

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

DISSERTAÇÃO DE MESTRADO

**Desenvolvimento de Metodologias para a Caracterização de Taninos
Empregando Técnicas de Espectrometria Molecular e Quimiometria**

FÁBIO HENRIQUE WEILER

Porto Alegre, agosto de 2017.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

FÁBIO HENRIQUE WEILER

**Desenvolvimento de Metodologias para a Caracterização de Taninos
Empregando Técnicas de Espectrometria Molecular e Quimiometria**

Dissertação apresentada como requisito parcial para a
obtenção do grau de Mestre em Química

Prof. Dr. Marco Flôres Ferrão
Orientador

Dr. Fábio dos Santos Grasel
Co-orientador

Porto Alegre, agosto de 2017.

AGRADECIMENTOS

À minha esposa, Vanessa Seibel, pelo amor, carinho, apoio, incentivo, companheirismo e paciência. Você esteve sempre ao meu lado para tudo.

À minha família, com destaque aos meus pais Lírio Weiler e Elaine Weiler pelo incentivo, carinho, vocês são exemplos de vida para mim, com suas vidas simples, mas de grandes princípios. Ao meu irmão Celson, pelo companheirismo, apoio, conselhos e amizade. Ao meu irmão André pela parceria, respeito e amizade. Ao meu irmão Roberto pela amizade, respeito e compreensão. A minha vó Lucia por todas as orações.

A família da minha esposa Vanessa por todo apoio e incentivo.

Ao meu orientador Marco Flores Ferrão pela orientação, ensinamentos, confiança, paciência e amizade durante a realização deste trabalho.

Ao meu co-orientador Fábio dos Santos Grasel pelos ensinamentos, paciência e amizade.

Ao meu amigo e colega gremista André.

A amiga e colega Camila por toda ajuda durante todo o trabalho.

A todos os colegas do IF que possibilitaram o meu afastamento para os estudos.

Aos amigos do IF, Rafael e Cláudia pelo apoio e parceria.

Aos meus colegas de laboratório e amigos: Carla, Marcelo, Fernanda, Layane, Juliana e Manuela.

A Universidade Federal do Rio Grande do Sul e a CAPES pelos recursos estruturais para realização deste trabalho.

A Empresa Tanac/SA pelo apoio no projeto.

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	ix
ABREVIATURAS	xii
RESUMO	xiv
ABSTRACT	xv
1 INTRODUÇÃO	1
2 OBJETIVO	3
2.1 GERAL.....	3
2.1.1 Específicos.....	3
3 REVISÃO BIBLIOGRÁFICA	4
3.1 TANINOS.....	4
3.2 ACÁCIA NEGRA	8
3.3 QUEBRACHO	9
3.4 ESPECTROSCOPIA NO INFRAVERMELHO	9
3.4.1 Espectroscopia por Reflexão Total Atenuada (ATR).....	12
3.5 QUIMIOMETRIA	13
3.5.1 Calibração Multivariada	13
3.5.2 Regressão por Mínimos Quadrados Parciais (PLS)	14
3.5.3 Métodos de Seleção de Variáveis em Modelos de Regressão.....	15
3.5.4 Seleção do Conjunto de Calibração e Previsão	17
3.5.5 Tratamento e Pré-Processamento dos Dados.....	17
3.5.6 Avaliação dos Modelos de Regressão	19
4 MATERIAIS E MÉTODOS	22
4.1 AMOSTRAS	22
4.2 ESPECTROSCOPIA DE INFRAVERMELHO MÉDIO	22

4.3	ESPECTROCOPIA DE INFRAVERMELHO PRÓXIMO	22
4.4	ANÁLISE MULTIVARIADA	23
4.4.1	Regressão multivariada do nitrogênio, enxofre, cinzas, carbono e hidrogênio por ATR-FTIR e NIR.....	23
4.5	TEORES DE NITROGÊNIO, ENXOFRE, CARBONO, HIDROGÊNIO E CINZAS.....	24
5	RESULTADOS E DISCUSSÃO.....	25
5.1	TEORES DE NITROGÊNIO, ENXOFRE, CARBONO, HIDROGÊNIO E CINZAS.....	25
5.2	ESPECTROSCOPIA DE INFRAVERMELHO	25
5.3	ANÁLISE MULTIVARIADA	27
5.3.1	Determinação do Teor de Nitrogênio por NIR, ATR-FTIR e Fusão por Regressão Multivariada.....	27
5.3.2	Determinação do Teor de enxofre por NIR, ATR-FTIR e Fusão por Regressão Multivariada.....	33
5.3.3	Determinação do Teor de Cinzas por NIR, ATR-FTIR e Fusão por Regressão Multivariada.....	40
5.3.4	Determinação do Teor de Carbono por NIR, ATR-FTIR por Regressão Multivariada.....	46
5.3.5	Determinação do Teor de Hidrogênio por NIR, ATR-FTIR por Regressão Multivariada.....	47
5.4	CONSIDERAÇÕES FINAIS	48
6	CONCLUSÕES	50
7	REFERÊNCIAS BIBLIOGRÁFICAS.....	52
8	ANEXOS	56
8.1	RESULTADOS DA DETERMINAÇÃO DO TEOR DE NITROGÊNIO, ENXOFRE, CARBONO, HIDROGÊNIO E CINZAS PELOS MÉTODOS DE REFERÊNCIA.....	56

8.2 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE NITROGÊNIO EM TANINOS.....	61
8.3 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE ENXOFRE EM TANINOS.....	66
8.4 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE CINZAS EM TANINOS.....	71
8.5 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE CARBONO EM TANINOS.....	76
8.6 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE HIDROGÊNIO EM TANINOS.....	80

LISTA DE FIGURAS

Figura 1: Estrutura dos taninos hidrolisáveis ²²	6
Figura 2: Monômero de taninos condensados ²²	6
Figura 3: Estrutura química de tanino condensado ²⁵	7
Figura 4: Representação das (a) vibrações de estiramento e (b) deformações angulares ^{a37}	11
Figura 5: Representação simplificada da análise por ATR ⁴²	13
Figura 6: Comportamento da variância e tendência (viés) nos modelos PLS em geral, indicado pelas linhas tracejadas. A linha cheia indica o comportamento de erro na previsão ⁵⁸	20
Figura 7: Espectros de NIR para amostras de taninos: a) espectro sem tratamentodos dos dados; b) espectro com tratamento dos dados; espectro suavizado e normalizado.	26
Figura 8: Espectros de ATR-FTIR para amostras de taninos: a) espectro sem tratamentodos dos dados; b) espectro com tratamento dos dados; espectro suavizado e normalizado.	26
Figura 9: Espectro NIR com destaque as regiões selecionadas (8, 24, 30) combinadas com regressão NIR/siPLS para determinação de nitrogênio em taninos.	31
Figura 10: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas pelo NIR/siPLS para nitrogênio combinando os intervalos 8, 24, 30 no modelo s3iPLS32M.	31
Figura 11: Gráfico dos valores de referência <i>versus</i> valores previstos por NIR/siPLS para determinação de nitrogênio em taninos.	32
Figura 12: Espectro NIR com destaque as regiões selecionadas (3, 4, 9) combinadas com regressão NIR/siPLS, para determinação de enxofre em taninos.	37
Figura 13: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas pelo NIR/siPLS para enxofre combinando os intervalos 3, 4, 9 no modelo s3iPLS16M.	38
Figura 14: Gráfico dos valores de referência <i>versus</i> valores previstos por NIR/siPLS para determinação de enxofre em taninos.	39
Figura 15: Espectro NIR com destaque as regiões selecionadas (3, 5, 6) combinadas com regressão NIR/siPLS para determinação de cinzas em taninos.	43

Figura 16: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas pelo NIR/siPLS para cinzas combinando os intervalos 3, 5, e 6 no modelo s3iPLS8M..... 44

Figura 17: Gráfico dos valores de referência *versus* valores previstos por NIR/siPLS para determinação de cinzas em taninos. 45

LISTA DE TABELAS

Tabela 1: Regiões espectrais do infravermelho ³⁷	10
Tabela 2: Resultados das figuras de mérito dos melhores modelos para determinação do teor de nitrogênio com dados NIR, ATR-FTIR, Fusão NIR-ATR-FTIR e Fusão NIR-ATR-FTIR resolução 1:2.	28
Tabela 3: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de nitrogênio por NIR com MSC.	30
Tabela 4: Resultados das figuras de mérito dos melhores modelos de determinação do teor de enxofre a com dados NIR, ATR-FTIR, Fusão NIR-ATR-FTIR e Fusão NIR-ATR-FTIR resolução 1:2.....	34
Tabela 5: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de enxofre por NIR com MSC.....	36
Tabela 6: Resultados das figuras de mérito dos melhores modelos de determinação do teor de cinzas com dados NIR, ATR-FTIR, Fusão NIR-ATR-FTIR e Fusão NIR-ATR-FTIR resolução 1:2.	40
Tabela 7: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de cinzas por NIR com MSC.....	42
Tabela 8: Resultados das figuras de mérito dos melhores modelos de determinação do teor de carbono com dados NIR e ATR-FTIR.	46
Tabela 9: Resultados das figuras de mérito dos melhores modelos de determinação do teor de hidrogênio com dados NIR e ATR-FTIR.....	47
Tabela 10 Resultados das figuras de mérito dos melhores modelos para os parâmetros de nitrogênio, enxofre, cinzas, carbono e hidrogênio.	49
Tabela 11: Resultados da determinação de Cinzas em amostras de taninos pelo método de referência ABNT NBR 11042.	56
Tabela 14: Resultados da determinação de Carbono em amostras de taninos pelo método de referência CHNS.	59
Tabela 15: Resultados da determinação de Hidrogênio em amostras de taninos pelo método de referência CHNS.	60
Tabela 16: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de nitrogênio por NIR sem MSC.	61

Tabela 17: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de nitrogênio por ATR-FTIR sem MSC.	62
Tabela 18: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de nitrogênio por ATR-FTIR com MSC.....	63
Tabela 19: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de nitrogênio por fusão dos dados NIR e ATR-FTIR sem MSC.	64
Tabela 20: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de nitrogênio por fusão dos dados NIR e ATR-FTIR com MSC.....	64
Tabela 21: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de nitrogênio por fusão dos dados NIR e ATR-FTIR com resolução 1:2 sem MSC.....	65
Tabela 22: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de nitrogênio por fusão dos dados NIR e ATR-FTIR com resolução 1:2 com MSC.....	65
Tabela 23: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de enxofre por NIR sem MSC.....	66
Tabela 24: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de enxofre por ATR-FTIR sem MSC.	67
Tabela 25: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de enxofre por ATR-FTIR com MSC.	68
Tabela 26: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de enxofre por fusão dos dados NIR e ATR-FTIR sem MSC.....	69
Tabela 27: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de enxofre por fusão dos dados NIR e ATR-FTIR com MSC.	69
Tabela 28: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de enxofre por fusão dos dados NIR e ATR-FTIR com resolução 1:2 sem MSC.....	70
Tabela 29: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de enxofre por fusão dos dados NIR e ATR-FTIR com resolução 1:2 com MSC.....	70
Tabela 30: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de cinzas por NIR sem MSC.....	71
Tabela 31: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de cinzas por ATR-FTIR sem MSC.....	72

Tabela 32: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de cinzas por ATR-FTIR com MSC.	73
Tabela 33: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de cinzas por fusão dos dados NIR e ATR-FTIR sem MSC.	74
Tabela 34: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de cinzas por fusão dos dados NIR e ATR-FTIR com MSC.	74
Tabela 35: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de cinzas por fusão dos dados NIR e ATR-FTIR com resolução 1:2 sem MSC.	75
Tabela 36: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de cinzas por fusão dos dados NIR e ATR-FTIR com resolução 1:2 com MSC.	75
Tabela 37: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de carbono por NIR sem MSC.	76
Tabela 38: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de carbono por NIR com MSC.	77
Tabela 39: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de carbono por ATR-FTIR sem MSC.	78
Tabela 40: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de carbono por ATR-FTIR com MSC.	79
Tabela 41: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de hidrogênio por NIR sem MSC.	80
Tabela 42: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de hidrogênio por ATR-FTIR sem MSC.	81
Tabela 43: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de hidrogênio por ATR-FTIR com MSC.	82

ABREVIATURAS

ABNT – Associação Brasileira de Normas Técnicas

ATR - Reflexão Total Atenuada (do inglês, “*Attenuated Total Reflection*”)

ATR-FTIR – Espectroscopia de Infravermelho Médio com Transformada de Fourier e Reflexão Total Atenuada (do inglês, “*Fourier Transform Infrared Spectroscopy with Attenuated Total Reflection*”)

FIR - Infravermelho distante (do inglês “*Far Infrared*”)

FTIR - Espectroscopia no Infravermelho Médio com Transformada de Fourier (do inglês “*Fourier Transform Infrared*”)

iPLS - Mínimos Quadrados Parciais por Intervalo (do inglês, “*Interval Partial Least Squares*”)

IR - Infravermelho (do inglês “*Infrared*”)

NIR - Infravermelho Próximo (do inglês “*Near Infrared*”)

MIR - Infravermelho Médio (do inglês “*Medium Infrared*”)

MSC - Correção do Espalhamento do Sinal (do inglês, “*Multiplicative scatter Correction*”)

PCA - Análise por Componentes Principais (do inglês “*Principal Component Analysis*”)

PLS - Mínimos Quadrados Parciais (do inglês, “*Partial Least Squares*”)

R^2 - Coeficiente de Determinação

Rc^2 - Coeficientes de Determinação de Calibração

RMSE - Raiz Quadrada dos Erros Médios (do inglês, *Root “Mean Square Error”*)

RMSEC - Raiz Quadrada do Erro Médio de Calibração (do inglês, “*Root Mean Square Error of Calibration*”);

RMSEP - Raiz Quadrada do Erro de Predição (do inglês, *Root Mean Square “Error of Prediction*”) e

RMSECV - Raiz Quadrada do Erro de Validação Cruzada (do inglês, “*Root Mean Square Error of Cross Validation*”)

Rp^2 - Coeficientes de Determinação de Previsão

siPLS - Mínimos quadrados parciais por sinergismo de intervalo (do inglês, “*Sinergy Interval Partial Least Squares*”)

VLs – Variáveis Latentes

RESUMO

Neste trabalho foram estudadas estratégias para caracterização de extratos de taninos condensados através de análises no infravermelho médio com transformada de Fourier (FTIR), obtidos com acessório de reflexão total atenuada (ATR), infravermelho próximo (NIR) e fusão das duas regiões do infravermelho. Foram utilizados um total de 64 amostras de taninos, 33 de acácia negra e 31 de quebracho. Os parâmetros estudados nos taninos foram os de nitrogênio, enxofre, carbono, hidrogênio e cinzas. Os modelos de regressão multivariada foram construídos por mínimos quadrados parciais (PLS), com avaliação dos métodos de mínimos quadrados parciais por intervalo (iPLS) e mínimos quadrados parciais por sinergismo de intervalos (siPLS) para seleção de variáveis. Avaliou-se a utilização do pré-tratamento de correção do espalhamento do sinal (MSC) nos modelos. A seleção de variáveis por siPLS juntamente com o pré-tratamento por MSC apresentaram os melhores resultados para os modelos construídos. Para os conjuntos de dados FTIR, NIR e fusão os melhores modelos foram construídos a partir dos dados NIR, sendo que em alguns modelos os dados com fusão apresentaram resultados semelhantes aos de NIR. Entre as propriedades avaliadas, nitrogênio, enxofre e cinzas apresentaram modelos com erros de validação cruzada e previsão (RMSECV e RMSEP) satisfatórios e coeficientes de determinação (R_p^2 , R_c^2) aceitáveis considerando-se que as amostras são reais. Para as propriedades de carbono e hidrogênio não foram obtidos bons resultados para os modelos de regressão nas duas regiões do infravermelho. Desta forma a metodologia proposta pode ser aplicada em análises de rotinas para determinação simultânea de nitrogênio, enxofre e cinzas em extratos de taninos, somando-se as vantagens da espectroscopia no infravermelho, tais como rapidez na análise, baixo custo, não ser destrutiva e a não utilização de solventes.

ABSTRACT

In this work, strategies for the characterization of condensed tannin extracts have been studied, through analysis of the medium infrared with Fourier Transform Infrared (FTIR), obtained with Attenuated Total Reflection (ATR) accessory, Near Infrared (NIR) and the fusion of both infrared regions. A total of 64 tannin samples were used, 33 of *Acacia mearnsii* and 31 of Quebracho (*Schinopsis balansae* and *Schinopsis Lorentzii*). The parameters studied in each tannin were nitrogen, sulfur, carbon, hydrogen and ashes. The multivariate regression models were constructed by partial least squares (PLS), with evaluation methods by interval partial least squares (iPLS) and synergy interval partial least squares (siPLS) to proceed variable selection. The use of a pre-treatment multiplicative scatter correction (MSC) has been observed in all models. The variable selection by siPLS, along with the pre-treatment by MSC, have presented the best results to the models. For data sets FTIR, NIR and fusion, the best models have been obtained from the NIR ones. Therefore, in some models, fusion data presented equivalent results to NIR. Among the evaluated properties, nitrogen, sulfur and ashes presented models with satisfactory errors in cross validation and prediction (RMSECV e RMSEP) and acceptable coefficients of determination (R_p^2 , R_c^2), regarding that the samples are real. Considering carbon and hydrogen properties, favourable results were not obtained to the regression models in both infrared regions. In this way, the proposed methodology can be applied in routine analysis to simultaneous determination of nitrogen, sulfur and ashes in tannin extracts, adding the advantages of infrared spectroscopy, such as rapid analysis, low cost, non-destructive and the non-use of solvents.

1 INTRODUÇÃO

Nas últimas décadas, a conscientização sobre a importância na preservação do meio ambiente resultou em um importante crescimento no uso de materiais que causem menor impacto ambiental para diferentes aplicações¹. As perspectivas de desenvolvimento sustentável estão sendo aplicadas cada vez mais no desenvolvimento de novos materiais provenientes de fontes renováveis². Neste cenário, o interesse em fontes renováveis, como os taninos, vem aumentando. Esses produtos naturais são matérias primas para diversos tipos de seguimentos, como a indústria de couro, alimentícia, bebidas, têxteis, farmacêutica e no tratamento de água e na agricultura ecológica³.

O controle de qualidade das propriedades físico-químicas dos taninos é de fundamental importância para sua aplicação industrial. A sua composição química pode ter grandes variações, pois são compostos que são extraídos de diferentes regiões das plantas e também por essas plantas serem cultivadas em diferentes regiões climáticas^{4,5,6}.

Algumas análises importantes realizadas em taninos são para a determinação do teor de enxofre, nitrogênio, cinzas e carbono. A presença do enxofre está relacionada com o processo de sulfitação, que auxilia na melhora de propriedades curtentes, como penetração mais rápida na pele e maior solubilidade em meio aquoso. Esse processo altera composições de cinzas, carbono e hidrogênio. Já a quantidade nitrogênio está inversamente relacionada com a concentração de polifenóis no taninos^{7,8,9}.

Os métodos de quantificação de nitrogênio, enxofre, carbono, hidrogênio e cinzas seguem métodos usuais como de análise multielementar CHNS e gravimétrico¹⁰. Métodos de análise multielementar CHNS necessitam de etapas de preparo das amostras e condições de combustão específicos para alguns parâmetros, tornando-se lento e de alto custo. Nos métodos gravimétricos as amostras devem passar por diversas etapas de secagem e calcinação, até massa constante, sendo um processo laborioso que demanda muito tempo.

Neste contexto, o presente trabalho vem contribuir no desenvolvimento de metodologias analíticas não destrutivas por meio de espectroscopia de infravermelho associada a ferramentas quimiométricas de regressão multivariada para caracterização de

taninos, constituindo-se em métodos mais rápidos, com mínimo preparo de amostras, baixo custo e que permitem a análise simultânea dos parâmetros mencionados anteriormente.

2 OBJETIVO

2.1 GERAL

Desenvolvimento de metodologias para a determinação de carbono, hidrogênio, enxofre, nitrogênio e cinzas em taninos empregando técnicas de espectroscopia de infravermelho médio e próximo em conjunto com ferramentas quimiométricas de regressão multivariada.

2.1.1 Específicos

- I) Avaliar o pré-tratamento dos dados com MSC nos modelos de regressão;
- II) Construir modelos de regressão empregando métodos de seleção de variáveis (iPLS e siPLS) e selecionar as faixas espectrais mais relacionadas para cada teor (nitrogênio, enxofre, carbono, hidrogênio e cinzas);
- III) Avaliar a partir dos dados de infravermelho próximo, médio e fusão dos dados o que apresenta os melhores modelos de regressão para as propriedades de N, C, H, S e cinzas.

3 REVISÃO BIBLIOGRÁFICA

3.1 TANINOS

Os taninos pertencem a um grupo de compostos fenólicos proveniente do metabolismo secundário vegetal ou metabolismo especial que apresentam um amplo valor nas interações entre a planta e seu ecossistema exercendo, por exemplo, o papel de fago-inibidores contra herbívoros ou como agentes antimicrobianos¹¹. Como metabólitos secundários, os taninos são compostos fenólicos de grande interesse econômico e ecológico. Apresentam solubilidade em água e peso molecular compreendido entre 500 e 3000 Dalton, possuindo a habilidade de formar complexos insolúveis em água com proteínas, gelatinas e alcalóides¹². Tais compostos são responsáveis pela adstringência de muitos frutos e produtos vegetais, devido à precipitação de glucoproteínas salivares, o que ocasiona a perda do poder lubrificante¹³.

A palavra tanino é largamente usada, particularmente em literatura botânica, originalmente derivada do termo “tanante”, implicando que o material vegetal produza couro a partir de peles¹⁴.

Produtos naturais, como os taninos, são matérias-primas muito importantes para diferentes indústrias, por apresentarem em sua composição um grande número de princípios ativos. As suas principais aplicações são: curtimento de couros naturais; na agricultura ecológica; na preparação para floculantes para o tratamento de água; na obtenção de polímeros naturais; na indústria alimentícia como antioxidantes nos sucos de frutas e bebidas; na clarificação de vinhos; como corantes têxteis e na síntese de produtos farmacêuticos³.

O conteúdo de taninos nas plantas pode variar de acordo com as condições climáticas e geográficas, podem apresentar uma composição química variada, sendo muitas vezes, pouco conhecida. Na forma não oxidada, os taninos reagem com proteínas através de ligação de hidrogênio ou ligações hidrofóbicas. Quando não oxidados, os taninos se transformam em quinonas, as quais formam ligações covalentes com alguns grupos funcionais das proteínas, principalmente os grupos sulfídricos de cisteína e ω -amino da lisina¹⁵.

A maioria dos vegetais é portador de taninos, podem ser encontrados nas raízes, no lenho, na casca, nas folhas, nos frutos, nas sementes e na seiva. O teor e a espécie de taninos variam, não só de um vegetal ao outro como também da uma parte para outra do mesmo vegetal^{5,6}.

Os taninos são substâncias polifenólicas complexas e de acordo com suas estruturas químicas, são divididos em dois grandes grupos nas plantas terrestres: taninos hidrolisáveis e taninos condensados¹⁶. Os taninos hidrolisáveis possuem baixa disponibilidade e com produção limitada em todo o mundo, sendo menos de 10% da produção mundial de taninos comerciais, devido ao preço elevado, tornando-os menos atraente que os taninos condensados^{17,18}. Por outro lado, os taninos condensados constituem mais de 90% da produção mundial.

Os taninos hidrolisáveis são ésteres de um açúcar (central) e de fenólicos, principalmente ácidos gálico e elágico e seus derivados. Os galotaninos resultam da união entre unidades de ácido gálico e um poliol. Os elagitaninos possuem um ou dois resíduos de hexa-hidroxidifenoila-D-glicose (HHDP) unidos a um açúcar central (Figura 1). Pode ocorrer a oxidação de HHDP a desidroxidifenoila (HHP) que é característico dos desidroelagitaninos^{19,13}.

Os taninos são amplamente distribuídos dentro do reino vegetal, sendo comuns em espécies gimnospermas como angiospermas. Dentro das angiospermas são mais comuns nas dicotiledônias do que nas monocotiledônias. Algumas famílias de dicotiledônias ricas em taninos são as leguminosas, anacardiáceas, combretáceae, rhizoparacea, mirtácea e polinaceae¹⁵. Os condensados ocorrem amplamente em gimnospermas e angiospermas, enquanto que os hidrolisáveis estão quase restritos a angiospermas dicotiledôneas¹⁹. Algumas madeiras apresentam teor de polifenóis acima de 35% como o cerne do quebracho²⁰, a casca da acácia negra e de espécies provenientes de manguezais também possuem altos teores de polifenóis²¹.

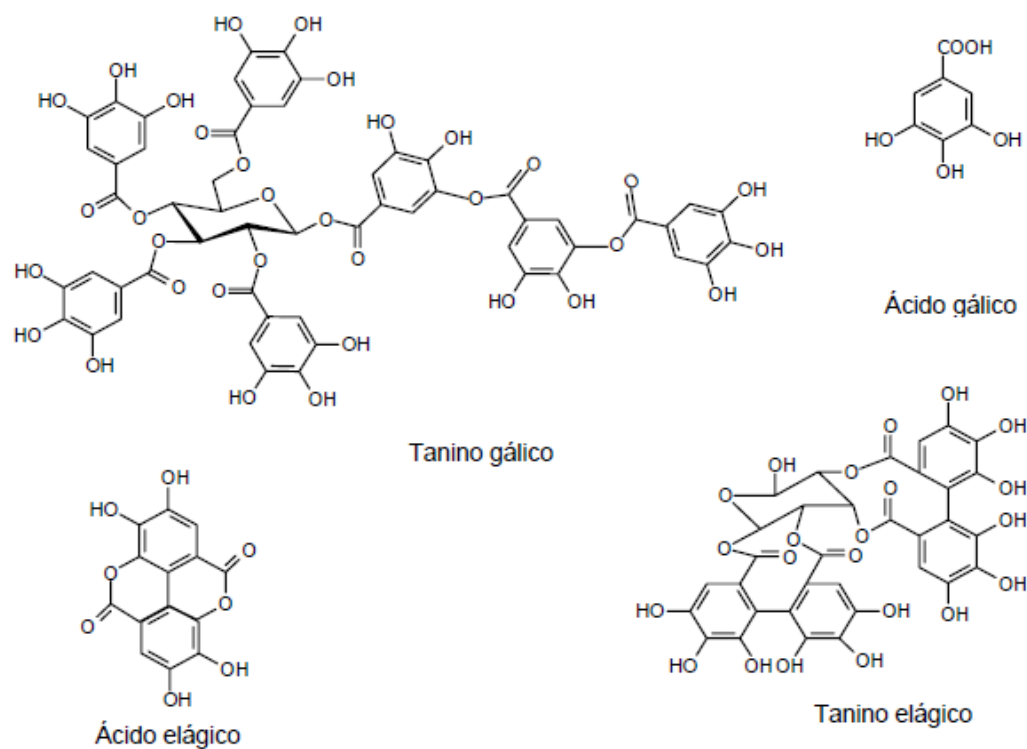


Figura 1: Estrutura dos taninos hidrolisáveis²².

Os taninos condensados são oligômeros e polímeros formando pela policondensação de duas unidades flavan-3-ol, tais como a catequina ou a epicatequina, respectivamente unidas por pares de ligações $C_4 \rightarrow C_8$ ou $C_4 \rightarrow C_6$ (Figura 2). Essas unidades também são denominadas proantocianidinas, devido à propriedade de produzirem pigmento avermelhado da classe das antocianidinas, após degradação à quente com ácido mineral diluído^{13,19,23}.

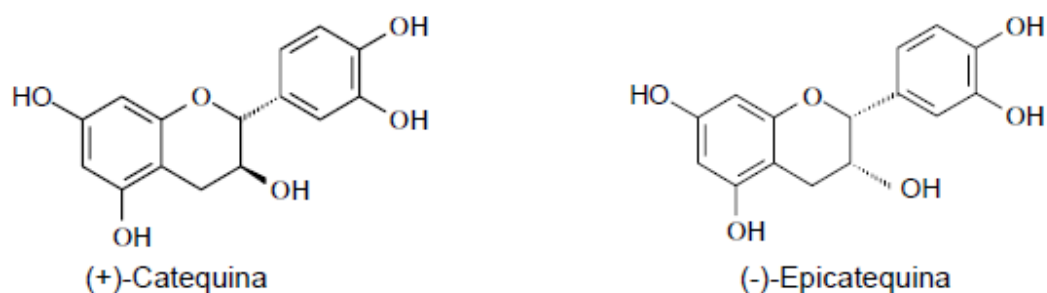


Figura 2: Monômero de taninos condensados²².

Os taninos condensados ou proantocianidinas podem conter duas a cinquenta unidades flavonoides, portanto possuem estrutura complexa e são resistentes a hidrólise, mas podem ser solúveis em solventes orgânicos aquosos, dependendo da sua estrutura. A Figura 3 ilustra a estrutura química dos taninos condensados²⁴.

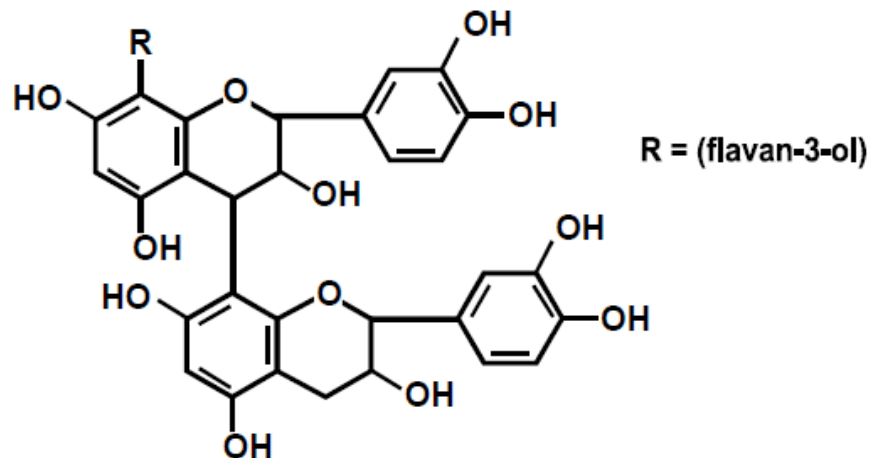


Figura 3: Estrutura química de tanino condensado²⁵.

As duas fontes de taninos condensados comerciais são a acácia negra (*Acacia mearnsii*) e o quebracho (*Shinopsis sp.*)^{26,27}. Ambos os taninos (acácia e quebracho) podem sofrer o processo de sulfitação, para facilitar o processo de extração, clareamento do extrato e melhoria nas propriedades curtentes, como penetração mais rápida na pele e maior solubilidade em meio aquoso^{7,8}.

O processo de sulfitação aumenta o teor de material inorgânico, resultando assim, em maior teor de cinzas, em contrapartida menores teores de H e C. A concentração de N tem uma relação inversa com o teor de polifenóis presentes, portanto, apresentando baixos valores de nitrogênio possuem altos valores de polifenóis. A concentração de nitrogênio indica a presença de carboidratos e outras substâncias diferentes de polifenóis, portanto uma análise importante para a caracterização destes extratos é pelo empregando o analisador elementar CHNS^{8,9}.

3.2 ACÁCIA NEGRA

A *Acacia mearnsii* é um arbusto de grande porte ou uma pequena árvore, comumente com altura de 6 e 10 metros, podendo, contudo, atingir até 15 metros. Apresenta um caule principal, geralmente reto e dominante. Nas árvores adultas, a casca é variável, geralmente preto-amarronzada, dura e fissurada. É uma espécie de crescimento rápido, que alcança taxa máxima entre 3 e 5 anos de idade. É uma cultura de ciclo curto, com colheitas em torno do sétimo ano de idade²⁸.

As acácias habitam as regiões tropicais e subtropicais da América, África, Ásia e Austrália, entretanto, não ocorrem nas floras da Europa e Nova Zelândia²⁹. Existem aproximadamente 1200 espécies do gênero Acácia, ocorrendo naturalmente nas savanas e nas regiões da Austrália, África, Índia e América. Boa parte das espécies tem a forma de árvores e arbustos perenifólios e muitas também suportam períodos secos prolongados^{11,30}.

Acácia-negra é a terceira espécie florestal mais plantada no Brasil, superada apenas por espécies dos gêneros Eucaliptus e Pinus³¹. No Brasil o primeiro plantio comercial foi feito em 1930, quando foram importados 30 kg de sementes da África do Sul. Acredita-se que, até meados dos anos 80, todos os plantios tenham sido originários desse material³².

A acácia-negra é uma das principais espécies florestais plantadas no Estado do Rio Grande do Sul, onde representa significativa parcela dos reflorestamentos do Estado. A plantação de acácia-negra tem característica multifuncional, ou seja, tem ação recuperadora dos solos de baixa fertilidade, através da fixação de nitrogênio, além de seu uso como matéria-prima para tanino, celulose e carvão, apresenta, ainda, grande importância social, por ser plantada em pequenas propriedades, beneficiando milhares de famílias na região. Também é usada em plantios consorciados no início da cultura e, após três anos de idade, a área é usada como pastagem de gado, que se alimenta da vegetação do sub-bosque³³.

Historicamente, a acácia negra é reconhecida pela qualidade de sua casca, onde são obtidos os extratos vegetais, ricos em tanantes e polifenóis, que originam os taninos, largamente utilizados no curtimento de peles, constituindo-se no mais conhecido grupo de produtos obtidos a partir da casca. Pode-se obter, também, os floculantes, utilizados em processos de tratamento de água e que por sua origem vegetal, não apresentam os inconvenientes e contraindicações dos similares de origem química, além de dispersantes, usados como aditivos nos processos onde é necessária a aspersão de líquidos, os adesivos e

resinas, à base fenólica, destonados, principalmente, às indústrias que utilizam madeiras, bem como componentes destinados à produção de agentes anticorrosivos³⁴.

3.3 QUEBRACHO

As florestas de quebracho selvagens na região do Gran Chaco, localizado na Argentina, Bolívia e Paraguai são utilizados a mais de 100 anos como importante fonte de taninos e madeira. A madeira é durável e extremamente dura e o nome quebracho é derivado da palavra espanhola *quiebra-hacha* que significa quebra machado⁷.

Os extratos de quebracho são obtidos a partir do cerne das árvores, que possuem normalmente de 15 à 25 metros de altura. A partir de suas folhas pode-se extrair tintura de resina, uma substância que pode ser usada com tônico. A casca pode ser triturada e cozida a qual possui propriedades medicinais³⁵.

Os extratos de quebracho são normalmente obtidos a partir da solubilização do cerne em água quente, onde esse é previamente separado da casca. Um extrato solúvel em água fria é obtido após tratamento desse extrato em água morna com bissulfito ou extração direta de aparas de madeira com uma solução aquosa de bissulfito. As taxas de extração mais elevadas são obtidas com solução aquosa de bissulfito do que somente água fervente⁷.

O extrato de quebracho consiste em cerca de 95% de proantocianidinas (PAs) e 5% de açúcares solúveis em água numa base seca. O extrato de quebracho obtido a partir da espécie *Schinopsis balansae*, da região do Chaco, contém 20-21% de taninos, enquanto a espécie de *Schinopsis Lorentzii*, dessa mesma região, possui um teor de 15-18% de taninos⁷.

3.4 ESPECTROSCOPIA NO INFRAVERMELHO

A espectroscopia no infravermelho (IR) é uma ferramenta versátil aplicada às determinações quantitativas e qualitativas de muitas espécies inorgânicas, orgânicas e bioquímicas. A radiação infravermelha corresponde aproximadamente à parte do espectro eletromagnético situada entre a região do visível e das micro-ondas³⁶.

O espectro na região do infravermelho é usualmente dividido em infravermelho próximo (NIR), infravermelho médio (MIR) e infravermelho distante (FIR). A Tabela 1 apresenta os intervalos correspondentes a cada uma destas regiões³⁷.

Tabela 1: Regiões espectrais do infravermelho³⁷.

Região	Número de onda (cm⁻¹)	Comprimento de onda (nm)
NIR	12.800 – 4.000	780 – 2.500
MIR	4.000 - 400	2.500 – 5.000
FIR	200 - 10	5.000 – 100.000

A radiação infravermelha não produz transições eletrônicas pois não possui energia suficiente; assim, para que tenha absorção da radiação infravermelha uma molécula precisa apresentar uma variação no momento de dipolo durante seu movimento rotacional ou vibracional. Apenas nessas situações a radiação eletromagnética é capaz de interagir com a molécula, sendo que a energia absorvida está associada aos modos vibracionais e rotacionais de grupos funcionais³⁸.

Em uma molécula, as posições relativas dos átomos não estão fixadas exatamente, estas variam continuamente em função de inúmeros tipos de vibrações e rotações em torno das ligações da molécula. As vibrações podem ser classificadas nas categorias de estiramento e deformação. Um modo de estiramento envolve uma variação contínua na distância interatômica ao longo do eixo da ligação entre dois átomos. Já os modos de deformação são caracterizados por variações no ângulo entre duas ligações e podem ser de quatro tipos: balanço no plano, tesoura no plano, sacudida fora do plano e torção fora do plano³⁷, como ilustrados na Figura 4.

Para NIR, estas distorções são devidas às interações de sobretons e combinações modais relativas a diversos grupos funcionais como, por exemplo, alifáticos (CH), aromáticos (CH), carboxil (CO), hidroxil (OH), amidas (NO), e aminas (NH). No entanto, o espectro é complexo porque muitas transições podem ocorrer levando a uma grande sobreposição de bandas de absorção³⁸.

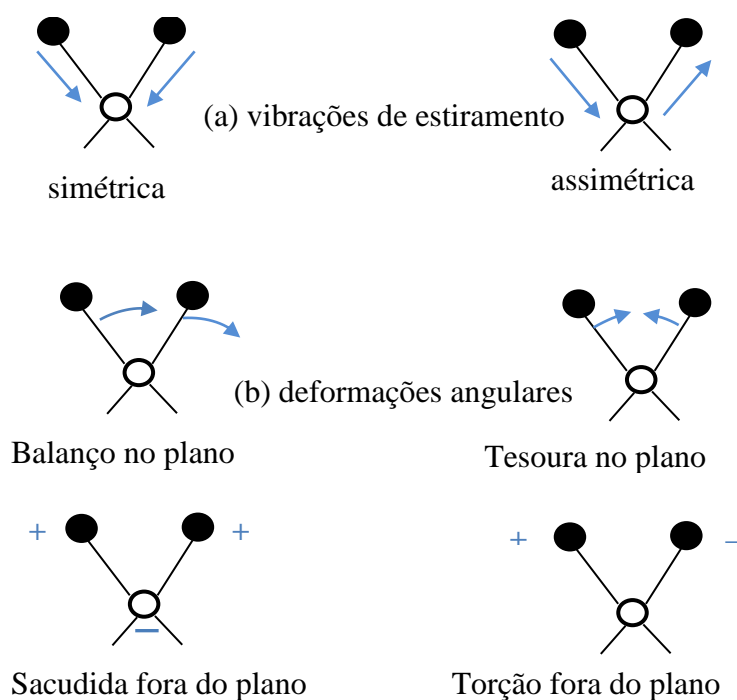


Figura 4: Representação das (a) vibrações de estiramento e (b) deformações angulares^{a37}.

Os espectros obtidos na região do MIR apresentam picos bem definidos, as bandas são intensas e os picos possuem altas absorbâncias facilitando a interpretação dos espectros. A faixa espectral do MIR conhecida como região de impressão digital, compreendida entre $1.200-600\text{ cm}^{-1}$, é bastante útil para identificação de compostos, visto que pequenas diferenças na estrutura e na constituição das moléculas resultam em mudanças significativas no perfil e na distribuição dos picos de absorção. Deste modo é possível identificar, por exemplo, um composto orgânico comparando-se seu espectro com banco de dados pré-existent³⁷.

Os métodos mais utilizados para obtenção dos espectros no infravermelho envolvem a transmissão ou a reflexão. No primeiro caso e, mais antigo, utilizado para fins analíticos, a radiação passa através da amostra, onde parte da radiação é absorvida e outra transmitida. Para isto, a amostra deve ser colocada em uma janela cujo material seja transparente na região utilizada do IR e compatível com a amostra. Já para o segundo caso as informações são obtidas através da reflexão da radiação incidente, por três diferentes modos: reflexão total atenuada (ATR), reflexão especular (RE) e reflexão difusa (DR)^{36,39}.

3.4.1 Espectroscopia por Reflexão Total Atenuada (ATR)

A espectroscopia por reflexão total atenuada é uma técnica extremamente robusta e segura para estudos quantitativos e qualitativos envolvendo líquidos. Destaca-se por não envolver processos laboriosos de preparo das amostras³⁶.

O princípio da espectroscopia por ATR está nas múltiplas reflexões da radiação infravermelha que ocorrem no interior de cristais, de materiais com alto índice de refração como, por exemplo, o seleneto de zinco (ZnSe), interagindo apenas com a amostra que estiver superficialmente no cristal⁴⁰.

Um acessório de reflexão total atenuada opera medindo as mudanças que ocorrem em um feixe de infravermelho que sofre reflexão interna total, ao entrar em contato com a amostra. Embora ocorra a reflexão, parte da radiação penetra além da superfície do cristal de ATR, interagindo com o material absorvente (amostra)⁴⁰.

A profundidade de penetração (dp) é dada pela Equação 1:

$$dp = \frac{\lambda}{2\pi\eta_1 \left[\text{sen}^2\theta_1 - \left(\frac{n_2}{n_1}\right)^2 \right]^{\frac{1}{2}}} \quad (1)$$

onde θ_1 é o ângulo de incidência, η_1 é o índice de refração do elemento de ATR e η_2 índice de refração da amostra.

Esta radiação penetrante é chamada de onda evanescente. Se a amostra absorve determinado comprimento de onda evanescente, ocorre uma atenuação do feixe cujo comprimento de onda corresponde às bandas de absorção no infravermelho. É em função deste mecanismo que surge o espectro⁴⁰. Na Figura 5 está à representação simplificada da análise por ATR.

Recentemente, a técnica de ATR tem revolucionado as análises de amostras líquidas e sólidas porque combate os aspectos mais desafiadores da análise no infravermelho, como a preparação das amostras e a reprodutibilidade espectral. Destaca-se por empregar pequena quantidade de amostra e por não envolver processos laboriosos no preparo destas⁴¹.

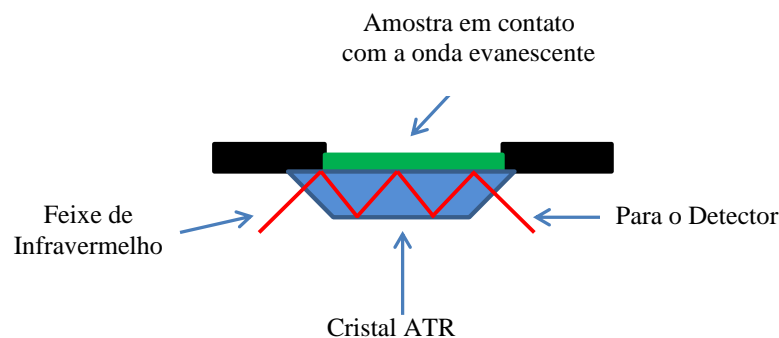


Figura 5: Representação simplificada da análise por ATR⁴².

3.5 QUIMIOMETRIA

A quimiometria pode ser definida como uma área da química que combina métodos matemáticos e estatísticos para maximizar as informações que podem ser extraídas de um conjunto de dados químicos. Os métodos quimiométricos podem ser aplicados em dados multivariados com os propósitos qualitativos (análise exploratória e reconhecimento de padrões) e quantitativos (calibração multivariada)⁴³.

3.5.1 Calibração Multivariada

A calibração multivariada emprega medidas realizadas em múltiplos canais na construção do modelo para relacionar concentração e sinal instrumental. Esse processo surge como alternativa capaz de superar as limitações univariadas, permitindo as determinações simultâneas de analitos com maior sensibilidade e confiabilidade na presença de interferentes. Para isso, os interferentes devem também estar presentes na etapa de calibração, o que é conhecido como vantagem de primeira ordem⁴⁴.

Na construção de modelos de calibração multivariada (calibração de primeira ordem) a informação analítica associada a cada amostra corresponde a um vetor. Os dados a serem modelados são dispostos em uma matriz, que contém, por exemplo, dados de espectros para o conjunto de amostras. A matriz de dados é composta de tal modo que cada linha corresponde a uma amostra e cada coluna contém a informação referente a um canal analítico. No exemplo de espectros, as colunas correspondem aos comprimentos de onda. Convencionalmente, a matriz de variáveis independentes são comumente chamadas de matriz de respostas instrumentais e é denominado matriz X. O vetor contendo a variável dependente ou parâmetro de referência é denominado por Y⁴⁵.

Na calibração multivariada umas das principais técnicas de análise utilizada para a elaboração dos modelos é por mínimos quadrados parciais (PLS, do inglês *Partial Least Squares*), e a utilização desse método permitiu significativo progresso na aplicação da quimiometria⁴⁶.

3.5.2 Regressão por Mínimos Quadrados Parciais (PLS)

Os métodos de regressão são utilizados para analisar a relação entre uma variável dependente (concentração ou propriedade de interesse) e um conjunto de variáveis independentes (sinais instrumentais), sendo utilizado para quantificar propriedades de interesse de um conjunto de amostras¹⁸. Entre os métodos de regressão, destaca-se o método por PLS, sendo o mais utilizado na atualidade para análise quantitativa⁴⁷.

O método PLS possui grande vantagem que é a robustez, isso permite que os parâmetros de modelo não se alterem de maneira significativa quando novas amostras são adicionadas ou subtraídas do conjunto de calibração⁴⁸.

Os princípios da regressão por mínimos quadrados parciais estão na decomposição de matriz de dados X (variáveis independentes) e de Y (variável dependente) em uma soma de produtos entre dois vetores: t (escores) e p (pesos), conforme é apresentado nas equações 2 e 3. A esses vetores, soma-se, ainda, uma matriz de resíduos (E) que corresponde à parte não modelada da matriz X. Na regressão por PLS, tanto a matriz de dados espectroscópicos X, como a matriz de propriedades de interesse Y são decompostas em um novo sistema de coordenadas⁴⁰.

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_h p_h^T + E \text{ ou } X = TP^T + E = \sum t_h p_h^T + E \quad (2)$$

$$Y = UQ^T + F = \sum u_h q_h^T + F \quad (3)$$

onde T e U são as matrizes de escores das matrizes X e Y, respectivamente; P e Q são as matrizes de pesos das matrizes X e Y, respectivamente; h corresponde ao número de variáveis latentes e E e F correspondem às matrizes de resíduos.

A relação interna dos escores das matrizes X e Y é obtida através do coeficiente de regressão linear para cada componente principal (b_h) de acordo com a equação 4:

$$u_h = b_h t_h \quad (4)$$

sendo b_h obtido através da Equação 5:

$$b_h = \frac{u_h^T t_h}{t_h^T t_h} \quad (5)$$

onde, u e t são elementos das matrizes U e T , respectivamente.

Os valores de b_h são agrupados na matriz diagonal B (matriz identidade), que contém os coeficientes de regressão entre a matriz de escores U e Y e a matriz de escores T de X . Entretanto, como cada matriz é decomposta separadamente, pode ocorrer uma relação não linear entre os escores dos dois blocos. A melhor relação linear possível entre os escores das matrizes é obtida através de pequenas rotações das variáveis latentes das matrizes X e Y , devendo existir um compromisso entre a capacidade de descrever as amostras individuais (modelagem dos blocos X e Y) e o aumento da correlação entre T e U ⁴⁹.

A matriz Y pode ser calculada por meio das informações contidas em u_h (3), conforme equação 6:

$$Y = TBQ^T + F \quad (6)$$

onde, T são os escores da matriz X , B é a matriz identidade de b_h , Q^T são os pesos da matriz Y e F é a matriz residual de Y . Dessa forma, a concentração das novas amostras pode ser prevista a partir dos novos escores de X , dado por T^* , substituindo na equação 7:

$$Y = T^*BQ^T \quad (7)$$

Para se obter o melhor modelo referente aos dados sob investigação devem ser avaliados o número de variáveis latentes (VLs) e a Raiz Quadrada dos Erros Médios (RMSE, do inglês, Root Mean Square Error).

3.5.3 Métodos de Seleção de Variáveis em Modelos de Regressão

As técnicas espectroscópicas geralmente originam um grande número de variáveis, tornando o processo de regressão muito demorado. Além disso, algumas variáveis contém informações que não são pertinentes ao analito (ruído) ou podem ser colineares (informações redundantes). Adicionalmente, existem regiões espectrais, nas quais as variações das concentrações de um determinado composto não causam variações nos valores de absorvância⁴⁹.

Deste modo, a habilidade de previsão do modelo de calibração é afetada pelo excesso de variáveis, podendo ter seu desempenho diminuído. Em contrapartida, a escolha das regiões espectrais ou de variáveis específicas pode melhorar a eficiência dos modelos de calibração.

Os benefícios da seleção espectral adequada relacionam-se com a estabilidade do modelo de calibração em relação à colinearidade, como também a vantagem desse método permitir produzir um modelo mais robusto, simples de interpretar e com menores erros de previsão^{49,50,51}.

Existem vários procedimentos para a seleção de regiões espectrais. Muitos trabalhos vêm usando alguns desses métodos: o Algoritmo Genético (GA, do inglês *Genetical Algorithm*), Algoritmo das Projeções Sucessivas (SPA, do inglês *Successive Projections Algorithm*), Método de Eliminação de Variáveis não Informativas por Mínimos Quadrados Parciais (UVE – PLS, do inglês *Elimintion of Uninformative Variable in Partial Least Square*), Método de Mínimos Quadrados Parciais por intervalo (iPLS, do inglês *Interval Partial Least Square*) e do Mínimo Quadrado Parcial por sinergismo de intervalos (siPLS, do inglês *synergy interval Partial Least Square*).

Neste trabalho foram utilizados os métodos de iPLS e siPLS para a seleção de variáveis, sendo que esses métodos selecionam faixas do espectro e não variáveis isoladas, e também, por se tratarem de métodos mais simples em relação aos demais métodos citados acima.

O iPLS é uma extensão desenvolvida para o PLS, onde é feita uma regressão por mínimos quadrados parciais em cada intervalo equidistante ao longo de toda a extensão do espectro. Desse modo, é avaliada a importância da informação nas diferentes subdivisões espectrais, de onde é possível identificar e selecionar o intervalo cujas variáveis exibem informações mais importantes. Um modelo de PLS é constituído para cada intervalo, sendo apresentados os resultados na forma de gráfico, dessa forma, facilitando a comparação com toda a faixa espectral^{52,53}.

O siPLS é uma extensão do algoritmo iPLS. Esse algoritmo consiste na divisão de espectro em regiões equidistantes (intervalos) seguido de combinações desses intervalos, possibilitando a obtenção de modelos que podem apresentar melhores coeficientes de determinação (R^2), menores erros de calibração e previsão que aqueles encontrados por iPLS⁵⁴.

3.5.4 Seleção do Conjunto de Calibração e Previsão

A seleção das amostras que irão compor os conjuntos de calibração e previsão é feita através do algoritmo Kennard-Stone⁵⁵. Nesse Algoritmo, a primeira amostra selecionada pelo algoritmo é a que apresenta a maior distância em relação à amostra média. A segunda amostra a ser selecionada será a que apresentar maior distância em relação à primeira amostra selecionada. A próxima amostra a ser selecionada apresentará maior distância em relação à última amostra selecionada, e assim sucessivamente até atingir o número de amostras desejadas⁵⁶. Esse algoritmo é aplicado para inicialmente formar o grupo de calibração, e por consequência a seleção das amostras de maior variabilidade, ou seja, as amostras mais “externas” do conjunto total.

3.5.5 Tratamento e Pré-Processamento dos Dados

Medidas em diferentes unidades e variáveis com diferentes variâncias podem afetar a distribuição e uniformidade dos dados espectrais. Este fato pode ocasionar dificuldades na análise dos dados, interferindo na extração de informações relevantes, bem como suas interpretações⁵⁷. Deste modo, há a necessidade de tratar os dados a serem analisados por meio de pré-tratamentos.

O pré-tratamento tem como objetivo reduzir matematicamente as variações indesejáveis que não foram removidas durante a aquisição dos dados e que não são eliminados naturalmente durante a análise, mas que podem influenciar os resultados finais. O tratamento dos dados é aplicado na fase de calibração para melhorar a eficiência do método, produzindo um modelo de regressão com maior capacidade de previsão das amostras, também têm por objetivo remover variações sistemáticas não desejadas do espectro, como mudanças na linha de base, efeito de espalhamento e fatores não controláveis⁵⁸. No entanto, qualquer tipo de tratamento deve ser feito com cuidado e criteriosamente, para não distorcer, comprometer ou até mesmo inutilizar os resultados da análise.

Existem várias técnicas matemáticas comuns que podem ser aplicadas com o objetivo de reduzir tanto as variações aleatórias quanto as informações sistemáticas indesejáveis. A componente aleatória (ruído) pode ser reduzida por meio das técnicas de alisamento (do inglês *Smoothing*)⁵⁸.

O alisamento tem grande aplicação na área espectroscópica para redução da componente aleatória dos dados e o conseqüente aumento da razão sinal ruído do sinal analítico. O alisamento também pode ser utilizado para destacar pequenos picos que estão escondidos numa linha de base onde o ruído é alto e, ainda, como um método de compressão de dados⁵⁸.

Entre os diversos métodos de alisamento existente, neste trabalho se utilizou o alisamento pelo método de Savitzky-Golay⁵⁹, que ajusta um polinômio de baixa ordem aos pontos de uma janela pelos mínimos quadrados. A escolha pelo número de pontos é de suma importância, pois um número elevado pode acarretar perda de informação e um número reduzido, a permanência de ruído.

Outro pré-tratamento utilizado foi o de Correção do Espalhamento do Sinal (MSC). A transformação MSC é usada para corrigir efeitos de espalhamento aditivo e multiplicativos na absorvância. Esses efeitos de espalhamento são causados por fenômenos físicos, como mudança no caminho ótico, na sensibilidade do detector e do amplificador, variações na temperatura e na pressão e diferenças no tamanho e na forma das partículas sólidas (pós e grãos), bem como de emulsão e dispersão, e não têm a ver com a composição das amostras⁶⁰.

Esses efeitos aditivos e multiplicativos de espalhamento são dependentes do comprimento de onda e devem ser removidos, pois constituem fontes de variabilidade irrelevante para o problema⁶⁰.

O pré-processamento é aplicado quando se deseja comparar variáveis com diferentes dimensões e consiste, basicamente, em centrar os dados na média ou autoescalar os dados. Centrar os dados na média consiste em calcular a média das intensidades para cada comprimento de onda e subtraírem-se do respectivo valor médio. Desta forma, ocorre a mudança do sistema de coordenadas para o centro dos dados, ocorrendo um deslocamento para o centro da informação descrita pelos dados originais, sendo que não ocorre distorção da informação disponível, não alterando a informação para o modelo. O segundo pré-processamento, autoescalar os dados, significa centrar os dados na média e dividi-los pelo respectivo desvio padrão, sendo um para cada comprimento de onda. O efeito causado pelo autoescalamamento é que todos os comprimentos de onda terão igual peso, não importando se eles representam um pico, um espalhamento ou simplesmente um ruído de linha de base, por conseqüência não se torna aconselhável para dados espectrais^{57,61}.

Um problema instrumental, de amostragem, ou até devido à morfologia das partículas pode deslocar um espectro como um todo, de um valor constante positivo ou negativo em relação ao zero de absorvância, podendo haver um deslocamento constante da linha de base (no eixo das ordenadas). Esse deslocamento pode ser corrigido tomando-se a primeira derivada de espectro. Como a primeira derivada de uma constante é zero, após a correção o espectro resultante estará deslocado para o zero de absorvância. Caso o espectro apresente uma inclinação da linha de base à medida que decresce o número de onda, essa variação pode ser corrigida tomando-se a segunda derivada. Com isso, as informações contidas ao longo dos diferentes comprimentos de onda são geralmente acentuadas. Contudo, a principal desvantagem desse processo é que o ruído uma vez presente nos espectros originais, pode se tornar acentuado, diminuindo a relação sinal/ruído por isso do cuidado com a qualidade dos espectros quando se pretende aplicar o cálculo das derivadas^{40,58}.

Outro pré-processamento muito empregado é a normalização. Nessa transformação os valores de cada uma das variáveis de uma dada amostra são divididos por um fator de normalização, como, por exemplo, pela norma dessa amostra. Dessa forma, todas as amostras estarão em uma escala pré-determinada⁵⁸.

3.5.6 Avaliação dos Modelos de Regressão

A avaliação de alguns parâmetros resultantes da construção dos modelos permite a escolha e a seleção de modelos mais adequados. Um aspecto importante da etapa de validação numa regressão multivariada é a decisão sobre a dimensionalidade do modelo⁶².

A determinação do número de variáveis latentes (VLs) é um fator relevante para a qualidade do modelo quando se utiliza uma técnica de calibração baseada no método PCA, como é o caso do PLS. Existe para cada modelo um número ótimo de fatores, com informações suficientes para o modelamento, mas podendo ocorrer os casos onde se contenha um número de VLs inferior ao ideal, ocorrendo o subajuste, como também o caso de um modelo com número superior de VLs ao ideal ocorrendo um sobreajuste⁶³.

Assim, num modelo com subajuste, com número de VLs inferior ao ideal, se terá um modelo que não irá considerar a totalidade da informação contida no conjunto de dados. Mas em geral, a decisão sobre um número de VLs superior ao ideal é o que ocorre com mais frequência, ocasionando a construção de modelos de regressão com resultados satisfatórios, isto é, com baixo valor de erro na determinação da propriedade de interesse de amostras do conjunto de calibração, mas que apresente elevado valor de erro na previsão da propriedade

de interesse de amostras de um conjunto de validação ou previsão, devido às informações desnecessárias incluídas no modelo ao selecionar maior número VLs que o necessário^{63,64}.

Um gráfico muito usado na literatura para ilustrar o comportamento do erro de previsão em modelos tendenciosos é apresentado na Figura 6. No início, quando o número de fatores é pequeno e o erro de previsão é alto, ocorrendo o subajuste. À medida que os fatores são adicionados, incorporam-se informações relevantes ao modelo, e naturalmente a variância explicada aumenta, enquanto a tendência e o erro de previsão diminuem. Após um número ótimo de fatores quando a curva passa pelo mínimo o erro de previsão volta a crescer, indicando que está ocorrendo um sobreajuste. A Figura 6 mostra esquematicamente como essas contribuições opostas na qualidade do modelo devem ser harmonizadas para que ele apresente um balanço adequado entre a tendência e a variância, e esteja otimizado para fazer previsão⁵⁸.

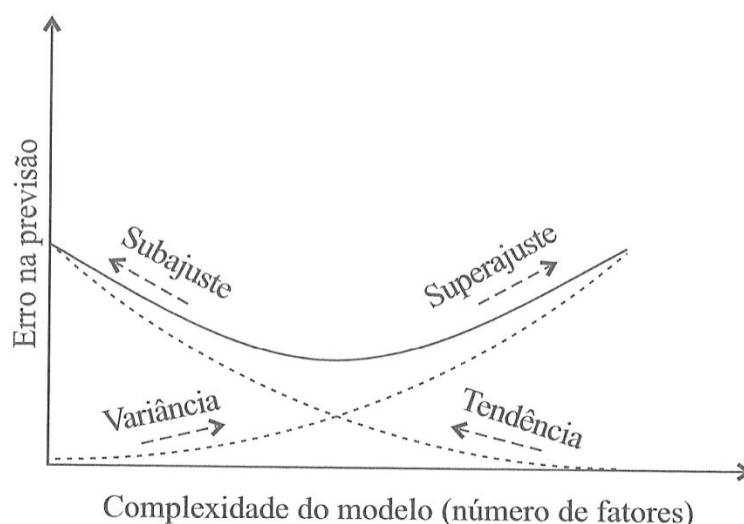


Figura 6: Comportamento da variância e tendência (viés) nos modelos PLS em geral, indicado pelas linhas tracejadas. A linha cheia indica o comportamento de erro na previsão⁵⁸.

Por meio do número de VLs é feito o cálculo da Raiz Quadrada dos Erros Médios (RMSE, do inglês, *Root Mean Square Error*), que é utilizado para avaliação dos modelos construídos. Este erro pode ser estimado através dos conjuntos: Raiz Quadrada do Erro Médio de Calibração (RMSEC, do inglês, *Root Mean Square Error of Calibration*); Raiz Quadrada do Erro de Predição (RMSEP, do inglês *Root Mean Square Error of Prediction*) e Raiz Quadrada do Erro de Validação Cruzada (RMSECV, do inglês *Root Mean Square Error of Cross Validation*)⁴⁴. O cálculo do RMSE é dado conforme equação 8:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (8)$$

onde, y_i é o valor de referência para a i -ésima amostra e \hat{y}_i é o valor de previsão para esta mesma amostra, sendo n o número total de amostras utilizadas para a construção do modelo.

A validação cruzada está baseada na avaliação da magnitude do erro de predição de um dado modelo de regressão e consiste na remoção de uma ou de um bloco de amostras do conjunto de calibração, e subsequente construção do modelo de regressão. As amostras removidas são previstas no modelo e calcula-se o erro. Esse procedimento ocorre até que todas as amostras do modelo sejam retiradas e previstas, e então, calcula-se o RMSECV. O procedimento é repetido utilizando n número de VLs ($n = 1, 2, 3, \dots, n$), onde n é definido previamente, sendo escolhido o número de VLs no qual se obtém o menor erro (RMSECV)⁶².

Após escolha do número de VLs através da validação cruzada, constroem-se os modelos de regressão a partir das amostras do conjunto de calibração. Os modelos gerados são avaliados pelo cálculo da RMSEC, segundo equação 8, onde são comparados os valores obtidos no método de referência com os valores estimados pelo modelo de regressão.

Através do cálculo da RMSEP, conforme equação 8, é realizada a avaliação da capacidade de predição do modelo de calibração para as amostras de predição. Nesse caso são utilizadas amostras independentes daquelas utilizadas para construção do modelo de regressão, permitindo avaliar a capacidade do modelo de prever amostras desconhecidas.

O coeficiente de determinação (R^2), definido na equação 9, é uma ferramenta estatística utilizada na avaliação de modelos de regressão. Ele indica o quão bem o modelo de regressão proposto se ajusta a um modelo linear, o quão mais próximo a valor da razão ser de 1 (um)⁶⁴.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

em que y_i é o valor de referência para a i -ésima amostra, \hat{y}_i é o valor de previsão para essa mesma amostra, sendo n o número total de amostras utilizadas para a construção do modelo, e \bar{y} é o valor médio das n amostras utilizadas na construção do modelo.

Portanto, o modelo ideal de regressão é o que possui baixo número de VLs, valores de RMSECV, de RMSEC e de RMSEP o mais baixo possível e de mesma magnitude, pois diferenças entre RMSECV e RMSEP indicam problemas de subajuste ou sobreajuste. Assim como o valor mais próximo de 1 (um) possível de R^2 .

4 MATERIAIS E MÉTODOS

4.1 AMOSTRAS

As amostras de taninos de acácia negra e quebracho foram fornecidas pela empresa TANAC/SA, do Município de Montenegro, Rio Grande do Sul. Essas amostras são extratos de taninos das plantas onde o processo de extração foi realizado na empresa. No total, foram utilizadas 64 amostras, sendo 31 amostras de taninos de quebracho e 33 amostras de taninos de acácia negra.

4.2 ESPECTROSCOPIA DE INFRAVERMELHO MÉDIO

Os espectros de Infravermelho Médio com Transformada de Fourier e Reflexão Total Atenuada (ATR-FTIR) foram obtidos mediante espectrofotômetro (Perkin Elmer modelo Spectrum 400), acoplado a um acessório de refletância total atenuada (ATR) com cristal de seleneto de zinco (ZnSe) empregando *laser* gás hélio/neon. Os espectros foram obtidos em duplicata para cada amostra de tanino na faixa de $4000 - 700 \text{ cm}^{-1}$, a temperatura ambiente, com resolução de 4 cm^{-1} e 16 varreduras.

4.3 ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO

Os espectros no NIR foram obtidos mediante espectrofotômetro por reflexão do infravermelho próximo (Foss Near-IR, modelo 6500). Todas as amostras foram medidas em duplicatas na faixa de 1200 a 2500 nm, a temperatura ambiente, com resolução de 4 nm e 32 varreduras. Posteriormente a escala em frequência (nm) dos espectros foi convertida para número de onda (cm^{-1}).

4.4 ANÁLISE MULTIVARIADA

A análise multivariada das amostras de taninos foi realizada a partir dos dados de ATR-FTIR e NIR, sendo com esses dados construídos modelos de regressão multivariada por iPLS e siPLS para as concentrações de nitrogênio, enxofre, cinzas, carbono e hidrogênio.

O software utilizado para estes estudos foi o Matlab 7,11 (MathWorks Inc., Natick, MA), pacote iToolbox e computador com processador Intel Core i7 2,40 GHz.

Todos os espectros foram centrados na média, suavizados (algoritmo Savistky-Golay, 15 pontos, polinômio de primeira ordem) e normalizados (entre 0-1).

4.4.1 Regressão multivariada do nitrogênio, enxofre, cinzas, carbono e hidrogênio por ATR-FTIR e NIR.

Considerando o total de 64 amostras de taninos, 42 amostras foram selecionadas para a construção dos modelos de calibração e 22 amostras foram utilizadas para avaliar a predição dos modelos. Foram utilizadas as médias das duplicatas dos espectros. A seleção das amostras de calibração e previsão foram realizadas utilizando o algoritmo de Kennard-Stone⁵⁵.

Para a construção dos modelos de regressão por iPLS e siPLS foram avaliados a aplicação de duas alternativas de pré-processamentos aos espectros: apenas centrar os dados na média e o tratamento de correção do espalhamento de luz (MSC) aliado a centrar na média.

Os modelos foram construídos a partir dos dados de ATR-FTIR e NIR separadamente, como também, a construção de modelos a partir da fusão das duas faixas espectrais. Os modelos iPLS foram construídos realizando 2, 4, 8, 16, 32 divisões equidistantes do espectro, para cada caso escolheu-se apenas um intervalo, de acordo com os menores valores de RMSECV, do Rc^2 e o menor número de variáveis latentes (VLs). Os modelos de regressão por siPLS foram obtidos realizando 2, 3 ou 4 combinações entre divisões espectrais que foram de 8, 16 e 32 intervalos, escolheu-se a combinação de intervalos que obteve o modelo com menor RMSECV, do Rc^2 e o menor número de variáveis latentes e também se considerou as regiões que possuíam informação espectral mais relevante.

O número de variáveis latentes para cada modelo de regressão foi determinado pelo procedimento da validação cruzada *leave-one-out*, escolhendo o menor número de VLs com que os modelos de regressão apresentam menor RMSECV. Tendo o cuidado para que não houvesse um subajuste ou um sobreajuste.

Após construção dos modelos de regressão por iPLS e siPLS, o melhor modelo foi definido como aquele que possui menores valores de RMSECV e RMSEP, sendo desejado que não ocorram grandes diferenças entre os erros para não ocorrer um subajuste ou sobreajuste. Também foi levado em consideração o número de VLs, onde o ideal é a utilização do menor número possível, assim como os maiores valores de Rc^2 e Rp^2 .

4.5 TEORES DE NITROGÊNIO, ENXOFRE, CARBONO, HIDROGÊNIO E CINZAS.

A determinação das concentrações de nitrogênio, enxofre, cinzas, carbono e hidrogênio nos extratos dos taninos foram realizados no CEPAC (Centro de Excelência em Pesquisa e Inovação sobre Petróleo), PUCRS (Pontifícia Universidade Católica do Rio Grande do Sul).

A determinação de cinzas foi realizada pelo método de gravimetria, conforme ABNT NBR 11042/2013¹⁰. Inicialmente, se calcinou a cápsula de porcelana a 800 °C por aproximadamente 1 h, desligou-se o forno-mufla (marca Lavoisier, modelo 402B) e se aguardou até a temperatura atingir aproximadamente 300 °C, colocou-se no dessecador até esfriar e se mediu a massa do cadinho. Foi pesado aproximadamente 1 g, com precisão de 0,0001g, da amostra sólida no cadinho previamente tarado e se iniciou a calcinação em bico de Bunsen, inicialmente de forma mais branda e posteriormente a alta temperatura. Em seguida foi levado a forno-mufla a 800 °C por um período de 3 h, desligou-se a mufla e quando atingiu 300 °C foi levado ao dessecador para esfriar, e posteriormente mediu-se a massa. O procedimento foi realizado em duplicata.

O cálculo é dado pela expressão:

$$TC\% = \frac{m_r}{M} \times 100$$

Onde:

TC é o teor de cinzas, expresso em percentual (%); m_r é a massa do resíduo (cinzas), expresso em gramas (g); e M é a massa da amostra, expressa em gramas (g).

A determinação de nitrogênio, carbono e hidrogênio foi realizada num analisador elementar Leco (modelo Truspec), o enxofre foi determinado num analisador elementar Leco (modelo SC-632). Inicialmente as amostras foram pulverizadas e peneiradas em peneira de 60

mesh. Até o momento da análise as amostras ficaram acondicionadas em um dessecador. As amostras foram colocadas em folhas de alumínio e introduzidas no equipamento. Ao ser introduzido no equipamento ocorreu a oxidação das amostras em altas temperaturas, de 950 °C para N, C e H, e de 1350 °C para S. Após a combustão total as concentrações foram medidas por um detector de infravermelho.

5 RESULTADOS E DISCUSSÃO

5.1 TEORES DE NITROGÊNIO, ENXOFRE, CARBONO, HIDROGÊNIO E CINZAS

A determinação de cinzas foi realizada pelo método de referência ABNT NBR 11042/2013¹⁰. Os resultados são apresentados no Anexo (Tabela 11). Os teores de cinzas variam na faixa de 2,5 a 9% de cinzas nas amostras de taninos. Esses resultados foram utilizados como dados de referência para a construção do modelo de regressão multivariada.

A determinação de nitrogênio (N), enxofre (S), carbono (C) e hidrogênio (H) foi realizada por análise elementar CHNS. Os resultados são apresentados nas Tabelas 12, 13, 14 e 15, em Anexo. O teor de N, S, C, H nas amostras variam na faixa de 0,0440 - 0,6282%, 0,0984 – 3,3825%, 46,5090 – 51,2830%, 4,9128 - 5,5588% respectivamente.

5.2 ESPECTROSCOPIA DE INFRAVERMELHO

Os espectros de NIR foram obtidos na região de 8500 – 4000 cm^{-1} a partir das amostras de taninos, Figura 7. A Figura 7a representa os espectros NIR sem tratamento dos dados e a Figura 7b representa os espectros NIR após tratamento dos dados, esses foram suavizados e normalizados. É possível verificar diferenças entre os espectros tratados e não tratados, sendo que nos tratados ocorreu a redução do ruído. Todos os espectros das amostras ficaram definidos numa escala estabelecida de 0-1.

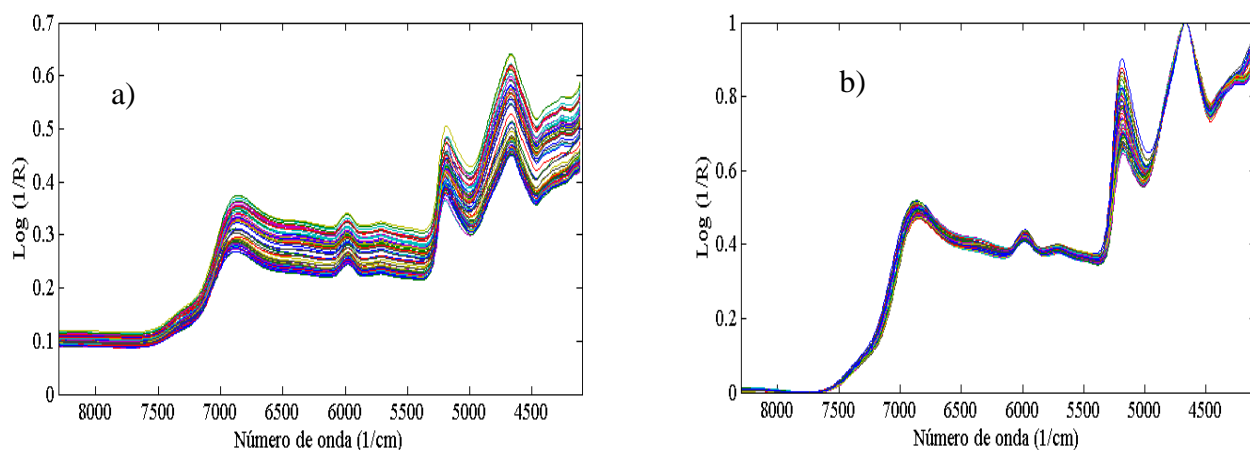


Figura 7: Espectros de NIR para amostras de taninos: a) espectro sem tratamentodos dos dados; b) espectro com tratamento dos dados; espectro suavizado e normalizado.

Os espectros ATR-FTIR foram obtidos na região de $4000 - 700 \text{ cm}^{-1}$ a partir das amostras de taninos, Figura 8. A Figura 8a representa os espectros ATR-FTIR sem tratamento dos dados e a Figura 8b representa os espectros ATR-FTIR após tratamentos dos dados, esses foram suavizados e normalizados. Assim como os espectros de NIR, também é possível verificar grandes diferenças entre os espectros tratados e não tratados.

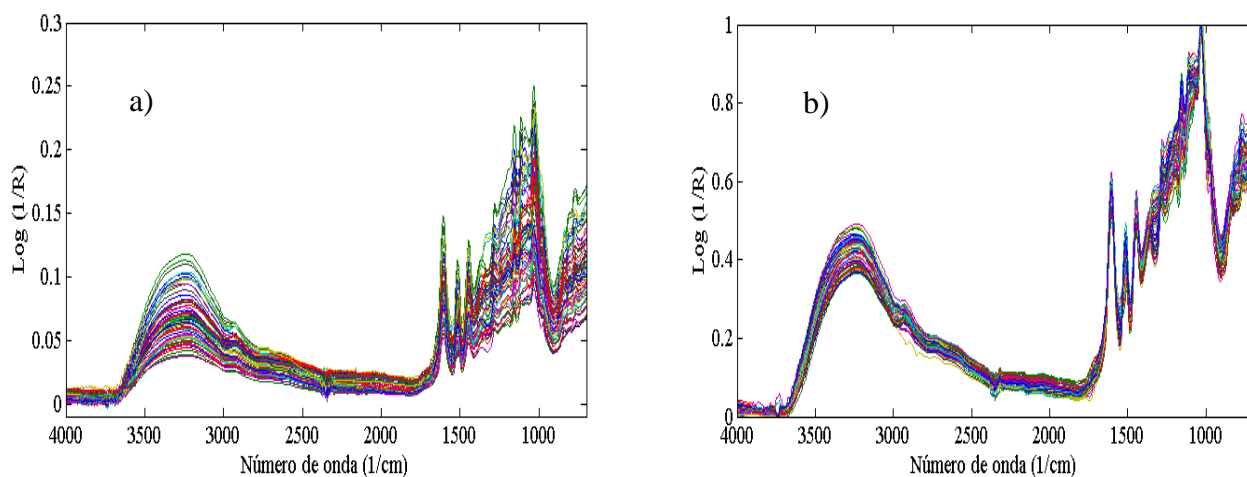


Figura 8: Espectros de ATR-FTIR para amostras de taninos: a) espectro sem tratamentodos dos dados; b) espectro com tratamento dos dados; espectro suavizado e normalizado.

Os dados NIR e ATR-FTIR foram avaliados tanto separadamente como também conjuntamente. A fusão das duas regiões teve como objetivo a seleção de regiões relevantes de cada conjunto de dados e a tentativa de construção de modelos ainda melhores, ou seja, com menores erros e maiores coeficientes de determinação.

Inicialmente os modelos de fusão foram construídos empregando a faixa de NIR (7396 – 4390 cm^{-1}) adicionando a faixa de FTIR (1700 – 761 cm^{-1}). Num segundo momento empregando método de diminuição da resolução para os dados FTIR, os modelos foram construídos combinando-se a região de NIR (7396 – 4390 cm^{-1}) com FTIR (3483 – 700 cm^{-1}).

5.3 ANÁLISE MULTIVARIADA

5.3.1 Determinação do Teor de Nitrogênio por NIR, ATR-FTIR e Fusão por Regressão Multivariada.

Foram construídos modelos de regressão por PLS a partir dos dados de NIR, ATR-FTIR como também a partir da fusão dos dois conjuntos de dados. Os modelos com fusão dos dados (NIR e ATR-FTIR) ainda foram avaliados com a redução de resolução dos dados ATR-FTIR.

Na Tabela 2 são destacados os melhores modelos de regressão, para determinação de nitrogênio, para cada conjunto de dados. A totalidade dos resultados dos modelos realizados é apresentado no anexo 8.2. É possível verificar as principais figuras de mérito para os quatro conjuntos de dados utilizados para construir os modelos de regressão multivariada.

A utilização do pré-tratamento com MSC apresentou resultados melhores nos modelos para os dados de NIR, Fusão NIR–ATR-FTIR e Fusão NIR–ATR-FTIR com resolução 1:2. Para os dados de ATR-FTIR não se observou melhoras nos modelos, além de aumentar o tempo de processamento.

Através da Tabela 2 é possível verificar os resultados dos melhores modelos de cada conjunto de dados. Os dados de NIR apresentaram modelos superiores comparados aos de ATR-FTIR, pois apresentaram valores de coeficientes de determinação maiores e erros de validação cruzada e de previsão menores. Avaliando os modelos gerados a partir da fusão dos dados, verificou-se que não apresentaram melhoras em comparação a utilização apenas dos dados de NIR. Portanto definiu-se que o melhor modelo gerado foi o que utilizou dados de NIR com MSC, para determinação de nitrogênio em taninos. O destaque em negrito na Tabela 2 sinaliza o melhor modelo construído.

Tabela 2: Resultados das figuras de mérito dos melhores modelos para determinação do teor de nitrogênio com dados NIR, ATR-FTIR, Fusão NIR-ATR-FTIR e Fusão NIR-ATR-FTIR resolução 1:2.

Dados	Tratamento ^a	Melhores Modelos ^b	Intervalos selecionados	VLS ^c	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
NIR	s/MSC	s2iPLS32	8, 30	9	0,956	0,036	0,946	0,043
	c/MSC	s3iPLS32M	8, 24, 30	7	0,962	0,034	0,944	0,043
ATR-FTIR	s/MSC	s3PLS32	29, 30, 31	6	0,936	0,039	0,927	0,050
	c/MSC	s3PLS32M	7, 26, 31	7	0,936	0,040	0,918	0,062
Fusão NIR-ATR-FTIR	s/MSC	s4iPLS8	1, 2, 3, 5	10	0,932	0,042	0,903	0,054
	c/MSC	s2iPLS32M	4, 7	7	0,947	0,037	0,923	0,047
Fusão NIR-ATR-FTIR resolução 1:2	s/MSC	s4iPLS16	2, 3, 4, 12	8	0,935	0,042	0,921	0,050
	c/MSC	s2iPLS8M	1, 2	5	0,933	0,042	0,929	0,047

^aTratamento - modelos com e sem tratamento por MSC; ^b iPLS – número de divisões do espectro, s – número de intervalos combinados, M – tratamento com MSC; ^c número de variáveis latentes; Formação em negrito ao melhor modelo de regressão.

5.3.1.1 Avaliação dos modelos de regressão por iPLS e siPLS a partir de dados NIR para determinação de nitrogênio em taninos.

Os dados NIR foram utilizados para produzir os modelos de regressão multivariada e as amostras dos conjuntos de calibração e previsão foram selecionados pelo algoritmo de Kennard-Stone⁵¹; os espectros foram suavizados, normalizados e centrados na média. Os resultados dos modelos de regressão por iPLS e siPLS foram construídos por validação cruzada *leave-one-out* para a predição do teor de nitrogênio. Os resultados estão organizados pelo número de divisões equidistantes feitas no espectro e, no caso do siPLS, de suas combinações de 2, 3 ou 4 entre os intervalos. Os modelos foram avaliados com e sem

tratamento por MSC, os resultados sem MSC são apresentados na Tabela 16 do anexo 8.2 e os com MSC são apresentados na Tabela 3.

Pelos resultados obtidos, Tabela 3, pode-se observar que quando ocorre a combinação dos intervalos (siPLS) obtém-se modelos com menores erros (RMSECV e RMSEP), assim como valores mais altos dos coeficientes de determinação comparados ao método de seleção de variáveis por iPLS.

A partir dos resultados das figuras de mérito dos modelos de regressões usando o método de iPLS e siPLS com e sem tratamento de MSC definiu-se que o melhor modelo foi o s3iPLS32M utilizando dados NIR, onde ocorreu a divisão em 32 parte equidistantes, combinação de 3 intervalos com tratamento de MSC. Modelos como o s4iPLS32M, com NIR, possuem valores de erros semelhantes e coeficientes de determinação um pouco superiores comparado ao modelo s3iPLS32M, mas um fator determinante para a não escolha desse é a utilização de 11 VLs o que torna o modelo sobreajustado, inserindo informações que podem não ser relevantes, além de usar faixas espectrais que não possui informações químicas. Os intervalos selecionados do modelo s3iPLS32M foram 8, 24 e 30, respectivamente, correspondem às faixas espectrais $6748 - 6588 \text{ cm}^{-1}$, $4785-4704 \text{ cm}^{-1}$ e $4314 - 4248 \text{ cm}^{-1}$, Figura 9. A faixa espectral de $6748 - 6588 \text{ cm}^{-1}$ representa a região do segundo sobreton para os compostos que apresentam em sua estrutura RNH e CCNH₂, a região de $4785-4704 \text{ cm}^{-1}$ representa a região de combinações dos compostos com função química CONH₂(R)^{2, 54}.

Tabela 3: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de nitrogênio por NIR com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
PLSM	Todo espectro	5	0.956	0.036	0.888	0.055
iPLS2M	1	8	0.953	0.037	0.934	0.046
iPLS4M	4	8	0.951	0.038	0.917	0.044
iPLS8M	2	6	0.947	0.040	0.916	0.052
iPLS16M	11	4	0.927	0.047	0.893	0.062
iPLS32M	13	4	0.918	0.050	0.919	0.050
s2iPLS8M	4, 6	7	0.959	0.035	0.927	0.047
s3iPLS8M	4, 5, 6	7	0.956	0.036	0.926	0.046
s4iPLS8M	2, 4, 6, 8	6	0.959	0.035	0.932	0.043
s2iPLS16M	13, 15	9	0.955	0.036	0.918	0.049
s3iPLS16M	7, 12, 13	6	0.963	0.033	0.932	0.043
s4iPLS16M	3, 9, 10, 16	6	0.966	0.032	0.935	0.043
s2iPLS32M	20, 30	8	0.954	0.037	0.932	0.046
s3iPLS32M	8, 24, 30	7	0.962	0.034	0.944	0.043
s4iPLS32M	2, 12, 27, 30	11	0.972	0.029	0.949	0.040

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC; ^b número de variáveis latentes; Formação em negrito ao melhor modelo de regressão.

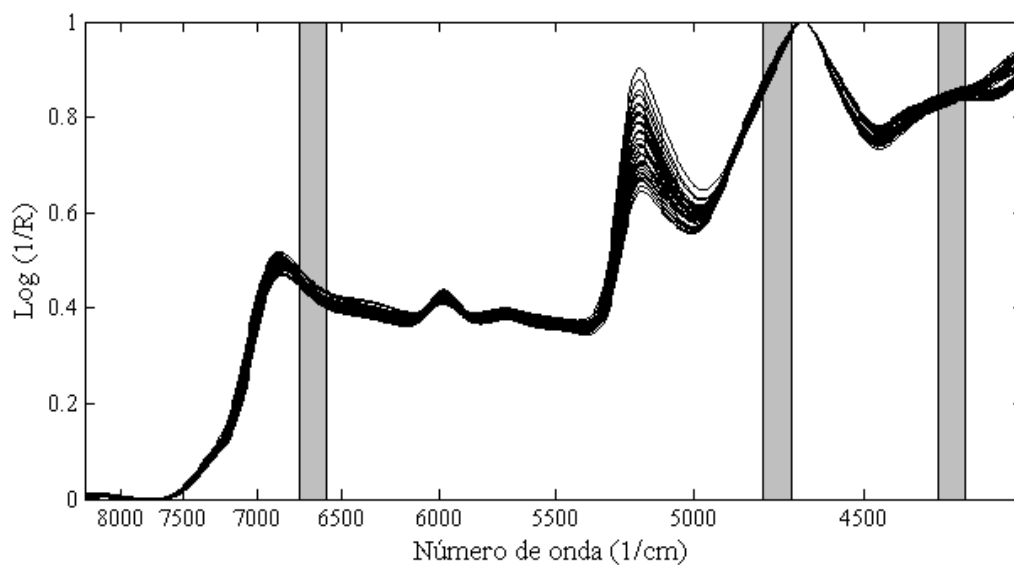


Figura 9: Espectro NIR com destaque as regiões selecionadas (8, 24, 30) combinadas com regressão NIR/siPLS para determinação de nitrogênio em taninos.

A definição do número de variáveis latentes é de fundamental importância para que não ocorra o sobreajuste ou subajuste sendo, pelos valores de RMSECV, apresentados na Figura 10, possível justificar a escolha pelo número de VLs do modelo s3iPLS32M.

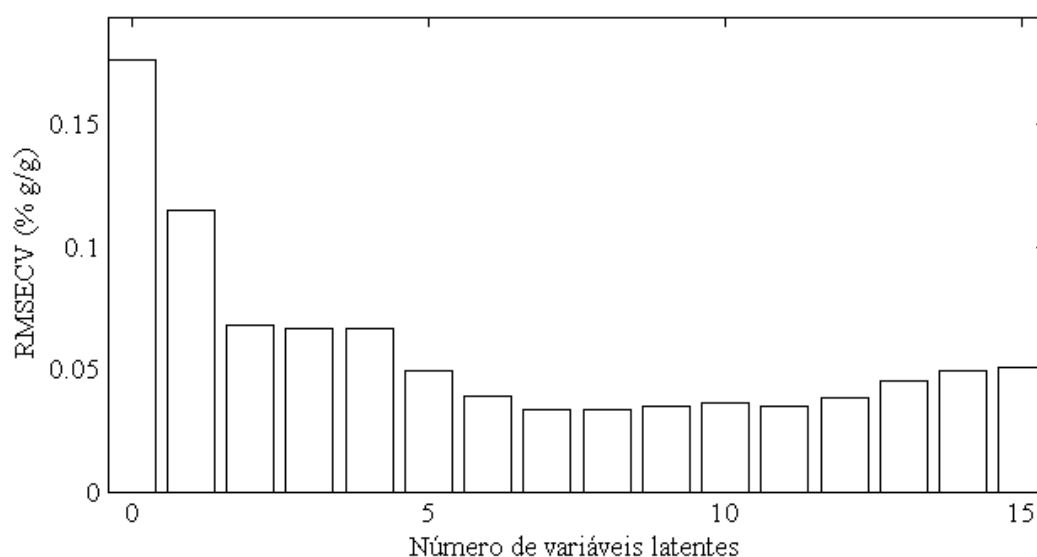


Figura 10: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas pelo NIR/siPLS para nitrogênio combinando os intervalos 8, 24, 30 no modelo s3iPLS32M.

Na Figura 10 é possível verificar que o número ideal de VLs para o modelo s3iPLS32M é 7, pois possui o menor erro validação cruzada (RMSECV), determinado por validação cruzada, sendo que a partir desse número de variáveis não ocorre mais a diminuição significativa do erro.

Na Figura 11, tem-se o gráfico de correlação dos valores de referência, obtido por análise elementar, com os valores obtidos pelo melhor modelo de regressão através do NIR/siPLS. Pode-se avaliar que o modelo possui valores satisfatórios de coeficiente de determinação para as amostras de calibração e previsão, e valores de RMSECV e RMSEP baixos e pouco discrepantes. Portanto, é possível constatar que esse modelo apresenta desempenho aceitável para determinação de nitrogênio para a maior parte das amostras de taninos estudadas.

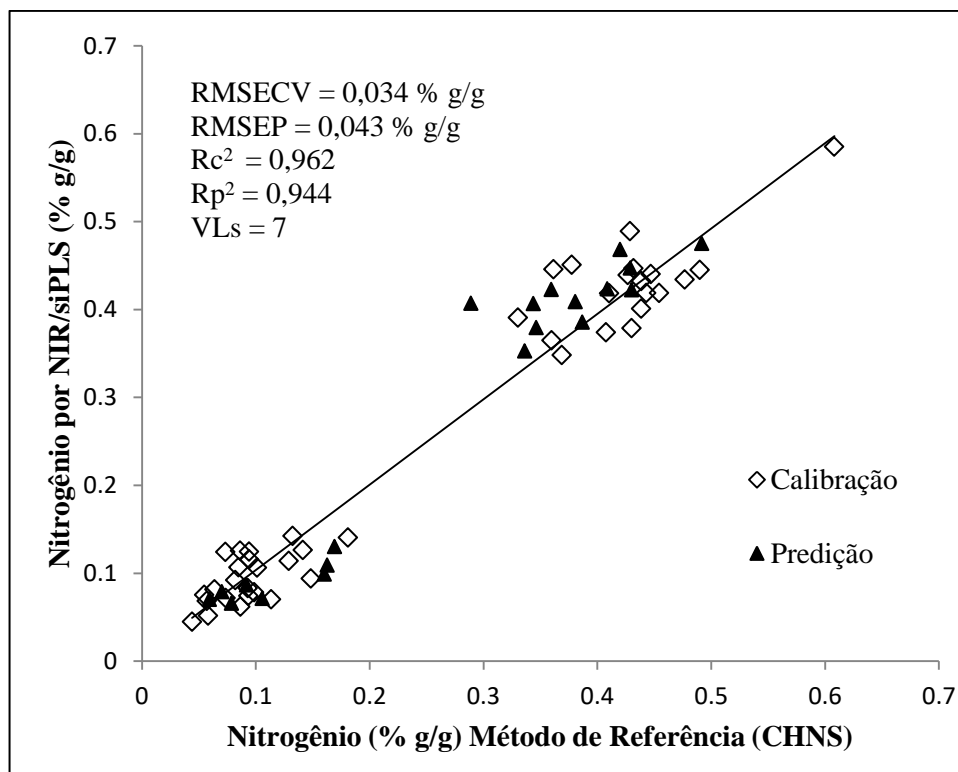


Figura 11: Gráfico dos valores de referência *versus* valores previstos por NIR/siPLS para determinação de nitrogênio em taninos.

5.3.1.2 Considerações quanto aos modelos de regressão multivariada a partir de dados NIR, ATR-FTIR e Fusão para determinação de nitrogênio em taninos.

Comparando-se os modelos obtidos por regressão PLS, verificou-se que a utilização do método de seleção de variáveis por siPLS é mais adequada que por iPLS, apresentando menores erros (RMSECV e RMSEP), maiores coeficientes de determinação (R_c^2 , R_p^2) para quantificar nitrogênio em taninos.

Pode-se também concluir que os melhores modelos foram gerados a partir dos dados NIR. Nos modelos de fusão, a partir dos dois conjuntos, os dados não apresentaram as melhoras esperadas, tendo esses modelos ainda a necessidade de utilizar duas regiões espectrais e na maioria dos casos a utilização de dois equipamentos, o que torna o trabalho mais lento e de maior custo.

As regiões selecionadas no espectro para a construção dos modelos de NIR correspondem a regiões importantes que contém nitrogênio como a das aminas primária, secundárias e amidas primárias. Nos modelos de regressão é desejado que as regiões selecionadas no modelo tenham relação com a composição química das amostras.

5.3.2 Determinação do Teor de enxofre por NIR, ATR-FTIR e Fusão por Regressão Multivariada.

Os modelos de regressão multivariada por PLS foram construídos a partir dos conjuntos de dados NIR, ATR-FTIR e também da fusão dos dados. Também foi avaliada a fusão dos dados com resolução menor dos dados de ATR-FTIR.

Na Tabela 4, é possível verificar os melhores modelos de regressão obtidos para determinação de enxofre em taninos para os quatro conjuntos de dados com a utilização ou não do pré-tratamento MSC. As principais figuras de mérito da totalidade dos modelos gerados são apresentadas no anexo 8.3.

Para a determinação do teor de enxofre a utilização do pré-processamento de MSC apresentou pequenas melhoras para os modelos a partir dos dados NIR e ATR-FTIR. Os modelos a partir da fusão dos dados NIR e ATR-FTIR não apresentaram ganhos com a utilização desse tratamento. Os dados de fusão de NIR e ATR-FTIR com resolução 1:2

apresentaram diminuição dos erros e aumento dos coeficientes de determinação, mas ocorreu o aumento do número de variáveis latentes o que não é desejado.

Tabela 4: Resultados das figuras de mérito dos melhores modelos de determinação do teor de enxofre a com dados NIR, ATR-FTIR, Fusão NIR– ATR-FTIR e Fusão NIR– ATR-FTIR resolução 1:2.

Dados	Tratamento ^a	Melhores Modelos ^b	Intervalos selecionados	VLs ^c	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
NIR	s/MSC	s4iPLS32	3, 5, 17, 18	8	0,957	0,189	0,959	0,193
	c/MSC	s3iPLS16M	3, 4, 9	6	0,966	0,172	0,989	0,105
ATR-FTIR	s/MSC	s2iPLS32	27, 31	8	0,910	0,275	0,906	0,301
	c/MSC	s3iPLS32M	3, 27, 31	9	0,923	0,254	0,914	0,284
Fusão NIR–ATR-FTIR	s/MSC	s4iPLS32	5, 6, 17, 29	9	0,965	0,171	0,957	0,227
	c/MSC	s4iPLS32M	5, 7, 11, 21	10	0,960	0,182	0,949	0,240
Fusão NIR - ATR-FTIR resolução 1:2	s/MSC	s2iPLS16	1, 6	8	0,937	0,228	0,955	0,225
	c/MSC	s4iPLS32M	1, 4, 5, 11	10	0,953	0,198	0,971	0,168

^a Tratamento - modelos com e sem tratamento por MSC; ^b iPLS – número de divisões do espectro, s – número de intervalos combinados, M – tratamento com MSC; ^c número de variáveis latentes; Formação em negrito ao melhor modelo de regressão.

A partir da Tabela 4 é possível verificar os resultados das figuras de mérito dos modelos para cada conjunto de dados avaliado. Pode-se observar que os modelos gerados a partir dos dados NIR apresentaram erros de validação cruzada (RMSECV) e previsão (RMSEP) menores que os modelos a partir dos dados ATR-FTIR, também foi possível observar valores maiores de coeficientes de determinação para calibração e previsão, assim como a utilização de um número menor de variáveis latentes. Os modelos gerados a partir da fusão dos dados, sem diminuição da resolução, com a de menor resolução para ATR- FTIR não apresentaram melhoras comparado aos modelos obtidos por NIR, nos modelos de fusão

também foi possível observar que as regiões preferencialmente selecionadas foram as do infravermelho próximo. Portanto definiu-se que o melhor modelo gerado foi o que utilizou dados NIR com MSC para determinação de enxofre em taninos. O destaque em negrito na Tabela 4 sinaliza o melhor modelo construído.

5.3.2.1 Avaliação dos modelos de regressão por iPLS e siPLS a partir de dados NIR para quantificação de enxofre em taninos.

Os resultados também estão organizados pelo número de divisões equidistantes feitas no espectro e no caso do siPLS de suas combinações de 2, 3 ou 4 intervalos. Os modelos foram avaliados com e sem tratamento por MSC os resultados sem MSC são apresentados na Tabela 23 do anexo 8.3 e os com MSC são apresentados na Tabela 5.

A partir dos dados apresentados na Tabela 5, é possível observar que a seleção de variáveis por siPLS apresentou modelos com menores erros de validação cruzada e previsão, como também maiores valores de coeficiente de determinação tanto para calibração como para previsão, comparados aos modelos gerados por iPLS para determinação de enxofre em taninos com dados de infravermelho próximo.

A partir dos resultados das regressões usando o método de iPLS e siPLS com e sem tratamento de MSC, pode-se definir que o melhor modelo foi o s3iPLS16M utilizando dados NIR, onde ocorreu a divisão em 16 parte equidistantes, combinação de 3 intervalos com tratamento de MSC. Verificou-se que diversos modelos apresentaram resultados considerados satisfatório utilizando NIR com MSC (Tabela 5). Modelos como s4iPLS32M apresentam inclusive valores de coeficientes de determinação (R_c^2) maiores e erros de validação cruzada (RMSECV) menores comparados aos do modelo s3iPLS16M, mas esse modelo não foi selecionado por possuir valores de erros de previsão muito discrepantes dos de validação cruzada, o que é um indicativo de possível sobreajuste, assim como a utilização de 15 VLs o que também é considerado elevado se compararmos ao modelo escolhido, que utiliza 6 VLs.

Tabela 5: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de enxofre por NIR com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
PLSM	Todo espectro	6	0,922	0,254	0,965	0,184
iPLS2M	1	6	0,901	0,287	0,973	0,166
iPLS4M	3	5	0,923	0,252	0,956	0,207
iPLS8M	2	5	0,873	0,325	0,974	0,168
iPLS16M	4	4	0,842	0,363	0,952	0,226
iPLS32M	8	3	0,832	0,373	0,918	0,286
s2iPLS8M	2, 6	6	0,931	0,240	0,947	0,219
s3iPLS8M	1, 2, 6	4	0,936	0,231	0,921	0,268
s4iPLS8M	1, 2, 7, 8	6	0,940	0,224	0,939	0,235
s2iPLS16M	3, 13	5	0,941	0,221	0,923	0,265
s3iPLS16M	3, 4, 9	6	0,966	0,172	0,989	0,105
s4iPLS16M	2, 3, 4, 9	9	0,959	0,184	0,982	0,130
s2iPLS32M	5, 19	10	0,948	0,208	0,971	0,176
s3iPLS32M	4, 5, 18	8	0,961	0,181	0,977	0,151
s4iPLS32M	5, 9, 17, 32	15	0,974	0,147	0,956	0,230

Os intervalos selecionados por siPLS para do modelo s3iPLS16M foram 3, 4 e 9, respectivamente, correspondem às faixas espectrais de 7364 – 6973, 6964 - 6614 e 5482 – 5269 cm⁻¹, conforme ilustrado na Figura 12.

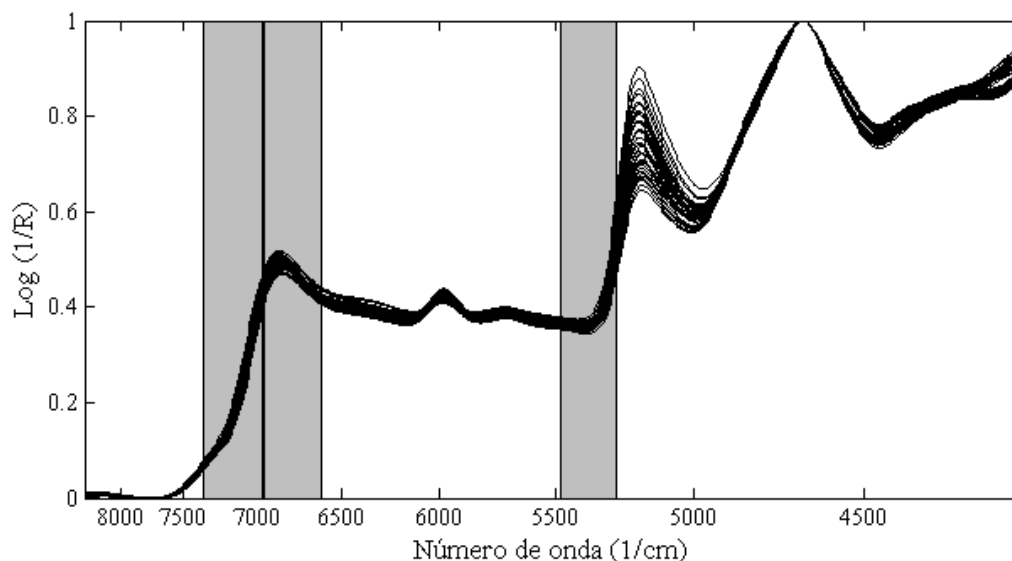


Figura 12: Espectro NIR com destaque as regiões selecionadas (3, 4, 9) combinadas com regressão NIR/siPLS, para determinação de enxofre em taninos.

A seleção das regiões que apresentam a composição química da propriedade estudada é de fundamental importância nos modelos de regressão por PLS. Na propriedade de enxofre investigou-se a relação das regiões selecionadas com possíveis compostos químicos contendo enxofre que estão presentes em amostras de taninos, não foi encontrada uma relação direta das faixas espectrais selecionadas com as regiões descritas na literatura relacionados com os compostos que contém enxofre⁵⁴.

As regiões dos compostos químicos nos espectros podem ter variações nas suas posições e segundo estudos de Xiaobo *et al.* (2010)⁵⁴, regiões onde ocorrem as ligações dos compostos químicos no NIR dos materiais analisados, são afetados por grupos vizinhos, ligações de hidrogênio, cristalinidade, separação de fases, térmica e mecânica. As atribuições das bandas NIR não são realizadas a partir de estudos de moléculas simples e sim a partir de dados empíricos de métodos de NIR. Em outros casos, foram estimadas as posições dos grupos funcionais por bandas conhecidas dos grupos funcionais no espectro de infravermelho. Portanto, as regiões das bandas dos grupos funcionais são apenas aproximações e as limitações das regiões selecionadas servem como referência.

Com base na Figura 13 foi possível definir que o número ideal de VLs para o modelo s3iPLS16M é 6, pois possui um baixo valor de erro de validação cruzada (RMSECV). É possível observar que utilizando 15 VLs temos um erro um pouco menor que utilizando 6 VLs, mas não se utilizou 15 VLs pois isso irá acarretar na inserção de informação não

relevante ao modelo, ocorrendo um sobreajuste e por consequência tendo um maior erro de previsão (RMSEP).

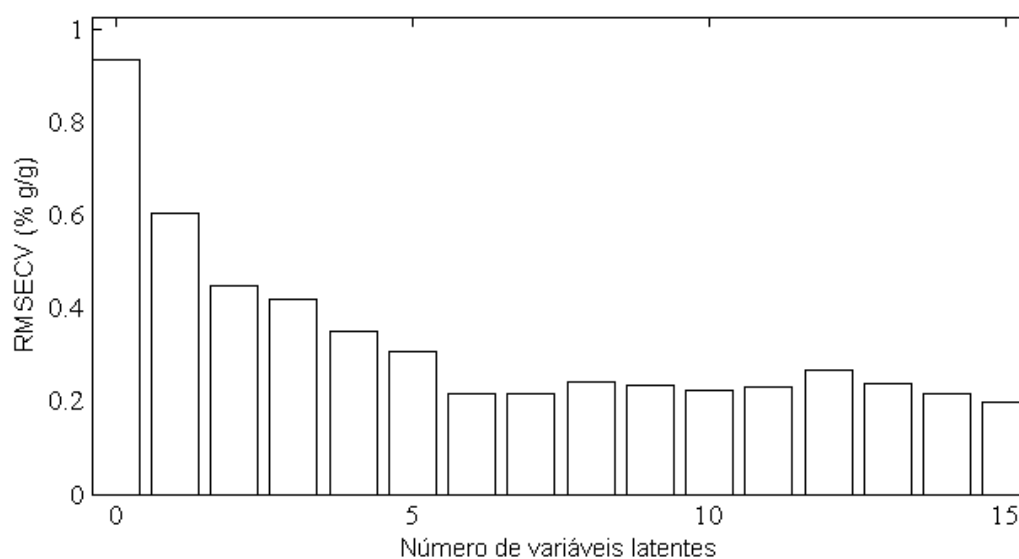


Figura 13: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas pelo NIR/siPLS para enxofre combinando os intervalos 3, 4, 9 no modelo s3iPLS16M.

Também foi construído um gráfico de correlação dos valores de referência, obtidos pelo método de referência para o teor de enxofre, com os valores obtidos pelo melhor modelo de regressão através do NIR/siPLS, conforme ilustrado na Figura 14. Pode-se avaliar que o modelo possui valores considerados satisfatórios de coeficiente de determinação para as amostras de calibração e previsão, e valores de RMSECV e RMSEP baixos e pouco discrepantes. Portanto, é possível constatar que esse modelo apresenta desempenho aceitável para determinação de enxofre em amostras de taninos.

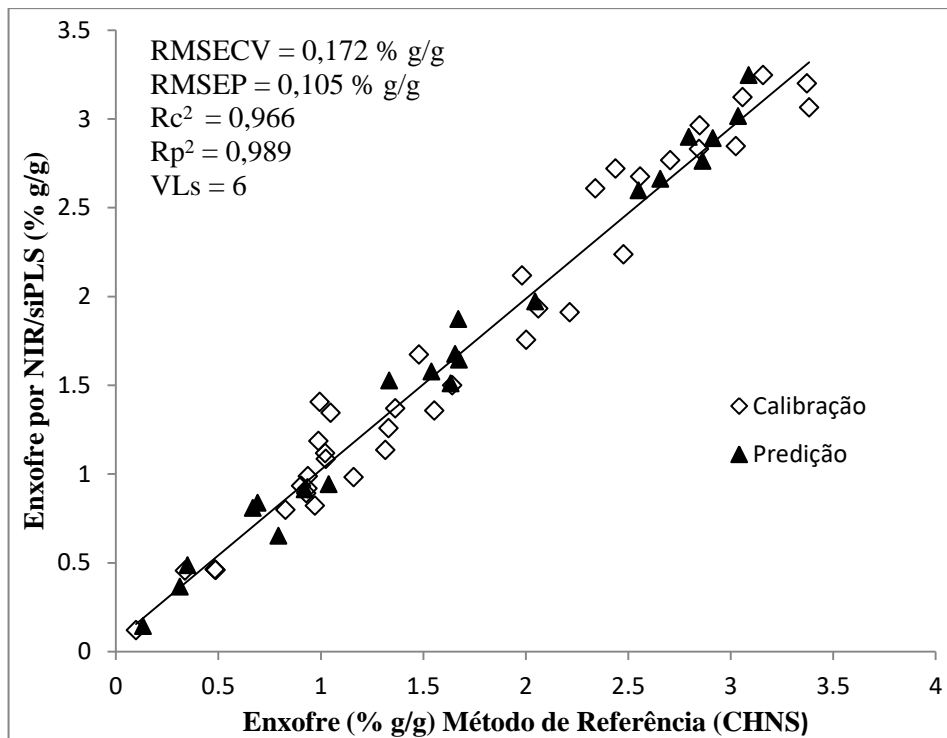


Figura 14: Gráfico dos valores de referência *versus* valores previstos por NIR/siPLS para determinação de enxofre em taninos.

5.3.2.2 Considerações quanto aos modelos de regressão multivariada a partir de dados NIR, ATR-FTIR e Fusão para determinação de enxofre em taninos.

A partir dos modelos de regressão por PLS foi possível concluir que o melhor método de seleção de variáveis, para determinação de enxofre, foi o siPLS. Os modelos gerados a partir desse método de seleção de variáveis apresentaram menores erros de validação cruzada e previsão e maiores valores de coeficientes de determinação quando comparado aos métodos de seleção de variáveis por iPLS.

Quando avaliamos a utilização de diferentes regiões espectrais, assim como a fusão desses, podemos concluir que os melhores modelos foram gerados a partir da utilização de dados da região do infravermelho próximo. Modelos de fusão de intervalos até apresentaram resultados próximos, mas com a utilização quase que exclusivamente de intervalos do NIR, reforçando os resultados obtidos na avaliação individual de cada intervalo.

5.3.3 Determinação do Teor de Cinzas por NIR, ATR-FTIR e Fusão por Regressão Multivariada.

Os modelos de regressão multivariada por PLS foram construídos a partir dos conjuntos de dados NIR, ATR-FTIR e também da fusão dos dados. Também foi avaliada a fusão dos dados com resolução menor dos dados de infravermelho médio.

Na Tabela 6, são apresentados os resultados dos melhores modelos de regressão por PLS obtidos para a determinação de cinzas em taninos, para os quatro conjuntos de dados com a utilização ou não utilização do pré-tratamento MSC. A totalidade dos modelos gerados são apresentados no anexo 8.4.

Tabela 6: Resultados das figuras de mérito dos melhores modelos de determinação do teor de cinzas com dados NIR, ATR-FTIR, Fusão NIR– ATR-FTIR e Fusão NIR– ATR-FTIR resolução 1:2.

Dados	Tratamento ^a	Melhores Modelos ^b	Intervalos selecionados	VLs ^c	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
NIR	s/MSC	s4iPLS16	3, 7, 9, 12	6	0,930	0,49	0,920	0,54
	c/MSC	s3iPLS8M	3, 5, 6	6	0,912	0,54	0,945	0,44
ATR-FTIR	s/MSC	s4iPLS16	2, 4, 7, 14	9	0,828	0,76	0,869	0,69
	c/MSC	s4iPLS8M	1, 3, 4, 8	9	0,906	0,79	0,877	0,60
Fusão NIR–ATR-FTIR	s/MSC	s3iPLS32	4, 5, 7	6	0,939	0,46	0,926	0,48
	c/MSC	s4iPLS32M	4, 5, 7, 19	8	0,940	0,46	0,929	0,47
Fusão NIR–ATR-FTIR resolução 1:2	s/MSC	s2iPLS16	3, 16	9	0,937	0,48	0,938	0,46
	c/MSC	s4iPLS16M	2, 4, 9, 16	9	0,940	0,46	0,941	0,48

^a Tratamento - modelos com e sem tratamento por MSC; ^b iPLS – número de divisões do espectro, s – número de intervalos combinados, M – tratamento com MSC; ^c número de variáveis latentes; Formação em negrito ao melhor modelo de regressão.

Os dados NIR e ATR-FTIR foram avaliados separadamente assim como a fusão deles. A fusão dos dados de infravermelho próximo e médio foi realizada de duas maneiras; a primeira utilizando os dados sem diminuição da resolução e a segunda com a diminuição da resolução dos dados de ATR-FTIR pela metade, pois os dados de infravermelho médio possuem número superior de variáveis.

A utilização do pré-tratamento de MSC também apresentou significativa melhora nos modelos para os dados ATR-FTIR. Na utilização dos dados NIR ocorreu uma pequena melhora nos modelos com o pré-tratamento. Já os modelos a partir da fusão de NIR e ATR-FTIR não apresentaram melhora e, para os modelos de fusão, com resolução 1:2, a melhora foi pouco representativa.

A partir da Tabela 6 é possível verificar os resultados das figuras de mérito dos modelos para cada conjunto de dados. Os melhores modelos foram gerados novamente a partir dos dados NIR, assim como as propriedades anteriormente avaliadas. Podemos verificar que os modelos gerados a partir dos dados NIR apresentaram baixos erros validação cruzada e os menores erros de previsão (RMSECV e RMSEP), como também maiores valores de coeficientes de determinação de calibração e previsão (R_c^2 e R_p^2) comparado aos modelos com dados de ATR-FTIR. Os modelos gerados a partir da fusão dos dados NIR e ATR-FTIR apresentaram semelhanças aos modelos utilizando apenas NIR, possuindo erros e coeficientes de determinação bem próximos. Nesses modelos as regiões selecionadas foram, na maior parte, as de infravermelho próximo, por esse motivo possuindo modelos semelhante aos que só utilizavam dados NIR. Portanto, definiu-se que o melhor modelo gerado foi o que utilizou dados NIR com a utilização do MSC, para determinação de cinzas em taninos. O destaque em **negrito** na tabela 6 sinaliza o melhor modelo construído.

5.3.3.1 Avaliação dos modelos de regressão por iPLS e siPLS a partir de dados NIR para quantificação de cinzas em taninos.

Os modelos para a determinação do teor de cinzas foram avaliados com e sem tratamento por MSC, os resultados sem MSC são apresentados na Tabela 30 do anexo 8.4 e os com MSC são apresentados na Tabela 7. A partir dos resultados dos modelos podemos concluir que a seleção de variáveis por siPLS foi mais eficiente que a por iPLS, apresentando

modelos com menores erros de validação cruzada e previsão, maiores coeficientes de determinação, conforme Tabela 7. Esses modelos mais satisfatórios para siPLS são resultados da combinação de regiões importantes para a construção dos modelos, o que não ocorre para o método de iPLS, que apenas seleciona uma região.

Tabela 7: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de cinzas por NIR com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
PLSM	Todo espectro	8	0,908	0,56	0,931	0,49
iPLS2M	1	8	0,905	0,57	0,958	0,41
iPLS4M	2	8	0,889	0,61	0,943	0,49
iPLS8M	2	5	0,834	0,75	0,918	0,54
iPLS16M	13	4	0,803	0,82	0,875	0,72
iPLS32M	23	3	0,752	0,91	0,806	0,83
s2iPLS8M	2, 6	6	0,898	0,58	0,891	0,61
s3iPLS8M	3, 5, 6	6	0,912	0,54	0,945	0,44
s4iPLS8M	1, 2, 6, 7	5	0,922	0,51	0,919	0,54
s2iPLS16M	2, 9	10	0,927	0,50	0,914	0,53
s3iPLS16M	9, 10, 12	4	0,912	0,54	0,899	0,59
s4iPLS16M	2, 8, 10, 12	7	0,927	0,50	0,930	0,52
s2iPLS32M	18, 24	7	0,912	0,54	0,929	0,50
s3iPLS32M	18, 20, 23	5	0,920	0,52	0,896	0,60
s4iPLS32M	5, 17, 18, 24	6	0,914	0,54	0,934	0,49

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC; ^b número de variáveis latentes; Formação em negrito ao melhor modelo de regressão.

A partir dos resultados das regressões usando o método de iPLS e siPLS com e sem tratamento de MSC definiu-se que o melhor modelo foi o s3iPLS8M utilizando dados NIR, onde ocorreu a divisão em 8 parte equidistantes, combinação de 3 intervalos com tratamento de MSC. Foi possível verificar que diversos modelos apresentaram resultados muito semelhantes. Entre estes os modelos gerados utilizando NIR com MSC (Tabela 7), o modelo de s4iPLS16M é um exemplo de modelo com resultado das figuras de mérito bem próximo ao modelo escolhido. Esse modelo possui uma pequena melhora para a calibração do modelo no erro (RMSECV) e no coeficiente de determinação de calibração (R_c^2) comparado ao modelo escolhido s3iPLS8M, mas o modelo s4iPLS16M possui uma pequena piora nos modelos de previsão, no erro (RMSEP) e no coeficiente de determinação (R_p^2). Avaliando as variáveis latentes podemos verificar que o modelo s3iPLS8M possui 6 e o modelo de s4iPLS16M possui 7. Quando avaliamos as regiões selecionadas pelo método de siPLS verificamos que as regiões escolhidas coincidem nos dois modelos. Portanto os dois modelos apresentam resultados considerados satisfatórios e a escolha por modelo s3iPLS8M foi realizado pelos melhores resultados dos modelos de previsão e menor número de variáveis latentes.

Os intervalos selecionados do modelo s3iPLS8M foram 3, 5, 6 respectivamente, correspondem às faixas espectrais 6623 - 6017, 5501 - 5076 e 5071 - 4708 cm^{-1} , conforme ilustrado na Figura 15.

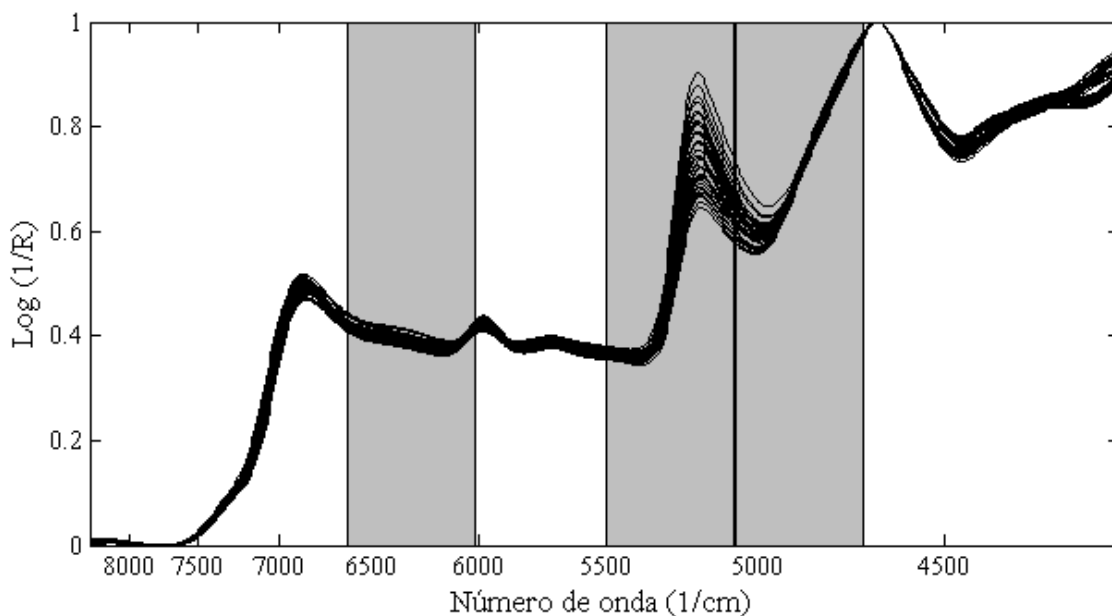


Figura 15: Espectro NIR com destaque as regiões selecionadas (3, 5, 6) combinadas com regressão NIR/siPLS para determinação de cinzas em taninos.

Na Figura 16 é possível verificar, a partir dos erros de validação cruzada, a escolha pelo número de VLs do modelo s4iPLS32M.

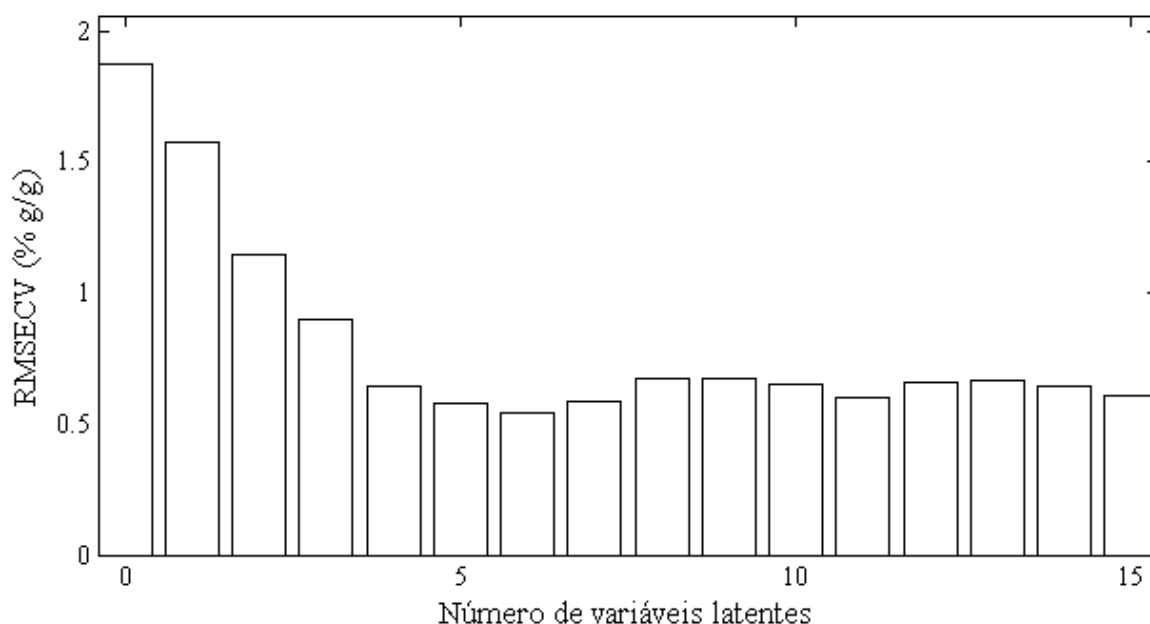


Figura 16: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas pelo NIR/siPLS para cinzas combinando os intervalos 3, 5, e 6 no modelo s3iPLS8M.

A partir da Figura 16 é possível observar que o menor erro de validação cruzada é obtido a partir da utilização de 6 VLs, e a utilização de um número maior irá acarretar na inserção de informação não relevantes para os modelos, aumentando os erros.

Na Figura 17, é possível observar a correlação dos valores de referência para o teor de cinzas, obtidos pelo método de referência, com os valores obtidos pelo melhor modelo de regressão através do NIR/siPLS. Pode-se avaliar que o modelo possui valores considerados satisfatórios de coeficiente de determinação para as amostras de calibração e previsão, considerando que a determinação de cinzas por esse método multivariado ocorre de forma indireta, além de valores de RMSECV e RMSEP baixos e pouco discrepantes. Portanto, é possível constatar que esse modelo apresenta desempenho satisfatório para determinação de cinzas em amostras de taninos.

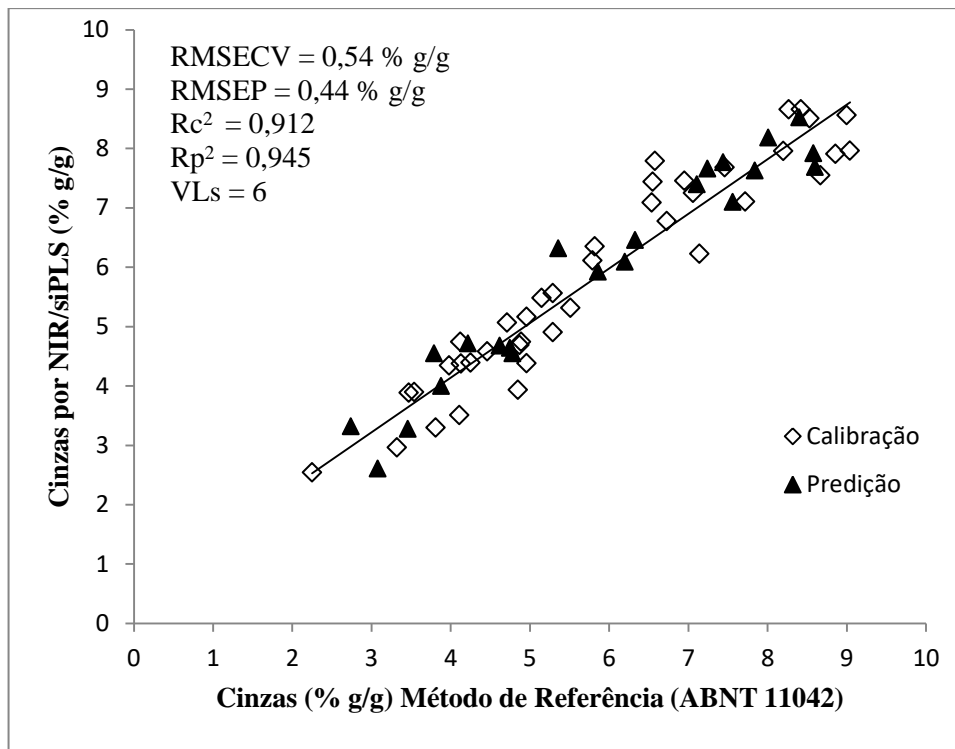


Figura 17: Gráfico dos valores de referência *versus* valores previstos por NIR/siPLS para determinação de cinzas em taninos.

5.3.3.2 Considerações quanto aos modelos de regressão multivariada a partir de dados NIR, ATR-FTIR e Fusão para determinação de cinzas em taninos.

A partir dos modelos de regressão por PLS foi possível concluir que assim como nas propriedades anteriores, o melhor método de seleção de variáveis foi o siPLS para determinação de cinzas. Os modelos gerados a partir desse método de seleção de variáveis apresentaram menores erros de validação cruzada e previsão e maiores valores de coeficientes de determinação quando comparado aos métodos de seleção de variáveis por iPLS.

Quando avaliamos os quatro conjuntos de dados utilizados para geração dos modelos de regressão por PLS concluiu-se que o melhor é o que utiliza somente dados NIR. Os modelos gerados por ATR-FTIR apresentaram piores resultados em todas as figuras de mérito comparado aos de NIR. Os modelos a partir da fusão de dados originais e com diminuição da resolução apresentaram resultados semelhantes aos de NIR, esses modelos de fusão

apresentam a problemática de utilizar duas regiões espectrais, o que acarreta em maior tempo de análise e aumento de custos.

5.3.4 Determinação do Teor de Carbono por NIR, ATR-FTIR por Regressão Multivariada.

Os modelos de regressão multivariada foram construídos a partir dos dados NIR e ATR-FTIR para a determinação de carbono em amostras de taninos condensados.

Na Tabela 8 são apresentados os melhores modelos obtidos a partir dos dados NIR e ATR-FTIR com e sem utilização do pré-tratamento de MSC. A totalidade dos resultados são apresentados nos anexos 8.5.

Tabela 8: Resultados das figuras de mérito dos melhores modelos de determinação do teor de carbono com dados NIR e ATR-FTIR.

Dados	Tratamento ^a	Melhores Modelos ^b	Intervalos selecionados	VLs ^c	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
NIR	s/MSC	s3iPLS8	2, 4, 6	5	0,770	0,495	0,651	0,481
	c/MSC	s3iPLS16M	6, 9, 12	5	0,794	0,467	0,604	0,526
ATR-FTIR	s/MSC	s4iPLS16	3, 5, 13, 14	7	0,794	0,462	0,711	0,404
	c/MSC	s3iPLS8M	2, 3, 8	8	0,700	0,563	0,738	0,404

^a Tratamento - modelos com e sem tratamento por MSC; ^b iPLS – número de divisões do espectro, s – número de intervalos combinados, M – tratamento com MSC; ^c número de variáveis latentes;

Nos modelos para a determinação do teor de carbono, a utilização do pré-tratamento MSC não resultou em melhoras nos modelos, tanto para dados NIR com para ATR-FTIR.

Avaliando os resultados dos modelos gerados a partir dos dois conjuntos de dados é possível verificar baixos valores de coeficientes de determinação para calibração e previsão.

5.3.4.1 Considerações quanto aos modelos de regressão multivariada a partir de dados NIR, ATR-FTIR para determinação de carbono em taninos.

Os modelos de regressão por PLS gerado a partir dos dados NIR e ATR-FTIR para determinação de carbono em taninos não apresentaram resultados satisfatórios, pois não foi possível gerar modelos que apresentassem uma correlação adequada entre valores previstos pelos modelos e os de referência. Nesses modelos foi possível observar baixos valores de coeficientes de determinação.

5.3.5 Determinação do Teor de Hidrogênio por NIR, ATR-FTIR por Regressão Multivariada.

Os modelos de regressão multivariada para a determinação de nitrogênio foram construídos a partir dos dados NIR e ATR-FTIR para as amostras de taninos condensados.

Na Tabela 9 são apresentados os melhores modelos obtidos a partir dos dados NIR e ATR-FTIR e utilização do pré-tratamento MSC para dados de infravermelho médio. A totalidade dos resultados são apresentados nos anexos 8.6.

Tabela 9: Resultados das figuras de mérito dos melhores modelos de determinação do teor de hidrogênio com dados NIR e ATR-FTIR.

Dados	Tratamento ^a	Melhores Modelos ^b	Intervalos selecionados	VLs ^c	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
NIR	s/MSC	s3iPLS8	5, 6, 7	3	0,580	0,108	0,618	0,095
ATR-FTIR	s/MSC	s2iPLS8	6, 8	5	0,451	0,093	0,383	0,082
	c/MSC	s2iPLS8M	6, 8	3	0,499	0,089	0,437	0,083

^a Tratamento - modelos com e sem tratamento por MSC; ^b iPLS – número de divisões do espectro, s – número de intervalos combinados, M – tratamento com MSC; ^c número de variáveis latentes;

A utilização do pré-tratamento MSC também não apresentou melhoras nos modelos gerados a partir dos dados de ATR-FTIR.

A partir da Tabela 9 foi possível constatar que os modelos de regressão utilizando dados NIR como ATR-FTIR apresentaram valores de coeficientes de determinação para calibração e previsão muito baixos.

5.3.5.1 Considerações quanto aos modelos de regressão multivariada a partir de dados NIR, ATR-FTIR para determinação de hidrogênio em taninos.

Os modelos de regressão por PLS gerados a partir dos dados NIR e ATR-FTIR para a determinação de hidrogênio em taninos apresentaram resultados insatisfatórios, pois assim como nos modelos para determinação de carbono, não apresentaram uma correlação adequada entre aos valores previstos pelo modelo com os de referência, gerando coeficientes de determinação para calibração e previsão extremamente baixos.

5.4 CONSIDERAÇÕES FINAIS

Na tabela 10 são apresentados os resultados das figuras de méritos para os melhores modelos selecionados para a quantificação dos parâmetros de nitrogênio, enxofre, cinzas, bem como os modelos que não foram eficientes, contemplando a quantificação do carbono e hidrogênio.

Cabe destacar que para o nitrogênio, enxofre e cinzas os modelos empregando espectroscopia no infravermelho apresentam diversas vantagens, como maior rapidez na análise, não destruição das amostras e a não utilização de solventes. Visto que o método de determinação de nitrogênio e enxofre por CHNS possui etapas de preparo de amostras e oxidação das amostras a altas temperaturas. Na quantificação de cinzas por gravimetria diversas etapas de secagem e calcinação são necessárias, sendo um processo que demanda muito tempo. Portanto essa metodologia proposta não gera efluentes, é de menor custo e muito rápida comparada aos métodos usuais.

Tabela 10 Resultados das figuras de mérito dos melhores modelos para os parâmetros de nitrogênio, enxofre, cinzas, carbono e hidrogênio.

Parâmetros	Dados	Tratamento ^a	Melhores Modelos ^b	Intervalos selecionados	VLs ^c	Rc ²	RMSECV (% g/g)	Rp ²	RMSEP (% g/g)
N	NIR	c/MSC	s3iPLS32M	8, 24, 30	7	0,962	0,034	0,944	0,043
S	NIR	c/MSC	s3iPLS16M	3, 4, 9	6	0,966	0,172	0,989	0,105
Cinzas	NIR	c/MSC	s3iPLS8M	3, 5, 6	6	0,912	0,54	0,945	0,44
C	ATR-FTIR	s/MSC	s4iPLS16	3, 5, 13, 14	7	0,794	0,462	0,711	0,404
H	NIR	s/MSC	s3iPLS8	5, 6, 7	3	0,580	0,108	0,618	0,095

^a Tratamento - modelos com e sem tratamento por MSC; ^b iPLS – número de divisões do espectro, s – número de intervalos combinados, M – tratamento com MSC; ^c número de variáveis latentes;

6 CONCLUSÕES

Através dos resultados obtidos na determinação das propriedades foi possível constatar a produção de modelos mais satisfatórios com a utilização do método de siPLS, que além de selecionar intervalos, como no método de iPLS, também faz a combinação dessas regiões fornecendo modelos com melhores resultados nas figuras de mérito.

Foi realizado um estudo sobre a utilização ou não do pré-tratamento de MSC e, constatou-se, que o emprego de MSC na grande maioria dos modelos apresentou significativa melhora nos resultados. A sua utilização torna o tempo de processamento maior, mas para amostras sólidas como no caso dos taninos condensados, a sua utilização foi considerada relevante.

Com o estudo realizado a partir da utilização dos conjuntos de dados de infravermelho próximo, médio e fusão foi possível observar que os melhores modelos foram gerados a partir da utilização dos dados de NIR para as determinações dos teores de nitrogênio, enxofre e cinzas. Modelos de fusão dos conjuntos de dados não apresentaram melhoras significativas comparados aos que somente utilizavam dados de NIR.

Os modelos a partir dos dados de NIR apresentaram melhores resultados na maioria das propriedades, pois uma mesma região no espectro de NIR pode abranger vários compostos químicos simultaneamente, ou seja, estes compostos ficam sobrepostos nessas regiões, sendo que para amostras complexas, como no caso dos taninos, pode ser considerado um fator positivo.

Para os teores de nitrogênio, enxofre e cinzas os resultados dos modelos gerados podem ser considerados satisfatórios, pois os valores para os coeficientes de determinação para calibração e previsão foram superiores a 0,9116 e 0,9451 %, respectivamente, bem como apresentaram baixos erros de validação cruzada e previsão.

Na avaliação das propriedades de carbono e hidrogênio a partir dos resultados dos coeficientes de determinação, através da análise multivariada com dados de infravermelho, foi possível constatar que os resultados não foram satisfatórios, portanto, não sendo uma ferramenta adequada para determinação dessas propriedades em amostras de taninos.

Portanto, para os teores de nitrogênio, enxofre e cinzas a técnica de análise multivariada por regressão PLS com dados de infravermelho apresentaram resultados satisfatórios. Essa técnica apresenta diversas vantagens sobre os métodos tradicionais de determinação desses parâmetros como: redução no tempo de análise, não destruição das amostras, possibilidade de análise em tempo real e simultânea dos parâmetros e a possibilidade de ser facilmente implementada no processo produtivo.

7 REFERÊNCIAS BIBLIOGRÁFICAS

1. Mohanty, A. K.; Miska, M.; Drzal, L. *Journal Polymers and the Environment* **2002** v.10, n. 112, p. 19-26.
2. Averous, L.; Digabel, F. *Carbohydr. polym.* **2006**. v. 66, p. 480-493.
3. Pansera, M. R.; Santos, A. C. A.; Pease, K.; Wasung, R.; Rossato, M.; Rota, L. D.; Pauletti, G. F.; Serafini, L. A. *Ver. Bras. Farmacogn.* **2003**. V. 13. p. 17-22.
4. Nozella, E. Determinação de taninos em plantas com potencial forrageiro para ruminantes. 2001. 72f. Dissertação (Mestrado) – ESALQ, USP, Piracicaba, São Paulo, 2001.
5. S. Quideau, D. Deffieux, C. *Angew. Chem. Int. Ed.*, **2011**, 50, 586.
6. A. Ricci, K. J.; Olejar, G. P.; Parpinello, P. A.; Kilmartin and A. Versari, *Appl. Spectrosc. Rev.* **2015**, 50, 407.
7. Venter, P. B.; Sisa, M.; Van Der Merwe, M. J.; Bonnet, S. L.; & Van Der Westhuizen, J. H. *Phytochemistry* **2012** 73, 95-105.
8. Tambi, L.; Frediani, P.; Frediani, M.; Rosi, L.; & Camaiti, M. *J. Appl. Polym. Sci.* **2008**. 108(3), 1797-1809.
9. Grein, A.; Da Silva B. C.; *et al.* *Carbohydr. polym.* **2013**, 92, 312-320.
10. ABNT NBR 11042:**2013** Insumos — Corantes, ligantes, pastas pigmentadas, substâncias graxas e tanantes — Determinação do teor de cinza. 2013
11. Lamprecht, H. *Silvicultura nos trópicos: ecossistemas florestais e respectivas espécies arbóreas-possibilidades e métodos de aproveitamento sustentado*. Eschoborn: Instituto de Silvicultura da Universidade de Göttingen. 1990. 343p.
12. Simões, C.M.O.; Schenkel, E.P.; Gosmann, G.; Mello, J.C.P.de; Mentz, L.A.; Petrovick, P.R. *Ver. Bras. Farmacogn.* 3.ed. Porto Alegre/Florianópolis, UFRGS/ EDUFSC, 2001
13. Bruneton, J. *Elementos de Fitoquímica y de Farmacognosia*, Ed. Acribia, SA: Espanha, 1991.
14. Haslam, E. *J. Chem. Ecol.* **1988**, 14, 1789.
15. Nozella, E. Determinação de taninos em plantas com potencial forrageiro para ruminantes. 2001. 72f. Dissertação (Mestrado) – ESALQ, USP, Piracicaba, São Paulo 2001.
16. Falcão, L.; & Araújo, M. E. M. *Vibrational Spectroscopy* **2014**, 74, 98-103.
17. Arbenz, A.; & Avérous, L. *Green Chem.* **2015**, 17(5), 2626-2646.
18. Dos Santos Grasel, F.; Ferrão, M. F.; & Wolf, C. R. *Industrial Crops and Products* **2016**, 91, 279-285.

19. Santos, S. Da C.; Mello, J. C. P. De. Taninos. in: Simões, C. M. O.; Schenkel, E. P.; Gosmann, G.; Mello, J. C. P. De.; Mentz, L. A.; Petrovick, P. R. *Ver. Bras. Farmacogn*, Porto Alegre: UFRGS, Florianópolis, UFSC, 2003, Cap. 24, p. 615-656.
20. Long, R. Quebracho based polyphenols for use in wood panel adhesive system. *Holz als Roh-und Werkstoff*. V. 49, p. 485-487, 1991.
21. Lelis, C. C. R.; Gonçalves, A. de C. Teores de taninos da casca da madeira de cinco leguminosas arbóreas. *Floresta e Ambiente*. v. 8, n. 1, p.167-173, jan/dez. 2001.
22. Verza, S. G. Avaliação das variáveis analíticas dos métodos de determinação do teor de taninos totais baseados na formação de complexos com substâncias protéicas e derivados da polivinilpirrolidona. 2006. 41f. Dissertação (Mestrado) – UFRGS, Porto Alegre, Rio Grande do Sul, 2006.
23. Shoji, T.; Masumoto, S.; Moriichi, N.; Kanda, T.; Ohtake, Y. *J. chromatogr. A* **2006**, v. 1102, p. 206-213.
24. Salunkhe, D. K.; Chavan, J.K.; Kadan, S.S. Dietary tannins: consequences and remedies. Boca Raton: CRC Press, 1990. 200p.
25. Lekha, P. K.; Lonsane, B. K. *Adv. App. Microbiol* **1997** V. 44.
30. Griffin, A. R., Nambiar, E. S., Harwood, C. E., & See, L. S. *Southern Forests: a Journal of Forest Science* **2015**, 77(1), v-viii.
27. J. M. Chan, P. Day, J. Feely, R. Thompson, K. M. Little and C. H. Norris, *J. For. Sci.* **2015**, 77, 19.
28. Boland, D. J.; Brooker, M. I. H.; Chippendale, G. M.; Hall, N.; Hyland, B. P. M.; Johnston, R. D.; Kleining, D. A.; Turner, J. D. *Forest trees of Australia*. 4th edition. Thomas Nelson Australia and CSIRO; Melbourne. 1984, p 687.
29. Burkart, A. *Las leguminosas argentinas silvestres y cultivadas*. Buenos Aires: ACME Agency, 1952, p 569.
30. Binkley, D.; Giardina, C. Nitrogen fixation in tropical forest plantation. In: Nambiar, E. K. S.; Brown, A.G. *Management of soil, nutrientes and water in plantation forests*. Canberra: ACIAR, monograph 43, Australia, 1997, p 297-337.
31. SBS. *Fatos e Números do Brasil Florestal*. Sociedade Brasileira de Silvicultura. 2008. Disponível em www.sbs.org.br/FatoseNumerosdoBrasilFlorestal.pdf. Acesso em 29 de março de 2017.
32. Oliveira, H. A. *Acácia-negra e taninos no Rio Grande do Sul*. Canoas: La Salle. 1968. V. 2. p.121.

33. Mora, A. L. Aumento da Produção de Sementes geneticamente melhoradas de *Acácia mearnsii* De Wild (Acácia Negra) no Rio Grande do Sul. Curitiba, 2002. 140f. Tese (Doutorado em Ciências Florestais) – Setor de Ciências Agrárias, Universidade Federal do Paraná, Curitiba. 2002.
34. Ciclo da industrialização da acácia negra. Montenegro: TANAC, [198-];
35. Saraiva, A. M. Avaliação da atividade biológica de extratos metanólicos de *Schinopsis Brasiliensis* Engl. em sistemas microemulsionados 2012. 28 f. Tese (Doutorado) – UFPE, Recife, Pernambuco.2012.
36. Barbosa, L. C. Espectroscopia no Infravermelho na Caracterização de Compostos Orgânicos, 2007.
37. Holler, F. J.; Skoog, D. A.; Crouch, S. R. Princípios de Análise Instrumental. 6. Ed. Porto Alegre: Bookman, 2009
38. Luypaert, J.; Massart, D.L.; Heyden, Y. V., *Talanta* 2007, 72(3), 865-883.
39. Stuart, B.; George, B.; McIntyre, P., *Modern Infrared Spectroscopy*. West Sussex, England, 1996.
40. Ferrão, M. F. Aplicação de técnicas espectroscópica de reflexão no infravermelho no controle de qualidade de farinha de trigo. 2000. 219 f. Dissertação (Doutorado) - Universidade Estadual de Campinas, Campinas, São Paulo.2000.
41. Ferrão, M. F.; Viera, M. D. S.; Pazos, R. E. P.; Fachini, D.; Gerbase, A. E.; Marder, L. *Fuel* **2011**, 90, 701-706.
42. Liske, M. A. Medidas de espectrometria no infravermelho médio para a determinação do teor de biodiesel em óleo diesel. 2010. 100 f. Dissertação (mestrado). Universidade Federal do Paraná, Curitiba, 2010.
43. Brereton, R. G. *Anal. Meth.* **2013**, 5, 3785-3789.
44. Brereton, R. G. *Analyst.* **2000**, 125, 2125:2154
45. Gomes, A. A.; Algoritmos das Projeções Sucessivas aplicado à seleção de variáveis em regressão PLS. 2012. 121 f. Dissertação (mestrado). Universidade Federal da Paraíba. João Pessoa. 2012
46. Burns, D.A.; Ciurczak, E. W. *Handbook of near-infrared analysis*. 3 Ed. Boca Raton: CRC, 2008. p. 808.
47. Huck, C. W. *J. pharm. biomed. anal.* **2014**, 87, 26-35.
48. Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta* **1986**. 185, 1, 1-17.
49. Geladi, P.; Macdougall, D.; Martens, H. *Appl. Spectrosc* **1985**, 39, 491-500.
50. Jiang, J.; Berry, R.J.; Siester, H.W.; Ozaki, Y. *Anal. Chem.* **2002**. 74, 14, 3555-3565.

51. Oliveira, F.C.C.; de Souza, A.; Dias, J.; Dias, S. C. L.; Rubim, J. C. *Quím. Nova* **2004**, *27*, 218-225.
52. Leardi, R.; Norgaard, L. *J. chemom.* **2004**, *18*, 486-497.
53. Norgaard, L.; Saudland, A.; Wagner, J.; Nielse, J. P.; Munck, L.; Engelsen, S. B. *Appl. Spectrosc.* **2000**, *54*, 413-419.
54. Xiaobo, Z.; Jiewen, Z.; Povey, M. J. W.; Holmes, M.; Hanpin, M. *Anal. Chim. Acta* **2010**, *667*, 14-32.
55. Kennard, R.W.; Stone L.A. *Technometrics* **1969**, *11*, 137-148.
56. Daszykowski, M.; Walczak, B.; Massart, D. L. *Anal. Chim. Acta* **2002**, *468*, 91-103.
57. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. *Quím. Nova* **1999**, *22*, 724-731.
58. Ferreira, M. M. C. *Quimiometria: conceitos, métodos e aplicações*. Editora: Unicamp Campinas/SP, 2015, p 496.
59. Savitzky, A., Golay, M. J. E., *Anal. Chem.* **1964**, *8*: 1627.
60. Isaksson, T.; Naes, T. *Appl. Spectrosc* **1988**. *42*, 1278-1284.
61. Morgano, M. A. *Aplicação de Métodos Quimiométricos em Análise de Alimentos*. 2005. Tese (Doutorado em Ciências), Universidade Estadual de Campinas – Unicamp. Campinas, São Paulo, 2005.
62. Brereton, R. G. *Chemometrics: data analysis for the laboratory and chemical plant*; John Wiley & Sons Ltd: Chichester, UK, 2003.
63. Gomes-Carracedo, M. P.; Andrade, J. M.; Rutledge, D. N.; Faber, N. M. *Anal. Chim. Acta* **2007**, *585*, 253-265.
64. Gemperline, P. J. *Practical guide to chemometrics*. 2ªEd., Nova York: CRC Press Taylor & Francis. 2006, p. 140-145.

8 ANEXOS

8.1 RESULTADOS DA DETERMINAÇÃO DO TEOR DE NITROGÊNIO, ENXOFRE, CARBONO, HIDROGÊNIO E CINZAS PELOS MÉTODOS DE REFERÊNCIA.

Tabela 111: Resultados da determinação de Cinzas em amostras de taninos pelo método de referência ABNT NBR 11042.

Código das Amostras	Teor de Cinzas (% g/g)	Desvio Padrão	Código das Amostras	Teor de Cinzas (% g/g)	Desvio Padrão
1	7,72	0,13	33	2,25	0,01
2	4,25	0,10	34	4,87	0,12
3	7,27	0,02	35	4,22	0,02
4	8,53	0,25	36	3,79	0,02
5	4,62	0,12	37	5,51	0,11
6	4,11	0,14	38	7,49	0,21
7	3,54	0,12	39	7,24	0,25
8	4,85	0,07	40	4,89	0,04
9	5,15	0,06	41	4,78	0,08
10	4,46	0,08	42	4,96	0,08
11	4,71	0,09	43	7,11	0,08
12	5,79	0,20	44	4,75	0,04
13	6,54	0,23	45	5,29	0,05
14	6,95	0,01	46	3,98	0,07
15	8,40	0,27	47	7,14	0,08
16	8,27	0,03	48	5,86	0,18
17	8,42	0,06	49	6,73	0,12
18	9,00	0,13	50	6,20	0,04
19	8,01	0,18	51	3,32	0,06
20	8,20	0,10	52	2,74	0,07
21	8,67	0,14	53	3,81	0,04
22	7,06	0,11	54	4,12	0,04
23	7,46	0,13	55	5,29	0,11
24	8,58	0,01	56	4,96	0,01
25	7,44	0,08	57	3,88	0,13
26	7,56	0,06	58	4,13	0,01
27	7,84	0,10	59	6,33	0,18
28	8,60	0,13	60	5,36	0,13
29	8,86	0,12	61	5,82	0,04
30	3,46	0,04	62	6,58	0,04
31	3,47	0,09	63	4,34	0,11
32	3,08	0,01	64	6,55	0,06

Tabela 12: Resultados da determinação de Nitrogênio em amostras de taninos pelo método de referência CHNS.

Código das Amostras	Teor de Nitrogênio (% g/g)	Desvio Padrão	Código das Amostras	Teor de Nitrogênio (% g/g)	Desvio Padrão
1	0,129	0,006	33	0,362	0,002
2	0,093	0,004	34	0,439	0,016
3	0,132	0,008	35	0,628	0,072
4	0,044	0,006	36	0,492	0,005
5	0,169	0,005	37	0,432	0,010
6	0,085	0,005	38	0,427	0,020
7	0,141	0,018	39	0,289	0,002
8	0,094	0,005	40	0,330	0,014
9	0,094	0,001	41	0,381	0,004
10	0,148	0,003	42	0,378	0,005
11	0,074	0,006	43	0,360	0,001
12	0,101	0,002	44	0,430	0,011
13	0,082	0,009	45	0,443	0,011
14	0,099	0,002	46	0,454	0,010
15	0,079	0,007	47	0,490	0,015
16	0,064	0,005	48	0,492	0,005
17	0,057	0,010	49	0,410	0,011
18	0,058	0,002	50	0,409	0,007
19	0,059	0,010	51	0,438	0,003
20	0,055	0,003	52	0,420	0,003
21	0,073	0,002	53	0,429	0,001
22	0,086	0,010	54	0,477	0,018
23	0,094	0,004	55	0,447	0,015
24	0,163	0,012	56	0,608	0,004
25	0,160	0,012	57	0,387	0,004
26	0,092	0,007	58	0,408	0,014
27	0,070	0,002	59	0,346	0,026
28	0,106	0,004	60	0,336	0,003
29	0,087	0,004	61	0,369	0,015
30	0,344	0,008	62	0,360	0,004
31	0,430	0,017	63	0,114	0,003
32	0,429	0,023	64	0,181	0,005

Tabela 13: Resultados da determinação de Enxofre em amostras de taninos pelo método de referência CHNS.

Código das Amostras	Teor de Enxofre (% g/g)	Desvio Padrão	Código das Amostras	Teor de Enxofre (% g/g)	Desvio Padrão
1	3,025	0,058	33	0,098	0,001
2	1,365	0,017	34	1,022	0,019
3	3,026	0,086	35	0,669	0,003
4	2,849	0,056	36	0,693	0,005
5	1,334	0,012	37	1,332	0,007
6	0,936	0,042	38	2,477	0,125
7	0,937	0,031	39	2,046	0,030
8	1,315	0,055	40	0,971	0,016
9	1,642	0,010	41	0,922	0,014
10	1,554	0,006	42	1,025	0,024
11	1,048	0,012	43	1,671	0,027
12	2,061	0,031	44	1,039	0,007
13	2,438	0,079	45	0,996	0,047
14	2,845	0,125	46	0,905	0,019
15	3,088	0,116	47	2,002	0,005
16	2,283	0,070	48	1,634	0,010
17	3,373	0,024	49	2,216	0,084
18	3,159	0,035	50	1,674	0,030
19	3,036	0,184	51	0,487	0,006
20	3,058	0,088	52	0,350	0,020
21	2,193	0,129	53	0,337	0,002
22	2,339	0,040	54	0,930	0,014
23	2,706	0,059	55	1,160	0,001
24	2,863	0,058	56	0,989	0,017
25	2,795	0,077	57	0,794	0,006
26	2,551	0,061	58	0,828	0,013
27	2,657	0,033	59	1,656	0,013
28	2,913	0,004	60	1,541	0,058
29	3,383	0,255	61	1,479	0,039
30	0,313	0,008	62	1,983	0,101
31	0,483	0,028	63	1,000	0,025
32	0,134	0,002	64	2,559	0,225

Tabela 12: Resultados da determinação de Carbono em amostras de taninos pelo método de referência CHNS.

Código das Amostras	Teor de Carbono (% g/g)	Desvio Padrão	Código das Amostras	Teor de Carbono (% g/g)	Desvio Padrão
1	47,970	0,071	33	50,609	0,430
2	48,766	0,084	34	49,004	0,021
3	46,509	0,061	35	48,634	0,027
4	47,268	0,055	36	49,243	0,320
5	49,711	0,005	37	48,932	0,074
6	50,373	0,204	38	49,181	0,267
7	49,964	0,028	39	47,944	0,086
8	50,547	0,040	40	51,283	0,834
9	48,987	0,183	41	48,753	0,406
10	50,338	0,083	42	48,666	0,660
11	50,881	0,718	43	47,815	0,117
12	50,595	0,092	44	48,679	0,083
13	48,489	0,344	45	48,692	0,223
14	48,006	0,199	46	48,340	0,295
15	47,336	0,089	47	47,993	0,192
16	47,383	0,113	48	47,815	0,129
17	47,793	0,073	49	48,182	0,022
18	47,223	0,080	50	48,072	0,039
19	48,242	0,008	51	49,873	0,059
20	48,341	0,204	52	48,548	0,088
21	47,856	0,096	53	48,419	0,148
22	48,496	0,194	54	49,013	0,059
23	48,124	0,030	55	49,296	0,225
24	47,668	0,071	56	49,673	0,165
25	47,065	0,071	57	49,567	0,194
26	48,689	0,162	58	49,699	0,180
27	48,335	0,004	59	48,000	0,042
28	48,566	0,267	60	48,878	0,065
29	47,777	0,182	61	48,900	0,226
30	50,223	0,048	62	47,152	0,255
31	49,275	0,270	63	49,845	0,153
32	48,960	0,235	64	47,989	0,093

Tabela 13: Resultados da determinação de Hidrogênio em amostras de taninos pelo método de referência CHNS.

Código das Amostras	Teor de Hidrogênio (% g/g)	Desvio Padrão	Código das Amostras	Teor de Hidrogênio (% g/g)	Desvio Padrão
1	5,076	0,098	33	5,559	0,129
2	5,338	0,085	34	5,214	0,106
3	5,154	0,114	35	5,054	0,047
4	5,116	0,090	36	5,080	0,037
5	5,268	0,047	37	5,205	0,090
6	5,447	0,083	38	5,125	0,106
7	5,362	0,133	39	5,028	0,107
8	5,347	0,094	40	5,427	0,033
9	5,281	0,097	41	5,176	0,052
10	5,303	0,103	42	5,213	0,098
11	5,159	0,066	43	4,930	0,085
12	5,156	0,081	44	5,129	0,061
13	5,340	0,049	45	4,983	0,157
14	5,179	0,018	46	5,133	0,103
15	5,214	0,073	47	5,045	0,077
16	5,214	0,068	48	5,040	0,094
17	5,135	0,092	49	5,119	0,090
18	5,279	0,107	50	5,021	0,093
19	5,211	0,055	51	5,293	0,101
20	5,108	0,090	52	5,307	0,033
21	5,171	0,103	53	5,267	0,122
22	5,207	0,076	54	5,191	0,104
23	5,235	0,042	55	5,206	0,052
24	5,137	0,081	56	5,337	0,015
25	5,106	0,099	57	5,327	0,092
26	5,205	0,084	58	5,342	0,087
27	5,138	0,088	59	5,019	0,052
28	5,102	0,089	60	5,170	0,060
29	5,082	0,060	61	5,104	0,109
30	5,481	0,088	62	4,913	0,101
31	5,477	0,126	63	5,378	0,107
32	5,349	0,131	64	5,271	0,081

8.2 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE NITROGÊNIO EM TANINOS.

Tabela 14: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de nitrogênio por NIR sem MSC.

Modelo ^a	Intervalos selecionados	VLS ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	2	6	0,956	0,036	0,921	0,048
iPLS4	3	8	0,944	0,041	0,912	0,050
iPLS8	2	7	0,940	0,042	0,924	0,050
iPLS16	4	8	0,943	0,041	0,911	0,050
iPLS32	22	4	0,923	0,048	0,872	0,066
s2iPLS8	4, 6	8	0,955	0,036	0,927	0,047
s3iPLS8	4, 6, 8	6	0,958	0,036	0,931	0,043
s4iPLS8	4, 6, 7, 8	5	0,961	0,034	0,934	0,043
s2iPLS16	4, 10	6	0,949	0,039	0,919	0,050
s3iPLS16	9, 12, 13	7	0,961	0,034	0,897	0,050
s4iPLS16	4, 9, 10, 16	10	0,962	0,034	0,936	0,043
s2iPLS32	8, 30	9	0,956	0,036	0,946	0,043
s3iPLS32	13, 26, 30	9	0,961	0,034	0,939	0,042
s4iPLS32	2, 15, 28, 30	10	0,971	0,029	0,949	0,038

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 15: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de nitrogênio por ATR-FTIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	2	8	0,925	0,043	0,924	0,055
iPLS4	4	7	0,931	0,041	0,917	0,556
iPLS8	8	4	0,926	0,042	0,917	0,055
iPLS16	15	9	0,925	0,043	0,933	0,048
iPLS32	26	6	0,908	0,047	0,091	0,062
s2iPLS8	7, 8	7	0,931	0,041	0,918	0,056
s3iPLS8	3, 7, 8	8	0,926	0,042	0,916	0,057
s4iPLS8	3, 4, 7, 8	9	0,923	0,044	0,916	0,056
s2iPLS16	15, 16	6	0,935	0,040	0,912	0,051
s3iPLS16	1, 13, 15	7	0,936	0,039	0,927	0,052
s4iPLS16	1, 10, 13, 15	8	0,938	0,039	0,933	0,053
s2iPLS32	26, 27	5	0,917	0,045	0,895	0,067
s3iPLS32	29, 30, 31	6	0,936	0,039	0,927	0,050
s4iPLS32	24, 25, 29, 32	9	0,946	0,036	0,914	0,060

^a iPLS – número de divisões do espectro; ^s – número de intervalos combinados; ^b número de variáveis latentes;

Tabela 16: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de nitrogênio por ATR-FTIR com MSC.

Modelo ^a	Intervalos selecionados	VLS ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2M	2	10	0,908	0,048	0,900	0,061
iPLS4M	3	3	0,873	0,055	0,873	0,066
iPLS8M	6	3	0,878	0,054	0,873	0,068
iPLS16M	13	5	0,921	0,044	0,914	0,058
iPLS32M	26	2	0,932	0,041	0,917	0,062
s2iPLS8M	3, 8	6	0,910	0,047	0,896	0,058
s3iPLS8M	2, 6, 8	6	0,901	0,042	0,876	0,062
s4iPLS8M	2, 3, 6, 8	6	0,900	0,049	0,883	0,062
s2iPLS16M	4, 13	2	0,910	0,047	0,911	0,060
s3iPLS16M	3, 15, 16	6	0,915	0,045	0,900	0,057
s4iPLS16M	5, 12, 15, 16	6	0,912	0,046	0,893	0,060
s2iPLS32M	26, 31	4	0,924	0,043	0,912	0,067
s3iPLS32M	7, 26, 31	7	0,936	0,040	0,918	0,062
s4iPLS32M	6, 24, 25, 26	7	0,936	0,040	0,910	0,062

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes.

Tabela 17: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de nitrogênio por fusão dos dados NIR e ATR-FTIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8	2, 6	9	0,931	0,042	0,842	0,070
s3iPLS8	1, 2, 6	8	0,924	0,044	0,875	0,063
s4iPLS8	1, 2, 3, 5	10	0,932	0,042	0,903	0,054
s2iPLS16	4, 5	9	0,939	0,040	0,075	0,089
s3iPLS16	3, 4, 5	9	0,951	0,035	0,879	0,060
s4iPLS16	3, 4, 5, 7	10	0,942	0,039	0,860	0,065
s2iPLS32	3, 4, 5, 7	10	0,942	0,039	0,860	0,065
s3iPLS32	4, 6	10	0,946	0,037	0,846	0,081
s4iPLS32	8, 10, 25	9	0,966	0,030	0,775	0,085

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 18: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de nitrogênio por fusão dos dados NIR e ATR-FTIR com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8M	2, 3	8	0,945	0,038	0,880	0,053
s3iPLS8M	1, 2, 6	8	0,928	0,043	0,864	0,058
s4iPLS8M	1, 2, 3, 5	10	0,937	0,040	0,901	0,050
s2iPLS16M	2, 3	14	0,961	0,032	0,851	0,060
s3iPLS16M	1, 2, 3	10	0,955	0,034	0,907	0,050
s4iPLS16M	2, 3, 4, 5	8	0,949	0,036	0,906	0,047
s2iPLS32M	4, 7	7	0,947	0,037	0,923	0,047
s3iPLS32M	4, 5, 6	12	0,960	0,032	0,857	0,060
s4iPLS32M	4, 8, 10, 25	9	0,966	0,030	0,880	0,055

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

Tabela 19: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de nitrogênio por fusão dos dados NIR e ATR-FTIR com resolução 1:2 sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8	1, 2	6	0,930	0,043	0,927	0,048
s3iPLS8	1, 2, 6	9	0,930	0,043	0,914	0,052
s4iPLS8	1, 2, 7, 8	12	0,941	0,040	0,882	0,059
s2iPLS16	4, 8	9	0,931	0,043	0,794	0,083
s3iPLS16	1, 3, 13	10	0,937	0,041	0,896	0,057
s4iPLS16	2, 3, 4, 12	8	0,935	0,042	0,921	0,050
s2iPLS32	7, 15	7	0,937	0,041	0,822	0,080
s3iPLS32	1, 3, 7	10	0,945	0,038	0,917	0,052
s4iPLS32	1, 9, 28, 30	10	0,948	0,037	0,827	0,075

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 20: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de nitrogênio por fusão dos dados NIR e ATR-FTIR com resolução 1:2 com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8M	1, 2	5	0,933	0,042	0,929	0,047
s3iPLS8M	1, 2, 6	8	0,929	0,043	0,917	0,052
s4iPLS8M	1, 2, 4, 7	11	0,936	0,041	0,917	0,054
s2iPLS16M	2, 3	6	0,938	0,041	0,917	0,055
s3iPLS16M	1, 3, 13	8	0,937	0,041	0,897	0,058
s4iPLS16M	1, 2, 13, 14	10	0,950	0,036	0,845	0,070
s2iPLS32M	4, 7	7	0,940	0,040	0,896	0,060
s3iPLS32M	4, 7, 15	10	0,949	0,037	0,902	0,064
s4iPLS32M	6, 24, 26, 30	11	0,952	0,036	0,886	0,059

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

8.3 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE ENXOFRE EM TANINOS.

Tabela 21: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de enxofre por NIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	1	5	0,938	0,227	0,935	0,242
iPLS4	1	8	0,931	0,239	0,936	0,241
iPLS8	2	5	0,926	0,248	0,927	0,264
iPLS16	4	5	0,892	0,300	0,975	0,160
iPLS32	6	5	0,736	0,473	0,930	0,261
s2iPLS8	2, 6	5	0,925	0,249	0,915	0,276
s3iPLS8	1, 2, 5	5	0,940	0,222	0,948	0,218
s4iPLS8	1, 2, 3, 8	6	0,940	0,223	0,940	0,233
s2iPLS16	3, 11	6	0,939	0,226	0,935	0,243
s3iPLS16	3, 4, 6	8	0,943	0,217	0,956	0,201
s4iPLS16	2, 3, 5, 9	9	0,947	0,211	0,972	0,161
s2iPLS32	5, 23	8	0,947	0,209	0,941	0,237
s3iPLS32	3, 5, 18	7	0,954	0,195	0,956	0,199
s4iPLS32	3, 5, 17, 18	8	0,957	0,189	0,959	0,193

^a iPLS – número de divisões do espectro; ^s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 22: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de enxofre por ATR-FTIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	2	7	0,837	0,371	0,952	0,246
iPLS4	4	6	0,871	0,328	0,968	0,219
iPLS8	7	5	0,887	0,306	0,963	0,277
iPLS16	14	6	0,889	0,305	0,946	0,205
iPLS32	27	4	0,846	0,358	0,935	0,238
s2iPLS8	2, 7	7	0,877	0,321	0,959	0,260
s3iPLS8	2, 4, 7	8	0,878	0,320	0,940	0,290
s4iPLS8	2, 3, 5, 7	9	0,871	0,329	0,946	0,264
s2iPLS16	6, 14	8	0,893	0,301	0,951	0,200
s3iPLS16	6, 7, 14	8	0,893	0,300	0,954	0,195
s4iPLS16	4, 6, 14, 16	10	0,898	0,292	0,935	0,243
s2iPLS32	27, 31	8	0,910	0,275	0,906	0,301
s3iPLS32	10, 27, 31	10	0,920	0,260	0,877	0,375
s4iPLS32	6, 13, 27, 31	10	0,927	0,247	0,865	0,423

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 23: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de enxofre por ATR-FTIR com MSC.

Modelo ^a	Intervalos selecionados	VLS ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2M	2	6	0,861	0,343	0,960	0,235
iPLS4M	4	6	0,865	0,338	0,981	0,179
iPLS8M	7	3	0,870	0,329	0,955	0,268
iPLS16M	14	9	0,878	0,321	0,927	0,284
iPLS32M	28	8	0,838	0,371	0,932	0,267
s2iPLS8M	3, 7	5	0,875	0,323	0,964	0,234
s3iPLS8M	1, 7, 8	7	0,869	0,331	0,974	0,206
s4iPLS8M	2, 3, 5, 7	7	0,873	0,326	0,952	0,259
s2iPLS16M	6, 14	5	0,871	0,328	0,939	0,237
s3iPLS16M	11, 13, 14	5	0,890	0,303	0,970	0,235
s4iPLS16M	10, 13, 14, 15	5	0,890	0,303	0,973	0,248
s2iPLS32M	27, 30	6	0,882	0,314	0,968	0,166
s3iPLS32M	3, 27, 31	9	0,923	0,254	0,914	0,284
s4iPLS32M	5, 10, 27, 31	11	0,939	0,227	0,839	0,418

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

Tabela 24: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de enxofre por fusão dos dados NIR e ATR-FTIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8	2, 5	8	0,947	0,209	0,968	0,176
s3iPLS8	1, 2, 5	9	0,949	0,206	0,985	0,124
s4iPLS8	2, 3, 5, 8	10	0,941	0,221	0,967	0,185
s2iPLS16	3, 15	10	0,959	0,184	0,936	0,249
s3iPLS16	1, 3, 15	11	0,959	0,184	0,949	0,224
s4iPLS16	1, 3, 6, 8	10	0,955	0,192	0,974	0,170
s2iPLS32	6, 12	11	0,959	0,182	0,751	0,490
s3iPLS32	1, 5, 18	9	0,958	0,185	0,947	0,240
s4iPLS32	5, 6, 17, 29	9	0,965	0,171	0,957	0,227

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 25: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de enxofre por fusão dos dados NIR e ATR-FTIR com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8M	2, 5	8	0,940	0,221	0,972	0,161
s3iPLS8M	1, 2, 5	8	0,947	0,209	0,985	0,129
s4iPLS8M	1, 2, 5, 8	8	0,941	0,221	0,985	0,122
s2iPLS16M	3, 4	9	0,937	0,227	0,593	0,688
s3iPLS16M	3, 4, 15	11	0,957	0,189	0,968	0,181
s4iPLS16M	3, 4, 9, 16	8	0,947	0,209	0,975	0,159
s2iPLS32M	6, 7	9	0,952	0,199	0,566	0,720
s3iPLS32M	5, 6, 12	10	0,955	0,192	0,536	0,798
s4iPLS32M	5, 7, 11, 21	10	0,960	0,182	0,949	0,240

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

Tabela 26: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de enxofre por fusão dos dados NIR e ATR-FTIR com resolução 1:2 sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8	1, 6	7	0,921	0,256	0,971	0,167
s3iPLS8	1, 2, 7	8	0,925	0,249	0,985	0,128
s4iPLS8	1, 2, 4, 7	9	0,932	0,237	0,981	0,141
s2iPLS16	1, 6	8	0,937	0,228	0,955	0,225
s3iPLS16	1, 4, 16	8	0,937	0,229	0,985	0,128
s4iPLS16	1, 10, 13, 14	9	0,935	0,232	0,969	0,182
s2iPLS32	1, 25	10	0,938	0,228	0,934	0,246
s3iPLS32	1, 8, 31	6	0,944	0,215	0,931	0,255
s4iPLS32	1, 11, 12, 19	10	0,951	0,202	0,943	0,232

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 27: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de enxofre por fusão dos dados NIR e ATR-FTIR com resolução 1:2 com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8M	2, 7	7	0,926	0,247	0,983	0,127
s3iPLS8M	1, 2, 7	7	0,932	0,238	0,989	0,103
s4iPLS8M	1, 2, 4, 7	8	0,933	0,236	0,982	0,136
s2iPLS16M	1, 3	6	0,913	0,266	0,925	0,276
s3iPLS16M	1, 4, 16	9	0,936	0,231	0,958	0,207
s4iPLS16M	1, 2, 4, 16	8	0,935	0,232	0,954	0,220
s2iPLS32M	1, 5	8	0,940	0,223	0,970	0,173
s3iPLS32M	1, 8, 31	7	0,943	0,217	0,963	0,267
s4iPLS32M	1, 4, 5, 11	10	0,953	0,198	0,971	0,168

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

8.4 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE CINZAS EM TANINOS.

Tabela 28: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de cinzas por NIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	1	9	0,920	0,52	0,907	0,57
iPLS4	2	10	0,909	0,55	0,939	0,48
iPLS8	2	5	0,905	0,57	0,875	0,68
iPLS16	3	5	0,870	0,66	0,870	0,71
iPLS32	6	5	0,884	0,62	0,886	0,66
s2iPLS8	2, 4	7	0,907	0,56	0,896	0,61
s3iPLS8	2, 6, 7	5	0,919	0,52	0,914	0,55
s4iPLS8	1, 2, 6, 7	5	0,921	0,54	0,911	0,56
s2iPLS16	3, 5	9	0,908	0,55	0,918	0,55
s3iPLS16	3, 6, 8	10	0,932	0,48	0,920	0,55
s4iPLS16	3, 7, 9, 12	6	0,930	0,49	0,920	0,54
s2iPLS32	6, 20	4	0,912	0,54	0,887	0,63
s3iPLS32	7, 14, 31	9	0,935	0,47	0,953	0,42
s4iPLS32	5, 11, 15, 16	9	0,940	0,45	0,922	0,55

^a iPLS – número de divisões do espectro; ^s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 29: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de cinzas por ATR-FTIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	2	5	0,680	1,04	0,810	0,76
iPLS4	4	8	0,727	0,97	0,870	0,68
iPLS8	7	5	0,751	0,91	0,886	0,64
iPLS16	14	5	0,747	0,92	0,831	0,75
iPLS32	28	8	0,763	0,89	0,729	0,92
s2iPLS8	2, 7	7	0,747	0,93	0,913	0,60
s3iPLS8	2, 4, 7	9	0,792	0,83	0,909	0,59
s4iPLS8	2, 3, 5, 7	9	0,763	0,90	0,907	0,59
s2iPLS16	5, 14	8	0,785	0,85	0,861	0,72
s3iPLS16	4, 7, 14	8	0,797	0,83	0,867	0,71
s4iPLS16	2, 4, 7, 14	9	0,828	0,76	0,869	0,69
s2iPLS32	10, 27	11	0,841	0,73	0,801	1,01
s3iPLS32	10, 13, 27	12	0,867	0,67	0,795	1,00
s4iPLS32	8, 10, 13, 27	11	0,879	0,63	0,815	0,89

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 30: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de cinzas por ATR-FTIR com MSC.

Modelo ^a	Intervalos selecionados	VLS ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2M	2	5	0,804	0,81	0,919	0,54
iPLS4M	4	4	0,770	0,97	0,888	0,60
iPLS8M	7	4	0,798	0,83	0,899	0,58
iPLS16M	14	6	0,835	0,75	0,857	0,66
iPLS32M	28	9	0,803	0,81	0,890	0,58
s2iPLS8M	6, 8	5	0,818	0,78	0,899	0,66
s3iPLS8M	6, 7, 8	5	0,810	0,80	0,916	0,55
s4iPLS8M	1, 3, 4, 8	9	0,906	0,79	0,877	0,60
s2iPLS16M	12, 14	7	0,840	0,74	0,855	0,69
s3iPLS16M	2, 5, 15	8	0,857	0,69	0,774	0,82
s4iPLS16M	2, 3, 5, 15	10	0,856	0,70	0,851	0,74
s2iPLS32M	27, 30	4	0,855	0,69	0,898	0,56
s3iPLS32M	10, 13, 27	12	0,919	0,52	0,861	0,74
s4iPLS32M	8, 10, 13, 27	10	0,918	0,52	0,848	0,72

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

Tabela 31: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de cinzas por fusão dos dados NIR e ATR-FTIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8	1, 8	8	0,927	0,50	0,874	0,65
s3iPLS8	1, 2, 8	8	0,919	0,53	0,921	0,50
s4iPLS8	1, 2, 5, 8	9	0,916	0,54	0,948	0,45
s2iPLS16	1, 16	6	0,932	0,48	0,877	0,63
s3iPLS16	2, 3, 16	7	0,933	0,48	0,909	0,52
s4iPLS16	2, 3, 6, 16	7	0,932	0,49	0,853	0,70
s2iPLS32	5, 7	5	0,929	0,50	0,897	0,56
s3iPLS32	4, 5, 7	6	0,939	0,46	0,926	0,48
s4iPLS32	5, 6, 8, 31	7	0,938	0,47	0,891	0,59

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 32: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de cinzas por fusão dos dados NIR e ATR-FTIR com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8M	2, 8	6	0,924	0,52	0,912	0,53
s3iPLS8M	1, 2, 8	8	0,930	0,49	0,918	0,51
s4iPLS8M	1, 2, 3, 8	8	0,920	0,53	0,911	0,53
s2iPLS16M	1, 3	6	0,923	0,52	0,910	0,55
s3iPLS16M	3, 4, 16	5	0,929	0,50	0,898	0,57
s4iPLS16M	2, 3, 4, 16	7	0,936	0,47	0,896	0,58
s2iPLS32M	1, 5	6	0,925	0,51	0,929	0,51
s3iPLS32M	3, 6, 15	8	0,931	0,49	0,909	0,58
s4iPLS32M	4, 5, 7, 19	8	0,940	0,46	0,929	0,47

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

Tabela 33: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de cinzas por fusão dos dados NIR e ATR-FTIR com resolução 1:2 sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8	1, 2	7	0,915	0,55	0,939	0,48
s3iPLS8	1, 2, 7	8	0,916	0,55	0,929	0,50
s4iPLS8	1, 2, 5, 7	9	0,920	0,54	0,938	0,47
s2iPLS16	3, 16	9	0,937	0,48	0,938	0,46
s3iPLS16	2, 3, 16	9	0,944	0,45	0,899	0,55
s4iPLS16	2, 3, 4, 16	9	0,941	0,46	0,919	0,51
s2iPLS32	5, 6	11	0,944	0,44	0,902	0,76
s3iPLS32	2, 5, 32	8	0,937	0,47	0,914	0,51
s4iPLS32	2, 5, 7, 32	9	0,942	0,46	0,899	0,56

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 34: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de cinzas por fusão dos dados NIR e ATR-FTIR com resolução 1:2 com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8M	2, 7	7	0,911	0,57	0,937	0,50
s3iPLS8M	2, 5, 7	9	0,920	0,54	0,915	0,56
s4iPLS8M	1, 2, 5, 7	8	0,923	0,53	0,936	0,47
s2iPLS16M	1, 3	6	0,927	0,51	0,919	0,55
s3iPLS16M	1, 3, 16	10	0,936	0,48	0,812	0,77
s4iPLS16M	2, 4, 9, 16	9	0,940	0,46	0,941	0,48
s2iPLS32M	1, 5	6	0,928	0,51	0,915	0,56
s3iPLS32M	2, 5, 7	6	0,929	0,50	0,927	0,49
s4iPLS32M	2, 5, 8, 32	7	0,947	0,43	0,891	0,57

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

8.5 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE CARBONO EM TANINOS.

Tabela 35: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de carbono por NIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	1	4	0,692	0,572	0,697	0,490
iPLS4	1	3	0,683	0,581	0,659	0,512
iPLS8	2	3	0,686	0,578	0,711	0,497
iPLS16	3	2	0,692	0,572	0,652	0,522
iPLS32	6	6	0,703	0,564	0,624	0,525
s2iPLS8	2, 6	4	0,735	0,531	0,630	0,499
s3iPLS8	2, 4, 6	5	0,770	0,495	0,651	0,481
s4iPLS8	2, 6, 7, 8	8	0,778	0,492	0,557	0,562
s2iPLS16	9, 12	5	0,800	0,462	0,583	0,539
s3iPLS16	9, 12, 13	6	0,794	0,472	0,616	0,520
s4iPLS16	3, 11, 13, 15	7	0,820	0,440	0,549	0,571
s2iPLS32	6, 23	4	0,767	0,498	0,606	0,508
s3iPLS32	8, 16, 21	8	0,843	0,410	0,527	0,606
s4iPLS32	8, 9, 17, 21	8	0,839	0,415	0,499	0,646

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 36: Resultados das figuras de mérito dos modelos de regressão por siPLS para determinação do teor de carbono por NIR com MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
s2iPLS8M	3, 6	7	0,774	0,493	0,548	0,562
s3iPLS8M	1, 4, 6	5	0,774	0,491	0,595	0,508
s4iPLS8M	2, 6, 7, 8	8	0,783	0,484	0,561	0,558
s2iPLS16M	9, 12	4	0,804	0,457	0,580	0,539
s3iPLS16M	6, 9, 12	5	0,794	0,467	0,604	0,526
s4iPLS16M	2, 11, 13, 15	6	0,835	0,419	0,561	0,554
s2iPLS32M	18, 23	6	0,822	0,437	0,589	0,536
s3iPLS32M	18, 23, 24	3	0,804	0,457	0,590	0,527
s4iPLS32M	5, 21, 26, 29	7	0,837	0,418	0,597	0,551

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC; ^b número de variáveis latentes.

Tabela 37: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de carbono por ATR-FTIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	2	8	0,669	0,592	0,687	0,497
iPLS4	4	7	0,604	0,649	0,779	0,485
iPLS8	7	5	0,681	0,637	0,723	0,436
iPLS16	14	6	0,664	0,594	0,466	0,601
iPLS32	27	4	0,679	0,575	0,502	0,562
s2iPLS8	2, 7	8	0,774	0,487	0,687	0,421
s3iPLS8	2, 3, 7	8	0,770	0,493	0,712	0,417
s4iPLS8	2, 3, 4, 7	8	0,763	0,502	0,711	0,424
s2iPLS16	12, 14	6	0,751	0,506	0,564	0,512
s3iPLS16	6, 14, 16	9	0,775	0,488	0,553	0,496
s4iPLS16	3, 5, 13, 14	7	0,794	0,462	0,711	0,404
s2iPLS32	23, 27	7	0,763	0,497	0,579	1,010
s3iPLS32	7, 20, 27	6	0,793	0,666	0,453	0,615
s4iPLS32	5, 11, 21, 27	14	0,832	0,417	0,411	0,560

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 38: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de carbono por ATR-FTIR com MSC.

Modelo ^a	Intervalos selecionados	VLS ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2M	2	6	0,673	0,583	0,685	0,460
iPLS4M	3	7	0,557	0,690	0,058	0,498
iPLS8M	7	4	0,683	0,575	0,582	0,481
iPLS16M	14	6	0,639	0,608	0,454	0,551
iPLS32M	24	8	0,665	0,596	0,413	0,608
s2iPLS8M	6, 7	5	0,708	0,548	0,656	0,480
s3iPLS8M	2, 3, 8	8	0,700	0,563	0,738	0,404
s4iPLS8M	1, 2, 3, 7	9	0,775	0,490	0,667	0,447
s2iPLS16M	12, 14	4	0,731	0,527	0,578	0,518
s3iPLS16M	6, 14, 16	8	0,749	0,510	0,544	0,530
s4iPLS16M	2, 5, 14, 16	9	0,757	0,505	0,683	0,421
s2iPLS32M	27, 32	7	0,769	0,492	0,426	0,591
s3iPLS32M	2, 11, 27	6	0,790	0,465	0,398	0,578
s4iPLS32M	2, 11, 27, 32	9	0,778	0,479	0,511	0,536

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes

8.6 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS NIR, ATR-FTIR E FUSÃO NA DETERMINAÇÃO DO TEOR DE HIDROGÊNIO EM TANINOS.

Tabela 39: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de hidrogênio por NIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	2	4	0,318	0,110	0,615	0,092
iPLS4	2	6	0,345	0,108	0,551	0,098
iPLS8	7	8	0,379	0,108	0,426	0,112
iPLS16	13	7	0,437	0,101	0,493	0,102
iPLS32	7	6	0,463	0,097	0,358	0,113
s2iPLS8	3, 7	3	0,322	0,109	0,615	0,098
s3iPLS8	5, 6, 7	3	0,580	0,108	0,618	0,095
s4iPLS8	5, 6, 7, 8	3	0,314	0,110	0,615	0,092
s2iPLS16	6, 13	4	0,396	0,103	0,563	0,099
s3iPLS16	7, 12, 13	3	0,362	0,106	0,665	0,086
s4iPLS16	8, 13, 14, 15	6	0,388	0,105	0,578	0,098
s2iPLS32	25, 26	7	0,437	0,101	0,493	0,102
s3iPLS32	11, 25, 26	4	0,413	0,101	0,575	0,098
s4iPLS32	11, 13, 25, 27	6	0,430	0,100	0,495	0,106

^a iPLS – número de divisões do espectro; ^s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 40: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de hidrogênio por ATR-FTIR sem MSC.

Modelo ^a	Intervalos selecionados	VLs ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2	2	5	0,473	0,091	0,246	0,089
iPLS4	4	4	0,426	0,095	0,119	0,102
iPLS8	8	4	0,377	0,099	0,139	0,099
iPLS16	14	6	0,455	0,093	0,330	0,083
iPLS32	30	5	0,380	0,099	0,313	0,092
s2iPLS8	6, 8	5	0,451	0,093	0,383	0,082
s3iPLS8	6, 7, 8	5	0,486	0,091	0,371	0,082
s4iPLS8	4, 6, 7, 8	5	0,501	0,089	0,352	0,083
s2iPLS16	14, 15	5	0,444	0,093	0,260	0,089
s3iPLS16	2, 3, 14	7	0,569	0,082	0,212	0,101
s4iPLS16	2, 13, 14, 16	8	0,556	0,085	0,278	0,093
s2iPLS32	10, 28	9	0,541	0,085	0,037	0,104
s3iPLS32	3, 5, 28	10	0,594	0,081	0,055	0,124
s4iPLS32	3, 5, 27, 28	8	0,577	0,082	0,027	0,089

^a iPLS – número de divisões do espectro; s – número de intervalos combinados, ^b número de variáveis latentes

Tabela 41: Resultados das figuras de mérito dos modelos de regressão por iPLS e siPLS para determinação do teor de hidrogênio por ATR-FTIR com MSC.

Modelo ^a	Intervalos selecionados	VLS ^b	Rc ²	RMSECV (g/g%)	Rp ²	RMSEP (g/g%)
iPLS2M	2	5	0,531	0,085	0,324	0,047
iPLS4M	4	4	0,450	0,093	0,171	0,097
iPLS8M	7	5	0,399	0,098	0,357	0,082
iPLS16M	15	3	0,396	0,097	0,088	0,104
iPLS32M	23	4	0,417	0,095	0,159	0,097
s2iPLS8M	6, 8	3	0,499	0,089	0,437	0,083
s3iPLS8M	6, 7, 8	5	0,524	0,086	0,321	0,085
s4iPLS8M	3, 6, 7, 8	4	0,534	0,085	0,363	0,081
s2iPLS16M	14, 16	5	0,465	0,091	0,155	0,098
s3iPLS16M	13, 14, 16	6	0,550	0,086	0,259	0,093
s4iPLS16M	12, 13, 14, 16	4	0,539	0,085	0,351	0,083
s2iPLS32M	27, 28	7	0,569	0,082	0,223	0,095
s3iPLS32M	6, 28, 32	8	0,578	0,081	0,147	0,098
s4iPLS32M	5, 6, 12, 27	6	0,366	0,101	0,176	0,099

^a iPLS – número de divisões do espectro; s – número de intervalos combinados; M – tratamento com MSC, ^b número de variáveis latentes