos quais a taxa de inclinação da variância residual se estabiliza.

Figura 21. Erro residual do modelo e a inclusão de fatores que o compõe



FONTE: O autor. Arte de Josué Pereira Dias



Por Anelise Maria Wasem Pires Schneider

3. METODOLOGIA

3.1 Amostras

Na execução de todas etapas do presente estudo, utilizou-se amostras de vinho tinto cedidas pela Embrapa Uva e Vinho – Bento Gonçalves. Os objetos de estudo de reconhecimento de padrões foram amostras previamente selecionadas como de destaque no concurso “Avaliação Nacional de Vinhos” dos anos 2017 e 2018. De um total de cento e cinco (105), noventa e duas (92) amostras representam a produção de duas safras de vinícolas gaúchas variadas. Tendo em vista a aleatoriedade das variedades participantes, é importante ressaltar que não necessariamente as amostras tenham exemplares em ambas safras dos anos de 2017 e 2018.

Além das diferentes localidades de procedência e safra, as amostras utilizadas nos estudos de reconhecimento de padrões e seleção de variáveis também são divididas em dez (10) diferentes variedades de uvas tintas. A tabela completa com informações de safra, procedência e tipo de uva pode ser consultada no Apêndice A.

Os estudos de regressão empregaram 4 (quatro) amostras, também cedidas pela Embrapa – Uva e Vinho. Todavia, diferentemente dos vinhos tintos da etapa de reconhecimento de padrões, as amostras foram produzidas sob coordenação do corpo técnico especialista da Embrapa. Uma vez que misturar variedades de uva na produção de vinhos é uma prática comum e prevista por lei, escolheu-se avaliar o perfil de vinhos, e suas variações em misturas conhecidas, com amostras seguramente monovarietais.

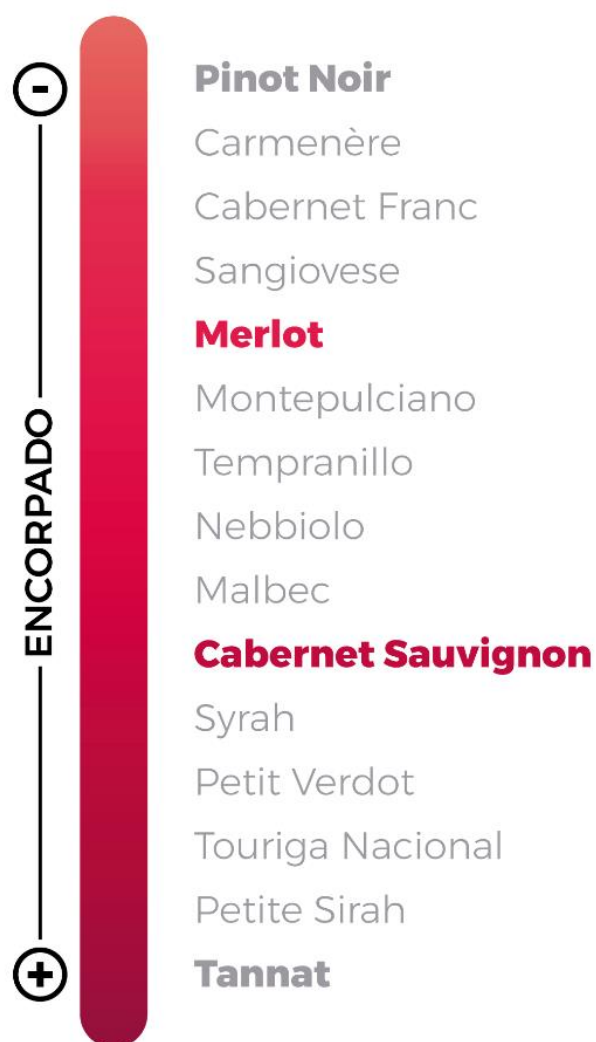
Após transportadas para o Laboratório de Análise e Instrumentação Analítica – LAQIA, no Campus do Vale da UFRGS, as amostras foram reservadas em local escuro até o momento da abertura. O objeto de estudo da presente dissertação é parte integrante de um grupo, que se dedicou a pesquisar outras vias analíticas e quimiométricas com esses vinhos. Assim, alíquotas de todas as amostras foram acondicionadas na geladeira do laboratório, no momento que foram abertas. Então, para o registro do espectro de infravermelho médio dessas amostras, permitiu-se que essas atingissem o equilíbrio térmico com o ambiente.

3.1.1 Preparo de amostras

As misturas aqui estudadas foram projetadas para simular condições comumente aplicadas na viticultura, como a variação de composição de variedades de uva, que visa a diversificação e complexidade nas suas características sensoriais do produto final. O *blend*

(do inglês, mistura) de vinhos pode ser complexo a ponto de representar a assinatura de uma empresa. Todavia, o processo aqui estudado considerou a adição controlada de um vinho mais encorpado em um vinho base, menos encorpado.

Por integrar o grupo das variedades de uvas produzidas no Estado do Rio Grande do Sul (ilustradas na Figura 22), uma garrafa monovarietal Merlot e três de Cabernet Sauvignon foram cedidas para o estudo em métodos de regressão. Considerando a diversidade amostral limitada, misturas secundárias foram preparadas integrando mais de uma garrafa na proporção de Cabernet Sauvignon em sua composição, assim ampliando possibilidades de avaliação dos métodos de regressão.



Fonte: Adaptado de Goldstein ⁶. Arte de Josué Pereira Dias

Figura 22. Relação de uvas e vinhos tintos produzidos no Rio Grande do Sul

Nas misturas preparadas, o monovarietal Merlot foi o componente majoritário em todos conjuntos de soluções. Ainda, essas misturas sempre consideraram a mesma proporção global de ambos monovariais envolvidos. Ou seja, mesmo nas misturas contendo vinho Cabernet Sauvignon de duas ou três garrafas diferentes, a concentração específica desse varietal é a mesma das misturas iniciais, compostas por uma garrafa de Cabernet Sauvignon e uma de Merlot.

De modo a evitar tendências e vícios por erros operacionais, foi necessário um planejamento das concentrações por meio do sorteio randômico executado por *software* computacional. As concentrações estão discriminadas na **Tabela 8**, incluindo alguns pontos com concentração deliberadamente escolhida para testes de validação.

Para cada conjunto de avaliação dos métodos de regressão, foram produzidas vinte e nove (29) misturas (ou alíquotas, no caso dos vinhos monovarietal), onde vinte e uma (21) foram nomeadas como conjunto de calibração e oito (8) como validação. Uma vez que cada ponto de calibração e validação foi produzido em duplicata, o conjunto de misturas totaliza cinquenta e oito (58) amostras. Assim, foram produzidas quatrocentas e seis (406) misturas.

Tabela 8. Concentrações de monovariais presentes nas misturas

| Modelo | Identificação | Concentração do monovarietal (% v/v) | |
|------------|---------------|--------------------------------------|--------|
| | | Cabernet Sauvignon | Merlot |
| (continua) | | | |
| CALIBRAÇÃO | C01 | 0,00 | 100,0 |
| | C02 | 6,08 | 93,92 |
| | C03 | 6,24 | 93,76 |
| | C04 | 8,08 | 91,92 |
| | C05 | 8,30 | 91,70 |
| | C06 | 8,66 | 91,34 |
| | C07 | 9,74 | 90,26 |
| | C08 | 9,82 | 90,18 |
| | C09 | 10,14 | 89,86 |
| | C10 | 11,76 | 88,24 |
| | C11 | 12,05 | 87,95 |

Tabela 8. Concentrações de monovarietais presentes nas misturas

| Modelo | Identificação | Concentração do monovietal (% v/v) | |
|------------|---------------|------------------------------------|--------|
| | | Cabernet Sauvignon | Merlot |
| CALIBRAÇÃO | C12 | 16,13 | 83,87 |
| | C13 | 17,23 | 82,77 |
| | C14 | 17,27 | 82,73 |
| | C15 | 17,48 | 82,52 |
| | C16 | 18,40 | 81,60 |
| | C17 | 18,47 | 81,53 |
| | C18 | 20,90 | 73,10 |
| | C19 | 22,40 | 77,60 |
| | C20 | 26,38 | 73,62 |
| | C21 | 26,80 | 73,20 |
| VALIDAÇÃO | V01 | 30,00 | 70,00 |
| | V02 | 25,00 | 75,00 |
| | V03 | 20,00 | 80,00 |
| | V04 | 15,00 | 85,00 |
| | V05 | 10,00 | 90,00 |
| | V06 | 5,00 | 95,00 |
| | V07 | 2,50 | 97,50 |
| | V08 | 0,00 | 100,00 |

3.2 Registro dos Espectros IR

A aquisição e registro dos espectros foi realizada mediante FTIR – ATR *Cary 630*, da Agilent, equipamento ilustrado pela Figura 23. Ao espectrômetro de infravermelho com transformada de Fourier foi acoplado o acessório de refletância total atenuada com o seletor de comprimento de caminho óptico, o *DialPath*^{162,163}. No Apêndice B pode ser consultada a tabela de dias e horários de aquisição dos espectros.

Uma porção inicial das amostras foi utilizada para testar o melhor caminho óptico a ser empregado. A indicação do fabricante concorda com a proporcionalidade presente na Lei

de Lambert-Beer, de que para soluções de baixas concentrações utiliza-se longos caminhos, enquanto em soluções concentradas aplica-se curtos caminhos ¹⁶².



Fonte: “Agilent Cary 630 FTIR sample interfaces”, 2011.¹⁶³

Figura 23. Espectrômetro FTIR – ATR, Cary 630

Os espectros foram registrados em laboratório de pesquisa do Instituto de Química da UFRGS, no LAQIA especificamente, onde a variabilidade das condições de temperatura e umidade exigem uma adaptabilidade do equipamento. Para sobressair maiores interferências, mesmo assim, além da atenção para que as amostras estivessem em equilíbrio térmico com o ambiente no momento do registro do espectro, cada amostra teve registrado o espectro em triplicata. A resolução do equipamento foi de 2 cm^{-1} , cada registro resultou de 12 *scans* (varreduras) e a faixa espectral de trabalho foi de $2000\text{ a }800\text{ cm}^{-1}$.

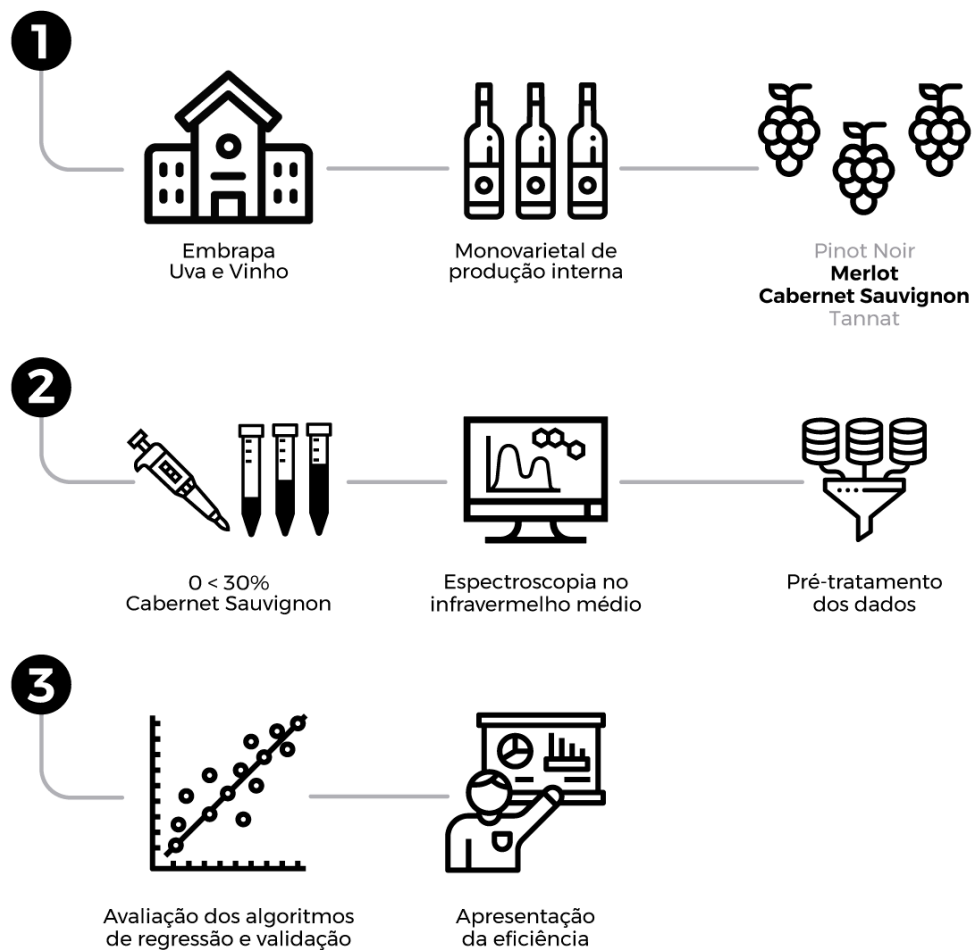
O registro das amostras foi feito ao longo dos anos de 2018 e 2019, por mais de um analista. Para os estudos de reconhecimento de padrões e regressão, as análises contaram com o apoio da aluna de pós-graduação em Química da UNIFESSPA, Kelly Cristina Mendes de Souza. Enquanto que algumas amostras de regressão foram analisadas pelo aluno de graduação em Química da UFRGS e iniciação científica, Fernando Diogo Sena.

3.3 Processamento de Dados

3.3.1 Regressão

Como esquematizado na Figura 24, as amostras de vinho do tipo monovarietal foram doadas pela Embrapa e, a partir destas foram produzidos os conjuntos de misturas para o estudo. Dos sete conjuntos de mistura, cada um continha vinte e nove concentrações diferentes, variando entre zero e trinta por cento de Cabernet Sauvignon. O preparo de soluções foi feito em duplicatas, totalizando quatrocentos e seis (406) amostras. Uma vez que cada amostra teve registrado seu espectro em triplicata, o tratamento de dados inicial partiu

de uma matriz de mil duzentos e dezoito amostras por mil duzentos e trinta e cinco (1235) variáveis.



FONTE: O autor. Arte de Josué Pereira Dias

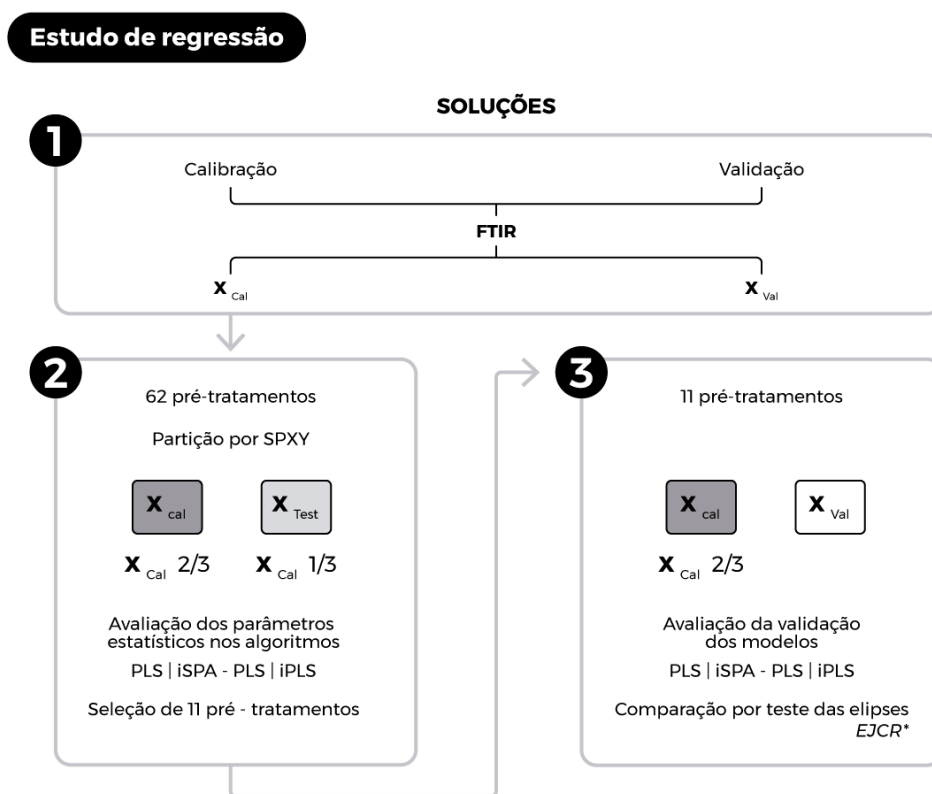
Figura 24. Fluxograma do Estudo de Regressão

Todos os conjuntos de dados foram processados conforme representado **Figura 25**. Primeiramente, regiões espectrais dos dois extremos foram subtraídas do conjunto de variáveis, por serem ruidosas ou porque continham pouca informação relevante. Partindo do conjunto de três replicatas por amostra (174 linhas ao todo), selecionou-se as duplicatas mais similares entre si por meio de uma rotina em Matlab, desenvolvida no laboratório. Uma vez que cada ponto de concentração possui soluções em duplicata, a seleção por espectros similares resultou em 4 espectros, dos quais foi calculada a média aritmética.

De posse dessa matriz de dados médios de concentração, foi necessário aplicar alguma transformação para suavizar o ruído espectral ainda existente. Contando com o

espectro de dados brutos, ao todo foram executados cinquenta e sete (57) pré-tratamentos, variações das componentes de suavização por Derivadas Savitzky – Golay (polinômio e número de pontos na janela), como consta na Tabela 9. Tais transformações foram executadas separadamente, primeiro para o grupo de soluções “calibração” seguido pelo grupo “validação”.

O conjunto de espectros das soluções “calibração”, agora transformados, é então particionado aplicando o algoritmo SPXY, designando quatorze (14) amostras para a calibração propriamente dita do modelo, enquanto sete (7) amostras são reservadas para teste. Após a partição, cada matriz de dados foi aplicada no método de regressão PLS, via caixa *toolbox* iSPA_GUI desenvolvida para execução de pesquisas acadêmicas ¹²⁹ em Matlab (disponibilizado pela UFRGS na versão R2012b).



Fonte: O autor. Arte de Josué Pereira Dias.

Figura 25. Esquema de tratamento das amostras de Regressão

Os conjuntos de dados, cujos modelos apresentaram um desempenho adequado nas etapas de calibração e teste em PLS, foram, então, submetidos a métodos de regressão por seleção de variáveis. Com apoio da mesma *toolbox*, em Matlab avaliou-se a eficiência da

calibração e teste dos métodos de regressão por mínimos quadrados parciais em intervalos da faixa espectral (iPLS) e de sucessivas projeções (iSPA-PLS).

Com a seleção dos melhores modelos, construídos com as soluções calibração, executou-se um segundo teste, agora com a matriz de dados das soluções de validação. Os espectros desse novo conjunto de soluções passaram pelas mesmas transformações (Tabela 9) visando a suavização de ruídos e correção de inclinações de linha de base, tal qual executada no seu respectivo conjunto de calibração.

A seleção do melhor modelo em cada um dos três métodos avaliados foi feita considerando a função multiobjetivo, demonstrada pela equação 14. Tal função é composta das respostas estatísticas, fatores que refletem direta ou inversamente a eficiência modelo, *e.g.*, coeficiente de correlação e erros de predição, respectivamente. A linha de respostas da função devolverá um ponto de mínimo que foi utilizado para guiar a escolha dos melhores modelos.

$$G = \sum_{m=1}^M g_{cal} + \sum_{l=1}^L g_{teste} \quad (14)$$

| Soluções | | Calibração | | |
|----------|---|------------------------|------------------------|---|
| 1 | | | RMSEC | REMSEP |
| 2 | | Calibração | 1/r | REP |
| 3 | | | 1/R² | t_{cal} - t_{crit} |
| 4 | | <i>g_{cal}</i> | RMSECV | <i>g_{teste}</i> |
| 5 | L | Validação cruzada | 1/r | 1/R² |
| 6 | M | | 1/R² | |

Tabela 9. Pré-processamento dos dados FTIR – ATR

| Transformação | Jan* | Pol* | Der* | Transformação | Jan* | Pol* | Der* | Transformação | Jan* | Pol* | Der* | |
|---------------|------------|--------|------|---------------|------------|-----------|------|---------------|--------------|----------|------|--|
| X02 | 7 | 2 | 1 | X06 | 9 | 2 | 1 | X10 | 11 | 2 | 1 | |
| X03 | | 2 | 2 | X07 | | 2 | 2 | X11 | | 2 | 2 | |
| X04 | | 3 | 1 | X08 | | 3 | 1 | X12 | | 3 | 1 | |
| X05 | | 3 | 2 | X09 | | 3 | 2 | X13 | | 3 | 2 | |
| X14 | 13 | 2 | 1 | X18 | 15 | 2 | 1 | X22 | 17 | 2 | 1 | |
| X15 | | 2 | 2 | X19 | | 2 | 2 | X23 | | 2 | 2 | |
| X16 | | 3 | 1 | X20 | | 3 | 1 | X24 | | 3 | 1 | |
| X17 | | 3 | 2 | X21 | | 3 | 2 | X25 | | 3 | 2 | |
| X26 | 19 | 2 | 1 | X30 | 21 | 2 | 1 | X34 | 23 | 2 | 1 | |
| X27 | | 2 | 2 | X31 | | 2 | 2 | X35 | | 2 | 2 | |
| X28 | | 3 | 1 | X32 | | 3 | 1 | X36 | | 3 | 1 | |
| X29 | | 3 | 2 | X33 | | 3 | 2 | X37 | | 3 | 2 | |
| X38 | 25 | 2 | 1 | X42 | 27 | 2 | 1 | X46 | 29 | 2 | 1 | |
| X39 | | 2 | 2 | X43 | | 2 | 2 | X47 | | 2 | 2 | |
| X40 | | 3 | 1 | X44 | | 3 | 1 | X48 | | 3 | 1 | |
| X41 | | 3 | 2 | X45 | | 3 | 2 | X49 | | 3 | 2 | |
| X50 | 31 | 2 | 1 | X54 | 33 | 2 | 1 | X01 | Dados Brutos | | | |
| X51 | | 2 | 2 | X55 | | 2 | 2 | | | | | |
| X52 | | 3 | 1 | X56 | | 3 | 1 | | | | | |
| X53 | | 3 | 2 | X57 | | 3 | 2 | | | | | |
| Notas: | Jan | Janela | | | Pol | Polinômio | | | Der | Derivada | | |

3.3.2 Reconhecimento de Padrões

Ao chegar nas dependências do laboratório, as noventa e nove (99) amostras de vinho tinto permaneceram em local escuro até o momento da execução das primeiras análises, tendo suas alíquotas reservadas em geladeira para as análises via infravermelho. O registro dos espectros ocorreu em três etapas, devido a demanda de trabalho requerida ou a chegada de novas remessas amostrais no laboratório, conforme resumido pela Figura 26.



FONTE: O autor. Arte de Josué Pereira Dias.

Figura 26. Esquema do estudo de Reconhecimento de Padrões

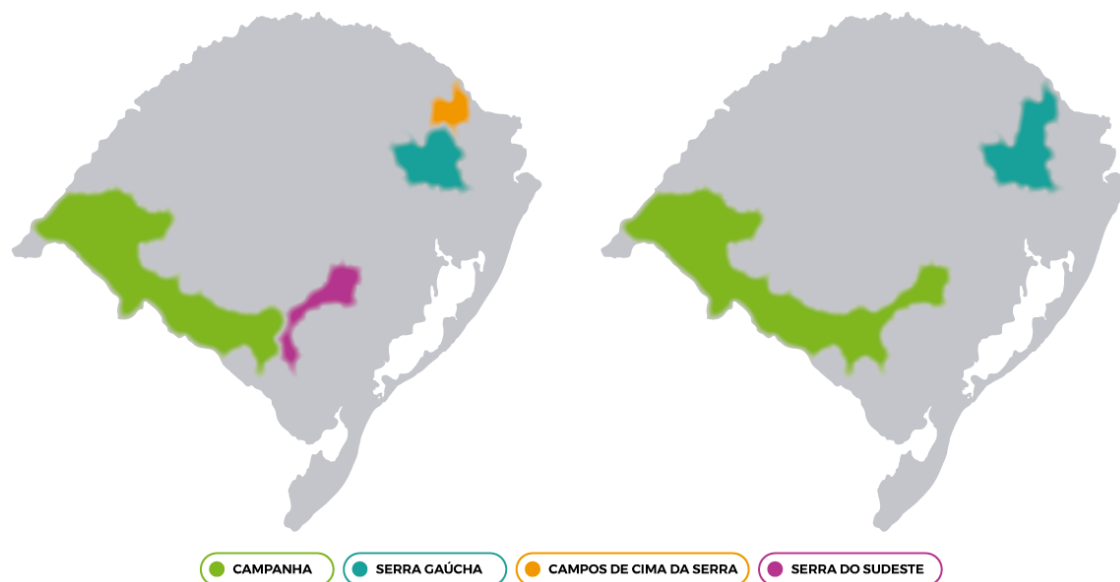
Do conjunto de dados total, restringiu-se a análise de reconhecimento de padrões às amostras de produção gaúcha. Para cada amostra, primeiramente selecionou-se as quatro (4) replicatas mais similares entre si, mediante rotina aplicada em Matlab. Após, a

média das replicatas foi aplicada em análise de componentes principais ¹⁶⁴ para visualização prévia dos dados. Assim, optou-se por segregar das análises posteriores as amostras consideradas outliers.

Devido à proximidade entre as subregiões tanto do nordeste quanto sudoeste gaúcho, e a reduzida representatividade amostral das áreas em ascensão produtiva, as amostras foram agrupadas em dois conjuntos maiores, como demonstrado na Figura 27. Assim, amostras provenientes da Serra do Sudeste passam a integrar a Campanha Gaúcha (Classe 1), ao passo que as amostras de Campos de Cima da Serra integram a Serra Gaúcha (Classe 2).

De um total de noventa e duas amostras, trinta e duas são consideradas oriundas da Classe 1, e sessenta são da Classe 2. De acordo com a Tabela 10, é possível perceber que, apesar de não balanceada, a composição amostral é diversificada em termos de safra, região e varietal.

Tal qual no estudo de regressão, a matriz de dados para reconhecimento de padrões teve reduzida sua dimensão relacionada às variáveis, os números de onda. Por exclusão de regiões espectrais periféricas, ruidosas ou sem informações significativas, o número de variáveis passou de 1235 para 645. Essa matriz foi empregada tanto para o processo não-supervisionado quanto para o supervisionado, esse último considerando a divisão entre Classe 1 e Classe 2.



Fonte: O autor. Arte de Josué Pereira Dias

Figura 27. Grandes regiões produtoras de uva e vinho

Tabela 10. Composição do conjunto amostral

| | | safra 2017 | safra 2018 | Região | Varietal | |
|------------------|---|------------|------------|--------|-----------|--------------|
| | | | | | Total | Total (%) |
| AA | c | 2 | 0 | 2 | 8 | 8,70 |
| | s | 2 | 4 | 6 | | |
| CF | c | 1 | 3 | 4 | 9 | 9,78 |
| | s | 1 | 4 | 5 | | |
| CS | c | 1 | 5 | 6 | 18 | 19,57 |
| | s | 1 | 11 | 12 | | |
| MC | c | 0 | 1 | 1 | 7 | 7,61 |
| | s | 3 | 3 | 6 | | |
| MN | c | 2 | 3 | 5 | 9 | 9,78 |
| | s | 2 | 2 | 4 | | |
| MT | c | 1 | 0 | 1 | 12 | 13,04 |
| | s | 3 | 8 | 11 | | |
| PN | c | 1 | 2 | 3 | 5 | 5,43 |
| | s | 2 | 0 | 2 | | |
| PS | c | 0 | 0 | 0 | 3 | 3,26 |
| | s | 3 | 0 | 3 | | |
| TO | c | 2 | 2 | 4 | 6 | 6,52 |
| | s | 2 | 0 | 2 | | |
| TT | c | 2 | 4 | 6 | 15 | 16,30 |
| | s | 2 | 7 | 9 | | |
| TOTAL | | 33 | 59 | 92 | 92 | 100,0 |
| TOTAL (%) | | 35,87 | 64,13 | | | |

| | | | |
|-----------|--------------------|-----------|--------------|
| AA | Ancelota | MT | Merlot |
| CF | Cabernet Franc | PN | Pinot Noir |
| CS | Cabernet Sauvignon | PS | Petit Shirah |
| MC | Malbec | TO | Teroldego |
| MN | Marselan | TT | Tannat |

Aplicados por DATA HAND GUI, a matriz de dados passou por etapas de pré-tratamentos similares aos apresentados na Tabela 9. Desta forma, os novos conjuntos de dados também são enumerados de X01 a X57, cada um com uma variação do tamanho da janela, do polinômio, ou do grau da derivada Savitzky-Golay.

O estudo de reconhecimento de padrões foi executado em Matlab 2012b. No modo não-supervisionado por meio da PCA TOOLBOX 1.5 ¹⁶⁴. Os métodos de reconhecimento supervisionado foram executados mediante CLASSIFICATION TOOLBOX 5.3 ¹³⁶ e LDA VS GUI ¹²⁹.

Além dos métricas para avaliação da eficiência da classificação produzida pela validação cruzada, foi aplicada a função multiobjetivo da equação 14, que considera o

conjunto de métricas e a relação (direta ou inversa) de desempenho do modelo. Essa função, auxiliou na escolha do modelo a ser aplicado no grupo amostral de teste.

Intrinsecamente, alguns fatores da função multiobjetivo consideram o número de ocorrências de falsos positivos e negativos para uma dada classe, ponderados pelo número de integrantes total de cada classe. Na comparação entre os modelos procura-se a existência de um mínimo, proveniente de baixas ocorrências de erro em classificação.

Os modelos selecionados foram, por fim, comparados em eficiência com o GA-LDA. Para tal, procedeu-se uma otimização do tamanho da população e o número de gerações empregando a função intitulada de aptidão. A superfície resultante dessa função fornece pontos de máximo, cujas coordenadas serviram de parâmetro para aplicação posterior em simulação de Monte Carlo (MC). De forma estocástica, essa simulação estima a probabilidade de uma correta classificação amostral, tomando como base as classes e os espectros de FTIR fornecidos.

A simulação de Monte Carlo aliada ao Algoritmo Genético pode fornecer uma estimativa das variáveis de maior influência, conforme sua probabilidade de ocorrência. Com os parâmetros otimizados, o algoritmo genético foi executado ao longo de cem (100) ciclos, quantidade deliberadamente selecionada, considerando a randomicidade da simulação (MC) e o custo computacional que cada ciclo necessita.

A musical score featuring a large, colorful graphic of the word 'LOVE' in a stylized, outlined font. The letters are filled with a rainbow gradient. The graphic is superimposed over a musical score. The score consists of several staves. The first staff has a treble clef and a key signature of one sharp (F#). It contains a melodic line with dynamics *cresc.* and *p*. The second staff has a bass clef and contains a melodic line with dynamics *f* and *sf*. The third staff is a continuation of the melodic line. The fourth staff is a continuation of the melodic line, starting at measure 84. The fifth staff is a continuation of the melodic line, starting at measure 89. The sixth staff is a continuation of the melodic line, starting at measure 93. The graphic 'LOVE' is positioned above the first two staves, with the 'L' and 'O' being the largest and most prominent.

Por Anelise Maria Wasem Pires Schneider

4. RESULTADOS E DISCUSSÃO

4.1 Escolha do Caminho óptico

Inicialmente, um breve teste fora executado para decidir o caminho óptico a ser empregado na aquisição dos espectros FTIR – ATR. De fato, o instrumento disponível no laboratório contém o acessório chamado *Dial Path*, um seletor de comprimento de caminho óptico. Estão disponíveis para uso três caminhos ópticos diferentes, 50, 100 e 150 μm . O teste foi aplicado somente nos caminhos menores.

Pela Figura 28 **Erro! Fonte de referência não encontrada.** é possível perceber a variação da Absorbância de uma amostra nos dois caminhos ópticos estudados. De acordo com a Lei de Lambert – Beer, o espectro de FTIR com caminho óptico maior apresenta absorbância maior. Todavia, apesar de acentuar bandas características já observadas via 50 μm , o caminho óptico de 100 μm foi preterido devido à oscilação no instante da aquisição do sinal e, conseqüentemente, no crescimento do ruído na média final dos espectros.

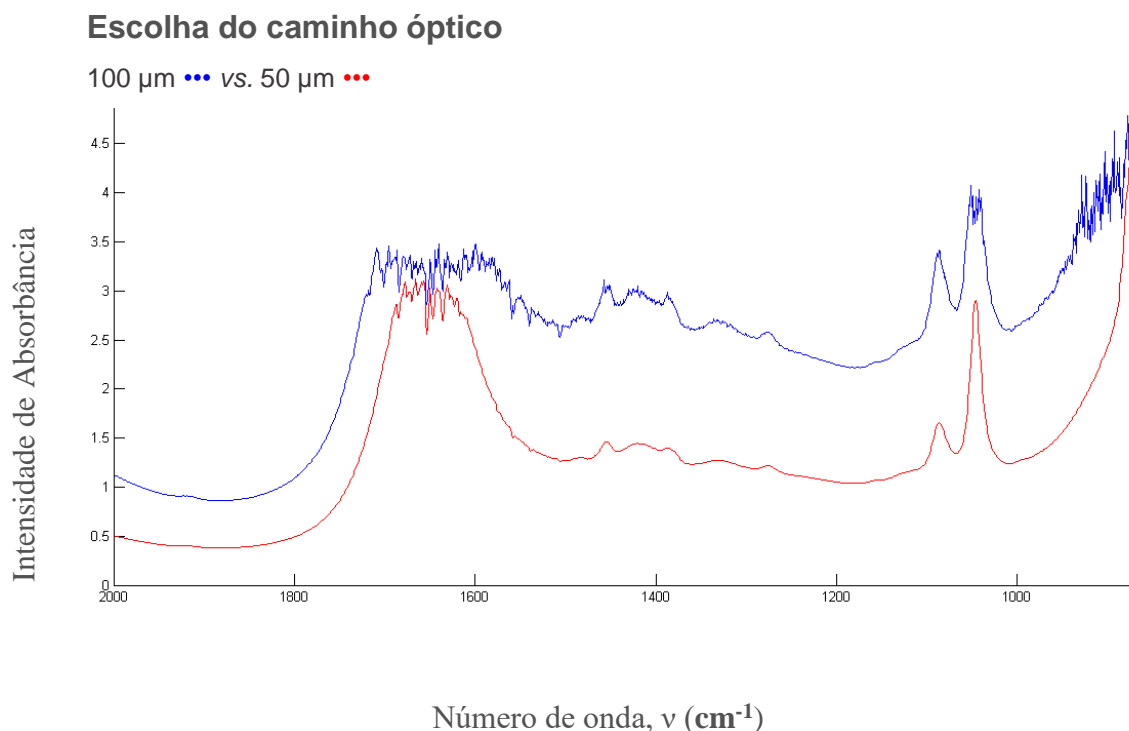


Figura 28. Absorbância com variação do caminho óptico

Após escolhido o tamanho do caminho óptico, os espectros foram registrados e observações iniciais puderam ser realizadas, mas serão melhor discutidas nas figuras das próximas seções (Figura 42). Em suma, a média amostral da Campanha é maior que a Serra Gaúcha em regiões do espectro no infravermelho que correspondem a presença de álcoois e compostos polifenólicos.

4.2 Reconhecimento de padrão

Modelos não-supervisionados

PCA

O processo de reconhecimento de padrões foi iniciado pelo método não-supervisionado de análise por componentes principais (PCA). Primeiramente, mediante uma breve inspeção no gráfico de escores dos dados brutos de noventa e nove amostras, algumas amostras foram excluídas, uma vez que se dispersaram de pequenos grupos, de maneira a considerá-las amostras anômalas.

Após a exclusão de amostras anômalas, as noventa e duas amostras restantes foram aplicadas em uma nova análise por PCA. Sem tratamentos no conjunto de espectros, os gráficos de escores com classificação segundo as safras, origens e ambos fatores são representados nas Figura 33, Figura 29 e Figura 31, respectivamente. Em conjunto com os escores dos dados brutos, há a comparação com a PCA aplicado aos dados pré-processados por derivada primeira do método de Savitsky – Golay, com janela de trinta e um pontos e adequação a polinômio de segundo grau (X50).

Os dados brutos, bem como os pré-processados, não apresentam clara separação de classes conforme sua origem. Ou seja, a PCA não é conclusivo no reconhecimento de padrões dos vinhos provenientes da Campanha e da Serra Gaúcha, Figura 29 e Figura 30.

Apesar de manter esse mesmo padrão inconclusivo com relação a procedência, é possível perceber uma tendência de agrupamento nas Figura 31 e Figura 32, que compreendem as classes Origem integradas ao ano da Safra. Os agrupamentos se dão por um lado, majoritariamente, com as amostras da Campanha e Serra Gaúcha da Safra de 2017, por outro lado, com as amostras dessas regiões, porém na safra de 2018.

O último teste via PCA foi, então, aplicar os conjuntos de dados a este método, mas considerando somente a variável safra. As Figura 33 e Figura 34 mostram os gráficos de escores com as classes 2017 e 2018. É possível generalizar que a capacidade de reconhecimento de padrão se deu em função da safra, onde as amostras de 2017 foram agrupadas nos quadrantes inferiores da PC2, enquanto as de 2018 se distribuíram, em grande maioria, na metade superior.

Dispersão amostral – Campanha vs. Serra Gaúcha

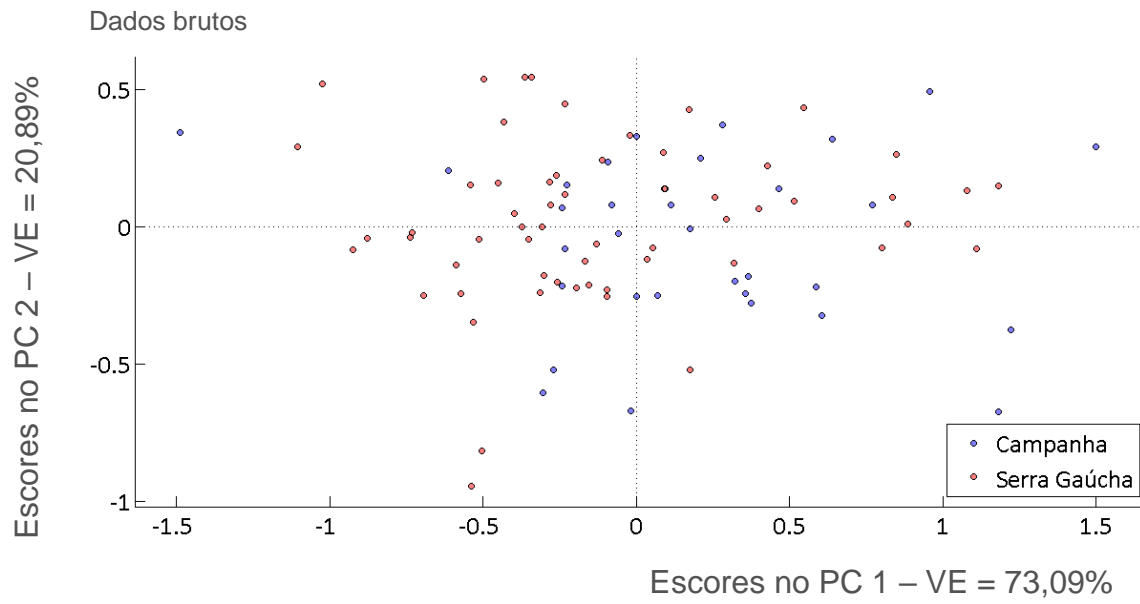


Figura 29. PCA, Dados Brutos; Origens

Dispersão amostral – Campanha vs. Serra Gaúcha

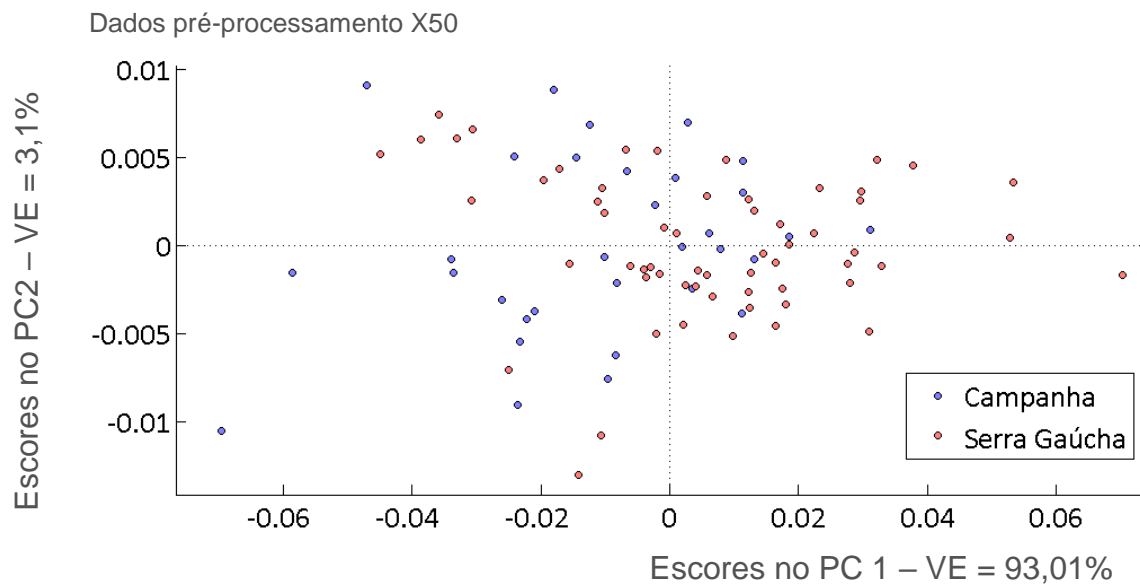


Figura 30. PCA, Dados pré-processados; Origens

Dispersão amostral – Safras vs. Origem

Dados brutos

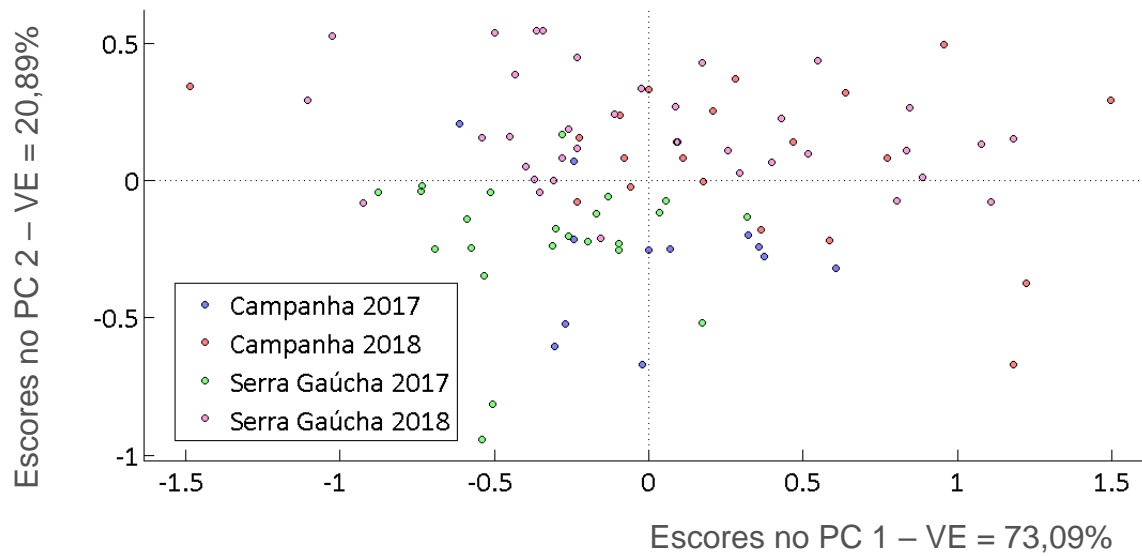


Figura 31. PCA, Dados Brutos; Safras vs. Origens

Dispersão amostral – Safras vs. Origem

Dados pré-processamento X50

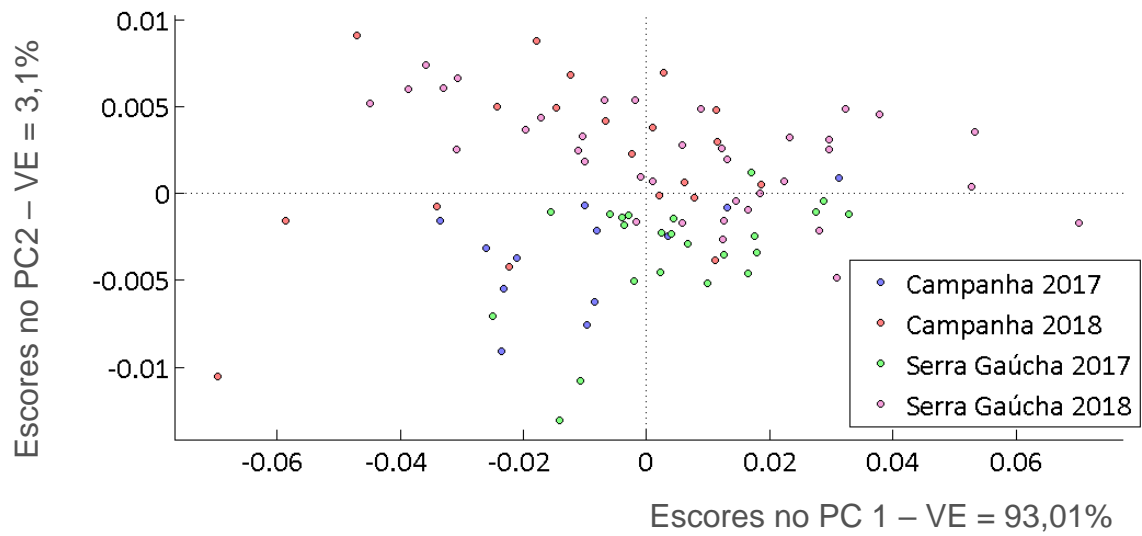


Figura 32. PCA, Dados Brutos; Safras vs. Origens

Dispersão amostral – Safras 2017 vs. 2018

Dados brutos

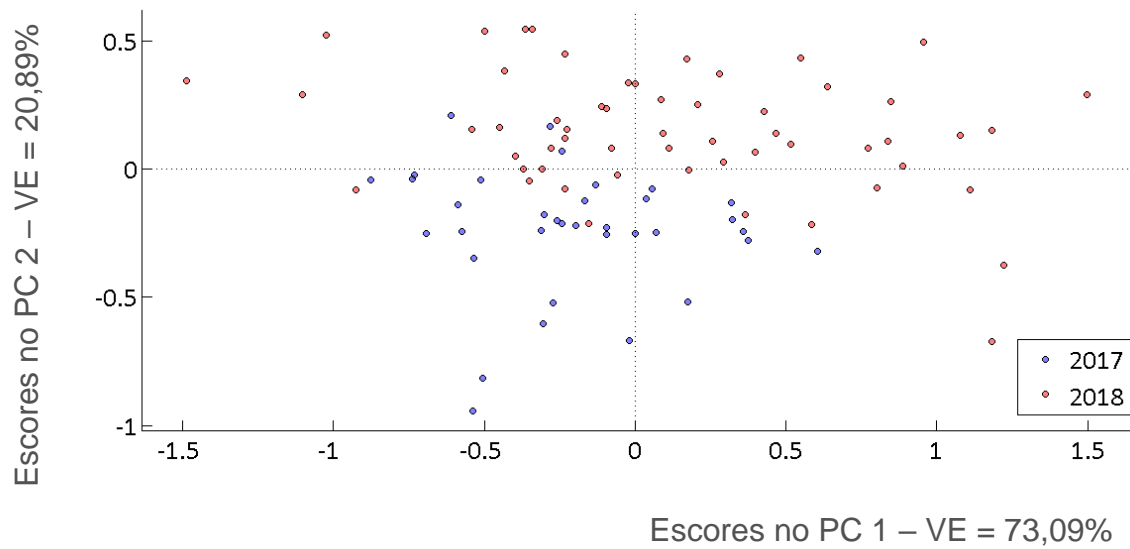


Figura 33. PCA, Dados Brutos; Safras

Dispersão amostral – Safras 2017 vs. 2018

Dados pré-processamento X50

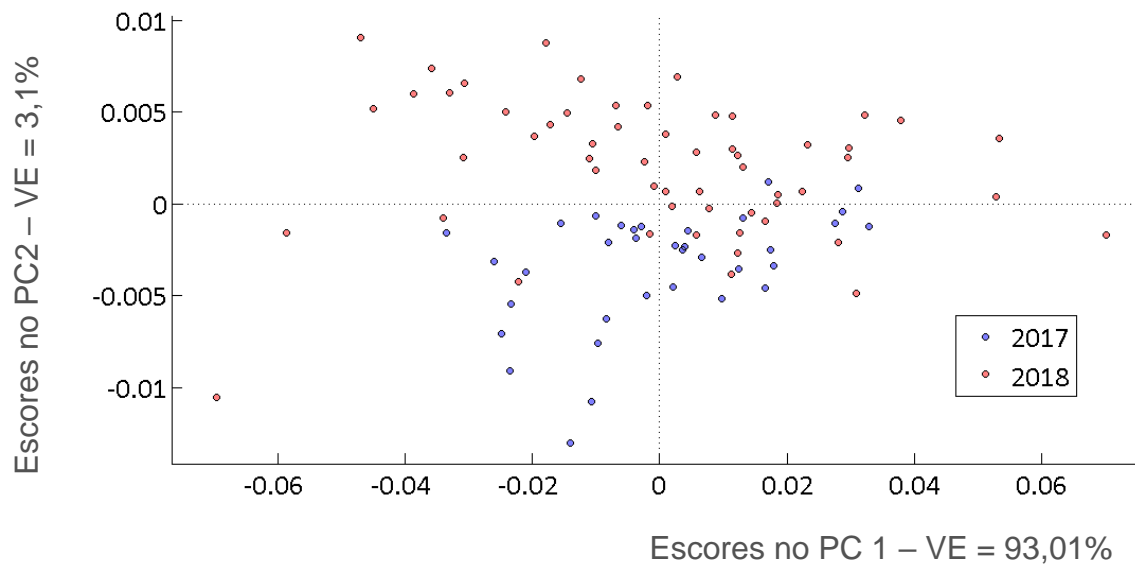


Figura 34. PCA, Dados pré-processados; Safras

É possível averiguar que uma das principais fontes de variação dos espectros é proveniente das diferenças entre as safras, onde as amostras de 2017 foram agrupadas nos quadrantes inferiores da PC2, enquanto as de 2018 se distribuíram, em grande maioria, na metade superior. A informação de origem geográfica, embora pareça estar contida nos

espectros, não se revelou como sendo a fonte de variabilidade proeminente, o que tornará a tarefa de desenvolvimento de modelos preditivos uma tarefa não trivial.

Uma das características do reconhecimento de padrões via PCA é a análise de *Loadings*, ou pesos. A Figura 35 permite constatar que os Números de Onda entre 1030 e 1090 (cm^{-1}), aproximadamente, são significativos na separação das classes, região referente à vibração C – O, de álcool primário.

Variáveis relevantes no processo de Reconhecimento de Padrões

Dados pré-processados; X50 (Janela 31 – Polinômio 2 – Derivada 1). Peso na variável 1.

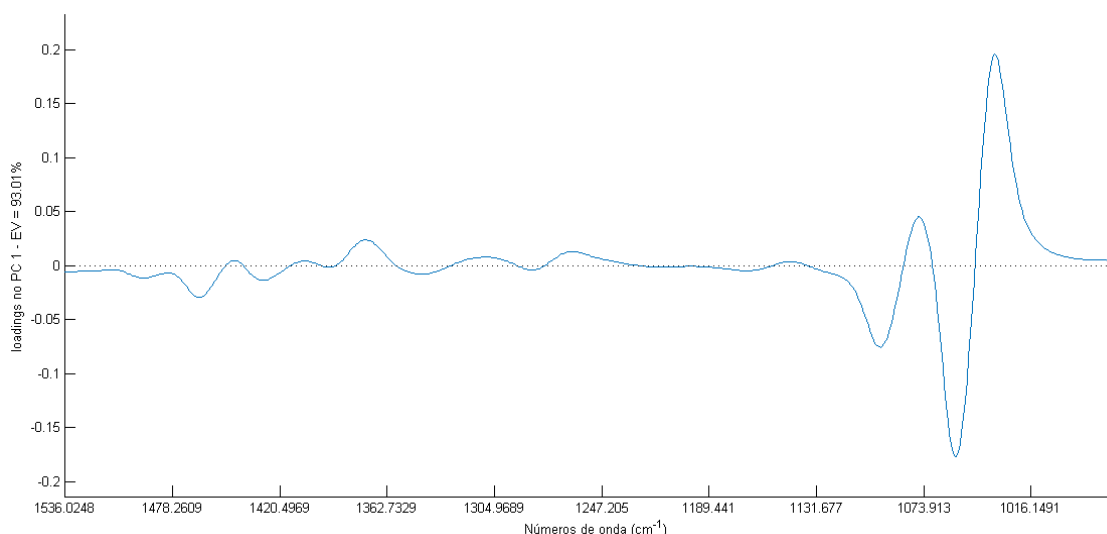


Figura 35. Loadings PCA

Modelos supervisionados

PLS – DA

Inicialmente, cada um dos conjuntos de dados designados à calibração do modelo foi inspecionado de modo a selecionar o número ótimo de fatores. Essa seleção é feita com base na menor taxa de erro na validação cruzada (RMSECV). Devido a resposta particular para cada conjunto de dados, se torna complexo definir tendências correlacionadas aos elementos desse estudo primário.

Apesar da inclusão gradativa de fatores caracterizar eficientemente o conjunto avaliado, o sobreajuste representa uma problemática na aplicação do modelo em novas amostras. Assim, a opção escolhida como mais segura para o treinamento do modelo via PLS – DA foi aquela apresentando o menor erro em validação cruzada, onde o número de fatores não ocasionasse amostras sem classe definida, como na Figura 36, onde escolheu-se quatro variáveis latentes.

Variação da Taxa de Erro de Predição na Validação Cruzada

Conjunto de dados bruto – X01

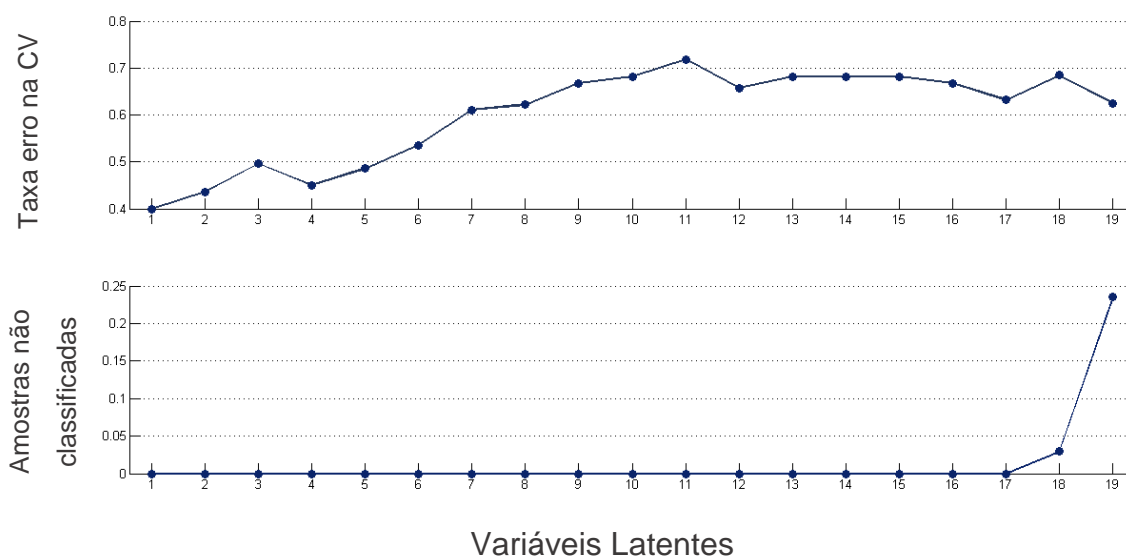


Figura 36. Variação do Erro com as Variáveis Latentes

Por ser um conjunto amostral diverso, mas com poucos representantes em seus subconjuntos, a validação escolhida foi do tipo *leave-one-out*, ou seja, que reveza de amostra a amostra a cada ciclo durante essa validação interna, até testar o conjunto como todo ^{116,165}. Então, a habilidade de classificação do modelo é avaliada com base nas métricas de eficiência desse processo como sensibilidade, especificidade, média geométrica (relação entre sensibilidade e especificidade), coeficiente de correlação de Matthew, entre outros.

a) Métricas de desempenho geral:

Para facilitar a visualização e comparação entre modelos, cada uma das métricas foi representada graficamente. Como o conjunto de dados foi trabalhado sob a consideração da existência de duas classes, a sensibilidade da classe Campanha pode ser interpretada como a especificidade da classe Serra Gaúcha, e vice e versa.

O gráfico representado na Figura 37, demonstra a sensibilidade, ou a habilidade de identificar os elementos de uma dada classe, frente ao número total de exemplares analisados dessa mesma classe. Por considerar essas métricas como variáveis, a média geométrica de cada modelo também foi representada.

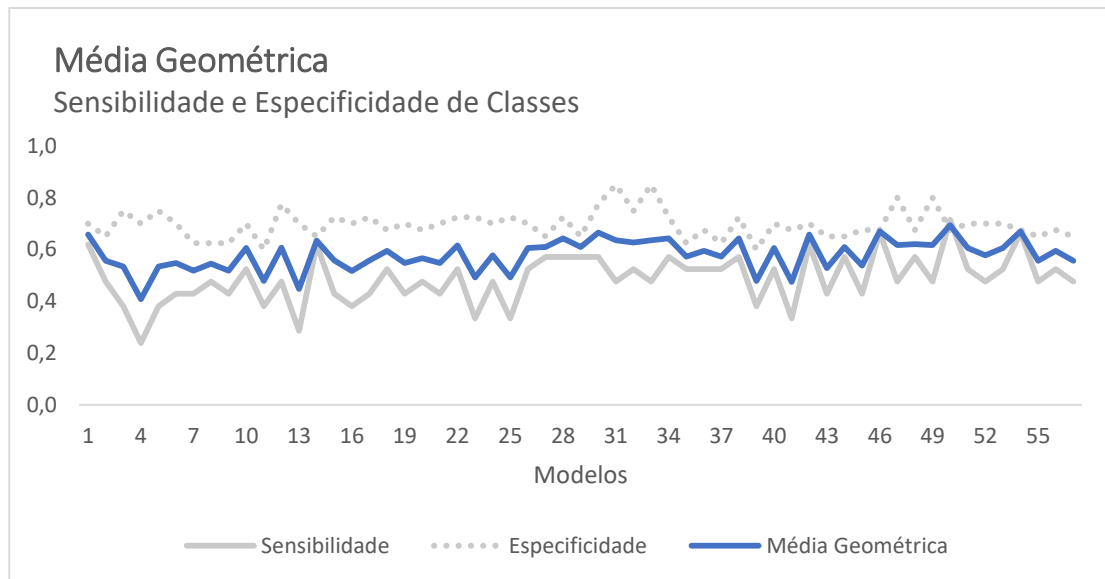


Figura 37. Média Geométrica do conjunto de modelos PLS – DA

Na inspeção do comportamento da sensibilidade é possível perceber um dos efeitos da maior proporção de amostras da Serra. Salvo algumas exceções, na validação cruzada a maioria dos modelos apresentou uma melhor sensibilidade para a Serra Gaúcha do que para a Campanha. Ou seja, os modelos testados apresentaram maior capacidade de discernimento das amostras consideradas como da Serra Gaúcha.

Ao combinar as métricas Sensibilidade e Especificidade de cada classe, pode-se avaliar a eficiência dos modelos em termos da Média Geométrica. Como o estudo envolve uma classificação dicotômica, a média geométrica (Gm_g) é equivalente para ambas classes. Refletindo o comportamento da sensibilidade da classe Campanha, a variação Gm_g também é crescente.

Outra métrica que demonstra os efeitos da desproporção amostral entre classes é a precisão. Diferentemente da sensibilidade, que considera a relação de amostras corretamente classificadas e a totalidade das amostras dessa classe envolvidas no estudo, a precisão esboça a pureza da classe predita, dependente da capacidade do modelo em termos de acerto em predição. Os modelos via PLS – DA (Figura 38) apresentaram precisão maior nas amostras da Serra Gaúcha do que da Campanha.

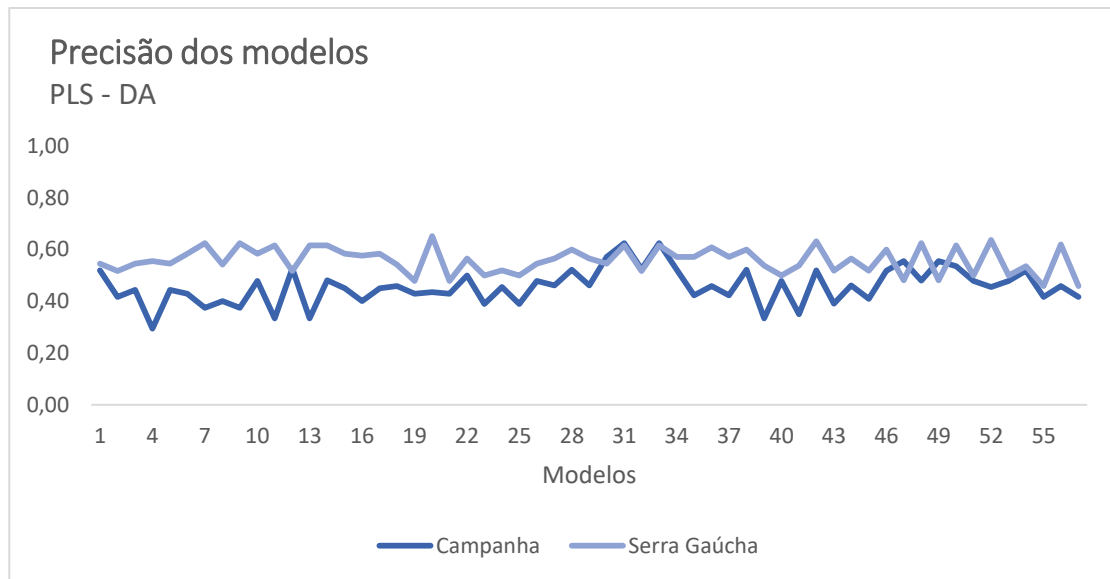


Figura 38. Precisão do conjunto de modelos PLS – DA

A frequência de falsos positivos também contribuiu para a definição do melhor modelo através da Taxa de falso alarme. Enquanto a Campanha apresenta um comportamento contínuo com média de 0,26, com variação de 0,15 a 0,38, a taxa de falsos positivos decresce na faixa de modelos estudados na Serra, com mínimo e máximo respectivamente de 0,29 e 0,67. Pela **Figura 51** é feita a comparação de desempenhos desse com outro método de reconhecimento de padrões.

Além da média geométrica (Gm), outra métrica secundária avaliada foi a área abaixo da curva intitulada Característica de Operação do Receptor (*AUC – ROC*). Por si só, a curva relaciona a taxa de falsos positivos com a sensibilidade do modelo, já a AUC serve de parâmetro para a capacidade de distinção entre as classes. Assim, para classificações dicotômicas cada modelo possui um valor de AUC. O desempenho via AUC pode ser visto na Figura 39.

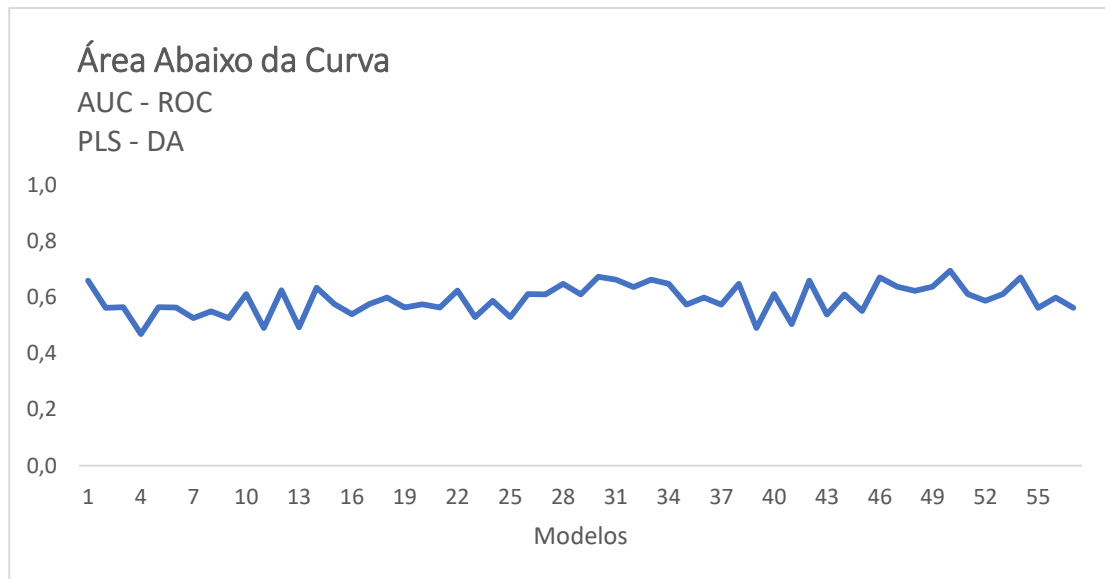


Figura 39. Área abaixo da Curva para os modelos PLS – DA.

A Exatidão e o coeficiente de correlação de Matthew (MCC) também refletem um comportamento global da classificação. Ambas expressões ilustram a taxa de corretas classificações e as ocorrências de falsos positivos e falsos negativos. Matematicamente, a exatidão varia na faixa de zero a um, o que sugere a existência de modelos com classificação randômica na inspeção feita, visto que o mínimo de exatidão encontrado é próximo de 0,5 (Figura 40). Contudo, alguns dos modelos PLS-DA atingem exatidão de 0,75, valores interessantes para seleção do modelo e aplicação do conjunto de dados teste.

O MCC, representado pela Figura 41, pode variar de valores entre -1 e 1. Com os valores deslocados para a região positiva da amplitude possível, é possível afirmar que uma maior parcela de modelos é capaz de obter classificações corretas e que sobressaiam às incorretas. Merecem atenção, todavia, os modelos cujo MCC foi próximo de zero, demonstrando que a classificação foi randômica nessas condições.

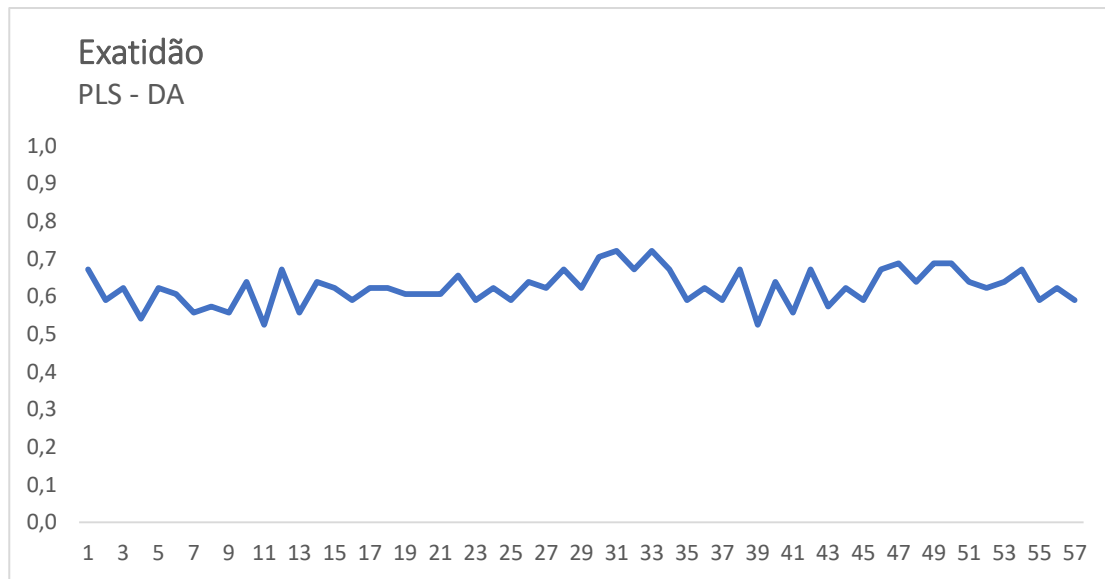


Figura 40. Exatidão dos modelos PLS – DA

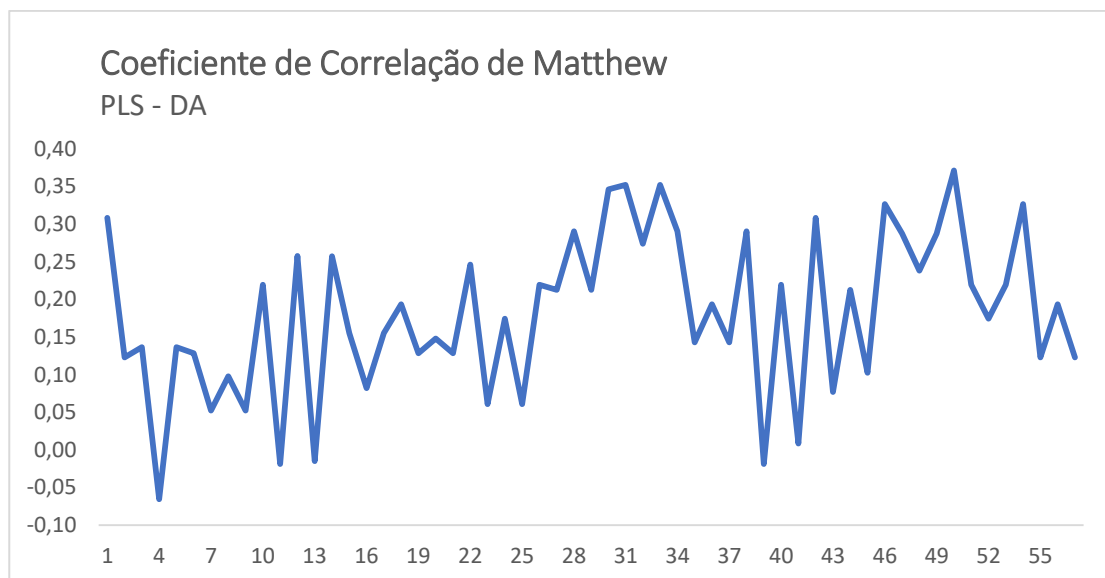


Figura 41. MCC dos modelos PLS – DA

b) Escolha do melhor modelo:

Visualmente, o comportamento dos modelos via PLS – DA sugere alguns pontos à aplicação na etapa posterior, com o conjunto amostral de teste. As métricas secundárias, como média geométrica, MCC e AUC reforçam a escolha do modelo com melhor desempenho através da visualização de seus picos. Todavia, apesar de majoritariamente apresentar comportamentos antagônicos, a sensibilidade e especificidade apontam modelos com desempenho relativamente de destaque.

Dentro desse conjunto de modelos está o identificado como **X50**, cuja diferença entre Sensibilidade e Especificidade da classificação da Campanha é mínima. Assim, a capacidade de classificação para as duas classes envolvidas é similar, ou seja, é um dos

melhores modelos estudados e capaz de diferenciar as amostras da classe minoritária quanto da majoritária.

Comparativamente, a Tabela 11 apresenta alguns dos modelos com melhor desempenho na etapa de validação cruzada do reconhecimento de padrões via PLS-DA. Cabe ressaltar que o primeiro modelo (X01), diferentemente dos demais ali representados, não passou sob pré-tratamentos.

Tabela 11. Comparação de desempenhos de modelos PLS – DA.

| MODELO | X01 | | X31 | | X42 | | X50 | |
|----------------|--|-----------|------------|-----------|----------------|-----------|------------|-----------|
| MÉTRICA | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> |
| GM | 0,70 | | 0,64 | | 0,72 | | 0,69 | |
| SEN | 0,62 | 0,80 | 0,48 | 0,85 | 0,62 | 0,82 | 0,71 | 0,68 |
| SPE | 0,80 | 0,62 | 0,85 | 0,48 | 0,82 | 0,62 | 0,68 | 0,71 |
| PRE | 0,62 | 0,80 | 0,62 | 0,76 | 0,65 | 0,80 | 0,54 | 0,82 |
| EXA | 0,74 | | 0,72 | | 0,75 | | 0,69 | |
| MCC | 0,42 | | 0,35 | | 0,45 | | 0,37 | |
| TFP | 0,20 | 0,38 | 0,15 | 0,52 | 0,18 | 0,38 | 0,32 | 0,29 |
| SEN | Sensibilidade | | SPE | | Especificidade | | | |
| PRE | Precisão | | EXA | | Exatidão | | | |
| TFP | Taxa de falso positivo | | | | | | | |
| GM | Média Geométrica <i>SEN</i> e <i>SPE</i> | | | | | | | |
| C1 | Campanha | | C2 | | Serra Gaúcha | | | |

c) Influência das variáveis químicas no resultado

A Figura 42 está dividida em duas partes principais, cada qual com três subunidades, uma representando a média de classes e duas com os espectros de cada classe centrados na média. Os três primeiros espectros são a representação dos dados sem pré-tratamentos, enquanto os três espectros da base da figura representam os dados após tratamento por derivada de Savitzky-Golay (X50). É perceptível o efeito das correções de linha de base, reduzindo ruído ao longo do espectro e facilitando a inspeção visual para identificação das regiões do espectro que mais influenciam na diferenciação das classes.

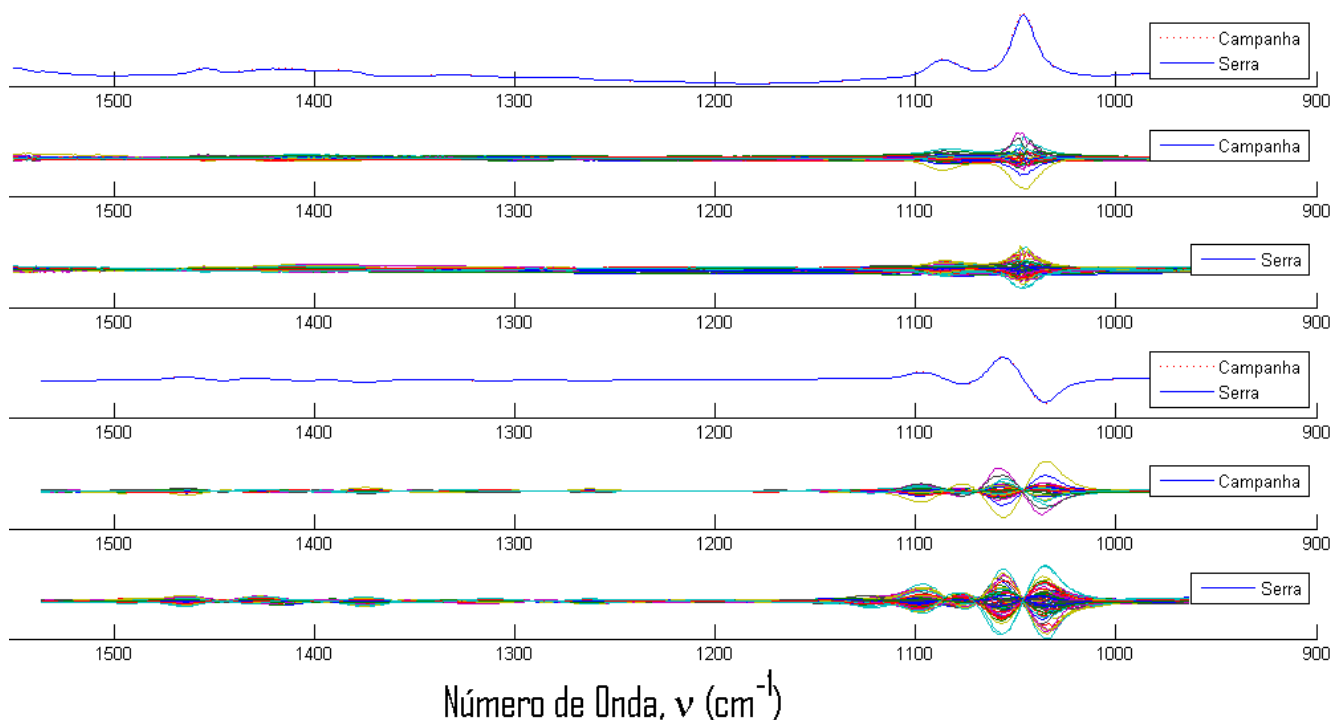


Figura 42. Dados Brutos

A otimização do número de fatores (variáveis latentes) apontou um menor erro de validação cruzada quando inclusos 16 fatores. Considerando um limite de 20 fatores como um limite a se evitar devido ao sobreajuste, o número 16 demonstra uma quantidade, apesar de adequada, elevada de combinações lineares necessárias para explicar o sistema. Em outras palavras, um reduzido erro de validação cruzada na classificação via PLS – DA baseado em 16 fatores representam um conjunto amostral complexo e multivariado. Todavia, a primeira variável latente se mostrou responsável por explicar mais de 90% da variância amostral.

O gráfico de *loadings* (Figura 43) apresenta uma estrutura similar ao espectro do qual foi originado. Mesmo assim, é possível deduzir quais os números de onda foram críticos para o agrupamento das classes. Majoritariamente, o gráfico de *loadings* aponta a influência das bandas com números de onda entre 1132 a 1016 cm^{-1} , região do espectro que compreende o alongamento C – O de álcoois primários. Mais ainda, regiões do gráfico (1478 a 1360 cm^{-1}) demonstram a influência, mesmo que mínima, de vibrações específicas dos polifenóis presentes no vinho¹⁶⁶⁻¹⁶⁸.

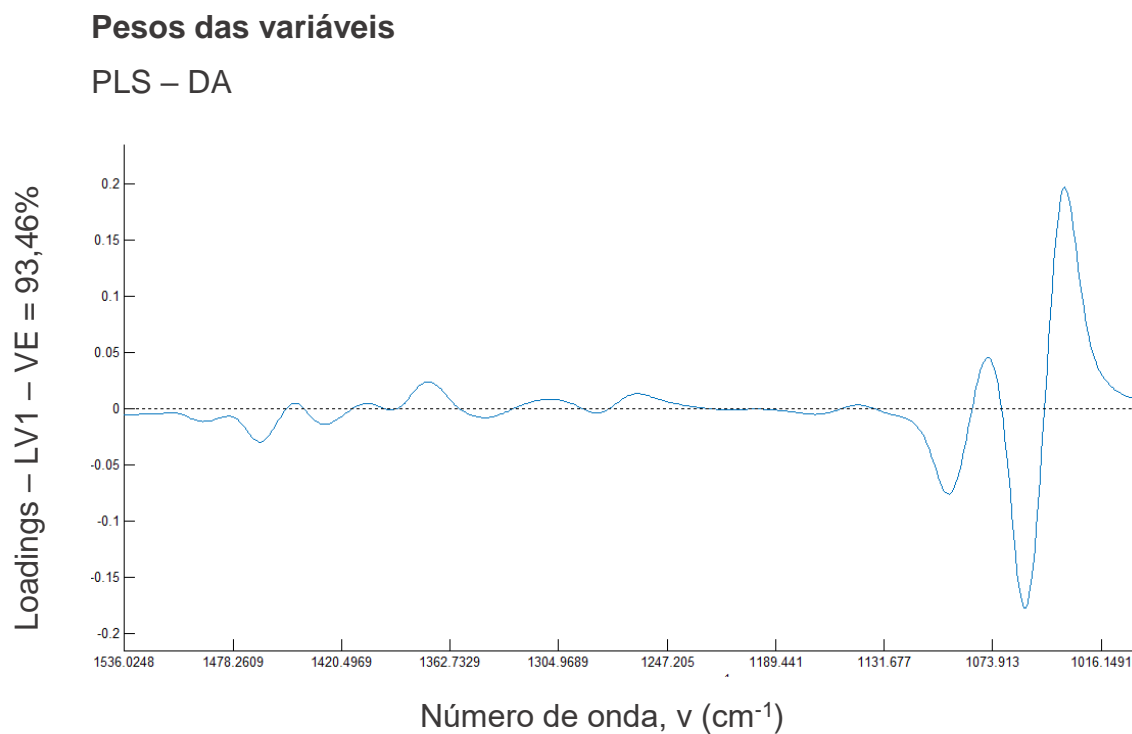


Figura 43. Variáveis do espectro significativas para a separação de classes

| Regiões significativas | |
|------------------------------------|--|
| cm^{-1} | <i>Bandas</i> |
| 1680 – 900 | Fenóis presentes em vinhos. |
| 1448 – 1444 | Deformação angular assimétrica no plano de $-\text{CH}_3$. |
| 1376 – 1373 | Deformação de $-\text{CH}_2$ – característica de núcleo fenólico |
| 1340 – 1339 | Deformação angular simétrica no plano de $-\text{CH}_3$. |
| 1340 – 1160 | Balço de CH_2 e Deformação angular de $\text{C} - \text{H}$. |
| 1285 | Deformação axial em $\text{C} - \text{OH}$ de fenóis |
| 1281 – 1278 | Deformação axial do anel pirano de flavonoide. |
| 1207 | Deformação angular no plano de $\text{O} - \text{H}$. |
| 1110 – 1107 | |
| 1068 – 1062 | Deformação axial $\text{C} - \text{O}$ |
| 1050 | Deformação axial $\text{C} - \text{O}$ de álcool primário. |

d) Desempenho desse modelo frente ao grupo Teste

Devido a sua relativamente boa sensibilidade para classificar corretamente ambas classes, o modelo PLS – DA definido como ponto 50 foi aplicado ao conjunto de amostras destinadas para teste. Cabe lembrar que o modelo aqui chamado de ponto 50 é resultado do tratamento do conjunto de dados brutos por derivada primeira Savitzky Golay, janela de 31 pontos e adequação da curva a polinômio de segundo grau. Também, o grupo de teste é a terça parte resultante da partição Kennard Stone, aplicado aos espectros pré-tratados das 92 amostras em estudo.

Tabela 12. Desempenho de modelos selecionados na etapa de predição via PLS – DA

| MODELO | X01 | | X31 | | X42 | | X50 | |
|----------------|--|-----------|------------|-----------|----------------|-----------|------------|-----------|
| MÉTRICA | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> |
| GM | 0,66 | | 0,57 | | 0,57 | | 0,72 | |
| SEN | 0,54 | 0,80 | 0,36 | 0,90 | 0,36 | 0,90 | 0,55 | 0,95 |
| SPE | 0,80 | 0,54 | 0,90 | 0,36 | 0,90 | 0,36 | 0,95 | 0,55 |
| PRE | 0,60 | 0,76 | 0,67 | 0,72 | 0,67 | 0,72 | 0,86 | 0,79 |
| EXA | 0,71 | | 0,71 | | 0,71 | | 0,81 | |
| MCC | 0,35 | | 0,32 | | 0,32 | | 0,57 | |
| TFP | 0,20 | 0,45 | 0,10 | 0,64 | 0,10 | 0,64 | 0,05 | 0,45 |
| SEN | Sensibilidade | | SPE | | Especificidade | | | |
| PRE | Precisão | | EXA | | Exatidão | | | |
| TFP | Taxa de falso positivo | | | | | | | |
| GM | Média Geométrica <i>SEN</i> e <i>SPE</i> | | | | | | | |
| C1 | Campanha | | C2 | | Serra Gaúcha | | | |

As métricas de desempenho do modelo selecionado (e de outros para comparação) estão dispostas na Tabela 12. Como esperado, pela tendência apresentada no conjunto das respostas da validação cruzada, a sensibilidade para predição das amostras oriundas da Serra Gaúcha é maior que da Campanha, 0,95 e 0,55, respectivamente. Ou seja, o modelo apresentou um número menor de falsos negativos para a Serra Gaúcha quando comparado com a Campanha.

Por considerar alocação dicotômica, e sem a possibilidade de não designar alguma amostra a quaisquer das classes envolvidas, alterações na Sensibilidade em uma classe em particular são observadas na Precisão da outra envolvida. Isso se deve ao

número de amostras não classificadas corretamente como Campanha (Falsos Negativos), que interferem na pureza da Serra Gaúcha (Falsos Positivos).

No modelo 50, por exemplo, a Precisão da Serra é de 0,79, enquanto a Sensibilidade da Campanha é de 0,55. Em contrapartida, a precisão de 0,86 da Campanha demonstra, também, a habilidade do modelo de definir as amostras da Serra, e consequentemente não incrementar o número de falsos positivos na Campanha.

Das 31 amostras designadas para teste, cerca de 81% dessas foram classificadas corretamente. Como ilustrado na Figura 44, o modelo 50 PLS – DA classificou corretamente 95% das amostras teste da Serra Gaúcha, uma vez que apenas uma (CS79) foi incorretamente reconhecida como integrante da outra classe. As amostras da Campanha separadas para teste, por sua vez, foram classificadas corretamente a uma taxa de 55%.

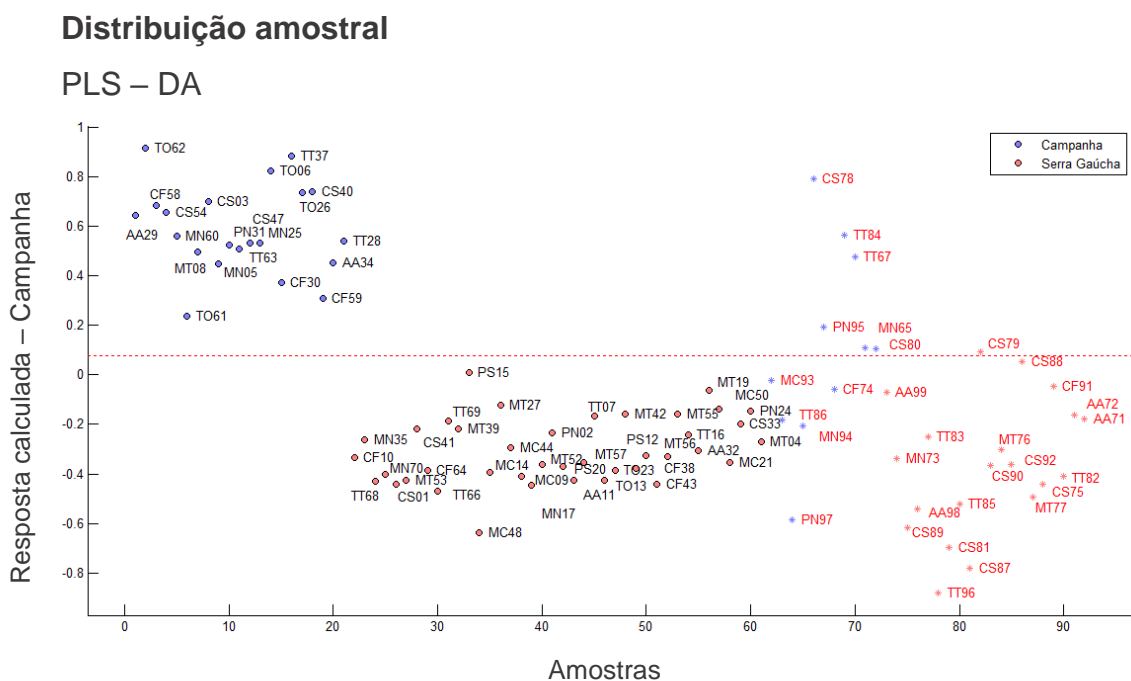


Figura 44. Distribuição das amostras de treino e teste via PLS – DA

Pela diferença em proporção de erros de classificação, além das variedades de uvas envolvidas nesses erros, é possível afirmar mais uma vez que a quantidade de amostras em ambas classes é um fator importante para a ocorrência de desvios. O algoritmo do Kennard – Stone não considera subclasses envolvidas na execução da partição, sendo assim, um conjunto amostral como o da Serra Gaúcha apresentou robusto desempenho devido ao volume amostral, que ocasionalmente é representativo mesmo após partição.

Nesse sentido, a Figura 44 pode reforçar esse argumento quando analisadas as amostras classificadas incorretamente. Como um todo, a Campanha possui uma amostra de varietal Malbec, que eventualmente foi alocada no grupo de teste após a partição por KS. Não surpreendente, a classificação por PLS-DA foi incorreta, uma vez que a Serra Gaúcha possui seis (6) amostras de vinhos dessa uva e todas elas no conjunto de treino dessa classe.

Outro exemplo é a amostra de Cabernet Franc (CF74), pertencente a safra 2018 da Campanha, mas classificada como da Serra. Frente à duas das três amostras da Campanha, três das quatro amostras de CF foram produzidas na Serra durante a safra de 2018. Ou seja, o conjunto maior de amostras na Serra, da mesma uva e mesma safra, pode ter conduzido ao erro de classificação da amostra da originária da Campanha.

De modo similar ao ocorrido com o conjunto de Cabernet Franc, a amostra de Tannat (TT86, safra 2018) foi classificada incorretamente como pertencente à Serra Gaúcha. A Campanha contém um conjunto de treino composto por três amostras desse varietal, sendo duas da safra de 2017 e uma de 2018. O grupo de treino da Serra Gaúcha é composto, por sua vez, por duas amostras da safra de 2017 e três de 2018. Ou seja, uma vez que a safra desse varietal em 2018 é três vezes maior na Serra do que na Campanha, é possível que essa proporção amostral exerça influência no método de reconhecimento de padrão, de modo a, eventualmente, considerar a amostra TT86 como integrante da Serra Gaúcha.

As amostras do varietal Marselan são divididas quase que igualmente entre as duas classes regionais e safras de produção. Enquanto a Campanha apresenta um total de cinco amostras, a Serra Gaúcha tem quatro amostras. Com a partição, ambos conjuntos de treino foram compostos por três amostras (duas da safra 2017 e uma da safra 2018). Os conjuntos de teste resultantes não possuem amostras de 2017 e a Campanha apresentou uma amostra a mais que a Serra Gaúcha. Ainda assim, uma amostra da Campanha (MN94) fora reconhecida como amostra proveniente da Serra Gaúcha.

Por fim, as amostras do varietal Pinot Noir são divididas igualmente entre classes, porém não entre conjuntos de treino e teste. Enquanto a Serra Gaúcha possui três amostras no grupo de treino, a Campanha possui apenas uma. O conjunto de teste da Campanha contém duas amostras, dentre essas a que fora incorretamente classificada como da Serra Gaúcha (PN97). Possivelmente, esse erro de classificação é resultado da desproporção de amostras entre grupos de treino e teste, que considerou mais modelos de uma classe que de outra.

SPA – LDA

a) Otimização do tamanho da cadeia da fase 1 do SPA

Diferentemente da otimização realizada no PLS-DA, onde cada conjunto de dados passou pelo processo de escolha do número de fatores com menor erro de validação cruzada, aqui foi escolhida uma condição de compromisso comum para todos os modelos. Considerando as duas classes envolvidas no estudo e a procura pelo uso reduzido de variáveis que expliquem o sistema, a condição de compromisso escolhida foi a de cadeias de variáveis de menor comprimento igual a 2, (este número leva em conta os graus de liberdade do modelo LDA) e máximo igual a 10 variáveis, de modo a evitar problemas de sobre-ajuste e assegurar modelos parcimoniosos.

b) Métricas de desempenho geral

O Algoritmo de Sucessivas Projeções aplicado aos conjuntos de dados, brutos e pré-processados, apresentou resultados promissores, ainda que similares aos do PLS – DA. Quando relacionado o desempenho das classes envolvidas, a Serra Gaúcha mais uma vez apresenta médias de sensibilidade e precisão maiores que a Campanha.

A sensibilidade da Campanha varia entre 0,52 e 0,76, ao passo que os dados da Serra a variação é de 0,65 e 0,80 (Figura 45). A Precisão, por sua vez, varia de 0,46 a 0,65 na Campanha e 0,73 a 0,86 na Serra (**Figura 46**). Novamente, a diferença de desempenho entre classes pode ser explicada pela diferença de composição amostral, seja por quantidade ou pluralidade.

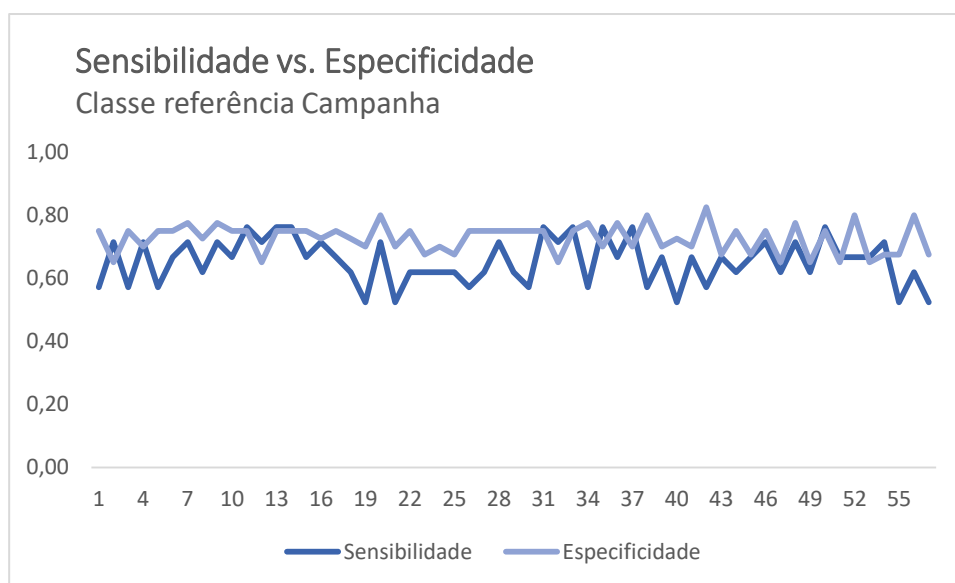


Figura 45. Relação Sensibilidade e Especificidade dos modelos SPA – LDA

Ao comparar o desempenho dos métodos PLS – DA e SPA – LDA, registrados pelas métricas Sensibilidade e Precisão, não se observa diferenças significativas na classificação das amostras da Serra Gaúcha. Através do SPA – LDA, todavia, o reconhecimento das amostras da Campanha foi mais eficaz, haja vista a elevação dos pontos mínimos nessa avaliação geral de Sensibilidade e Precisão na classificação após validação cruzada.

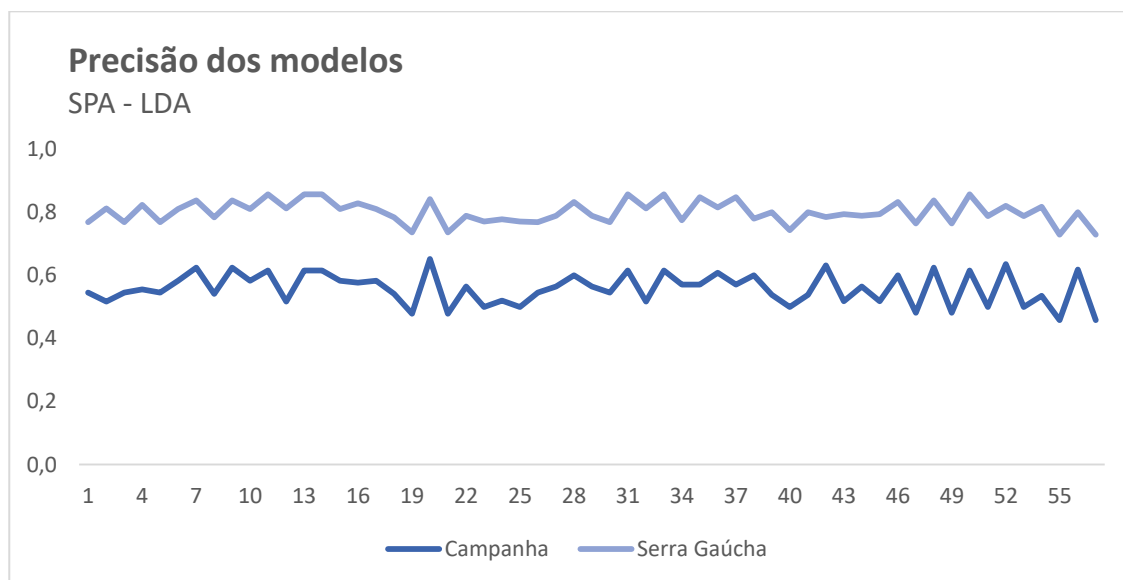


Figura 46. Precisão do conjunto de modelos SPA – LDA

Essa elevação nos mínimos e máximos nas métricas de desempenho do método SPA – LDA também se faz percebida pela Média Geométrica (Figura 47), AUC (Figura 48), Exatidão e MCC. Em ambas Gm e AUC, seus mínimos passaram respectivamente de 0,48 e 0,52 para 0,60. Já seus máximos saíram de 0,72 para 0,76. Em termos gerais, é possível dizer que houve uma redução na diferença entre Sensibilidade e Especificidade de ambas classes em análise.

A redução na diferença entre sensibilidade e especificidade demonstra uma queda no número de classificações incorretas. O MCC e a exatidão, que consideram fatores como os falsos positivos e negativos, tiveram médias mais elevadas e limites de máximo e mínimo mais estreitos. O ponto de mínimo do MCC (Figura 49) passou de 0,03 para 0,20, e o máximo de 0,45 para 0,50. A exatidão mínima saiu de 0,57 para 0,62, enquanto que a máxima passou de 0,75 para 0,77. Assim, considerando a relação matemática de cada uma das métricas, pode-se dizer que há, ao menos, uma queda no número de falsos negativos.

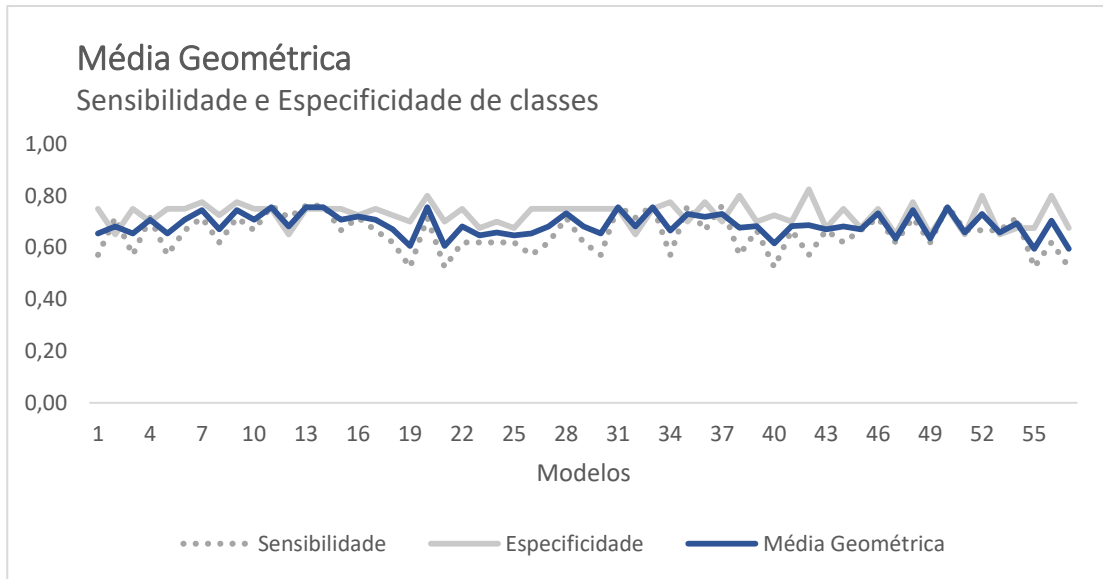


Figura 47. Média Geométrica do conjunto de modelos SPA – LDA

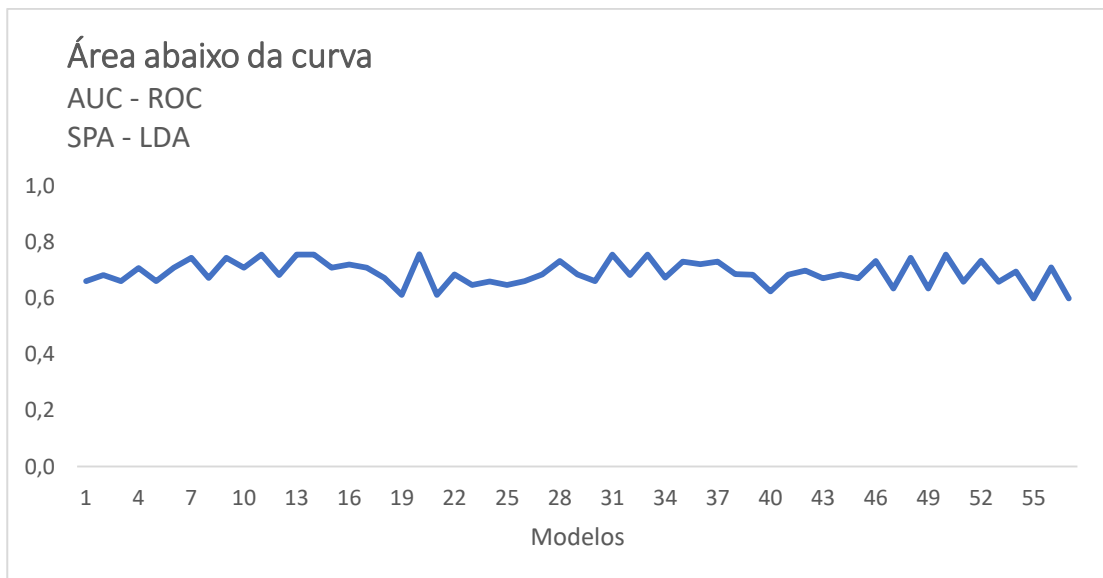


Figura 48. Área abaixo da Curva para os modelos SPA – LDA

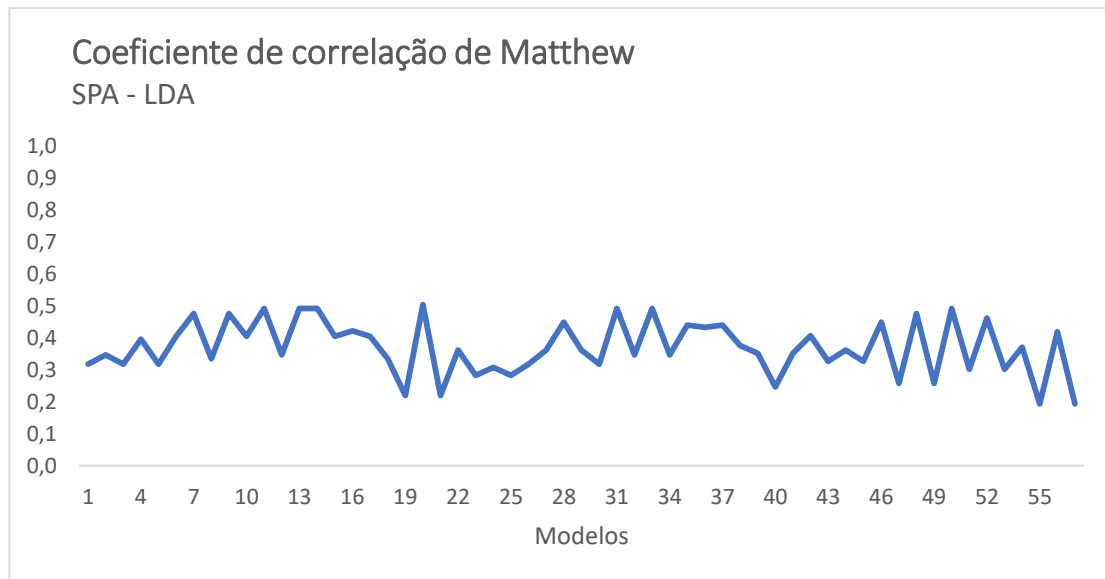


Figura 49. MCC para conjunto de modelos SPA – LDA

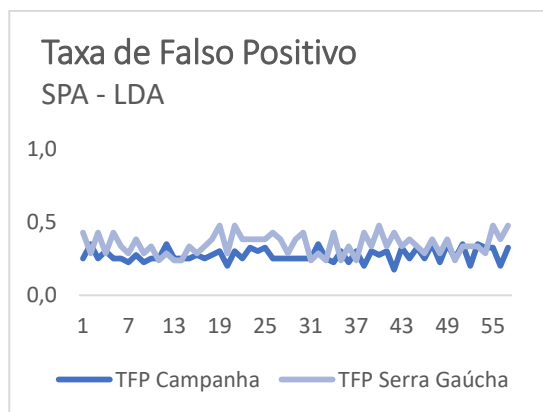


Figura 50. Taxa de Falso Positivo nos modelos SPA – LDA

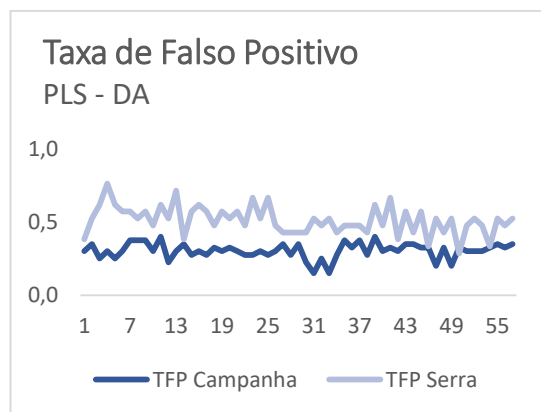


Figura 51. Taxa de Falso Positivo nos modelos PLS – DA

A taxa de falsos positivos, por sua vez, ficou mais estreita na sua diferença entre mínimos e máximos, como pode ser visto nas Figura 50 e Figura 51. A Campanha reduziu seu ponto máximo na taxa de falsos positivos de 0,38 para 0,35, mas elevou seu ponto mínimo de 0,15 para 0,20. A Serra Gaúcha teve seus pontos mínimo e máximo reduzidos de 0,29 para 0,24 e de 0,67 para 0,48, respectivamente. A reduzida ocorrência de falsos positivos na Serra Gaúcha reflete um melhor desempenho nos modelos em classificar corretamente as amostras da Campanha.

Assim como realizado na PLS – DA, a avaliação e escolha do modelo com melhor desempenho no método SPA – LDA precisa ponderar as métricas como um todo, mas principalmente priorizando a máxima sensibilidade da Campanha, sem significativo detrimento de sua especificidade. Nesse sentido, o modelo escolhido é o referente ao

ponto 31. Esse modelo apresenta uma das maiores sensibilidades e especificidades da Campanha, além de mínimas taxas de falso positivo similares para as duas classes.

O modelo 31 foi escolhido por ter apresentado melhor desempenho nas métricas discutidas acima e em consideração ao princípio da parcimônia, tendo em vista a similaridade nas métricas dos modelos 33 e 50. A título de comparação, todavia, estão dispostas métricas de outros modelos SPA – LDA na Tabela 13.

Tabela 13. Desempenho de modelos selecionados na etapa de predição via SPALDA.

| MODELO | X01 | | X31 | | X42 | | X50 | |
|----------------|--|-----------|------------|-----------|----------------|-----------|------------|-----------|
| MÉTRICA | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> |
| GM | 0,70 | | 0,76 | | 0,73 | | 0,76 | |
| SEN | 0,67 | 0,72 | 0,76 | 0,75 | 0,71 | 0,75 | 0,76 | 0,75 |
| SPE | 0,72 | 0,67 | 0,75 | 0,76 | 0,75 | 0,71 | 0,75 | 0,76 |
| PRE | 0,56 | 0,81 | 0,62 | 0,86 | 0,60 | 0,83 | 0,62 | 0,86 |
| EXA | 0,70 | | 0,75 | | 0,74 | | 0,75 | |
| MCC | 0,38 | | 0,49 | | 0,45 | | 0,49 | |
| TFP | 0,28 | 0,33 | 0,25 | 0,24 | 0,25 | 0,29 | 0,25 | 0,24 |
| SEN | Sensibilidade | | SPE | | Especificidade | | | |
| PRE | Precisão | | EXA | | Exatidão | | | |
| TFP | Taxa de falso positivo | | | | | | | |
| GM | Média Geométrica <i>SEN</i> e <i>SPE</i> | | | | | | | |
| C1 | Campanha | | C2 | | Serra Gaúcha | | | |

c) Influência das variáveis químicas no resultado

Diferentemente do visualizado no modelo X50 (discutido no PLS – DA), ao tratar o conjunto de dados brutos com segunda derivada SG, como no pré-tratamento X31, o espectro assume curvatura tal qual apresenta a Figura 52. Ainda, a rotina do algoritmo SPA – LDA fornece o perfil médio de todas as amostras envolvidas no processo de calibração do modelo, bem como as variáveis de maior influência no reconhecimento de padrões.

As variáveis selecionadas pelo SPA – LDA são, de forma similar ao resultado apresentado pelo PLS – DA, relacionáveis à presença de compostos polifenólicos e álcool primário. A saber, o modelo decorrente dos dados X31 destacou os números de onda a seguir: 1317, 1045 e 1040 (cm⁻¹).

O conjunto de dados utilizado como treinamento e calibração do modelo foi o chamado X31. É importante lembrar que esse conjunto de dados passou pelo pré-tratamento de suavização de ruídos e correção da linha de base, Savitzky – Golay de janela de 21 pontos, com adequação do polinômio e derivada de segundo grau. O desempenho do modelo na predição das amostras do grupo Teste é comparado, na Tabela 14 a outros modelos de SPA – LDA.

Variáveis selecionadas e Perfil Médio

SPA – LDA

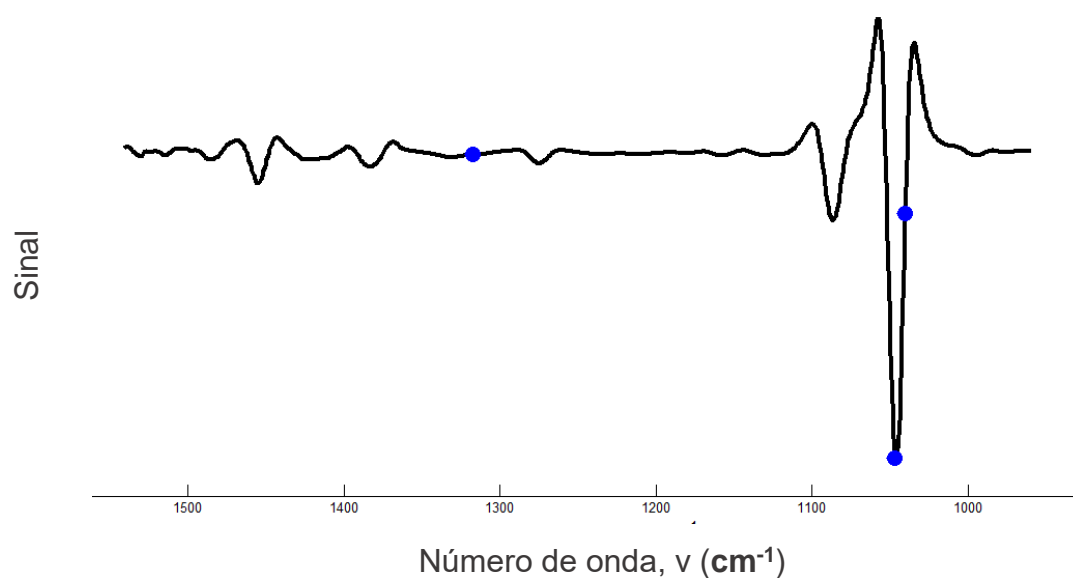


Figura 52. Espectro no Infravermelho após SG com 2ª Derivada

| Regiões significativas | |
|------------------------|---|
| cm^{-1} | Bandas |
| 1340 – 1339 | Balanço de CH_2 e Deformação angular de C – H. |
| 1340 – 1160 | Deformação axial em C – OH de fenóis |
| 1110 – 1107 | Deformação axial C – O |
| 1068 – 1062 | |
| 1050 | Deformação axial C – O de álcool primário. |

Tabela 14. Desempenho de modelos SPA – LDA selecionados para teste.

| MODELO | X01 | | X31 | | X42 | | X50 | |
|----------------|--|-----------|------------|-----------|----------------|-----------|------------|-----------|
| MÉTRICA | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> | <i>C1</i> | <i>C2</i> |
| GM | 0,37 | | 0,66 | | 0,68 | | 0,62 | |
| SEN | 0,18 | 0,75 | 0,55 | 0,80 | 0,55 | 0,85 | 0,45 | 0,85 |
| SPE | 0,75 | 0,18 | 0,80 | 0,55 | 0,85 | 0,55 | 0,85 | 0,45 |
| PRE | 0,29 | 0,62 | 0,60 | 0,76 | 0,67 | 0,77 | 0,62 | 0,74 |
| EXA | 0,55 | | 0,71 | | 0,74 | | 0,71 | |
| MCC | -0,08 | | 0,35 | | 0,42 | | 0,33 | |
| TFP | 0,25 | 0,82 | 0,20 | 0,45 | 0,15 | 0,45 | 0,15 | 0,55 |
| SEN | Sensibilidade | | SPE | | Especificidade | | | |
| PRE | Precisão | | EXA | | Exatidão | | | |
| TFP | Taxa de falso positivo | | | | | | | |
| GM | Média Geométrica <i>SEN</i> e <i>SPE</i> | | | | | | | |
| C1 | Campanha | | C2 | | Serra Gaúcha | | | |

Frente a classificação do grupo teste, o modelo 31 SPA – LDA interpretou incorretamente como provenientes da Serra Gaúcha as amostras CS78, MN65, TT67 e TT86. Também, o modelo incorretamente classificou como da Campanha as amostras CS90, MT76, TT82 e TT83. Na Validação Cruzada da Serra Gaúcha, 14 amostras não foram corretamente classificadas: AA32, CF64, CS33, MC48, MN70, MT19, MT55, PS15, PS20, PN24, TO23, TT07, TT16 e TT66. Ainda na etapa de Validação Cruzada, o modelo não classificou corretamente as amostras da Campanha CF59, MN05, MN60 e TO62. Esses resultados podem ser vistos na Figura 53.

Através da análise dos modelos testados, percebe-se que a escolha do pré-processamento X31, ainda que não tenha reproduzido o melhor, resultou em um dos melhores desempenhos. De modo geral, em comparação com os modelos testados ao aplicar PLS – DA, as métricas de desempenho apontam uma dificuldade de lidar com um conjunto amostral limitado e diverso. Tal fato pode ser percebido pela redução em boa parte das métricas expressas. A sensibilidade da classe majoritária, por exemplo, chegou a 0,95 no modelo X50 PLS – DA, enquanto que no X31 SPA – LDA foi de 0,80.

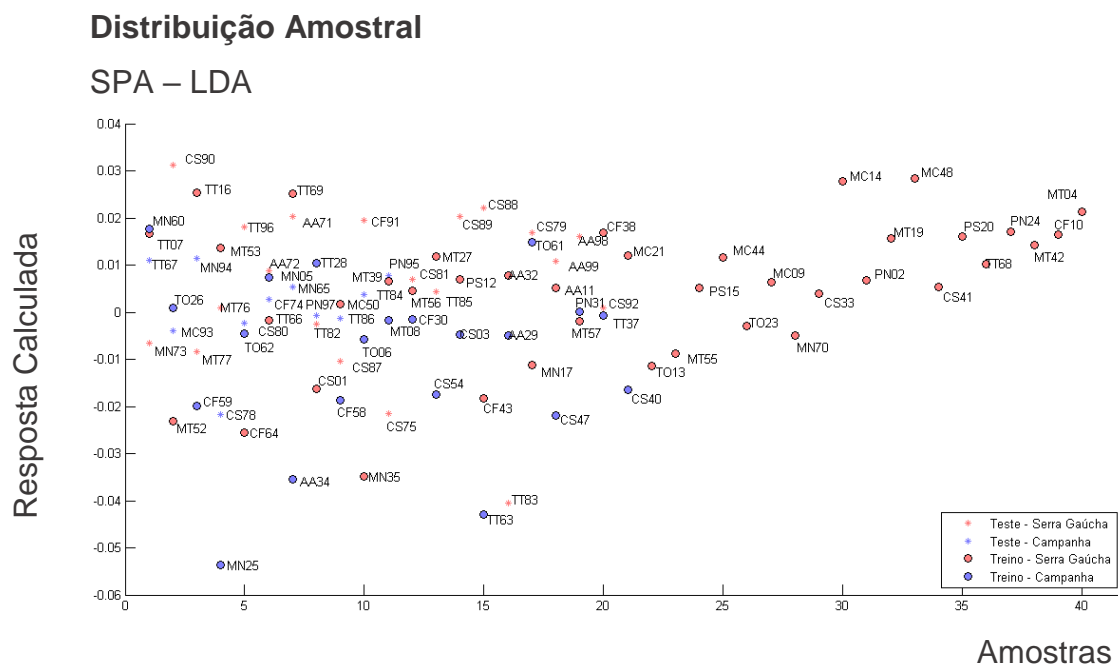


Figura 53. Distribuição de amostras de treino e teste segundo SPA – LDA

Ao comparar o desempenho dos métodos pelas métricas dos seus melhores modelos, percebe-se que a sensibilidade da Campanha permanece a mesma, 0,55. Por sua vez, o método SPA – LDA teve decréscimo na especificidade da Campanha, ou seja, a sensibilidade da Serra foi reduzida. Assim, com uma capacidade de classificação reduzida na Serra, o número de falsos negativos se eleva.

A elevação da quantidade de falsos negativos da Serra pode ser percebida na comparação entre Precisões de modelos. O modelo 50 PLS – DA apresentou para a Campanha uma precisão de 0,86, contra 0,60 no modelo 31 SPA – LDA. Ou seja, a elevação da quantidade de Falsos Negativos na Serra reduz a Precisão da Campanha, que entende esses valores como Falsos Positivos.

Essa queda em sensibilidade da Serra Gaúcha pode estar relacionada com a abordagem do modelo de reconhecimento de padrões. O PLS – DA utiliza toda região do espectro informada, organizando as variáveis originais em novas variáveis latentes, com pesos adequando a variável a sua influência na separação das classes. Por sua vez, ao selecionar variáveis de maior destaque, com o limite imposto pela condição de compromisso, o SPA – LDA se torna mais específico para a generalização da classe, logo excluindo amostras sem dada representatividade de origem, safra e varietal.

GA-LDA

a) Aplicação em conjuntos de dados específicos

A Análise Discriminante Linear com seleção de variáveis por algoritmo genético (GA – LDA) foi executada em conjuntos de dados específicos, isso devido ao elevado tempo médio em processamento por computador a que cada análise depende. Também, os conjuntos de dados selecionados para essa etapa são, basicamente, aqueles que tiveram desempenho destaque na Validação Cruzada de PLS – DA e SPA – LDA: a saber X15, X31, X38, X42 e X50.

b) Função aptidão na escolha do número de gerações e tamanho da população

A aptidão é a habilidade de adaptação ao espaço amostral disponível, adaptação essa dependente do tamanho da população e da quantidade de ciclos de cruzamento, ou gerações. A aptidão foi avaliada, então, em uma superfície de resposta, cujo ponto máximo forneceu as condições para execução do modelo de reconhecimento de padrões via método de Monte Carlo.

Do conjunto de dados X31, por exemplo, obteve-se a superfície representada na Figura 54. Pela inspeção da superfície se identificou um ponto de máxima aptidão (população 800 e gerações 500) e seus parâmetros foram aplicados na próxima etapa. Essa, que subentende a construção do modelo e ao teste, teve o número de ciclos como condição de compromisso, tendo em vista a limitação do tempo de processamento computacional a ser empregado. O número de ciclos é igual a 100 e foi o mesmo para todas as classificações estudadas via GA – LDA.

A título de comparação, a Figura 55 demonstra a superfície de aptidão resultante do GA – LDA aplicado ao conjunto de dados X50. Mais evidente que a situação anterior, a superfície apresenta um pico destacado na posição 800 e 1000 (gerações por tamanho de população), posição essa escolhida para execução do modelo pelo método de Monte Carlo.

Após finalização de todos os ciclos do algoritmo, a avaliação do desempenho ao classificar corretamente as amostras foi iniciada pela análise das variáveis selecionadas ao decorrer do método de Monte Carlo. Como esse método registra resultados numéricos de ordem estocástica, a quantidade de ciclos executados influencia a proporção das variáveis selecionadas. Pelo gráfico de barras, presente na Figura 56 e acompanhado do espectro tratado do modelo X31, é possível verificar os números de onda selecionados a cada ciclo, bem como a frequência dessa seleção ao longo de todos os ciclos.

Superfície de Aptidão do Algoritmo Genético – X31

Desempenho do algoritmo como função de População e Gerações

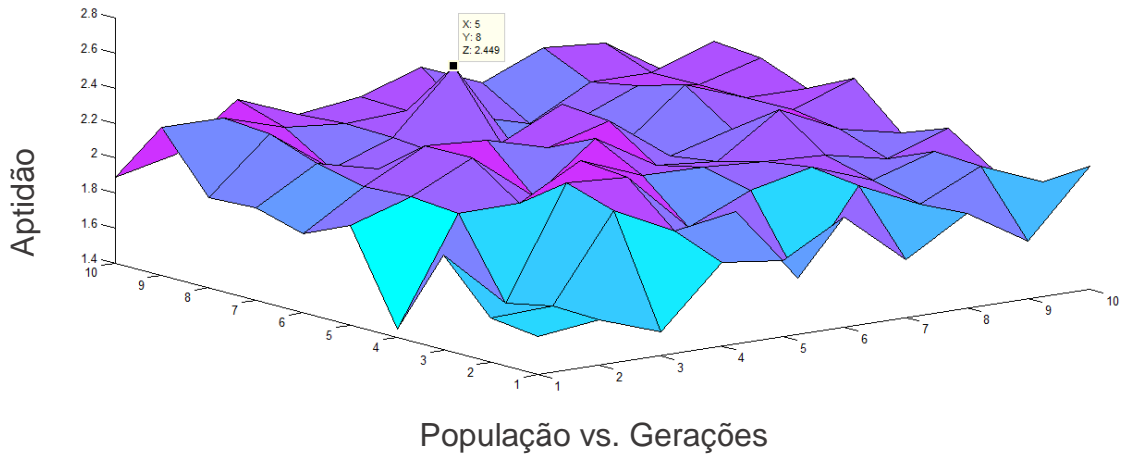


Figura 54. Aptidão do Algoritmo Genético aplicado aos dados X31

Superfície de Aptidão do Algoritmo Genético – X50

Desempenho do algoritmo como função de População e Gerações

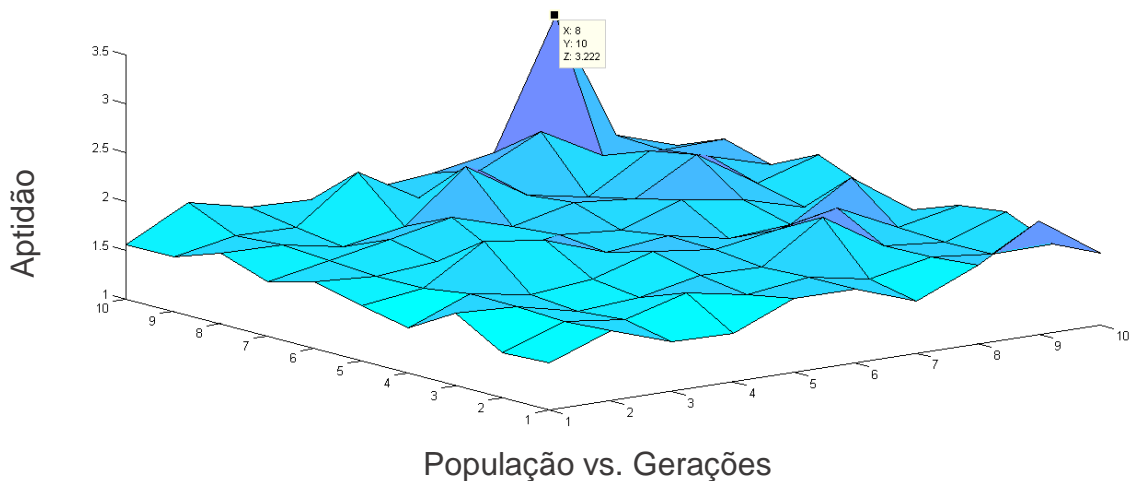


Figura 55. Aptidão do Algoritmo Genético aplicado aos dados X50

A frequência de ocorrência dos números de onda, apesar da aleatoriedade do método, representa a influência de dadas variáveis na construção de um modelo que, através do conjunto de dados de treinamento, atendesse o critério de classificação segundo sua origem. Ainda assim, partindo da variedade de frequências, um conjunto de números de onda de maior destaque pode ser filtrado ao se definir um limite de tolerância. Na Figura 56, essa tolerância é representada pela barra horizontal azul.

Ao elevar os valores de tolerância, restringe-se a classificação de acordo com as variáveis que ultrapassam essa barreira de frequências. Dessa maneira, é importante perceber que, quanto menor for a tolerância, mais sujeito ao sobreajuste estará o modelo.

Por outro lado, uma tolerância muito elevada também poderia reduzir o desempenho do modelo, haja vista a falta de informação útil para detecção de características relevantes à diferenciação entre classes.

Como parâmetro de desempenho da classificação, escolheu-se a sensibilidade e especificidade dos modelos. Os gráficos da Figura 56 representam a variação da sensibilidade (e especificidade) conforme o valor de tolerância. Enquanto o gráfico do topo representa esse comportamento no grupo de treino, o gráfico da base apresenta o desempenho na etapa de teste.

Na Figura 57 está representado o desempenho do modelo com X50. Em comparação com o modelo resultante do conjunto de dados X31, o comportamento da métrica em análise se mostrou semelhante, também, em todos outros testes, com pontos de máximo (demonstrados com círculo vermelho) e estabilizando a variação de sensibilidade após 0,625.

Não havendo uma regra clara na relação tolerância e sensibilidade nos conjuntos treino e teste, priorizou-se estabilidade das métricas em ambas etapas e definiu-se a tolerância como 0,625 (ponto 50). Na Tabela 15 está discriminado o desempenho fruto dessa escolha.

Mais ainda, a Tabela 15 apresenta o conjunto de variáveis significativas no reconhecimento de padrões via GA – LDA. Dadas bandas, apesar do pré-processamento aplicado, sinalizam a importante contribuição da diferença de concentração de álcool primário e polifenóis para a classificação das amostras segundo sua origem de procedência.

Tabela 15. Desempenho de Grupos de Teste e variáveis significativas

| TOLERÂNCIA FIXA 0,625 | | | | | | |
|------------------------------|------------------------------------|-------|------|-------|-----------------------|-------|
| ETAPA | Modelo | X15 | X31 | X38 | X42 | X50 |
| TESTE | Número de onda (cm ⁻¹) | 974,2 | 1309 | 972,4 | 972,4 1078 1300 | 974,2 |
| | SEN | 0,64 | 0,73 | 0,55 | 0,64 | 0,55 |
| | SPE | 0,85 | 0,85 | 0,50 | 0,85 | 0,65 |
| | PRE – 01 | 0,70 | 0,73 | 0,38 | 0,70 | 0,46 |
| | PRE – 02 | 0,81 | 0,85 | 0,67 | 0,81 | 0,72 |
| | Exatidão | 0,77 | 0,81 | 0,52 | 0,77 | 0,61 |

Comportamento Tolerância e Predição

Tolerância fixa em 0,250 – X31

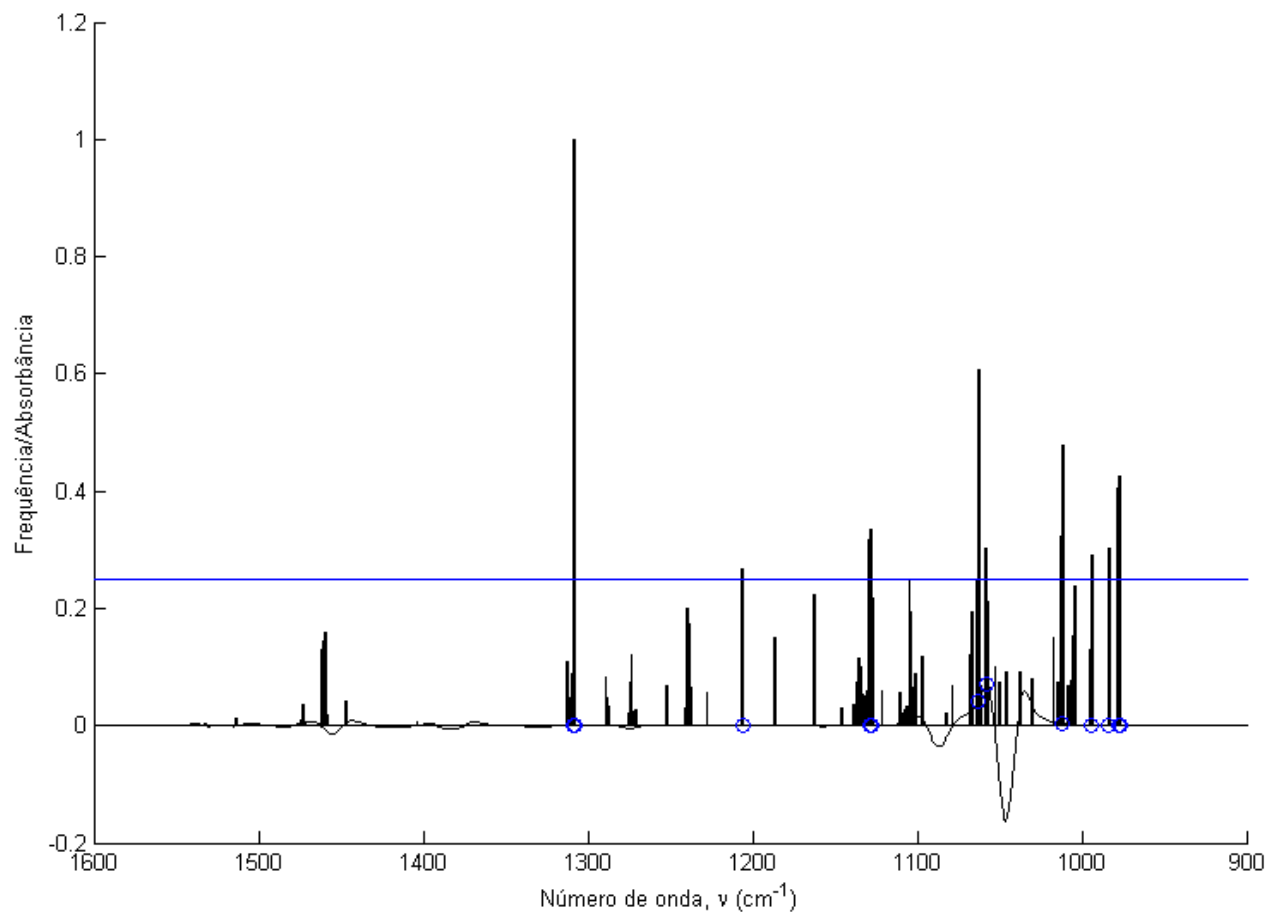
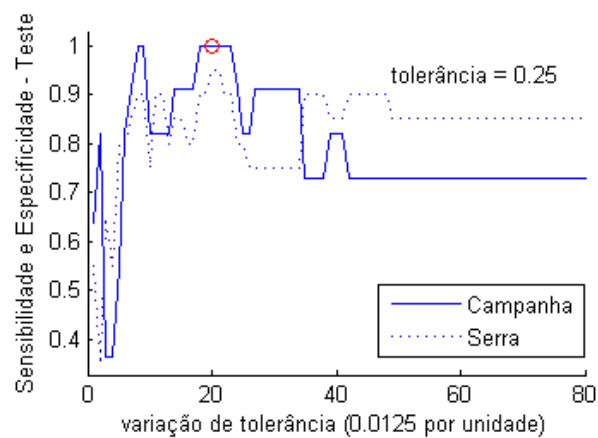
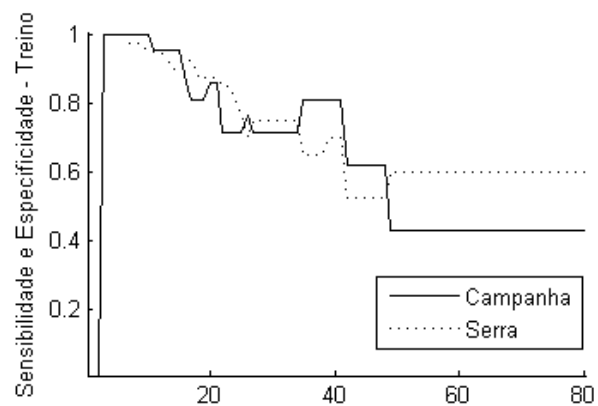


Figura 56. Frequência de variáveis selecionadas e espectro médio de X31

Comportamento Tolerância e Predição

Tolerância fixa em 0,325 – X50

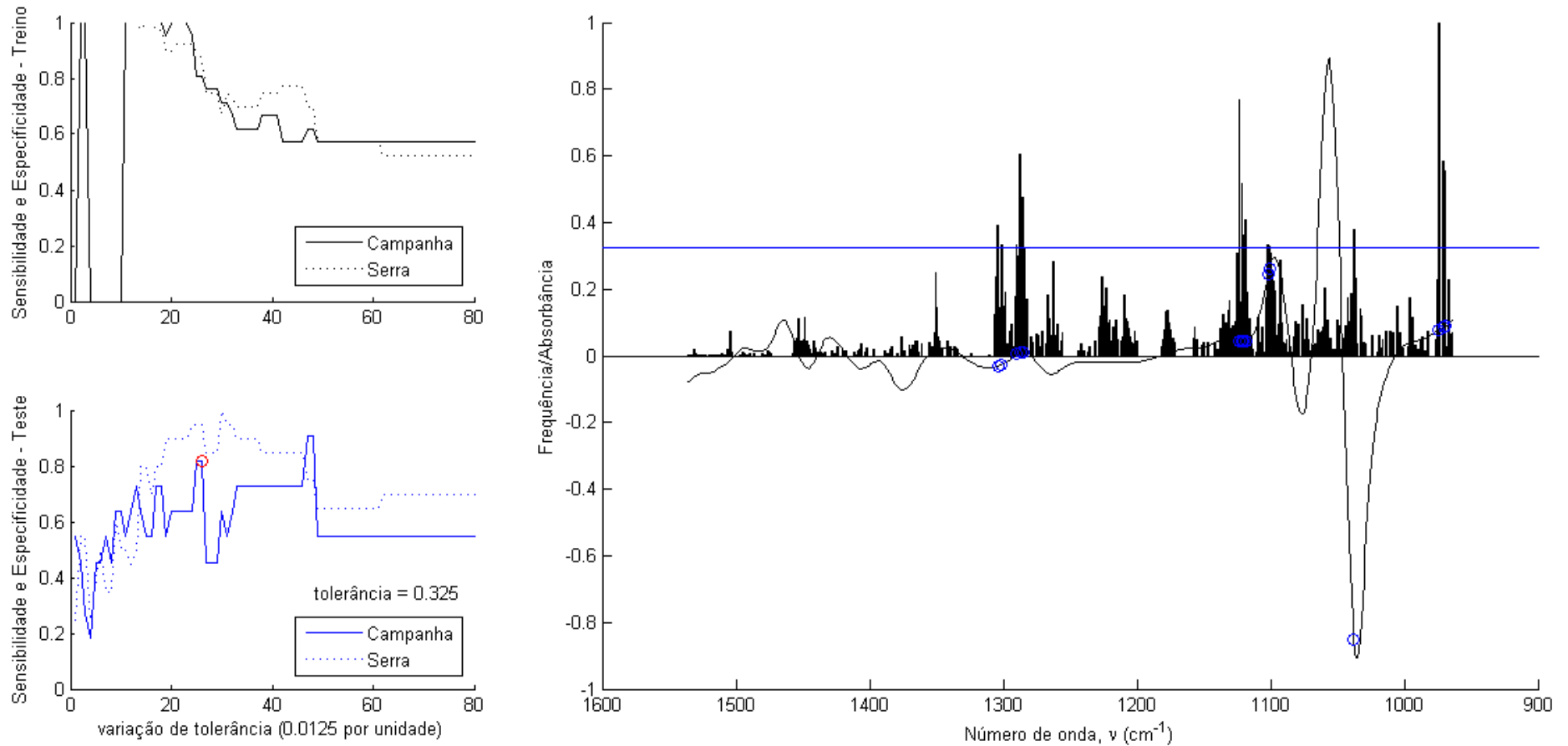


Figura 57. Frequência de variáveis selecionadas e espectro médio de X50

4.3 Estudos extra em Reconhecimento de Padrões

4.3.1 Efeitos na redução do conjunto amostral

Um conjunto amostral diverso e limitado eleva o teor de complexidade nos estudos de reconhecimento de padrões. Tendo em vista que os resultados apresentados até agora são fruto de aproximadamente 90 amostras Gaúchas, produzidas por dez tipos de uvas e de duas safras diferentes, avaliou-se a redução desse conjunto amostral na busca de uma uniformidade de distribuição nos seus componentes.

Para compor um conjunto amostral mais uniforme, foram mantidas as amostras cujos varietais apresentavam um exemplar, pelo menos, em cada safra de produção. Assim, houve a redução do conjunto para quatro varietais, com cinquenta e uma amostras ao todo. Como pode ser visto na Tabela 16, o conjunto é formado por amostras de Cabernet Franc, Cabernet Sauvignon, Marselan e Tannat. Mais uma vez, contudo, a Serra Gaúcha representa um subconjunto majoritário.

Tabela 16. Conjunto de amostras reduzido

| | | safra 2017 | safra 2018 | Região | Total | Total (%) |
|------------------|---|-------------------|-------------------|---------------|--------------|------------------|
| CF | c | 1 | 3 | 4 | 9 | 17,65 |
| | s | 1 | 4 | 5 | | |
| CS | c | 1 | 5 | 6 | 18 | 35,29 |
| | s | 1 | 11 | 12 | | |
| MN | c | 2 | 3 | 5 | 9 | 17,65 |
| | s | 2 | 2 | 4 | | |
| TT | c | 2 | 4 | 6 | 15 | 29,41 |
| | s | 2 | 7 | 9 | | |
| TOTAL | | 12 | 39 | 51 | 51 | 100,0 |
| TOTAL (%) | | 23,53 | 76,47 | | | |

| | | | |
|-----------|--------------------|-----------|----------|
| CF | Cabernet Franc | MN | Marselan |
| CS | Cabernet Sauvignon | TT | Tannat |

Como no processo discutido anteriormente, esse reduzido conjunto de dados foi tratado pelos métodos PLS – DA, SPA – LDA e GA – LDA. No reconhecimento de padrões via PLS – DA e SPA – LDA, foi executado o método para cada conjunto pré-tratado (de X01 a X57) com as derivadas de Savitzky – Golday. Devido ao tempo computacional despendido, o GA – LDA foi aplicado somente como comparação ao modelo entendido como de melhor desempenho nos métodos prévios.

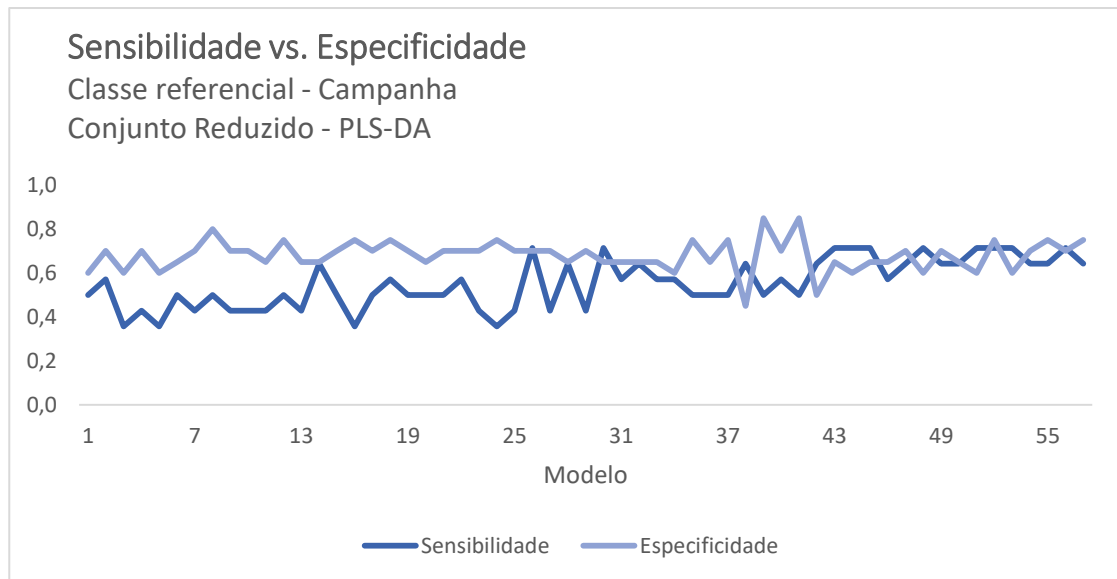


Figura 58. Relação Sensibilidade e Especificidade dos modelos PLS – DA

Ao avaliar a sensibilidade e especificidade na etapa de validação cruzada do PLS – DA e SPA – LDA, pelas Figura 58 e Figura 59, respectivamente, é possível perceber um comportamento similar ao do conjunto de dados anterior. Primeiramente, os modelos via PLS – DA apresentam uma crescente sensibilidade para a classe Campanha, com uma diferença reduzida ao longo dos modelos finais. Já pelo SPA – LDA, a diferença entre sensibilidade das classes não segue uma tendência fora uma média constante.

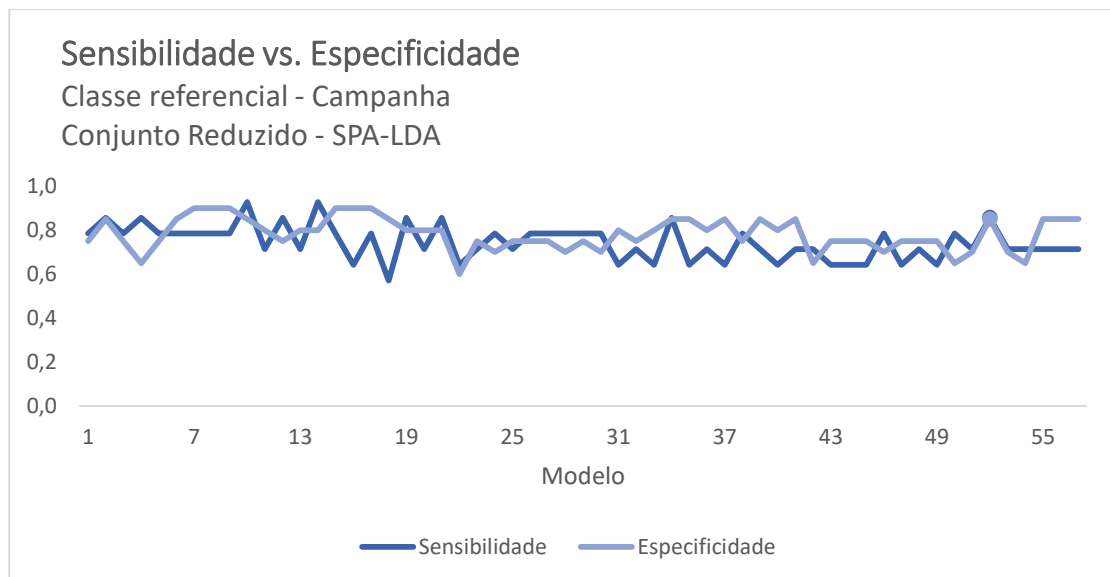


Figura 59. Relação Sensibilidade e Especificidade dos modelos SPA – LDA

Na classificação das amostras estudadas, em média, o SPA – LDA teve maior taxa de acerto durante a etapa de validação cruzada (Figura 60). Todavia, os valores de MCC (Figura 61) de ambos (PLS-DA e SPA-LDA) suportam a necessidade de janelas de alisamento maiores, no intuito de obter melhores desempenhos de classificação.

Bem como no conjunto amostral trabalhado anteriormente, as amostras passaram por partição via algoritmo de Kennard – Stone, reservando um terço dessas para teste do modelo formado. Uma vez modelado, testou-se o novo conjunto amostral nos respectivos métodos de reconhecimento de padrões.

Devido à sua baixa diferença entre sensibilidades das classes estudadas em ambos métodos, bem como por representar um ponto de máximo, o modelo escolhido para teste foi o correspondente ao ponto 52. Cabe recordar que o conjunto de dados trabalhado aqui foi pré-processado por primeira derivada aplicada ao espectro no infravermelho, com função ajustada a polinômio de terceiro grau e uma janela de alisamento com trinta e um pontos.

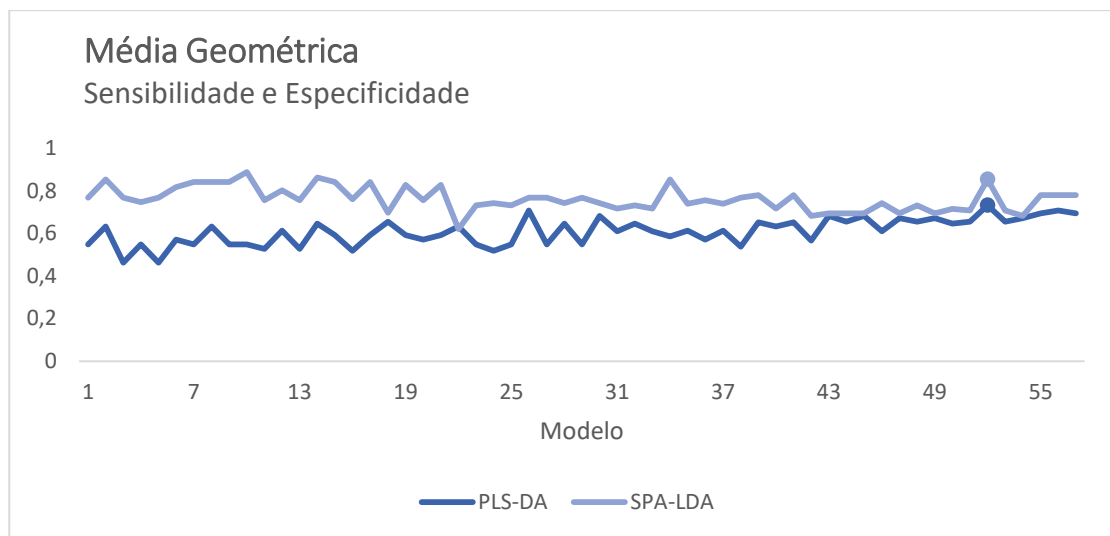


Figura 60. Média Geométrica do Conjunto Reduzido

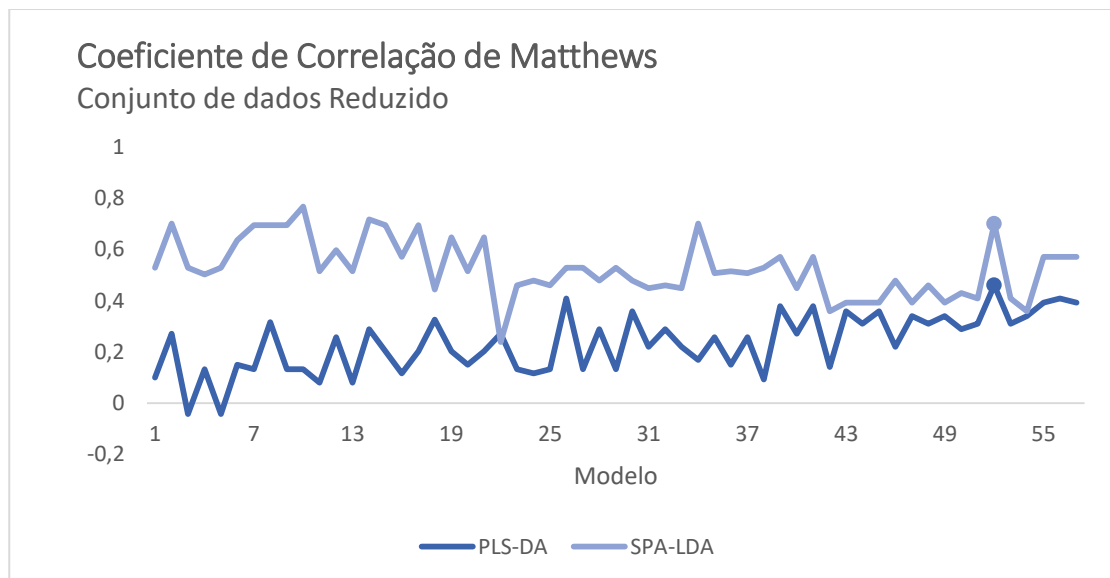


Figura 61. Coeficiente de Correlação de Matthews do Conjunto amostral reduzido

A Tabela 17 mostra as figuras de mérito resultantes da classificação na etapa de teste. Mais ainda, a tabela traz uma comparação com as métricas de desempenho obtidas por GA – LDA. Entre modelos via PLS – DA e SPA – LDA, o primeiro é levemente melhor que o segundo, considerando nível de acerto. Porém, ambas obtiveram altas taxas de falsos positivos, quando em comparação com o GA – LDA.

Tabela 17. Desempenho dos métodos de Reconhecimento de Padrão

| MODELO | PLS – DA | | SPA – LDA | | GA – LDA (*) | |
|----------------|------------------------|-----------|-----------------------|-----------|---------------------|-----------|
| | X52 | | | | | |
| MÉTRICA | C1 | C2 | C1 | C2 | C1 | C2 |
| GM | 0,68 | | 0,66 | | 0,88 | |
| SEN | 0,57 | 0,80 | 0,71 | 0,60 | 0,86 | 0,90 |
| SPE | 0,80 | 0,57 | 0,60 | 0,71 | 0,90 | 0,86 |
| PRE | 0,67 | 0,73 | 0,56 | 0,75 | 0,86 | 0,90 |
| EXA | 0,71 | | 0,65 | | 0,88 | |
| MCC | 0,38 | | 0,31 | | 0,76 | |
| TFP | 0,20 | 0,43 | 0,40 | 0,29 | 0,1 | 0,14 |
| SEM | Sensibilidade | | SPE | | Especificidade | |
| PRE | Precisão | | EXA | | | |
| TFP | Taxa de Falso Positivo | | (*) TOLERÂNCIA | | 0,625 | |
| GM | Média Geométrica | | | | | |
| C1 | Campanha | | C2 | | Serra Gaúcha | |

Respectivamente, as Figura 62, Figura 63 e Figura 64 apresentam o gráfico de loadings do PLS – DA, o espectro médio com variáveis selecionadas do SPA – LDA e o histograma das frequências do GA – LDA. Similar ao encontrado no conjunto original de dados, as variáveis de maior influência na classificação integram bandas no infravermelho referentes a álcool primário e polifenóis presentes no vinho.

Pesos das variáveis

PLS – DA

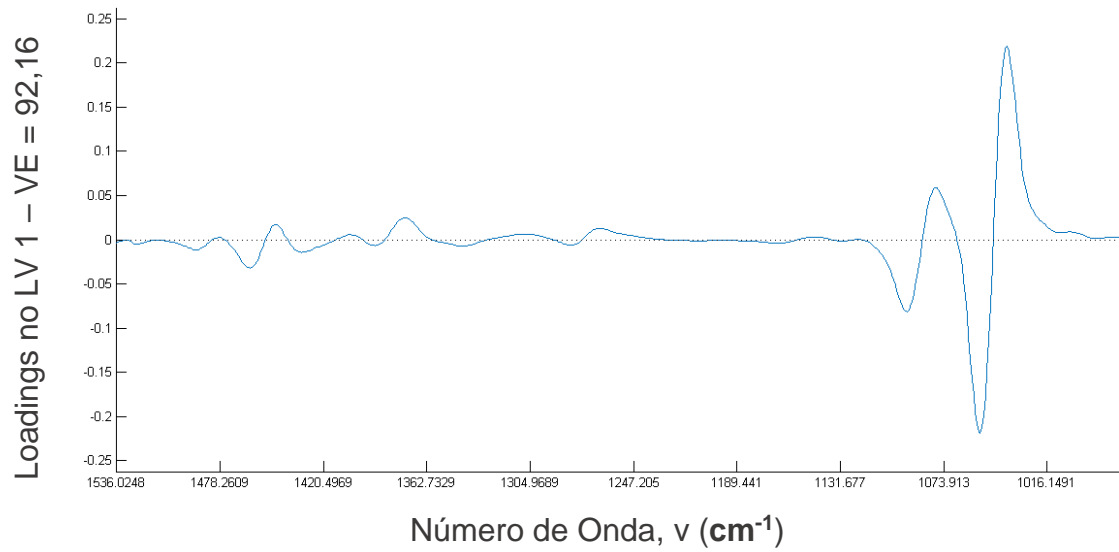


Figura 62. Gráfico de Loadings PLS – DA; Conjunto reduzido.

Variáveis selecionadas e Perfil Médio

SPA – LDA

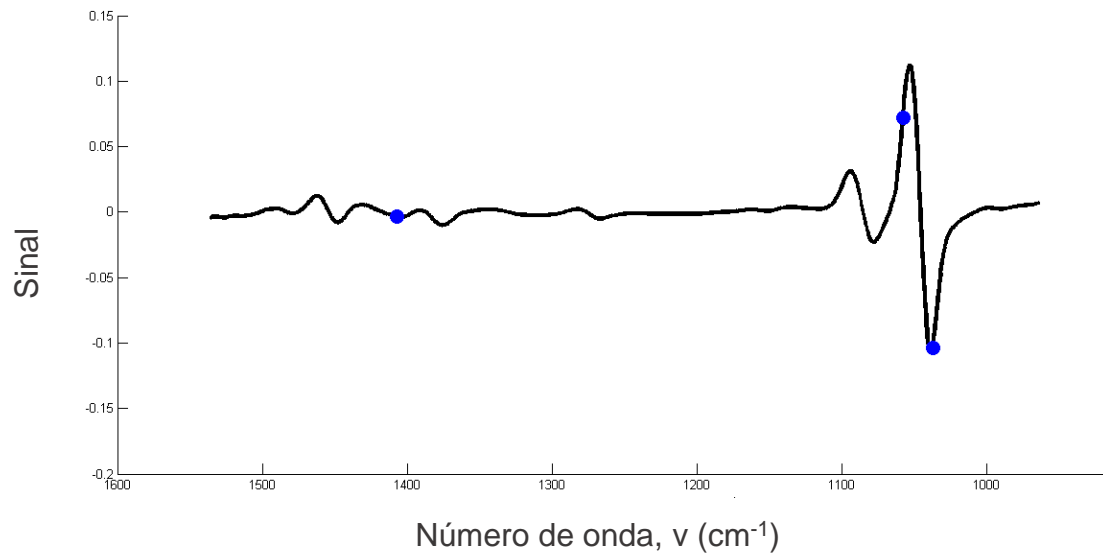


Figura 63. Perfil médio do IR e variáveis selecionadas por SPA – LDA

Frequências e Variáveis selecionadas

Perfil Médio aplicado em GA – LDA

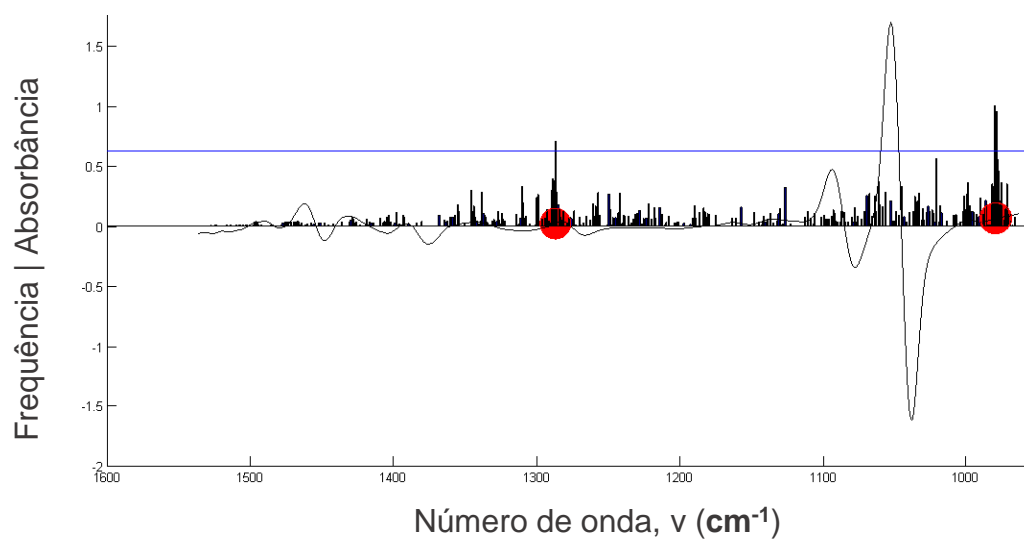


Figura 64. Perfil médio do IR e frequência de seleção de variáveis via GA – LDA

| Regiões Significativas | |
|------------------------|--|
| cm ⁻¹ | Bandas |
| 1448 – 1444 | Deformação angular assimétrica no plano de –CH ₃ . |
| | Deformação de – CH ₂ – característica de núcleo fenólico |
| 1376 – 1373 | Deformação angular simétrica no plano (tesoura) de – CH ₃ . |
| 1340 – 1339 | Balço de CH ₂ e Deformação angular de C – H. |
| 1340 – 1160 | Deformação axial em C – OH de fenóis |
| 1285 | Deformação axial do anel pirano de flavonoide. |
| 1281 – 1278 | |
| 1207 | Deformação angular no plano de O – H. |
| 1110 – 1107 | |
| 1068 – 1062 | Deformação axial C – O |
| 1050 | Deformação axial C – O de álcool primário. |

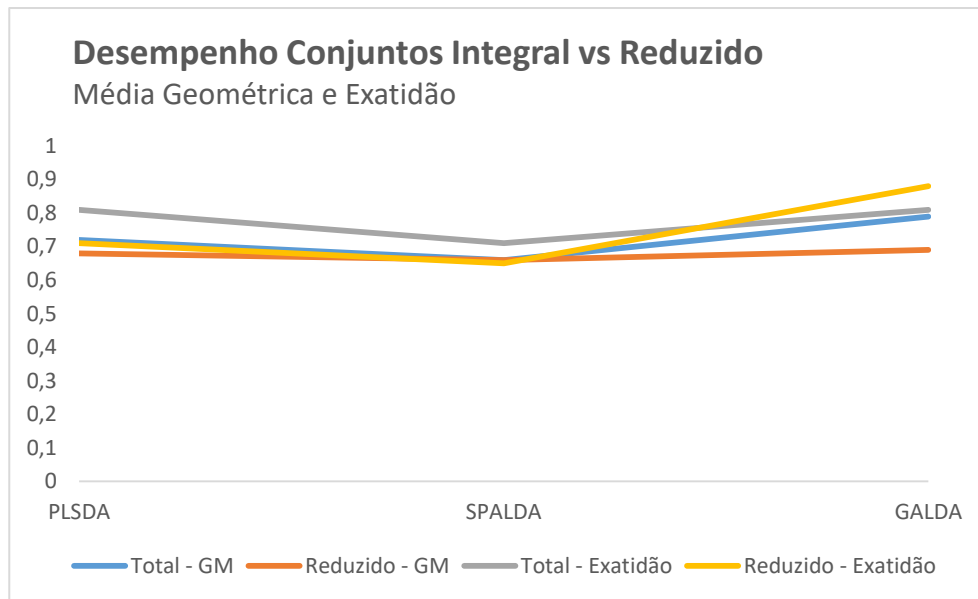


Figura 65. Comparação de desempenhos: conjunto integral vs. reduzido

Contra os valores encontrados no conjunto original e no intuito de visualizar o efeito da redução do conjunto amostral no desempenho dos modelos PLS – DA, SPA – LDA e GA – LDA, na Figura 65 comparou-se as métricas Média Geométrica e Exatidão. As métricas cuja legenda é acrescida da palavra total sinalizam a resposta dos dados originais, obtidos pelo método de melhor desempenho. Por exemplo, o conjunto de dados original, melhor classificado através de PLS – DA foi o X50.

Ainda, a Figura 65 mostra que a redução do conjunto amostral não conduz a melhorias significativas do reconhecimento de padrões. Apesar disso, a não redução nos valores aponta que as amostras remanescentes foram representativas o suficiente para manter a o padrão observado no conjunto amostral maior.

4.4 Regressão

Inicialmente, os dados foram organizados em planilhas conforme a mistura em estudo. As sete planilhas resultantes contêm os dados de intensidade obtidos pelo FTIR – ATR e as informações referentes à concentração (relação monovarietal A sobre monovarietal B) de cada replicata. Evidenciou-se as datas e horários de cada análise, para eventuais comparações, uma vez que diversos analistas auxiliaram na aquisição dos dados.

A inspeção inicial das tabelas organizadas foi feita de modo a assegurar que cada amostra analisada tivesse, no mínimo, três replicatas registradas. Raramente encontra-se mais de três análises por amostra, e mais raro ainda é encontrar menos que três. Ambas ocasiões foram contornadas pela aplicação da média de triplicatas nos tratamentos e estudos futuros ¹¹⁹.

Para cada um dos sete conjuntos de dados, as etapas de avaliação de desempenho dos métodos de regressão foram essencialmente as mesmas. Os dados obtidos das vinte e uma soluções (calibração) foram pré-processados por derivadas de Savitzky – Golay, variando o número de pontos da janela, o grau do polinômio e derivada. Os cinquenta e seis novos conjuntos de dados e o original foram individualmente particionados por SPXY, reservando dois terços das amostras para a calibração e validação cruzada dos modelos, enquanto o terço restante foi destinado à predição externa.

Um por um, os modelos foram calibrados via PLS, iSPA – PLS e iPLS. O desempenho desses foi avaliado visualmente, tomando como base as métricas resultantes da validação cruzada (do tipo *Leave one out*) e da predição externa. Como resultado, de cada método de regressão foram selecionados dez melhores modelos, mantendo os provenientes dos dados brutos a título de comparação.

Posterior à seleção dos dados em destaque, comparou-se o desempenho desses métodos frente a variação da função multiobjetivo proposta. A comparação pode ser observada nas Figura 66 (a, b, c, d, e, f, g). Nessas também estão demarcados os pontos de mínimo de cada método, selecionados para validação com conjunto amostral externo, por simbolizarem um menor custo na validação interna, ou seja, uma maior aptidão de aplicabilidade preditiva com relação aos demais.

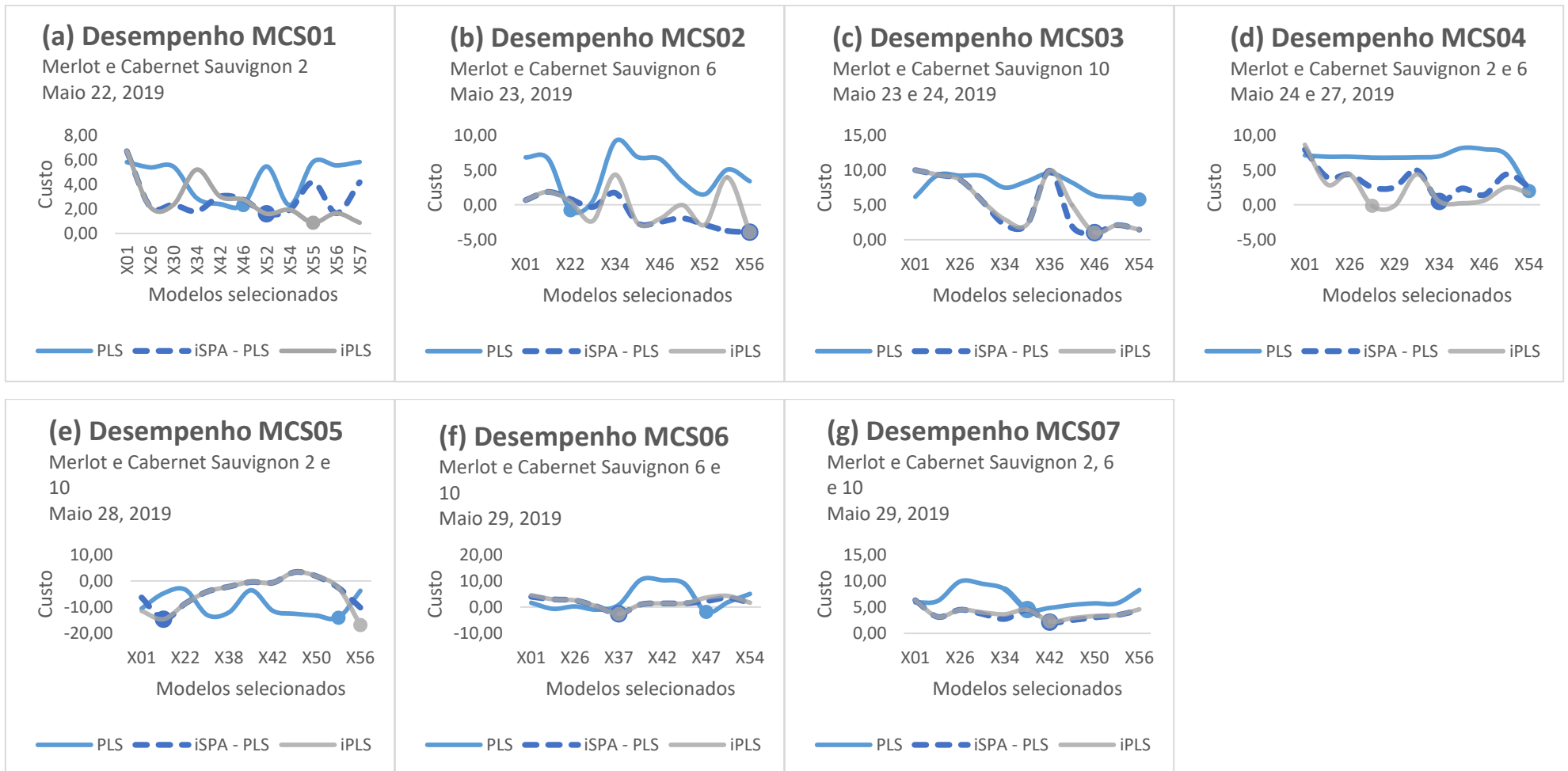


Figura 66. Funções Multiobjetivo

Após inspeção do conjunto de modelos via função multiobjetivo, testou-se cada modelo selecionado por meio da região de confiança elíptica (EJCR). Essa inspeção possibilita a identificação de tendências e desvios como, também, serve de estimativa para precisão relativa aos modelos em estudo. A comparação entre modelos referentes aos conjuntos de soluções pode ser observado nas Figura 67, Figura 68, Figura 69, Figura 70, Figura 71, Figura 72 e Figura 73.

Considerando o ponto ideal no interior das elipses como adequação das amostras validadas ao modelo construído, é possível dizer que, dos 21 modelos estudados, apenas um apresentou desvio (Figura 68). Diferente dos demais métodos, o PLS utiliza todas as informações fornecidas pelo espectro, eventualmente não conseguindo filtrar as variações introduzidas devido às diferentes condições ambientais e particularidades do dia. Inclusive, um dos operadores do FTIR era o aluno de iniciação científica, infelizmente desassistido durante uma parte do seu turno daquele dia.

Com desvio presente ou não na etapa de validação cruzada, todos os modelos foram testados com os conjuntos amostrais de predição e validação externa. Como dito anteriormente, uma das etapas de decisão para aplicabilidade dos modelos na etapa de validação externa foi o desempenho, relativamente adequado, na predição das amostras do conjunto inicialmente particionado. Assim, dos 1197 modelos iniciais (57 pré-processamentos, 3 métodos, 7 conjuntos de misturas), 231 foram previamente selecionados (11:3:7) mas 21 foram levados à etapa de validação externa.

Método OLS
Merlot e Cabernet Sauvignon 2
Maio 22, 2019
PLS: X46; iSPA-PLS: X52; iPLS: X55

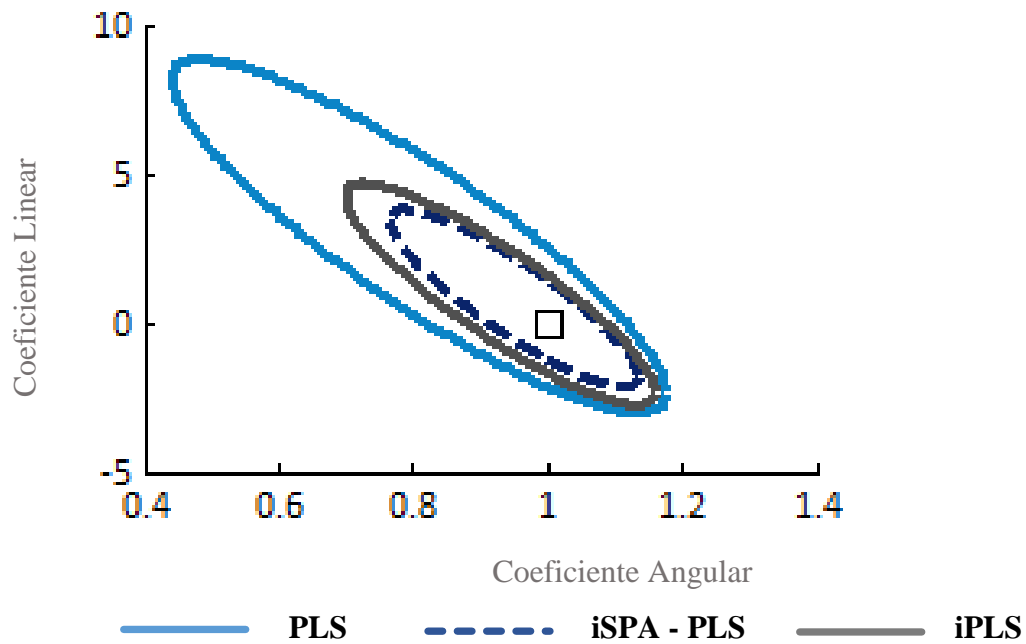


Figura 67. EJCRC - MCS01

Método OLS
Merlot e Cabernet Sauvignon 6
Maio 23, 2019
PLS: X22; iSPA-PLS e iPLS: X56

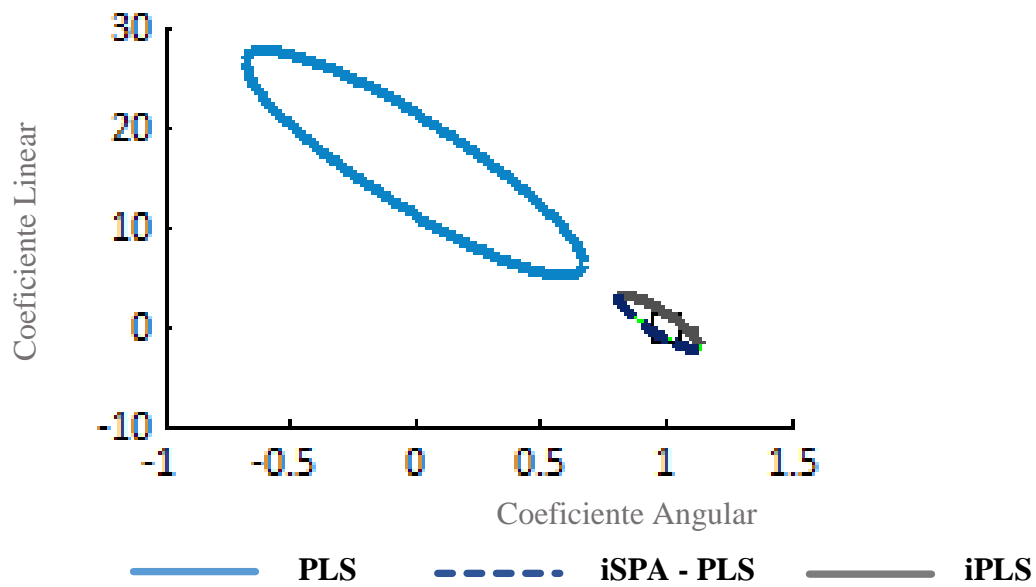


Figura 68. EJCRC - MCS02

Método OLS
Merlot e Cabernet Sauvignon 10
Maio 23 e 24, 2019
PLS: X54; iSPA-PLS e iPLS: X46

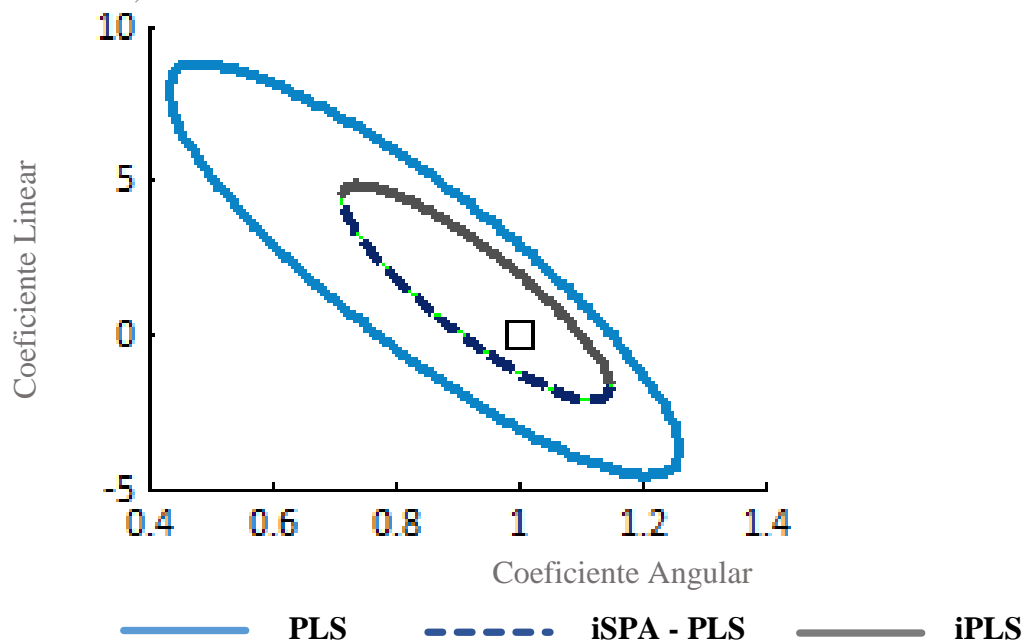


Figura 69. EJCR - MCS03

Método OLS
Merlot e Cabernet Sauvignon 2 e 6
Maio 24 e 27, 2019
PLS: X54; iSPA-PLS: X34; iPLS: X27

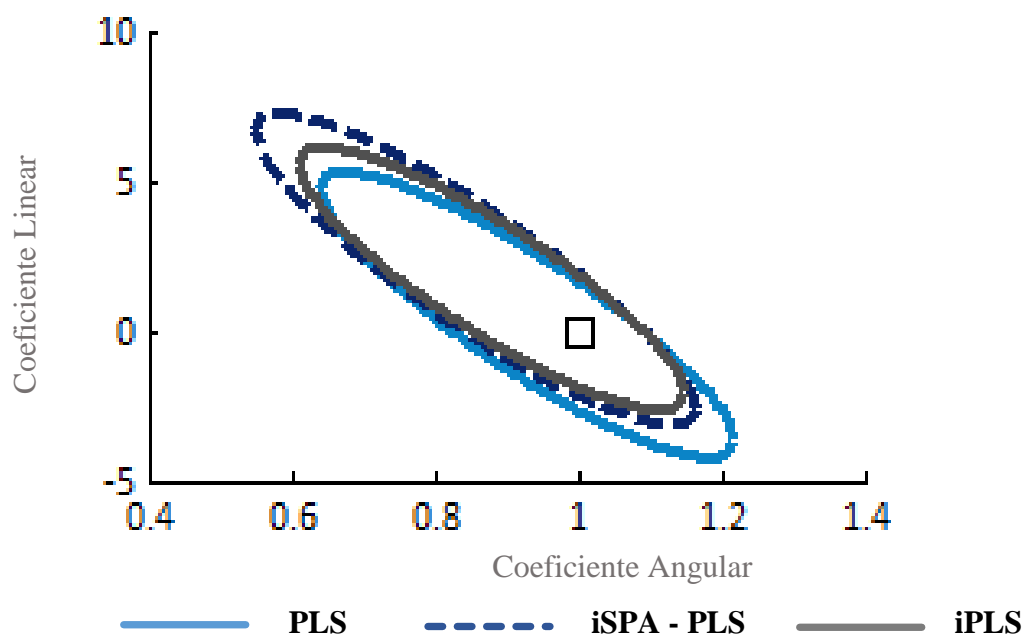


Figura 70. EJCR - MCS04

Método OLS
Merlot e Cabernet Sauvignon 2 e 10
Maio 28, 2019
PLS: X54; iSPA-PLS: X18; iPLS: X56

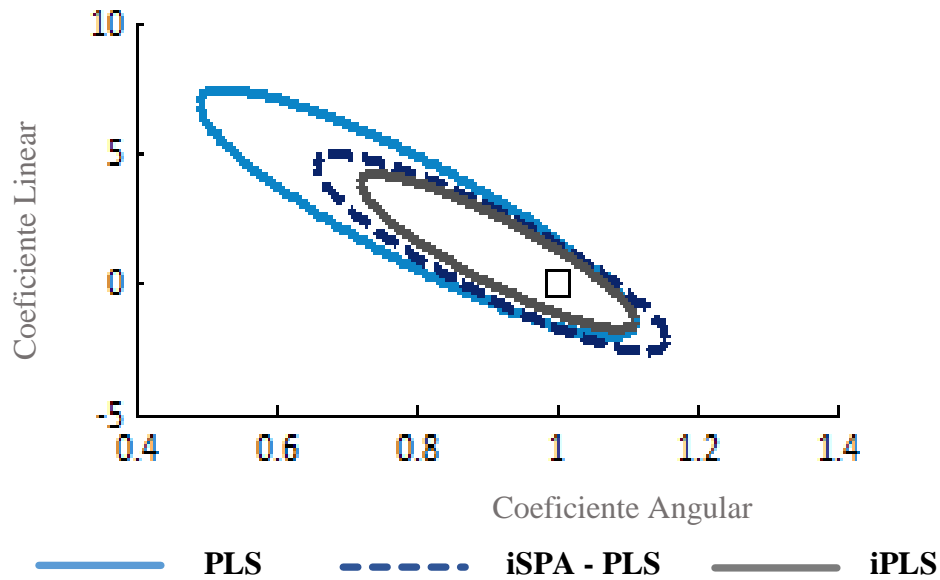


Figura 71. EJCRC - MCS05

Método OLS
Merlot e Cabernet Sauvignon 6 e 10
Maio 29, 2019
PLS: X47; iSPA-PLS e iPLS: X37

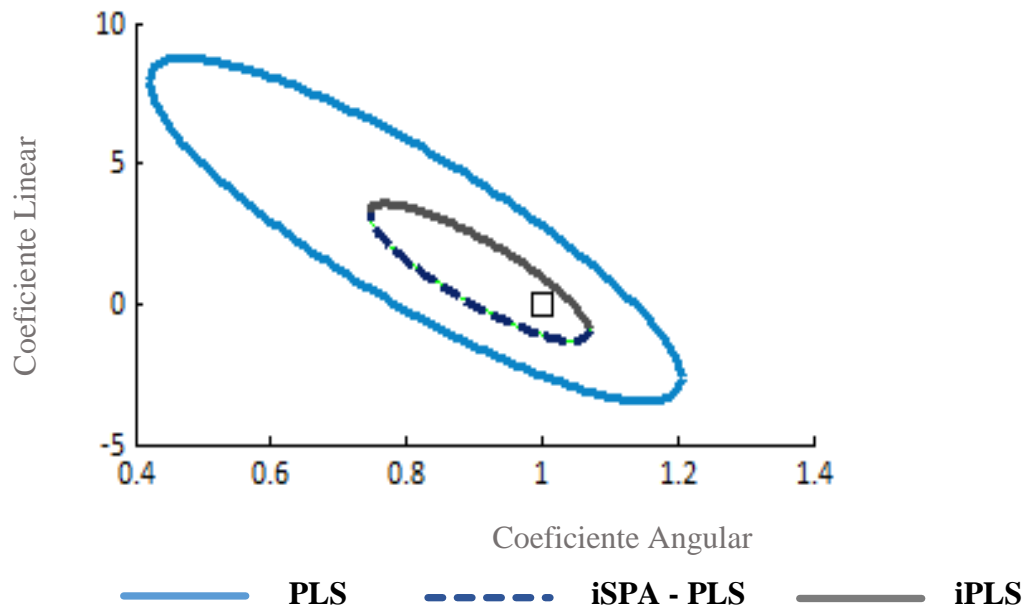


Figura 72. EJCRC - MCS06

Método OLS
Merlot e Cabernet Sauvignon 2, 6 e 10
Maio 29, 2019
PLS: X38; iSPA-PLS e iPLS: X42

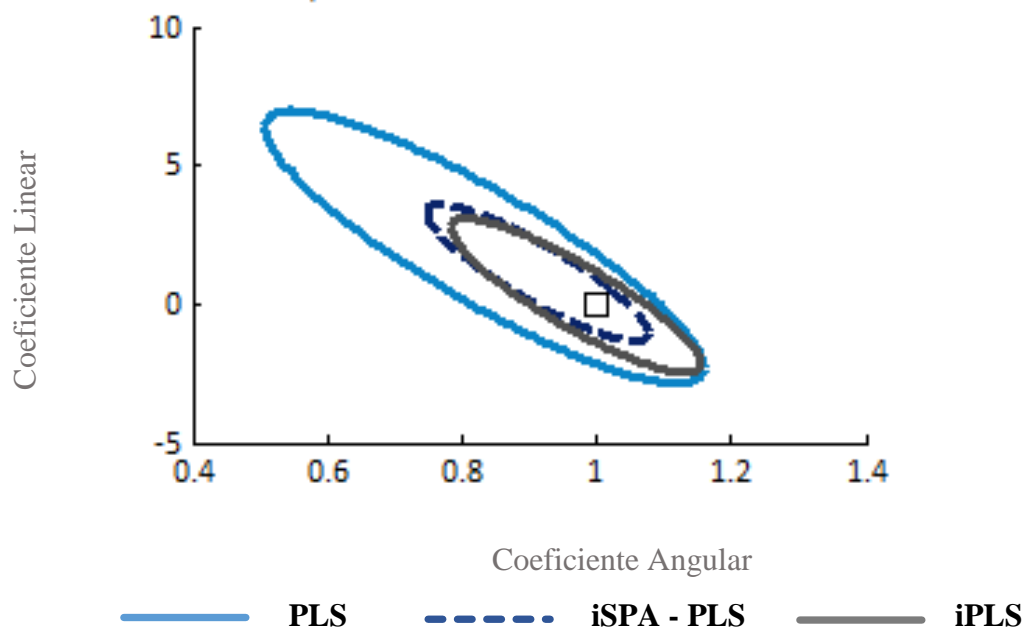


Figura 73. EJCR - MCS07

Tal qual abordado nos modelos de reconhecimento de padrões, os métodos de regressão aplicados podem tanto utilizar toda como uma fração da extensão de números de onda mais significativos do espectro em estudo. Os modelos via regressão PLS extraem informação de todo o espectro infravermelho fornecido, enquanto os iSPA – PLS e iPLS procuram uma região do espectro que corrobore com as informações suplementares, como a concentração do varietal de interesse.

Para os modelos de desempenho destaque, que serão discutidos posteriormente, o trabalho apresenta as regiões de variáveis selecionadas para a regressão. Pelas Figura 74, Figura 78, Figura 80 e Figura 81, é possível constatar a importância da região de números de onda entre 1137 e 1080 (cm^{-1}). Já as Figura 75, Figura 76, Figura 77 e Figura 79 destacam a região próxima a essa, entre 1194 e 1137 (cm^{-1}). Ambas regiões são características de vibrações carbono – oxigênio de álcoois (ν , C – O).

Há também o modelo iSPA – PLS, construído pela seleção de múltiplas regiões (Figura 81). Além da região comentada anteriormente (1137 e 1080 cm^{-1}), esse método selecionou as regiões dos números de onda entre 1531 e 1481, de 1424 até 1367, e de 1026 a 970 cm^{-1} . Essas regiões representam a vibração de polifenóis presentes no vinho quando

sob irradiação no infravermelho médio. Mais especificamente, as duas primeiras regiões (1531 até 1367 cm^{-1}) refletem alongamento da ligação dupla carbono – carbono (típicos de um sistema aromático) e deformações angulares das ligações metila ³².

A literatura apresenta algumas comparações entre tipos de uvas que refletem na caracterização dos vinhos e que sugerem uma explicação para a seleção das variáveis supracitadas. Por exemplo, a separação de classes entre esses vinhos com PCA, sendo as primeiras componentes a concentração alcóolica e conteúdo polifenólico ^{80,111}.

Nesse contexto, e considerando que os vinhos aqui analisados foram produzidos sob condições controladas, é possível afirmar que a variação observada reflete as espécies de uva envolvidas. Assim, a informação extraída dos espectros no FTIR através dos algoritmos de seleção de variáveis é de uma diluição alcoólica, e em alguns casos de conteúdo polifenólico.

Posteriormente à etapa de validação com o conjunto amostral externo, os vinte e um (21) modelos tiveram seu desempenho comparado de acordo com o conjunto de soluções em questão e com o método de regressão aplicado. As métricas resultantes podem ser observadas na Tabela 18.

Dentro dos fatores representados na Tabela 18 estão os erros de predição (RMSEP e REP), os coeficientes de determinação e correlação (R^2 e r), o desvio padrão das amostras de validação (SDV) e, finalmente, a confirmação da ausência de tendências na aplicação das amostras de validação. Essas métricas representam o desempenho dos modelos selecionados, todavia é importante ressaltar que alguns resultados foram atingidos somente por meio da exclusão de amostras anômalas.

Na validação, foram desconsideradas tais amostras problemáticas para, assim, suprimir seu efeito deletério no desempenho da validação. Eventualmente, desconsiderou-se de uma a quatro amostras dos conjuntos estudados. Apesar de não serem sempre as mesmas, todas se apresentam como anômalas pelo menos em uma ocasião.

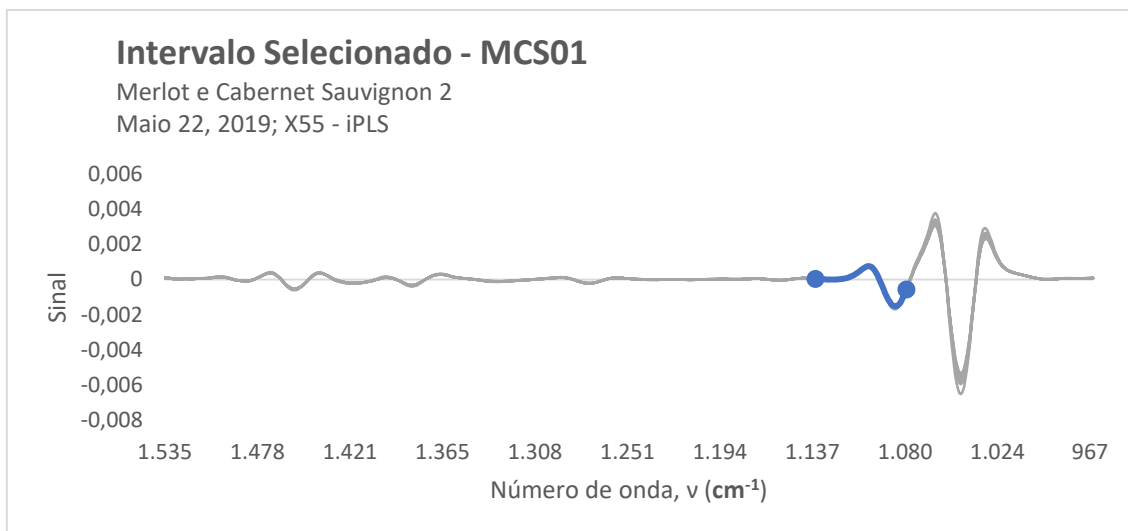


Figura 74. Seleção de intervalo em FTIR pré-processado em MCS01

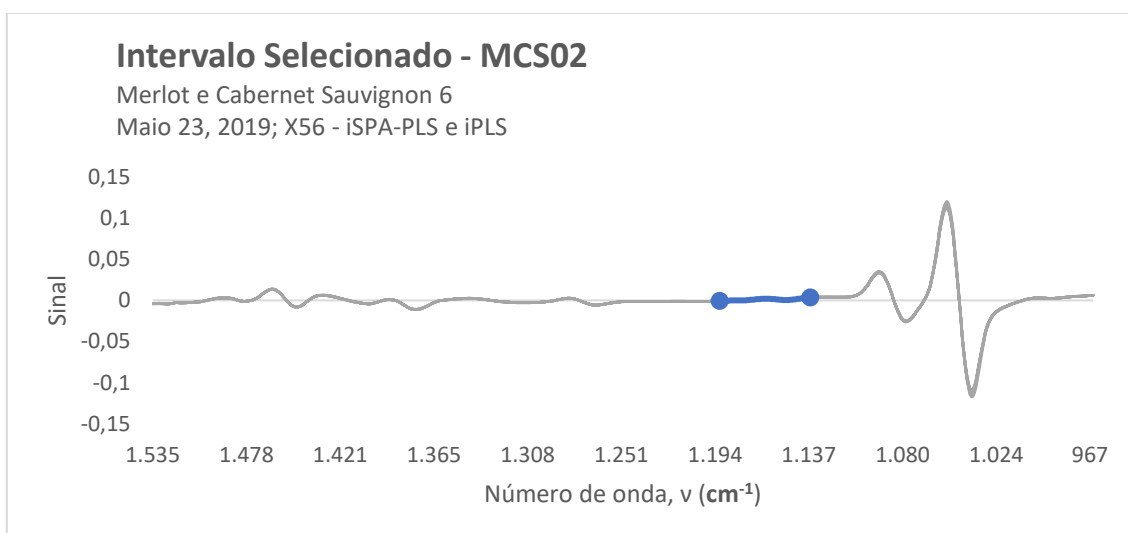


Figura 75. Seleção de intervalo em FTIR pré-processado em MCS02

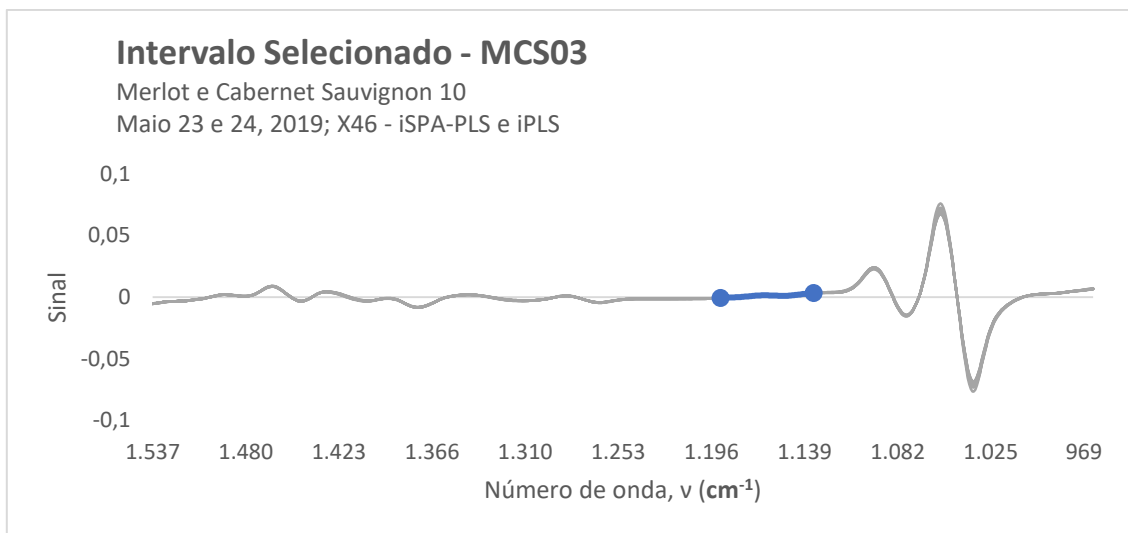


Figura 76. Seleção de intervalo em FTIR pré-processado em MCS03

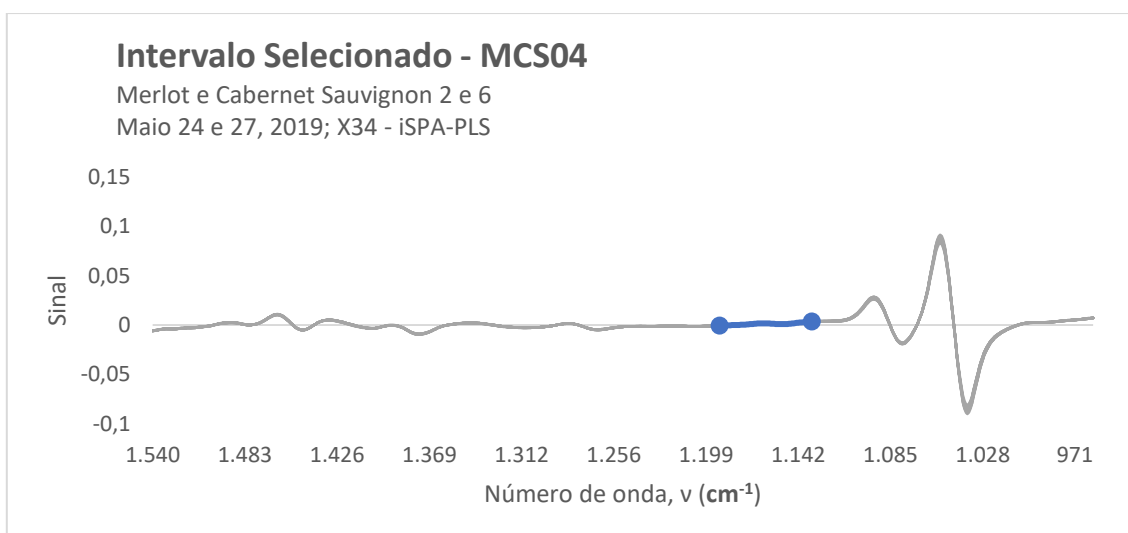


Figura 77. Seleção de intervalo em FTIR pré-processado em MCS04

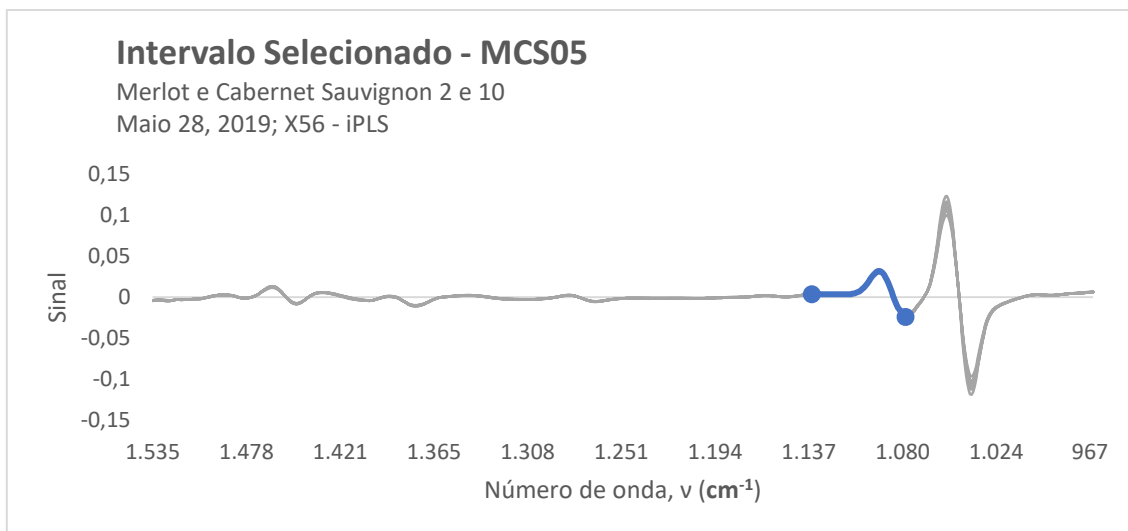


Figura 78. Seleção de intervalo em FTIR pré-processado em MCS05

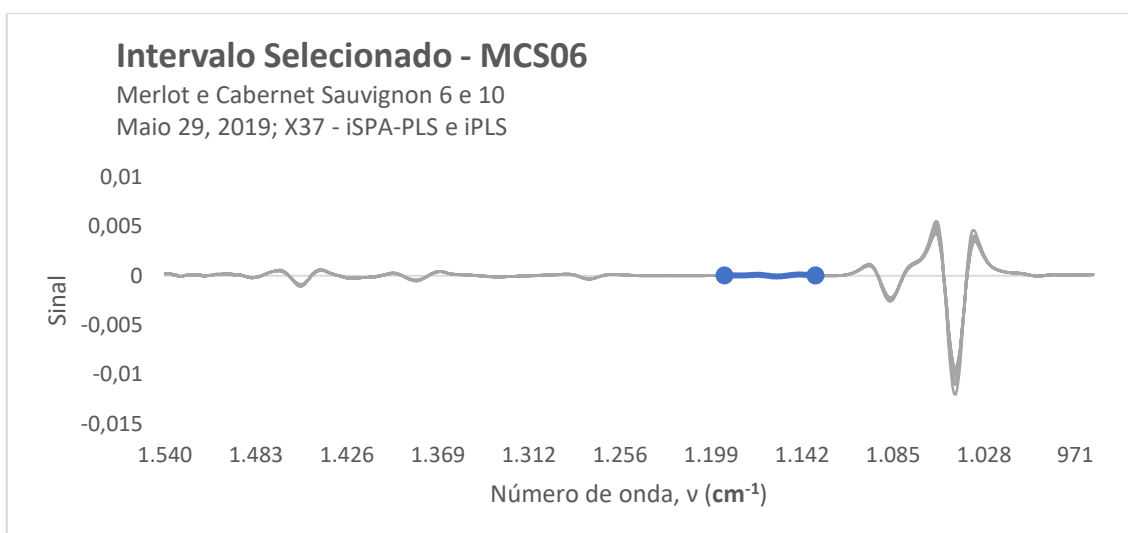


Figura 79. Seleção de intervalo em FTIR pré-processado em MCS06

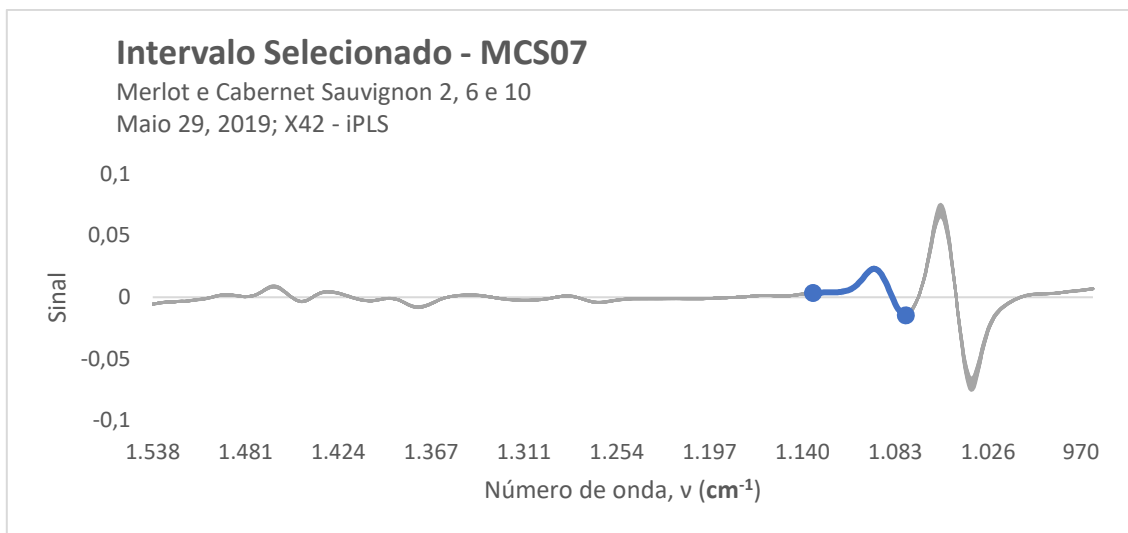


Figura 80. Seleção de intervalo em FTIR pré-processado em MCS07

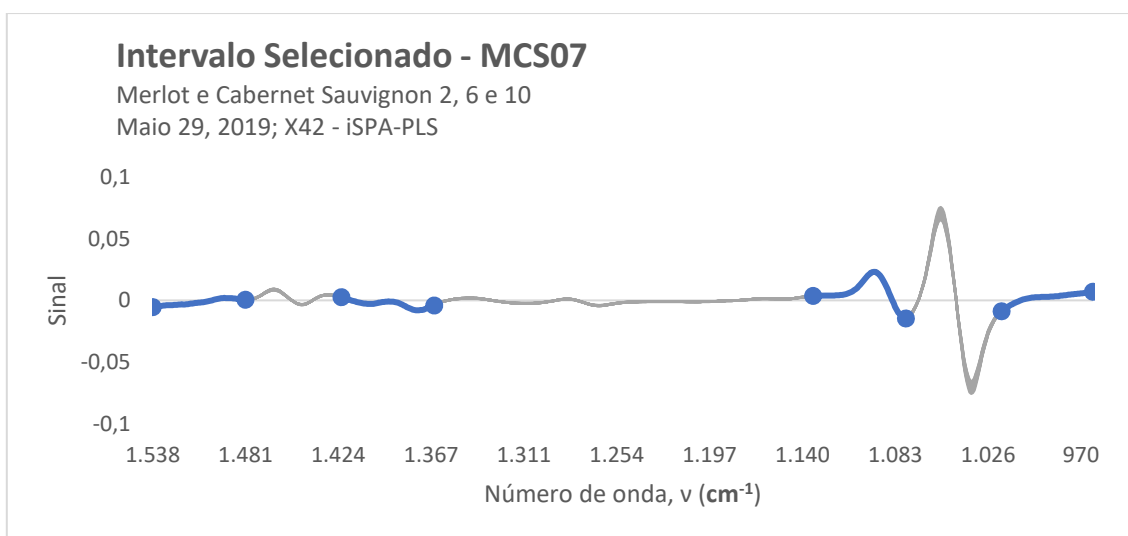


Figura 81. Seleção de intervalo em FTIR pré-processado em MCS07b

Tabela 18. Métodos de Regressão em conjuntos de dados selecionados

| (continua) | | | | | | | |
|-------------------|-------------------------------|------------|-------------------|-------------|--------|---------|--------|
| SOLUÇÕES | MÉTRICAS MODELOS | PLS | ISPA – PLS | IPLS | | | |
| MCS01 | RMSEP | | 6,951 | | 3,904 | 1,516 | |
| | REP | | 51,72 | | 29,06 | 11,28 | |
| | R ² | | 0,5301 | | 0,6720 | 0,8036 | |
| | <i>r</i> | X46 | 0,9691 | X52 | 0,9984 | X55 | 0,9938 |
| | SDV | | 11,68 | | 6,992 | 1,635 | |
| | Ausência de <i>Bias</i> | | Sim | | Sim | Sim | |
| MCS02 | RMSEP | | 6,017 | | 1,716 | 1,716 | |
| | REP | | 44,78 | | 12,77 | 12,77 | |
| | R ² | | 0,9990 | | 0,9999 | 0,9999 | |
| | <i>r</i> | X22 | 0,9554 | X56 | 0,9904 | X56 | 0,9904 |
| | SDV | | 3,376 | | 1,707 | 1,707 | |
| | Ausência de <i>Bias</i> | | Não | | Sim | Sim | |
| MCS03 | RMSEP | | 1,970 | | 1,646 | 1,827 | |
| | REP | | 15,25 | | 12,25 | 16,24 | |
| | R ² | | 0,7821 | | 0,7864 | 0,7906 | |
| | <i>r</i> | X54 (*) | 0,9966 | X46 | 0,9928 | X46 (*) | 0,9911 |
| | SDV | | 1,642 | | 1,787 | 2,149 | |
| | Ausência de <i>Bias</i> | | Sim | | Sim | Sim | |
| MCS04 | RMSEP | | 4,397 | | 1,666 | 2,031 | |
| | REP | | 31,04 | | 11,76 | 13,54 | |
| | R ² | | 0,5964 | | 0,7254 | 0,8351 | |
| | <i>r</i> | X54 (*) | 0,9351 | X34 (*) | 0,9989 | X27 (*) | 0,9849 |
| | SDV | | 5,481 | | 1,795 | 2,036 | |
| | Ausência de <i>Bias</i> | | Sim | | Sim | Sim | |
| MCS05 | RMSEP | | 2,467 | | 4,2485 | 5,145 | |
| | REP | | 10,96 | | 21,24 | 39,29 | |
| | R ² | | 0,5871 | | 0,8383 | 0,8421 | |
| | <i>r</i> | X54 (*) | 0,9762 | X18 (*) | 0,9869 | X56 | 0,9941 |
| | SDV | | 4,234 | | 8,644 | 10,16 | |
| | Ausência de <i>Bias</i> | | Sim | | Sim | Sim | |
| MCS06 | RMSEP | | 16,00 | | 4,829 | 6,531 | |
| | REP | X47 | 119,03 | X37 (*) | 27,60 | X37 | 48,61 |

Tabela 18. Métodos de Regressão em conjuntos de dados selecionados

| | | | | | | | (Conclusão) |
|--------------|-------------------------|---|---------|-----------------------------------|---------------------------|---------|-------------|
| MCS06 | R^2 | | 0,9999 | | 0,2914 | | 0,4006 |
| | r | X47 | -0,1561 | X37 (*) | 0,9964 | X37 | 0,9963 |
| | SDV | | 27,96 | | 6,487 | | 10,17 |
| | Ausência de <i>Bias</i> | | Sim | | Sim | | Sim |
| MCS07 | RMSEP | | 3,994 | | 2,013 | | 1,5219 |
| | REP | | 29,72 | | 16,10 | | 11,07 |
| | R^2 | X38 | 0,4532 | X42 (*) | 0,7311 | X42 (*) | 0,9553 |
| | r | | 0,9803 | | 0,9988 | | 0,9975 |
| | SDV | | 4,790 | | 1,808 | | 0,7842 |
| | Ausência de <i>Bias</i> | | Sim | | Sim | | Não |
| NOTAS | R^2 | Coeficiente de determinação | | REP | Erro Relativo de Predição | | |
| | r | Coeficiente de correlação | | | | | |
| | SDV | Desvio Padrão da Validação | (*) | Exclusão de amostra(s) anômala(s) | | | |
| | RMSEP | Raiz quadrada da média de erros em predição | | | | | |

Em suma, pela Tabela 18 é possível perceber que os modelos com melhor desempenho são, em sua maioria, os construídos com base em seleção de variáveis. Ao considerar os erros de predição (RMSEP e REP), a seleção de regiões do espectro foi eficiente no propósito da predição de concentração das misturas em estudo, exceto nos modelos MCS05, onde a validação teve um desempenho relativo melhor mediante regressão PLS.

Os coeficientes de determinação e correlação também mostram um melhor desempenho dos métodos de regressão com seleção de variáveis. Mesmo com diferenças não tão expressivas, tais coeficientes suportam a percepção de que os modelos via iSPA – PLS e iPLS designam adequadamente as amostras em análise às suas regiões de concentração. Uma exceção é o conjunto de dados MCS06 que, além da randomicidade da predição ($R^2 < 0,5$), apresenta correlação baixa, senão inversa, entre concentração e intensidade do FTIR.

De modo geral, os modelos MCS01 à MCS03 tendem a ser mais estáveis e melhor preditivos do que o restante dos conjuntos. Fora as variações agregadas em decorrência do ambiente e operadores do FTIR, as misturas de MCS04 à MCS07 são preparadas com volumes menores de cada garrafa de Cabernet Sauvignon. Ou seja, mesmo que a proporção geral do varietal minoritário seja a mesma dos primeiros conjuntos de misturas, eleva-se a

incerteza associada às micropipetas, o que reflete no erro de predição e deslocamento de inclinação da linha de tendência das amostras em validação.

Ainda na etapa de validação, os métodos de regressão aplicados nas sete misturas resultaram em vinte e um modelos, que foram inspecionados por meio do teste F¹⁰⁹ de modo a entender as frequências das amostras anômalas. Desses modelos, apenas 15,5% das amostras foram consideradas anômalas. Em ordem percentual decrescente de ocorrências de desvio estão as soluções de Cabernet Sauvignon das seguintes concentrações: 2,5% (26,92%), 0% (19,23%), 25% (15,39%), 5% e 20% (11,54%), 15% (7,69%), 10% e 30% (3,85%). Uma comparação de regiões modeladas e validadas pode ser vista na Figura 82.

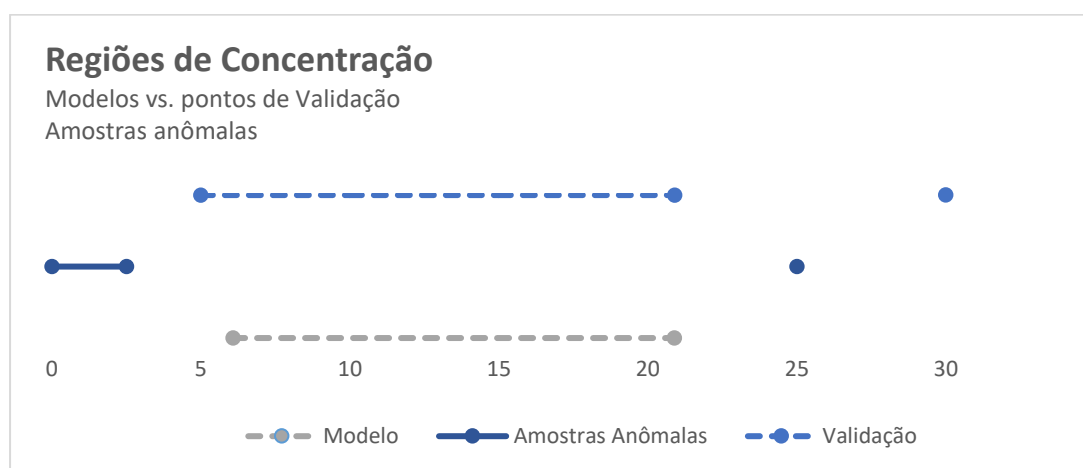


Figura 82. Regiões de calibração, validação e pontos de anomalia

O conjunto e a frequência desses desvios podem expressar dificuldades dos métodos em extrair informações úteis apesar do ruído. Merecem atenção as misturas de concentração de Cabernet Sauvignon a 2,5% e 0%. De forma conjunta, ambas são responsáveis por 46,2% dos desvios. Não menos importante, porém, mesmo com a reduzida fração de desvios no ponto de concentração máxima, alguns modelos consideram como anômala a amostra de validação do ponto 25%. Ou seja, por um lado alguns modelos estão majoritariamente limitados à curva de calibração, por outro, o desvio se agrava à medida que se aproxima de um limite de sensibilidade do equipamento.

Sob outro aspecto, é possível que os desvios em pontos de baixa concentração tenham refletido, também, um erro sistemático inerente à troca de micropipetas. Considerando as frequências dos pontos 0%, 2,5%, 5% e 10%, chega-se a 61,5% das amostras anômalas. No processo de preparo das misturas de vinhos, não foi utilizada a prática

de diluição por solução intermediária, em consequência disso, tanto a utilização de micropipetas de valor nominal distintas, quanto a coleta de alíquotas próximas ao limite de trabalho inferior, podem ter afetado o resultado da validação nessa região de concentrações.

De qualquer forma, ao inspecionar os modelos optou-se por representar graficamente somente os melhores de cada conjunto de soluções. Como dito anteriormente, o resultado obtido nas métricas da Tabela 18 é reflexo da habilidade do modelo, bem como da exclusão de amostras anômalas. Ainda assim, mesmo não sendo contabilizadas nas etapas de comparação de desempenho, as amostras anômalas são representadas e discriminadas quando ocorrem; por exemplo, nas soluções de MCS04 (Figura 86), MCS05 (Figura 87) e MCS07 (Figura 89).

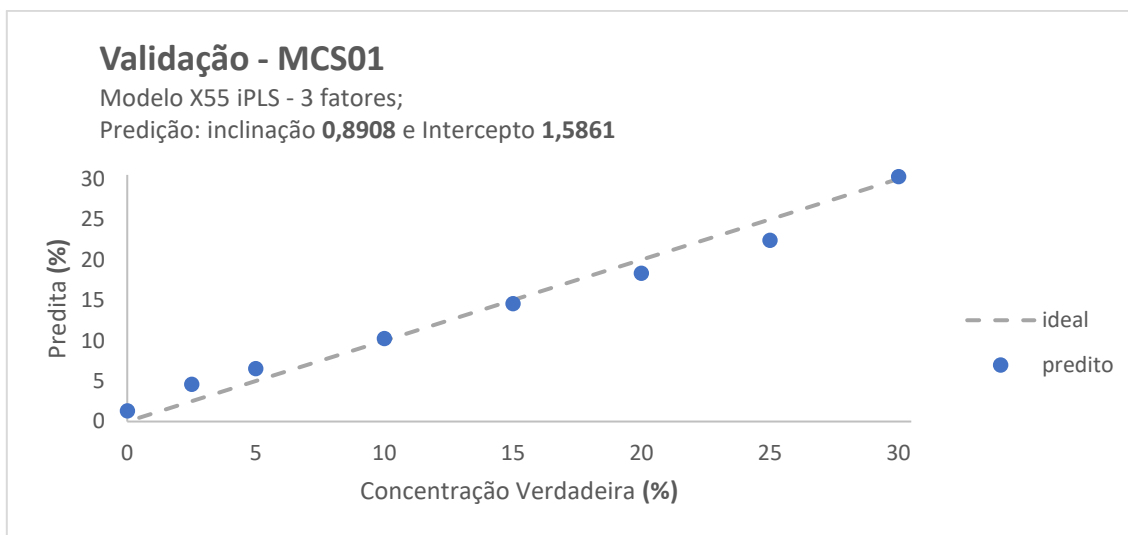


Figura 83. Amostras de Validação MCS01

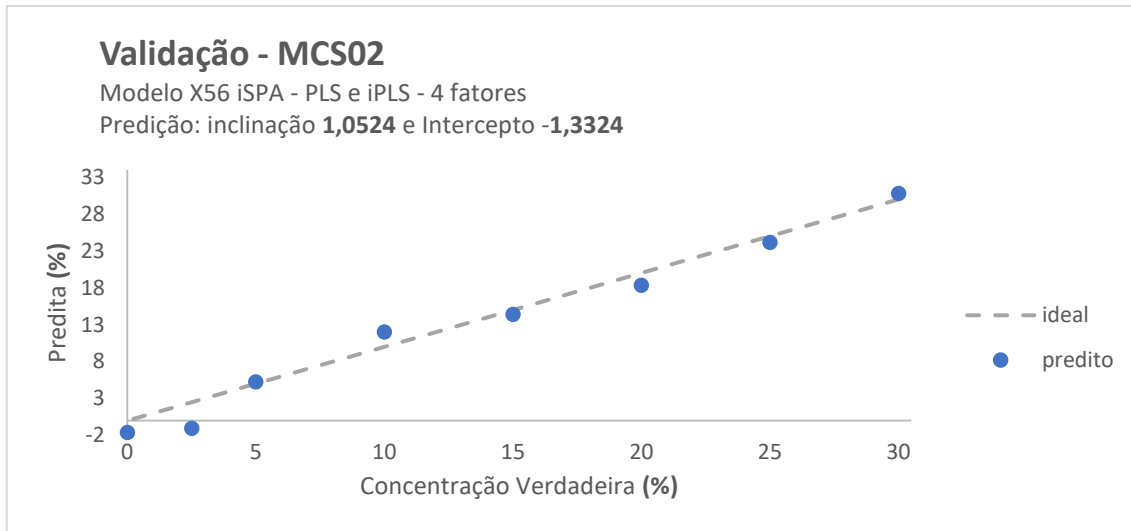


Figura 84. Amostras de Validação MCS02

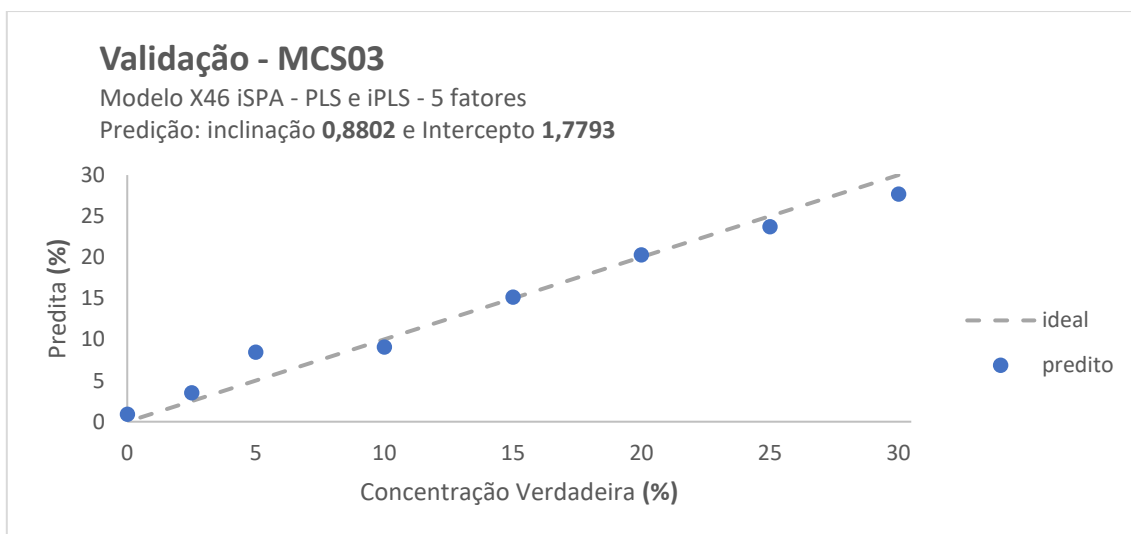


Figura 85. Amostras de Validação MCS03

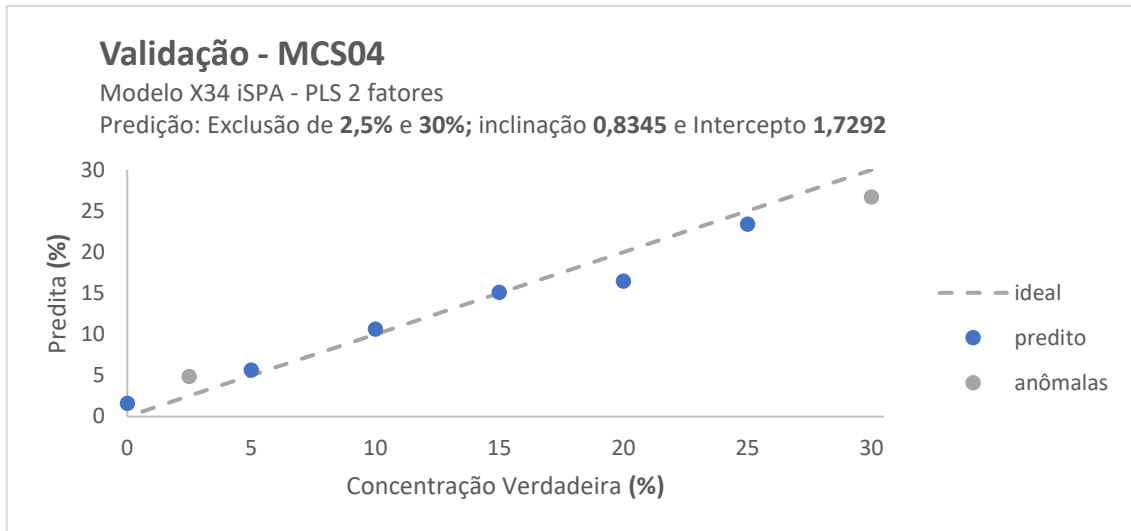


Figura 86. Amostras de Validação MCS04

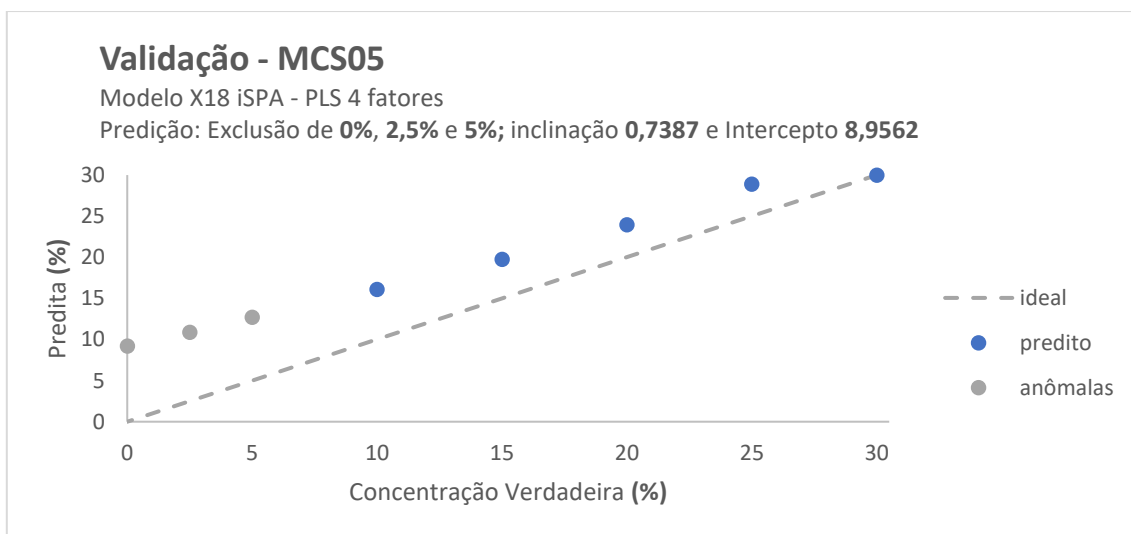


Figura 87. Amostras de Validação MCS05

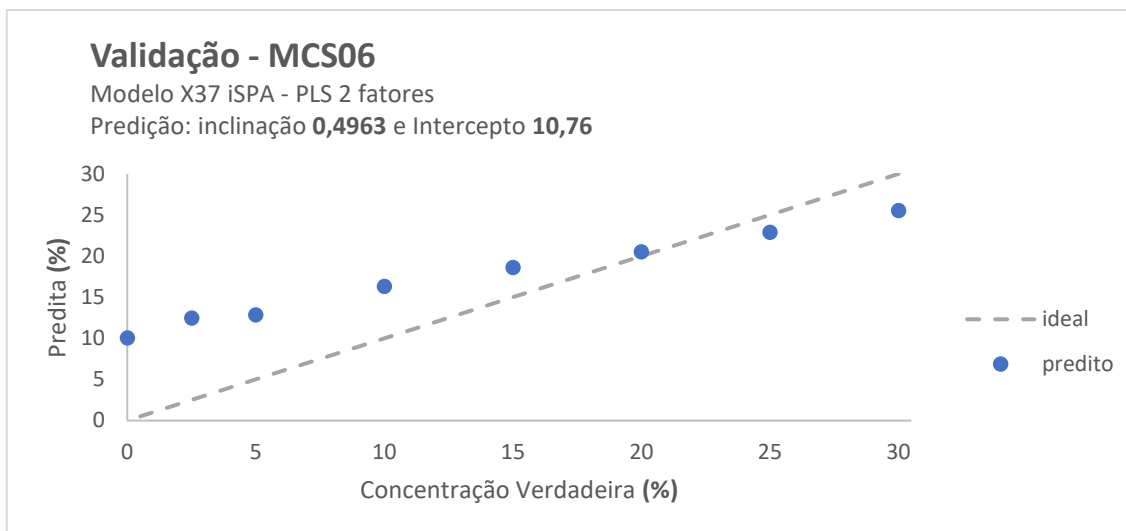


Figura 88. Amostras de Validação MCS06

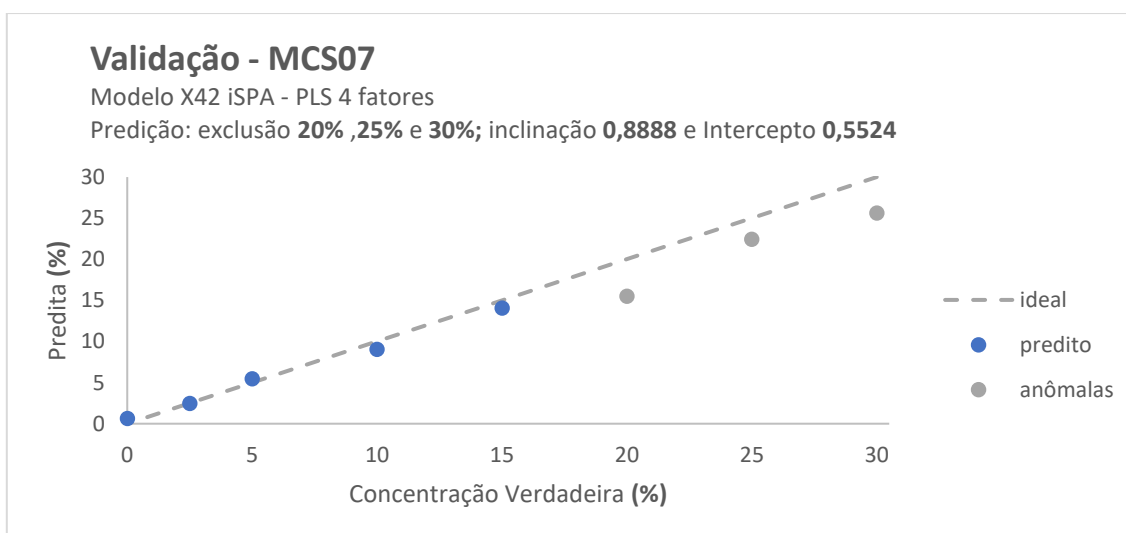


Figura 89. Amostras de Validação MCS07

Os gráficos representados trazem também a linha de idealidade da relação entre as concentrações verdadeira e predita. A inclinação dessa linha é igual a unidade e seu intercepto extrapolado para atingir a concentração zero. Os pontos representados em azul são as amostras de validação cuja concentração fora predita e considerada na avaliação de adequação do modelo. As amostras representadas com pontos cinza são as interpretadas pelos modelos como anômalas, sendo assim desconsideradas da estimativa de tendência e validação do modelo.

Os modelos representados possuem habilidade promissora, apesar dos desvios encontrados em algumas condições. As misturas contendo mais de uma garrafa de Cabernet Sauvignon apresentaram regiões de anomalia (Figura 89), ou até um deslocamento ou tendência (Figura 87 e Figura 88, respectivamente). Todavia, principalmente pelos modelos das Figura 83, Figura 84 e Figura 85, pode-se perceber a capacidade de predição correta da concentração de vinhos em misturas na aplicação de dados FTIR e ferramentas quimiométricas.



Por Anelise Maria Wasem Pires Schneider

5. CONCLUSÕES

A indicação de procedência representa um dos fatores de diferenciação na qualidade de produtos como o vinho. Ao valorizar a vitivinicultura local e especializada possibilita-se o crescimento econômico baseado no reconhecimento cultural, social e tecnológico. Para tanto, é necessário desenvolver metodologias que facilitem a extração de informação pertinente para o controle de qualidade, haja vista a constante adaptabilidade das fraudes e a crescente pressão imposta pela lógica da competitividade.

Apesar de sua potencialidade, além de economicamente restritivas, as áreas tradicionais da química analítica são subutilizadas e ecologicamente não sustentáveis. Por isso, no presente trabalho se estudou a performance na classificação e na regressão de vinhos tintos gaúchos ao empregar espectroscopia no infravermelho, uma técnica analítica secundária, associada à quimiometria.

Amostras de vinho tinto produzidas no Estado do Rio Grande do Sul foram analisadas mediante FTIR – ATR sob ferramentas de reconhecimento de padrões e regressão. Os resultados obtidos no presente trabalho mostram que essas duas técnicas possuem ação sinérgica nas finalidades de classificação e quantificação e, assim, sugerem uma continuidade dos estudos com amostras reais em quantidades mais expressivas e controladas para assegurar repetibilidade, reprodutibilidade e robustez.

Nos estudos de reconhecimento de padrões via PCA se percebe formações de agrupamentos que diferenciam as safras analisadas. A pluralidade amostral nas classes de origem geográfica pode ter dificultado a extração de informação suficiente para separação desses grupos pelo método não-supervisionado, tanto nos dados brutos quanto nos pré-processados. Contudo, a leve tendência de formação de agrupamento Campanha em relação à Serra Gaúcha sugere continuidade de testes com um conjunto amostral seletivo ou mais representativo.

Quando aplicado aos métodos supervisionados, o conjunto amostral pôde ser classificado com relação à origem geográfica na etapa de teste a uma exatidão de 81% via PLS – DA e 71% por SPA – LDA. O algoritmo genético, por sua vez, também apresentou exatidão de 81% em condições controladas de tolerância na aceitação das variáveis classificatórias de maior frequência. Porém, o custo computacional exigido pelo Algoritmo Genético o desclassifica das metodologias analíticas alternativas rápidas.

Dentre os modelos de classificação formados pelos três métodos supervisionados, a diferença em sensibilidade, especificidade e precisão entre Campanha e Serra Gaúcha é perceptível. Na tentativa de padronizar o conjunto, segregando amostras minoritárias, o grupo amostral total foi reduzido a apenas variedades que apresentavam ao menos um exemplar por safra e região. Todavia, a diferença em desempenho (superior para a Serra Gaúcha) se manteve, assim como a proporção amostral.

Nesse contexto, com o intuito de contornar o efeito da desproporção entre classes, adotou-se uma sistemática que primeiramente valorizasse o desempenho da Campanha, ainda considerando as melhores condições possíveis para a Serra Gaúcha. Assim, os melhores modelos de classificação foram aqueles pré-processados por derivada de Savitzky Golay, com janela de alisamento contendo entre 21 e 31 pontos (SPA – LDA e PLS – DA, respectivamente), demonstrando a necessidade de tratamento do sinal para correções de desvios da linha de base, possivelmente causadas pela inconstância de temperatura e umidade da infraestrutura disponível para a prática.

Ainda nesses termos, na análise por FTIR – ATR se observou uma média espectral que aponta uma maior concentração de álcoois e de compostos polifenólicos na Campanha, dado suportado na literatura como consequência dos níveis regionais de precipitação anual e de insolação durante a maturação dos frutos. Ou seja, apesar de toda variabilidade intrínseca da produção do vinho aqui analisado, as ferramentas da quimiometria preservaram, acentuaram e, eventualmente, confirmaram a segregação das classes como efeito das diferenças químicas causadas pelas diferenças geográficas.

Adicionalmente, como no estudo para reconhecimento de padrões, os estudos de regressão também precisaram considerar etapas de pré-processamento de sinais na tentativa de suprimir os efeitos de variabilidades das condições ambientais e de execução do processo de aquisição do espectro. As derivadas de Savitzky Golay foram aplicadas e seus efeitos avaliados nas etapas de validação cruzada dos modelos construídos. Uma função multiobjetivo foi criada para fundamentar a escolha dos melhores modelos, aos quais as amostras de teste foram aplicadas para validar o sistema.

A aplicação de dados de espectroscopia por FTIR aliada a ferramentas da quimiometria também se mostrou uma alternativa promissora para quantificar a variação em concentração de monovarietais em misturas de vinhos. Na faixa de adição prevista

legalmente, a composição de Cabernet Sauvignon em Merlot foi corretamente predita pela maioria dos modelos de regressão selecionados, com destaque aos com seleção de variáveis, iSPA – PLS e iPLS, que apresentaram aproximadamente 12% de erro de predição.

As variáveis cruciais para o desempenho adequado dos modelos de regressão se assemelham as de reconhecimento de padrões. Frequentemente, a região do espectro selecionada pelos algoritmos foi aquela que corresponde as vibrações C – O de álcoois. Assim, o FTIR e os métodos iSPA – PLS e iPLS foram capazes de quantificar concentrações de misturas de monovarietais por meio da diferença de grau alcoólico particular de cada uva.

Para estudos futuros em que se aplique métodos de regressão, novas condições de análise precisam ser consideradas. Entre essas estão as interações entre grupos amostrais de treinamento e validação de dias distintos, bem como a composição de grupos de treinamento maiores, com registros espectrais de dias diferentes. Adicionalmente, para validar reprodutibilidade e robustez da metodologia, novas amostras desses e de outros varietais precisam ser analisadas seguindo o método proposto para quantificação.

Por tanto, as técnicas integradas da espectroscopia no infravermelho médio e ferramentas quimiométricas se mostram promissoras à caracterização e quantificação do perfil de vinhos tintos do Rio Grande do Sul. Superando expectativas de uma técnica analítica tida como de segunda ordem, a metodologia proposta tem potencial para continuidade de estudos de aplicação, visando sustentabilidade e acesso de execução na obtenção de informação segura e democrática.

6. PERSPECTIVAS FUTURAS

Primeiramente, a continuidade da parceria e fornecimento de amostras de vinhos pela Embrapa Uva e Vinho é primordial para o progresso no melhoramento das metodologias propostas no presente trabalho. Essa parceria contribui com o interesse da sociedade como um todo, considerando que seus resultados agregam às possibilidades de controle de qualidade e reconhecimento e valorização do vinho produzido nesta terra.

Os dados aqui estudados podem ser submetidos a ferramentas não lineares da quimiometria. Considerando a diversidade de tipos de uva que compõem os vinhos gaúchos, o desempenho no reconhecimento de padrões precisa ser avaliado por métodos que superem esse incremento ao conjunto de variáveis, já importante devido ao reduzido conjunto amostral disponível.

Ainda sobre os dados, esses podem ser agregados a outras técnicas analíticas, para avaliar os efeitos em metodologias conhecidas por fusão de dados. Uma vez apresentando um efeito sinérgico melhor, a fusão de dados acrescentaria informações úteis ao reconhecimento de padrões de distinção entre regiões ou ainda ao melhor entendimento da composição do vinho local.

Por fim, uma remodelagem dos dados estudados pode considerar a safra de procedência principalmente nas técnicas supervisionadas. Ao incrementar amostras de safras futuras, as novas análises poderiam inspecionar a diferença conjunta entre regiões de produção e safra.

REFERÊNCIAS

1. Brasil. *Decreto nº 8198*. (2014).
2. OIV. *International code of oenological practices*. (2020).
3. Phillips, R. *Uma Breve História do Vinho*. (Editora Record, 2014).
4. This, P., Lacombe, T. & Thomas, M. R. Historical origins and genetic diversity of wine grapes. *Trends in Genetics* **22**, 511–519 (2006).
5. Dresch, R. R. *et al.* Phenolic Compounds from the Leaves of *Vitis labrusca* and *Vitis vinifera* L. as a Source of Waste Byproducts: Development and Validation of LC Method and Antichemotactic Activity. *Food Anal. Methods* **7**, 527–539 (2014).
6. Goldstein, E. *Wines of South America*. (University of California Press, 2014).
7. OIV. *RESOLUTION OIV/VITI 333/2010*. (2010).
8. Cejudo-Bastante, M. J., del Barrio-Galán, R., Heredia, F. J., Medel-Marabolí, M. & Peña-Neira, Á. Location effects on the polyphenolic and polysaccharidic profiles and colour of Carignan grape variety wines from the Chilean Maule region. *Food Res. Int.* **106**, 729–735 (2018).
9. Mazzei, P., Francesca, N., Moschetti, G. & Piccolo, A. NMR spectroscopy evaluation of direct relationship between soils and molecular composition of red wines from Aglianico grapes. *Anal. Chim. Acta* **673**, 167–172 (2010).
10. Pisano, P. L., Silva, M. F. & Olivieri, A. C. Exploration of liquid chromatographic-diode array data for Argentinean wines by extended multivariate curve resolution. *Chemom. Intell. Lab. Syst.* **132**, 1–7 (2014).
11. Barbará, J. A. *et al.* Maturation and maceration effects on tropical red wines assessed by chromatography and analysis of variance - Principal component analysis. *J. Braz. Chem. Soc.* **30**, 1357–1377 (2019).
12. de Villiers, A., Alberts, P., Tredoux, A. G. J. & Nieuwoudt, H. H. Analytical techniques for wine analysis: An African perspective; a review. *Anal. Chim. Acta* **730**, 2–23 (2012).

13. dos Santos, C. A. T., Páscoa, R. N. M. J. & Lopes, J. A. A review on the application of vibrational spectroscopy in the wine industry: From soil to bottle. *TrAC - Trends in Analytical Chemistry* **88**, 100–118 (2017).
14. Moreno-Arribas, M. V. & Polo, M. C. *Wine chemistry and biochemistry. Wine Chemistry and Biochemistry* (2009). doi:10.1007/978-0-387-74118-5
15. González-Barreiro, C., Rial-Otero, R., Cancho-Grande, B. & Simal-Gándara, J. Wine Aroma Compounds in Grapes: A Critical Review. *Crit. Rev. Food Sci. Nutr.* **55**, 202–218 (2015).
16. Garde-Cerdán, T. *et al.* Implications of nitrogen compounds during alcoholic fermentation from some grape varieties at different maturation stages and cultivation systems. *Food Chem.* **124**, 106–116 (2011).
17. Coelho, E. M. *et al.* Simultaneous analysis of sugars and organic acids in wine and grape juices by HPLC: Method validation and characterization of products from northeast Brazil. *J. Food Compos. Anal.* **66**, 160–167 (2018).
18. Soleas, G. J., Diamandis, E. P. & Goldberg, D. M. Wine as a biological fluid: History, production, and role in disease prevention. *J. Clin. Lab. Anal.* **11**, 287–313 (1997).
19. Chemistry, E. *Enological Chemistry. Enological Chemistry* (2012). doi:10.1016/c2011-0-69661-9
20. Fragoso, S., Aceña, L., Guasch, J., Mestres, M. & Busto, O. Quantification of phenolic compounds during red winemaking using FT-MIR spectroscopy and PLS-regression. *J. Agric. Food Chem.* **59**, 10795–10802 (2011).
21. Mazzei, P., Francesca, N., Moschetti, G. & Piccolo, A. NMR spectroscopy evaluation of direct relationship between soils and molecular composition of red wines from Aglianico grapes. *Anal. Chim. Acta* **673**, 167–172 (2010).
22. De Andrade, R. H. S., Do Nascimento, L. S., Pereira, G. E., Hallwass, F. & Paim, A. P. S. Anthocyanic composition of Brazilian red wines and use of HPLC-UV-Vis associated to chemometrics to distinguish wines from different regions. *Microchem. J.* **110**, 256–262 (2013).

23. Cook, N. C. & Samman, S. Flavonoids - Chemistry, metabolism, cardioprotective effects, and dietary sources. *J. Nutr. Biochem.* **7**, 66–76 (1996).
24. Heras-Roger, J., Díaz-Romero, C. & Darias-Martín, J. What Gives a Wine Its Strong Red Color? Main Correlations Affecting Copigmentation. *J. Agric. Food Chem.* **64**, 6567–6574 (2016).
25. Kähkönen, M. P., Heinämäki, J., Ollilainen, V. & Heinonen, M. Berry anthocyanins: Isolation, identification and antioxidant activities. *J. Sci. Food Agric.* **83**, 1403–1411 (2003).
26. Sivilotti, P. *et al.* Impact of Leaf Removal, Applied before and after Flowering, on Anthocyanin, Tannin, and Methoxypyrazine Concentrations in Merlot' (*Vitis vinifera* L.) Grapes and Wines. *J. Agric. Food Chem.* **64**, 4487–4496 (2016).
27. Koak, J. H. *et al.* Blending of different domestic grape wines using mixture design and optimization technique. *Food Sci. Biotechnol.* **19**, 1011–1018 (2010).
28. Guerra, C. C. & Zanus, M. C. Uvas Viníferas para Processamento em Regiões de Clima Temperado. *Embrapa Uva e Vinho* 1–3 (2003). Disponível em: <https://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Uva/UvasViniferasRegioesClimaTemperado/colheita.htm>. (Accessed: 26th November 2020)
29. Sartori, G. V. Maturação fenólica de uvas tintas cultivadas no Rio grande do Sul. (UFSM, 2011). <https://repositorio.ufsm.br/handle/1/5712>
30. Arenhart, M. Caracterização físico-química, fenólica e sensorial da cv. Marselan de diferentes regiões do Rio Grande do Sul. (UFSM, 2015). <https://repositorio.ufsm.br/bitstream/handle/1/5786/ARENHART%2C%20MARCIA.pdf?sequence=1&isAllowed=y>
31. Guerra, C., Mandelli Francisco, Tonietto, J., Zanus, M. C. & Camargo, U. A. *Conhecendo o essencial do vinho. Embrapa* **1**, (2005).
32. Basalekou, M. *et al.* Authenticity Determination of Greek-Cretan Mono-Varietal White and Red Wines Based on their Phenolic Content Using Attenuated Total Reflectance Fourier Transform Infrared Spectroscopy and Chemometrics. *Curr. Res.*

- Nutr. Food Sci. J.* **4**, 54–62 (2016).
33. Belmiro, T. M. C., Pereira, C. F. & Paim, A. P. S. Red wines from South America: Content of phenolic compounds and chemometric distinction by origin. *Microchem. J.* **133**, 114–120 (2017).
 34. Wrege, M. S. *et al.* *Atlas Climático do Rio Grande do Sul.* (2011). <https://www.agricultura.rs.gov.br/upload/arquivos/202005/13110034-atlas-climatico-rs.pdf>
 35. Niederle, P. A. *Indicações Geográficas: qualidade e origem nos mercados alimentares.* (Editora da UFRGS, 2013).
 36. Brasil. *Lei N.º 9279, de 14 de maio de 1996.* 1–20 (1996).
 37. IBRAVIN. Brasil Vitivinícola - Regiões Produtoras. (2019). Disponível em: <https://www.ibravin.org.br/Regioes-Produtoras>. (Accessed: 19th May 2020)
 38. Tuon, L. Acordo Mercosul-UE pode deixar vinho europeu até 30 % mais barato no Brasil. *Exame* (2020). Disponível em: <https://exame.com/economia/acordo-mercosul-ue-pode-deixar-vinho-europeu-ate-30-mais-barato-no-brasil/>. (Accessed: 8th December 2020)
 39. Brigada Militar. Em duas ações da Operação Hórus BM apreende grande quantidade de vinho em Tuparendi. (2020). Disponível em: <https://www.brigadamilitar.rs.gov.br/em-duas-acoes-da-operacao-horus-bm-apreende-grande-quantidade-de-vinho-em-tuparendi>. (Accessed: 8th December 2020)
 40. Mello, L. M. R. de. *Viticultura brasileira : panorama 2018.* (2019).
 41. Hamm, G. & Flach, M. Setor vitivinícola prevê crescimento com alteração na cobrança de tributo. *Governo do Rio Grande do Sul* (2019). Disponível em: <https://estado.rs.gov.br/setor-vitivinicola-preve-crescimento-com-alteracao-na-cobranca-de-tributo>. (Accessed: 9th December 2020)
 42. Sperb, P. Entre robôs e tradição, cultura do vinho gira R\$ 3 bilhões no RS. *Folhapress* 1–3 (2020).

43. Sebrae & INPI. *Indicações geográficas brasileiras: vinho*. (2016).
44. INPI. Campanha Gaúcha ganha indicação geográfica para vinhos e espumantes. *INPI* 1 (2020). Disponível em: <http://www.inpi.gov.br/noticias/campanha-gaucha-ganha-indicacao-geografica-para-vinhos-e-espumantes>. (Accessed: 13th May 2020)
45. Associação da Campanha Gaúcha. Regulamento de Uso da Indicação de Procedência Campanha Gaúcha. 14 (2019).
46. PROGOETHE. Regulamento de Uso de Indicação de Procedência Vales da Uva Goethe. 1–9 (2012).
47. Tonietto, J., Zanús, M. C., Falcade, I. & Guerra, C. C. *O Regulamento de Uso da indicação geográfica Pinto Bandeira*. (Embrapa, 2013).
48. Tonietto, J., Zanús, M. C., Falcade, I. & Guerra, C. C. *O Regulamento de Uso da indicação geográfica Altos Montes*. (Embrapa, 2013).
49. Tonietto, J., Zanús, M. C., Falcade, I. & Guerra, C. C. *O Regulamento de Uso da denominação de origem Vale dos Vinhedos*. (Embrapa, 2013).
50. Tonietto, J., Zanús, M. C., Falcade, I. & Guerra, C. C. *O Regulamento de Uso da indicação geográfica Monte Belo*. (Embrapa, 2016).
51. Tonietto, J., Zanús, M. C., Falcade, I., Taffarel, J. C. & Guerra, C. C. *O Regulamento de Uso da indicação geográfica Farroupilha*. (Embrapa, 2017).
52. APROBELO. Regulamento de Uso de Indicação de Procedência Monte Belo. 11 (2012).
53. ASPROVINHO. Regulamento de Uso da Indicação de Procedência Pinto Bandeira. 9 (2008).
54. APROMONTES. Regulamento de Uso da Indicação de Procedência Altos Montes. 1–9 (2012).
55. APROVALE. Regulamento de Indicação de Procedência Vale dos Vinhedos. 1–8 (2001).

56. Brasil. *Lei de nº 7678, de 8 de novembro de 1988*. 16 (1988).
57. Bruch, K. L. *Lei do Vinho Sistematizada*. (SEBRAE, 2012).
58. Brasil. *Lei Nº 10970, de 12 de novembro de 2004*. 1–2 (2004).
59. Brasil. *Decreto nº 5.863*. 5 (2006).
60. Brasil. *Consolidação das Normas de Bebidas, Fermentado Acético, Vinho e Derivados da Uva e do Vinho*. 813 (2020).
61. Brasil & Ministério da Agricultura. *Portaria nº 1012, de 17 de novembro de 1978*. 3 (1978).
62. Brasil. *Instrução Normativa nº14, de 8 de fevereiro de 2018*. 4–6 (2018).
63. Brasil. *Instrução Normativa nº 24, de 8 de setembro de 2005*. 80 (2005).
64. Kamiloglu, S. Authenticity and traceability in beverages. *Food Chem.* **277**, 12–24 (2019).
65. Porto, L. F. de A., Lopes, M. A. & Zambalde, A. L. Desenvolvimento de um sistema de rastreabilidade aplicado à cadeia de produção do vinho. *Ciência e Agrotecnologia* **31**, 1310–1319 (2007).
66. Rizzon, L. A. & Salvador, M. B. G. *Metodologia para análise de vinho*. (2010).
67. Cuadros-Rodríguez, L., Ruiz-Samblás, C., Valverde-Som, L., Pérez-Castaño, E. & González-Casado, A. Chromatographic fingerprinting: An innovative approach for food ‘identification’ and food authentication - A tutorial. *Anal. Chim. Acta* **909**, 9–23 (2016).
68. Ballin, N. Z. & Laursen, K. H. To target or not to target? Definitions and nomenclature for targeted versus non-targeted analytical food authentication. *Trends Food Sci. Technol.* **86**, 537–543 (2019).
69. Versari, A., Laurie, V. F., Ricci, A., Laghi, L. & Parpinello, G. P. Progress in authentication, typification and traceability of grapes and wines by chemometric approaches. *Food Res. Int.* **60**, 2–18 (2014).

70. Daniel, C. & Smyth, H. *Analytical and chemometric-based methods to monitor and evaluate wine protected designation*. *Comprehensive Analytical Chemistry* **60**, (Copyright © 2013 Elsevier B.V. All rights reserved., 2013).
71. Ebeler, S. E., Takeoka, G. R. & Winterhalter, P. Progress in authentication of food and wine. *ACS Symp. Ser.* **1081**, (2011).
72. Ghanem, E. *et al.* Predicting the composition of red wine blends using an array of Multicomponent peptide-based sensors. *Molecules* **20**, 9170–9182 (2015).
73. Villano, C. *et al.* Wine varietal authentication based on phenolics, volatiles and DNA markers: State of the art, perspectives and drawbacks. *Food Control* **80**, 1–10 (2017).
74. Brasil. *Instrução normativa nº 4 de 05 de fevereiro de 2001*. 5–8 (2001).
75. Brasil. *Instrução Normativa nº 10, de 14 de abril de 2009*. *Diário Oficial da União* 1–2 (2009).
76. OIV. *RESOLUTION OIV-OENO 353-2009*. 1–9 (2009).
77. OIV. *RESOLUTION OIV-OENO 479-2017*. 1–10 (2014).
78. Roßmann, A. *et al.* Stable oxygen isotope content of water of EU data-bank wines from Italy, France and Germany. *Eur. Food Res. Technol.* **208**, 400–407 (1999).
79. Primrose, S., Woolfe, M. & Rollinson, S. Food forensics: Methods for determining the authenticity of foodstuffs. *Trends Food Sci. Technol.* **21**, 582–590 (2010).
80. Dutra, S. V. *et al.* Characterization of wines according the geographical origin by analysis of isotopes and minerals and the influence of harvest on the isotope values. *Food Chem.* **141**, 2148–2153 (2013).
81. Geana, E. I. *et al.* Verifying the red wines adulteration through isotopic and chromatographic investigations coupled with multivariate statistic interpretation of the data. *Food Control* **62**, 1–9 (2016).
82. Vinciguerra, V. *et al.* Strontium isotope characterization of wines from Quebec, Canada. *Food Chem.* **210**, 121–128 (2016).

83. Bora, F. D. *et al.* Elemental content and lead-strontium isotope characterization of wine. *Stud. Univ. Babeş-Bolyai Chem.* **63**, 137–155 (2018).
84. Coelho, E. M. *et al.* Simultaneous analysis of sugars and organic acids in wine and grape juices by HPLC: Method validation and characterization of products from northeast Brazil. *J. Food Compos. Anal.* **66**, 160–167 (2018).
85. Serrano-Lourido, D., Saurina, J., Hernández-Cassou, S. & Checa, A. Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis. *Food Chem.* **135**, 1425–1431 (2012).
86. Ziólkowska, A., Wąsowicz, E. & Jeleń, H. H. Differentiation of wines according to grape variety and geographical origin based on volatiles profiling using SPME-MS and SPME-GC/MS methods. *Food Chem.* **213**, 714–720 (2016).
87. Sagandykova, G. N., Alimzhanova, M. B., Nurzhanova, Y. T. & Kenessov, B. Determination of semi-volatile additives in wines using SPME and GC-MS. *Food Chem.* **220**, 162–167 (2017).
88. Brasil. RDC n° 42 de 29 de Agosto de 2013. *Diário Oficial da União* 33 (2013).
89. Pohl, P. What do metals tell us about wine? *TrAC - Trends Anal. Chem.* **26**, 941–949 (2007).
90. McGrath, T. F. *et al.* What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed? – Spectroscopy case study. *Trends Food Sci. Technol.* **76**, 38–55 (2018).
91. González, A., Llorens, A., Cervera, M. L., Armenta, S. & de la Guardia, M. Elemental fingerprint of wines from the protected designation of origin Valencia. *Food Chem.* **112**, 26–34 (2009).
92. Fabani, M. P. *et al.* Evaluation of elemental profile coupled to chemometrics to assess the geographical origin of Argentinean wines. *Food Chem.* **119**, 372–379 (2010).
93. Soares, F. *et al.* Element selection and concentration analysis for classifying South

- America wine samples according to the country of origin. *Comput. Electron. Agric.* **150**, 33–40 (2018).
94. Sádecká, J. & Tóthová, J. Fluorescence spectroscopy and chemometrics in the food classification - A review. *Czech J. Food Sci.* **25**, 159–173 (2007).
 95. Suciú, R. C., Zarbo, L., Guyon, F. & Magdas, D. A. Application of fluorescence spectroscopy using classical right angle technique in white wines classification. *Sci. Rep.* **9**, 1–10 (2019).
 96. Airado-Rodríguez, D., Galeano-Díaz, T., Durán-Merás, I. & Wold, J. P. Usefulness of fluorescence excitation-emission matrices in combination with parafac, as fingerprints of red wines. *J. Agric. Food Chem.* **57**, 1711–1720 (2009).
 97. Tan, J. *et al.* Geographical classification of Chinese Cabernet Sauvignon wines by data fusion of ultraviolet–visible and synchronous fluorescence spectroscopies: the combined use of multiple wavelength differences. *Aust. J. Grape Wine Res.* **22**, 358–365 (2016).
 98. Aleixandre-Tudo, J. L., Nieuwoudt, H., Olivieri, A., Aleixandre, J. L. & du Toit, W. Phenolic profiling of grapes, fermenting samples and wines using UV-Visible spectroscopy with chemometrics. *Food Control* **85**, 11–22 (2018).
 99. Martelo-Vidal, M. J. & Vázquez, M. Classification of red wines from controlled designation of origin by ultraviolet-visible and near-infrared spectral analysis. *Cienc. e Tec. Vitivinic.* **29**, 35–43 (2014).
 100. Cozzolino, D. *et al.* Relationship between wine scores and visible-near-infrared spectra of Australian red wines. *Anal. Bioanal. Chem.* **391**, 975–981 (2008).
 101. Martelo-Vidal, M. J. & Vázquez, M. Evaluation of ultraviolet, visible, and near infrared spectroscopy for the analysis of wine compounds. *Czech J. Food Sci.* **32**, 37–47 (2014).
 102. Sen, I. & Tokatli, F. Authenticity of wines made with economically important grape varieties grown in Anatolia by their phenolic profiles. *Food Control* **46**, 446–454 (2014).

103. Musingarabwi, D. M., Nieuwoudt, H. H., Young, P. R., Eyéghè-Bickong, H. A. & Vivier, M. A. A rapid qualitative and quantitative evaluation of grape berries at various stages of development using Fourier-transform infrared spectroscopy and multivariate data analysis. *Food Chem.* **190**, 253–262 (2016).
104. Genisheva, Z. *et al.* New PLS analysis approach to wine volatile compounds characterization by near infrared spectroscopy (NIR). *Food Chem.* **246**, 172–178 (2018).
105. Banc, R., Loghin, F., Miere, D., Fetea, F. & Socaciu, C. Romanian wines quality and authenticity using FT-MIR spectroscopy coupled with multivariate data analysis. *Not. Bot. Horti Agrobot. Cluj-Napoca* **42**, 556–564 (2014).
106. Esteki, M. *et al.* A review on the application of chromatographic methods, coupled to chemometrics, for food authentication. *Food Control* **93**, 165–182 (2018).
107. Halliday, D., Resnick, R. & Walker, J. *Fundamentos de Física.* **4**, (LTC, 2012).
108. Encyclopaedia Britannica. Electromagnetic Spectrum. (2019). Disponível em: <https://www.britannica.com/science/electromagnetic-spectrum>. (Accessed: 4th June 2020)
109. Skoog, D. A., West, D. M., Holler, F. J. & Crouch, S. R. *Fundamentos de Química Analítica.* (Thomson, 2006).
110. Barbosa, L. C. A. *Espectroscopia no infravermelho na caracterização de compostos Orgânicos.* (Edidora UFV, 2007).
111. Bevin, C. J., Dambergs, R. G., Fergusson, A. J. & Cozzolino, D. Varietal discrimination of Australian wines by means of mid-infrared spectroscopy and multivariate analysis. *Anal. Chim. Acta* **621**, 19–23 (2008).
112. Da Silva, F. E. B., Flores, É. M. M., Parisotto, G., Müller, E. I. & Ferrão, M. F. Green method by diffuse reflectance infrared spectroscopy and spectral region selection for the quantification of sulphamethoxazole and trimethoprim in pharmaceutical formulations. *An. Acad. Bras. Cienc.* **88**, 1–15 (2016).

113. Ferro, L., Gojkovic, Z. & Funk, C. Statistical Methods for Rapid Quantification of Proteins , Lipids , and Carbohydrates in Nordic. *Molecules* **2019**, (2019).
114. Fanelli, S., Zimmermann, A., Totóli, E. G. & Salgado, H. R. N. FTIR Spectrophotometry as a Green Tool for Quantitative Analysis of Drugs: Practical Application to Amoxicillin. *J. Chem.* **2018**, (2018).
115. Laghi, L., Versari, A., Parpinello, G. P., Nakaji, D. Y. & Boulton, R. B. FTIR Spectroscopy and Direct Orthogonal Signal Correction Preprocessing Applied to Selected Phenolic Compounds in Red Wines. *Food Anal. Methods* **4**, 619–625 (2011).
116. Silva, S. D., Feliciano, R. P., Boas, L. V. & Bronze, M. R. Application of FTIR-ATR to Moscatel dessert wines for prediction of total phenolic and flavonoid contents and antioxidant capacity. *Food Chem.* **150**, 489–493 (2014).
117. Sigma-Aldrich. IR Spectrum Table and Chart. 1–5 (2019). Disponível em: <https://www.sigmaaldrich.com/technical-documents/articles/biology/ir-spectrum-table.html>. (Accessed: 9th June 2020)
118. Versari, A., Parpinello, G. P., Scazzina, F. & Rio, D. Del. Prediction of total antioxidant capacity of red wine by Fourier transform infrared spectroscopy. *Food Control* **21**, 786–789 (2010).
119. Ferreira, M. M. C. *Quimiometria*. (Editora da Unicamp, 2015).
120. Bradley, M. S. & ThermoFisher Scientific. *FT-IR Sample Analysis*.
121. Pasquini, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* **14**, 198–219 (2003).
122. Valand, R., Tanna, S., Lawson, G. & Bengtström, L. A review of Fourier Transform Infrared (FTIR) spectroscopy used in food adulteration and authenticity investigations. *Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess.* **37**, 19–38 (2020).
123. Shimadzu Corporation. *Spectral characteristics dependent on ATR Crystal selection - Differences in properties (shape, hardness, refractive index) according to sample -*.

- (2015).
124. Neto, B. D. B., Scarminio, I. S. & Bruns, R. E. 25 Anos de quimiometria no Brasil. *Quim. Nova* **29**, 1401–1406 (2006).
 125. Fadock, M., Brown, R. B. & Reynolds, A. G. Visible-near infrared reflectance spectroscopy for nondestructive analysis of red wine grapes. *Am. J. Enol. Vitic.* **67**, 38–45 (2016).
 126. Fernández, K. & Agosin, E. Quantitative analysis of red wine tannins using Fourier-transform mid-infrared spectrometry. *J. Agric. Food Chem.* **55**, 7294–7300 (2007).
 127. Alamar, P. D., Caramês, E. T. S., Poppi, R. J. & Pallone, J. A. L. Detection of Fruit Pulp Adulteration Using Multivariate Analysis: Comparison of NIR, MIR and Data Fusion Performance. *Food Anal. Methods* (2020). doi:10.1007/s12161-020-01755-x
 128. GALVAO, R. *et al.* A method for calibration and validation subset partitioning. *Talanta* **67**, 736–740 (2005).
 129. De Araújo Gomes, A., Galvão, R. K. H., De Araújo, M. C. U., Vêras, G. & Da Silva, E. C. The successive projections algorithm for interval selection in PLS. *Microchem. J.* **110**, 202–208 (2013).
 130. Agatonovic-Kustrin, S., Milojković-Opsenica, D., Morton, D. W. & Ristivojević, P. Chemometric characterization of wines according to their HPTLC fingerprints. *Eur. Food Res. Technol.* **243**, 659–667 (2017).
 131. Majchrzak, T., Wojnowski, W. & Płotka-Wasyłka, J. Classification of Polish wines by application of ultra-fast gas chromatography. *Eur. Food Res. Technol.* **244**, 1463–1471 (2018).
 132. de Lima, C. M. *et al.* Digital image-based tracing of geographic origin, winemaker, and grape type for red wine authentication. *Food Chem.* **312**, (2020).
 133. Bro, R. & Smilde, A. K. Principal component analysis. *Anal. Methods* **6**, 2812–2831 (2014).
 134. Marini, F. Classification Methods in Chemometrics. *Curr. Anal. Chem.* **6**, 72–79

- (2009).
135. *Chemometrics in Practical Applications*. (InTech, 2012).
 136. Ballabio, D. & Consonni, V. Classification tools in chemistry. Part 1: linear models. PLS-DA. *Anal. Methods* **5**, 3790 (2013).
 137. Lee, L. C., Liong, C. Y. & Jemain, A. A. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: A review of contemporary practice strategies and knowledge gaps. *Analyst* **143**, 3526–3539 (2018).
 138. Martinez, W. L. & Martinez, A. R. *Computational Statistics Handbook with Matlab*. (Chapman & Hall, 2008).
 139. Paulino, C. D. *et al.* *Glossário Inglês-Português de Estatística*. (Sociedade Portuguesa de Estatística e Associação Brasileira de Estatística, 2011).
 140. Medina, S., Pereira, J. A., Silva, P., Perestrelo, R. & Câmara, J. S. Food fingerprints – A valuable tool to monitor food authenticity and safety. *Food Chem.* **278**, 144–162 (2019).
 141. Teixeira dos Santos, C. A. *et al.* Raman spectroscopy for wine analyses: A comparison with near and mid infrared spectroscopy. *Talanta* **186**, 306–314 (2018).
 142. Páscoa, R. N. M. J., Porto, P. A. L. S., Cerdeira, A. L. & Lopes, J. A. The application of near infrared spectroscopy to wine analysis: An innovative approach using lyophilization to remove water bands interference. *Talanta* **214**, (2020).
 143. Oliveri, P. & Simonetti, R. Chemometrics for Food Authenticity Applications. in *Advances in Food Authenticity Testing* 702–728 (Elsevier Ltd, 2016). doi:10.1016/b978-0-08-100220-9.00025-4
 144. Berrueta, L. A., Alonso-Salces, R. M. & Héberger, K. Supervised pattern recognition in food analysis. *J. Chromatogr. A* **1158**, 196–214 (2007).
 145. Malakar, G. P. Linear Discriminant Analysis vs Principal Component Analysis. (2017). Disponível em: <https://youtu.be/M4HpyJHPYBY>. (Accessed: 30th June 2020)

146. Brereton, R. G. & Lloyd, G. R. Partial least squares discriminant analysis : taking the magic away. *J. Chemom.* **48**, 213–225 (2014).
147. Ivanović, M., Petek, A., Razboršek, M. I. & Kolar, M. Chemometric characterization of Slovenian red wines. *Acta Chim. Slov.* **64**, 537–542 (2017).
148. Louw, L. *et al.* Characterization of selected South African young cultivar wines using FT-MIR Spectroscopy, Gas chromatography, and multivariate data analysis. *J. Agric. Food Chem.* **57**, 2623–2632 (2009).
149. Monakhova, Y. B. & Mushtakova, S. P. Methodology of chemometric modeling of spectrometric signals in the analysis of complex samples. *J. Anal. Chem.* **72**, 147–155 (2017).
150. Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2001).
151. Di Egidio, V., Sinelli, N., Giovanelli, G., Moles, A. & Casiraghi, E. NIR and MIR spectroscopy as rapid methods to monitor red wine fermentation. *Eur. Food Res. Technol.* **230**, 947–955 (2010).
152. Andersson, M. A comparison of nine PLS1 algorithms. *J. Chemom.* **23**, 518–529 (2009).
153. Lletí, R., Meléndez, E., Ortiz, M. C., Sarabia, L. A. & Sánchez, M. S. Outliers in partial least squares regression: Application to calibration of wine grade with mean infrared data. *Anal. Chim. Acta* **544**, 60–70 (2005).
154. Michalewicz, Z. *Genetic algorithms + data structures = evolution programs.* (Springer, 1992).
155. Araújo, M. C. U. *et al.* The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemom. Intell. Lab. Syst.* **57**, 65–73 (2001).
156. Friedel, M., Patz, C. D. & Dietrich, H. Comparison of different measurement techniques and variable selection methods for FT-MIR in wine analysis. *Food Chem.* **141**, 4200–4207 (2013).

157. Lletí, R., Ortiz, M. C., Sarabia, L. A. & Sánchez, M. S. Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Anal. Chim. Acta* **515**, 87–100 (2004).
158. Ballabio, D., Grisoni, F. & Todeschini, R. Multivariate comparison of classification performance measures. *Chemom. Intell. Lab. Syst.* **174**, 33–44 (2018).
159. Cuadros-Rodríguez, L., Pérez-Castaño, E. & Ruiz-Samblás, C. Quality performance metrics in multivariate classification methods for qualitative analysis. *TrAC Trends Anal. Chem.* **80**, 612–624 (2016).
160. BPA. Regression for Measurement and Verification: Reference Guide. 54 (2012).
161. Valderrama, P., Braga, J. W. B. & Poppi, R. J. Estado da arte de figuras de mérito em calibração multivariada. *Quim. Nova* **32**, 1278–1287 (2009).
162. Higgins, F. & Rein, A. Significantly improved measurement of liquid samples using the Cary 630 FTIR spectrometer with DialPath accessory. (2011).
163. Agilent Cary 630 FTIR sample interfaces. 6 (2011).
164. Ballabio, D. A MATLAB toolbox for Principal Component Analysis and unsupervised exploration of data structure. *Chemom. Intell. Lab. Syst.* **149**, 1–9 (2015).
165. Urbano-Cuadrado, M., Luque De Castro, M. D., Pérez-Juan, P. M., García-Olmo, J. & Gómez-Nieto, M. A. Near infrared reflectance spectroscopy and multivariate analysis in enology: Determination or screening of fifteen parameters in different types of wines. *Anal. Chim. Acta* **527**, 81–88 (2004).
166. Tarantilis, P. A., Troianou, V. E., Pappas, C. S., Kotseridis, Y. S. & Polissiou, M. G. Differentiation of Greek red wines on the basis of grape variety using attenuated total reflectance Fourier transform infrared spectroscopy. *Food Chem.* **111**, 192–196 (2008).
167. Jensen, J. S., Egebo, M. & Meyer, A. S. Identification of spectral regions for the quantification of red wine tannins with fourier transform mid-infrared spectroscopy. *J. Agric. Food Chem.* **56**, 3493–3499 (2008).

168. Bevin, C. J., Fergusson, A. J., Perry, W. B., Janik, L. J. & Cozzolino, D. Development of a Rapid “Fingerprinting” System for Wine Authenticity by Mid-infrared Spectroscopy. *J. Agric. Food Chem.* **54**, 9713–9718 (2006).

APÊNDICES

Apêndice A – Vinhos Embrapa

| Safras | Origem da Uva | Garrafa | Controle Interno Adotado | | |
|------------|---------------------------|---------|--------------------------|-------------|--------------|
| | | | Rótulo | Procedência | |
| (continua) | | | | | |
| 2017 | Serra Gaúcha | 283 | CS | 01 | Serra |
| 2017 | Serra Gaúcha | 193 | PN | 02 | Serra |
| 2017 | Campanha Gaúcha | 317 | CS | 03 | Campanha |
| 2017 | Serra Gaúcha | 249 | MT | 04 | Serra |
| 2017 | Campanha Gaúcha | 188 | MN | 05 | Campanha |
| 2017 | Campanha Gaúcha | 325 | TO | 06 | Campanha |
| 2017 | Guaporé | 255 | TT | 07 | Serra |
| 2017 | Campanha Gaúcha | 266 | MT | 08 | Campanha |
| 2017 | Vale dos Vinhedos | 295 | MC | 09 | Serra |
| 2017 | Pinto Bandeira | 222 | CF | 10 | Serra |
| 2017 | Serra Gaúcha | 273 | AA | 11 | Serra |
| 2017 | Flores da Cunha | 259 | PS | 12 | Serra |
| 2017 | Alto Feliz | 319 | TO | 13 | Serra |
| 2017 | Alto Feliz | 286 | MC | 14 | Serra |
| 2017 | Vale dos Vinhedos | 262 | PS | 15 | Serra |
| 2017 | Alto Feliz | 332 | TT | 16 | Serra |
| 2017 | Serra Gaúcha | 246 | MN | 17 | Serra |
| 2017 | São Paulo | 248 | CF | 18 | |
| 2017 | Antônio Prado | 301 | MT | 19 | Serra |
| 2017 | Flores da cunha | 346 | PS | 20 | Serra |
| 2017 | Alto Feliz | 290 | MC | 21 | Serra |
| 2017 | São Joaquim | 235 | CS | 22 | |
| 2017 | Alto Feliz | 327 | TO | 23 | Serra |
| 2017 | Campos de Cima da Serra | 203 | PN | 24 | Serra |
| 2017 | Encruzilhada | 251 | MN | 25 | Campanha |
| 2017 | Encruzilhada | 244 | TO | 26 | Campanha |
| 2017 | Campos de Cima da Serra | 227 | MT | 27 | Serra |
| 2017 | Livramento | 252 | TT | 28 | Campanha |
| 2017 | Livramento – Encruzilhada | 213 | AA | 29 | Campanha |
| 2017 | Campanha Gaúcha | 230 | CF | 30 | Campanha |
| 2017 | Serra Sudeste | 202 | PN | 31 | Campanha |
| 2017 | Flores da Cunha | 217 | AA | 32 | Serra |
| 2017 | Vale dos Vinhedos | 250 | CS | 33 | Serra |
| 2017 | Campanha Gaúcha | 261 | AA | 34 | Campanha |
| 2017 | São Valentin | 260 | MN | 35 | Serra |
| 2017 | São Joaquim | 220 | PN | 36 | Serra |
| 2017 | Livramento | 256 | TT | 37 | Campanha |
| 2017 | Mato Perso | 238 | CF | 38 | Serra |
| 2017 | Embrapa | | MT | 39 | Serra |
| 2018 | Campanha Gaúcha | 109 | CS | 40 | Campanha |
| 2018 | Campestre da Serra | 147 | CS | 41 | Serra |
| 2018 | Serra Gaúcha | 211 | MT | 42 | Serra |
| 2018 | Pinto Bandeira | 140 | CF | 43 | Serra |
| 2018 | Alto Feliz | 144 | MC | 44 | Serra |
| 2018 | Serra Gaúcha | 70 | MT | 45 | Serra |
| 2018 | Serra Gaúcha | 101 | MT | 46 | Serra |
| 2018 | Campanha Gaúcha | 209 | CS | 47 | Campanha |

AA – Ancelota

CF – Cabernet Franc

CS – Cabernet Sauvignon

MC – Malbec

MT – Merlot

PN – Pinot Noir

PS – Petit Syrah

TO – Teroldego

TT – Tannat

| Safra | Origem da Uva | Garrafa | Controle Interno Adotado | | |
|------------|----------------------------|---------|--------------------------|-------------|----------|
| | | | Rótulo | Procedência | |
| (continua) | | | | | |
| 2018 | Vale dos Vinhedos | 125 | MC | 48 | Serra |
| 2018 | Caxias do Sul | 49 | CS | 49 | Serra |
| 2018 | Alto Feliz | 158 | MC | 50 | Serra |
| 2018 | Serra Catarinense | 106 | MC | 51 | |
| 2018 | Vale dos Vinhedos | 198 | MT | 52 | Serra |
| 2018 | Vale dos Vinhedos | 180 | MT | 53 | Serra |
| 2018 | Campanha Gaúcha | 133 | CS | 54 | Campanha |
| 2018 | Vale dos Vinhedos | 191 | MT | 55 | Serra |
| 2018 | Campos de Cima da Serra | 123 | MT | 56 | Serra |
| 2018 | Serra Gaúcha | 42 | MT | 57 | Serra |
| 2018 | Serra do Sudeste | 86 | CF | 58 | Campanha |
| 2018 | Campanha Gaúcha | 114 | CF | 59 | Campanha |
| 2018 | Serra do Sudeste | 169 | MN | 60 | Campanha |
| 2018 | Encruzilhada do Sul | 130 | TO | 61 | Campanha |
| 2018 | Campanha Gaúcha | 182 | TO | 62 | Campanha |
| 2018 | Campanha Gaúcha | 214 | TT | 63 | Campanha |
| 2018 | Serra Gaúcha | 142 | CF | 64 | Serra |
| 2018 | Encruzilhada do Sul | 113 | MN | 65 | Campanha |
| 2018 | Serra Gaúcha | 207 | TT | 66 | Serra |
| 2018 | Campanha Gaúcha | 146 | TT | 67 | Campanha |
| 2018 | Serra Gaúcha | 187 | TT | 68 | Serra |
| 2018 | Serra Gaúcha | 38 | TT | 69 | Serra |
| 2018 | Serra Gaúcha | 190 | MN | 70 | Serra |
| 2018 | Campestre da Serra | 96 | AA | 71 | Serra |
| 2018 | Campestre da Serra | 53 | AA | 72 | Serra |
| 2018 | Serra Gaúcha | 193 | MN | 73 | Serra |
| 2018 | Campanha Gaúcha | 79 | CF | 74 | Campanha |
| 2018 | Serra Gaúcha | 157 | CS | 75 | Serra |
| 2018 | Vale dos Vinhedos | 186 | MT | 76 | Serra |
| 2018 | Mato Perso | 121 | MT | 77 | Serra |
| 2018 | Campanha Gaúcha | 73 | CS | 78 | Campanha |
| 2018 | Vale dos Vinhedos | 131 | CS | 79 | Serra |
| 2018 | Campanha Meridional | 108 | CS | 80 | Campanha |
| 2018 | Vale dos Vinhedos | 145 | CS | 81 | Serra |
| 2018 | Guaporé | 74 | TT | 82 | Serra |
| 2018 | Serra Gaúcha | 192 | TT | 83 | Serra |
| 2018 | Campanha | 120 | TT | 84 | Campanha |
| 2018 | Monte Belo do Sul e Caxias | 77 | TT | 85 | Serra |
| 2018 | Campanha Gaúcha | 206 | TT | 86 | Campanha |
| 2018 | Serra Gaúcha | 151 | CS | 87 | Serra |
| 2018 | Campos de Cima da Serra | 95 | CS | 88 | Serra |
| 2018 | Serra Gaúcha | 172 | CS | 89 | Serra |
| 2018 | Serra Gaúcha | 184 | CS | 90 | Serra |
| 2018 | Serra Gaúcha | 117 | CF | 91 | Serra |
| 2018 | Serra Gaúcha | 166 | CS | 92 | Serra |
| 2018 | Serra do Sudeste | 135 | MC | 93 | Campanha |
| 2018 | Serra do Sudeste | 148 | MN | 94 | Campanha |
| 2018 | Campanha Gaúcha | 216 | PN | 95 | Campanha |
| 2018 | Alto Feliz | 200 | TT | 96 | Serra |
| 2018 | Campanha Gaúcha | 129 | PN | 97 | Campanha |

AA – Ancelota

CF – Cabernet Franc

CS – Cabernet Sauvignon

MC – Malbec

MT – Merlot

PN – Pinot Noir

PS – Petit Syrah

TO – Teroldego

TT – Tannat

| Safra | Origem da Uva | Garrafa | Controle Interno Adotado | | |
|-------|-----------------------------|---------|--------------------------|----|-------------|
| | | | Rótulo | | Procedência |
| | | | | | (conclusão) |
| 2018 | Dois Lajeados e Vila Flores | 56 | AA | 98 | Serra |
| 2018 | Serra Gaúcha | 52 | AA | 99 | Serra |

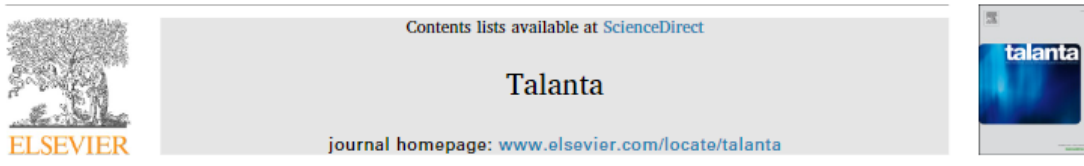
Apêndice B – Datas e Horários no FTIR – ATR

| Amostra | Data | Horário da Replicata | | |
|------------|------------|----------------------|----------|----------|
| (continua) | | | | |
| CS01 | | 13:43:32 | 13:46:11 | 13:48:23 |
| PN02 | | 14:00:03 | 14:02:21 | 14:04:45 |
| CS03 | | 14:14:06 | 14:16:00 | 14:18:01 |
| MT04 | | 14:28:08 | 14:29:52 | 14:31:41 |
| MN05 | | 14:40:09 | 14:42:11 | 14:44:21 |
| TO06 | | 14:53:00 | 14:54:59 | 14:56:44 |
| TT07 | | 15:28:05 | 15:29:48 | 15:31:26 |
| MT08 | | 15:34:24 | 15:37:10 | 15:39:21 |
| MC09 | | 15:42:03 | 15:43:52 | 15:45:50 |
| CF10 | | 15:47:54 | 15:49:48 | 15:51:42 |
| AA11 | | 15:53:55 | 15:56:19 | 15:58:35 |
| PS12 | | 16:01:02 | 16:06:21 | 16:08:41 |
| TO13 | | 16:10:40 | 16:15:48 | 16:17:33 |
| MC14 | | 16:19:54 | 16:21:46 | 16:23:45 |
| PS15 | | 16:26:06 | 16:28:19 | 16:30:06 |
| TT16 | | 16:32:18 | 16:34:00 | 16:35:57 |
| MN17 | | 16:38:10 | 16:40:10 | 16:42:20 |
| MT19 | | 16:49:55 | 16:51:58 | 16:53:53 |
| PS20 | 05/07/2018 | 17:01:47 | 17:03:44 | 17:05:23 |
| MC21 | | 17:07:53 | 17:09:47 | 17:12:01 |
| CS22 | | 17:14:23 | 17:16:15 | 17:18:15 |
| TO23 | | 17:20:22 | 17:22:09 | 17:23:55 |
| PN24 | | 17:28:09 | 17:29:39 | 17:31:07 |
| MN25 | | 17:33:15 | 17:35:57 | 17:38:03 |
| TO26 | | 17:40:07 | 17:41:55 | 17:43:30 |
| MT27 | | 17:45:29 | 17:47:31 | 17:49:06 |
| TT28 | | 17:51:03 | 17:52:36 | 17:54:22 |
| AA29 | | 17:56:25 | 17:57:51 | 17:59:46 |
| CF30 | | 18:01:34 | 18:03:05 | 18:04:51 |
| PN31 | | 18:06:40 | 18:08:18 | 18:10:05 |
| AA32 | | 18:12:17 | 18:14:14 | 18:16:08 |
| CS33 | | 19:02:35 | 19:04:53 | 19:09:58 |
| AA34 | | 19:12:23 | 19:14:26 | 19:16:21 |
| MN35 | | 19:18:28 | 19:20:08 | 19:21:54 |
| PN36 | | 19:21:54 | 19:25:37 | 19:27:25 |
| TT37 | | 19:29:14 | 19:31:00 | 19:33:17 |
| CF38 | | 19:35:03 | 19:36:39 | 19:38:16 |
| CS40 | | 10:47:24 | 10:49:41 | 10:49:41 |
| CS41 | | 10:54:20 | 11:07:49 | 11:09:17 |
| MT42 | | 10:59:43 | 11:11:02 | 11:14:15 |
| CF43 | | 12:26:31 | 12:29:02 | 12:41:14 |
| MC44 | | 12:43:37 | 12:47:33 | 12:51:33 |
| MT45 | | 15:37:40 | 15:39:11 | 15:40:51 |
| MT46 | 02/05/2019 | 15:44:06 | 15:47:13 | 15:50:09 |
| CS47 | | 15:51:48 | 15:53:20 | 15:53:20 |
| MC48 | | 16:04:51 | 16:07:45 | 16:11:32 |
| CS49 | | 16:16:44 | 16:18:29 | 16:21:59 |
| MC50 | | 16:27:58 | 16:32:10 | 16:33:51 |
| MC51 | | 16:37:27 | 16:39:10 | 16:44:29 |
| MT52 | | 16:46:22 | 16:47:56 | 16:53:53 |
| MT53 | | 16:55:47 | 16:57:15 | 16:59:04 |

| Amostra | Data | Horário da Replicata | | | | | (conclusão) | |
|---------|------------|----------------------|----------|----------|----------|----------|-------------|----------|
| | | | | | | | | |
| CS54 | 02/05/2019 | 17:05:03 | 17:06:47 | 17:10:32 | | | | |
| MT55 | | 17:16:28 | 17:17:55 | 17:22:42 | | | | |
| MT56 | | 17:24:25 | 17:27:09 | 17:31:47 | | | | |
| MT57 | | 17:37:58 | 17:39:29 | 17:41:05 | | | | |
| CF58 | | 17:44:47 | 17:46:13 | 17:49:16 | | | | |
| CF59 | | 17:52:49 | 17:54:28 | 17:56:04 | | | | |
| MN60 | | 18:05:20 | 18:07:05 | 18:08:33 | | | | |
| TO61 | | 03/05/2019 | 09:02:01 | 09:06:30 | 09:10:00 | | | |
| TO62 | 09:14:19 | | 09:17:45 | 09:21:30 | | | | |
| TT63 | 09:24:02 | | 09:25:57 | 09:27:56 | | | | |
| CF64 | 09:37:39 | | 09:39:38 | 09:46:30 | | | | |
| MN65 | 09:49:04 | | 09:54:58 | 09:56:39 | | | | |
| TT66 | 10:04:16 | | 10:06:03 | 10:07:24 | | | | |
| TT67 | 10:11:13 | | 10:16:21 | 10:21:46 | | | | |
| TT68 | 10:26:26 | | 10:29:51 | 10:38:49 | | | | |
| TT69 | 10:49:48 | | 10:52:11 | 10:54:23 | | | | |
| MN70 | 12:22:29 | | 12:24:26 | 12:28:39 | | | | |
| AA71 | 12:33:37 | | 12:41:59 | 12:46:00 | | | | |
| AA72 | 12:52:58 | | 12:56:36 | 13:01:04 | | | | |
| MN73 | 13:07:43 | | 13:12:14 | 13:14:03 | | | | |
| CF74 | 13:18:38 | | 13:21:36 | 13:26:20 | | | | |
| CS75 | 01/11/2019 | | 17:22:00 | 17:23:38 | 17:25:15 | 17:27:00 | 17:29:58 | 17:31:32 |
| MT76 | | | 17:35:04 | 17:36:41 | 17:38:15 | 17:40:04 | 17:41:31 | 17:43:01 |
| MT77 | | 17:45:02 | 17:46:35 | 17:48:02 | 17:49:57 | 17:52:00 | 17:54:19 | |
| CS78 | | 17:56:06 | 17:57:48 | 17:59:13 | 18:01:40 | 18:03:17 | 18:04:43 | |
| CS79 | | 18:07:06 | 18:08:38 | 18:10:18 | 18:12:31 | 18:14:09 | 18:15:47 | |
| CS80 | | 18:18:46 | 18:20:34 | 18:22:19 | 18:24:43 | 18:26:22 | 18:28:14 | |
| CS81 | | 18:30:58 | 18:32:35 | 18:34:01 | 18:37:14 | 18:39:14 | 18:41:02 | |
| TT82 | | 18:43:53 | 18:45:55 | 18:47:40 | 18:51:08 | 18:52:50 | 18:54:27 | |
| TT83 | | 18:57:09 | 18:59:00 | 19:00:38 | 19:02:58 | 19:04:40 | 19:06:19 | |
| TT84 | | 19:08:39 | 19:10:38 | 19:12:15 | 19:15:24 | 19:17:16 | 19:18:56 | |
| TT85 | | 19:23:03 | 19:24:48 | 19:26:29 | 19:29:02 | 19:30:50 | 19:32:27 | |
| TT86 | | 19:35:51 | 19:37:48 | 19:39:30 | 19:42:21 | 19:44:14 | 19:46:25 | |
| CS87 | | 19:48:41 | 19:50:50 | 19:52:35 | 19:54:50 | 19:56:26 | 19:58:06 | |
| CS88 | | 20:00:21 | 20:02:04 | 20:03:36 | 20:05:35 | 20:07:09 | 20:08:47 | |
| CS89 | | 20:11:08 | 20:12:34 | 20:14:09 | 20:15:55 | 20:17:22 | 20:18:57 | |
| CS90 | | 20:21:23 | 20:22:58 | 20:24:24 | 20:26:38 | 20:28:13 | 20:29:40 | |
| CS91 | | 20:31:46 | 20:37:23 | 20:33:23 | 20:38:43 | 20:35:00 | 20:40:15 | |
| CS92 | | 20:42:15 | 20:43:59 | 20:46:14 | 20:48:17 | 20:49:58 | 20:51:39 | |
| MC93 | | 20:54:03 | 20:55:40 | 20:57:19 | 20:59:40 | 21:01:17 | 21:02:49 | |
| MN94 | | 21:04:50 | 21:06:11 | 21:07:44 | 21:10:12 | 21:11:59 | 21:13:26 | |
| PN95 | | 21:15:20 | 21:16:51 | 21:18:25 | 21:20:01 | 21:21:34 | 21:23:02 | |
| TT96 | | 21:24:42 | 21:26:27 | 21:27:48 | 21:29:59 | 21:31:31 | 21:32:58 | |
| PN97 | | 21:34:52 | 21:36:28 | 21:37:50 | 21:39:45 | 21:41:08 | 21:42:31 | |
| AA98 | | 21:44:33 | 21:46:15 | 21:47:39 | 21:50:02 | 21:51:30 | 21:52:59 | |
| AA99 | 21:55:07 | 21:56:35 | 21:58:03 | 22:00:07 | 22:01:50 | 22:03:16 | | |

Apêndice C – Participações em Produção Científica

Talanta 194 (2019) 86–89



Speciation analysis based on digital image colorimetry: Iron (II/III) in white wine



Joao H. Santos Neto^{a,b}, Icaro S.A. Porto^{a,b}, Mateus P. Schneider^c, Ana M.P. dos Santos^{a,b}, Adriano A. Gomes^{c,*}, Sergio L.C. Ferreira^{a,b,**}

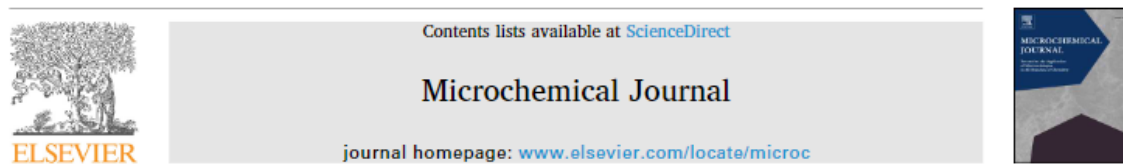
^a Universidade Federal da Bahia, Instituto de Química, Grupo de Pesquisa em Química e Quimiotoméria, Campus Ondina, 40170-115 Salvador, Bahia, Brazil

^b Instituto Nacional de Ciência e Tecnologia, INCT, de Energia e Ambiente, Universidade Federal da Bahia, 40170-115 Salvador, Bahia, Brazil

^c Universidade Federal do Rio Grande do Sul, Instituto de Química, 90650-001 Porto Alegre, Rio Grande do Sul, Brazil

DOI: <https://doi.org/10.1016/j.talanta.2018.09.102>

Microchemical Journal 147 (2019) 622–627



Detection oxidative degradation in lubricating oil under storage conditions using digital images and chemometrics



Flaviana C.P. Ribeiro^a, Alan S. Oliveira^a, Alisson Araújo^a, Weverton Marinho^a, Mateus P. Schneider^c, Licarion Pinto^{b,*}, Adriano A. Gomes^{a,c,**}

^a Instituto de Ciências Exatas, Universidade Federal do Sul e Sudeste do Pará - Unifesspa, Marabá, PA 68507-590, Brazil

^b Departamento de Química Fundamental, Universidade Federal de Pernambuco - UFPE, Recife, PE 50670-901, Brazil

^c Instituto de Química, Universidade Federal do Rio Grande do Sul - UFRGS, Porto Alegre, RS 90040-060, Brazil

DOI: <https://doi.org/10.1016/j.microc.2019.03.087>



Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc



Chromatographic quantification of seven pesticide residues in vegetable: Univariate and multiway calibration comparison



Emanuella Santos Sousa^a, Mateus P. Schneider^b, Licarion Pinto^{c,*},
Mario Cesar Ugulino de Araujo^{a,*}, Adriano de Araújo Gomes^{b,*}

^a Laboratório de Automação e Instrumentação em Química Analítica e Quimiometria (LAQA) Universidade Federal da Paraíba, CCEM, Departamento de Química, Caixa Postal 5093, CEP 58051-970, João Pessoa, PB, Brasil

^b Laboratório de Quimiometria e Instrumentação Analítica-LQIA, Instituto de Química da Universidade Federal do Rio Grande do Sul, Agronomia 9500, Porto Alegre -RS, Brasil

^c Departamento de Química Fundamental, Universidade Federal de Pernambuco - UFPE, Recife, PE 50670-901, Brazil

DOI: <https://doi.org/10.1016/j.microc.2019.104301>



Contents lists available at ScienceDirect

Microchemical Journal

journal homepage: www.elsevier.com/locate/microc



Bio-inspired algorithm for variable selection in i-PLSR to determine physical properties, thorium and rare earth elements in soils from Brazilian semiarid region



Danubio Leonardo Bernardino Oliveira^a, Lavinia Hannah de Souza Pereira^a,
Mateus Pires Schneider^b, Ygor Jacques Agra Bezerra Silva^c,
Clístenes Williams Araújo Nascimento^c, Peter van Straaten^d, Yuri Jacques Agra Bezerra Silva^e,
Adriano de Araujo Gomes^{b,*}, Germano Veras^{a,*}

^a Chemistry Department, State University of Paraíba (UEPB), Campina Grande, PB 58.429-500, Brazil

^b Institute of Chemistry, Federal University of Rio Grande do Sul (UFRGS), Porto Alegre, RS, 91501-970, Brazil

^c Agronomy Department, Federal Rural University of Pernambuco (UFRPE), Dom Manuel de Medeiros street, s/n - Dois Irmãos, Recife, PE 52171-900, Brazil

^d Professor Emeritus, University of Guelph, Guelph, Ontario, Canada

^e Agronomy Department, Federal University of Piauí (UFPI), Planalto Horizonte, Bom Jesus, PI 64900-000, Brazil

DOI: <https://doi.org/10.1016/j.microc.2020.105640>

Apêndice D – Participação em eventos

- 19º Encontro Nacional de Química Analítica – ENQA, e 7º Congresso Iberoamericano de Química Analítica – CIAQA
 - Caldas Novas, Goiás. 2018
 - Apresentação de Pôster: “*Brazilian red wine classification with respect of grape variety by means of FTIR and Chemometric tools*”
- X Workshop de Quimiometria
 - Salvador, Bahia. 2019
 - Apresentação Oral e Destaque de Seção: “*Perfil de Vinhos através de FTIR – ATR e Regressão PLS*”
- Escola de Inverno de Quimiometria
 - Porto Alegre, Rio Grande do Sul. 2019
 - Equipe de Apoio
- XI Workshop de Quimiometria
 - Campina Grande, Paraíba – versão web conferência. 2020
 - Presença em Minicurso: Design de Experimentos