

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

DISSERTAÇÃO DE MESTRADO

**Caracterização de Bebidas à Base de Soja Empregando
Espectroscopia no Infravermelho Médio com Transformada de
Fourier por Reflexão Total Atenuada e Quimiometria**

ANDRÉ MACHADO RECH

Porto Alegre, abril de 2018.

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

ANDRÉ MACHADO RECH

**Caracterização de Bebidas à Base de Soja Empregando
Espectroscopia no Infravermelho Médio com Transformada de
Fourier por Reflexão Total Atenuada e Quimiometria**

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

Prof. Dr. Marco Flôres Ferrão
Orientador

Porto Alegre, abril de 2018.

AGRADECIMENTOS

À minha família, com destaque aos meus pais, Nadir Antonio Rech e Maria Júlia Machado Rech, pelo incentivo e carinho. Vocês são exemplos de vida para mim, com suas vidas simples, mas de grandes princípios. Às minhas irmãs, Simone, Raquel e Juliana, pelo companheirismo, apoio, respeito, amizade, compreensão e conselhos. À minha sobrinha, Victória, por toda alegria, diversão, risadas e amor. À minha namorada, Simone, pelo amor, carinho, apoio, incentivo, companheirismo e paciência.

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

Ao meu grande amigo e colega gremista Fábio por toda ajuda.

À amiga e colega Camila por toda ajuda durante o trabalho.

A todos os colegas da FACFAR que possibilitaram meu estudo.

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

Ao Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul e ao MAPA pelo compartilhamento de equipamentos e de conhecimentos.

À Universidade Federal do Rio Grande do Sul, à CAPES, e ao CNPq pelos recursos materiais e estruturais para realização deste trabalho.

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	ix
RESUMO	xii
ABSTRACT	xiii
1 INTRODUÇÃO	1
2 OBJETIVOS	3
2.1 GERAL	3
2.1.1 Específicos	3
3 REVISÃO BIBLIOGRÁFICA	4
3.1 SOJA	4
3.2 ALIMENTOS À BASE DE SOJA	4
3.3 GLICÍDIOS	7
3.4 PROTEÍNAS	11
3.5 LEGISLAÇÃO ALIMENTAR	12
3.6 ESPECTROSCOPIA NO INFRAVERMELHO	13
3.6.1 Espectroscopia por Reflexão Total Atenuada (ATR)	15
3.7 QUIMIOMETRIA	16
3.7.1 Calibração Multivariada	17
3.7.2 Regressão por Mínimos Quadrados Parciais (PLS)	17
3.7.3 Métodos de Seleção de Variáveis em Modelos de Regressão	19
3.7.4 Seleção do Conjunto de Calibração e Previsão	20
3.7.5 Pré-tratamento dos Dados	21
3.7.6 Avaliação dos Modelos de Regressão	23
4 MATERIAIS E MÉTODOS	26
4.1 AMOSTRAS	26

4.2	ESPECTROSCOPIA DE INFRAVERMELHO MÉDIO	26
4.3	ANÁLISE MULTIVARIADA	26
4.3.1	Regressão multivariada dos teores de glicídios totais, glicídios redutores, glicídios não redutores e proteínas totais a partir dos dados de FTIR-ATR.	27
4.4	TEORES DE GLICÍDIOS TOTAIS, GLICÍDIOS REDUTORES, GLICÍDIOS NÃO REDUTORES E PROTEÍNA.	28
5	RESULTADOS E DISCUSSÃO	32
5.1	TEORES DE GLICÍDIOS TOTAIS, GLICÍDIOS REDUTORES, GLICÍDIOS NÃO REDUTORES E PROTEÍNA TOTAIS.	32
5.2	ESPECTROSCOPIA DE INFRAVERMELHO	32
5.3	ANÁLISE MULTIVARIADA	33
5.3.1	Determinação do Teor de Glicídios Totais por FTIR-ATR Combinada com Regressão Multivariada.	33
5.3.2	Determinação do Teor de Glicídios Redutores por FTIR-ATR Combinada com Regressão Multivariada.	37
5.3.3	Determinação do Teor de Glicídios Não Redutores por FTIR-ATR Combinada com Regressão Multivariada.	39
5.3.4	Determinação do Teor de Proteínas Totais por FTIR-ATR Combinada com Regressão Multivariada.	41
6	CONCLUSÕES	51
7	REFERÊNCIAS BIBLIOGRÁFICAS.....	52
8	APÊNDICES.....	56
8.1	RESULTADOS DA DETERMINAÇÃO DO TEOR DE GLICÍDIOS REDUTORES, GLICÍDIOS TOTAIS, GLICÍDIOS NÃO REDUTORES E DE PROTEÍNAS TOTAIS, PELOS MÉTODOS DE REFERÊNCIA.	56
8.2	RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS FTIR-ATR NA DETERMINAÇÃO DO TEOR DE GLICÍDIOS TOTAIS EM BBS.	61

8.3 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS FTIR-ATR NA DETERMINAÇÃO DO TEOR DE GLICÍDIOS REDUTORES EM BBS.....	62
8.4 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS FTIR-ATR NA DETERMINAÇÃO DO TEOR DE GLICÍDIOS NÃO REDUTORES EM BBS.	63
8.5 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS FTIR-ATR NA DETERMINAÇÃO DO TEOR DE PROTEÍNAS TOTAIS EM BBS.	64
8.6 RESULTADOS DAS DETERMINAÇÃO DO TEOR DE PROTEÍNAS TOTAIS EM BBS POR FTIR-ATR, COMBINADA COM REGRESSÃO MULTIVARIADA, UTILIZANDO O ALGORITMO KENNARD-STONE PARA SELEÇÃO DAS AMOSTRAS DE CALIBRAÇÃO E PREVISÃO.....	65

LISTA DE FIGURAS

Figura 1: Fluxograma de produção de bebidas à base de soja, a partir do EHS da Campinas Vital ¹³	5
Figura 2: Representações do polihidroxialdeído D-glicose e da polihidroxicetona D-frutose ¹⁷	7
Figura 3: Representações tridimensionais e projeção de Fischer dos enantiômeros do gliceraldeído. ¹⁷	8
Figura 4: Representações da mutarrotação entre a alfa glicose e a beta glicose ¹⁷	9
Figura 5: Representação da ligação glicosídica entre duas D-glicoses, originando uma α -1-4-maltose ¹⁷	10
Figura 6: Representação da mutarrotação da maltose ¹⁷	10
Figura 7: Representação da molécula de sacarose. ¹⁷	10
Figura 8: Representação da fórmula estrutural básica de um aminoácido.	11
Figura 9: Representação das vibrações de (a) estiramento e (b) deformações angulares. ²³	14
Figura 10: Esquema simplificada da análise por ATR ³⁰	16
Figura 11: A linha cheia indica o comportamento de erro na previsão. As linhas tracejadas indicam o comportamento da variância e tendência (viés) nos modelos PLS em geral ³²	24
Figura 12: Espectros das bebidas à base de soja a) espectro sem tratamentodos dos dados; b) espectro com tratamento dos dados, espectro suavizado, normalizado, derivados e realizadas suas médias.....	33
Figura 13: Espectro com destaque as regiões selecionadas (6, 7) combinadas com regressão siPLS para determinação de glicídios totais em BBS.	35
Figura 14: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de glicídios totais, por meio de espectroscopia no infravermelho, combinando os intervalos 6, 7 no modelo s2iPLS8.	36
Figura 15: Valores de referência pelo método Lane-Eynon <i>versus</i> valores previstos do modelo siPLS para glicídios totais usando os intervalos 6 e 7 correspondentes aos intervalos espectrais 1025 - 1150 cm^{-1} e 1151 - 1476 cm^{-1}	37

Figura 16: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de proteínas totais, por meio de espectroscopia no infravermelho, combinando to espectro no modelo PLS.....	44
Figura 17: Valores de referência pelo método Kjeldhal <i>versus</i> valores previstos do modelo PLS para proteínas totais utilizando todo o espectro correspondente ao intervalo espectral de 1899 – 900 cm ⁻¹	45
Figura 18: Espectro com destaque as regiões selecionadas (3, 5 e 6) combinadas com regressão siPLS para determinação de proteínas totais em BBS.	46
Figura 19: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de proteínas totais, por meio de espectroscopia no infravermelho, combinando os intervalos 3, 5 e 6 no modelo s3iPLS8.	46
Figura 20: Valores de referência pelo método Kjeldhal <i>versus</i> valores previstos do modelo s3iPLS8 para proteínas totais usando os intervalos 3, 5 e 6 correspondentes aos intervalos espectrais 1649 – 1525 cm ⁻¹ , 1399 – 1275 cm ⁻¹ e 1274 – 1150cm ⁻¹	47
Figura 21: Espectro com destaque as regiões selecionadas (2, 3, 5 e 6) combinadas com regressão siPLS para determinação de proteínas totais em BBS.	48
Figura 22: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de proteínas totais, por meio de espectroscopia no infravermelho, combinando os intervalos 2, 3, 5 e 6 no modelo s4iPLS8.	48
Figura 23: Valores de referência pelo método Kjeldhal <i>versus</i> valores previstos do modelo s4iPLS8 para proteínas totais usando os intervalos 2, 3, 5 e 6 correspondentes aos intervalos espectrais 1774 – 1650cm ⁻¹ , 1649 – 1525 cm ⁻¹ , 1399 – 1275 cm ⁻¹ e 1274 – 1150cm ⁻¹	49
Figura 24: Espectro com destaque as regiões selecionadas (2, 3, 5, e 6) combinadas com regressão siPLS para determinação de proteínas totais em BBS.	68
Figura 25: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de proteínas totais, por meio de espectroscopia no infravermelho, combinando os intervalos 2, 3, 5 e 6 no modelo s4iPLS8.	69
Figura 26: Valores de referência pelo método Kjeldhal <i>versus</i> valores previstos do modelo siPLS para proteínas totais usando os intervalos 2, 3, 5 e 6 correspondentes aos intervalos espectrais 1775 – 1650 cm ⁻¹ , 1649 – 1525 cm ⁻¹ , 1399 – 1275 cm ⁻¹ , 1274 – 1150 cm ⁻¹	70

LISTA DE TABELAS

Tabela 1: Expressão dos valores a serem declarados nos rótulos ²²	12
Tabela 2: Quantidades não significativas por porção ²²	13
Tabela 3: Regiões espectrais do infravermelho ²³	13
Tabela 4: Fatores de conversão de nitrogênio total em proteína ³	31
Tabela 5: Resultados das figuras de mérito dos melhores modelos de regressão, para determinação do teor de glicídios totais com dados MIR, FTIR-ATR.....	34
Tabela 6: Resultados das figuras de mérito dos melhores modelos de regressão, para determinação do teor de glicídios redutores com dados FTIR-ATR.....	38
Tabela 7: Resultados das figuras de mérito dos melhores modelos de regressão, para determinação do teor de glicídios não redutores com dados FTIR-ATR.....	40
Tabela 8: Resultados das figuras de mérito dos melhores modelos de regressão, para determinação do teor de proteínas totais com dados FTIR-ATR.....	42
Tabela 9: Resultados da determinação de glicídios redutores, em amostras de BBS pelo método de referência Lane-Eynon.....	56
Tabela 10: Resultados da determinação de glicídios totais, em amostras de BBS pelo método de referência Lane-Eynon.....	57
Tabela 11: Resultados da determinação de glicídios não redutores, em amostras de BBS pelo método de referência Lane-Eynon.....	58
Tabela 12: Resultados da determinação de proteínas totais, em amostras de BBS pelo método de referência Kjeldhal.....	59
Tabela 13: Resultados das figuras de mérito dos modelos de regressão por PLS, iPLS e siPLS para determinação do teor de glicídios totais.....	61
Tabela 14: Resultados das figuras de mérito dos modelos de regressão por PLS, iPLS e siPLS para determinação do teor de glicídios redutores.....	62
Tabela 15: Resultados das figuras de mérito dos modelos de regressão por PLS, iPLS e siPLS para determinação do teor de glicídios não redutores.....	63
Tabela 16: Resultados das figuras de mérito dos modelos de regressão por PLS, iPLS e siPLS para determinação do teor de proteínas totais.....	64
Tabela 17: Resultados das figuras de mérito dos modelos de regressão, para determinação do teor de proteínas totais com dados MIR, FTIR-ATR.....	66

ABREVIATURAS

AA – Aminoácido

ANVISA – Agência Nacional de Vigilância Sanitária

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

BBS – Bebidas a Base de Soja

CLS – Quadrados Mínimo Clássicos (do inglês “*Classical Least Squares*”)

EHS – Extrato Hidrossolúvel de Soja

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

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

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

GA - . Algoritmo Genético (do inglês, “*Genetic Algorithm*”),

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

HCA – Análise de Agrupamentos Hierárquicos (do inglês “*Hierarchical Cluster Analysis*”)

ILS – Quadrados Mínimos Inversos (do inglês “*Inverse Least Squares*”)

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

IUPAC (do inglês, “*International Union of Pure and Applied Chemistry*”)

KNN – k-ésimo vizinho mais próximo (do Inglês “*K-Nearest Neighbor*”)

LDA – Análise Discriminante Linear (do inglês “*Linear Discriminant Analysis*”)

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

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

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

PCR – Regressão por Componentes Principais (do inglês “*Principal Component Regression*”)

PI – Proteína Isolada de soja

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

PLS-DA - Análise Discriminante por Mínimos Quadrados Parciais (do inglês “*Partial Least Squares Discriminant Analysis*”)

R^2 - Coeficiente de Determinação

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

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

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

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

SPA – Algoritmo das Projeções Sucessivas (do inglês, “*Successive Projections Algorithm*”)

UEV-PLS –Mínimos Quadrados Parciais com Eliminação de Variáveis não Informativas (do inglês, “*Partial Least Square with Uninformative Variable Elimintion*”),

VLs – Variáveis Latentes

RESUMO

Neste trabalho, foram estudadas estratégias para caracterização de bebidas à base de soja (BBS), por meio de análises por espectroscopia no infravermelho médio com transformada de Fourier com acessório de reflexão total atenuada (FTIR-ATR). Foram utilizadas 20 amostras comerciais de BBS, de 7 diferentes sabores e 3 diferentes marcas. Os teores estudados nas BBS foram glicídios totais, glicídios redutores, glicídios não redutores, e proteínas totais. Os modelos de regressão multivariada foram construídos por mínimos quadrados parciais (PLS), empregando como seleção de variáveis os métodos de mínimos quadrados parciais por intervalo (iPLS) e mínimos quadrados parciais por sinergismo de intervalos (siPLS). As seleções de variáveis por siPLS apresentaram os melhores resultados para os modelos construídos. Entre as propriedades avaliadas, a de glicídios totais apresentou modelos com erros de calibração e previsão (RMSECV e RMSEP) baixos, e coeficientes de determinação (R^2_{cv} e R^2_{prev}) próximos de um. Para proteínas totais, os modelos apresentaram resultados promissores, pois também tiveram erros de calibração e previsão (RMSECV e RMSEP) baixos, e coeficientes de determinação (R^2_{cv} e R^2_{prev}) próximos de um, considerando-se que as amostras reais e não apresentavam uma variabilidade de concentração de proteínas ideal. Para as propriedades de glicídios redutores e glicídios não redutores, não foram obtidos bons resultados para os modelos de regressão. Desta forma, a metodologia proposta apresenta potencial em análises de rotinas para determinação simultânea de glicídios totais e proteínas, atendendo aos requisitos referente às informações nutricionais na rotulagem das BBS, somando-se às vantagens da espectroscopia no infravermelho, tais como rapidez na análise, elevada frequência analítica, pequena quantidade de amostra necessária, baixo custo, não ser destrutiva e ser ambientalmente amigável.

Palavras chave: Soja. Bebidas à Base de Soja. Espectroscopia no Infravermelho. Análise Multivariada. Regressão por Mínimos Quadrados Parciais. Glicídios. Proteínas.

ABSTRACT

In this work, strategies were studied for the characterization of soy-based beverages (SBB), by means of Fourier transform infrared spectroscopy with attenuated total reflectance (FTIR-ATR). Twenty commercial samples of SBB were used, of 7 different flavors and 7 different brands. The contents studied in SBB were total sugar, reducing sugar, non-reducing sugars, and total proteins. The multivariate regression models were constructed by partial least squares (PLS), with evaluation of the methods by interval partial least squares (iPLS) and by synergy interval partial least squares (siPLS), for selection of variables. The selections of variables per siPLS presented the best results for the constructed models. Among the evaluated properties, the total sugar content presented models with low calibration and prediction errors (RMSECV and RMSEP), and determination coefficients (R^2_{cv} and R^2_{prev}) close to one. For total proteins, the models presented promising results, as they also had low calibration and prediction errors (RMSECV and RMSEP), and determination coefficients (R^2_{cv} and R^2_{prev}) close to one, considering that the actual samples did not present an ideal protein concentration variability. For the properties of reducing sugars and non-reducing sugars, good results were not obtained for the regression models. In this way, the proposed methodology presents potential in routine analysis for simultaneous determination of total glycogen and protein, taking into account the requirements referring to the nutritional information in the SBB labeling, adding to the advantages of the infrared spectroscopy, such as speed in the analysis, high analytical frequency, small amount of sample required, low cost, non destructive and environmentally friendly.

Key words: Soy. Soy based beverages. Infrared Spectroscopy. Multivariate analysis. Partial Least Square Regression. Sugars. Proteins.

1 INTRODUÇÃO

O crescimento do uso de técnicas e materiais que causem o menor ou nenhum impacto ambiental para diversas aplicações é uma consequência da conscientização da importância na preservação do meio ambiente¹. Ao desenvolver um método analítico, tão importante quanto o custo e o tempo de análise, outras características também devem ser consideradas, como as preocupações com a quantidade e toxicidade dos reagentes utilizados e os resíduos gerados². Nesse cenário, o interesse em análises físico-químicas, que sejam ambientalmente amigáveis, vem aumentando.

O principal meio de determinar a qualidade nutricional de alimentos é a determinação por componentes majoritários, conhecida como composição centesimal (fibra, cinza, lipídio, proteína, carboidrato e umidade). Ela é muito importante e fornece as informações nutricionais dos alimentos, presente nos rótulos dos mesmo. Para bebidas à base de soja (BBS), as análises abrangem as determinações de substâncias voláteis, lipídios, resíduos por incineração (cinzas), proteínas, glicídios e fibra alimentar³.

Para a determinação de componentes majoritários em BBS, são utilizados métodos tradicionais, como o termogravimétrico para os voláteis; o gravimétrico para gorduras, após a desnaturação da amostra; a incineração para cinzas; digestão, destilação e titulação para proteínas; digestão e titulação para glicídios e enzimático-gravimétrico para fibra alimentar³. No entanto, os métodos tradicionais de análise são demorados e utilizam materiais, equipamentos e diversos reagentes químicos, que além de oferecerem risco ao analista e ao meio ambiente, ainda geram resíduos tóxicos.

Uma alternativa para a análise físico-química de alimentos, como as bebidas à base de soja, de baixo custo, rápida, sem uso de reagentes químicos e sem preparo de amostra é a espectroscopia na região do infravermelho com transformada de Fourier (FTIR), com acessório de reflexão total atenuada (ATR) combinada com a análise multivariada por métodos de regressão por mínimos quadrados parciais. Encontram-se exemplos na literatura da utilização dessas técnicas combinadas para análises de alimentos, como determinação quantitativa de antioxidantes em produtos agrícolas⁴, medição do conteúdo de soja adicionada em carne moída⁵, detecção do teor de nitrogênio em molho de soja⁶, análise quantitativa da adulteração de lecitina com soja⁷.

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 bebidas à base de soja, constituindo-se em métodos mais rápidos, sem preparo de amostras, com baixo custo e que permitem a análise simultânea de propriedades, como glicídios e proteínas. Ambas propriedades são determinadas através de métodos de referência trabalhosos (no caso dos glicídios pelo método Lane-Eynon e nas proteínas o método Kjeldhal), onde se utilizam diversos materiais e reagentes, o que tornam os métodos demorados, caros e poluentes.

2 OBJETIVOS

2.1 GERAL

Desenvolver metodologias alternativas para a caracterização de bebidas à base de soja empregando técnicas de espectroscopia de infravermelho médio, em conjunto com ferramentas quimiométricas de regressão multivariada como PLS, iPLS e siPLS, atendendo à legislação referente às informações nutricionais na rotulagem das bebidas a base de soja.

2.1.1 Específicos

Construir e avaliar modelos de regressão, empregando métodos de seleção de variáveis (iPLS e siPLS) para determinar os teores de glicídios totais, glicídios redutores e glicídios não redutores em amostras de bebidas à base de soja;

Construir e avaliar modelos de regressão, empregando métodos de seleção de variáveis (iPLS e siPLS) para determinar os teores proteínas em amostras de bebidas à base de soja.

3 REVISÃO BIBLIOGRÁFICA

3.1 SOJA

A soja (*Glycine max (L.) Merrill*) é a semente oleaginosa mais produzida no mundo e consequentemente a oleaginosa de maior expressão econômica entre os países produtores e exportadores. Atualmente os Estados Unidos são os maiores produtores de soja, sendo o Brasil o segundo⁸.

O grão é um ingrediente essencial na produção de rações animais e na alimentação humana. A soja pode ser vista como um alimento completo, pois têm em sua composição, em média, proteínas (38%), carboidratos (27%), lipídios (19%), umidade (11%) e cinzas (5%), as quais contêm vitaminas, sais minerais e fibra⁹. Assim, ela pode gerar diversos produtos para alimentação humana, como soja cozida, farinha não desengordurada, manteiga de soja, alimentos dietéticos, flocos de soja e bebidas à base de soja, entre outros¹⁰. A soja também é uma fonte alternativa para a produção de biodiesel¹¹.

3.2 ALIMENTOS À BASE DE SOJA

Os alimentos à base de soja podem ser classificados em 4 grupos. O grupo dos produtos não desengordurados, nos quais estão incluídos o extrato de soja e a farinha torrada integral, contém todos os nutrientes da soja, podendo-se excluir a casca ou não. Já o grupo dos produtos desengordurados, como as farinhas e os farelos, resultam da trituração do grão e da remoção do óleo, ficando com cerca de 40 a 54% de proteínas, podendo ter granulação grossa, média ou fina. Nesse grupo, estão incluídos os produtos concentrados, os quais possuem, pelo menos 65% de proteínas, e os isolados, que são mais refinados e possuem mais de 90% de proteínas. O grupo dos produtos do óleo bruto, por sua vez, é constituído pelo óleo comestível, pelas margarinas e pela lecitina. E por fim, o grupo dos alimentos naturais, que pode ser fermentado, como o molho de soja, ou não fermentado, como o grão fresco verde, grão seco, leite de soja e tofu¹².

A soja rompeu os limites dos públicos vegetarianos ou alérgicos ao leite de vaca e sua presença na dieta aumentou muito nos últimos anos. Isso se deve a evolução tecnológica, que melhorou seu sabor e ampliou seu uso, fazendo parte de inúmeros alimentos industrializados¹². Hoje temos no comércio os mais variados alimentos à base de soja, desde

bebidas até uma linha completa de embutidos, iogurtes e doces, que buscam atrair todos os públicos, tanto os vegetarianos como os que buscam uma alimentação mais saudável.

Os fatores que promoveram o crescimento do uso da soja na indústria de alimentos foram a sua contribuição e a de seus derivados no benefício da saúde, o melhoramento do processo tecnológico da indústria para a sua utilização e o aumento da produção mundial desse grão¹².

As bebidas à base de soja (BBS) podem ser elaboradas a partir de dois produtos de soja, o extrato hidrossolúvel de soja (EHS) ou a proteína isolada de soja (PI).

A obtenção do EHS, se dá a partir da emulsão aquosa resultante da hidratação dos grãos de soja. Tal emulsão é obtida por meio do processamento tecnológico adequado, no qual é preservado o maior valor nutricional do grão, como proteínas, parte dos carboidratos solúveis, gorduras insaturadas, vitaminas e minerais. Para transformar o EHS e assemelhá-lo a um suco, adiciona-se aditivos e ingredientes, afim de dar diferentes cores e sabores ao extrato. Um fluxograma do processo é apresentado na Figura 1.

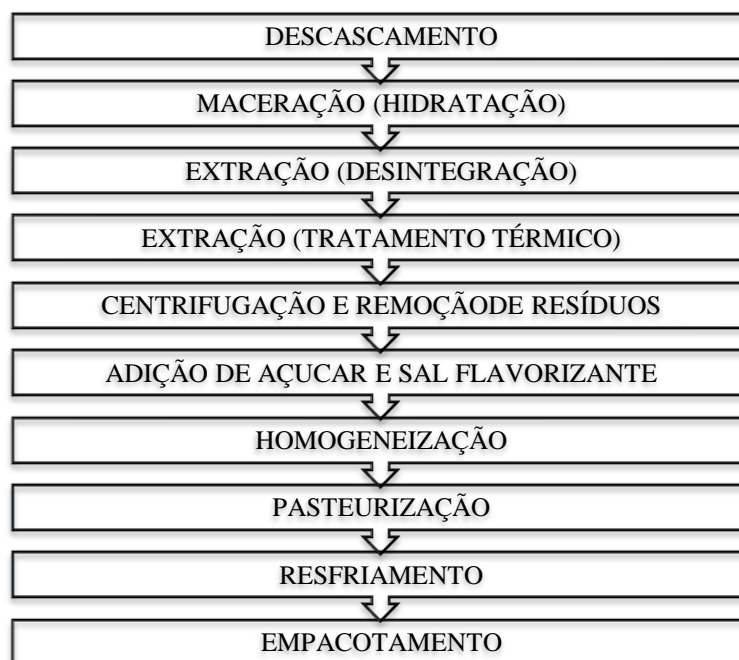


Figura 1: Fluxograma de produção de bebidas à base de soja, a partir do EHS da Campinas Vital¹³.

Uma etapa anterior ao descascamento é o processo de seleção do grão, em que a soja é separada, por meio de peneiramento de sujeiras. Neste etapa de seleção, também é importante retirar a soja quebrada, pois nela já ocorreram as reações enzimáticas responsáveis pelo sabor amargo e adstringente do extrato de soja. O descascamento é realizado por meio de moinhos de discos paralelos e de turbina de aspiração de cascas e sujeiras menores. A retirada da casca visa também melhorar o sabor do EHS, pois assim são eliminados os compostos adstringentes da casca, como sujeiras e pesticidas¹⁴.

A etapa em que o grão absorve água é a maceração. A maceração a quente auxilia na inativação de enzimas lipoxigenases, as quais são liberadas de forma parcial nessa etapa e total na trituração, a partir do rompimento das estruturas internas dos grãos, atuam sobre os ácidos graxos presentes no óleo de soja. Essas enzimas são as principais responsáveis pelo sabor amargo dos extratos de soja¹⁴.

A extração do extrato hidrossolúvel é realizada, numa primeira etapa, por meio da desintegração do grão, que envolve moagem ou trituração, seguido de adição de água quente, para extração de proteínas em uma segunda etapa. Para separar a massa de sólidos insolúveis do EHS, é utilizada a filtragem por centrifugação. A massa retirada do processo pode ser utilizada para preparo de outros alimentos. Após são adicionados os aditivos e ingredientes ao extrato sob agitação, para completa homogeneização. É nessa etapa que são adicionados açúcares ou adoçantes, base mascarante, antiespumante, aroma, corante, entre outros, quando a finalidade visa à produção de bebidas à base de soja¹⁴.

Na etapa de tratamento térmico é realizada a pasteurização. Nesta operação o extrato sofre aplicação de calor, por certo tempo, para inativar microrganismos patogênicos contidos no extrato. Após o tratamento térmico, o extrato deve ser pré-resfriado e resfriado, para continuar no processo até a etapa de embalagem¹⁴.

A proteína isolada da soja é extraída da farinha, ou do farelo, de soja desengordurada, obtida por remoção dos componentes não protéicos dos grãos. Contém aproximadamente 85 a 90% de proteínas em base seca e possui baixos teores de óleos, gorduras e carboidratos. Apesar de conter maior teor de proteínas do que o EHS, por ser concentrada, a PI é diluída para o consumo, adquirindo basicamente as mesmas propriedades do extrato hidrossolúvel; portanto o produto final, acaba oferecendo, por porção, quantidades de proteínas similares às do extrato¹⁵.

3.3 GLICÍDIOS

Os glicídios também conhecidos como açúcares, carboidratos ou hidratos de carbono, são substâncias orgânicas constituídas fundamentalmente por átomos de carbono, hidrogênio e oxigênio. Eles abrangem um dos grandes grupos de biomoléculas na natureza, além de serem a mais abundante fonte de energia. Quimicamente, os glicídios podem ser polihidroxiáldeídos, como a glicose, e polihidroxicetonas, como a frutose, ou substâncias que liberam esses por hidrólise¹⁶, como mostrado na Figura 2.

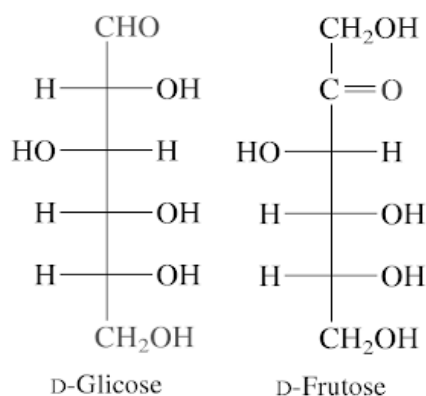


Figura 2: Representações do polihidroxiáldeído D-glicose e da polihidroxicetona D-frutose¹⁷.

Neste grupo de compostos, tem-se os mais variados tipos de substâncias, desde os monossacarídeos, representados pela glicose e frutose, os oligossacarídeos, onde se encontram os dissacarídeos, dos quais os mais frequentes em alimentos são a sacarose e a lactose, até os polissacarídeos, como amido e celulose¹⁶.

Os monossacarídeos são as unidades mais simples de carboidratos, são açúcares simples, não hidrolisáveis em unidades menores. Podem ser divididos quanto à função orgânica presente, cetose (função orgânica cetona) e aldose (função orgânica aldeído), e quanto ao número de átomos de carbono na cadeia, triose (3 átomos de carbonos), tetrose (4 átomos de carbono), pentose (5 átomos de carbono), hexose (6 átomos de carbonos)¹⁶.

Podem apresentar isomeria de função como a glicose (ceto-hexose) e a frutose (ldo-hexose). Como pode ser observado na Figura 2. Feita exceção a dihidroxicetona (ceto-triose), todos os outros monossacarídeos e por extensão, todos os outros carboidratos - possuem centros de assimetria (ou quirais), onde ocorre isomeria óptica. A Figura 3 mostra os enantiômeros do gliceraldeído¹⁷.

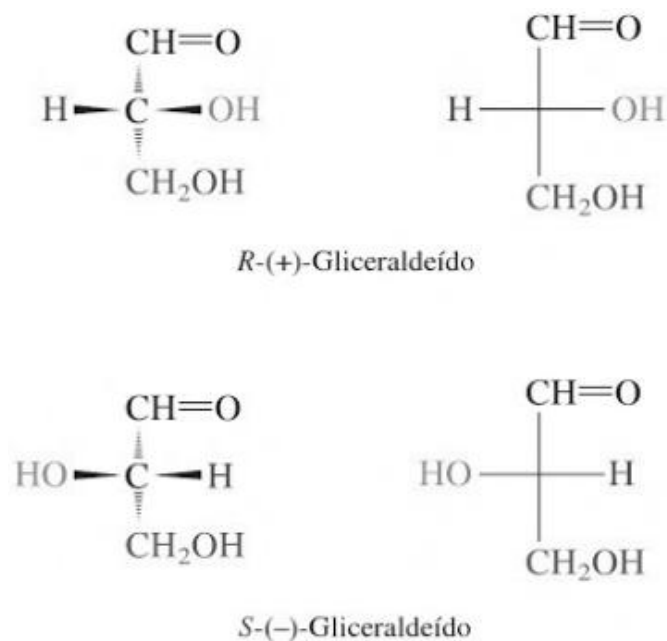


Figura 3: Representações tridimensionais e projeção de Fischer dos enantiômeros do gliceraldeído.¹⁷

Aumentando a complexidade das moléculas, como as tetroses, as pentoses e as hexoses, compostos que possuem mais de um centro de quiralidade, teremos 2^n estereoisômeros, onde n é o número de carbonos quirais da molécula¹⁷.

Os monossacarídeos, de 5 e 6 carbonos, em solução aquosa estão presentes na sua forma aberta em uma pequena proporção. O restante das moléculas está na forma de um anel hemiacetal, para aldeídos, ou hemicetal, para o caso de cetonas, de 5 ou de 6 vértices. Na estrutura do anel, o carbono onde ocorre a formação do hemiacetal, ou hemicetal, é denominado "carbono anomérico", e sua "hidroxila anomérica" pode assumir duas formas: alfa e beta. Anômeros são formas isoméricas de monossacarídeos que diferem entre si apenas na configuração ao redor do átomo de carbono anomérico, se a sua hidroxila está na posição alfa ou beta. A forma alfa acontece quando a hidroxila anomérica fica para "baixo do plano" do anel. A beta é quando ela fica para "cima do plano" do anel. A interconversão entre estas formas é dinâmica e denomina-se Mutarrotação¹⁷, como apresentado na Figura 4.

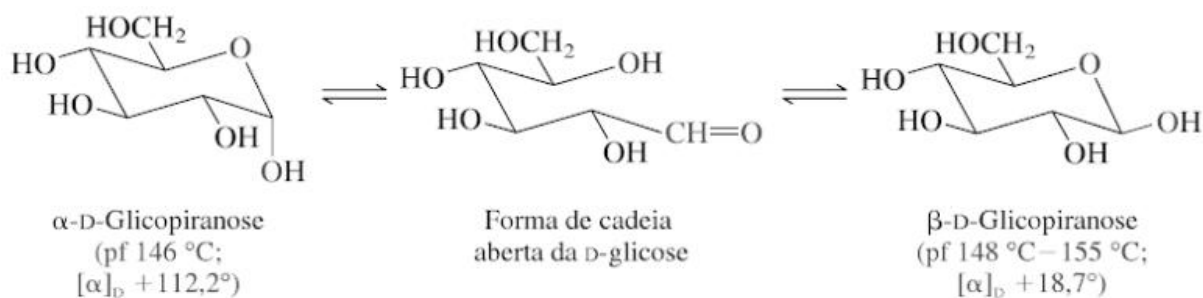


Figura 4: Representações da mutarotação entre a alfa glicose e a beta glicose¹⁷.

A hidroxila anomérica confere poder redutor aos monossacarídeos, pois permite que o grupo aldeído ou cetona, fique livre. Mas nem todos os dissacarídeos são redutores, para ser um açúcar redutor, necessitam ter uma hidroxila anomérica livre, ou seja, sem ligação glicosídica. A maltose é redutora, já a sacarose não¹⁷.

Oligossacarídeos são monossacarídeos unidos através da ligação glicosídica. Assim a International Union of Pure and Applied Chemistry (IUPAC) classifica os dissacarídeos, trissacarídeos e outros como subcategorias dos oligossacarídeos. A IUPAC não especifica o número de monossacarídeos que separa os oligossacarídeos dos polissacarídeos. Na prática o oligossacarídeo é homogêneo. Cada molécula de determinado oligossacarídeo tem o mesmo número de unidades de monossacarídeos. Já os polissacarídeos, são misturas de moléculas com comprimentos de cadeias parecidos, mas não iguais¹⁷.

Os oligossacarídeos mais comuns são os dissacarídeos, pois são formados a partir da ligação de dois monossacarídeos através de ligações especiais denominadas "ligações glicosídicas". A ligação glicosídica ocorre entre o carbono anomérico de um monossacarídeo com a hidroxila de qualquer carbono de outro monossacarídeo, produzindo uma molécula de água¹⁷.

O tipo de ligação glicosídica é definido pelos carbonos envolvidos e pelas configurações de suas hidroxilas. Se o carbono anomérico se ligar a hidroxila do carbono 4, por exemplo, (1-4) indica uma ligação entre o carbono C1 de uma molécula e o C4 da segunda. Além dessa designação, as ligações glicosídicas também são chamadas de alfa ou beta. Essa nomenclatura depende da configuração do carbono anomérico envolvido na ligação. Ou seja, se o carbono anomérico estiver na configuração alfa, a ligação é do tipo alfa, se estiver na configuração beta, a ligação é do tipo beta. A Figura 5 representa uma ligação glicosídica¹⁷.

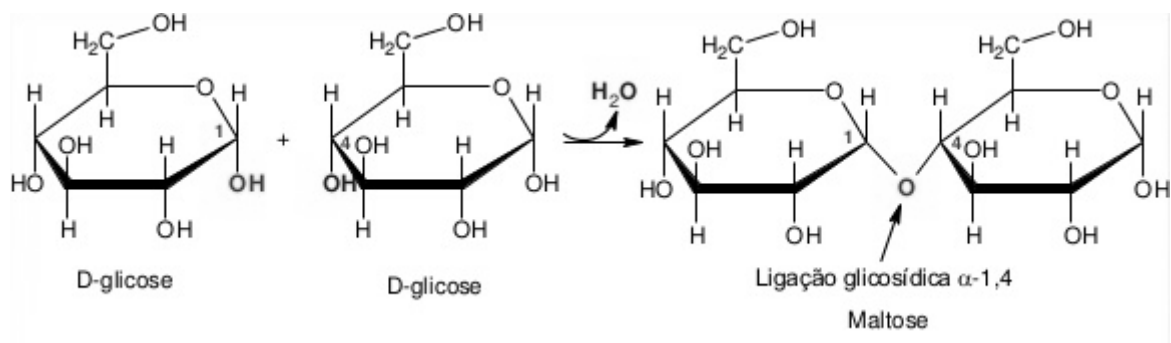


Figura 5: Representação da ligação glicosídica entre duas D-glicoses, originando uma α -1-4-maltose¹⁷.

Podemos observar a maltose não possui o grupo aldeído, mas, em solução, o grupo aldeído aparece por isomerismo, como mostrado na Figura 6, sendo a maltose um glicídio redutor.

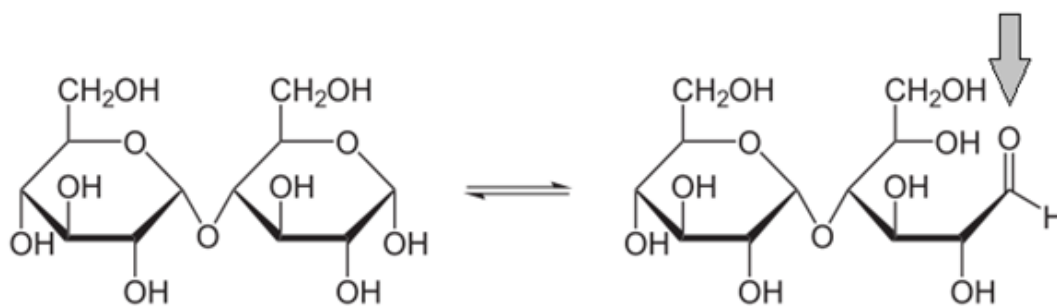


Figura 6: Representação da mutarrotação da maltose¹⁷.

A sacarose, Figura 7, é um exemplo de açúcar não redutor. Contudo, por hidrólise, produz dois açúcares redutores, a glicose e a frutose que possuem o grupo aldeído e cetona, respectivamente. Ela é um dissacarídeo no qual a D-glicose e a D-frutose são conectadas pelos seus carbonos anoméricos através de uma ligação glicosídica. Como a sacarose não tem um grupo hidroxila anomérico livre, ela não sofre mutarrotação. A hidrólise da sacarose catalisada por ácido produz uma mistura 1:1 de D-glicose e D-frutose¹⁷.

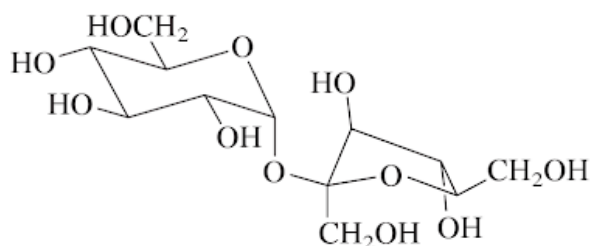


Figura 7: Representação da molécula de sacarose.¹⁷

Por fim, os polissacarídeos são monossacarídeos unidos através da ligação glicosídica, apresentando milhares de monossacarídeos. Eles podem ser de origem vegetal (celulose, amido e fibras) e animal (glicogênio). Podemos classificar os polissacarídeos em homopolissacarídeos, quando formado por um só tipo de monossacarídeo, por exemplo, a celulose, formada por ligações glicosídicas beta-1,4-D-glicose e heteropolissacarídeos, quando formado por mais de um tipo de monossacarídeo.¹⁶

3.4 PROTEÍNAS

A presença das proteínas é essencial para a vida. Elas constituem as biomoléculas mais diversas, complexas e de maior peso molecular.¹⁸

Os aminoácidos (AA) são os elementos que constituem as proteínas. Eles contêm carbono (C), hidrogênio (H), oxigênio (O) e nitrogênio (N), e alguns contêm enxofre (S). Cada AA é composto de um carbono central (chamado de carbono alfa), ligado a quatro grupos, um grupo amino (-NH₂), um grupo carboxílico (-COOH) e um hidrogênio (-H). O quarto grupo é variável e, de acordo com este, como apresentado na Figura 8. Vinte AA primários diferentes podem ser descritos¹⁸.

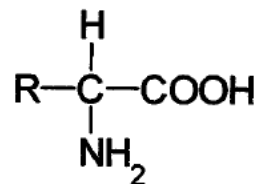


Figura 8: Representação da fórmula estrutural básica de um aminoácido.

Em termos gerais, são peptídeos quando a molécula é formada por até 100 AA (dipeptídios, tripeptídios, oligopeptídios, polipeptídios) e se trata de uma proteína quando a quantidade de aminoácidos é maior que 100. Em quase todas as proteínas, o conteúdo de nitrogênio é de cerca de 16% da massa total da molécula (com variações de 15,7%, no leite, a 19%, nas nozes), ou seja, cada 6,25 g de proteína contêm 1 g de nitrogênio. Por isso, o fator 6,25 é usado para calcular a quantidade de proteína numa amostra, levando-se em consideração o cálculo de nitrogênio em cada uma.¹⁸

A soja é uma das principais fontes de proteína de alta qualidade na alimentação de humanos e, comparada com alimentos de origem animal, contém menos ácidos graxos

saturados, além de ser livre de colesterol.¹⁹ É matéria-prima para o desenvolvimento, por processos de vários níveis de sofisticação tecnológica, de grande quantidade de alimentos.¹⁸

Os produtos baseados na soja, como farinha, grãos, leite de soja, proteína vegetal texturizada e mesmo suplementos nutricionais e fórmulas para bebês, são amplamente disponíveis para a população. Os grãos da soja correspondem a 35%-40% do peso seco de proteína²⁰. As proteínas da soja fornecem todos os aminoácidos essenciais para a nutrição humana, o que a caracteriza como uma fonte vegetal de proteína de alta qualidade.²¹

3.5 LEGISLAÇÃO ALIMENTAR

Não existe nenhuma legislação específica sobre as quantidades máximas e mínimas de proteínas ou glicídios em BBS. A única exigência, é a rotulagem nutricional de alimentos embalados, que visa proteger a saúde da população. Considerando que a rotulagem nutricional facilita ao consumidor conhecer as propriedades nutricionais dos alimentos, contribuindo para um consumo adequado dos mesmos.

Conforme a ANVISA, através da RDC Nº 360, DE 23 DE DEZEMBRO DE 2003, na rotulagem nutricional devem ser declarados os seguintes nutrientes, por porção de 100 g ou mL: valor energético, carboidratos(glicídios), proteínas, gorduras totais, gorduras saturadas, gorduras trans e sódio. O valor energético deve ser expresso em, quilocaloria (Kcal) ou quilojoule (KJ), o sódio em miligramas (mg) e os demais nutrientes em gramas (g)²².

Para a expressão dos valores, os nutrientes serão declarados de acordo com o estabelecido na Tabela 1 e as cifras deverão ser expressas nas unidades indicadas.

Tabela 1: Expressão dos valores a serem declarados nos rótulos²².

Valores	Expressões dos valores
Maiores ou igual 100	Declarar em números inteiros com três cifras
Menores que 100 e maiores ou iguais a 10	Declarar em números inteiros com duas cifras
Menores que 10 e maiores ou iguais a 1	Declarar com uma cifra decimal
Menores que 1	Declarar com uma cifra decimal

A informação nutricional será expressa como “zero” ou “0” ou “não contém” para valor energético e ou nutrientes quando o alimento contiver quantidades menores ou iguais as estabelecidas como “não significativas” de acordo com a Tabela 2:

Tabela 2: Quantidades não significativas por porção²².

Valor energético/nutriente	Quantidades não significativas por porção (em g ou mL)
Valore energético	Menor ou igual a 4 Kcal ou 17 KJ
Carboidratos (glicídios)	Menor ou igual a 0,5 g
Proteínas	Menor ou igual a 0,5 g
Gorduras totais	Menor ou igual a 0,5 g
Gorduras saturadas	Menor ou igual a 0,2 g
Gorduras trans	Menor ou igual a 0,2 g
Fibra alimentar	Menor ou igual a 0,5 g
Sódio	Menor ou igual a 0,5 mg

3.6 ESPECTROSCOPIA NO INFRAVERMELHO

Nosso campo de estudo é a espectroscopia no infravermelho (IR), que é uma ferramenta polivalente aplicada às determinações quantitativas e qualitativas de muitas espécies orgânicas. A radiação infravermelha corresponde aproximadamente à parte do espectro eletromagnético situada entre a região do visível e das microondas²³.

O espectro na região do infravermelho é habitualmente dividido em infravermelho próximo (NIR), infravermelho médio (MIR) e infravermelho distante (FIR). A Tabela 3 apresenta os intervalos correspondentes a cada uma dessas regiões.

Tabela 3: Regiões espectrais do infravermelho²³.

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 possui energia suficiente para produzir transições eletrônicas, assim, para que uma molécula absorva a radiação infravermelha ela precisa ser capaz de apresentar uma variação no momento de dipolo durante seu movimento vibracional ou rotacional. 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 funcionais presentes nas moléculas em determinados comprimentos de onda²³.

As posições relativas dos átomos não estão exatamente fixas em uma molécula, elas mudam continuamente em função de inúmeros tipos de rotações e vibrações das ligações da molécula. As vibrações são classificadas como deformação e estiramento. As deformações são caracterizadas 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 plano. Já o estiramento abrange uma variação contínua na distância interatômica ao longo do eixo da ligação entre dois átomos, podendo ser simétrico e assimétrico²³, como ilustrados na Figura 9.

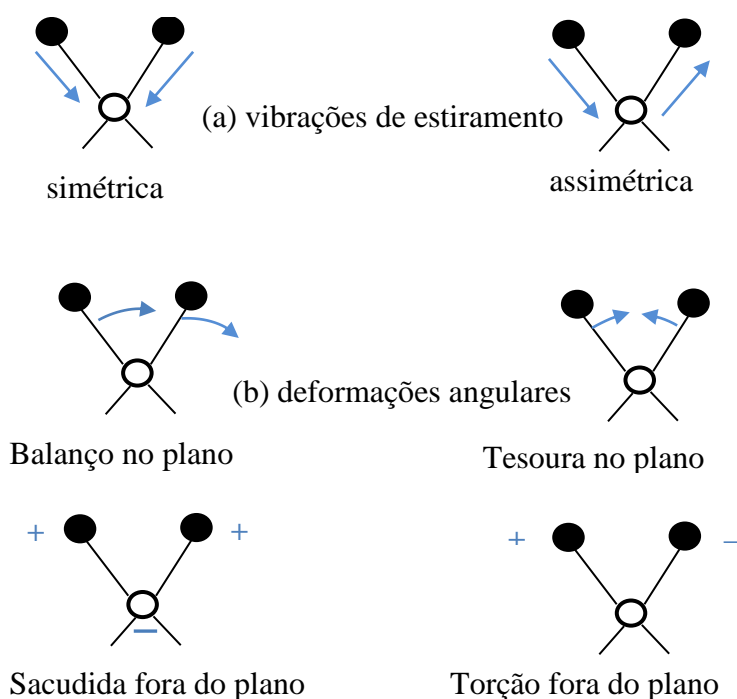


Figura 9: Representação das vibrações de (a) estiramento e (b) deformações angulares.²³

Os espectros obtidos na região do MIR apresentam picos bem definidos e de alta absorvâncias, com bandas intensas, facilitando a interpretação dos mesmos. A faixa espectral do MIR, compreendida entre $1.200-600\text{ cm}^{-1}$, conhecida como região de impressão digital é

bastante útil para identificação de compostos, uma vez que pequenas diferenças na estrutura e na constituição das funções orgânicas das moléculas resultam em mudanças na distribuição e no perfil dos picos de absorção. Deste modo, é possível identificar, por exemplo, um composto orgânico, comparando-se seu espectro com banco de dados pré-existentes.²³

A espectroscopia no IR é uma das técnicas mais promissoras no campo de análises não destrutivas, permitindo a obtenção de resultados rápidos e sem pré-tratamento da amostra. A aplicação do IR para fins analíticos foi complementada pelos avanços da eletrônica e também pelos ocorridos na área de tratamento de dados, hoje conhecida, entre os químicos, como quimiometria.²⁴

Os métodos mais empregados para obtenção dos espectros no infravermelho, para fins analíticos, envolvem as técnicas de transmissão ou a de reflexão. No primeiro caso, a radiação passa através da amostra, onde parte da radiação é absorvida e outra transmitida. Para isso, a amostra deve ser colocada em uma janela cujo material seja transparente na região utilizada do IR e compatível com a amostra. Para o segundo caso, as informações espectrais são adquiridas por meio da reflexão da radiação incidente, por três diferentes técnicas: reflexão especular, reflexão difusa e reflexão total atenuada (ATR)^{25,26}.

3.6.1 Espectroscopia por Reflexão Total Atenuada (ATR)

A espectroscopia por reflexão total atenuada é uma técnica robusta e segura para estudos qualitativos e quantitativos, em amostras líquidas. Destaca-se por não envolver processos laboriosos de preparo das amostras²⁵.

O princípio da técnica de espectroscopia por ATR está nas múltiplas reflexões da radiação infravermelha, que ocorrem no interior de cristais, com alto índice de refração como, por exemplo, o seleneto de zinco (ZnSe), para que somente uma pequena parte do feixe incidente seja refletido, interagindo apenas com a amostra que estiver na superfície do cristal.^{27,28}

Um acessório de reflexão total atenuada mede as mudanças que ocorrem em um feixe de radiação infravermelha, o qual sofre reflexão interna ao entrar em contato com a amostra. Embora ocorra a reflexão, parte da radiação penetra no meio menos denso, acima da superfície do cristal de ATR, interagindo com o material absorvente (amostra)²⁷.

A profundidade de penetração (dp , do inglês *depth of penetration*) é dada pela Equação 1:

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

onde λ é o comprimento de onda da radiação, θ_i é o ângulo de incidência, n_1 é o índice de refração do elemento de ATR e n_2 índice de refração da amostra.

A onda evanescente é a radiação que penetra a amostra. Se houver uma absorção em determinado comprimento de onda evanescente, ocorre uma atenuação do feixe corresponde às bandas de absorção no infravermelho. É de acordo com esse mecanismo que surge o espectro²⁷. Na Figura 10, está representado um esquema simplificado da análise por ATR.

A técnica de ATR tem avançado em análises de amostras sólidas e líquidas, porque soluciona os pontos mais desafiadores da análise no infravermelho, como a reprodutibilidade espectral e o preparo de amostras. Sua principal característica se destaca por utilizar pequena quantidade de amostra e por não envolver processos laboriosos no preparo desta, ou, ainda, não envolver nenhum preparo, podendo-se utilizar a amostra in natura²⁹.

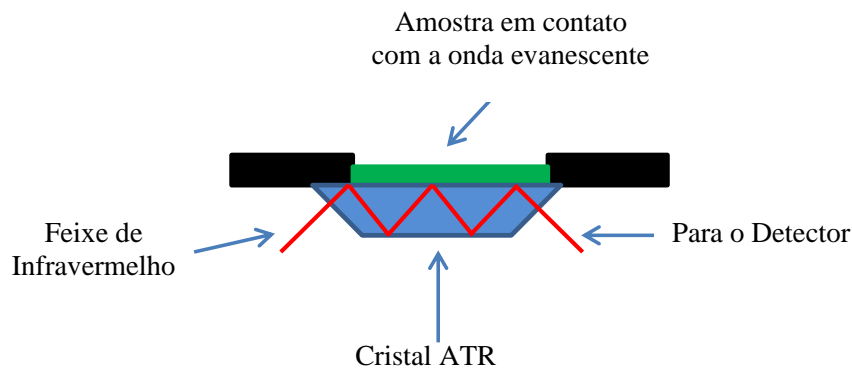


Figura 10: Esquema simplificada da análise por ATR³⁰.

3.7 QUIMIOMETRIA

A quimiometria pode ser definida como uma área que combina métodos estatísticos e matemáticos para potencializar as informações, referente à uma propriedade de interesse, que podem ser extraídas de um conjunto de dados químicos. Os métodos quimiométricos, são

empregados em dados multivariados e podem ser supervisionados, como os métodos PLS, CLS, PCR e ILS, que possuem propósito quantitativo (calibração multivariada), bem como os métodos LDA, k-NN e PLS-DA, com propósitos qualitativos (análise exploratória e reconhecimento de padrões) ou métodos não supervisionados, como os métodos qualitativos PCA e HCA.^{31,32,33,34}

3.7.1 Calibração Multivariada

A calibração multivariada utiliza medidas realizadas em múltiplos canais, para a construção do modelo que relaciona o sinal digital com a concentração. Esse método aparece como uma alternativa capaz de vencer as limitações das análises univariadas, permitindo determinações simultâneas de analitos, na presença de interferentes, com maior sensibilidade e confiabilidade. Para esse fim, os interferentes devem, também, estar presentes na etapa de calibração, o que é denominado como vantagem de primeira ordem³⁵.

Os modelos de calibração multivariada (calibração de primeira ordem) são construídos com a informação analítica relacionada a cada amostra correspondente a um vetor. Os dados químicos obtidos em uma análise, por exemplo, dados de espectros no infravermelho para um conjunto de amostras, são dispostos em uma matriz, para serem modelados. A matriz de dados é composta por linhas e colunas, onde cada linha corresponde a uma amostra e cada coluna contém a informação referente a um canal analítico das amostras, que, no exemplo de espectros, são os comprimentos de onda. Por convenção, a matriz de variáveis independentes é denominada de matriz \mathbf{X} e usualmente chamada de matriz de respostas. O vetor contendo a variável dependente ou parâmetro de referência é denominado de matriz \mathbf{y} ³⁶.

Uma das principais ferramentas de análise utilizada para a elaboração de modelos na calibração multivariada é a técnica por mínimos quadrados parciais (PLS, do inglês *Partial Least Squares*). Esse método permitiu significativo progresso na aplicação da quimiometria³⁷.

3.7.2 Regressão por Mínimos Quadrados Parciais (PLS)

O método PLS é o mais utilizado na atualidade para análises quantitativas em dados multivariados³⁸. Os métodos de regressão são empregados para verificar a relação entre uma variável \mathbf{y} (que é a propriedade de interesse, como, por exemplo, a concentração) e um conjunto de variáveis \mathbf{X} (sinais instrumentais)³⁹.

O método PLS é robusto, permitindo que os parâmetros do modelo não se modifiquem significativamente quando novas amostras são adicionadas ou subtraídas do conjunto de calibração³⁹. Além de ser um algoritmo simples e com excelente poder de previsão⁴⁰.

O princípio da regressão por mínimos quadrados parciais é a decomposição de matriz de dados \mathbf{X} (sinais instrumentais) em uma soma de várias matrizes \mathbf{M}_i , que tem dimensionalidade um, mais uma matriz de erros (\mathbf{E}), que corresponde à parte não modelada da matriz \mathbf{X} . As matrizes \mathbf{M}_i formam as variáveis latentes (VLs) e são formadas pelo produto de dois vetores \mathbf{t} e \mathbf{p} , sendo respectivamente os escores e os pesos^{27,40}, conforme é apresentado nas equações 2 e 3. No PLS, a matriz \mathbf{Y} , também é representada pelos escores e pelos pesos⁴⁰, conforme a equação 4.

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \dots + \mathbf{M}_a + \mathbf{E} \text{ ou } \mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_a\mathbf{p}_a^T + \mathbf{E} \quad (2)$$

$$\mathbf{X} = \sum \mathbf{t}_a\mathbf{p}_a^T + \mathbf{E} \text{ ou } \mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (3)$$

$$\mathbf{Y} = \sum \mathbf{u}_a\mathbf{q}_a^T + \mathbf{F} \text{ ou } \mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (4)$$

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

A dimensão do espaço original é igual ao número de colunas em \mathbf{X} , ou seja, o número de variáveis originais. No novo modelo, a dimensionalidade é descrita pelo número de matrizes \mathbf{M}_i necessárias para descrever a matriz \mathbf{X} . Assim, se for possível descrever uma matriz \mathbf{X} que tenha muitas variáveis, por um número menor dessas matrizes \mathbf{M}_i , haverá um decréscimo na dimensionalidade, sem perda da informação estudada⁴⁰.

A relação entre as duas matrizes de dados, \mathbf{X} e \mathbf{Y} , é obtida com a correlação dos escores de cada uma, para obter uma relação linear de formas equivalentes, descrita na equação 5.

$$\mathbf{y} = \mathbf{bX} + \mathbf{e} \quad (5)$$

Os valores de \mathbf{b} correspondem aos coeficientes de regressão entre a matriz de escores \mathbf{U} de \mathbf{Y} e a matriz de escores \mathbf{T} de \mathbf{X} , assim como \mathbf{e} contém o que não foi modelado denominado de resíduo do modelo.

Para se alcançar o melhor modelo referente aos dados examinados deve ser avaliado o número de variáveis latentes (VLs) e a Raiz Quadrada dos Erros Médios (RMSE, do inglês, *Root Mean Square Error*).

3.7.3 Métodos de Seleção de Variáveis em Modelos de Regressão

Os métodos espectroscópicos podem originar um grande número de variáveis, deixando o processo de regressão complexo e demorado. Além disso, algumas variáveis contêm informações que podem ser colineares, ou seja, contêm as mesmas informações, sendo, portanto, redundantes, ou ainda, podem conter o ruído. Também existem regiões espectrais que não variam os valores de absorbância, mesmo com a variação das concentrações de um determinado composto⁴¹.

A habilidade de previsão do modelo de calibração pode ser afetado pelo excesso de variáveis, podendo ter seu desempenho diminuído. Em compensação, a escolha das regiões espectrais ou de variáveis específicas pode aprimorar a eficiência dos modelos de calibração. As vantagens da seleção espectral adequada estão relacionadas com a estabilidade do modelo de calibração, em relação à colinearidade, e também com produção de um modelo mais robusto, simples de interpretar e com menores erros de previsão^{41,42,43}.

Há vários procedimentos para a seleção de regiões espectrais. Diversos trabalhos vêm utilizando métodos como: o Algoritmo das Projeções Sucessivas (SPA, do inglês “*Successive Projections Algorithm*”), utilizado para determinação de ácidos orgânicos em vinagre de ameixa⁴⁴. Algoritmo Genético (GA, do inglês “*Genetical Algorithm*”), aplicado na otimização da codigestão anaeróbica de resíduos de babta e plantas aquáticas⁴⁵. Método de Eliminação de Variáveis não Informativas por Mínimos Quadrados Parciais (UVE – PLS, do inglês “*Uninformative Variable Elimination by Partial Least Square*”), utilizado para a determinação de ácidos alfa-linolênico e linoléico em óleos comestíveis⁴⁶. Os métodos 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*), utilizados para a quantificação de benzoilmetronidazol em insumos farmacêuticos⁴⁷.

No iPLS inicialmente é informado o número que intervalos equidistantes que se deseja dividir os espectros obtidos. O iPLS emprega o algoritmo PLS, no qual é efetuada uma regressão por mínimos quadrados parciais em cada intervalo equidistante ao longo de todo o

espectro. Dessa forma um modelo de PLS é construído para cada intervalo, bem como um para o espectro global, apresentando os resultados na forma de gráfico que contém o raiz quadrada dos erros médios (RMSE, do inglês, “*Root Mean Square Error*”) obtido para cada modelo (intervalo), o que facilita a comparação com toda a faixa espectral^{48,49}. Assim, é possível avaliar a importância da informação em cada uma das subdivisões espectrais, sendo possível selecionar o intervalo cujas variáveis exibem as informações mais relevantes para a regressão desejada.

O siPLS é uma extensão do algoritmo iPLS. Esse algoritmo constitui-se na divisão do espectro em regiões equidistantes, seguido da combinação desses intervalos, possibilitando a obtenção de modelos que podem apresentar melhores coeficientes de determinação (R^2), menores erros de calibração e previsão do que aqueles encontrados empregando o iPLS⁵⁰, pois pode combinar informações relevantes de diferentes intervalos que não são vizinhos, as quais podem não são contempladas pelo iPLS.

Neste trabalho foram utilizados os métodos PLS, iPLS e siPLS, sendo estes dois últimos para realizar a seleção de variáveis, sendo que esses modelos selecionam faixas espectrais e não variáveis isoladas, como também por se tratarem de métodos mais simples, em relação aos demais modelos citados.

3.7.4 Seleção do Conjunto de Calibração e Previsão

A escolha das amostras que irão constituir os conjuntos de calibração e previsão é realizada através do algoritmo Kennard-Stone⁵¹. A metodologia desse algoritmo consiste em selecionar a primeira amostra, como a que apresentar a maior distância em relação à amostra média. A segunda amostra a ser selecionada será a que exibir maior distância em relação à primeira. A próxima amostra a ser designada apresentará maior distância em relação à última amostra selecionada, e assim sucessivamente até atingir o número de amostras almejado⁵². Esse algoritmo é aplicado para formar o conjunto de calibração, e por consequência eleger as amostras de maior variabilidade, ou seja, as amostras mais “extremas” do conjunto total. Como consequência as amostras não selecionadas serão empregadas para compor o conjunto de previsão externa.

3.7.5 Pré-tratamento dos Dados

Variáveis com diferentes variâncias e medidas com diferentes unidades, podem afetar a disposição e uniformidade dos dados espectrais. Essa circunstância pode gerar dificuldades na análise dos dados, acometendo à extração de informações relevantes, bem como suas interpretações⁵³. Assim, existe a necessidade de tratar os dados, que serão modelados, pelo emprego de pré-tratamentos.

O pré-tratamento tem como propósito reduzir as variações indesejáveis, de forma matemática, que não foram extintas no decorrer da obtenção dos dados e que não são eliminados naturalmente durante a análise, mas que podem interferir nos resultados finais. O tratamento dos dados é aplicado na etapa de calibração para refinar a eficiência do método, resultando num modelo de regressão com melhor capacidade de previsão. O tratamento prévio também desempenha a função de remover variações sistemáticas não desejadas do espectro, como o efeito de espalhamento, mudanças na linha de base e fatores não controláveis³². Contudo, qualquer tipo de tratamento deve ser realizado criteriosamente, para não comprometer, distorcer ou até mesmo inutilizar os resultados da análise.

Encontram-se vários métodos matemáticas comuns que podem ser empregados com o propósito de reduzir as variações aleatórias, como as informações sistemáticas indesejáveis. Existem dois tipos de pré-tratamento: transformação, que é aplicado às linhas da matriz de dados (as amostras), e pré-processamento, empregado às colunas de X (as variáveis)³².

O alisamento, do inglês *Smoothing*, é um método de transformação e tem grande aplicação na área espectroscópica para redução da componente aleatória (ruído) dos dados e o consequente aumento da razão sinal ruído do sinal analítico⁵⁰. Também pode ser empregado para evidenciar pequenos picos que estão cobertos na linha de base pelo o ruído e, ainda, como um método de compressão de dados³².

Utilizou-se neste trabalho o método de Savitzky-Golay⁵⁴ para a realização do alisamento. Essa ferramenta ajusta um polinômio de baixa ordem aos pontos de uma janela pelos mínimos quadrados. É muito importante a escolha pelo número de pontos, pois um número alto pode provocar perda de informação e um número reduzido, a permanência de ruído.

A normalização, se caracteriza pela divisão dos valores de cada uma das variáveis de uma dada amostra, por um fator de normalização, como a norma dessa amostra. Deste modo, todas as amostras estarão em uma escala pré-determinada³².

A finalidade dessa normalização é equiparar a magnitude de cada amostra. Ela desloca a informação da distância de cada amostra à origem dos dados, mas mantém a direção. Essa transformação é recomendada quando se quer somente a informação que qualitativamente discrimina uma amostra da outra e remove toda a informação que poderia distinguir duas amostra de composição idêntica, mas com intensidades diferentes para o sinal analítico, que pode ter ocorrido pela diferença de quantidade de amostra utilizada ou mudança da intensidade instrumental oriunda de fatores aleatórios como temperatura e pressão³².

Outro pré-tratamento utilizado foi a técnica de correção de linha base. Um problema instrumental, de amostragem, ou até devido à estrutura das partículas, pode distanciar 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 aplicando-se a primeira derivada do espectro. Como a primeira derivada de uma constante é zero, após a correção o espectro resultante estará deslocado para zero de absorvância. Dessa forma, as informações apresentadas ao longo dos diferentes comprimentos de onda são acentuadas. Entretanto, a principal desvantagem desse processo é que o ruído, uma vez presente nos espectros originais, pode se acentuar, diminuindo a relação sinal/ruído. Esse problema pode ser amenizado, através da aplicação de um alisamento antes da derivação. Por isso se deve ter todo cuidado com a qualidade dos espectros quando se pretende aplicar o cálculo das derivadas,^{26,32}.

Já o pré-processamento é executado quando se deseja comparar variáveis com diferentes dimensões. Consiste, principalmente, em centrar os dados na média ou autoescalar os dados. Centrar os dados na média significa em calcular a média das intensidades para cada comprimento de onda e subtraí-los do respectivo valor médio. Desse modo, ocorre a mudança do sistema de coordenadas para o centro dos dados, com o deslocamento da informação descrita pelos dados originais, para o centro. Não ocorre distorção da informação disponível, não havendo alteração da informação para o modelo.

O pré-processamento, autoescalar os dados, é baseado em 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 o mesmo peso, não

importando se eles representam um pico, um espalhamento ou um ruído de linha de base, por consequência não se torna aconselhável para dados espectrais^{32,55}.

3.7.6 Avaliação dos Modelos de Regressão

A análise de alguns parâmetros, oriundos da construção dos modelos, proporciona a escolha e a seleção de modelos mais adequados. Uma questão importante da etapa de validação de uma regressão multivariada é sobre a determinação da dimensionalidade do modelo³³.

Um fator relevante, para a qualidade do modelo quando se utiliza uma técnica de calibração baseada na regressão linear, como é o PLS, é a determinação do número de variáveis latentes (VLs). Para cada modelo existe um número ótimo de fatores, com dados suficientes para o modelamento. Pode ocorrer casos onde se tenha um número de VLs inferior ao ideal ocorrendo o subajuste, como também casos de um modelo com número superior de VLs ao ideal, ocorrendo um sobreajuste⁵⁶.

Com um subajuste, se terá um modelo que não avaliará a totalidade da informação contida no conjunto de dados, o número de fatores é considerado muito abaixo do ideal e o erro de previsão será alto. A sentença sobre um número de VLs superior ao ideal é o que ocorre com mais frequência, gerando a construção de modelos de regressão com resultados satisfatórios, mas sobreajustado, em outras palavras, com baixo valor de erro na determinação da propriedade de interesse da amostras do conjunto de calibração, mas com elevado valor de erro na previsão da propriedade de interesse de amostras de um conjunto de validação ou previsão. Isso é explicado pois 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^{32,34}.

Na figura 11 é apresentado um gráfico para demonstrar o comportamento do erro de previsão em modelos tendenciosos. Ela apresenta um esquema de como as contribuições das informações contidas nas VLs, devem ser harmonizadas para que ele apresente um balanço adequado entre a variância e a tendência, estando otimizado para fazer previsões³².

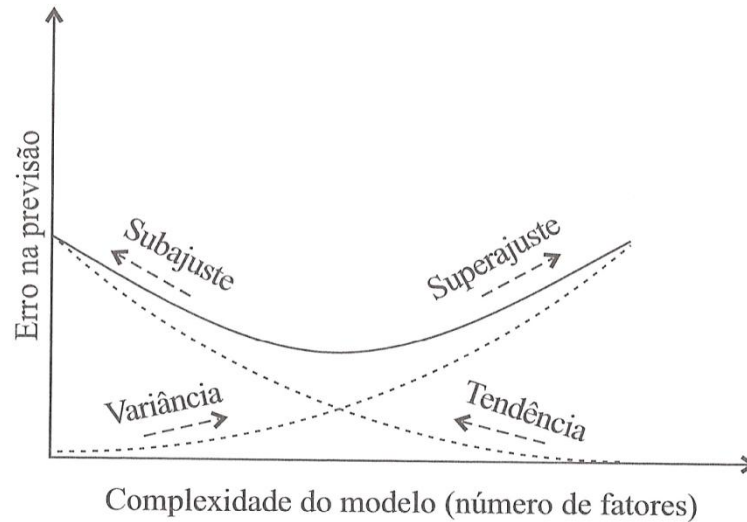


Figura 11: A linha cheia indica o comportamento de erro na previsão. As linhas tracejadas indicam o comportamento da variância e tendência (viés) nos modelos PLS em geral³².

A definição do número de VLs utilizadas no modelo, para que não ocorra sub ou sobreajuste, é realizada através da validação do modelo com um conjunto teste e validação externa. Os valores da propriedade de interesse são estimados e comparados com os valores reais calculando-se os resíduos $(y_i - \hat{y}_i)$, que são uteis para a avaliação da qualidade e confiabilidade dos modelos de regressão, através dos erros estimados pela raiz quadrada do erro médio de calibração cruzada (RMSECV, do inglês *Root Mean Square Error of Cross Validation*) e raiz quadrada do erro médio de previsão (RMSEP, do inglês *Root Mean Square Error of Prediction*). A prática geral é utilizar o gráfico dos valores de RMSECV e RMSEP, versus o número de fatores no modelo para determinar o número ótimo de fatores^{32,35}.

O cálculo dos erros RMSECV e RMSEP são dados conforme as equações (6) e (7), respectivamente

$$RMSECV = \sqrt{\frac{\sum_{i=1}^I (y_i - \hat{y}_i)^2}{I}} \quad (6)$$

onde, y_i é o valor de referência para a i -ésima amostra do conjunto de calibração cruzada e \hat{y}_i é o valor de previsão para esta mesma amostra, sendo I o numero total de amostras, do conjunto de calibração cruzada, utilizadas para a construção do modelo³².

$$RMSEP = \sqrt{\frac{\sum_{p=1}^P (y_p - \hat{y}_p)^2}{P}} \quad (7)$$

onde, y_p é o valor de referência para a i -ésima amostra do conjunto teste e \hat{y}_p é o valor de previsão para esta mesma amostra, sendo P o número total de amostras, do conjunto de previsão externa, utilizadas para a avaliação do modelo³².

A validação cruzada está fundamentada na avaliação da grandeza do erro de previsão de um dado modelo de regressão e baseia-se na remoção de uma amostra do conjunto de calibração, e subsequente construção do modelo de regressão. As amostras retiradas são previstas no modelo e calcula-se o erro. Esse mecanismo acontece até que todas as amostras do modelo sejam removidas e previstas, e então, calcula-se o RMSECV, conforme equação 6. O procedimento é repetido utilizando n número de VLs ($n = 1, 2, 3, \dots n$), sendo escolhido o número de VLs no qual se obtém o menor erro RMSECV, que não sobreajuste o modelo³³.

Por meio do cálculo da RMSEP, conforme equação 7, é executada a avaliação da capacidade de previsão do modelo de calibração para as amostras de previsão. Para isso, são utilizadas amostras independentes daquelas utilizadas para construção do modelo de calibração, permitindo avaliar a capacidade do modelo de prever amostras desconhecidas.

Outro fator relevante é o coeficiente de determinação (R^2), que é uma metodologia estatística utilizada na avaliação de modelos de regressão. Ele aponta o quão bem o modelo de regressão construído, quanto mais próximo de 1 (entre 0 e 1) for o seu valor, mais explicativo é o modelo³⁴.

Consequentemente, o modelo ideal de regressão é aquele que possui valor R^2 mais próximo de 1, menor número possível de VLs, com valores, de RMSECV e de RMSEP, o mais baixo possível e de mesma magnitude, pois diferenças significativas entre RMSECV e RMSEP indicam problemas de subajuste ou sobreajuste.

Existem autores que defendem que os parâmetros para a avaliação dos modelos de regressão, devem ser visto com cuidado, pois o R^2 mostra somente qual a fração de variância é explicada pelo modelo, e dependendo do parâmetro, é aceitável valores não tão próximos de 1. Defendem também que o erro relativo, deve ser levado mais em consideração que os RMSECV e RMSEP, pois o mesmo da uma melhor dimensão do erro do modelo³⁶.

4 MATERIAIS E MÉTODOS

4.1 AMOSTRAS

Foram adquiridas, no comércio local, 20 amostras de bebidas de soja, de diferentes fornecedores e sabores, representando mais de 80% dos produtos do gênero vendidos no comércio local de Porto Alegre - RS. A partir de cada uma das amostras, foram preparadas mais 3 amostras com adição crescente de água de 5%, 10% e 20%, totalizando 80 amostras. A adição de água tem o objetivo de diminuir a concentração dos terores estudados, para ampliar a curva de calibração, e detectar também possíveis adulterações de bebidas à base de soja, por adição de água.

4.2 ESPECTROSCOPIA DE INFRAVERMELHO MÉDIO

Os espectros das amostras BBS foram obtidos num espectrofotômetro de infravermelho com transformada de Fourier (FTIR) Perkin Elmer Spectrum 400, acoplado a um acessório de refletância total atenuada (ATR) com cristal de seleneto de zinco (ZnSe), a temperatura ambiente, com resolução de 2 cm^{-1} e 16 varreduras na região do infravermelho médio, na faixa espectral de $1899\text{ a }900\text{ cm}^{-1}$, em triplicata.

4.3 ANÁLISE MULTIVARIADA

A análise multivariada das amostras de bebidas à base de soja foi realizada a partir dos dados de FTIR-ATR, e com esses dados, após pré tratamento, construídos modelos de regressão multivariada por PLS, iPLS e siPLS para as concentrações de glicídios totais, glicídios redutores, glicídios não redutores e proteínas totais.

Como pré-tratamento, os espectros foram alisados (empregando Savitsky-Golay com janela de 13 pontos, polinômio de primeira ordem), normalizados (entre 0-1), aplicada a primeira derivada (com janela de 13 pontos) e realizada a média das triplicatas utilizando o software ChemoStat.⁵⁷ Já para a construção dos modelos de regressão foram utilizados os algoritmos de regressão por mínimos quadrados parciais (PLS), regressão por mínimos quadrados parciais por intervalos (iPLS) e por sinergismo de intervalos (siPLS), empregando o pacote iToolbox em ambiente Matlab 8.1 (MathWorksInc., Natick, MA). Os conjuntos de

calibração e previsão externa foram selecionados pelo emprego do algoritmo Kennard-Stone⁵¹, como primeira opção.

4.3.1 Regressão multivariada dos teores de glicídios totais, glicídios redutores, glicídios não redutores e proteínas totais a partir dos dados de FTIR-ATR.

Considerando o total de 80 amostras de BBS, inicialmente, como primeira opção, 60 amostras foram designadas para a construção dos modelos de calibração e as 20 amostras restantes foram utilizadas para avaliar a predição dos modelos. A seleção das amostras de calibração e previsão foram realizadas pela o algoritmo de Kennard-Stone⁵¹. Posteriormente, somente para a determinação de proteínas totais, considerando-se que as amostras selecionadas para os conjuntos de calibração e validação não apresentavam uma variabilidade de concentração de proteínas ideal foram selecionadas, aleatoriamente, 54 amostras para a construção de modelos de calibração e 26 para os modelos de previsão.

Os modelos de regressão por PLS, iPLS e siPLS, foram construídos a partir dos dados de FTIR-ATR. O PLS foi aplicado no espectro inteiro, usando a totalidade da informação (faixa espectral de 1899 a 900 cm^{-1}). Os modelos iPLS foram construídos realizando 2, 4 e 8 divisões equidistantes do espectro. Escolheu-se apenas intervalos, de acordo com os menores valores de RMSECV, do coeficiente de determinação de calibração cruzada (R^2_{cv}) mais próximo de um e o menor número de variáveis latentes (VLs), para cada modelo de regressão, considerando regiões espectrais mais relevantes e com melhores resultados. O modelo PLS funciona igualmente o iPLS, mas é aplicado no espectro como um todo, sem divisões de intervalos. Os modelos de regressão por siPLS foram obtidos realizando 2, 3 ou 4 combinações, de 8 intervalos. Escolheu-se as combinações que obteve o modelo com os menores valores de RMSECV e de variáveis latentes e valores R^2_{cv} mais próximos de um, para cada modelo de regressão, de cada combinação de intervalos. Também considerou-se as regiões que possuíam informação espectral mais relevante.

A determinação do número de variáveis latentes, para cada modelo de regressão, foi determinada pela método da validação cruzada, *leave-one-out*, elegendo o menor número de VLs com que os modelos de regressão apresentem menor RMSECV. Havendo todo cuidado para que não ocorresse nenhum sobreajuste ou um subajuste.

Após a modelagem de calibração, afim de verificar a predição dos modelos que apresentaram melhor resultado, foi produzido os modelos de previsão para cada modelagem. Também buscou-se os menores valores de RMSEP, de coeficiente de determinação de predição (R^2_{prev}) e de VLs.

Após a elaboração dos modelos de regressão por iPLS e siPLS, o melhor modelo, de cada regressão para cada intervalo espectral selecionado, foi estabelecido como o que possui menores valores de RMSECV e RMSEP, desejando que não existam grandes diferenças entre os erros, para não ocorrer um subajuste ou sobreajuste. O número de VLs, também foi levado em consideração, onde o ideal é a utilização do menor numero possível, assim como os maiores valores de R^2_{cv} e R^2_{prev} , desde que, sempre se tenha cuidado para não sub ou sobreajustar o modelo.

4.4 TEORES DE GLICÍDIOS TOTAIS, GLICÍDIOS REDUTORES, GLICÍDIOS NÃO REDUTORES E PROTEÍNA.

A metodologia utilizada para determinação de glicídios (açúcares) em bebidas, foi a do procedimento titulométrico denominado Método Lane-Eynon. O princípio do método baseia-se na reação dos açúcares redutores com os íons cúpricos da solução de Fehling, reduzindo-se a íons cuprosos, sob a ação do calor em meio alcalino. Ao reagir com os íons cúpricos, os açúcares sofrem oxidação, enquanto que o Cu(II) é reduzido a Cu(I), formando-se um precipitado vermelho de óxido cuproso. Os glicídios não redutores devem sofrer uma prévia hidrólise em meio ácido, dissociando o dissacarídeo em seus monossacarídeos.

Para a realização das análises dos teores de glicídios, inicialmente é necessário realizar a padronização da solução de Fehling. Primeiramente foi preparada uma solução padrão de glicose 1%, essa solução é preparada todo dia, com glicose seca a temperatura de 80°C por 1 hora e esfriada em dessecador. Após é realizada a padronização da solução de Fehling onde, se transfere para um erlenmeyer 10mL de solução A de Fehling (34,639g de sulfato de cobre penta hidratado em um litro de água), 10mL de solução B de Fehling (173g de tartarato de sódio e potássio tetra hidratado, 250mL de hidróxido de sódio 20% e completado volume até um litro) e 40mL de água. A mistura é aquecida até fervura e titulada, pela solução padrão de glicose 1% até clarear. Após é adicionado 1 gota de azul de metileno 1%, e continuada a titulação, até o ponto de virada, terminando a titulação³.

O cálculo do título da solução de Fehling é dado pela expressão:

$$T_{Fehling} = \frac{V_{Glicose\ 1\%} \times m}{100} \quad (8)$$

onde, $T_{Fehling}$ é o título da solução de Fehling, $V_{Glicose\ 1\%}$ é o volume de solução padrão de glicose gasto na titulação em mililitros e m é a massa de glicose em gramas.

O aquecimento da solução é fator crucial na análise, se a fervura não for permanente durante a titulação, o ar oxida o azul de metileno, mantendo a coloração azul, o que facilmente induz a erro quanto ao ponto final da titulação. O ponto de viragem do indicador azul de metileno, muda a cor da solução de azul para incolor, mas como ocorre a formação de um precipitado de cor tijolo (óxido cuproso), a cor visível da viragem é de azul para vermelho tijolo. A adição do indicador azul de metileno é feita somente próximo ao ponto de viragem. O descoramento é quase instantâneo e a ação do indicador é reversível, já que a adição de traços de sais de cobre restaura sua cor imediatamente. A titulação do teor de glicídios deve ser realizada em até 3 minutos (desde o início do gotejamento do filtrado até o descoramento do indicador).

A determinação do teor de glicídios redutores nas BBS, foi realizada por método titulométrico, a partir do procedimento tradicional *Lane-Eynon* (038/IV, Instituto Adolf Lutz)³, que usa como referência o método 958.06 da *American Association of Cereal Chemists* (1995). Inicialmente a amostra, in natura, é pesada de 10 a 25 g de amostra, com precisão de 0,0001 g, em balão volumétrico de 100 mL, e adicionados 5 mL de ferrocianeto de potássio 6% (m/v), 5mL de acetato de zinco 12% (m/v), completado volume e homogeneizado. A mistura foi deixada sedimentando por 15min e após filtrada. O filtrado foi transferido para uma bureta. Em um erlenmeyer foram adicionados 5ml de Solução A de Fehling e igual volume de Solução B de Fehling, mais 40 mL de água destilada. Essa solução é aquecida até fervura. Em seguida foi titulado sob fervura, pelo filtrado, até clarear. Após é adicionado 1 gota de azul de metileno 1%, e continuada a titulação, até o ponto de viragem, terminando a titulação. A análise foi realizada em triplicata.

O cálculo é dado pela expressão:

$$\%G_{Redutores} = \frac{100 \times 100 \times T_{Fehling}}{V \times m} \quad (9)$$

onde, $\%G_{Redutores}$ é o teor de glicídios redutores, expresso em % (m/m) de glicose, $T_{Fehling}$ é o título da solução de Fehling, V é o volume gasto na titulação em mililitros e m é a massa da amostra em gramas.

A determinação do teor de glicídios não-redutores e de glicídios totais, nas BBS foram realizadas pelo método de referência *Lane-Eynon* (respectivamente 039/IV e 040/IV, Instituto Adolf Lutz)³, que usam como referência o método 958.06 da *American Association of Cereal Chemists* (1995). Aproveitando o filtrado dos glicídios redutores, foi transferido 20 ml do mesmo e adicionado 1 mL de ácido clorídrico concentrado e deixado em banho Maria a 100°C por 1 hora. Após esfriar a solução é neutralizada com hidróxido de sódio e transferida, quantitativamente, para um balão volumétrico de 100 mL e completado o volume. O filtrado tratado foi transferido para uma bureta. Em um erlenmeyer foram adicionados 5 ml de cada Solução A e B de Fehling mais 40 mL de água destilada e aquecida até fervura. Em seguida foi realizada a titulação sob fervura até clarear. Após é adicionado 1 gota de azul de metileno 1%, e continuada a titulação, até o ponto de virada, terminando a titulação. A análise também foi realizada em triplicata.

Os cálculos são dados pelas expressões:

$$\%G_{Totais} = \frac{100 \times 100 \times T_{Fehling}}{V_1 \times m_1} \quad (10)$$

onde, $\%G_{Totais}$ é o teor de glicídios totais, expresso em % (m/m) de glicose, $T_{Fehling}$ é o título da solução de Fehling, V_1 é o volume gasto na titulação em mililitros e m_1 é a massa da amostra em gramas, na alíquota.

$$\%G_{N\tilde{a}o-redutores} = (\%G_{Totais} - \%G_{Redutores}) \times 0,95 \quad (11)$$

onde, $\%G_{N\tilde{a}o-redutores}$ é o teor de glicídios não-redutores, expresso em % (m/m) de sacarose, $\%G_{Totais}$ é o teor de glicídios totais, expresso em % (m/m) de glicose, $\%G_{Redutores}$ é o teor de glicídios redutores, expresso em % (m/m) de glicose e 0,95 é o fator de conversão de glicose para sacarose.

A determinação de proteínas é baseada na determinação de nitrogênio total. Geralmente é realizado pelo método de digestão Kjeldahl, que é um métodos idealizado em 1883 e tem sofrido inúmeras modificações e adaptações, porém sempre baseado em três etapas, digestão, destilação e titulação³.

Na digestão a matéria orgânica existente na amostra é decomposta com ácido sulfúrico e um catalisador, onde o nitrogênio é transformado em sal amoniacal. Já na destilação a amônia é liberada do sal amoniacal pela reação com hidróxido e recolhida numa solução ácida de volume e concentração conhecidas. Por ultimo, titula-se o excesso do ácido utilizado na destilação com hidróxido e se determina a quantidade de nitrogênio presente na amostra.

O teor de nitrogênio das diferentes proteínas é aproximadamente 16%, por isso, é adotado o fator empírico 6,25 para transformar o numero de g de nitrogênio encontrado em numero de g de protídeos. Em alguns casos, emprega-se um fator diferenciado de 6,25, conforme descrito na Tabela 4.

Tabela 4: Fatores de conversão de nitrogênio total em proteína³.

Alimento	Fator	Alimento	Fator
Farinha de centeio	5,83	Castanha do Pará	5,46
Farinha de trigo	5,83	Avelã	5,30
Macarrão	5,70	Coco	5,30
Cevada	5,83	Outras nozes	5,30
Aveia	5,83	Leite e derivados	6,38
Amendoim	5,46	Margarina	6,38
Soja	6,25	Gelatina	5,55
Arroz	5,95	Outros alimentos	6,25
Amêndoas	5,18		

A determinação do teor de proteína das amostras empregadas neste estudo foi realizada no Laboratório de Nutrição Animal (LNA) da Faculdade de Agronomia da UFRGS, pelo método Kjeldahl, a partir do procedimento (037/IV, Instituto Adolf Lutz)³, que usa como referência o método 991.20 da *American Association of Cereal Chemists* (1995).

5 RESULTADOS E DISCUSSÃO

5.1 TEORES DE GLICÍDIOS TOTAIS, GLICÍDIOS REDUTORES, GLICÍDIOS NÃO REDUTORES E PROTEÍNA TOTAIS.

Os resultados da determinação de glicídios redutores, glicídios totais e glicídios não redutores realizada pelo método de referência *Lane-Eynon* (respectivamente 038/IV, 039/IV e 040/IV, Instituto Adolf Lutz)³, que usam como referência o método 958.06 da *American Association of Cereal Chemists* (1995), para as 80 amostras de BBS analisadas, são apresentados no Apêndice 8.1 (Tabelas 9, 10, 11). As análises foram realizadas em triplicatas, e realizado sua média. Os teores de glicídios redutores, totais e não redutores, nas BBS, variam na faixa de 0,45 – 3,80%, 1,00 - 7,92% e 0,04 – 3,91% respectivamente. Esses resultados foram utilizados como dados de referência para a construção e avaliação dos modelos de regressão multivariada.

Os resultados da determinação de proteína (N), realizada pelo método de referência Kjeldahl, são apresentados no Apêndice 8.1 (Tabela 12). As análises também foram realizadas em triplicatas, e realizadas suas médias. O teor de proteína nas amostras variam na faixa de 0,35 até 2,54% (m/m). As análises foram realizadas pelo Laboratório de Nutrição Animal da Faculdade de Agronomia da Universidade Federal do Rio Grande do Sul. Como anteriormente, esses resultados foram utilizados como dados de referência para a construção e avaliação dos modelos de regressão multivariada.

5.2 ESPECTROSCOPIA DE INFRAVERMELHO

Os espectros de MIR foram obtidos em triplicata, na região de 1899 – 900 cm^{-1} , a partir das amostras de BBS. A Figura 12a, representa os espectros sem tratamento dos dados e a Figura 12b, representa os espectros após receberem o tratamento de suavização, normalização, derivação e média das triplicatas. É possível verificar diferenças entre os espectros tratados e não tratados.

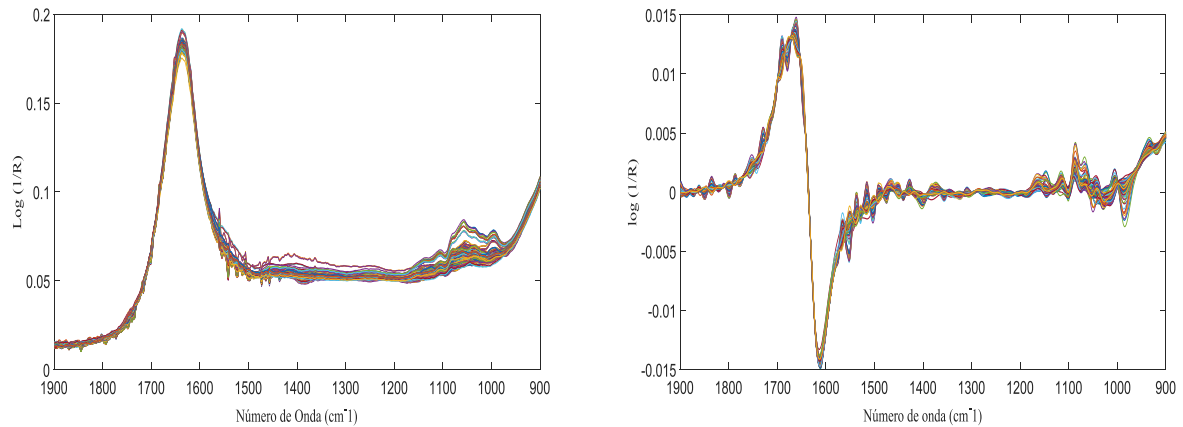


Figura 12: Espectros das bebidas à base de soja a) espectro sem tratamentos dos dados; b) espectro com tratamento dos dados, espectro suavizado, normalizado, derivados e realizadas suas médias.

Os dados espectrais foram utilizados separadamente para na construção de cada regressão para a determinação de cada analito proposto, com o objetivo selecionar regiões relevantes e de construir modelos melhores para cada, ou seja, com menores erros e maiores coeficientes de determinação.

5.3 ANÁLISE MULTIVARIADA

5.3.1 Determinação do Teor de Glicídios Totais por FTIR-ATR Combinada com Regressão Multivariada.

Foram construídos modelos de regressão por PLS, iPLS e siPLS a partir do conjunto de dados, pré tratados, obtidos por MIR, FTIR-ATR.

Por meio da Tabela 5, é possível verificar os resultados de RMSECV, RMSEP, R^2_{cv} , R^2_{prev} e número de variáveis latentes (VLs) dos melhores modelos desenvolvidos, de cada regressão para cada intervalo espectral selecionado, para a determinação de glicídios totais. Nela é possível verificar as principais figuras de mérito para construir os modelos de regressão multivariada. O erro relativo, para o conjunto das amostras de previsão, também é apresentado, para se ter uma melhor visualização do baixo erro de previsão do modelo, para a determinação dos glicídios totais. A totalidade dos modelos gerados, que foram avaliados, são apresentadas no Apêndice 8.2 (Tabela 13).

Os modelos foram escolhidos pelos menores valores de RMSECV e RMSEP e maiores coeficientes de correlação (R^2). Para escolher o número de VLs avaliamos o erro dos modelos de calibração interna cruzada RMSECV, tendo cuidado para escolher um número que não causasse subajuste nem superajuste dos modelos. Para avaliar a capacidade de previsão entre os modelos, foi utilizado o erro do conjunto de previsão externa RMSEP.⁵⁰

Tabela 5: Resultados das figuras de mérito dos melhores modelos de regressão, para determinação do teor de glicídios totais com dados MIR, FTIR-ATR.

Modelos ^a	Intervalos selecionados ^b	VLs ^c	RMSECV (% m/m)	R^2_{cv}	RMSEP (% m/m)	R^2_{prev}	Erro relativo de previsão (% m/m)
PLS	Global	9	0,18	0,9821	0,12	0,9492	3,23
iPLS2	2	7	0,17	0,9839	0,13	0,9320	3,25
iPLS4	4	9	0,18	0,9803	0,17	0,9012	3,95
iPLS8	7	3	0,24	0,9663	0,19	0,8887	4,06
s2iPLS8	6, 7	6	0,20	0,9779	0,13	0,9337	3,44
s3iPLS8	2, 5, 7	7	0,19	0,9785	0,13	0,9341	3,45
s4iPLS8	2, 3, 6, 7	7	0,19	0,9783	0,14	0,9239	3,48

^a s—número de intervalos combinados iPLS—número de divisões do espectro; ^b Intervalos selecionados; ^c número de variáveis latentes; Formação em negrito aos melhores modelos de regressão.

A partir dos resultados das regressões usando os métodos iPLS e siPLS pode-se definir que os modelos mais adequados para a determinação de glicídios totais, expressos em % (m/m) de glicídios totais em glicose, foram o PLS, iPLS2 e o S2iPLS8. Com exceção dos modelos iPLS4 e iPLS8, que tiveram menores R^2_{prev} , e os maiores RMPSEP, entre os modelos gerado, foi possível verificar, que os outros modelos apresentaram resultados muito semelhantes, principalmente entre os modelos siPLS, assim qualquer um, desses modelos, pode ser utilizado para a determinação de glicídios totais. O modelo s2iPLS8 foi selecionado, por ser o mais simples e por utilizara uma faixa mais específica do analito.

Para o modelo PLS, o melhor resultado obtido apresentou um RMSECV de 0,18% para 9 VLs e R^2_{cv} de 0,9821. Para as amostras de previsão, o modelo PLS obteve RMSEP de 0,12% e R^2_{prev} de 0,9492.

O melhor modelo iPLS, foi o iPLS2, que apresentou menor valor de RMSECV quando comparado ao modelo PLS, mas um valor maior de RMSEP. O modelo construído, com o intervalo 2, que corresponde a faixa de $1399 - 900 \text{ cm}^{-1}$, para o espectro dividido em 2 partes,

produziu os melhores valores, obtendo um RMSECV de 0,17% para 7 VLs e R^2_{cv} de 0,9839. Para as amostras de previsão, o modelo iPLS2 obteve RMSEP de 0,13% e R^2_{prev} de 0,9320. A região espectral selecionada engloba os estiramentos axiais C-O característico dos glicídios, que são polialcoóis, de $1260 - 1050 \text{ cm}^{-1}$.⁵⁸

O modelo siPLS escolhido, o qual será exemplificado e detalhado, foi o s2iPLS8, que dividiu o espectro em 8 partes e combinou os subintervalos selecionados 6 e 7, que correspondem aos intervalos espectrais $1275 - 1151 \text{ cm}^{-1}$, $1150 - 1025 \text{ cm}^{-1}$, com 6 VLs, conforme ilustrado na Figura 13. Apresentou RMSECV = 0,20%, RMSEP = 0,13% e coeficientes de determinação (R^2_{cv} e R^2_{prev}) de 0,9779 e 0,9337, respectivamente. Os intervalos selecionados, englobam os estiramentos axiais C-O característicos dos glicídios, de $1260 - 1050 \text{ cm}^{-1}$,⁵⁸ indicando que esta é uma faixa representativa na determinação dos glicídios totais nas amostras de bebidas a base de soja, ao selecionar uma região específica de absorção.

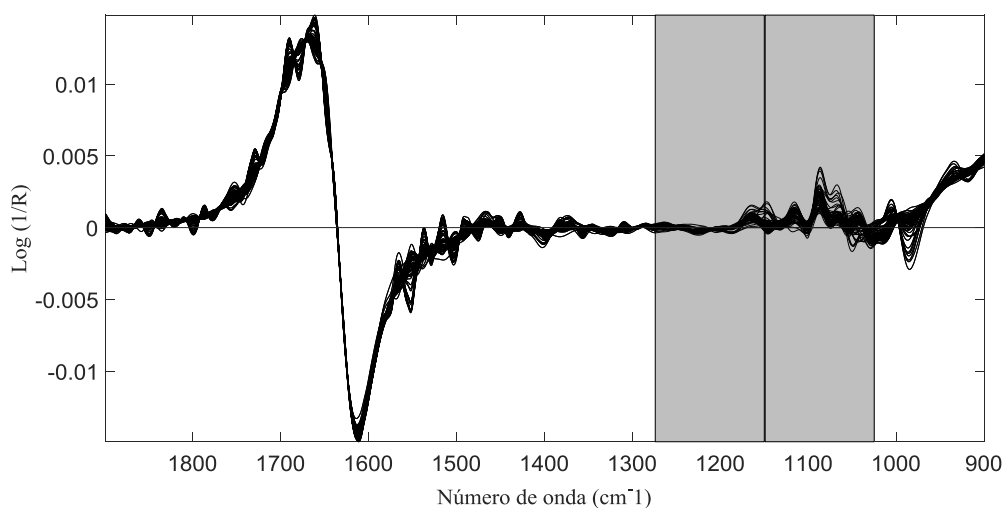


Figura 13: Espectro com destaque as regiões selecionadas (6, 7) combinadas com regressão siPLS para determinação de glicídios totais em BBS.

A definição do número de variáveis latentes é de fundamental importância para que não ocorra o sobreajuste ou subajuste, sendo pela Figura 14 possível justificar a escolha pelo número de 6 VLs do modelo s2iPLS8, possuindo menor erro de calibração, determinado por validação cruzada (RMSECV), e a partir desse número de variáveis não ocorre decréscimo significativo do erro. Vale ressaltar que a matriz de análise, a BBS, é bastante complexa, tendo a presença de diferentes substâncias como proteínas, vitaminas, gorduras, fibras, metais,

aromatizantes, corantes e glicídios como frutose, glicose, sacarose, gomas entre outros carboidratos, o que pode elevar o número de VLs do modelo.

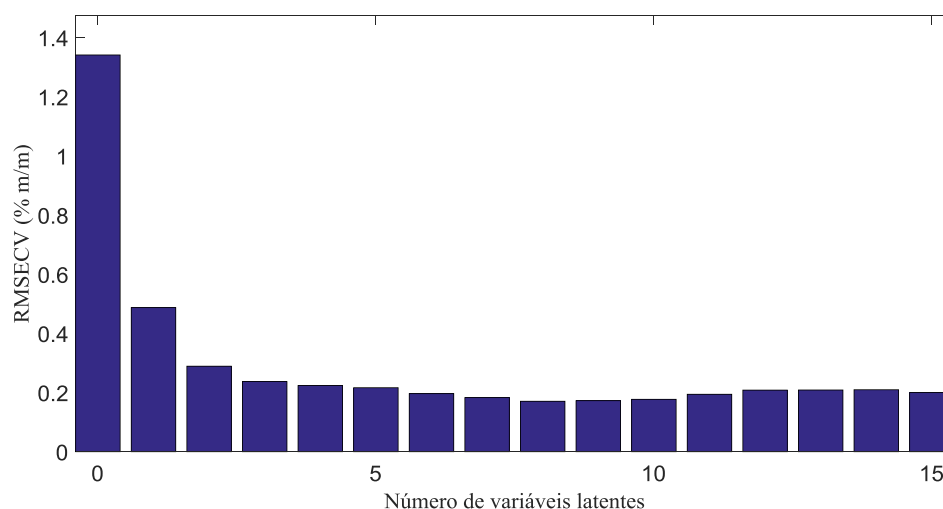


Figura 14: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de glicídios totais, por meio de espectroscopia no infravermelho, combinando os intervalos 6, 7 no modelo s2iPLS8.

Também foi construído um gráfico de correlação dos valores de referência, obtidos pelo método oficial para o teor de glicídios totais, com os valores obtidos pelo melhor modelo de regressão através do FTIR-ATR conforme ilustrado na Figura 15. 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 do teor de glicídios totais em amostras de BBS.

Os resultados obtidos, se mostram satisfatórios para a determinação de glicídios totais, para a rotulagem nutricional das BBS, uma vez que, os modelos apresentaram um baixo erro e a legislação vigente, não requer um baixo limite de detecção nem uma declaração mais detalhada dos valores dos nutrientes listados nas embalagens.

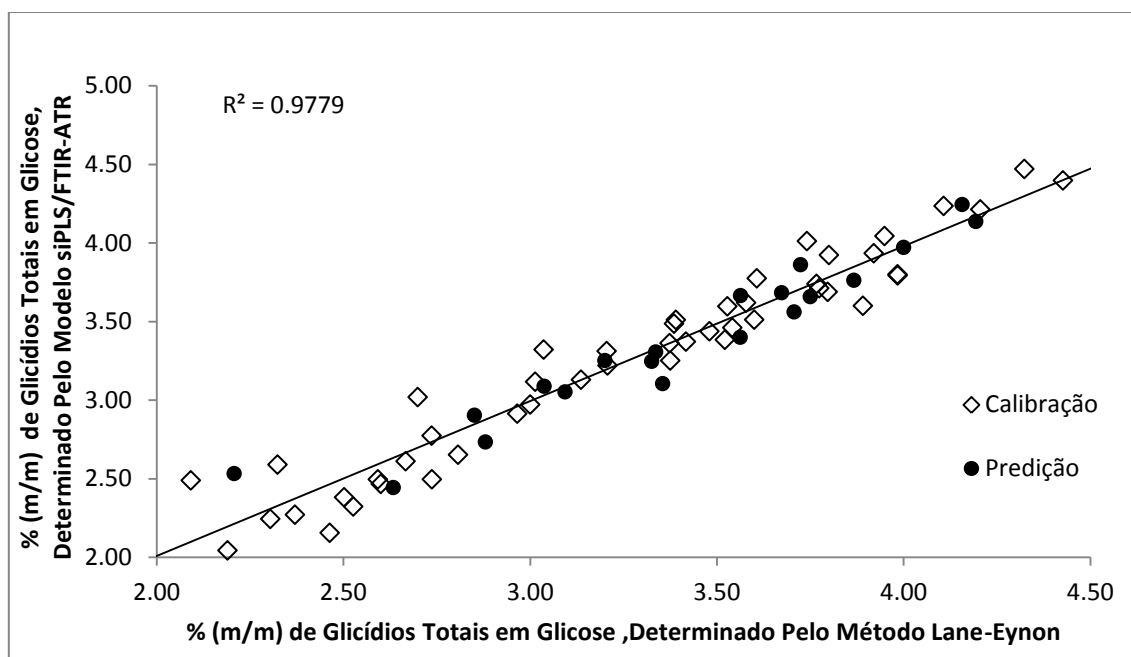


Figura 15: Valores de referência pelo método Lane-Eynon *versus* valores previstos do modelo siPLS para glicídios totais usando os intervalos 6 e 7 correspondentes aos intervalos espectrais $1025 - 1150 \text{ cm}^{-1}$ e $1151 - 1476 \text{ cm}^{-1}$.

5.3.2 Determinação do Teor de Glicídios Redutores por FTIR-ATR Combinada com Regressão Multivariada.

Como no primeiro caso, foram construídos modelos de regressão por PLS, iPLS e siPLS a partir do conjunto de dados, pré tratados, obtidos por FTIR-ATR.

Por meio da Tabela 6, também verificamos os resultados de RMSECV, RMSEP, R^2_{cv} , R^2_{prev} e número de variáveis latentes (VLs) dos melhores modelos desenvolvidos, de cada regressão para cada intervalo espectral selecionado, para a determinação de glicídios redutores. Nela, é possível verificar as principais figuras de mérito para construir os modelos de regressão multivariada. O erro relativo, para o conjunto das amostras de previsão, também é apresentado, para ter uma melhor visualização do erro do modelo. A totalidade dos modelos gerados, que foram avaliados, são apresentadas no Apêndice 8.3 (Tabela 14).

O critério de escolha dos modelos apresentados, foram os mesmos da determinação de glicídios totais, menores valores de RMSECV e RMSEP e maiores coeficientes de correlação

(R^2), assim como a avaliação do erro dos modelos de calibração cruzada, RMSECV, para determinação do número de VLs. Como no primeiro caso, para avaliar a capacidade de previsão entre os modelos, utilizamos o erro do conjunto de previsão externa RMSEP.⁵⁰

Tabela 6: Resultados das figuras de mérito dos melhores modelos de regressão, para determinação do teor de glicídios redutores com dados FTIR-ATR.

Modelos ^a	Intervalos selecionados ^b	VLs ^c	RMSECV (% m/m)	R^2_{cv}	RMSEP (% m/m)	R^2_{prev}	Erro relativo de previsão (% m/m)
PLS	Global	7	0,54	0,7242	0,44	0,6862	14,26
iPLS2	1	10	0,56	0,7091	0,43	0,6929	18,74
iPLS4	4	5	0,55	0,7068	0,43	0,7681	14,37
iPLS8	7	4	0,58	0,6732	0,42	0,7937	17,28
s2iPLS8	7, 8	5	0,55	0,7068	0,43	0,7681	7,90
s3iPLS8	2, 3, 7	5	0,54	0,7172	0,43	0,7985	16,39
s4iPLS8	1, 2, 3, 7	5	0,55	0,7164	0,44	0,8787	8,04

^a s—número de intervalos combinados iPLS—número de divisões do espectro; ^b Intervalos selecionados; ^c número de variáveis latentes.

Avaliando os resultados dos modelos gerados a partir do conjunto de dados, é possível verificar baixos valores de coeficientes de determinação para calibração e previsão e altos valores de erros, RMSECV e RMSEP, sendo maiores que os valores das amostras de menor concentração apresentados no Apêndice 8.1 Tabela 9.

5.3.2.1 Considerações quanto aos modelos de regressão multivariada a partir de dados FTIR-ATR para determinação do teor de glicídios redutores.

Os modelos de regressão por PLS gerados a partir dos dados FTIR-ATR para determinação de glicídios redutores em BBS, 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.

Isso pode ter ocorrido, porque os intervalos de comprimento de onda dos estiramentos axiais C-O, característicos dos glicídios, de $1260 - 1050 \text{ cm}^{-1}$,⁵⁸ sejam eles redutores e dos

totais, estão sobrepostos. Ou seja, a sobreposição da absorvância dos glicídios totais, em relação aos redutores, pode ter impedido a determinação de glicídios redutores. Também pode ser explicado através dos intervalos selecionados pelos modelos de determinação de glicídios redutores, com exceção dos modelos iPLS4 e iPLS8, não correspondem com os mesmos dos modelos de determinação dos glicídios totais. Ou seja, os modelos buscaram outras regiões do espectro para a tentativa de correlacionar adequadamente os valores previstos pelos modelos com os obtidos pelo método de referência.

Assim, podemos observar que os intervalos selecionados pelos modelos, iPLS4, intervalo 4, e s2iPLS8, intervalos 7 e 8, são respectivamente iguais, correspondendo a faixa espectral de $1150 - 900\text{cm}^{-1}$, e não englobando por completo a faixa dos estiramentos axiais C-O característicos dos glicídios, de $1260 - 1050\text{cm}^{-1}$,⁵⁸ indicando que o modelo está retirando informações mais representativas de regiões não seletivas do espectro.

5.3.3 Determinação do Teor de Glicídios Não Redutores por FTIR-ATR Combinada com Regressão Multivariada.

Foi seguido o mesmo procedimento dos casos anteriores. Foram construídos modelos de regressão por PLS, iPLS e siPLS a partir do conjunto de dados, pré tratados, obtidos por FTIR-ATR.

Por meio da Tabela 7 é possível verificar os resultados de RMSECV, RMSEP, R^2_{cv} , R^2_{prev} e número de variáveis latentes (VLs) dos melhores modelos desenvolvidos, de cada regressão para cada intervalo espectral selecionado, para a determinação de glicídios não redutores. Nela, é possível verificar as principais figuras de mérito para construir os modelos de regressão multivariada. O erro relativo, para o conjunto das amostras de previsão, também é apresentado, para ter uma melhor visualização do erro do modelo. A totalidade dos modelos gerados, que foram avaliados, são apresentadas no Apêndice 8.4 (Tabela 15).

O critério de escolha dos modelos apresentados, foram os mesmos da determinação de glicídios totais e glicídios redutores, menores valores de RMSECV e RMSEP e maiores coeficientes de correlação (R^2), assim como a avaliação do erro dos modelos de calibração interna cruzada, RMSECV, para determinação do número de VLs. Como nos casos anteriores, para avaliar a capacidade de previsão entre os modelos, utilizamos o erro do conjunto de previsão externa RMSEP.⁵⁰

Tabela 7: Resultados das figuras de mérito dos melhores modelos de regressão, para determinação do teor de glicídios não redutores com dados FTIR-ATR.

Modelos ^a	Intervalos selecionados ^b	VLS ^c	RMSECV (% m/m)	R ² _{cv}	RMSEP (% m/m)	R ² _{prev}	Erro relativo de previsão (% m/m)
PLS	Global	9	0,58	0,6597	0,43	0,2983	172,98
iPLS2	2	5	0,61	0,6239	0,50	0,3262	160,66
iPLS4	4	5	0,60	0,6252	0,51	0,2848	163,38
iPLS8	8	4	0,57	0,6667	0,51	0,3161	189,64
s2iPLS8	5, 8	4	0,58	0,6574	0,51	0,3276	194,09
s3iPLS8	4, 5, 8	4	0,58	0,6529	0,50	0,4222	196,74
s4iPLS8	1, 4, 5, 8	4	0,58	0,6522	0,51	0,4151	195,63

^a s—número de intervalos combinados iPLS—número de divisões do espectro; ^b Intervalos selecionados; ^c número de variáveis latentes.

Avaliando os resultados dos modelos gerados a partir do conjunto de dados, é possível verificar, assim como no glicídios redutores, baixos valores de coeficientes de determinação para calibração e previsão e altos valores de erros, RMSECV e RMSEP, sendo maiores que os valores das amostras de menor concentração apresentados no Apêndice 8.1 Tabela 11.

5.3.3.1 Considerações quanto aos modelos de regressão multivariada a partir de dados FTIR-ATR para determinação do teor de glicídios não redutores.

Os modelos de regressão por PLS gerados a partir dos dados FTIR-ATR para determinação de glicídios não redutores em BBS, assim como para glicídios redutores, 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.

Como a determinação do teor de glicídios não redutores é indireta, ou seja, obtida pela diferença entre o teor de glicídios totais, menos o teor de glicídios redutores, conforme equação 14, era esperado que a regressão não apresentasse resultado satisfatório, uma vez que o mesmo ocorreu com os glicídios redutores.

Como no caso anterior, pode ter ocorrido, a sobreposição da absorbância dos glicídios totais, em relação aos não redutores, o que acaba impedido a determinação. Com exceção do iPLS2 e iPLS4, que são modelos que usam faixas maiores, os demais buscaram outras regiões do espectro para a tentativa de correlacionar adequadamente os valores previstos pelos modelos com os obtidos pelo método de referência, assim como nos glicídios redutores.

5.3.4 Determinação do Teor de Proteínas Totais por FTIR-ATR Combinada com Regressão Multivariada.

Como na análise dos glicídios, também foram construídos modelos de regressão por PLS, iPLS e siPLS a partir do conjunto de dados, pré tratados, obtidos por FTIR-ATR.

Foram analisados modelos de regressão PLS, a partir de dois tipos de seleção de amostras para calibração e previsão. Como mostrado no Apêndice 8.1 (Tabela 12), temos amostras com pouca variabilidade de concentração de proteínas, havendo 72 amostras com concentração entre 0,35 até 0,66% (m/m) e somente 8 amostras com teores entre 1,79 até 2,54% (m/m), ocorrendo um vazio de informação da faixa de teores de proteínas totais de 0,67 até 1,78 % (m/m). A seleção por meio do algoritmo de Kennard-Stone⁵¹, onde das 80 amostras de BBS, 60 delas foram designadas para a construção dos modelos de calibração e as 20 amostras restantes foram utilizadas para avaliar a predição dos modelos. Como os resultados Kennard-Stone⁵¹, que se encontram no Apêndice 8.6, não foram eficazes para a previsão, visto que o algoritmo selecionou somente amostras de baixa concentração, foi realizada outra forma de seleção.

Como as amostras com concentrações mais elevadas eram em número pequeno, comparadas com as de teores mais baixos, e a variedade de concentração das amostras era pouca, sendo essa a realidade das amostras adquiridas no comércio, foi realizada uma seleção de amostras por escolha aleatória, onde 54 amostras, das 80 de BBS, foram designadas para a construção dos modelos de calibração e as 26 amostras restantes foram utilizadas para avaliar a predição dos modelos. A seleção dessas amostras é apresentada na Tabela 12. Assim se obteve amostras de mais variadas concentrações, nos dois modelos, o de calibração e o de previsão, obtendo melhores resultados para os parâmetros RMSECV, RMSEP, R^2_{cv} e R^2_{prev} , que serão apresentados e discutidos a seguir.

Por meio da Tabela 8 é possível verificar os resultados de RMSECV, RMSEP, R^2_{cv} , R^2_{prev} e número de variáveis latentes (VLs) dos melhores modelos desenvolvidos, de cada regressão para cada intervalo espectral selecionado, para a determinação de proteínas totais. Nela é possível verificar as principais figuras de mérito para construir os modelos de regressão multivariada. O erro relativo, para o conjunto das amostras de previsão, também é apresentado, para se ter uma melhor visualização do erro do modelo, para a determinação dos proteínas totais. O erro relativo alto, pode ser devido a grande faixa sem informação já mencionada. A totalidade dos modelos gerados, que foram avaliados, são apresentadas no Apêndice 8.5 (Tabela 16).

Como nas análises anteriores para escolher o número de VLs, de cada regressão para cada intervalo espectral selecionado, avaliamos o erro dos modelos de calibração cruzada RMSECV, tendo cuidado para escolher um número que não causasse subajuste nem superajuste dos modelos. Para avaliar a capacidade de previsão entre os modelos, foi utilizado o erro do conjunto de previsão externa RMSEP.⁵⁰

Tabela 8: Resultados das figuras de mérito dos melhores modelos de regressão, para determinação do teor de proteínas totais com dados FTIR-ATR.

Modelos ^a	Intervalos selecionados ^b	VLs ^c	RMSECV (% m/m)	R^2_{cv}	RMSEP (% m/m)	R^2_{prev}	Erro relativo de previsão (% m/m)
PLS	Global	9	0,13	0,9341	0,11	0,9592	12,85
iPLS2	1	9	0,17	0,9012	0,17	0,8955	15,33
iPLS4	3	9	0,12	0,9426	0,25	0,8028	16,92
iPLS8	3	8	0,24	0,8132	0,18	0,8819	19,33
s2iPLS8	3, 6	8	0,16	0,9055	0,13	0,9361	12,35
s3iPLS8	3, 5, 6	9	0,11	0,9512	0,10	0,9843	10,68
s4iPLS8	2, 3, 5, 6	9	0,12	0,9450	0,08	0,9833	10,40

^a s—número de intervalos combinados iPLS—número de divisões do espectro; ^b Intervalos selecionados; ^c número de variáveis latentes; Formação em negrito aos melhores modelos de regressão.

A partir dos resultados das regressões, usando os métodos PLS, iPLS e siPLS pode-se definir que as melhores opções de modelos promissores, que poderiam ser utilizados para a determinação de proteínas totais, para a rotulagem nutricional das BBS foram o PLS, o s3iPLS8 e o s4iPLS8. A combinação siPLS conseguiu suprimir algumas regiões do modelo

global, que não foram representativas, obtendo assim, melhorias nos resultados das figuras de mérito, justificando o uso do modelo siPLS. A vantagem da seleção siPLS, é que tem menos vulnerabilidade a flutuações quaisquer de equipamentos, tendo maior vida útil, do que um modelo global tão bom quanto. Os demais modelos não apresentaram resultados que pudessem apresentá-los como soluções, possuem valores de erros elevados e coeficientes de determinação baixos, quando comparados com os selecionados, mesmo os que utilizam 8 VLs, não se justifica sua escolha.

Quando avaliamos as regiões espectrais selecionadas pelos métodos de regressão iPLS e siPLS, podemos observar que as faixas selecionadas para a construção dos modelos, correspondem a regiões importantes, que contém absorção de ligações com nitrogênio, indicando que são faixas representativas na determinação das proteínas totais nas amostras de bebidas a base de soja. Nos modelos de regressão, é desejado que as regiões selecionadas, tenham relação com a composição química. As amidas secundárias, presentes nas proteínas, apresentam uma banda muito forte (longa) do estiramento do grupo carbonila (C=O), que aparece na faixa de 1680 a 1630 cm^{-1} ,⁵⁸ a banda pode sobrepor, parcialmente, a de dobramento N-H, de amidas secundárias, que aparece na faixa de 1640 a 1550 cm^{-1} .⁵⁸ Essas faixas, estão presentes, de maneira integral ou parcial, nos modelos iPLS2, iPLS8, s2iPLS8, s3iPLS8 e s4iPLS8. Na faixa de 1350 a 1000 cm^{-1} , ocorre o estiramento C-N,⁵⁸ e uma faixa da mesma, esta selecionada nos modelos iPLS4, s2iPLS8, s3iPLS8 e s4iPLS8. Nenhum modelo da regressão iPLS apresentou resultado satisfatório, provavelmente por não combinar duas regiões do espectro infravermelho, as regiões de estiramento do grupo carbonila e do dobramento da ligação N-H (1640 – 1550 cm^{-1}) que se sobrepõem, com a região do estiramento da ligação C-N (1350 – 1000 cm^{-1}). A combinação delas, provavelmente têm relevância na determinação do teor de proteínas totais em BBS. As mesmas, estão presentes, total ou parcialmente, em todos os modelos de regressão siPLS, e no PLS.

Para todos os modelos, vale ressaltar que a matriz de análise, a BBS, é bastante complexa, tendo a presença de diferentes substâncias como proteínas, vitaminas, gorduras, fibras, metais, aromatizantes, corantes e glicídios como frutose, glicose, sacarose, gomas entre outros carboidratos, o que pode elevar o número de VLs do modelo. Os modelos, que podem ser utilizados para a determinação de proteínas totais em BBS, já mencionados anteriormente, são apresentados a seguir.

O modelo PLS, que é o modelo global, utiliza todo o espectro, sem selecionar uma região específica, corresponde ao intervalo espectral $1899 - 900 \text{ cm}^{-1}$, com 9 VLs. Apresentou $\text{RMSECV} = 0,13\%$, $\text{RMSEP} = 0,11\%$ e coeficientes de determinação (R^2_{cv} e R^2_{prev}) de 0,9341 e 0,9592, respectivamente. Engloba todas as regiões do espectro, representativas para a determinação do teor de proteína, os estiramentos da ligação do grupo C=O característicos das amidas secundárias, de $1680 - 1630 \text{ cm}^{-1}$, a de dobramento N-H, de amidas secundárias, de 1640 a 1550 cm^{-1} , e o estiramento C-N, de 1350 a 1000 cm^{-1} ,⁵⁸ mas também pode considerar regiões não tão relevantes.

Como no estudo dos glicídios, a definição do número de variáveis latentes é de fundamental importância para que não ocorra o sobreajuste ou subajuste, sendo pela Figura 16 possível justificar a escolha pelo número de 9 VLs do modelo PLS, possuindo menor erro de calibração, determinado por validação cruzada (RMSECV), e a partir desse número de variáveis não ocorre decréscimo significativo do erro.

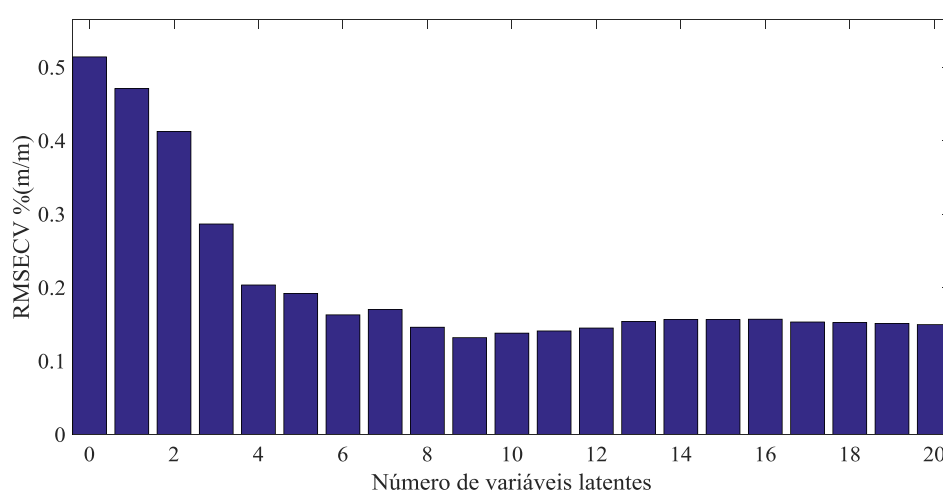


Figura 16: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de proteínas totais, por meio de espectroscopia no infravermelho, combinando to espectro no modelo PLS.

Um gráfico de correlação dos valores de referência, obtidos pelo método oficial para o teor de proteínas totais, com os valores obtidos pelo melhor modelo de regressão PLS através do FTIR-ATR é ilustrado na Figura 17. 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 pode apresentar desempenho aceitável para determinação do teor de proteínas totais em amostras de BBS.

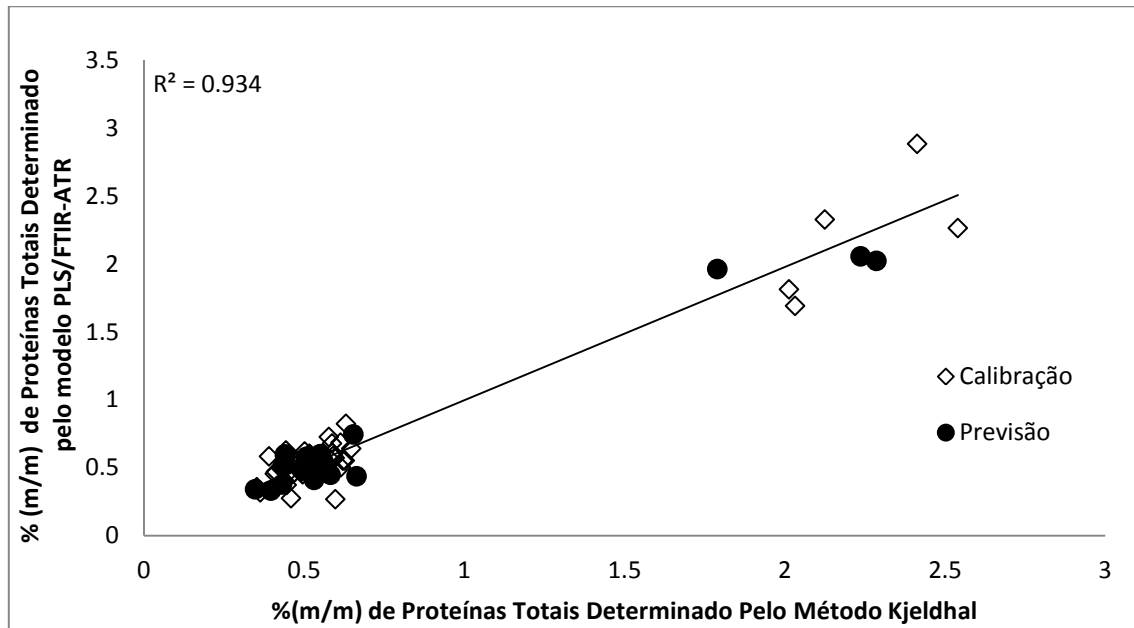


Figura 17: Valores de referência pelo método Kjeldhal *versus* valores previstos do modelo PLS para proteínas totais utilizando todo o espectro correspondente ao intervalo espectral de 1899 – 900 cm^{-1} .

O modelo s3iPLS8 dividiu o espectro em 8 partes e combinou os subintervalos selecionados 3, 5 e 6, que correspondem aos intervalos espectrais 1649 – 1525 cm^{-1} , 1399 – 1275 cm^{-1} e 1274 – 1150 cm^{-1} , com 9 VLS, conforme ilustrado na Figura 18. Apresentou RMSECV = 0,11%, RMSEP = 0,10% e coeficientes de determinação (R^2_{cv} e R^2_{prev}) de 0,9512 e 0,9843, respectivamente. Os intervalos selecionados, englobam os estiramentos da ligação do grupo C=O característicos das amidas secundárias, de 1680 – 1630 cm^{-1} , a de dobramento N-H, de amidas secundárias, de 1640 a 1550 cm^{-1} , e o estiramento C-N, de 1350 a 1000 cm^{-1} ,⁵⁸ indicando que esta é uma faixa representativa na determinação das proteínas totais nas amostras de bebidas a base de soja, ao selecionar uma região específica de absorção.

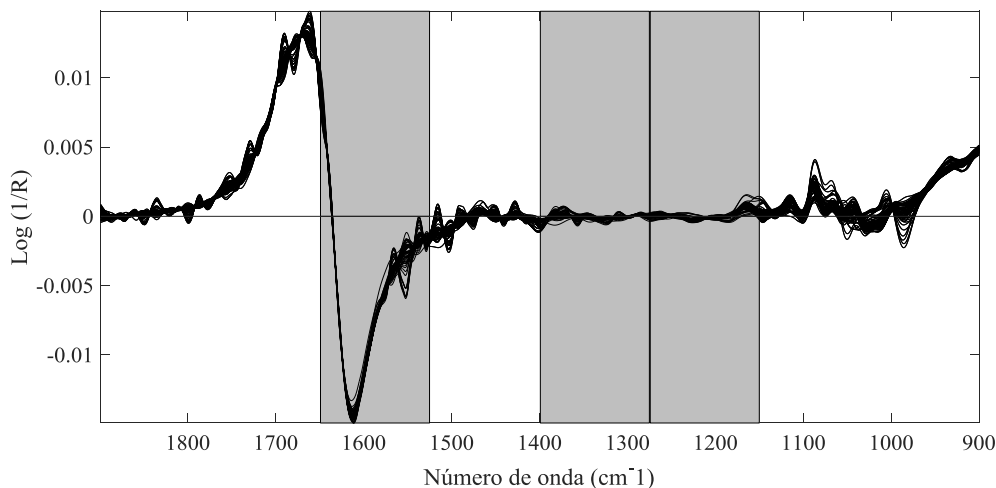


Figura 18: Espectro com destaque as regiões selecionadas (3, 5 e 6) combinadas com regressão siPLS para determinação de proteínas totais em BBS.

Na Figura 19, é possível justificar a escolha pelo número de 9 VLs do modelo s3iPLS8, possuindo menor erro de calibração, determinado por validação cruzada (RMSECV), e a partir desse número de variáveis não ocorre decréscimo significativo do erro.

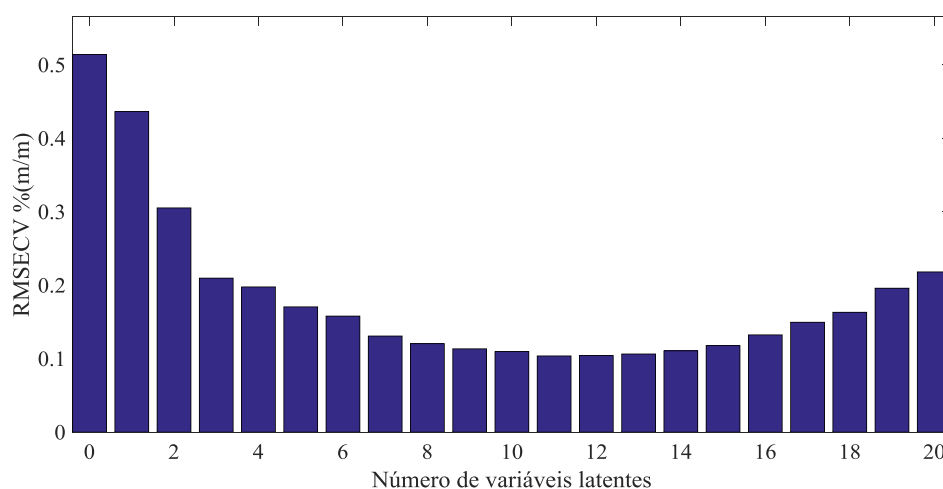


Figura 19: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de proteínas totais, por meio de espectroscopia no infravermelho, combinando os intervalos 3, 5 e 6 no modelo s3iPLS8.

Também foi construído um gráfico de correlação dos valores de referência, obtidos pelo método oficial para o teor de proteínas totais, com os valores obtidos pelo melhor modelo de regressão através do FTIR-ATR conforme ilustrado na Figura 20. 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 também pode apresentar desempenho aceitável para determinação do teor de proteínas totais em amostras de BBS.

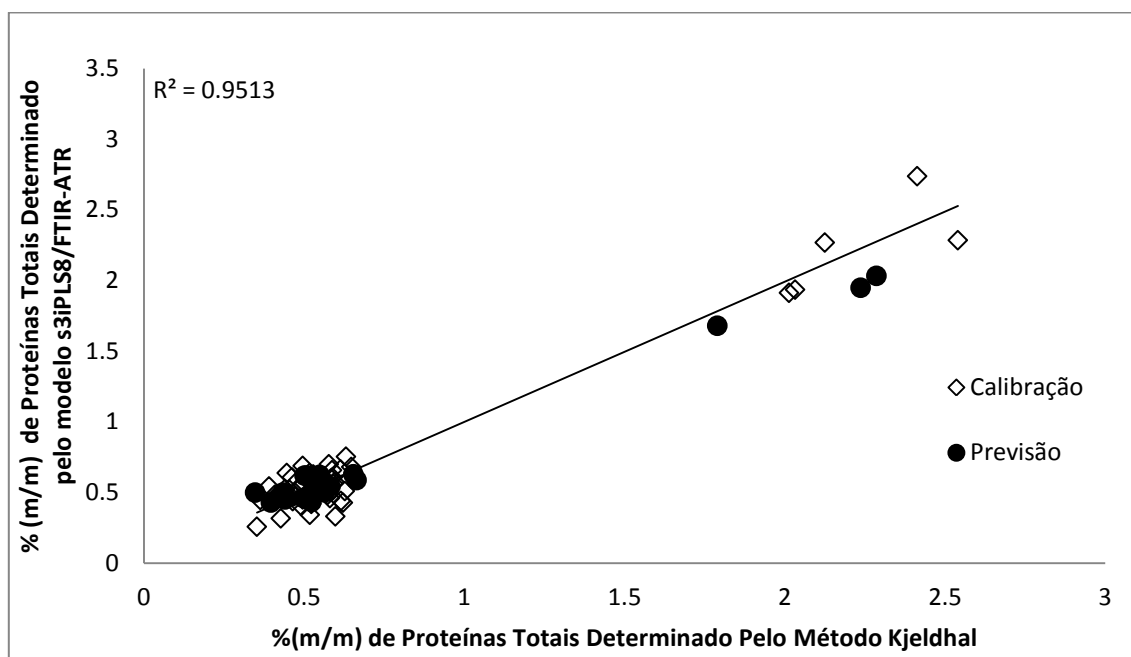


Figura 20: Valores de referência pelo método Kjeldhal *versus* valores previstos do modelo s3iPLS8 para proteínas totais usando os intervalos 3, 5 e 6 correspondentes aos intervalos espectrais $1649 - 1525 \text{ cm}^{-1}$, $1399 - 1275 \text{ cm}^{-1}$ e $1274 - 1150 \text{ cm}^{-1}$.

O modelo s4iPLS8 dividiu o espectro em 8 partes e combinou os subintervalos selecionados 2, 3, 5 e 6, que correspondem aos intervalos espectrais $1774 - 1650 \text{ cm}^{-1}$, $1649 - 1525 \text{ cm}^{-1}$, $1399 - 1275 \text{ cm}^{-1}$ e $1274 - 1150 \text{ cm}^{-1}$, com 9 VLs, conforme ilustrado na Figura 21. Apresentou RMSECV = 0,12%, RMSEP = 0,08% e coeficientes de determinação (R^2_{cv} e R^2_{prev}) de 0,9450 e 0,9833, respectivamente. Os intervalos selecionados, como nos casos anteriores, também englobam os estiramentos da ligação do grupo C=O característicos das amidas secundárias, de $1680 - 1630 \text{ cm}^{-1}$, a de dobramento N-H, de amidas secundárias, de 1640 a 1550 cm^{-1} , e o estiramento C-N, de 1350 a 1000 cm^{-1} ,⁵⁸ indicando que esta, também é, uma faixa representativa na determinação das proteínas totais nas amostras de bebidas a base de soja, ao selecionar uma região específica de absorção.

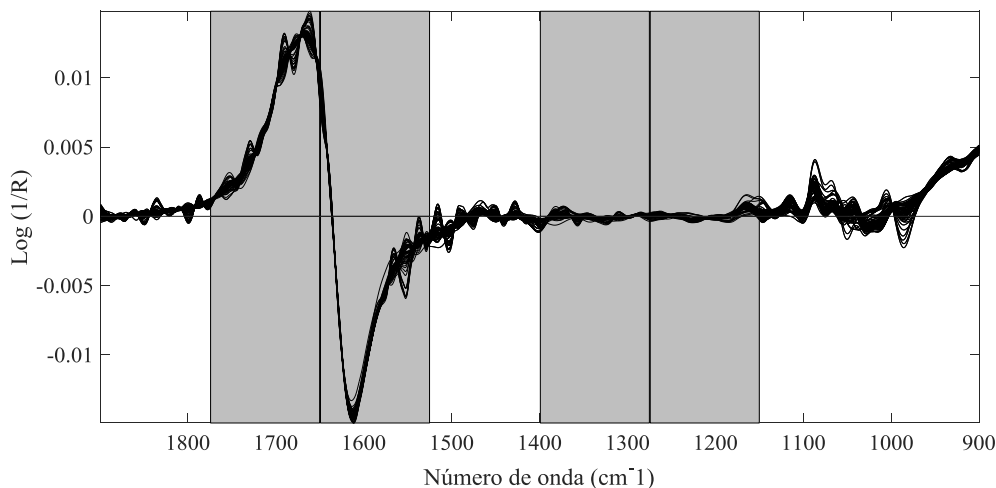


Figura 21: Espectro com destaque as regiões selecionadas (2, 3, 5 e 6) combinadas com regressão siPLS para determinação de proteínas totais em BBS.

Na Figura 22, é possível justificar a escolha pelo número de 9 VLs do modelo s4iPLS8, possuindo menor erro de calibração, determinado por validação cruzada (RMSECV), e a partir desse número de variáveis não ocorre decréscimo significativo do erro.

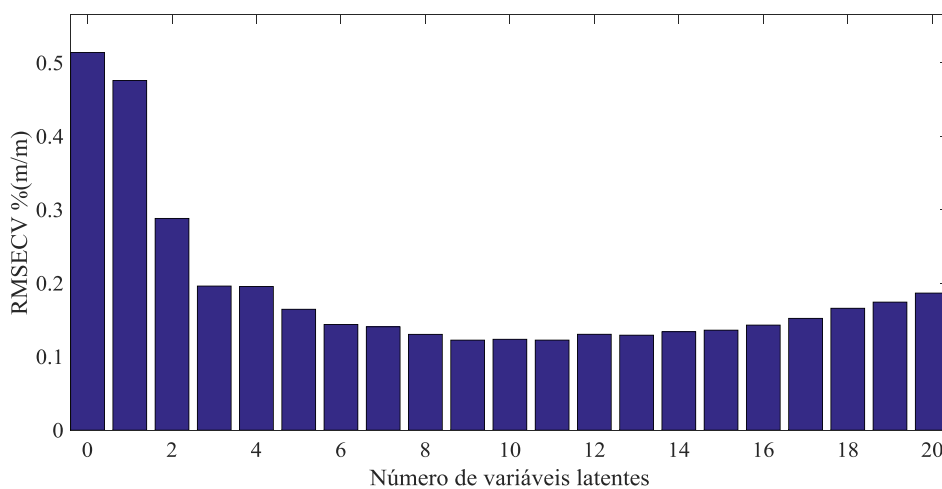


Figura 22: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de proteínas totais, por meio de espectroscopia no infravermelho, combinando os intervalos 2, 3, 5 e 6 no modelo s4iPLS8.

Também foi construído um gráfico de correlação dos valores de referência, obtidos pelo método oficial para o teor de proteínas totais, com os valores obtidos pelo modelo de regressão s4iPLS8, através do FTIR-ATR conforme ilustrado na Figura 23.

Como nos estudos anteriores, 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 também pode apresentar desempenho aceitável para determinação do teor de proteínas totais em amostras de BBS.

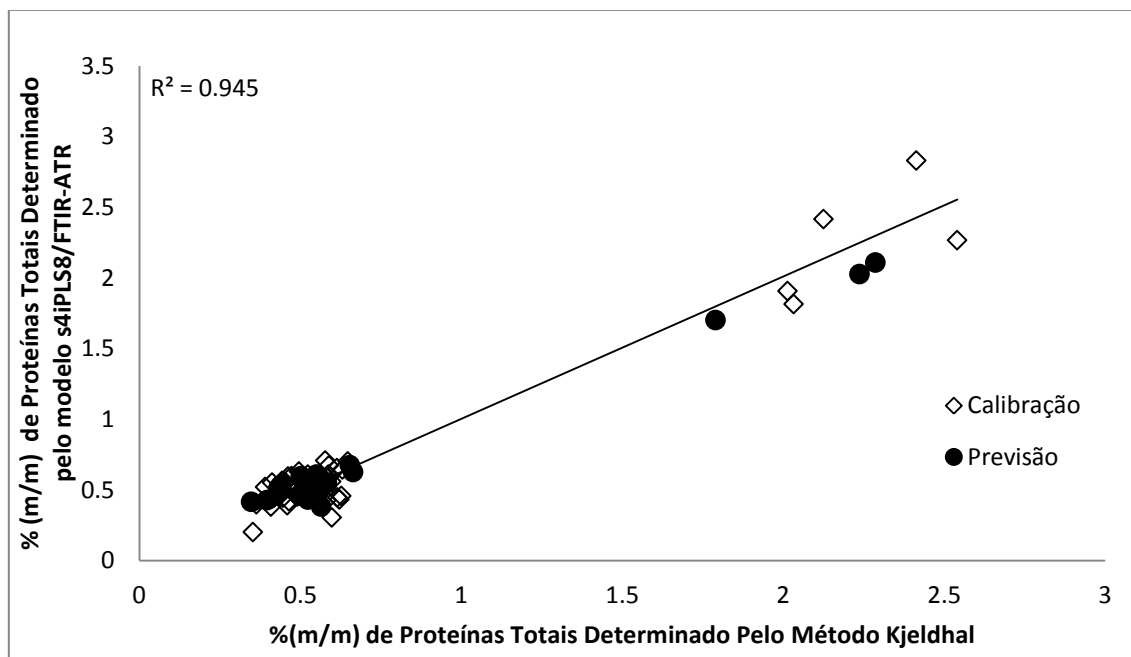


Figura 23: Valores de referência pelo método Kjeldhal *versus* valores previstos do modelo s4iPLS8 para proteínas totais usando os intervalos 2, 3, 5 e 6 correspondentes aos intervalos espectrais $1774 - 1650\text{cm}^{-1}$, $1649 - 1525\text{cm}^{-1}$, $1399 - 1275\text{cm}^{-1}$ e $1274 - 1150\text{cm}^{-1}$.

5.3.4.1 Considerações quanto ao modelo de regressão multivariada a partir de dados FTIR-ATR para determinação do teor de proteínas totais.

Com base nesses resultados, os modelos que apresentaram serem os mais adequados para a determinação de proteínas totais, expressos em % (m/m) foram o PLS, o s3iPLS8 e s4iPLS8, todos com 9 VLs.

Outro fator que justifica a escolha dos modelos, principalmente os modelos s3iPLS8 e s4iPLS8, são as regiões espectrais selecionadas, que incluem os intervalos de absorção característico do estiramento do grupo carbonila (C=O) de amidas secundárias, do dobramento N-H, de amidas secundárias, e os estiramentos C-N.

Por fim, as Figuras 18, 20 e 23 apresentam a correlação entre o valor real e o valor previsto para os modelos PLS, s3iPLS8 e s4iPLS8 desenvolvidos, que demonstraram serem os modelos com as melhores capacidades de determinar a % (m/m) de proteínas totais em glicose presentes em bebidas a base de soja.

Os modelos para determinação do teor de proteínas totais % (m/m) em BBS, mostraram um ótimo potencial, mesmo com um vazio de informação da faixa de teores de proteínas totais, os modelos se mostraram promissores, pois os resultados obtidos, são satisfatórios para a determinação de proteínas totais, para a rotulagem nutricional das BBS, uma vez que, os modelos apresentaram um baixo erro e a legislação vigente, não requer um baixo limite de detecção nem uma declaração mais detalhada dos valores dos nutrientes listados nas embalagens.

6 CONCLUSÕES

Por meio dos resultados alcançados na determinação das propriedades estudadas, foi possível constatar que os modelos gerados, com o objetivo de caracterizar bebidas à base de soja empregando técnicas de espectroscopia de infravermelho médio, em conjunto com ferramentas quimiométricas de regressão multivariada, como PLS, iPLS e siPLS, que podem atender à legislação referente às informações nutricionais na rotulagem das BBS, tiveram êxito.

Os modelos obtidos com base nos dados FTIR-ATR, se mostraram promissores para a determinação de mais de uma propriedade simultaneamente, pois uma mesma região no espectro de infravermelho pode conter informações sobre vários compostos químicos, ou seja, os sinais espectrais destes compostos ficam sobrepostos nessas regiões, sendo que para amostras complexas, como no caso das BBS, pode ser considerado um fator positivo.

Na avaliação das propriedades de glicídios redutores e não redutores, 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 desses parâmetros com base nos modelos construídos no presente estudo.

Portanto, para o teor de glicídios totais, a técnica de análise multivariada por regressão PLS, com dados de infravermelho, apresentou resultados satisfatórios, para a rotulagem nutricional das BBS. E para a determinação do teor de proteínas totais, a técnica de análise multivariada por regressão PLS, com dados do infravermelho, apresentou resultados promissores, para o mesmo objetivo. Essa técnica apresenta diversas vantagens sobre os métodos tradicionais de determinação desses parâmetros como: redução no tempo e no custo de análise, não destruição das amostras, possibilidade de análise em tempo real e simultâneo de ambos parâmetros e a possibilidade de ser facilmente implementada no processo produtivo.

7 REFERÊNCIAS BIBLIOGRÁFICAS

1. Mohanty, A. K.; Misra, M.; Drzal, L.; *J. Polym. Envir.* **2002**, 10, 19.
2. Doretto, K. M.; Gotardo, M. A.; Pezza, H. R.; Pezza, L.; *Quim. Nova.* **2010**, 33, 2, 461.
3. Zenebon, O.; Pascuet, N. S.; Tiglea, P.; Métodos físico-químicos para análise de alimentos; Instituto Adolfo Lutz: São Paulo, 2008.
4. Cozzolino, D.; *Antioxidants.* **2015**, 4, 482.
5. Jiang, H.; Zhuang, H.; Sohn, M.; Wang, W.; *Appl. Sci.* **2017**, 7, 97.
6. Xu, J.; Huang, F.; Li, Y.; Chen, Z.; Wang, Y.; *Czech J. Food Sci.* **2015**, 33, 518.
7. Monakhova, Y. B.; Duiehl, B. W. K.; *J. Am. Oil Chem. Soc.* **2015**, 93, 27.
8. USDA. Oilseeds: World markets and Trade. United States Department of Agriculture. 2017. Disponível em <http://usda.mannlib.cornell.edu/usda/fas/oilseed-trade//2010s/2017/oilseed-trade-07-12-2017.pdf>. Acesso em 30 de outubro de 2017.
9. EMBRAPA. Empresa Brasileira de Pesquisa Agropecuária. Soja na alimentação. Composição da soja. 2017. Disponível em http://www.cnpso.embrapa.br/soja_alimentacao/index.php?pagina=7. Acesso em 30 de outubro de 2017.
10. Rigo, A. A.; Dahmer, A. M.; Steffens, C.; Steffens, J.; *Int. J. Food Eng.* **2015**, 1, 1, 1.
11. Castanheira, E. G.; Grisoli, R.; Coelho, S.; Silva, G. A.; Freire, F.; *J. of Cleaner Prod.* **2015**, 102, 188.
12. Amaral, V. M. G.; A importância da soja como alimento funcional para qualidade de vida e saúde. 2007. 71f. Dissertação (Mestrado em Gestão da Qualidade) – Universidade Estadual de Campinas, Campinas, 2006.
13. Papaleo, V. T.; As propriedades da proteína de soja na alimentação humana. 2004. 44f. Monografia (Qualidade em Alimentos). Universidade Brasília, Brasília, 2004.
14. UFRGS. Universidade Federal do Rio Grande do Sul. Alimentos e novas tecnologias na UFRGS. Extrato de soja. 2017. Disponível em <http://www.ufrgs.br/alimentus/objetos-de-aprendizagem/soja/extrato-de-soja>. Acesso em 07 de novembro de 2017.
15. UFRGS. Universidade Federal do Rio Grande do Sul. Alimentos e novas tecnologias na UFRGS. Isolado proteico. 2017. Disponível em <http://www.ufrgs.br/alimentus/objetos-de-aprendizagem/soja/isolado-proteico>. Acesso em 07 de novembro de 2017.
16. Pinheiro, D. M.; Porto, K. R. A.; Menezes, M. E. S.; A química dos alimentos: carboidratos, lipídios, proteínas e minerais; EDUFAL: Maceió, 2005.

17. Francis, A. C.; Química orgânica; AMGH: Porto Alegre, 2011
18. Llanos, P.I.; Rogero, M. M.; García, L. A.; Boasi, L. E. V.; Torres, N.; Soja como uma fonte de proteína de alta qualidade; ILSI: São Paulo, 2016.
19. Anthony, M. S.; Clarkson, T. B.; Hughes, C. L.; Jr, Morgan, T. M.; Burke, G.L.; *J. Nutr.* **1996**, 126, 1, 43.
20. Miniello, V. L.; Moro, G. E.; Tarantino, M.; Natile, M.; Granieri, L.; Armenio, L.; *Acta. Paediatr. Suppl.* **2003**, 91, 441, 93.
21. Kostelac, D.; Rechkemmer, G.; Briviba, K.; *J. Agric. Food Chem.* **2003**, 51, 26, 7632.
22. ANVISA. AGENCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. Resolução da diretoria colegiada- RDC nº 360, de 23 de Dezembro de 2003. Disponível em:< www.anvisa.gov.br/legis> Acesso em 10 de maio de 2018
23. Holler, F. J.; Skoog, D. A.; Crouch, S. R. Princípios de Análise Instrumental; Bookman: Porto Alegre, 2009.
24. Dantas Filho, H. A.; Desenvolvimento de técnicas quimiométricas de compressão de dados e de redução de ruído instrumental aplicadas a óleo diesel e madeira de eucalipto usando espectroscopia NIR. 2007. 158f. Tese (Doutorado em Química). Universidade estadual de Campinas, Campinas, 2007.
25. Barbosa, L. C. A.; Espectroscopia no Infravermelho na Caracterização de Compostos Orgânicos; Ufv: Viçosa, 2007.
26. Stuart, B.; George, B.; McIntyre, P.; Modern Infrared Spectroscopy; West Sussex: England, 1996.
27. 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.
28. Harrick, N.; *J. J. Phys. Chem.* **1960**, 64, 1110.
29. Ferrão, M. F.; Viera, M. D. S.; Pazos, R. E. P.; Fachini, D.; Gerbase, A. E.; Marder, L. *Fuel.* **2011**, 90, 701.
30. 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.
31. Brereton, R. G.; *Anal. Meth.* **2013**, 5, 3785.
32. Ferreira, M. M. C.; Quimiometria: conceitos, métodos e aplicações; Unicamp: Campinas, 2015.

33. Brereton, R. G.; Chemometrics: data analysis for the laboratory and chemical plant; John Wiley & Sons Ltd: Chichester, 2003.
34. Gemperline, P. J.; Practical guide to chemometrics. 2^oEd.; CRC Press Taylor & Francis: Nova York, 2006.
35. Brereton, R. G.; *Analyst*. **2000**, 125, 2125.
36. 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.
37. Burns, D. A.; Ciurczak, E. W.; Hand book of near-infrared analysis. 3 Ed.; CRC: Boca Raton, 2008.
38. Huck, C. W.; *J. Pharm. Biomed. Anal.* **2014**, 87, 26.
39. Geladi, P.; Kowalski, B. R.; *Anal. Chim. Acta.* **1986**. 185, 1, 1.
40. Zamora, P. P.; Ponce, L. C.; Nagata, N.; Poppi, R. J.; *Quím. Nova.* **1997**, 20, 5, 469.
41. Geladi, P.; Macdougall, D.; Martens, H.; *Appl. Spectrosc.* **1985**, 39, 491.
42. Jiang, J.; Berry, R.J.; Siester, H.W.; Ozaki, Y.; *Anal. Chem.* **2002**. 74, 14, 3555.
43. Oliveira, F. C. C.; de Souza, A.; Dias, J.; Dias, S. C. L.; Rubim, J. C.; *Quím. Nova.* **2004**, 27, 218.
44. Liu, F.; He, Y.; *Food Chem.* **2009**, 115, 4, 1430.
45. Jacob, S.; Banerjee, R.; *Bioresour. Technol.* **2016**, 214, 386.
46. Wu, D.; Chen, X.; Shi, P.; Wang, S.; Feng, F.; He, Y.; *Anal. Chim. Acta.* **2009**, 634, 2, 166.
47. Almeida, W. R. R.; Silva, A. S.; Paula, F. R.; Silva, F. E. B.; Foresti, G. R.; Ferrão, M. F.; *Orbital: The Elect. J. Chem.* **2015**, 7, 4, 341.
48. Leardi, R.; Norgaard, L.; *J. Chemom.* **2004**, 18, 486.
49. Norgaard, L.; Saudland, A.; Wagner, J.; Nielse, J. P.; Munck, L.; Engelsen, S. B.; *Appl. Spectrosc.* **2000**, 54, 413.
50. Xiaobo, Z.; Jiewen, Z.; Povey, M. J. W.; Holmes, M.; Hanpin, M.; *Anal. Chim. Acta.* **2010**, 667, 14.
51. Kennard, R. W.; Stone, L. A.; *Technometrics.* **1969**, 11, 137.
52. Daszykowski, M.; Walczak, B.; Massart, D. L.; *Anal. Chim. Acta.* **2002**, 468, 91.
53. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O.; *Quim. Nova.* **1999**, 22, 724.
54. Savitzky, A.; Golay, M. J. E.; *Anal. Chem.* **1964**, 8, 1627.

55. 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.
56. Gomes-Carracedo, M. P.; Andrade, J. M.; Rutledge, D. N.; Faber, N. M.; *Anal. Chim. Acta.* **2007**, 585, 253.
57. Helfer G. A.; Bock, F.; Marder, L.; Furtado, J. C.; Costa, A. B.; Ferrão, M. F.; *Quim. Nova.* **2015**. 38, 4, 575.
58. Pavia, D. L.; Lampman, G. M.; Kriz, G. S.; Vyvyan, J. R.; Introduction to Spectroscopy, 4th ed.; Brooks/Cole: Belmont, 2009.

8 APÊNDICES

8.1 RESULTADOS DA DETERMINAÇÃO DO TEOR DE GLICÍDIOS REDUTORES, GLICÍDIOS TOTAIS, GLICÍDIOS NÃO REDUTORES E DE PROTEÍNAS TOTAIS, PELOS MÉTODOS DE REFERÊNCIA.

Tabela 9: Resultados da determinação de glicídios redutores, em amostras de BBS pelo método de referência Lane-Eynon.

Código das Amostras	Teor de Glicídios redutores (% m/m) de glicose*	Código das Amostras	Teor de Glicídios redutores (% m/m) de glicose*
BBS01/00	3,53 (±0,01)	BBS01/10	3,17 (±0,01)
BBS02/00	3,23 (±0,01)	BBS02/10	2,91 (±0,01)
BBS03/00	2,95 (±0,01)	BBS03/10	2,66 (±0,01)
BBS04/00	2,89 (±0,01)	BBS04/10	2,60 (±0,01)
BBS05/00	2,85 (±0,01)	BBS05/10	2,57 (±0,01)
BBS06/00	3,37 (±0,02)	BBS06/10	3,03 (±0,02)
BBS07/00	2,67 (±0,02)	BBS07/10	2,41 (±0,02)
BBS08/00	1,01 (±0,01)	BBS08/10	0,91 (±0,01)
BBS09/00	2,19 (±0,02)	BBS09/10	1,97 (±0,02)
BBS10/00	2,87 (±0,02)	BBS10/10	2,59 (±0,02)
BBS11/00	0,59 (±0,01)	BBS11/10	0,53 (±0,01)
BBS12/00	3,80 (±0,04)	BBS12/10	3,42 (±0,04)
BBS13/00	0,64 (±0,03)	BBS13/10	0,57 (±0,03)
BBS14/00	0,56 (±0,01)	BBS14/10	0,50 (±0,01)
BBS15/00	0,73 (±0,01)	BBS15/10	0,66 (±0,01)
BBS16/00	2,27 (±0,03)	BBS16/10	2,04 (±0,03)
BBS17/00	1,35 (±0,01)	BBS17/10	1,21 (±0,01)
BBS18/00	2,89 (±0,03)	BBS18/10	2,60 (±0,03)
BBS19/00	2,71 (±0,02)	BBS19/10	2,44 (±0,02)
BBS20/00	2,91 (±0,03)	BBS20/10	2,62 (±0,03)
BBS01/05	3,35 (±0,01)	BBS01/20	2,82 (±0,01)
BBS02/05	3,07 (±0,01)	BBS02/20	2,59 (±0,01)
BBS03/05	2,81 (±0,01)	BBS03/20	2,36 (±0,01)
BBS04/05	2,74 (±0,01)	BBS04/20	2,31 (±0,01)
BBS05/05	2,71 (±0,01)	BBS05/20	2,28 (±0,01)
BBS06/05	3,20 (±0,02)	BBS06/20	2,70 (±0,02)
BBS07/05	2,54 (±0,02)	BBS07/20	2,14 (±0,02)
BBS08/05	0,96 (±0,01)	BBS08/20	0,81 (±0,01)
BBS09/05	2,08 (±0,02)	BBS09/20	1,75 (±0,02)
BBS10/05	2,73 (±0,02)	BBS10/20	2,30 (±0,02)
BBS11/05	0,56 (±0,01)	BBS11/20	0,47 (±0,01)
BBS12/05	3,61 (±0,04)	BBS12/20	3,04 (±0,04)
BBS13/05	0,60 (±0,03)	BBS13/20	0,51 (±0,03)

BBS14/05	0,53 ($\pm 0,01$)	BBS14/20	0,45 ($\pm 0,01$)
BBS15/05	0,70 ($\pm 0,01$)	BBS15/20	0,59 ($\pm 0,01$)
BBS16/05	2,15 ($\pm 0,03$)	BBS16/20	1,81 ($\pm 0,03$)
BBS17/05	1,28 ($\pm 0,01$)	BBS17/20	1,08 ($\pm 0,01$)
BBS18/05	2,74 ($\pm 0,03$)	BBS18/20	2,31 ($\pm 0,03$)
BBS19/05	2,57 ($\pm 0,02$)	BBS19/20	2,17 ($\pm 0,02$)
BBS20/05	2,77 ($\pm 0,03$)	BBS20/20	2,33 ($\pm 0,03$)

*Média das triplicas \pm desvio padrão

Tabela 10: Resultados da determinação de glicídios totais, em amostras de BBS pelo método de referência Lane-Eynon.

Código das Amostras	Teor de Glicídios totais (% m/m) de glicose*	Código das Amostras	Teor de Glicídios totais (%m/m) de glicose*
BBS01/00	3,92 ($\pm 0,05$)	BBS01/10	3,53 ($\pm 0,05$)
BBS02/00	3,56 ($\pm 0,01$)	BBS02/10	3,21 ($\pm 0,01$)
BBS03/00	3,87 ($\pm 0,01$)	BBS03/10	3,48 ($\pm 0,01$)
BBS04/00	3,75 ($\pm 0,03$)	BBS04/10	3,38 ($\pm 0,03$)
BBS05/00	3,71 ($\pm 0,01$)	BBS05/10	3,34 ($\pm 0,01$)
BBS06/00	3,80 ($\pm 0,01$)	BBS06/10	3,42 ($\pm 0,01$)
BBS07/00	3,37 ($\pm 0,01$)	BBS07/10	3,04 ($\pm 0,01$)
BBS08/00	1,25 ($\pm 0,01$)	BBS08/10	1,13 ($\pm 0,01$)
BBS09/00	3,77 ($\pm 0,01$)	BBS09/10	3,39 ($\pm 0,01$)
BBS10/00	4,00 ($\pm 0,02$)	BBS10/10	3,60 ($\pm 0,02$)
BBS11/00	2,74 ($\pm 0,02$)	BBS11/10	2,46 ($\pm 0,02$)
BBS12/00	7,92 ($\pm 0,04$)	BBS12/10	7,13 ($\pm 0,04$)
BBS13/00	2,63 ($\pm 0,01$)	BBS13/10	2,37 ($\pm 0,01$)
BBS14/00	2,88 ($\pm 0,01$)	BBS14/10	2,59 ($\pm 0,01$)
BBS15/00	4,32 ($\pm 0,02$)	BBS15/10	3,89 ($\pm 0,02$)
BBS16/00	2,32 ($\pm 0,02$)	BBS16/10	2,09 ($\pm 0,02$)
BBS17/00	2,81 ($\pm 0,03$)	BBS17/10	2,53 ($\pm 0,03$)
BBS18/00	4,19 ($\pm 0,03$)	BBS18/10	3,77 ($\pm 0,03$)
BBS19/00	4,16 ($\pm 0,03$)	BBS19/10	3,74 ($\pm 0,03$)
BBS20/00	4,43 ($\pm 0,02$)	BBS20/10	3,98 ($\pm 0,02$)
BBS01/05	3,72 ($\pm 0,05$)	BBS01/20	3,14 ($\pm 0,05$)
BBS02/05	3,39 ($\pm 0,01$)	BBS02/20	2,85 ($\pm 0,01$)
BBS03/05	3,67 ($\pm 0,01$)	BBS03/20	3,09 ($\pm 0,01$)
BBS04/05	3,56 ($\pm 0,03$)	BBS04/20	3,00 ($\pm 0,03$)
BBS05/05	3,52 ($\pm 0,01$)	BBS05/20	2,97 ($\pm 0,01$)
BBS06/05	3,61 ($\pm 0,01$)	BBS06/20	3,04 ($\pm 0,01$)
BBS07/05	3,20 ($\pm 0,01$)	BBS07/20	2,70 ($\pm 0,01$)
BBS08/05	1,19 ($\pm 0,01$)	BBS08/20	1,00 ($\pm 0,01$)

BBS09/05	3,58 ($\pm 0,01$)	BBS09/20	3,01 ($\pm 0,01$)
BBS10/05	3,80 ($\pm 0,02$)	BBS10/20	3,20 ($\pm 0,02$)
BBS11/05	2,60 ($\pm 0,02$)	BBS11/20	2,19 ($\pm 0,02$)
BBS12/05	7,53 ($\pm 0,04$)	BBS12/20	6,34 ($\pm 0,04$)
BBS13/05	2,50 ($\pm 0,01$)	BBS13/20	2,11 ($\pm 0,01$)
BBS14/05	2,74 ($\pm 0,01$)	BBS14/20	2,30 ($\pm 0,01$)
BBS15/05	4,11 ($\pm 0,02$)	BBS15/20	3,46 ($\pm 0,02$)
BBS16/05	2,21 ($\pm 0,02$)	BBS16/20	1,86 ($\pm 0,02$)
BBS17/05	2,67 ($\pm 0,03$)	BBS17/20	2,25 ($\pm 0,03$)
BBS18/05	3,98 ($\pm 0,03$)	BBS18/20	3,35 ($\pm 0,03$)
BBS19/05	3,95 ($\pm 0,03$)	BBS19/20	3,33 ($\pm 0,03$)
BBS20/05	4,21 ($\pm 0,02$)	BBS20/20	3,54 ($\pm 0,02$)

*Média das triplicas \pm desvio padrão

Tabela 11: Resultados da determinação de glicídios não redutores, em amostras de BBS pelo método de referência Lane-Eynon.

Código das Amostras	Teor de Glicídios não redutores (% m/m) de sacarose *	Código das Amostras	Teor de Glicídios não redutores (% m/m) de sacarose*
BBS01/00	0,37 ($\pm 0,04$)	BBS01/10	0,34 ($\pm 0,04$)
BBS02/00	0,31 ($\pm 0,02$)	BBS02/10	0,28 ($\pm 0,02$)
BBS03/00	0,87 ($\pm 0,01$)	BBS03/10	0,78 ($\pm 0,01$)
BBS04/00	0,82 ($\pm 0,01$)	BBS04/10	0,74 ($\pm 0,01$)
BBS05/00	0,81 ($\pm 0,01$)	BBS05/10	0,73 ($\pm 0,01$)
BBS06/00	0,41 ($\pm 0,01$)	BBS06/10	0,36 ($\pm 0,01$)
BBS07/00	0,67 ($\pm 0,01$)	BBS07/10	0,60 ($\pm 0,01$)
BBS08/00	0,23 ($\pm 0,01$)	BBS08/10	0,21 ($\pm 0,01$)
BBS09/00	1,49 ($\pm 0,03$)	BBS09/10	1,35 ($\pm 0,03$)
BBS10/00	1,07 ($\pm 0,04$)	BBS10/10	0,96 ($\pm 0,04$)
BBS11/00	2,04 ($\pm 0,02$)	BBS11/10	1,83 ($\pm 0,02$)
BBS12/00	3,91 ($\pm 0,01$)	BBS12/10	3,52 ($\pm 0,01$)
BBS13/00	1,90 ($\pm 0,02$)	BBS13/10	1,71 ($\pm 0,02$)
BBS14/00	2,21 ($\pm 0,01$)	BBS14/10	1,99 ($\pm 0,01$)
BBS15/00	3,41 ($\pm 0,05$)	BBS15/10	3,07 ($\pm 0,05$)
BBS16/00	0,05 ($\pm 0,04$)	BBS16/10	0,05 ($\pm 0,04$)
BBS17/00	1,39 ($\pm 0,03$)	BBS17/10	1,25 ($\pm 0,03$)
BBS18/00	1,24 ($\pm 0,01$)	BBS18/10	1,12 ($\pm 0,01$)
BBS19/00	1,37 ($\pm 0,04$)	BBS19/10	1,24 ($\pm 0,04$)
BBS20/00	1,44 ($\pm 0,04$)	BBS20/10	1,29 ($\pm 0,04$)
BBS01/05	0,35 ($\pm 0,04$)	BBS01/20	0,30 ($\pm 0,04$)
BBS02/05	0,30 ($\pm 0,02$)	BBS02/20	0,25 ($\pm 0,02$)
BBS03/05	0,82 ($\pm 0,01$)	BBS03/20	0,69 ($\pm 0,01$)

BBS04/05	0,78 ($\pm 0,01$)	BBS04/20	0,66 ($\pm 0,01$)
BBS05/05	0,77 ($\pm 0,01$)	BBS05/20	0,65 ($\pm 0,01$)
BBS06/05	0,39 ($\pm 0,01$)	BBS06/20	0,32 ($\pm 0,01$)
BBS07/05	0,63 ($\pm 0,01$)	BBS07/20	0,53 ($\pm 0,01$)
BBS08/05	0,22 ($\pm 0,01$)	BBS08/20	0,18 ($\pm 0,01$)
BBS09/05	1,42 ($\pm 0,03$)	BBS09/20	1,20 ($\pm 0,03$)
BBS10/05	1,02 ($\pm 0,04$)	BBS10/20	0,86 ($\pm 0,04$)
BBS11/05	1,93 ($\pm 0,02$)	BBS11/20	1,63 ($\pm 0,02$)
BBS12/05	3,72 ($\pm 0,01$)	BBS12/20	3,13 ($\pm 0,01$)
BBS13/05	1,80 ($\pm 0,02$)	BBS13/20	1,52 ($\pm 0,02$)
BBS14/05	2,10 ($\pm 0,01$)	BBS14/20	1,77 ($\pm 0,01$)
BBS15/05	3,24 ($\pm 0,05$)	BBS15/20	2,73 ($\pm 0,05$)
BBS16/05	0,05 ($\pm 0,04$)	BBS16/20	0,04 ($\pm 0,04$)
BBS17/05	1,32 ($\pm 0,03$)	BBS17/20	1,11 ($\pm 0,03$)
BBS18/05	1,18 ($\pm 0,01$)	BBS18/20	0,99 ($\pm 0,01$)
BBS19/05	1,31 ($\pm 0,04$)	BBS19/20	1,10 ($\pm 0,04$)
BBS20/05	1,37 ($\pm 0,04$)	BBS20/20	1,15 ($\pm 0,04$)

*Média das triplicas \pm desvio padrão

Tabela 12: Resultados da determinação de proteínas totais, em amostras de BBS pelo método de referência Kjeldhal.

Código das Amostras	Teor de proteínas Totais (% m/m)*	Código das Amostras	Teor de Proteínas Totais (% m/m)*
BBS01/00	0,58 ($\pm 0,01$)	BBS01/10	0,52 ($\pm 0,01$)
BBS02/00	0,55 ($\pm 0,02$)	BBS02/10	0,50 ($\pm 0,02$)
BBS03/00	0,55 ($\pm 0,01$)	BBS03/10	0,50 ($\pm 0,01$)
BBS04/00	0,53 ($\pm 0,03$)	BBS04/10	0,48 ($\pm 0,03$)
BBS05/00	0,56 ($\pm 0,02$)	BBS05/10	0,50 ($\pm 0,02$)
BBS06/00	0,55 ($\pm 0,02$)	BBS06/10	0,50 ($\pm 0,02$)
BBS07/00	0,59 ($\pm 0,03$)	BBS07/10	0,53 ($\pm 0,03$)
BBS08/00	2,54 ($\pm 0,04$)	BBS08/10	2,29 ($\pm 0,03$)
BBS09/00	0,65 ($\pm 0,01$)	BBS09/10	0,59 ($\pm 0,01$)
BBS10/00	0,59 ($\pm 0,03$)	BBS10/10	0,53 ($\pm 0,02$)
BBS11/00	0,65 ($\pm 0,06$)	BBS11/10	0,58 ($\pm 0,05$)
BBS12/00	0,66 ($\pm 0,08$)	BBS12/10	0,60 ($\pm 0,07$)
BBS13/00	0,58 ($\pm 0,01$)	BBS13/10	0,52 ($\pm 0,01$)
BBS14/00	0,61 ($\pm 0,02$)	BBS14/10	0,55 ($\pm 0,01$)
BBS15/00	2,24 ($\pm 0,01$)	BBS15/10	2,01 ($\pm 0,01$)
BBS16/00	0,57 ($\pm 0,01$)	BBS16/10	0,52 ($\pm 0,01$)
BBS17/00	0,63 ($\pm 0,02$)	BBS17/10	0,56 ($\pm 0,01$)
BBS18/00	0,43 ($\pm 0,02$)	BBS18/10	0,39 ($\pm 0,01$)
BBS19/00	0,45 ($\pm 0,04$)	BBS19/10	0,41 ($\pm 0,04$)

BBS20/00	0,44 ($\pm 0,01$)	BBS20/10	0,40 ($\pm 0,01$)
BBS01/05	0,55 ($\pm 0,01$)	BBS01/20	0,46 ($\pm 0,01$)
BBS02/05	0,53 ($\pm 0,02$)	BBS02/20	0,44 ($\pm 0,02$)
BBS03/05	0,52 ($\pm 0,01$)	BBS03/20	0,44 ($\pm 0,01$)
BBS04/05	0,51 ($\pm 0,03$)	BBS04/20	0,43 ($\pm 0,02$)
BBS05/05	0,53 ($\pm 0,02$)	BBS05/20	0,45 ($\pm 0,02$)
BBS06/05	0,52 ($\pm 0,02$)	BBS06/20	0,44 ($\pm 0,02$)
BBS07/05	0,56 ($\pm 0,03$)	BBS07/20	0,47 ($\pm 0,02$)
BBS08/05	2,41 ($\pm 0,03$)	BBS08/20	2,03 ($\pm 0,03$)
BBS09/05	0,62 ($\pm 0,01$)	BBS09/20	0,52 ($\pm 0,01$)
BBS10/05	0,56 ($\pm 0,02$)	BBS10/20	0,47 ($\pm 0,02$)
BBS11/05	0,61 ($\pm 0,05$)	BBS11/20	0,52 ($\pm 0,05$)
BBS12/05	0,63 ($\pm 0,07$)	BBS12/20	0,53 ($\pm 0,06$)
BBS13/05	0,55 ($\pm 0,01$)	BBS13/20	0,46 ($\pm 0,01$)
BBS14/05	0,58 ($\pm 0,01$)	BBS14/20	0,49 ($\pm 0,01$)
BBS15/05	2,12 ($\pm 0,01$)	BBS15/20	1,79 ($\pm 0,01$)
BBS16/05	0,54 ($\pm 0,01$)	BBS16/20	0,46 ($\pm 0,01$)
BBS17/05	0,60 ($\pm 0,01$)	BBS17/20	0,50 ($\pm 0,01$)
BBS18/05	0,41 ($\pm 0,01$)	BBS18/20	0,35 ($\pm 0,01$)
BBS19/05	0,43 ($\pm 0,04$)	BBS19/20	0,36 ($\pm 0,03$)
BBS20/05	0,42 ($\pm 0,01$)	BBS20/20	0,35 ($\pm 0,01$)

*Média das triplicas \pm desvio padrão. Formação em negrito às selecionadas para o modelo de calibração.

8.2 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS FTIR-ATR NA DETERMINAÇÃO DO TEOR DE GLICÍDIOS TOTAIS EM BBS.

Tabela 13: Resultados das figuras de mérito dos modelos de regressão por PLS, iPLS e siPLS para determinação do teor de glicídios totais.

Modelos ^a	Intervalos selecionados ^b	VLS ^c	RMSECV (% m/m)	R ² _{cv}	RMSEP (% m/m)	R ² _{prev}
PLS	Global	4	0,24	0,9665	0,18	0,8767
PLS	Global	5	0,21	0,9742	0,16	0,8983
PLS	Global	6	0,20	0,9763	0,14	0,9314
PLS	Global	9	0,18	0,9821	0,12	0,9492
iPLS2	2	4	0,22	0,9734	0,14	0,9281
iPLS2	2	5	0,19	0,9781	0,15	0,9128
iPLS2	2	6	0,18	0,9807	0,14	0,9224
iPLS2	2	7	0,17	0,9839	0,13	0,9320
iPLS4	4	4	0,24	0,9679	0,17	0,9172
iPLS4	4	5	0,22	0,9712	0,15	0,9124
iPLS4	4	6	0,21	0,9750	0,15	0,9099
iPLS4	4	9	0,19	0,9803	0,17	0,9012
iPLS8	7	3	0,24	0,9663	0,19	0,8887
iPLS8	7	9	0,23	0,9700	0,15	0,9128
iPLS8	7	11	0,22	0,9726	0,18	0,8714
s2iPLS8	6, 7	6	0,20	0,9779	0,13	0,9337
s2iPLS8	6, 7	7	0,18	0,9809	0,13	0,9382
s2iPLS8	6, 7	8	0,17	0,9833	0,13	0,9403
s3iPLS8	2, 5, 7	6	0,19	0,9781	0,15	0,9241
s3iPLS8	2, 5, 7	7	0,19	0,9785	0,13	0,9341
s3iPLS8	2, 5, 7	10	0,17	0,9833	0,13	0,9419
s4iPLS8	2, 3, 6, 7	6	0,20	0,9775	0,17	0,9016
s4iPLS8	2, 3, 6, 7	7	0,19	0,9783	0,14	0,9239
s4iPLS8	2, 3, 6, 7	10	0,16	0,9849	0,13	0,9415

^a s—número de intervalos combinados iPLS—número de divisões do espectro; ^b Intervalos selecionados; ^c número de variáveis latentes; Formação em negrito ao melhor modelo de regressão.

8.3 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS FTIR-ATR NA DETERMINAÇÃO DO TEOR DE GLICÍDIOS REDUTORES EM BBS.

Tabela 14: Resultados das figuras de mérito dos modelos de regressão por PLS, iPLS e siPLS para determinação do teor de glicídios redutores.

Modelos ^a	Intervalos selecionados ^b	VLS ^c	RMSECV (% m/m)	R ² _{cv}	RMSEP (% m/m)	R ² _{prev}
PLS	Global	4	0,55	0,7106	0,40	0,8105
PLS	Global	7	0,54	0,7242	0,44	0,6862
PLS	Global	9	0,54	0,7358	0,42	0,6742
iPLS2	1	9	0,58	0,6884	0,43	0,7022
iPLS2	1	10	0,56	0,7091	0,43	0,6929
iPLS2	1	11	0,57	0,7056	0,45	0,6459
iPLS4	4	5	0,55	0,7068	0,43	0,7681
iPLS4	4	9	0,52	0,7510	0,52	0,5295
iPLS8	7	3	0,59	0,6717	0,43	0,7628
iPLS8	7	4	0,59	0,6732	0,44	0,7937
iPLS8	7	5	0,60	0,6567	0,41	0,8057
s2iPLS8	7, 8	5	0,55	0,7068	0,43	0,7681
s2iPLS8	7, 8	9	0,52	0,7510	0,52	0,5295
s3iPLS8	2, 3, 7	4	0,54	0,7154	0,41	0,8212
s3iPLS8	2, 3, 7	5	0,54	0,7172	0,43	0,7985
s3iPLS8	2, 3, 7	6	0,56	0,7078	0,44	0,7280
s4iPLS8	1, 2, 3, 7	4	0,54	0,7157	0,41	0,8147
s4iPLS8	1, 2, 3, 7	5	0,55	0,7164	0,44	0,7871
s4iPLS8	1, 2, 3, 7	9	0,55	0,7235	0,40	0,6967

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

8.4 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS FTIR-ATR NA DETERMINAÇÃO DO TEOR DE GLICÍDIOS NÃO REDUTORES EM BBS.

Tabela 15: Resultados das figuras de mérito dos modelos de regressão por PLS, iPLS e siPLS para determinação do teor de glicídios não redutores.

Modelos ^a	Intervalos selecionados ^b	VLS ^c	RMSECV (% m/m)	R ² _{cv}	RMSEP (% m/m)	R ² _{prev}
PLS	Global	3	0,62	0,6025	0,48	0,4590
PLS	Global	8	0,58	0,6688	0,46	0,2649
PLS	Global	9	0,59	0,6597	0,43	0,2983
iPLS2	2	1	0,63	0,5977	0,47	0,6642
iPLS2	2	5	0,61	0,6239	0,50	0,3262
iPLS2	2	6	0,60	0,6300	0,53	0,1547
iPLS4	4	3	0,62	0,6065	0,47	0,3851
iPLS4	4	5	0,61	0,6552	0,51	0,2851
iPLS4	4	9	0,57	0,6731	0,54	0,0878
iPLS8	8	4	0,57	0,6667	0,51	0,3161
iPLS8	8	8	0,56	0,6790	0,56	0,0382
s2iPLS8	5, 8	4	0,58	0,6574	0,51	0,3276
s2iPLS8	5, 8	6	0,59	0,6438	0,51	0,1655
s3iPLS8	4, 5, 7	4	0,58	0,6529	0,50	0,4222
s4iPLS8	1, 4, 5, 7	4	0,58	0,6522	0,51	0,4151

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

8.5 RESULTADOS DAS FIGURAS DE MÉRITO DOS MODELOS DE REGRESSÃO COM DADOS FTIR-ATR NA DETERMINAÇÃO DO TEOR DE PROTEÍNAS TOTAIS EM BBS.

Tabela 16: Resultados das figuras de mérito dos modelos de regressão por PLS, iPLS e siPLS para determinação do teor de proteínas totais.

Modelos ^a	Intervalos selecionados ^b	VLS ^c	RMSECV (% m/m)	R ² _{cv}	RMSEP (% m/m)	R ² _{prev}
PLS	Global	8	0,15	0,9199	0,10	0,9681
PLS	Global	9	0,13	0,9341	0,11	0,9592
PLS	Global	10	0,14	0,9310	0,12	0,9489
PLS	Global	11	0,14	0,9293	0,12	0,9491
iPLS2	1	8	0,16	0,9036	0,18	0,8859
iPLS2	1	9	0,17	0,9012	0,17	0,8955
iPLS2	1	10	0,18	0,8891	0,18	0,8855
iPLS2	1	11	0,20	0,8656	0,18	0,8877
iPLS4	3	8	0,12	0,9409	0,25	0,8096
iPLS4	3	9	0,12	0,9426	0,25	0,8028
iPLS4	3	10	0,12	0,9399	0,25	0,7989
iPLS8	3	8	0,24	0,8132	0,18	0,8819
iPLS8	3	9	0,24	0,8075	0,19	0,9391
iPLS8	3	10	0,24	0,8141	0,21	0,8457
s2iPLS8	3, 6	8	0,16	0,9055	0,13	0,9361
s2iPLS8	3, 6	9	0,18	0,8906	0,13	0,9675
s2iPLS8	3,6	10	0,19	0,9394	0,11	0,9551
s3iPLS8	3, 5, 6	8	0,12	0,9446	0,10	0,9708
s3iPLS8	3, 5, 6	9	0,11	0,9512	0,10	0,9843
s3iPLS8	3, 5, 6	10	0,11	0,9559	0,11	0,9866
s3iPLS8	3, 5, 6	11	0,10	0,9606	0,10	0,9819
s4iPLS8	2, 3, 5, 6	8	0,13	0,9366	0,09	0,9757
s4iPLS8	2, 3, 5, 6	9	0,12	0,9450	0,08	0,9833
s4iPLS8	2, 3, 5, 6	10	0,12	0,9456	0,10	0,9835
s4iPLS8	2, 3, 5, 6	11	0,12	0,9469	0,11	0,9754

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

8.6 RESULTADOS DAS DETERMINAÇÃO DO TEOR DE PROTEÍNAS TOTAIS EM BBS POR FTIR-ATR, COMBINADA COM REGRESSÃO MULTIVARIADA, UTILIZANDO O ALGORITMO KENNARD-STONE PARA SELEÇÃO DAS AMOSTRAS DE CALIBRAÇÃO E PREVISÃO.

Como nos estudos anteriores, foram construídos modelos de regressão por PLS, iPLS e siPLS a partir do conjunto de dados, pré tratados, obtidos por MIR, FTIR-ATR.

Nela é possível verificar as principais figuras de mérito para construir os modelos de regressão multivariada. O erro relativo, para o conjunto das amostras de previsão, também é apresentado, para se ter uma melhor visualização do erro do modelo, para a determinação dos proteínas totais. O erro relativo alto, pode ser devido a grande faixa sem informação já mencionada. A totalidade dos modelos gerados, que foram avaliados, são apresentadas no Apêndice 8.5 (Tabela 16).

Por meio da Tabela 15, é possível verificar os resultados de RMSECV, RMSEP, R^2_{cv} , R^2_{prev} e número de variáveis latentes (VLs) dos melhores modelos desenvolvidos, de cada regressão para cada intervalo espectral selecionado, para a determinação de proteínas totais, obtidos pelo seleção de conjuntos de dados de calibração e previsão pelo algoritmo Kennard-Stone⁵¹. Nela são apresentadas as principais figuras de mérito para a construção dos modelos de regressão multivariada.

Como nos casos anteriores, os modelos foram escolhidos pelos menores valores de RMSECV e RMSEP e maiores coeficientes de correlação (R^2). Para escolher o número de VLs avaliamos o erro dos modelos de calibração interna cruzada RMSECV, tendo cuidado para escolher um número que não causasse subajuste nem superajuste dos modelos. Para avaliar a capacidade de previsão entre os modelos, foi utilizado o erro do conjunto de previsão externa RMSEP.³²

Tabela 17: Resultados das figuras de mérito dos modelos de regressão, para determinação do teor de proteínas totais com dados MIR, FTIR-ATR.

Modelos ^a	Intervalos selecionados ^b	VLs ^c	RMSECV (% m/m)	R ² _{cv}	RMSEP (% m/m)	R ² _{prev}	Erro Relativo de previsão (% m/m)
PLS	Global	9	0,13	0,9516	0,0916	0,68	13,25
iPLS2	1	6	0,18	0,9065	0,1130	0,36	18,76
iPLS4	2	7	0,19	0,8961	0,1114	0,49	21,95
iPLS8	3	4	0,21	0,8696	0,1179	0,72	20,33
s2iPLS8	3, 6	8	0,17	0,9181	0,1100	0,64	21,24
s3iPLS8	3, 5, 6	9	0,11	0,9659	0,0941	0,61	15,16
s4iPLS8	2, 3, 5, 6	8	0,11	0,9659	0,0951	0,65	16,94

^a s—número de intervalos combinados iPLS—número de divisões do espectro; ^b Intervalos selecionados; ^c número de variáveis latentes; Formação em negrito ao melhor modelo de regressão.

A partir dos resultados das regressões usando os métodos PLS, iPLS e siPLS, podemos observar, que para ambos os modelos, tivemos baixos valores de RMSECV e RMSEP, mas valores de R² discrepantes, quando comparados o R²_{cal} e o R²_{prev}. Foram obtidos resultados satisfatórios para o R²_{cal}, mas o mesmo não ocorreu com o R²_{prev}.

Mesmo assim um fator positivo, que podemos observar, é que as regiões selecionadas no espectro para a construção dos modelos, correspondem a regiões importantes que contém nitrogênio como a das amidas e aminas secundárias. Nos modelos de regressão, é desejado que as regiões selecionadas, tenham relação com a composição química. As amidas secundárias, presentes nas proteínas, apresentam uma banda muito forte (longa) do estiramento do grupo carbonila (C=O), que aparece na faixa de 1680 a 1630 cm⁻¹, a banda sobrepõe, parcialmente, a banda de dobramento N-H, de amidas secundárias, que aparece na faixa de 1640 a 1550 cm⁻¹. Essas faixas, estão presentes em todos os modelos⁵⁸. Na faixa de 1350 a 1000 cm⁻¹, ocorre o estiramento C-N⁵⁸, e uma faixa da mesma, esta selecionada nos modelos s2iPLS8, s3iPLS8 e s4iPLS8.

Assim, apesar do modelo não ter levando em consideração uma faixa mais diversificada do teor de proteínas totais, ele obteve baixos valores de erros RMSECV e RMSEP, e selecionou regiões características do espectro para proteínas, obtendo um coeficiente de determinação satisfatório na etapa de calibração.

Realizadas as considerações, pode-se exemplificar que o modelo com melhores resultados, para a determinação de proteínas totais, expressos em % (m/m), em BBS foi o s4iPLS8. Ele apresentou, no conjunto, melhores valores de RMSECV, RMSEP, R^2_{cal} e R^2_{prev} . Foi possível verificar que alguns modelos apresentaram resultados muito semelhantes, entre os modelos gerados (Tabela 15). O s2iPLS8 e o s3iPLS8 são um exemplos de modelos com resultados das figuras de mérito bem próximo ao modelo escolhido. Esses modelos possuem uma pequena diferença para os erros de calibração e previsão e do coeficiente de determinação de calibração e previsão, comparado com o modelo escolhido s4iPLS8. Avaliando as variáveis latentes podemos verificar que os modelos de s2iPLS8 e s4iPLS8 possuem 8 cada e o modelo s3iPLS8 possui 9. Por utilizar mais variáveis latentes, o modelo s3iPLS8 não foi selecionado. Portanto, os modelos s2iPLS8 e s4iPLS8 apresentam resultados considerados melhores, dentro das condições, e a escolha pelo modelo s4iPLS8 se deu pelos melhores resultados do modelo para a previsão.

O modelo PLS, aplicado no espectro todo, apresentou melhores valores de coeficientes de determinação e erros para a previsão (RMSEP e R^2_{prev}) comparados aos do modelo s4iPLS8. Mas, esse modelo não foi selecionado, como exemplo, por utilizar maior número de variáveis latentes, 9 se compararmos ao modelo escolhido que utiliza 8 VLs, e por maiores diferenças entre os valores de RMSECV e RMSEP.

O modelo usado como exemplo, foi o s4iPLS8, que dividiu o espectro em 8 partes e combinou os subintervalos selecionados 2, 3, 5 e 6, que correspondem aos intervalos espectrais $1775 - 1650 \text{ cm}^{-1}$, $1649 - 1525 \text{ cm}^{-1}$, $1399 - 1275 \text{ cm}^{-1}$, $1274 - 1150 \text{ cm}^{-1}$, com 8 VLs, conforme ilustrado na Figura 24. Apresentou RMSECV = 0,11%, RMSEP = 0,65% e coeficientes de determinação (R^2_{cv} e R^2_{prev}) de 0,9659 e 0,6490, respectivamente. Os intervalos selecionados, englobam os estiramento do grupo carbonila (C=O) de amidas secundárias, faixa de $1680 \text{ a } 1630 \text{ cm}^{-1}$, a banda de dobramento N-H, de amidas secundárias, na faixa de $1640 \text{ a } 1550 \text{ cm}^{-1}$ e, e os estiramentos C-N no intervalo de $1350 \text{ a } 1000 \text{ cm}^{-1}$,⁵⁸ indicando que esta é uma faixa representativa na determinação das proteínas totais nas amostras de bebidas a base de soja, ao selecionar regiões específicas de absorção.

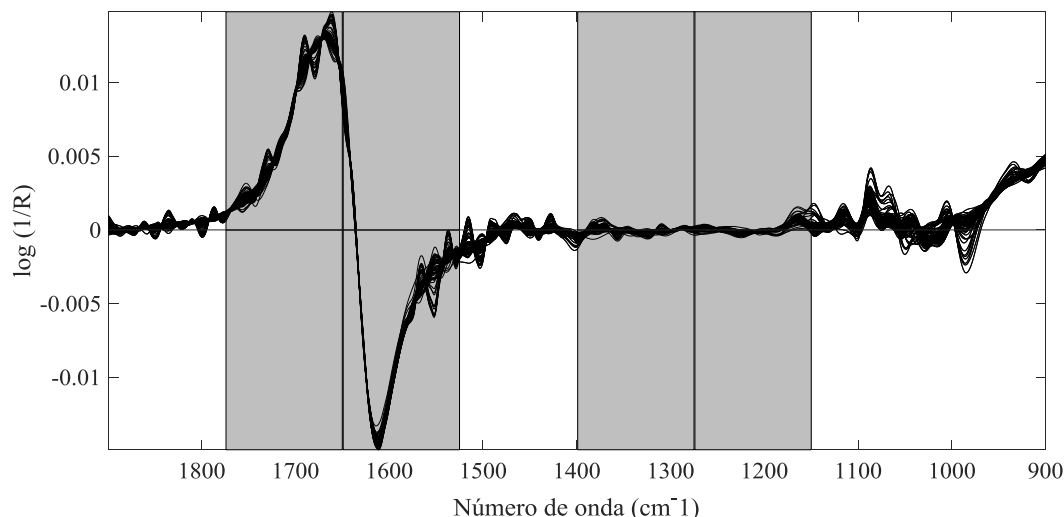


Figura 24: Espectro com destaque as regiões selecionadas (2, 3, 5, e 6) combinadas com regressão siPLS para determinação de proteínas totais em BBS.

A definição do número de variáveis latentes é de fundamental importância para que não ocorra o sobreajuste ou subajuste, sendo pela Figura 25, possível justificar a escolha pelo número de 8 VLs do modelo s4iPLS8, possuindo menor erro de calibração, determinado por validação cruzada (RMSECV), e a partir desse número de variáveis não ocorre decréscimo significativo do erro e nem uma melhora significativa do R^2_{cal} . Vale ressaltar que a matriz de análise, a BBS, é bastante complexa, tendo a presença de diferentes substâncias como proteínas, vitaminas, gorduras, fibras, metais, aromatizantes, corantes e glicídios como frutose, glicose, sacarose, gomas entre outros carboidratos, o que pode elevar o número de VLs do modelo.

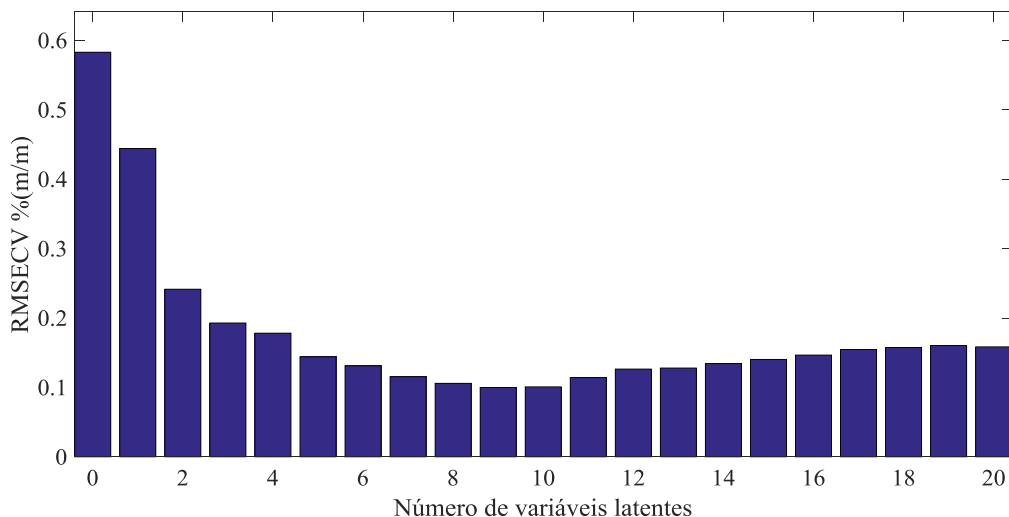


Figura 25: Gráfico dos valores de RMSECV por número de variáveis latentes utilizadas para determinação de proteínas totais, por meio de espectroscopia no infravermelho, combinando os intervalos 2, 3, 5 e 6 no modelo s4iPLS8.

Também foi construído um gráfico de correlação dos valores de referência, obtidos pelo método oficial para o teor de proteínas totais, com os valores obtidos pelo melhor modelo de regressão através do FTIR-ATR conforme ilustrado na Figura 26. Pode-se avaliar que o modelo possui valores considerados satisfatórios de coeficiente de determinação para as amostras de calibração e valores de RMSECV e RMSEP baixos e pouco discrepantes. Portanto, é possível constatar que esse modelo para determinação de proteínas em BBS, tem potencial, mas para obter melhores resultados, necessita-se de mais amostras e com uma faixa diversificada de concentração.

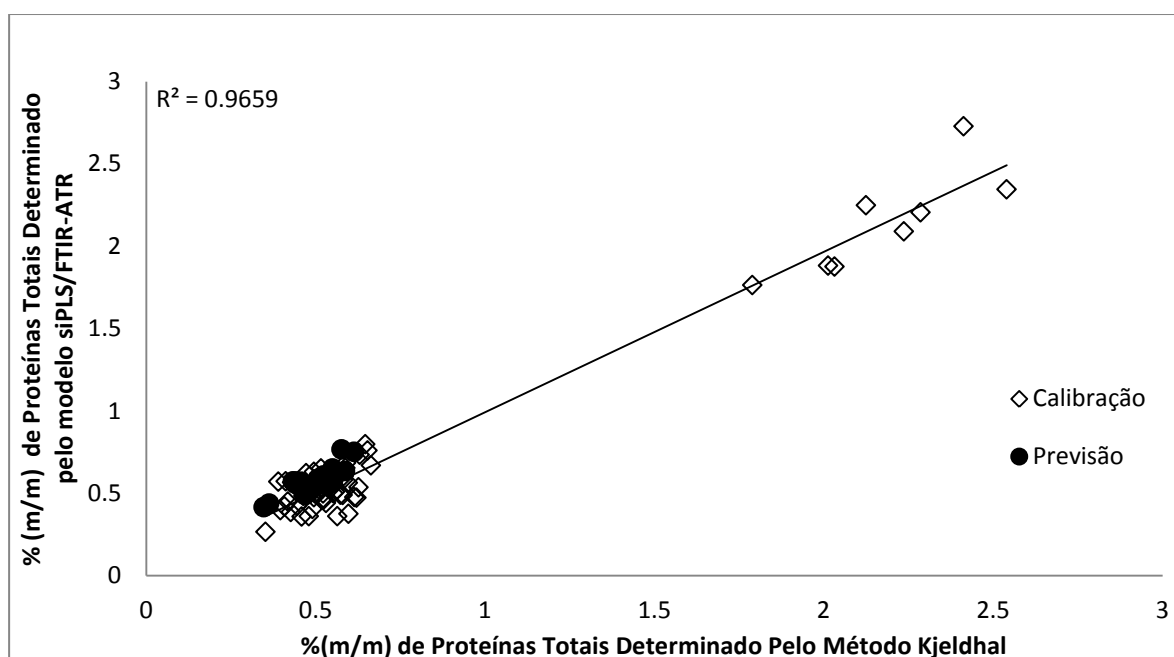


Figura 26: Valores de referência pelo método Kjeldhal *versus* valores previstos do modelo siPLS para proteínas totais usando os intervalos 2, 3, 5 e 6 correspondentes aos intervalos espectrais $1775 - 1650 \text{ cm}^{-1}$, $1649 - 1525 \text{ cm}^{-1}$, $1399 - 1275 \text{ cm}^{-1}$, $1274 - 1150 \text{ cm}^{-1}$.

8.6.1.1 Considerações quanto ao modelo de regressão multivariada a partir de dados FTIR-ATR para determinação do teor de proteínas totais em BBS, utilizando o algoritmo Kennard-Stone para seleção das amostras de calibração e previsão.

Com base nesses resultados, o modelo que apresentou poder ser o mais adequado para a determinação de proteínas totais, expressos em % (m/m) foi o s4iPLS8 com 8 VLs. Como mostrado no gráfico número de LVs versus RMSECV (Figura 25), o erro sofre uma grande redução, indicando que 8 VLs é um número adequado, para o modelo estudado, pois a diminuição do erro é inexpressível quando se incluem novas VLs no modelo e o que também, poderia super ajustá-lo.

Outro fator que justifica a escolha do modelo como exemplo, é a região espectral selecionada, que incluem os intervalos de absorção característico do estiramento do grupo carbonila (C=O) de amidas secundárias, do dobramento N-H, de amidas secundárias, e os estiramentos C-N.

A Figura 26 apresenta a correlação entre o valor real e o valor previsto para os modelo siPLS desenvolvido, com os subintervalos 2, 3, 5 e 6, que se mostrou o modelo com a melhor capacidade de determinar a % (m/m) de proteínas totais em glicose presentes em bebidas a base de soja utilizando o algoritmo Kennard-Stone para seleção das amostras de calibração e previsão.