

**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA
LAQQA- LABORATÓRIO DE QUIMIOMETRIA EM
QUÍMICA ANALÍTICA**



DISSERTAÇÃO DE MESTRADO

**DETERMINAÇÃO DE PROTEÍNA BRUTA E AMINOÁCIDOS EM
FARELO DE SOJA POR ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO**

***AUTORA: Rosineide Costa Simas
ORIENTADOR : Prof. Dr. Ronei Jesus Poppi***

UNICAMP , JUNHO DE 2005

FICHA CATALOGRÁFICA ELABORADA PELA
**BIBLIOTECA DO INSTITUTO DE QUÍMICA DA
UNICAMP**

Si42d

Simas, Rosineide Costa.

Determinação de proteína bruta e aminoácidos em farelo de soja por espectroscopia no infravermelho próximo / Rosineide Costa Simas. -- Campinas, SP: [s.n], 2005.

Orientador: Ronei Jesus Poppi.

Dissertação – Universidade Estadual de Campinas, Instituto de Química.

1. Proteína. 2. Farelo de soja. 3. Aminoácidos.
4. Infravermelho próximo. I. Poppi, Ronei Jesus. II.
Instituto de Química. III. Título.

À minha avó Maria das Virgens Araújo (in memoriam)
Por ter me educado e mostrado que a dignidade, a honestidade,
a perseverança e a vontade são os grandes segredos do sucesso...

“... porque não morre quem nos olhos vive,
não morre quem nos vivos vive...”

AGRADECIMENTOS

Ao professor, amigo e orientador Ronei Jesus Poppi por ter me conduzido no caminho da pesquisa, pela paciência , pelo voto de confiança , pela liberdade de expressão e pensamento que fez desse trabalho científico um exercício de realização profissional e pessoal;

Aos colegas do grupo LAQQA antigos e novos, especialmente Jez, Alessandra, Marcelo Sena, Waldomiro e Patrícia. Agradeço o apoio, dicas e paciência que me ajudaram na realização deste momento;

Aos professores Roy Bruns, Márcia M. Castro Ferreira, Célio Pasquini, Celso Davanzo, César Mello e especialmente ao professor Antônio Marsaioli Júnior (FEA), pelo apoio, ajuda e por acreditar que seria possível apresentar bons resultados;

À Polimate , Foss, Embrapa , Adisseo, CBO, Mogiana Alimentos e Labtec pela atenção no fornecimento de dados e material científico. Especialmente aos amigos da Adisseo Washington e Márcio Ceccantini, do CBO Oneida e Eder (CBO/DEPAN) e aos analistas do Labtec Meire , Silvana e Edinilson por toda ajuda prestada;

À minha pequena e amada família (Dona Margarida: a melhor mãe do mundo e Calixto: o grande Caco, meu irmão e amigo de todas as horas) minha maior força , incentivo e com certeza meu maior tesouro e orgulho...

Ao pessoal do laboratório da Nutron antigos e novos por incentivar, acreditar e apoiar meu sonho e por ser meu maior fã clube;

Ao meu primo Adriano Araújo da Silva (o Juninho), que sempre me ajuda em tudo;

Aos amigos Cristina, Daniel, Júlio e Renato que sempre estiveram comigo;

À Aldair de Souza Santos (o *Daio*) por ter sido meu referencial e ter me ajudado quando eu não podia mais continuar sozinha, com seu jeito sábio e bem humorado de ensinar, doutor na ciência da boa vontade e mestre na amizade...

Ao amigo e maior responsável pelo meu sucesso Edson Brás de Brito (o vizinho). Que me deu apoio financeiro e principalmente humano, investindo num sonho pelo simples fato de ajudar alguém que só precisava de incentivo para seguir em frente quando as portas se fecharam, meus mais sinceros agradecimentos , consideração e respeito;

À toda comunidade científica e funcionários da UNICAMP. E a todas as pessoas que de alguma forma contribuíram para esse trabalho;

Agradeço ao meu grande e poderoso DEUS, minha força, meu consolo e auxílio bem presente na hora das minhas maiores dificuldades, por ter colocado todas essas pessoas maravilhosas no meu caminho , por estar me ajudando hoje e sempre tendo me revelado à sua graça ...

SIMAS, R.C., 2005. Determinação de proteína bruta e aminoácidos em farelo de soja por espectroscopia no infravermelho próximo. Campinas: tese de mestrado – UNICAMP. 119 p.

RESUMO:

DETERMINAÇÃO DE PROTEÍNA BRUTA E AMINOÁCIDOS EM FARELO DE SOJA POR ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO

Foram determinados neste trabalho a quantidade de proteína bruta e aminoácidos em farelo de soja, empregando-se espectroscopia de reflectância no infravermelho próximo (NIR) e calibração multivariada baseada no método dos mínimos quadrados parciais (PLS). Foram utilizadas 150 amostras provenientes de várias regiões brasileiras divididas em 15 fornecedores diferentes entre eles Bunge, Caramuru, Cocamar, Cargill, Ceval e ADM que garantem uma boa abrangência e representatividade das amostras. Os espectros foram coletados na faixa de 1300 à 2400 nm, obtendo-se 551 variáveis por espectro . As análises de proteína bruta foram determinadas através do analisador de nitrogênio LECO FP-428 segundo o método de referência AACC 946-30 (*American Association of Cereal Chemists*,1995). As determinações de aminoácidos foram realizadas no equipamento Hitachi L-8500, utilizando a derivatização pós-coluna com hidrólise ácida com HCl 6 mol/L para todos os aminoácidos exceto triptofano analisado por hidrólise alcalina com hidróxido de lítio 4 mol/L e os aminoácidos sulfurados metionina e a cistina analisados por oxidação com ácido perfórmico, segundo método proposto para um estudo colaborativo do jornal da AOAC (*Association of Official Analytical Chemists*). A avaliação dos modelos mostra que foi possível determinar com confiabilidade proteína bruta, ácido glutâmico, ácido aspártico, arginina, treonina, valina, leucina, tirosina, triptofano, cistina e metionina. Os aminoácidos lisina, alanina, isoleucina e fenilalanina cujas faixas de calibração foram menores que 3 vezes o desvio padrão do método de referência, estão fora das recomendações da norma ASTM E1655-00, essas faixas devem ser aumentadas e os modelos aperfeiçoados. Já os modelos para os aminoácidos histidina, serina, glicina e prolina devem ser desconsiderados e refeitos, porém com outro método de referência, como por exemplo o de derivatização pré-coluna, pois mesmo aumentando a faixa de valores das amostras, o desvio padrão do método de referência (pós-coluna) é alto e não pode satisfazer a condição recomendada pela ASTM. Na maioria dos casos, foram obtidos bons resultados em farelo de soja por modelos de calibração para proteína bruta e aminoácidos desenvolvidos por espectroscopia no infravermelho próximo, além do que esse método também tem a vantagem de ser mais rápido, econômico e não gerar resíduos tóxicos.

SIMAS, R.C., 2005. Determination of crude protein and amino acids in soybean meal by Near Infrared Spectroscopy. Campinas: master degree – UNICAMP. 119 p.

ABSTRACT

DETERMINATION OF CRUDE PROTEIN AND AMINO ACIDS IN SOYBEAN MEAL BY NEAR INFRARED SPECTROSCOPY

In this work it was determined the amount of crude protein and amino acids in soybean meal by using near infrared spectroscopy and multivariate calibration base on partial least squares method (PLS). It was used 150 samples from different brazilian regions divided in several suppliers as Bunge, Caramuru, Cargill, Ceval and ADM that insure a good distribution and representivity of the samples. The spectra were collected in the range of 1300 to 2400 nm, obtaining 551 variables. The crude protein analysis were performed by a nitrogen analyser LECO FP-428 following the reference method AACC 946-30 (American Association of Cereal Chemists ,1995). The amino acids determination were performed by Hitachi L-8500 equipament by using post column derivatization with acid hidrolisis (HCl 6 mol/L) for all amino acids, except tryptophan analysed by alcalin hydrolisis (LiOH 4 mol/L) and the methionine and cystine by performic acid oxidation, following ia collaborative study of the journal of AOAC (Association of Official Analytical Chemists). The models valuation show that is possible to determine with confidence crude protein, glutamic acid, aspartic acid, arginine, threonine, valine, leucine, tyrosine, tryptophan, cystine and methionine. The amino acids lysine, alanine, isoleucine and phenylalanine that present concentration ranges least than three times the standard deviation of reference method are out of recomendation of ASTM E1655-00 norm, and these ranges must be increased and the models improved. By other hand, the models for the amino acids histidine, serine, glycine and proline must be descarted and rebuilt, since the increase of concentration range of samples in not enough due the standard deviation of reference method employed (post column) is high and it can not satisfy the ASTM recommendation, then another reference method, as pre column derivation, must be used. In the majory of the cases, it was obtained good results for crude protein and several amino acids for the models developed by using near infrared spectroscopy, with the advantage that this methodology is very fast, economic and not produce toxic residues.

ÍNDICE DE TABELAS

Tabela 1: Dados sobre farelo de soja (ano comercial –1000 toneladas).....	11
Tabela 2 : Classificação das frações protéicas da soja.....	15
Tabela 3: Range aproximado dos valores de aminoácidos de farelo em soja (43 à 54% de PB).....	37
Tabela 4 : Estudo de desvio padrão e coeficiente de variação para aminoácidos em farelo de soja.....	39
Tabela 5: Dados de calibração dos modelos PLS de proteína bruta e aminoácidos.....	45
Tabela 6: Dados de validação dos modelos PLS de proteína bruta e aminoácidos.....	46
Tabela 7 : Dados do modelo PLS para proteína bruta em farelo de soja.....	50
Tabela 8: Valores e estatísticas do conjunto de validação externa PLS para proteína bruta.....	52
Tabela 9: Dados do modelo PLS para ácido aspártico em farelo de soja.....	55
Tabela 10: Valores e estatísticas do conjunto de validação externa PLS para ASP....	57
Tabela 11: Dados do modelo PLS para ácido glutâmico em farelo de soja.....	58
Tabela 12: Valores e estatísticas do conjunto de validação externa PLS para GLU.....	59
Tabela 13: Dados do modelo PLS para valina em farelo de soja.....	60
Tabela 14: Valores e estatísticas do conjunto de validação externa PLS para VAL...62	
Tabela 15: Dados do modelo PLS para arginina em farelo de soja.....	63
Tabela 16: Valores e estatísticas do conjunto de validação externa PLS para ARG...65	
Tabela 17: Dados do modelo PLS para tirosina em farelo de soja.....	66
Tabela 18: Valores e estatísticas do conjunto de validação externa PLS para TYR.....	67
Tabela 19: Dados do modelo PLS para cistina em farelo de soja.....	68
Tabela 20: Valores e estatísticas do conjunto de validação externa PLS para CYS.....	70
Tabela 21: Dados do modelo PLS para metionina em farelo de soja.....	71
Tabela 22: Valores e estatísticas do conjunto de validação externa PLS para MET....	72
Tabela 23: Dados do modelo PLS para leucina em farelo de soja.....	73
Tabela 24: Valores e estatísticas do conjunto de validação externa PLS para LEU.....	75
Tabela 25 : Dados do modelo PLS para triptofano em farelo de soja.....	76

Tabela 26: Valores e estatísticas do conjunto de validação externa PLS para TRP...	77
Tabela 27: Dados do modelo PLS para treonina em farelo de soja.....	78
Tabela 28: Valores e estatísticas do conjunto de validação externa PLS para THR.....	80
Tabela 29: Dados do modelo PLS para lisina em farelo de soja.....	81
Tabela 30: Valores e estatísticas do conjunto de validação externa PLS para LYS.....	83
Tabela 31: Dados do modelo PLS para fenilalanina em farelo de soja.....	84
Tabela 32: Valores e estatísticas do conjunto de validação externa PLS para PHE...	85
Tabela 33: Dados do modelo PLS para isoleucina em farelo de soja.....	86
Tabela 34: Valores e estatísticas do conjunto de validação externa PLS para ILE.....	88
Tabela 35: Dados do modelo PLS para alanina em farelo de soja.....	89
Tabela 36: Valores e estatísticas do conjunto de validação externa PLS para ALA.....	90
Tabela 37: Dados do modelo PLS para glicina em farelo de soja.....	92
Tabela 38: Dados do modelo PLS para prolina em farelo de soja.....	94
Tabela 39: Dados do modelo PLS para serina em farelo de soja.....	97
Tabela 40: Dados do modelo PLS para histidina em farelo de soja.....	99

ÍNDICE DE GRÁFICOS

Gráfico1: Produção de soja no Brasil de 2000 à 2003 (azul) e previsão estimada para 2004 e 2005 (vermelho) segundo a ABIOVE.....	10
Gráfico 2: Composição média do grão de soja e seus principais subprodutos.....	10
Gráfico 3: Histograma das % de PB do modelo de calibração	36
Gráfico 4- Histogramas da % de PB do modelo de validação.....	37
Gráfico 5 - Espectros originais dos conjuntos de calibração para amostras de farelo de soja.....	40
Gráfico 6 : Espectros de calibração após o pré tratamento completo dos dados.....	41
Gráfico 7: Gráfico dos escores do PCA para os primeiros 3 componentes principais...41	
Gráfico 8: Pesos do modelo de PB, que seguem a mesma tendência para todos modelos demonstrados no primeiro estudo, com menos ou mais variáveis latentes	48
Gráfico 9: Coeficientes de regressão B do modelo de PB, que são idênticos para os 18 modelos apresentados no primeiro estudo	49
Gráfico 10: Pesos do modelo PLS para proteína bruta em farelo de soja.....	51
Gráfico 11 : Coeficientes de regressão B do modelo PLS para proteína bruta em farelo de soja.....	51
Gráfico 12: Pesos do modelo PLS para ácido aspártico em farelo de soja.....	56
Gráfico13: Coeficientes de regressão B do modelo PLS para ácido aspártico em farelo de soja.....	56
Gráfico 14: Pesos do modelo PLS para ácido glutâmico em farelo de soja.....	58
Gráfico 15: Coeficientes de regressão B do modelo PLS para ácido glutâmico em farelo de soja.....	59
Gráfico 16: Pesos do modelo PLS para valina em farelo de soja.....	61
Gráfico 17: Coeficientes de regressão B do modelo PLS para valina em farelo de soja.....	61
Gráfico 18: Pesos do modelo PLS para arginina em farelo de soja.....	64
Gráfico 19: Coeficientes de regressão B do modelo PLS para arginina em farelo de soja.....	64

Gráfico 20: Pesos do modelo PLS para tirosina em farelo de soja.....	66
Gráfico 21: Coeficientes de regressão B do modelo PLS para tirosina em farelo de soja.....	67
Gráfico 22: Pesos do modelo PLS para cistina em farelo de soja.....	69
Gráfico 23: Coeficientes de regressão B do modelo PLS para cistina em farelo de soja.....	69
Gráfico 24: Pesos do modelo PLS para metionina em farelo de soja.....	71
Gráfico 25: Coeficientes de regressão B do modelo PLS para metionina em farelo de soja.....	72
Gráfico 26: Pesos do modelo PLS para Leucina em farelo de soja.....	74
Gráfico 27 : Coeficientes de regressão B do modelo PLS para leucina em farelo de soja.....	74
Gráfico 28: Pesos do modelo PLS para triptofano em farelo de soja.....	76
Gráfico 29: Coeficientes de regressão B do modelo PLS para triptofano em farelo de soja.....	77
Gráfico 30: Pesos do modelo PLS para treonina em farelo de soja.....	79
Gráfico 31: Coeficientes de regressão B do modelo PLS para treonina em farelo de soja.....	79
Gráfico 32: Pesos do modelo PLS para lisina em farelo de soja.....	82
Gráfico 33: Coeficientes de regressão B do modelo PLS para lisina em farelo de soja.....	82
Gráfico 34: Pesos do modelo PLS para fenilalanina em farelo de soja.....	84
Gráfico 35: Coeficientes de regressão B do modelo PLS para fenilalanina em farelo de soja.....	85
Gráfico 36: Pesos do modelo PLS para isoleucina em farelo de soja.....	87
Gráfico 37: Coeficientes de regressão B do modelo PLS para isoleucina em farelo de soja.....	87
Gráfico 38: Pesos do modelo PLS para alanina em farelo de soja.....	89
Gráfico 39: Coeficientes de regressão B do modelo PLS para alanina em farelo de soja.....	90
Gráfico 40: Pesos do modelo PLS para glicina em farelo de soja.....	92

Gráfico 41: Coeficientes de regressão B do modelo PLS para glicina em farelo de soja.....	93
Gráfico 42: Pesos do modelo PLS para prolina em farelo de soja.....	95
Gráfico 43: Coeficientes de regressão B do modelo PLS para prolina em farelo de soja.....	95
Gráfico 44: Pesos do modelo PLS para serina em farelo de soja.....	97
Gráfico 45: Coeficientes de regressão B do modelo PLS para serina em farelo de soja.....	97
Gráfico 46: Pesos do modelo PLS para histidina em farelo de soja.....	99
Gráfico 47: Coeficientes de regressão B do modelo PLS para histidina em farelo de soja.....	99
Gráfico 48 : Comparativo do teor de aminoácidos para um farelo de soja normal.....	100
Gráfico 49: Comparativo do teor de aminoácidos para um farelo de soja contaminado com uréia.....	101
Gráfico 50: Comparativo do teor de aminoácidos para um farelo de soja superaquecido.....	102
Gráfico 51: Espectros dos 3 farelos de soja utilizados nos comparativos. (a amostra 1 e 2 correspondem a amostra normal e contaminada com uréia respectivamente. A amostra 3 corresponde ao farelo super aquecido).....	103

ÍNDICE DE FIGURAS

Figura 1: Representação esquemática da decomposição por PCA.....	23
Figura 2 : Vizinho mais próximo NH e H global.....	34

ÍNDICE GERAL

Prefácio.....	01
 Capítulo 1: Nutrição animal, farelo de soja, proteínas e aminoácidos: uma visão geral.....	 06
1.1 A soja e seus subprodutos.....	09
1.2 Farelo de soja.....	11
1.3 Proteínas e aminoácidos.....	12
1.4 A importância das proteínas e aminoácidos na nutrição animal.....	13
1.5 As proteínas do farelo de soja : estrutura e composição.....	14
1.6 Os aminoácidos do farelo de soja.....	15
 Capítulo2: Fundamentos de Espectroscopia no Infravermelho Próximo e Quimiometria.....	 19
2.1 Princípios da Espectroscopia no Infravermelho Próximo (NIR).....	21
2.2 Análise de Componentes Principais(PCA).....	23
2.3 Calibração Multivariada.....	24
2.4 Regressão por Mínimos Quadrados Parciais(PLS).....	25
 Capítulo 3: Desenvolvimento e avaliação dos modelos de regressão PLS.....	 28
3.1 Monitoramento e performance dos modelos de calibração multivariada.....	30
3.1.1 Medidas de Tendência Central.....	30
3.1.2 Medidas de dispersão.....	31
3.1.3 Medidas de qualidade do modelo.....	31
3.1.4 Identificação de anomalias (medidas de H e T).....	33
3.2 Desenvolvimento dos modelos de calibração PLS.....	34
3.2.1 Amostragem.....	35
3.2.2 Aquisição de espectros e análise de laboratório.....	35

3.2.3 Métodos de Referência.....	38
3.2.4 Pré-tratamento dos espectros.....	40
3.2.5 Análise de PCA dos dados de Calibração.....	41

Capítulo 4: Avaliação dos modelos de regressão PLS para proteína bruta e aminoácidos em farelo de soja.....42

4.1 Primeiro Estudo: espectro completo.....	44
4.2 Segundo Estudo: seleção de variáveis com ênfase na proteína bruta do farelo de soja.....	49
4.3 Terceiro Estudo: seleção de variáveis com ênfase nos aminoácidos do farelo de soja.....	53
4.3.1– Aminoácidos Determinados.....	54
4.3.1.1 Aminoácido Ácido Áspartico.....	54
4.3.1.2 Aminoácido Ácido Glutâmico.....	57
4.3.1.3 Aminoácido Valina.....	60
4.3.1.4 Aminoácido Arginina.....	62
4.3.1.5 Aminoácido Tirosina.....	65
4.3.1.6 Aminoácido Cistina.....	68
4.3.1.7 Aminoácido Metionina.....	70
4.3.1.8 Aminoácido Leucina.....	73
4.3.1.9 Aminoácido Triptofano.....	75
4.3.1.10 Aminoácido Treonina.....	78
4.3.2 Aminoácidos Não Determinados.....	80
4.3.2.1 Aminoácidos cuja faixa de concentração precisa ser ampliada.....	80
4.3.2.1.1 Aminoácido Lisina.....	80
4.3.2.1.2 Aminoácido Fenilalanina.....	83
4.3.2.1.3 Aminoácido Isoleucina.....	86
4.3.2.1.4 Aminoácido Alanina.....	88
4.3.2.2 Aminoácidos cujo método de referência precisa ser modificado.....	91
4.3.2.2.1 Aminoácido Glicina.....	91
4.3.2.2.2 Aminoácido Prolina.....	93

4.3.2.2.3 Aminoácido Serina.....	95
4.3.2.2.4 Aminoácido Histidina.....	98
4.4 Vantagens do uso de modelos desenvolvidos via Espectroscopia no Infravermelho Próximo sobre as tabelas nutricionais de aminoácidos.....	100
Conclusões.....	104
Perspectivas Futuras.....	108
Referências Bibliográficas.....	111

PREFÁCIO

Hoje, a produção eficiente e formulações que possibilitem o máximo de performance animal são muito importantes. O conhecimento do valor de proteína bruta, assim como de aminoácidos nos ingredientes de uma ração balanceada possibilitam ótimos resultados em campo. A determinação do teor de aminoácidos é fundamental visto que os animais não podem sintetizar todos os aminoácidos que necessitam para formar suas próprias proteínas. Por isso, é importante fornecer uma dieta que contenha as quantidades apropriadas para o crescimento e manutenção do organismo e um dos únicos mecanismos para conhecer a composição em aminoácidos de um alimento ou ingrediente é a análise química.

O grande problema para o nutricionista é reduzir as incertezas de uma formulação traduzidas por margens de segurança elevadas, super formulações e desperdício de recursos que levam sempre ao aumento de custos.

As análises clássicas para determinação de proteína bruta são geralmente demoradas e geram grande quantidade de resíduos agressivos ao meio ambiente como o consagrado método de Kjeldahl ⁽¹⁾, ou são de elevado custo como o método de combustão que utiliza um detector de condutividade térmica ⁽²⁾.

Os aminoácidos são determinados por cromatografia líquida de alta eficiência (do inglês *High Performace Liquid Chromatography* , *HPLC*) sendo que os métodos oficiais de derivatização pré ou pós- coluna são demorados e de custos extremamente altos.

Existem tabelas nutricionais que também são utilizadas para se conhecer a composição média em aminoácidos de um ingrediente desde que se conheça previamente o teor de proteína bruta do mesmo. Contudo, essas tabelas podem dar apenas uma idéia do nível de aminoácido , pois ingredientes têm variações muitas vezes maiores que 10% no valor do conteúdo de aminoácidos e, em alguns casos, verifica-se variações maiores que 20% (lisina em farelo de soja de diferentes qualidades, por exemplo) e utilizar a média de um valor fixo pode implicar em até 50% de erro numa formulação para nutrição animal.

Diante de tal situação, o estudo de análises rápidas , confiáveis , de baixo custo e que gerem uma menor quantidade de resíduos tóxicos para o ramo de agronegócio é hoje uma área de grande interesse. Por isso, esse trabalho vem propor uma

metodologia analítica secundária, com todas essas características desejáveis para a análise de proteína bruta e de alguns aminoácidos no ingrediente farelo de soja que compõe cerca de até 40% das formulações para nutrição animal pela técnica de espectroscopia no infravermelho próximo (do inglês *Near Infrared, NIR*), que será aqui chamada de método NIR de análise.

O método NIR pode ser empregado na determinação de substâncias orgânicas (como proteínas, aminoácidos, lipídios e carboidratos) que são detectados na região de 780 à 2500 nm.

A Quimiometria, uma ciência que nasceu na década de 70 dentro da química analítica para possibilitar o tratamento e compreensão da grande quantidade de dados analíticos complexos, quando utilizada com o NIR, pode fornecer modelos de calibração que relacionam a medida espectral obtida no infravermelho próximo com uma propriedade química determinada analiticamente por um método de referência. Para essa análise multivariada vai ser utilizada a técnica de regressão por mínimos quadrados parciais (do inglês *Partial Least Squares, PLS*), considerada padrão em química analítica.

Foi escolhido o equipamento da NIRSystem modelo 5000, por ser amplamente utilizado nas indústrias de nutrição animal, humana e farmacêutica, juntamente com o software Winisi versão 1.5 da ISI (*Infrasoft Internacional*), que utiliza alguns parâmetros específicos para garantir a validação e a performance dos modelos de calibração desenvolvidos.

Assim essa dissertação foi dividida em quatro capítulos, além das conclusões, perspectivas futuras e referências bibliográficas.

O primeiro capítulo chamado *Nutrição animal, farelo de soja, proteínas e aminoácidos: uma visão geral*, trata dos fundamentos teóricos da nutrição animal bem como da origem, processamento e importância do farelo de soja no Brasil. Além de salientar também a importância das proteínas de soja e sua composição em aminoácidos nas formulações para nutrição animal.

O segundo capítulo: *Fundamentos de Espectroscopia no Infravermelho Próximo e Quimiometria*, trata da teoria da técnica analítica secundária utilizada neste trabalho, numa abordagem simples e geral que comenta o princípio do infravermelho próximo

(NIR), as análises multivariadas, a análise de componentes principais (do inglês *Principal Component Analysis*, *PCA*) e a regressão por mínimos quadrados parciais (do inglês *Partial Least Squares*, *PLS*).

O terceiro capítulo: *Desenvolvimento e Validação dos Modelos de Regressão PLS*, reúne não só a parte de materiais e métodos como também as explicações de monitoramento e performance dos modelos utilizando o software Winisi.

O quarto capítulo: *Avaliação dos modelos de regressão PLS para proteína bruta e aminoácidos em farelo de soja*, trata dos resultados e discussão dos modelos obtidos pelo método NIR, suas limitações, vantagens e desvantagens apresentados para cada constituinte calibrado, assim serão 19 regressões analisadas separadamente antes de fundamentar uma avaliação total.

A presente dissertação de mestrado encerra-se com as conclusões gerais evidenciadas neste estudo e propõe algumas perspectivas futuras que podem ser desenvolvidas a partir do raciocínio aqui apresentado. Por fim é apresentada a lista de referências bibliográficas que fundamentaram todo esse trabalho.

**CAPÍTULO 1 -
NUTRIÇÃO ANIMAL, FARELO DE SOJA ,
PROTEÍNAS E AMINOÁCIDOS:
UMA VISÃO GERAL**

Capítulo 1 - Nutrição animal, farelo de soja, proteínas e aminoácidos: uma visão geral

A nutrição animal é a parte aplicada da nutrição que visa estudar os nutrientes, seu metabolismo e suas interações especificamente para as espécies animais. Esse estudo busca conhecer os alimentos utilizados na dieta de animais, considerando suas fases de vida e produção atendendo a interesses do mercado de consumidores.

Existe um grande interesse econômico na alimentação animal que movimenta o mercado e a indústria pecuária. Assim, a alimentação racional dos animais tem por objetivo fornecer a um indivíduo ou um grupo de indivíduos de uma determinada raça ou espécie as melhores condições de nascimento, crescimento, produção e abate .

O conceito de alimento é muito amplo e engloba todas as substâncias que podem ser incluídas nas dietas dos animais por conterem nutrientes. O teor de nutrientes é avaliado separando a água (umidade) da matéria mineral que contém os macros e microelementos (sódio, fósforo, cálcio, cloro, ferro, cobre, manganês, zinco, entre outros) e a matéria orgânica que é composta por glicídios, lipídios, vitaminas e proteínas. O papel da nutrição animal é o estudo detalhado de cada categoria citada que constitui a composição química dos alimentos. ⁽³⁾

Atualmente, o Brasil ocupa, no cenário mundial, um lugar de destaque no desempenho com o agronegócio. Neste contexto, é natural que os esquemas de criação assumam características de verdadeiras empresas e a criação de animais passe a ser considerada como nítida indústria de produção de alimentos.

Do ponto de vista econômico, é justamente sobre o fator alimento que recai a maior parcela dos significativos ônus de produção, fazendo com que os benefícios finais de uma criação sejam sensíveis, conforme variem a eficiência das rações utilizadas e , mais especificamente, os custos de produção destas mesmas rações. Daí a preocupação de melhorar a eficiência das rações, não apenas utilizando matérias primas de melhor qualidade, mas introduzindo paralelamente, tecnologia na análise de nutrientes, com o objetivo de utilizar o máximo de potencialidade destas mesmas matérias-primas.

Dentre os alimentos mais comuns utilizados na nutrição animal, temos a soja e seus subprodutos os concentrados, farinhas e farelos. O elevado teor protéico está

associado a um excelente equilíbrio de aminoácidos, que torna a soja e seus derivados o mais adequado suplemento protéico vegetal disponível para a formulação de rações.

1.1 - A soja e seus subprodutos

A soja, originária da China Continental, é uma planta que pertence à família das leguminosas denominadas cientificamente de ***Glycine max*** (L), compreende mais de 7000 espécies conhecidas a mais de 5000 anos ⁽⁴⁾. Em 1712 foi introduzida na Europa e em 1804 chegou aos Estados Unidos. No Brasil, o primeiro registro de introdução da soja data 1882 na Bahia sem alcançar êxito. Diversos outros registros demonstram que a soja amarela foi inicialmente plantada na Estação agrônômica de Campinas em 1891. Mas só na década de 60 seu desenvolvimento começou a ser significativo no Brasil. ⁽⁵⁾

O grande incremento nacional na produção de soja é devido, especialmente, aos incentivos governamentais, à boa adaptação da cultura às nossas condições ecológicas, às possibilidades de total mecanização no seu manejo e ao fato da colheita coincidir com a entressafra dos Estados Norte-americanos, o que assegura bom mercado de exportação. ⁽⁵⁾

A soja é uma excelente cultura de rotação, uma vez que sendo uma leguminosa, fixa nitrogênio do ar atmosférico ao solo, através da fixação simbiótica dos nódulos radiculares. O valor nutricional da soja prende-se ao seu alto teor de proteínas de fácil digestão, rica em aminoácidos essenciais e fonte de óleo de boa qualidade. ⁽⁶⁾

Aproximadamente 70% das proteínas da soja, em sua maioria globulinas, são proteínas de reserva do grão. As outras são hemaglutininas, inibidores de tripsina, lipoproteínas e enzimas. Os inibidores de tripsina e as hemaglutininas são componentes antinutricionais e devem ser inativados por processos térmicos. ⁽⁷⁾

Segundo dados da ABIOVE (Associação Brasileira das Indústrias de Óleos Vegetais de São Paulo), houve um crescimento significativo na produção brasileira de soja nos últimos anos e a previsão para 2005 é de 61500 toneladas como mostra o gráfico 1. A soja e seus derivados movimentam boa parte da economia brasileira. ⁽⁸⁾

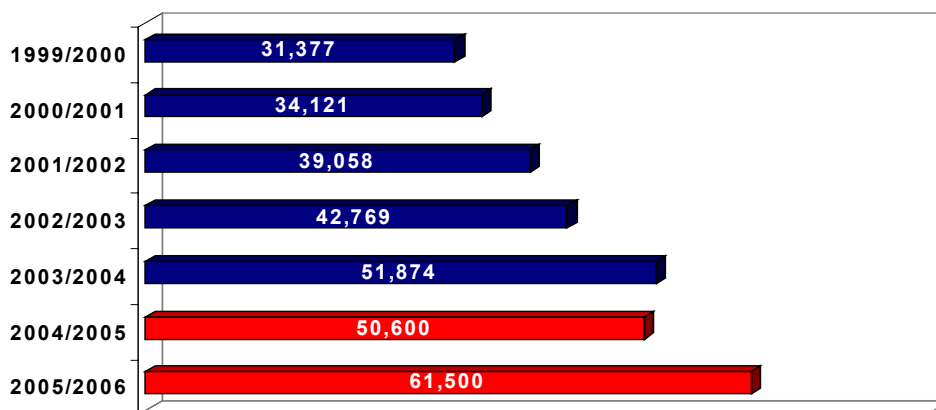


Gráfico1: Produção de soja no Brasil de 2000 à 2003 (azul) e previsão estimada para 2004 e 2005 (vermelho) segundo a ABIOVE.

A soja crua pode ser utilizada apenas para ruminantes desde que não seja em conjunto com uréia, em virtude da urease contida nas sementes desdobrar a uréia em amônia que pode ser tóxica em grandes quantidades. Para aves e suínos o ideal é utilizar a soja processada por aquecimento, que estará isenta de proteínas tóxicas e de ação antinutricional ⁽⁵⁾. O gráfico 2 mostra a composição aproximada dos grãos de soja e seus principais subprodutos:

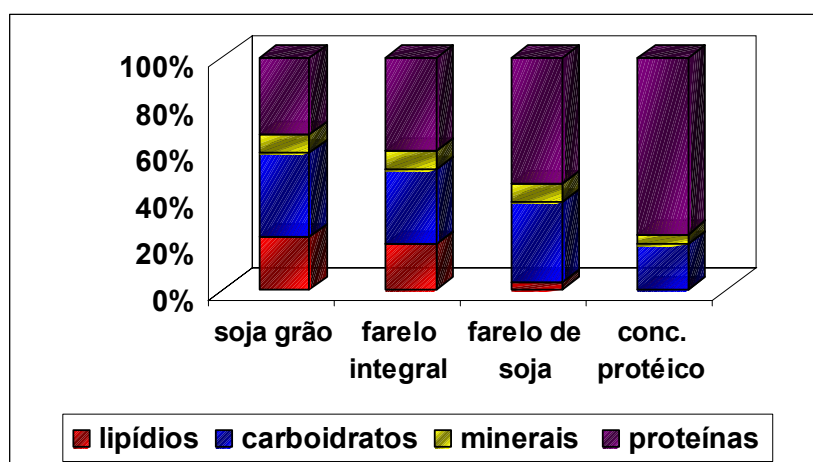


Gráfico 2: Composição média do grão de soja e seus principais subprodutos.

1.2 - Farelo de soja

Dentre os subprodutos de soja o mais utilizados na agroindústria é o farelo de soja, utilizado na nutrição de ruminantes, suínos e aves; sendo um ótimo complemento ao milho para formar a base de uma ração. O processamento do farelo de soja inicia-se com o recebimento dos grãos de soja que passam pelo processo de secagem, limpeza e retirada da casca. Essa medida prévia ao processo de tostagem e remoção do solvente no processamento do farelo assegura, também que os aminoácidos essenciais não sejam inativos pela ligação com compostos fibrosos. A soja sem casca é condicionada e laminada para a obtenção dos flocos com gordura. Algumas esmagadoras de soja usam atualmente expansoras para melhorar a eficiência da extração com solvente. Sua temperatura de 105-120°C melhora a percolação e a drenagem do solvente nos flocos. Extrai-se o óleo com solvente (hexano 65-70°C), para a obtenção do farelo sem gordura, que passam pela dessolventização (33-35°C) de hexano à 105-110°C de 15 à 30 minutos. O material é secado e novamente esfriado à -10°C e moído. ⁽⁵⁾

A produção de farelo de soja cresceu desde o ano 2000 e estima-se a produção de 22.450 toneladas para a safra 2004/2005 e 23.700 toneladas na safra de 2005/2006. O consumo interno manteve-se relativamente estável nos últimos anos mas acredita-se que sofrerá um ligeiro aumento a partir de 2004/2005 ⁽⁸⁾. O aumento também será visível na exportação do produto como mostra na tabela 1.

Tabela 1: Dados sobre farelo de soja (ano comercial –1000 toneladas)

FARELO	2005/06 (P)	2004/05 (P)	2003/04	2002/03	01/02	00/01	99/00	98/99	97/98
Estoque Inicial	790	862	622	358	460	438	417	361	408
Produção	23.700	22.450	21.407	20.040	17.699	16.831	16.868	17.135	14.786
Importação	100	178	288	372	213	119	75	135	308
Consumo Interno	9.100	8.600	7.878	7.569	7.211	7.066	6.945	6.434	5.387
Exportação	14.700	14.100	13.577	12.579	10.803	9.861	9.977	10.780	9.754
Estoque Final	790	790	862	622	358	460	438	417	361

FONTE: SECEX / (P) Previsão Abiove (março 2005) para farelo de soja no Brasil.

Dentre a série de fatores que prejudicam a composição e a qualidade do farelo de soja destaca-se a qualidade do grão de soja utilizado, que é influenciado por condições ambientais, daí a necessidade de se conhecer a composição média dos farelos de diferentes regiões brasileiras ⁽⁹⁾. O controle do tratamento térmico também é de fundamental importância pois pode comprometer a qualidade protéica, verificada pela análise de solubilidade em KOH, diminuindo seu aproveitamento no processo digestivo animal. ⁽¹⁰⁾

1.3 - Proteínas e aminoácidos

As proteínas e os aminoácidos são importantes elementos nutricionais. As proteínas são compostos formados fundamentalmente por carbono (C), hidrogênio (H), nitrogênio (N) e oxigênio (O). Elas têm funções importantes nos organismos tanto animais quanto vegetais. Estruturalmente as moléculas protéicas apresentam-se complexas sendo seu peso molecular bastante elevado. É possível afirmar que essas moléculas são polímeros constituídas de moléculas mais simples: os aminoácidos, ácidos orgânicos que apresentando uma função amina. O termo proteína foi criado por Mulder, em 1840 para designar a substância fundamental da albumina da clara de ovo. A palavra vem do grego *proteus* que significa primeiro. Desde a substância protoplasmática das bactérias até os tecidos mais altamente organizados têm a sua formação, crescimento e manutenção ligados estreitamente as proteínas. Mesmo os vírus que são considerados as formas mais simples de vida são fundamentalmente protéicos. Nos indivíduos mais desenvolvidos as proteínas são a base da estrutura histológica sendo ainda indispensáveis à formação de hormônios e anticorpos. ⁽¹¹⁾

As proteínas se diferem quimicamente dos glicídios e lipídios por conterem, além do carbono hidrogênio e oxigênio, também o nitrogênio. O teor de nitrogênio varia de 15 a 19 % sendo sua média 16 %. ⁽¹¹⁾

Os aminoácidos são os produtos da hidrólise das moléculas protéicas, como o nome sugere são compostos orgânicos contendo grupo ácido (carboxílico) e aminico. O primeiro aminoácido isolado foi a cistina em 1810 por Wallanston, porém só em 1820, Branconnot isolou a glicina por hidrólise ácida e demonstrou que os aminoácidos

são produtos da decomposição primária das proteínas. Em 1846, Liebig isolou a tirosina pela hidrólise alcalina da caseína.⁽¹²⁾

Os aminoácidos e proteínas têm várias funções como: composição do tecido animal, composição dos sistemas enzimáticos, fornecimento de energia de sistemas específicos e armazenamento de nutrientes (lipídios e minerais), efeito tampão e regulação da pressão osmótica, reprodução, imunidade e transporte de oxigênio.⁽¹³⁾

Portanto, os animais devem receber durante toda a vida uma quantidade mínima diária de proteínas para atender suas necessidades que podem ser de crescimento, recuperação dos tecidos e gestação. A ingestão de proteínas é estudada individualmente para cada tipo de animal. Para aves e suínos por exemplo os aminoácidos limitantes (que são aqueles requeridos em quantidade em que os alimentos não conseguem suprir) são a metionina para aves e a lisina para suínos. Geralmente, as proteínas são constituídas por 20 aminoácidos ou mais, que são designados pelas primeiras três letras do seu nome em inglês, com exceção de 4 deles: glutamina (GLN), asparagina (ASN), isoleucina (ILE) e o triptofano (TRP).^(14,15)

O desequilíbrio ou desbalanceamento dos aminoácidos pode causar aumento da síntese de gordura, aumento da sensibilidade aos contaminantes de alimentos (toxinas, patógenos e fatores antinutricionais), crescimento reduzido, redução da eficiência alimentar, reprodução ineficiente e problemas com a aparência de pelos, pele e penas. As dietas precisam ser convenientemente equilibradas para manter o balanço nitrogenado que regula a intensidade da urogênese e a taxa de transformação de aminoácidos em lipídios e glicídios.⁽¹⁶⁾

1.4 - A importância das proteínas e aminoácidos na nutrição animal

As proteínas e aminoácidos são de fundamental importância na nutrição animal, porque estão intimamente relacionadas com os processos vitais das células e, consequentemente de todo organismo.

Como as proteínas são formadas por vários aminoácidos, o organismo animal necessita dos mesmos para sintetizar as suas próprias proteínas. Embora alguns aminoácidos mais simples sejam sintetizados pelo organismo, pelo menos 10 deles não

são sintetizados na velocidade necessária para atender as necessidades orgânicas , sendo portanto necessária sua presença na dieta. Portanto, durante toda vida os animais devem receber quantidades mínimas diárias de proteínas e aminoácidos para atender as suas necessidades de crescimento, recuperação dos tecidos, gestação e reprodução. Para os animais monogástricos tão importante quanto à quantidade é a qualidade da proteína fornecida. ⁽¹⁷⁾

Para nutrição animal, o conceito essencial e não essencial tem sido revisto em função do desenvolvimento de trabalhos de pesquisa. Um aminoácido pode ser considerado essencial ou não dependendo do critério empregado na conceituação tais como crescimento, manutenção do equilíbrio de nitrogênio no adulto, espécie animal, presença de outros fatores como vitaminas, existência de condições fisiológicas alteradas ou patológicas. Em determinadas circunstâncias um aminoácido pode substituir o outro, como o caso da cistina que se torna essencial quando ocorre falta de metionina na dieta, ou a fenilalanina que substitui a tirosina, não ocorrendo o inverso. É comum em nutrição animal procurar definir para cada espécie os aminoácidos limitantes. ⁽¹⁸⁾

A bioquímica dos aminoácidos é complexa tanto na nutrição humana quanto na animal, o interesse aqui contudo é apenas demonstrar que conhecer as variações do conteúdo de aminoácidos e proteínas em um determinado alimento são de extrema importância para se melhorar a performance de uma determinada criação, evitar doenças por falta o excesso dos mesmos e diminuir os custos de produção.

1.5 – As proteínas do farelo de soja: estrutura e composição

O farelo de soja , sob o ponto de vista da nutrição, possui proteínas de alto valor biológico, e neste aspecto, assemelha-se mais a proteína animal do que a qualquer outra vegetal⁽¹⁹⁾. Normalmente o valor da proteína é dado multiplicando-se o valor de nitrogênio por 6,25. Os farelo de soja comuns tem aproximadamente de 43,5 à 48,5% de proteína bruta, dependendo do teor de fibras (5 à 9%) . Existem também os farelos “hipro” que possuem de 49 até 54% de proteína bruta com teor mínimo de fibras (menor que 4%), geralmente destinados à exportação. ⁽²⁰⁾

As proteínas de soja constituem uma mistura de macromoléculas de tamanhos, densidades de carga e estrutura diferentes. O peso molecular das proteínas é uma característica importante para o estudo das suas propriedades físicas, entre elas pode-se destacar o coeficiente de sedimentação dado pela constante S (velocidade de sedimentação em um campo gravitacional unitário), essa medida surgiu graças ao desenvolvimento das técnicas de centrifugação sugeridas em grande parte por Svedberg, esse parâmetro valor S é expresso em unidades Svedberg ($1S = 10^{-13}$ segundos). A ultracentrifugação separa as proteínas da soja em 4 frações de sedimentação equivalentes a 2,7,11 e 15S ⁽¹¹⁾. A quantidade relativa e a faixa de pesos moleculares são descritas na tabela 2 :

Tabela 2 : Classificação das frações protéicas da soja.

fração	2S	7S					11S	15S
Proporção (%)	15	35					40	10
Proteína	Inibidores de tripsina	Citocromo C	β -amilase	lipoxigenase	Hemato-glutininas	Globulina 7S β -conglucina	Globulina 11S glicina	Não classificada
Peso molecular	7860-21500	12000	62000	1020000	110000	140000-175000	320000-350000	600000

A fração 2S constituída por inibidores de tripsina é praticamente destruída na formação do farelo, assim como as hemaglutininas, que também são inativadas pelo tratamento térmico. Assim 90% das proteínas do farelo são globulinas 7S e 11S. O restante, são proteínas da fração 15S que não foram ainda classificadas ⁽¹¹⁾.

1.6 - Os aminoácidos do farelo de soja

As proteínas de farelo de soja são formadas por 18 aminoácidos. O farelo possui ainda uma pequena quantidade de nitrogênio não protéico (NH_3), que impede que a somatória de todos os aminoácidos seja igual ao teor total de proteína bruta. O único problema dos aminoácidos do farelo de soja é sua baixa concentração de metionina e cistina (aminoácidos sulfurados). Outra característica marcante do farelo de soja são os

altos níveis de lisina, que o qualifica ainda mais como excelente fonte de proteína vegetal. ^(21,22)

Os aminoácidos presente no farelo de soja são ^(23,24,25) :

- O ácido glutâmico (GLU), nome químico: ácido α - aminoglutâmico, tem peso molecular 147,10 é um aminoácido de cadeia lateral carboxilada, que constitui de 16 à 20 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_5H_9NO_4$.
- O ácido aspártico (ASP), nome químico: ácido α - amino-succínico, tem peso molecular 133,10 é um aminoácido de cadeia lateral carboxilada, que constitui de 8,0 à 9,5 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_4H_7NO_4$.
- A leucina (LEU), nome químico: ácido α - amino-isocapróico, tem peso molecular 131,20 é um aminoácido de cadeia lateral aberta, que constitui de 6,3 à 8,0 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_6H_{13}NO_2$.
- A arginina (ARG), nome químico: ácido α - amino-ureinovaleriânico, tem peso molecular 174,20 é um aminoácido de cadeia lateral básica, que constitui de 6,5 à 7,2 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_6H_{14}N_4O_2$.
- A lisina (LYS), nome químico: ácido α - ϵ diamino-capróico, tem peso molecular 146,20 é um aminoácido de cadeia lateral básica, que constitui de 5,5 à 6,0 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_6H_{14}N_2O_2$.
- A serina (SER), nome químico: ácido α - amino- β - hidróxi-propiónico, tem peso molecular 133,10 é um aminoácido de cadeia lateral hidroxilada, que constitui de 5,0 à 5,8 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_3H_7NO_3$.
- A fenilalanina (PHE), nome químico ácido α - amino- β - fenil- propiónico, tem peso molecular 165,20 é um aminoácido de cadeia lateral aromática, que constitui de 4,8 à 5,0 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_9H_{11}NO_2$.
- A valina (VAL), nome químico: ácido α - amino-isovaleriânico, tem peso molecular 117,10 é um aminoácido de cadeia lateral aberta, que constitui de 4,7 à 5,0 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_5H_{11}NO_2$.

- A prolina (PRO), nome químico: ácido pirrolidin-2-carboxílico, tem peso molecular 115,10 é um iminoácido, que constitui de 4,5 à 5,0 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_5H_9NO_2$.
- A isoleucina (ILE), nome químico: ácido α -amino- β -metil-valeriânico, tem peso molecular 131,20 é um aminoácido de cadeia lateral aberta, que constitui de 4,0 à 4,6 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_6H_{13}NO_2$.
- A alanina (ALA), nome químico: ácido α -aminopropiônico, tem peso molecular 89,10 é um aminoácido de cadeia lateral aberta, que constitui de 4,0 à 4,5 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_3H_7NO_2$.
- A glicina (GLY), nome químico: ácido α -aminoacético, tem peso molecular 75,10 é um aminoácido de cadeia lateral aberta, que constitui de 3,5 à 4,0 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_2H_5NO_2$.
- A treonina (THR), nome químico: ácido α -amino- β -hidróxi-n-butírico, tem peso molecular 119,10 é um aminoácido de cadeia lateral hidroxilada, que constitui de 3,2 à 4,0 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_4H_9NO_3$.
- A tirosina (TYR), nome químico: ácido α -amino- β -p-hidróxi-fenil-propiônico, tem peso molecular 181,20 é um aminoácido de cadeia lateral aromática, que constitui de 3,0 à 3,9 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_9H_{11}NO_3$.
- A histidina (HIS), nome químico: ácido α -amino- β -imidazol-propiônico, tem peso molecular 155,20 é um aminoácido de cadeia lateral básica, que constitui de 2,0 à 2,5 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_6H_9N_3O_2$.
- O triptofano (TRP), nome químico: ácido α -amino- β -3-indol-propiônico, tem peso molecular 204,20 é um aminoácido de cadeia lateral aromática, que constitui de 1,0 à 1,8 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_{11}H_{12}N_2O_2$.
- A metionina (MET), nome químico: ácido α -amino- γ -metil-n-butírico, tem peso molecular 149,20 é um aminoácido de cadeia lateral sulfurada, que constitui de 1,0 à 1,7 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_5H_{11}NO_2S$.

- A cistina (CYS), nome químico: ácido α - amino- β - carboxi-di-tio-propiónico, tem peso molecular 240,30 é um aminoácido de cadeia lateral sulfurada, que constitui de 1,0 à 1,5 % do valor das proteínas do farelo de soja, cuja fórmula molecular é $C_6H_{12}N_2O_4S_2$. Esse aminoácido é o produto da oxidação da cisteína para formar uma ponte dissulfurada S-S e apresenta-se de forma abundante em proteínas estruturais.

***CAPÍTULO 2 -
FUNDAMENTOS DE ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO
E QUIMIOMETRIA***

Capítulo 2 : Fundamentos de Espectroscopia no Infravermelho Próximo e Quimiometria

A espectroscopia no infravermelho próximo já pode ser considerada uma técnica consagrada para detecção de substâncias orgânicas. A região do NIR compreende a faixa de 780 à 2500 nm. As absorções moleculares nesta região são causadas à priori por ligações (X-H), onde X pode ser C, N ou O. ^(26,27)

As determinações analíticas no infravermelho próximo são baseadas no espectro inteiro do material analisado que necessita do estabelecimento de uma relação matemática entre o espectro e valores de concentrações medidos por um método de referência. Essa relação quando bem estabelecida, fornece à baixo custo, com alta precisão e repetibilidade, resultados de uma dada concentração através de um modelo de calibração. O raciocínio matemático, estatístico e químico que fundamenta estas conclusões deram origem à Quimiometria. ^(28,29)

No final dos anos 60 surgiu dentro da química analítica esta nova área de pesquisa que busca extrair de uma grande quantidade de dados químicos complexos, resultados analíticos confiáveis. O termo Quimiometria é derivado do inglês “*chemometrics*” sendo hoje utilizada para a análise de dados, com finalidade específica dentro de um estudo químico como a otimização de um processo, classificação de dados, modelagem e monitoramento de processos multivariados, construção de modelos de regressão e desenvolvimento de métodos de inteligência artificial. ⁽³⁰⁾

2.1 – Princípios da Espectroscopia no Infravermelho Próximo (NIR)

A natureza composta da luz branca foi demonstrada pela primeira vez por Newton, em 1664, quando decompôs a luz solar por meio de um prisma, projetando-a numa tela. A imagem alongada e colorida do sol foi chamada por ele de *espectro*. ⁽³¹⁾

Em 1800 o astrônomo inglês Willian Herschel repetiu a experiência de Newton, com a finalidade de descobrir qual das cores do arco-íris daria mais resultado no aquecimento do bulbo de um termômetro. Assim foi descoberta a radiação infravermelha, pois foi observado que a temperatura aumentava após o vermelho. ⁽³¹⁾

Na faixa de radiações do infravermelho distinguem-se três regiões: Infravermelho próximo, médio e distante. O infravermelho próximo possui algumas propriedades em comum com a luz visível, com a diferença de que não é percebido pelo olho humano. As absorções moleculares ou sobretons (do inglês *overtones*) ocorrem com frequência de 780 à 1100 nm e as bandas de combinação ocorrem de 1800 à 2500 nm. ^(32,33)

O espectro é composto de picos de absorções individuais para cada substância e pode ser calculado devido ao conceito de que as ligações entre átomos vibram com frequências que podem ser descritas por leis físicas, sejam moléculas diatômicas simples ou moléculas poliatômicas. Essa frequência é dada pelo estiramento, torção ou deformação das ligações NH, CH, OH ou aromáticos. Outros grupos funcionais importantes incluem estiramentos carbono-oxigênio, grupos carbonila, carbono-nitrogênio ,entre outros. Algumas vezes é possível identificar por essa técnica aminas primárias secundárias e terciárias. Entretanto, uma atribuição precisa de bandas na região NIR é difícil devido ao fato de uma simples banda de combinação poder ser atribuída a vários sobretons que são severamente sobrepostos. Isso faz do MID (Infravermelho Médio), a melhor região do infravermelho para se trabalhar com identificação de compostos. ^(33,34)

Os equipamentos mais utilizados atualmente para obtenção de medidas no NIR são equipados com interferômetros, que através da operação matemática com transformada de Fourier origina o espectro de absorção, o que gera uma grande eficiência no transporte de radiação até o detector, aumenta a resolução e a reprodutibilidade , além de realizar medidas simultaneamente. Porém ainda utiliza-se com grande eficácia equipamentos baseados em filtros ou monocromadores. ⁽³¹⁾

As medidas feitas no NIR em sólidos podem ser por reflectância especular, reflectância total atenuada ou reflectância difusa e possuem a vantagem de não serem destrutivas para amostra. O espectro de reflectância não é idêntico ao de absorção, mas apresenta uma certa semelhança e carrega as mesmas informações. Das reflectâncias citadas a difusa, conhecida como DRIFTS (do inglês, *Diffuse Reflectance Infrared Fourier Transform Spectrometry*) não requer grande preparo da amostra , apenas moagem para transforma-la em pó. Deve ter tamanho de partícula uniforme e a moagem deve ser refrigerada para não alterar as características químicas da amostra.

O grau de compactação não deve sofrer variações significativas e a superfície da amostra deve ser plana. ⁽³³⁾

A reflectância difusa é obtida pela penetração do feixe de radiação infravermelha na superfície da amostra, ocorrendo então a interação com as partículas sólidas que permite transferir as informações sobre a amostra para a radiação que será refletida de forma difusa. ⁽³¹⁾

2.2 - Análise de Componentes Principais (PCA)

Análise de componentes principais é um tipo de análise exploratória de dados que visa extrair o máximo de informações de uma tabela de dados convertendo-a em gráficos informatizados que mostram a relação entre amostras (linhas de uma matriz) e as variáveis (colunas de uma matriz) visando transformar dados complexos para que as informações, mais importantes e relevantes sejam realçadas. O novo conjunto de variáveis (fatores, componentes principais ou variáveis latentes) é a combinação linear das originais. Os novos eixos são ortogonais entre si e são constituídos em ordem decrescente da quantidade de variância que as descrevem. Assim, podemos dizer que o PCA tenta agrupar aquelas variáveis que estão altamente correlacionados numa nova variável chamada componente principal. Como mostra a figura 1 :

$$\begin{array}{ccccccc}
 & m & & 1 & & m & & 1 & & m & & 1 & & m \\
 & \boxed{\mathbf{X}} & = & \boxed{\mathbf{T}_1} & \boxed{\mathbf{P}_1^T} & + & \boxed{\mathbf{T}_2} & \boxed{\mathbf{P}_2^T} & + \dots + & \boxed{\mathbf{T}_A} & \boxed{\mathbf{P}_A^T} \\
 n & & & n & 1 & & n & 1 & & n & 1
 \end{array}$$

Figura 1: Representação esquemática da decomposição por PCA.

O termo PCA vem do inglês (*Principal Componentes Analysis*) sua função é decompor uma matriz $\mathbf{X}$ em duas matrizes de escores ($\mathbf{T}_{n \times A}$) e pesos ($\mathbf{P}_{m \times A}$), onde A é o número de componentes principais (CP).

A análise de componentes principais transforma a matriz X de maneira que:

$$X = TP^t \quad [1]$$

onde:

T= Escores que são as coordenadas das amostras em um novo sistema de eixos

P= Pesos que são as colunas que contém informação do peso de cada variável original na formação dos novos eixos.

Como usualmente a grande fração da variância é descrita nos primeiros componentes principais, é possível visualizar os dados pelo gráfico dos escores de um componente contra o outro. Os algoritmos usados para os cálculos com o PCA em química analítica são o NIPALS (do inglês, *Nonlinear Iterative Partial Least Squares*) mínimos quadrados parciais não linear interativo e o SVD (do inglês, *Singular Value Decomposition*) decomposição em valores singulares. ^(34,35)

2.3 - Calibração Multivariada

Os métodos de calibração usados em química analítica são classificados de acordo com a dimensão dos dados analisados, podendo ser de ordem zero (univariados) de primeira ordem (multivariados) e de segunda ordem (multidimensionais). A calibração, em geral, é uma operação que relaciona uma grandeza de saída com uma grandeza de entrada para um sistema de medida sob determinadas condições. ⁽³⁶⁾

O objetivo da maioria das análises multivariadas é desenvolver modelos para prever uma propriedade de interesse. A maior parte dos instrumentos fornecem um sinal como resposta, como por exemplo a quantidade de luz que chega a um detector. A relação da intensidade de luz e a concentração do analítico de interesse deve ser determinada por um modelo de calibração que correlaciona os dados medidos ou

calculados independentemente (x) e alguma propriedade da amostra (y). O processo consiste em duas etapas conhecidas como modelagem e validação que possibilita posteriormente a previsão de valores. Os métodos de calibração multivariada são particularmente úteis devido ao acréscimo de precisão decorrente do uso de vários canais de dados. Assim é possível analisar espécies mesmo na presença de interferentes, devido ao uso de todos os canais simultaneamente. ⁽³⁷⁾

Como exemplo de métodos de calibração temos o CLS (Quadrados Mínimos Clássicos), o MLR (Regressão Linear Múltipla), o PCR (Regressão por Componentes Principais) e o PLS (Regressão por Mínimos Quadrados Parciais). Cada um apresenta vantagens e desvantagens que devem servir de critério de escolha para o desenvolvimento de um modelo de calibração. De forma geral os mais usados são PCR e PLS. Porém, existem observações na literatura que afirmam que o método PLS é tendencioso, mas analisando a fundo a questão é possível notar que desde que os valores de laboratório sejam robustos, não há motivos para preocupação, além do mais qualquer método secundário pode ter desvantagens, por isso existe a necessidade de conhecer bem as incertezas das medidas com as quais se pretende trabalhar e eleger o melhor método matemático para reproduzir essas medidas. ^(36,37)

2.4 - Regressão por Mínimos Quadrados Parciais (PLS)

Este método de calibração foi desenvolvido por Herman Wold na década de 70. As informações espectrais e as informações das concentrações são usadas ao mesmo tempo na fase de calibração. A matriz de espectros é decomposta em matrizes de variações dos espectros (*loadings* ou pesos) e a posição das amostras (*scores*). Os espectros originais podem ser considerados como combinações lineares das variações dos espectros (*pesos*) onde os *scores* representam suas contribuições⁽³⁸⁾. No PLS tanto a matriz das variáveis independentes X, como das variáveis dependentes Y são representadas pela análise de componentes principais:

$$\mathbf{X} = \mathbf{TP} + \mathbf{E} \quad [2]$$

$$\mathbf{Y} = \mathbf{UQ} + \mathbf{F} \quad [3]$$

As informações são incorporadas de modo que cada componente principal (CP) do modelo seja modificada para que covariância de $\mathbf{T}$ e $\mathbf{U}$ seja maximizada. Nesta etapa, a CP recebe o nome de variável latente (VL). Quando as matrizes $\mathbf{X}$ e $\mathbf{Y}$ são decompostas, $\mathbf{T}$ são os escores. $\mathbf{P}$ e $\mathbf{Q}$ são os pesos. $\mathbf{E}$ e $\mathbf{F}$ são as matrizes de erro das decomposições de $\mathbf{X}$ e $\mathbf{Y}$ respectivamente. ⁽³⁸⁾

A decomposição pode ser feita utilizando NIPALS ou SIMPLS, (do inglês, *Straight Forward Implementation of statistically Inspired modification of PLS*) ⁽³⁹⁾. Os valores de scores são relacionados com os loadings para cada componente de cada vez pela equação:

$$\mathbf{q} = \mathbf{y} \cdot (\mathbf{t}^T \cdot \mathbf{t}) \cdot \mathbf{y}^T \quad [4]$$

O vetor de regressão é calculado pela equação :

$$\hat{\mathbf{b}} = \mathbf{W}(\mathbf{PW}^T)^{-1} \cdot \mathbf{q}^T \quad [5]$$

Em que $\mathbf{W}$ é a matriz de pesos do PLS.

Em seguida é feita a validação do modelo com as novas amostras. A validação cruzada é baseada na avaliação da grandeza dos erros de previsão comparando os valores das variáveis dependentes das amostras do conjunto de calibração com as

respectivas previsões, quando as mesmas não participam na construção do modelo de regressão. Em paralelo é realizado a validação com um conjunto externo que deve apresentar amostras com valores que compreendam o intervalo de dados do conjunto de calibração e que devem apresentar performance muito semelhante em todos os parâmetros de avaliação dos modelos de calibração. ⁽⁴⁰⁾

***CAPÍTULO 3 -
DESENVOLVIMENTO E AVALIAÇÃO
DOS MODELOS DE REGRESSÃO PLS***

Capítulo 3 : Desenvolvimento e avaliação dos modelos de regressão PLS

O desenvolvimento de uma boa calibração leva em consideração parâmetros de medidas de tendência central, dispersão e qualidade dos modelos que serão descritos para que se possa avaliar os modelos desenvolvidos. É necessário também compreender a identificação de amostras anômalas (do inglês, *outliers*), através das medidas de H que descrevem quanto a amostra está distante do centro do modelo (H global) ou dos seus vizinhos mais próximos (H vizinho ou neighborhood) e T que fornece o grau de incerteza da medida em relação ao valor numérico original. Esses valores são utilizados para entender as seleções de amostras utilizadas nos modelos de calibração, bem como avaliar posteriormente a rotina operacional do equipamento. Por isso, serão apresentados os recursos utilizados pelo software Winisi para tornar melhor a compreensão dos dados apresentados neste trabalho.

3.1 - Monitoramento e performance dos modelos de calibração multivariada

3.1.1 - Medidas de Tendência Central

A média representa o somatório de todos os valores do conjunto de dados dividido pelo número dos dados informados. ^(41,42)

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad [6]$$

Em que n é o número de medidas.

Bias é um caso especial da média, conhecido como a diferença entre duas médias, ou seja, um tipo de deslocamento da média:

$$\text{Bias} = \text{média das diferenças} \quad [7]$$

3.1.2 - Medidas de dispersão

O desvio padrão SD (do inglês, *Standard Deviation*) é uma medida de estatística primária usada para medir dispersão.

$$SD = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} \quad [8]$$

3.1.3 - Medidas de qualidade do modelo

A fórmula geral usada é o erro padrão da diferença SED (do inglês, *standard error of difference*). Ele pode ser aplicado entre valores de laboratório versus laboratório, valores de previsão versus previsão, valores de laboratório versus valores previsão, valores de espectro versus valores de espectro. ^(41,42)

$$SED = \frac{[\sum (D)^2]}{n} \quad [9]$$

O erro padrão da calibração SEC (do inglês, *standard error of calibration*) é a raiz quadrada da soma dos quadrados de erros entre os valores de referência e os valores previstos, divididos pelo número de graus de liberdade. ^(41,42)

$$SEC = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - p - 1}} \quad [10]$$

y_i é o valor de referência

$\hat{y}_i$ é o valor dado pelo modelo

n é o número de amostras na equação

p é o número de VLs do PLS

O coeficiente de determinação (R^2) ou coeficiente de correlação chamado RSQ, é o parâmetro que informa a correlação entre os valores previstos e os valores de referência. O valor de RSQ desejável é o mais próximo de 1. Esses dados são calculados para o conjunto de calibração. ^(41,42)

$$RSQ = \frac{SDy^2 - SEC^2}{SDy^2} \quad [11]$$

Onde SDy^2 é o quadrado dos desvios padrões dos valores previstos.

O erro padrão da validação cruzada SECV (do inglês, *standard error of cross validation*) é calculado segundo a fórmula:

$$SECV = \sqrt{\frac{SEP_1^2 + SEP_2^2 + SEP_3^2 + \dots + SEP_K^2}{K}} \quad [12]$$

Onde K é o número de segmentos ou população de amostras, utilizado na validação cruzada.

Usualmente o SECV informa a qualidade da previsão. Pode ser considerado um valor muito mais importante e relevante que o SEC. A população de amostras que pode ser usada para construir a calibração é geralmente dividida em 4 ou 5 grupos de amostras. Cada grupo escolhido é previsto dentro da equação utilizando outro grupo de amostras. O SECV é a média quadrática dos erros dos grupos. O valor de SECV é sempre mais alto que o SEC porque as amostras que são previstas não fazem parte do conjunto de calibração. ^(41,42)

O termo 1-VR é o coeficiente de determinação da validação cruzada. É um parâmetro que informa a qualidade dos valores previstos. Deve ser o mais próximo de 1. Um bom modelo de validação é aquele que possui o 1-VR próximo ao RSQ ou R^2 ($\pm 10\%$ de diferença, com tolerância máxima de 12% em casos onde a medida de valor analisado é menor que 1%). ^(41,42,43)

$$1 - VR = \frac{SDy^2 - SECV^2}{SDy^2} \quad [13]$$

O erro padrão de previsão SEP (do inglês, *standard error of prediction*) é dado pela equação:

$$SEP = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} \quad [14]$$

Onde n é o número de amostras do conjunto de validação. Os valores de referência são comparados aos valores de previsão. Assim, se o SEP for menor que o SECV os valores de previsão estão bons⁽⁴²⁾ Se o SEP for maior ao SECV existe algum problema que pode ser :

- a) As amostras não são representadas na calibração.
- b) Pode haver um erro sistemático no método de referência.
- c) Pode haver erro no tipo de dado inserido no sistema.

O SEP(C) é um erro padrão de performance que deve ser menor ou igual a 1,3 vezes o valor de SEC .

$$SEP(C) = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2 \cdot (\sum y_i - \hat{y})^2 / n}{n - 1}} \quad [15]$$

3.1.4 - Identificação de anomalias (medidas de H e T)

A detecção de amostras anômalas é muito importante nas análises multivariadas. Essa detecção é utilizada para excluir dados de referência ou espectros de amostras que não seguem valores padrões ideais para o desenvolvimento de um bom modelo de calibração e a posterior avaliação de resultados previstos na rotina de análise. Esses valores são as estatísticas de H (global e vizinho mais próximo) e T.⁽⁴³⁾

O teste para identificar amostras anômalas H pode ser o NH (*neighborhood*, ou vizinho mais próximo) e o global H. O NH é calculado tomando uma amostra e medindo sua distância entre os vizinhos mais próximos. Esta distância deve ser menor que 1,0.

A distância de uma amostra do ponto central dos eixos é chamada de global “H”. Os anéis na figura 2 representam as distâncias 1,00, 2,00 e 3,00 da média. ⁽⁴⁴⁾

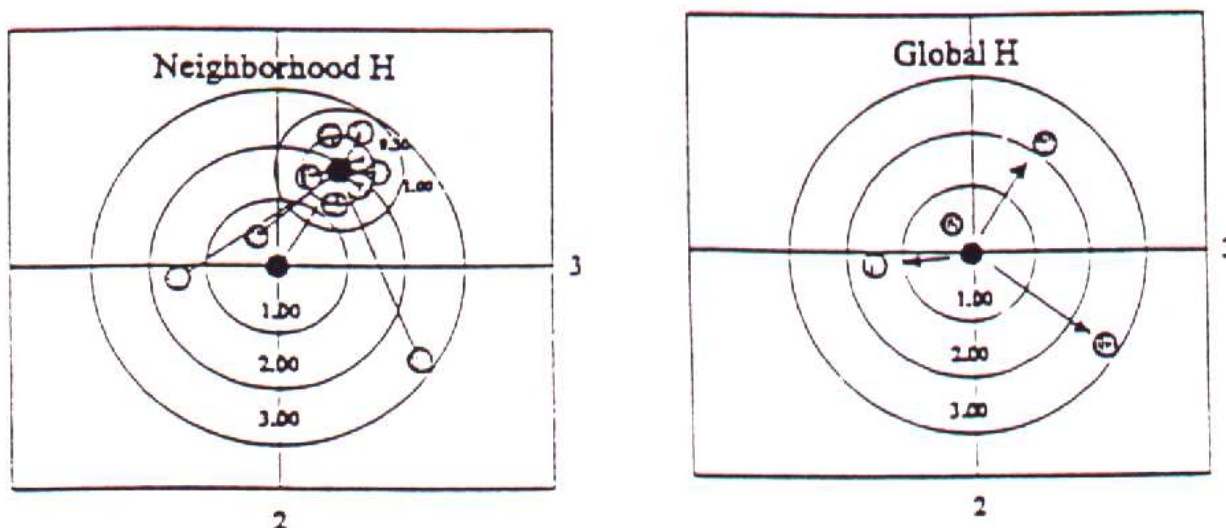


Figura 2 : Vizinho mais próximo NH e H global.

Normalmente os valores de H global devem estar abaixo de 4. A recomendação é que a faixa limite é de 3 a 4, dependendo das características dos conjuntos de amostras que se deseja calibrar. Os valores de NH são calculados pela distância de Mahalanobis ⁽⁴²⁾. A fórmula geral para calcular distância H é:

$$H = x_i(X^tX)^{-1}x_i^t \quad [16]$$

O valor de T define as diferenças (**D**) no conjunto das variáveis dependentes (concentrações), ou seja, entre os valores de previsão e os valores de referência. O valor máximo de T é 2,5 em alguns casos, onde sabe-se que a análise apresenta um erro analítico alto pode ser aceito até 3,0. ^(42,44)

$$T = \frac{D}{SED} \quad [17]$$

3.2 - Desenvolvimento dos modelos de calibração PLS

No emprego da técnica NIR, o algoritmo para o desenvolvimento (obtenção) dos modelos de calibração, usará um banco de dados gerados pelas análises químicas com amostras de farelo de soja. Os espectros das amostras foram coletados num equipamento NIRS (sistema de infravermelho próximo), foram feitas a validação cruzada com o próprio arquivo de calibração e uma validação externa com um arquivo de amostras adicionais que não participaram da construção dos modelos. A seleção do modelo matemático PLS final para representar a melhor equação para cada constituinte estudado será baseado nos parâmetros SEC, RSQC, 1-VR, SEP, SEPC e RSQV que foram detalhados no item 3.1 desse capítulo e serão avaliados tanto na calibração quanto na validação externa. Posteriormente, para o dia-a-dia são utilizadas as medidas de distância e seleção de valores (NH, GH e T) para validar os resultados obtidos e estipular o grau de incerteza analítica da medida. ^(44,45)

3.2.1 – Amostragem

As amostras de farelo de soja são provenientes de várias regiões brasileiras e foram cedidas pelo LABTEC (Campinas, SP) . Este laboratório terceiriza serviços de análises instrumentais e bromatológicas, sendo responsável também pelas análises da Mogiana Alimentos S.A., uma empresa do segmento de nutrição animal conhecida como Rações Guabi. Um total de 150 amostras foram relacionados de 15 fornecedores diferentes que atuam no Brasil fornecendo soja e seus subprodutos para abastecer pequenos médios e grandes produtores do agronegócio. Entre eles podemos citar Bunge, Olvego, Brejeiro, Caramuru, Cocamar, ABC Inco, Carol, Cargill, ADM e Ceval que garantem a boa representatividade das amostras utilizadas neste estudo.

3.2.2 - Aquisição dos espectros e análises de laboratório

Os espectros no infravermelho próximo foram coletados no equipamento da Foss NIRSystem (NIRS) , modelo 5000 , módulo spinning ⁽⁴²⁾. As seguintes condições de medidas foram padronizadas:

- 16 varreduras da cerâmica de referência antes da leitura da amostra;

- 32 varreduras da amostra para calcular a média;
- 16 varreduras da cerâmica de referência depois da leitura da amostra;
- opção de filtro e suavizamento com 4 pontos .

A região medida foi de 1300 a 2400 nm com resolução de 2 nm, obtendo-se 551 variáveis por espectro. Os espectros foram coletados das amostras “como recebidas” (desconsiderando portanto a possibilidade de corrigir os espectros pela matéria seca), uma vez que este é o padrão para esse tipo de amostra utilizado pelos zootecnistas na formulação de rações. As médias do espectro da amostra com os da cerâmica são utilizados para fornecer o espectro de refletância R que é transformado, para garantir a linearidade, pela relação $\text{Log}1/R$.

Foram realizadas as análises de laboratório para medir a porcentagem de proteína bruta (PB) das amostras ⁽²⁾. A faixa escolhida foi de 44,87 à 50,95% sendo que a maior representatividade está na faixa de 46,50 à 48% em 116 amostras do conjunto de calibração como mostra o gráfico 3. O conjunto de validação foi composto por 34 amostras de 45,90 a 50,53 % de proteína bruta (PB). É importante lembrar que esta faixa está inclusa no modelo de calibração visto que é necessário utilizar amostras diferentes mas que estejam dentro dos limites pré estabelecidos pelo modelo. Como mostra o gráfico 4, as amostras estão com distribuição bastante próxima à normal.

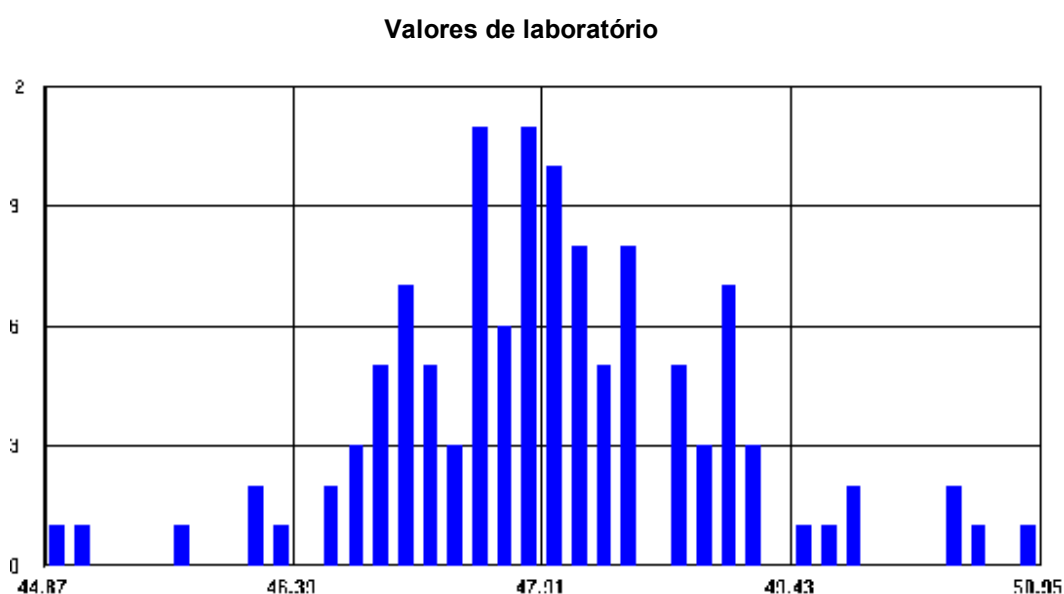


Gráfico 3: Histograma das % de PB do modelo de calibração .

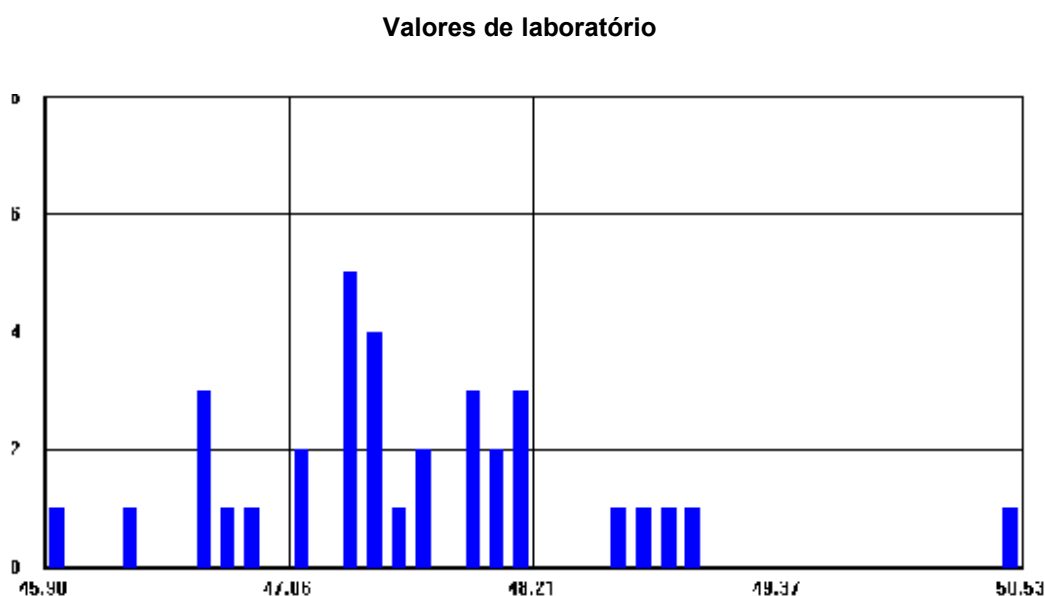


Gráfico 4: Histograma das % de PB do modelo de validação .

As análises de aminoácidos foram realizadas via analisador de aminoácidos Hitachi segundo metodologia específica para triptofano (TRP), cistina (CYS), metionina (MET). Todos os demais foram analisados pela mesma técnica de hidrólise com HCl 6 mol/L que são: glicina (GLY), isoleucina (ILE), ácido aspártico (ASP), leucina (LEU), arginina (ARG), treonina (THR), tirosina (TYR), prolina (PRO), serina (SER), valina (VAL), fenilalanina (PHE) e ácido glutâmico (GLU), histidina (HIS), alanina (ALA) e lisina (LYS)⁽⁴⁶⁾. Como estes valores possuem uma dada correlação com o valor de proteína bruta é possível afirmar que seguem a mesma distribuição normal, sendo que o aminoácido que apresenta maior valor (10%) é dado pelo ácido glutâmico que seria aproximadamente 20% do valor da proteína bruta e o menor valor (0,50%) é dado pelo aminoácido cistina que seria aproximadamente 1% do valor da proteína bruta total.

Existe uma relação entre o valor de proteína bruta e os valores dos aminoácidos que pode ser usada para estimar o valor aproximado de aminoácido numa amostra

conhecida ⁽⁴⁷⁾. Contudo, por não ser um valor fixo, pois depende do material pode trazer erros significativos, não sendo a técnica mais aconselhada para determinar o valor de aminoácidos, embora alguns nutricionistas acabem usando esse raciocínio para definir tabelas usadas em formulações. A tabela 3, mostra a faixa aproximada dos valores de aminoácidos no farelo de soja:

Tabela 3: Faixa de concentração dos valores de aminoácidos em farelo de soja (43 à 54% de PB).

ORDEM DE CONC.	AA	FAIXA		VARIAÇÃO	
		VALOR REAL		% VALOR PB	
		MIN% AA	MAX % AA	MIN	MAX
1	CYS	0,430	0,810	1,0	1,5
2	MET	0,430	0,918	1,0	1,7
3	TRP	0,430	0,972	1,0	1,8
4	HIS	0,860	1,350	2,0	2,5
5	TYR	1,290	2,106	3,0	3,9
6	THR	1,376	2,160	3,2	4,0
7	GLY	1,505	2,160	3,5	4,0
8	ALA	1,720	2,430	4,0	4,5
9	ILE	1,720	2,484	4,0	4,6
10	PRO	1,935	2,700	4,5	5,0
11	VAL	2,021	2,700	4,7	5,0
12	PHE	2,064	2,700	4,8	5,0
13	SER	2,150	3,132	5,0	5,8
14	LYS	2,365	3,240	5,5	6,0
15	ARG	2,795	3,888	6,5	7,2
16	LEU	2,709	4,320	6,3	8,0
17	ASP	3,440	5,130	8,0	9,5
18	GLU	6,880	10,800	16,0	20,0

3.2.3 - Métodos de referência

A quantificação dos níveis de proteína bruta no farelo de soja foi feita a partir da determinação do conteúdo de nitrogênio total na amostra. A técnica de determinação é conhecida como dosagem de proteína através do analisador de nitrogênio LECO FP-428 , segundo o método de referência AACC 946-30 (*American Association of Cereal Chemists, 1995*), que utiliza um detector de condutividade térmica para dosar o nitrogênio produzido por pirólise à 900°C, utilizando hélio como gás inerte de arraste. Este método é atualmente usado pelo LABTEC sendo aplicável à matérias primas de origem animal e vegetal, rações, leite, enlatados e fezes. Essa quantificação de nitrogênio é convertida em porcentagem de proteínas bruta através de um fator. Devido as cadeias de proteínas serem diferentes para cada estrutura temos alguns fatores de conversão sendo a média adotada 6,25, que é utilizada para calcular todos os valores

de proteína bruta neste trabalho. Esse método é extremamente sensível e por ser inteiramente automatizado produz uma variação analítica menor que 1,2% ⁽²⁾. Contudo, a variação analítica máxima aceita pelos padrões nacionais e internacionais de nutrição animal é 2%. ⁽¹⁾

As determinações dos aminoácidos foram realizados também pelo LABTEC, utilizando o analisador de aminoácidos modelo L-8500 e marca Hitachi com derivatização pós-coluna, segundo método proposto para um estudo colaborativo do jornal da AOAC (*Association of Official Analytical Chemists*) ⁽⁴⁶⁾. O princípio operacional consiste na separação dos aminoácidos livres em coluna de troca iônica. Os aminoácidos após separação na coluna, são levados ao reator em aquecimento, onde entrarão em contato com uma solução de ninidrina e desenvolverão uma reação colorimétrica que poderá ser detectada e registrada em detector específico. São utilizados 2 comprimentos de onda (440nm e 570 nm) para analisar os 18 aminoácidos em questão. Para o desenvolvimento analítico utiliza-se padrões e aminoácidos de fluídos fisiológicos. A hidrólise de proteínas é feita de acordo com o comportamento do aminoácido. Para determinação da maioria dos aminoácidos é feita uma hidrólise ácida com HCl 6 mol/L que não é aplicável ao triptofano, a metionina e a cistina. O triptofano é analisado através de hidrólise básica com hidróxido de lítio 4 mol/L; a cistina e a metionina são analisadas oxidação com ácido perfórmico 88%. A análise de aminoácidos devido principalmente ao tipo de abertura da amostra (processo químico de hidrólise e oxidação) e ao próprio tipo do aminoácido pode apresentar coeficientes de variação relativamente altos ^(46,47). O estudo feito com uma amostra de farelo de soja realizados em 10 dias diferentes mostrou os seguintes dados:

Tabela 4 : Estudo de desvio padrão e coeficiente de variação para aminoácidos em farelo de soja.

AA	média %	SD%	CV%	AA	média %	SD%	CV%	AA	média %	SD%	CV%
PRO	2,13	2,01x10 ⁻¹	9,44	ALA	2,08	1,34x10 ⁻¹	6,46	ILE	2,80	1,20x10 ⁻¹	3,35
HIS	1,78	4,30x10 ⁻¹	8,00	CYS	0,637	2,60x10 ⁻²	4,06	LYS	3,99	1,42x10 ⁻¹	3,49
SER	2,42	2,08x10 ⁻¹	8,52	PHE	2,98	1,28x10 ⁻¹	4,29	ASP	4,36	2,60x10 ⁻²	2,40
GLY	2,08	1,54x10 ⁻¹	7,40	THR	2,70	7,70x10 ⁻²	2,84	ARG	3,87	7,10x10 ⁻²	1,83
MET	0,364	2,40x10 ⁻²	6,54	LEU	3,92	1,42x10 ⁻¹	3,49	VAL	2,78	4,20x10 ⁻²	1,52
TRP	0,623	2,50x10 ⁻²	6,70	TYR	1,80	5,90x10 ⁻²	3,28	GLU	7,13	7,30x10 ⁻²	1,02

3.2.4 - Pré-tratamento dos Espectros

Como pode ser visto no gráfico 5, os dados originais do conjunto de calibração e validação apresentam problemas de espalhamento e variação de linha de base que precisam ser corrigidos. Como os espectros de validação são idênticos e por isso vão receber o mesmo tratamento. A razão dos espectros não apresentarem ruído foi a utilização do alisamento com transformada de Fourier feito na etapa de coleta de espectros.

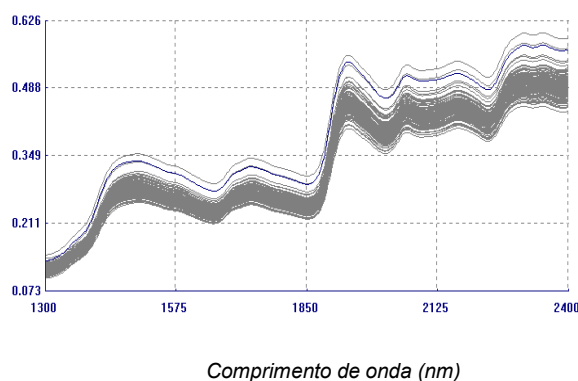


Gráfico 5 - Espectros originais dos conjuntos de calibração para amostras de farelo de soja.

Neste tratamento foi empregado o software Winisi versão 1.5, da ISI (infrasoftware Internacional) desenvolvido para equipamentos NIRSystem e representados no Brasil pela POLIMATE.

A correção de linha de base foi feita neste software com primeira derivada, que é calculada utilizando um intervalo de comprimentos de onda que pode ser de 0 à 99 nm. (chamado de GAP). Neste caso foi utilizado o GAP é igual a 4. O algoritmo utilizado para calcular a derivada foi o Lorentz. ⁽⁴⁷⁾

Foi feita a correção multiplicativa de sinal MSC (do inglês, *multiplicative scatter correction*), para diminuir os efeitos de espalhamento de luz que é comum em espectroscopia por reflectância difusa devido aos diferentes tamanhos de partícula. ⁽⁴¹⁾ Os espectros corrigidos são mostrados no gráfico 6 :

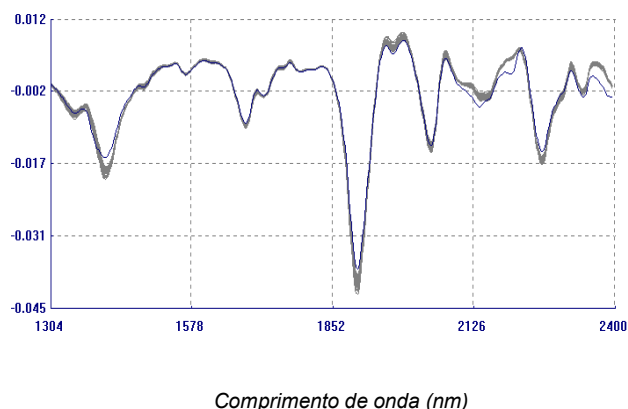


Gráfico 6 : Espectros de calibração após o pré tratamento completo dos dados.

3.2.5 - Análise de PCA dos dados de Calibração

Os dados foram centrados na média antes da execução do o PCA. Os resultados escores dos primeiros 3 componentes principais do conjunto de calibração com 116 amostras são mostrados no gráfico 7. É possível notar que a amostra 81 é anômala, e por isso foi excluída antes do desenvolvimento dos modelos de calibração PLS para a proteína bruta e demais aminoácidos estudados no farelo de soja. Portanto, o conjunto de calibração passou a ter 115 amostras. O conjunto de validação foi composto de 34 amostras. Foi desconsiderada a amostra 81 apenas na etapa de construção dos modelos, sendo que a mesma permaneceu no conjunto original de dados para que não houvesse alteração nos resultados do PCA que avaliou os conjuntos para cada aminoácido.

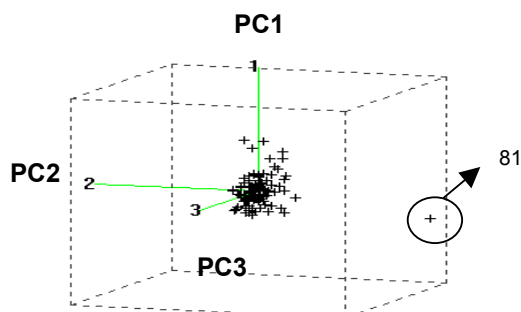


Gráfico 7: Gráfico dos escores do PCA para os primeiros 3 componentes principais.

CAPÍTULO 4 -
AVALIAÇÃO DOS MODELOS DE REGRESSÃO
PLS PARA PROTEÍNA BRUTA E AMINOÁCIDOS
EM FARELO DE SOJA

Capítulo 4: Avaliação dos modelos de regressão PLS para proteína bruta e aminoácidos em farelo de soja

Avaliou-se os modelos de previsão utilizando PLS, com os dados a partir de 115 amostras de calibração e 34 amostras do conjunto de validação. Foram feitos 3 tipos de estudos utilizando a organização das variáveis:

- No primeiro estudo foi considerado o espectro inteiro (551 variáveis) e foram concluídos simultaneamente os 19 modelos de calibração.

- No segundo estudo foi realizada uma seleção de variáveis e utilizou-se 90 variáveis com ênfase na análise de proteína bruta .

- No terceiro estudo foram selecionados conjuntos de variáveis diferentes para cada aminoácido e foram desenvolvidos 18 modelos de calibração para cada aminoácido separados para tentar sugerir uma região do espectro que melhor identificasse cada tipo de aminoácido do farelo de soja .

Os resultados serão demonstrados e discutidos a seguir, com ênfase para o terceiro estudo, por ser o mais completo e englobar a idéia dos demais. Uma discussão detalhada dos dois primeiros estudos ficaria longa e repetitiva e por isso não é apresentada neste trabalho.

4.1- Primeiro Estudo : Espectro Completo

Os resultados finais dos modelos de calibração para cada constituinte estudado estão resumidos na tabela 5 , onde:

AA = aminoácidos

ACAL = amostras de calibração

AVAL= amostras de validação

VL= variáveis latentes

SEC= erro padrão de calibração

RSQC= coeficiente de correlação do conjunto de calibração

SECV= erro padrão da validação cruzada

1-VR= coeficiente de correlação da validação cruzada

MAX= valor máximo dos dados de calibração

MIN = valor mínimo dos dados de calibração

SD = desvio padrão do conjunto de calibração

Tabela 5: Dados de calibração dos modelos PLS de proteína bruta e aminoácidos.

RESUMO DOS MODELOS DE CALIBRAÇÃO PLS PARA PROTEÍNA BRUTA E AMINOÁCIDOS										
AA	ACal	AVai	VL	SEC	RSQC	SECV	1-VR	MAX	MIN	SD
PB	112	32	8	0,3079	0,9070	0,3424	0,8880	51,0040	44,870	1,011
GLY	102	25	8	0,0170	0,8660	0,0200	0,8220	2,1496	1,880	0,037
ILE	103	26	4	0,0187	0,8410	0,0193	0,8370	2,3387	2,060	0,047
HIS	104	26	4	0,0112	0,8010	0,0120	0,7899	1,2101	1,060	0,025
ASP	111	31	7	0,0372	0,8980	0,0413	0,8810	5,8262	5,120	0,117
ALA	104	26	4	0,0194	0,8160	0,0207	0,8001	2,2488	1,970	0,045
LEU	104	25	7	0,0267	0,8870	0,0293	0,8710	3,9114	3,440	0,079
ARG	105	27	7	0,0298	0,8410	0,0324	0,8160	3,6856	3,240	0,075
THR	103	25	6	0,0132	0,8860	0,0140	0,8770	1,9358	1,700	0,039
CYS	107	28	4	0,0063	0,8300	0,0065	0,8320	0,7047	0,610	0,015
TYR	106	28	8	0,0163	0,8610	0,0184	0,8301	2,0975	1,840	0,044
PRO	102	25	8	0,0180	0,8920	0,0209	0,8602	2,6124	2,300	0,055
SER	99	23	4	0,0194	0,8400	0,0205	0,8300	2,5584	2,260	0,049
VAL	103	25	6	0,0164	0,8920	0,0174	0,8810	2,4816	2,190	0,050
PHE	106	27	8	0,0217	0,8630	0,0246	0,8302	2,6655	2,290	0,059
GLU	102	23	6	0,0619	0,9020	0,0680	0,8912	9,7369	8,550	0,197
MET	106	27	7	0,0063	0,8220	0,0070	0,7901	0,6930	0,600	0,015
LYS	106	27	9	0,0238	0,8480	0,0298	0,7740	3,1580	2,790	0,061
TRP	112	32	8	0,0040	0,9070	0,0045	0,8830	0,6630	0,580	0,013

O mesmo foi feito para o conjunto de validação resumido na tabela 6 , onde:

BIAS = diferença entre as médias

SEP= erro padrão de previsão

SEPC= erro padrão de performace

RSQV= coeficiente de correlação do conjunto de validação

SLOPE= inclinação da reta entre o valor previsto e o valor de referência

Tabela 6: Dados de validação dos modelos PLS de proteína bruta e aminoácidos

RESUMO DOS MODELOS DE VALIDAÇÃO PARA PB E AA					
AA	SEP	SEPC	SLOPE	BIAS	RSQV
PB	0,1820	0,1550	0,9520	-0,0100	0,8861
GLY	0,0130	0,0130	0,9040	-0,0040	0,8171
ILE	0,0120	0,0120	1,0020	-0,0040	0,8320
HIS	0,0050	0,0050	0,9490	0,0000	0,7720
ASP	0,0260	0,0260	0,9560	0,0030	0,8761
ALA	0,0090	0,0090	1,0400	-0,0020	0,7903
LEU	0,0130	0,0130	1,0190	-0,0010	0,8634
ARG	0,0120	0,0120	0,9500	-0,0010	0,8133
THR	0,0090	0,0090	1,0090	-0,0010	0,8711
CYS	0,0060	0,0040	0,8740	-0,0010	0,8214
TYR	0,0110	0,0110	0,9000	-0,0030	0,8203
PRO	0,0140	0,0140	0,8810	-0,0030	0,8533
SER	0,0140	0,0140	0,9700	-0,0020	0,8214
VAL	0,0150	0,0150	0,9520	-0,0020	0,8789
PHE	0,0190	0,0170	0,9590	0,0030	0,8266
GLU	0,0540	0,0540	0,9890	-0,0090	0,8814
MET	0,0040	0,0040	0,9540	-0,0040	0,7800
LYS	0,0600	0,0150	0,9110	0,0000	0,7644
TRP	0,0030	0,0030	0,9010	-0,0070	0,8861

Avaliando a performance dos modelos PLS, é possível notar que todos os parâmetros dos aminoácidos apresentaram valores bem parecidos. A razão é o fato dos valores dos dados experimentais serem menores que 10%. O mesmo não ocorre com a proteína bruta devido a faixa medida ser maior que 44%. A validação está dentro do esperado para os valores estatísticos obtidos, ou seja, o SEPC está menor ou igual ao SEP em todos os modelos. Outro ponto importante ainda considerando o SEPC é a constatação que todos os valores estão menores que o SECV mostrado na tabela 6. Todos os slopes estão próximos a 1 que é a faixa ideal pois, teoricamente este valor deve ser maior do que 0,7 e menor que 1,3. O valor de bias deve ser até 0,6(SEC), todos os modelos apresentam valores bem menores que essa regra o que reforça a boa qualidade de ajuste dos dados deste estudo.

O número de variáveis latentes é escolhido automaticamente pelo programa, porém percebe-se que o valor oscilou de 4 à 9. Essa oscilação ocorreu provavelmente devido ao fato de alguns aminoácidos apresentarem um alto coeficiente de variação, possibilitado a modelagem do erro analítico e um conseqüente aumento do número de variáveis latentes.

Os valores de correlação estão bem próximos a 1 para os modelos de calibração RSQC, validação RSQV e validação cruzada 1-VR. Ainda dentro do contexto de

correlação vemos que a diferença do RSQC e RSQV não ultrapassa 12%, o que indica que há uma grande similaridade entre os valores de calibração e validação.

Os valores de máximos e mínimos para cada modelo cobrem a maioria da faixa de concentração de proteína bruta e aminoácidos no farelo de soja brasileiro, o que garante a amplitude dos modelos para prever amostras desconhecidas.

O modelo para proteína bruta aparenta ter números de incerteza maiores, porém considerando a concentração medida que representa a soma de todos os aminoácidos referidos, vemos que assim como para os aminoácidos todos os valores estão dentro dos limites estatísticos desejáveis.

Avaliando os resultados de proteína bruta frente ao método de referência é possível notar que a diferença dos valores de previsão é menor que 1,5%, o que é considerado excelente pelas tabelas de desvio analítico ANFAL (Associação Nacional dos Fabricantes de Alimentos para Animais) ⁽¹⁾ seguida dentro dos laboratórios de nutrição animal, esta instituição admite até 2% de variação para valores de proteína bruta.

Já os valores de aminoácidos apresentaram uma variação analítica um pouco maior, pois quanto menor é o valor maior é variação aceitável. A literatura não é muito específica para esses valores, porém é possível afirmar que nos dados de validação a variação foi menor que 10% num faixa de valores de concentração que varia de 0,6 à 9% em teor de aminoácidos. Como já foi citado há erro analítico modelado junto com os dados de calibração mas são perfeitamente aceitáveis para as formulações de nutrição animal, visto que seguem o que de melhor se pode obter no método de referência.

No geral os modelos de calibração aqui apresentados atendem as necessidades de qualquer laboratório que queira quantificar o teor de proteína bruta e aminoácidos em farelo de soja a baixo custo e com boa precisão por um método secundário de quantificação. Esse é o raciocínio utilizado pela grande maioria das indústrias que trabalham com o sistema de infravermelho próximo (NIRS), pois a grande preocupação está nas estatísticas que validam o conjunto de medidas. Se matematicamente o modelo é válido, quimicamente deixa a desejar do ponto de vista da análise dos pesos (gráfico 8) e dos coeficientes de regressão B (gráfico 9). O grande problema é que

os coeficientes de regressão B para os 19 modelos são iguais, por isso são representados apenas por um único gráfico. Embora, devido aos pesos apresentarem de 4 à 9 variáveis latentes, justificados provavelmente pelo erro analítico ser diferente para cada constituinte (aminoácido ou proteína) a variação nos gráficos é bastante sutil, por isso foram representados também por um único gráfico. Como as proteínas são compostas por seqüência de aminoácidos que normalmente se repetem, é natural que seja utilizada a mesma relação matemática para calcular todas as medidas.

Outra questão são as medidas relacionadas a aminoácidos como cistina, metionina e triptofano que apresentam uma pequena representatividade na estrutura total de uma proteína, o que pode levar a crer que o que está sendo modelado é puramente erro analítico, já que o desvio padrão do método de referência é relativamente alto, ou seja, é preciso definir bandas espectrais específicas para esses aminoácidos.

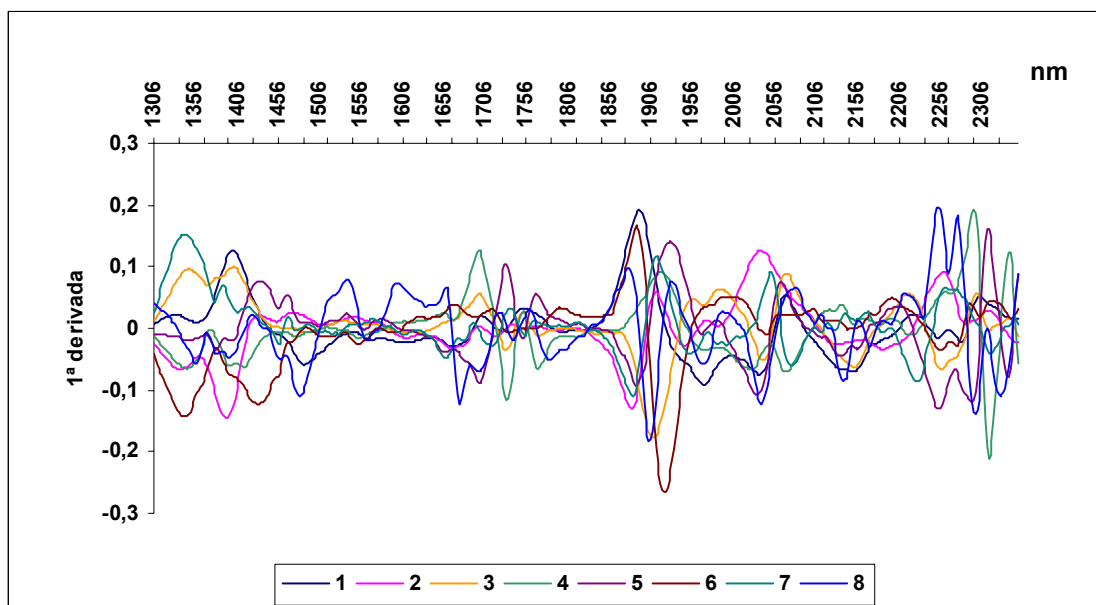


Gráfico 8: Pesos do modelo de PB, que seguem a mesma tendência para todos modelos demonstrados no primeiro estudo, com menos ou mais variáveis latentes.

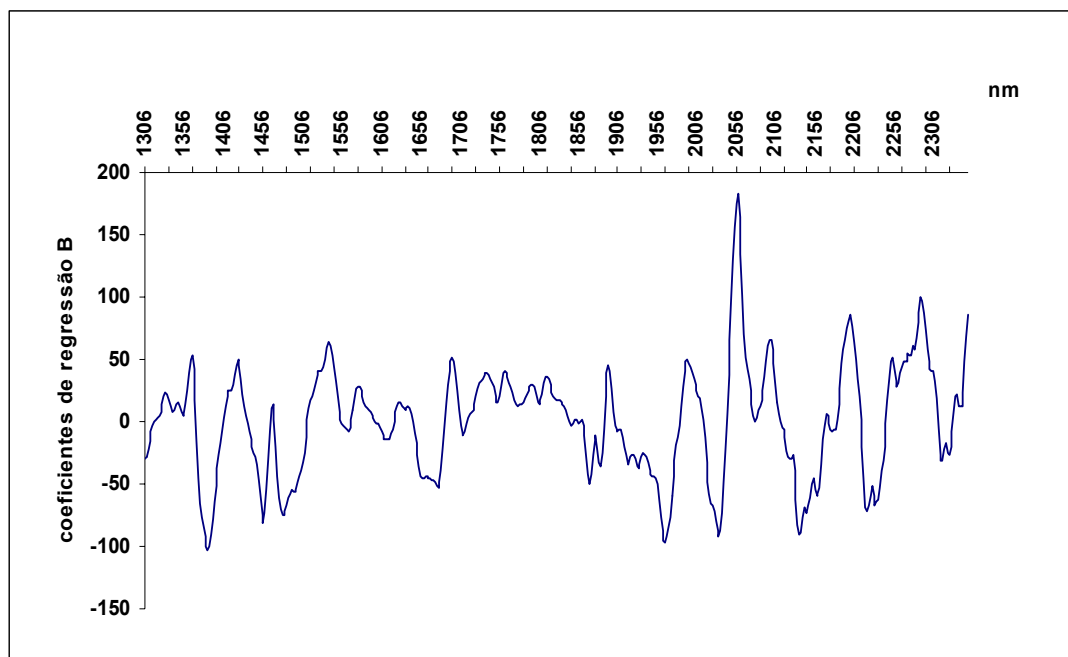


Gráfico 9: Coeficientes de regressão B do modelo de PB, que são idênticos para os 18 modelos apresentados no primeiro estudo.

Além disso um grande número de variáveis pode indicar o uso de informação repetida, e nos estudos atuais de quimiometria a seleção de variáveis busca cada dia mais a otimização dos cálculos e explicação química dos dados gerados nos modelos de calibração matemáticos, validados estatisticamente.

4.2 – Segundo Estudo: seleção de variáveis com ênfase na proteína bruta do farelo de soja

Quando surge a questão de otimizar um modelo de calibração, a redução do número de variáveis é a primeira opção para se trabalhar. Como aminoácidos são as partes de uma proteína, logo a região de bandas que será selecionada para proteína bruta também apresentará bom resultados para todos os 18 aminoácidos, por isso será analisada a questão da proteína para reduzir variáveis e analisar os pesos e os coeficientes de regressão B para depois, dentre essas bandas selecionadas, buscar a melhor região para cada aminoácido.

No espectro total é possível obter informações sobre toda a constituição química de uma dada matéria-prima, sabe-se que o NIR é a melhor técnica para se trabalhar

com umidade, gordura, fibra além da proteína bruta. As melhores bandas de absorção para esses constituintes estão bem estabelecidas na literatura. Assim é possível isolar as regiões que melhor descrevem as proteínas com uma boa margem de segurança .

Nesse caso foram selecionadas 90 variáveis, que apresentaram bons resultados estatísticos tanto na calibração quanto na validação das amostras de farelo de soja. Essas variáveis foram escolhidas com base em valores de literatura ⁽⁴²⁾. Foram utilizadas 110 amostras para o conjunto de calibração e 27 amostras para o conjunto de validação, o modelo foi desenvolvido com 4 variáveis latentes. O valor de coeficiente de correlação foi maior que 0,88 para o conjunto de calibração, validação cruzada e validação externa. Os erros padrões de calibração e validação foram abaixo de 0.34, e a faixa de concentração utilizada é bem ampla (de 44,94 à 51,02 %), como mostra a tabela 7:

Tabela 7 : Dados do modelo PLS para proteína bruta em farelo de soja.

Dado	Valor
Conjunto de calibração	110
Conjunto de validação	27
Número de Variáveis Latentes	4
Erro Padrão de Calibração (%)	0,3372
Coeficiente de Correlação do Conjunto de Calibração	0,8905
Coeficiente de Correlação da Validação Cruzada	0,8895
Erro Padrão de Previsão (%)	0,2570
Slope	0,9390
Bias	-0,0060
Coeficiente de Correlação do Conjunto de Validação	0,8720
Número de variáveis	90
1356-1366,2 nm	CH/CH ₂
1416-1448,2 nm	AROMÁTICO (AR)
1506-1516,2 nm	PB
1680-1690,2 nm	AROMÁTICO (AR)
1692-1700,2 nm	CH/CH ₃
1736-1746,2 nm	SH
1760-1770,2 nm	CH/CH ₂
1896-1906,2 nm	COOH
1946-1956,2 nm	COOH
2050-2066,2 nm	NH ₂ /PB
2086-2096,2 nm	OH
2176-2186,2 nm	NH ₂ /PB
2296-2304,2 nm	NH ₂ /PB
Teor mínimo (%)	44,9411
Teor máximo (%)	51,0284
Diferença entre os teores (%)	6,0878

Os pesos que descrevem o modelo são mostrados no gráfico 10:

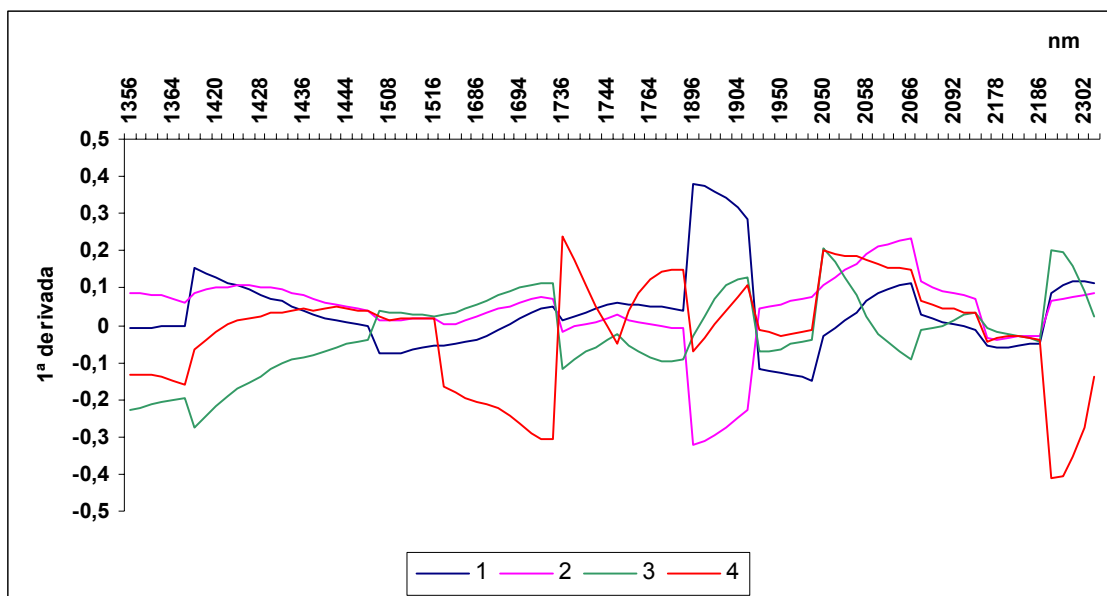


Gráfico 10: Pesos do modelo PLS para proteína bruta em farelo de soja.

Os Coeficientes de regressão B utilizados são mostrados no gráfico 11:

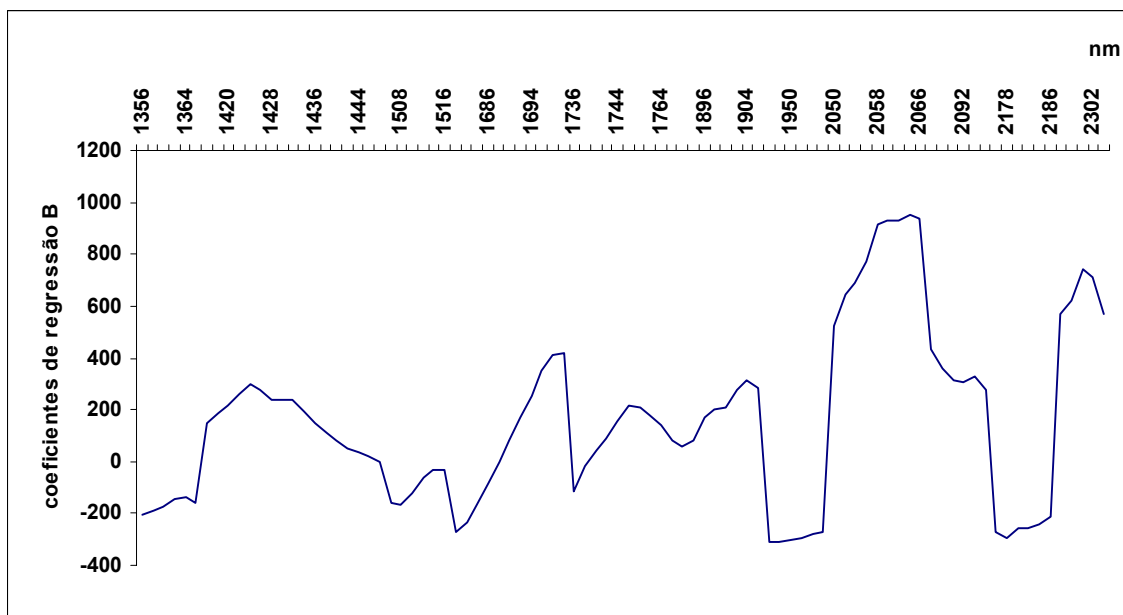


Gráfico 11 : Coeficientes de regressão B do modelo PLS para proteína bruta em farelo de soja.

Os valores previstos e as estatísticas do conjunto de validação externa para proteína bruta são mostrados na tabela 8.

Tabela 8: Valores e estatísticas do conjunto de validação externa PLS para proteína bruta.

nº	Real (%)	Previsto (%)	CV%	dif (real -prev)
1	45,170	45,172	0,003	-0,002
2	50,529	50,875	0,483	-0,346
3	50,200	50,250	0,070	-0,050
4	47,500	47,443	0,085	0,057
5	47,220	47,224	0,006	-0,004
6	47,500	47,330	0,254	0,170
7	50,300	50,584	0,398	-0,284
8	47,409	47,867	0,680	-0,458
9	48,100	48,429	0,482	-0,329
10	48,600	48,180	0,614	0,420
11	47,410	47,711	0,448	-0,301
12	47,700	47,967	0,395	-0,267
13	48,110	48,114	0,006	-0,004
14	48,020	47,529	0,727	0,491
15	47,900	47,852	0,071	0,048
16	47,200	47,273	0,109	-0,073
17	47,500	47,266	0,349	0,234
18	46,800	46,825	0,038	-0,025
19	46,500	46,554	0,082	-0,054
20	48,500	48,485	0,022	0,015
21	46,600	46,978	0,571	-0,378
22	47,300	47,252	0,072	0,048
23	48,100	48,088	0,018	0,012
24	47,877	48,138	0,384	-0,261
25	48,030	47,447	0,864	0,583
26	47,602	47,743	0,209	-0,141
27	48,010	48,085	0,110	-0,075
			MIN-MAX (%)	44,940 à 51,030
SD Método de Referência (%)			0,100	dif. Ideal (%)
				3,00 à 5,00
CV Método de Referência (%)			2,00	dif. Real (%)
				60,87

Visualizando os pesos (gráfico 10) e os coeficientes de regressão B (gráfico 11) é possível notar a importância dos comprimentos de onda 1506-1516 nm, 1680-1690 nm, 2050-2066 nm e 2176-2186nm que mostram uma boa correlação entre os dados. Já a análise do desvio padrão de 0,001 à 0,412% mostra que os valores estão bem abaixo dos 2% de variação analítica aceitos pelas tabelas da ANFAL. Esses valores se devem à técnica de proteína bruta por detector de condutividade térmica apresentar boa repetibilidade e robustez. Comparando com o primeiro estudo é possível notar que estavam sendo utilizadas 4 variáveis a mais para descrever informações que não eram necessárias ao modelo visto que os valores estatísticos de calibração e validação

(externa e cruzada), são tão satisfatórios quanto o modelo sugerido no primeiro estudo que era bem mais complexo.

Existe uma recomendação da ASTM para que a faixa de concentração entre os valores a serem calibrados apresentem diferença mínima igual à 3 ou 5 vezes o coeficiente de variação da análise ⁽⁵⁰⁾ para evitar que a resposta possa ser resultado de erro analítico puro. É possível observar que a diferença é superior a 6 vezes o valor necessário (como mostra a tabela 8), o que dá uma boa margem de segurança.

Uma observação importante quando se compara os resultados de proteína bruta com o valor da soma dos valores de aminoácido é o fato do valor da soma ser sempre inferior ao valor de proteína bruta. Essa diferença é atribuída ao fator empírico de conversão do nitrogênio em proteína (6,25). Este fator foi determinado por Kjeldahl na albumina da carne bovina, sendo generalizado para todos os tipos de alimentos embora seja claro que cada tipo de alimento possua um valor característico que pode variar de aproximadamente 5,2 à 6,9. Outra consideração a ser notada é que o método dosa nitrogênio total, que pode vir não só das proteínas como de purinas, creatinina, pirimidinas, vitaminas e amidas (como uréia), contidas nos alimentos. É muito comum o erro de transformar o nitrogênio de fontes não protéicas em proteínas. Contudo, essas diferenças são bem aceitas desde que não sejam ocasionadas por fraudes como a adição intencional de uréia para aumentar o valor de proteína de um farelo de soja rico em casca, o que diminui seu valor nutricional em proteínas, por exemplo.

4.3 – Terceiro Estudo : seleção de variáveis com ênfase nos aminoácidos do farelo de soja

Cada aminoácido de uma cadeia protéica tem sua importância nutricional e está presente numa determinada quantidade em cada alimento. O farelo de soja é sem dúvida a maior, melhor e mais completa fonte de aminoácidos entre as matérias-primas utilizadas em nutrição animal ⁽⁴⁹⁾. Sua composição vai de 1 à 20% do seu conteúdo de proteína bruta e o grande problema é uma determinação precisa dos valores que estão na faixa de até 2% (metionina, triptofano e cistina).

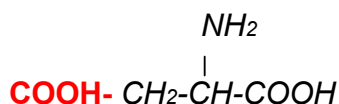
Os resultados serão apresentados em três grupos:

- O primeiro grupo compreende os 10 aminoácidos (ASP, GLU, VAL, ARG, TYR, MET, CYS, LEU, TRP e THR) cujos modelos de calibração seguem as normas ASTM para análises quantitativas no infravermelho próximo ⁽⁵⁰⁾ , além de apresentar boa correlação matemática e estatística.
- O segundo grupo é formado por 4 aminoácidos (LYS, PHE, ILE e ALA) cujos modelos de calibração precisam ter suas faixas de concentração ampliadas, para se enquadrarem dentro da recomendação ASTM, porém os mesmos podem ser utilizados com cautela pois apresentam limitações estatísticas.
- O terceiro grupo trás 4 aminoácidos (GLY, PRO, SER e HIS) que não apresentaram modelos de calibração satisfatórios devido ao desvio padrão apresentado no método de referência ser alto e não atender as normas ASTM. Esses modelos para serem refeitos seria necessário utilizar o método de derivatização pré-coluna que apresenta baixos desvios para esses aminoácidos ⁽⁵¹⁾ e não levaria a resultados tendenciosos que aparentemente são bons , mas que analisados com maior cuidado são produto da modelagem de puro erro analítico.

4.3.1- Aminoácidos Determinados:

4.3.1.1 – Aminoácido Ácido Aspártico

Fórmula molecular:



O ácido aspártico é um aminoácido carboxilado. Além das duas bandas utilizadas para grupo amina e ácido carboxílico estará sendo incluída uma banda na região 1356-1400 nm, específica para ligações do tipo CH - CH₂ , essa modificação visa criar uma condição analítica específica para quantificar esse aminoácido.

Foram utilizadas 114 amostras para o conjunto de calibração e 25 amostras para o conjunto de validação externa, sendo necessárias 6 variáveis latentes para descrever o modelo com 45 variáveis selecionadas. Os coeficientes de correlação estão entre 0,75 à 0,82 com erros de previsão iguais ou inferiores à 0,049, como mostra a tabela 9:

Tabela 9: Dados do modelo PLS para ácido aspártico em farelo de soja.

Dado	Valor
Conjunto de calibração	114
Conjunto de validação	25
Número de Variáveis Latentes	6
Erro Padrão de Calibração (%)	0,0495
Coeficiente de Correlação do Conjunto de Calibração	0,8220
Coeficiente de Correlação da Validação Cruzada	0,7751
Erro Padrão de Previsão (%)	0,0490
Slope	0,9870
Bias	-0,0010
Coeficiente de Correlação do Conjunto de Validação	0,7510
Número de variáveis	45
1356-1400,2 nm	CH/CH ₂
1896-1906,2 nm	C=O/COOH
2040-2070,2 nm	NH ₂ / PB
Teor mínimo (%)	5,1212
Teor máximo (%)	5,8237
Diferença entre os teores (%)	0,7025

Os pesos que descrevem o modelo são mostrados no gráfico 12, onde é possível notar que existem uma boa quantidade de informações relacionado os valores com os coeficientes de regressão B (gráfico 13), sendo estas variáveis uma combinação de bandas de absorção região que pode ser melhor explorada para o aminoácido ASP. Os valores previstos e as estatísticas do conjunto de validação externa são mostrados na tabela 10, onde é possível notar que o modelo apresentou coeficientes de variação entre 0,013 e 0,971%, além do faixa de concentração ser 25 vezes maior do que o desvio padrão do método de referência, o que garante uma boa segurança nas medidas geradas por esse modelo.

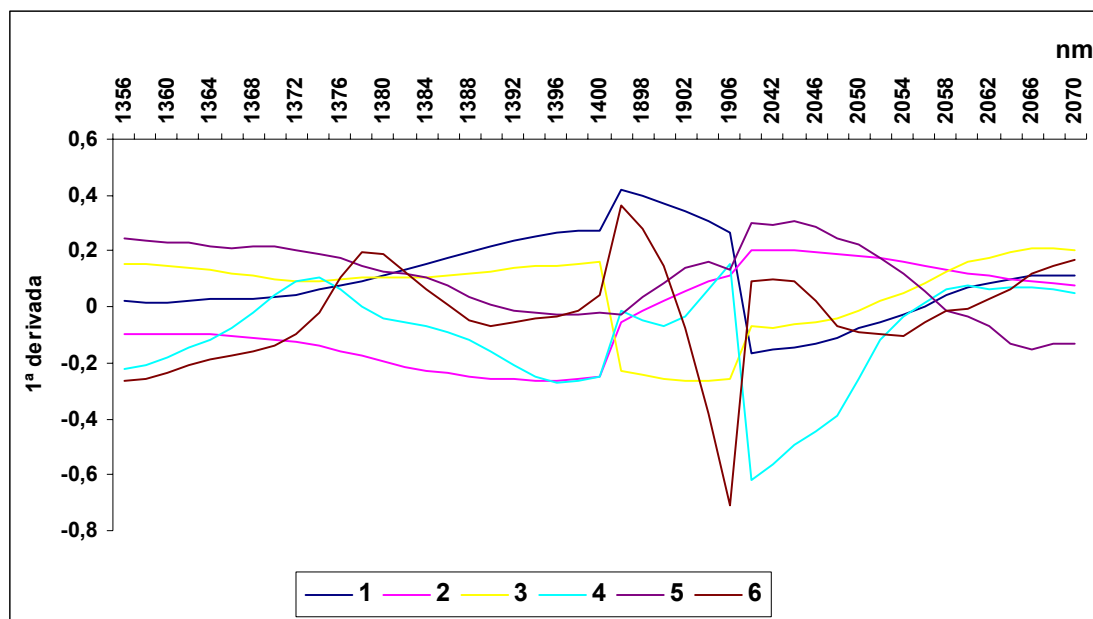


Gráfico 12: Pesos do modelo PLS para ácido aspártico em farelo de soja.

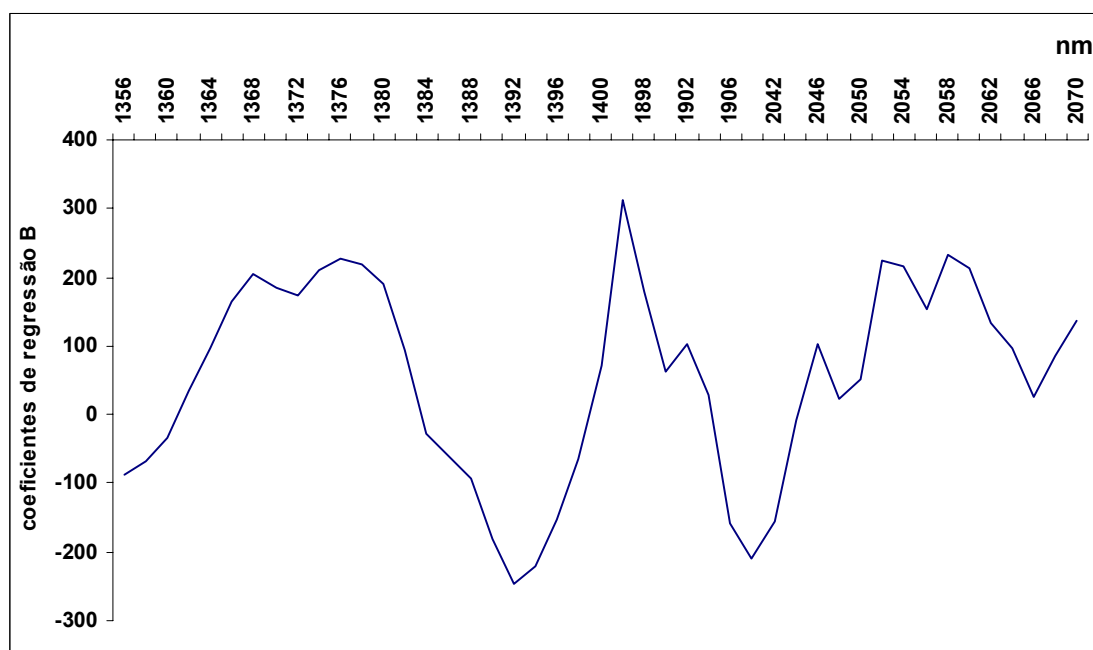


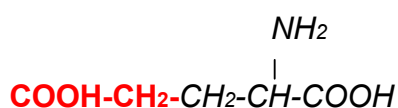
Gráfico 13: Coeficientes de regressão B do modelo PLS para ácido aspártico em farelo de soja.

Tabela 10: Valores e estatísticas do conjunto de validação externa PLS para ASP.

nº	Real (%)	Previsto (%)	CV%	dif (real -prev)
1	5,290	5,274	0,214	0,016
2	5,775	5,797	0,269	-0,022
3	5,416	5,440	0,313	-0,024
4	5,506	5,504	0,025	0,002
5	5,403	5,419	0,209	-0,016
6	5,440	5,405	0,456	0,035
7	5,419	5,431	0,156	-0,012
8	5,444	5,448	0,052	-0,004
9	5,498	5,508	0,128	-0,010
10	5,570	5,494	0,971	0,076
11	5,486	5,485	0,013	0,001
12	5,480	5,451	0,375	0,029
13	5,466	5,469	0,039	-0,003
14	5,361	5,409	0,630	-0,048
15	5,422	5,350	0,945	0,072
16	5,479	5,475	0,052	0,004
17	5,618	5,610	0,101	0,008
18	5,326	5,394	0,897	-0,068
19	5,406	5,378	0,367	0,028
20	5,500	5,450	0,646	0,050
21	5,472	5,526	0,694	-0,054
22	5,507	5,495	0,154	0,012
23	5,441	5,440	0,013	0,001
24	5,458	5,500	0,542	-0,042
25	5,484	5,476	0,103	0,008
			MIN - MAX (%)	5,824 à 5,121
SD Método de Referência (%) 0,026			dif. Ideal (%)	0,078 à 0,130
CV Método de Referência (%) 2,40			dif. Real (%)	0,703

4.3.1.2 – Aminoácido Ácido Glutâmico

Fórmula molecular:



O ácido glutâmico é um aminoácido carboxilado. Foram utilizadas 110 amostras para o conjunto de calibração e 27 amostras para o conjunto de validação externa, sendo necessárias 5 variáveis latentes para descrever o modelo com 28 variáveis selecionadas. As regiões selecionadas compreendem duas bandas de absorção marcantes para ácido carboxílico e uma com ênfase no grupo amina. Os coeficientes

de correlação estão entre 0,73 à 0,77 com erros de previsão iguais ou inferiores à 0,10, como mostra a tabela 11:

Tabela 11: Dados do modelo PLS para ácido glutâmico em farelo de soja.

Dado	Valor
Conjunto de calibração	110
Conjunto de validação	27
Número de Variáveis Latentes	5
Erro Padrão de Calibração (%)	0,0969
Coeficiente de Correlação do Conjunto de Calibração	0,7730
Coeficiente de Correlação da Validação Cruzada	0,7421
Erro Padrão de Previsão (%)	0,1000
Slope	0,8110
Bias	-0,001
Coeficiente de Correlação do Conjunto de Validação	0,7380
Número de variáveis	28
1896-1906,2 nm	C=O/COOH
1946-1956,2 nm	C=O/COOH
2040-2070,2 nm	NH ₂
Teor mínimo (%)	9,7275
Teor máximo (%)	8,5488
Diferença entre os teores (%)	1,1787

Os pesos que descrevem o modelo são mostrados no gráfico 14, é possível notar que existe informação relevantes nas regiões selecionadas observando os gráficos de pesos e coeficientes de regressão B (gráfico 15) .

Segundo a tabela 12, os coeficientes de variação estão entre 0,00 e 0,884%, além da faixa de concentração ser 16 vezes maior que o desvio padrão do método de referência o que garante uma boa segurança nas medidas geradas por esse modelo.

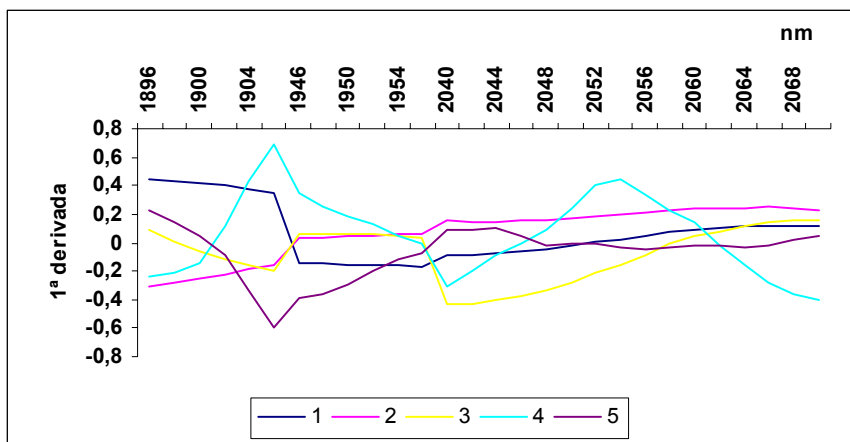


Gráfico 14: Pesos do modelo PLS para ácido glutâmico em farelo de soja.

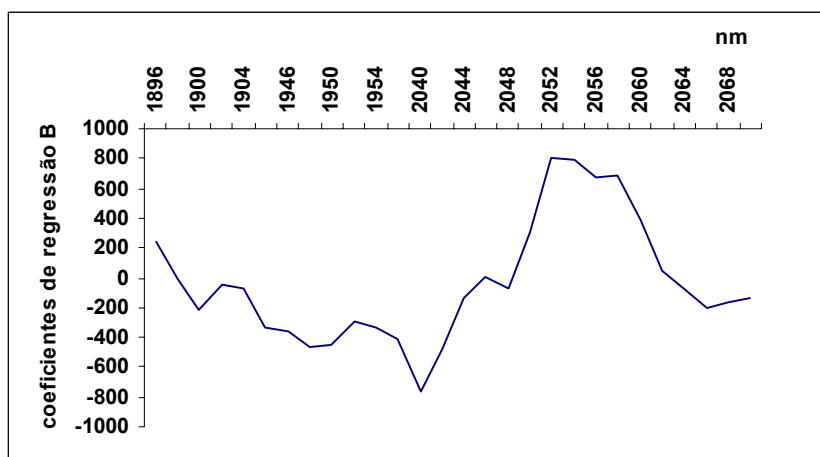


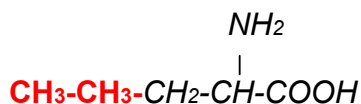
Gráfico 15: Coeficientes de regressão B do modelo PLS para ácido glutâmico em farelo de soja.

Tabela 12: Valores e estatísticas do conjunto de validação externa PLS para GLU.

nº	Real (%)	Previsto (%)	CV%	dif (real -prev)
1	8,677	8,708	0,252	-0,031
2	8,963	8,980	0,134	-0,017
3	9,231	9,200	0,238	0,031
4	9,015	9,002	0,102	0,013
5	9,100	8,987	0,884	0,113
6	9,041	9,096	0,429	-0,055
7	9,304	9,266	0,289	0,038
8	9,173	9,188	0,116	-0,015
9	9,293	9,300	0,054	-0,007
10	9,074	9,032	0,328	0,042
11	9,154	9,116	0,294	0,038
12	9,135	9,159	0,186	-0,024
13	8,963	8,960	0,024	0,003
14	9,120	9,122	0,016	-0,002
15	8,944	9,041	0,763	-0,097
16	8,986	8,980	0,047	0,006
17	9,020	9,012	0,063	0,008
18	9,170	9,150	0,154	0,020
19	8,791	8,851	0,481	-0,060
20	9,154	9,142	0,093	0,012
21	8,887	8,981	0,744	-0,094
22	9,020	9,068	0,375	-0,048
23	9,200	9,200	0,000	0,000
24	9,130	9,213	0,640	-0,083
25	9,009	9,004	0,039	0,005
26	9,078	9,166	0,682	-0,088
27	9,150	9,172	0,170	-0,022
			MIN-MAX (%)	9,723 à 8,549
SD Método de Referência (%) 0,073			dif. Ideal (%)	0,219 à 0,365
CV Método de Referência (%) 1,02			dif. Real (%)	1,179

4.3.1.3 – Aminoácido Valina

Fórmula molecular:



A valina é um aminoácido alifático ou de cadeia aberta. Foram escolhidas 4 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 108 amostras para o conjunto de calibração e 28 amostras para o conjunto de validação externa, sendo necessárias 4 variáveis latentes para descrever o modelo (gráfico 16), com 30 variáveis selecionadas. Os coeficientes de correlação estão entre 0,72 à 0,77 com erros de previsão iguais ou inferiores à 0,024%, (tabela 13) e os coeficientes de regressão B (gráfico 17) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,06 e 1,77%, a faixa de concentração utilizada é 7,5 vezes maior que o desvio padrão da análise.

Tabela 13: Dados do modelo PLS para valina em farelo de soja.

Dado	Valor
Conjunto de calibração	108
Conjunto de validação	28
Número de Variáveis Latentes	4
Erro Padrão de Calibração (%)	0,0241
Coeficiente de Correlação do Conjunto de Calibração	0,7769
Coeficiente de Correlação da Validação Cruzada	0,7476
Erro Padrão de Previsão (%)	0,0240
Slope	0,7260
Bias	-0,0030
Coeficiente de Correlação do Conjunto de Validação	0,7210
Número de variáveis	30
1506-1516,nm	PB-NH ₂
1690-1700, nm	CH-CH ₃
1896-1906,nm	C=O/COOH
2166-2188,nm	PB-NH ₂
Teor mínimo (%)	2,1768
Teor máximo (%)	2,4829
Diferença entre os teores (%)	0,3061

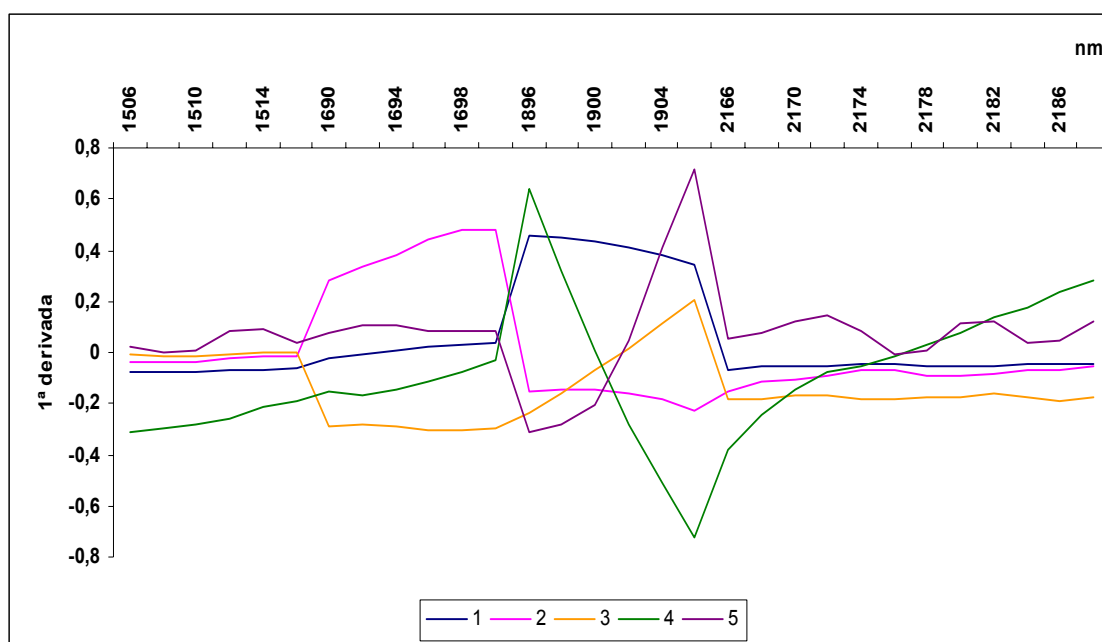


Gráfico 16: Pesos do modelo PLS para valina em farelo de soja.

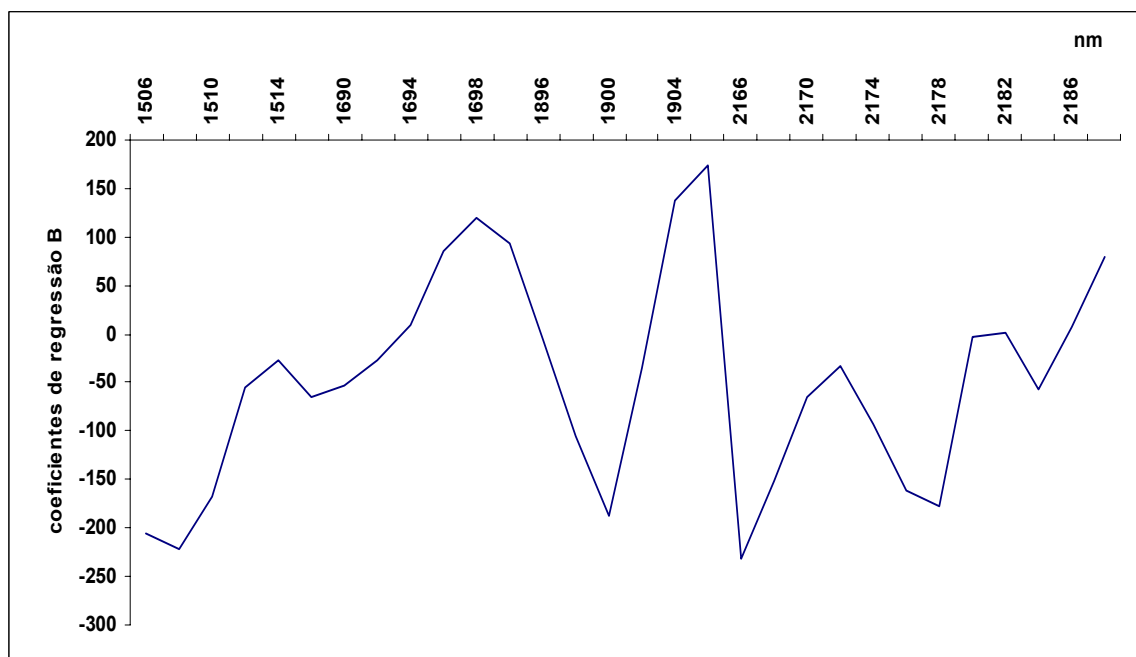


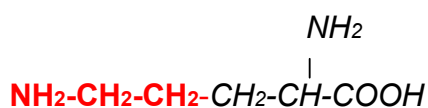
Gráfico 17: Coeficientes de regressão B do modelo PLS para valina em farelo de soja.

Tabela 14: Valores e estatísticas do conjunto de validação externa PLS para VAL.

nº	Real (%)	Previsto (%)	CV%	dif (real -prev.)
1	2,216	2,214	0,06	0,002
2	2,310	2,304	0,18	0,006
3	2,357	2,388	0,92	-0,031
4	2,302	2,285	0,52	0,017
5	2,300	2,286	0,43	0,014
6	2,309	2,318	0,28	-0,009
7	2,376	2,364	0,36	0,012
8	2,342	2,345	0,09	-0,003
9	2,370	2,339	0,93	0,031
10	2,317	2,329	0,37	-0,012
11	2,338	2,332	0,18	0,006
12	2,333	2,392	1,77	-0,059
13	2,330	2,328	0,06	0,002
14	2,329	2,326	0,09	0,003
15	2,284	2,304	0,62	-0,020
16	2,295	2,323	0,86	-0,028
17	2,320	2,296	0,74	0,024
18	2,335	2,299	1,10	0,036
19	2,245	2,265	0,63	-0,020
20	2,338	2,347	0,27	-0,009
21	2,269	2,298	0,90	-0,029
22	2,304	2,298	0,18	0,006
23	2,346	2,313	1,00	0,033
24	2,332	2,328	0,12	0,004
25	2,301	2,332	0,95	-0,031
26	2,318	2,312	0,18	0,006
27	2,320	2,375	1,66	-0,055
28	2,337	2,324	0,39	0,013
valores			MIN-MAX (%)	2,177 à 2,483
SD Método de Referência (%) 0,042			dif. Ideal (%)	0,126 à 0,210
CV Método de Referência (%) 1,520			dif. Real (%)	0,306

4.3.1.5 – Aminoácido Arginina

Fórmula molecular:



A arginina é um aminoácido básico. Foram escolhidas 4 faixas espectrais de absorção para tentar descrever a estrutura molecular desse aminoácido. Utilizou-se 108 amostras para o conjunto de calibração e 27 amostras para o conjunto de validação externa, sendo necessárias 3 variáveis latentes para descrever o modelo, com 28 variáveis selecionadas. Os coeficientes de correlação estão entre 0,71 à 0,76 com erros de previsão iguais ou inferiores à 0,035%, como mostra a tabela 15.

Os pesos que descrevem o modelo são mostrados no gráfico 18, as regiões escolhidas apresentaram bons resultados para o aminoácido arginina. Os resultados para a previsão do conjunto externo são mostrados na tabela 16, por essa tabela, pode-se notar que os valores são 6,28 vezes maiores que o desvio padrão do método de referência. Esse modelo apresenta boa concordância entre os dados uma vez que é baixo o coeficiente de variação da análise de referência (1,83%), sendo as regiões selecionadas aparentemente ideais uma vez que reduziram o número de variáveis latentes de 7, primeiro estudo, para 3 nas condições atuais.

Tabela 15: Dados do modelo PLS para arginina em farelo de soja.

Dado	Valor
Conjunto de calibração	108
Conjunto de validação	27
Número de Variáveis Latentes	3
Erro Padrão de Calibração (%)	0,0359
Coeficiente de Correlação do Conjunto de Calibração	0,7676
Coeficiente de Correlação da Validação Cruzada	0,7392
Erro Padrão de Previsão (%)	0,0350
Slope	0,7190
Bias	0,000
Coeficiente de Correlação do Conjunto de Validação	0,7190
Número de variáveis	27
1506-1516, nm	NH-PB
1720-1730, nm	CH/CH ₂
1896-1906, nm	COOH
2050-2066, nm	PB-NH
Teor mínimo (%)	3,2373
Teor máximo (%)	3,6835
Diferença entre os teores (%)	0,4462

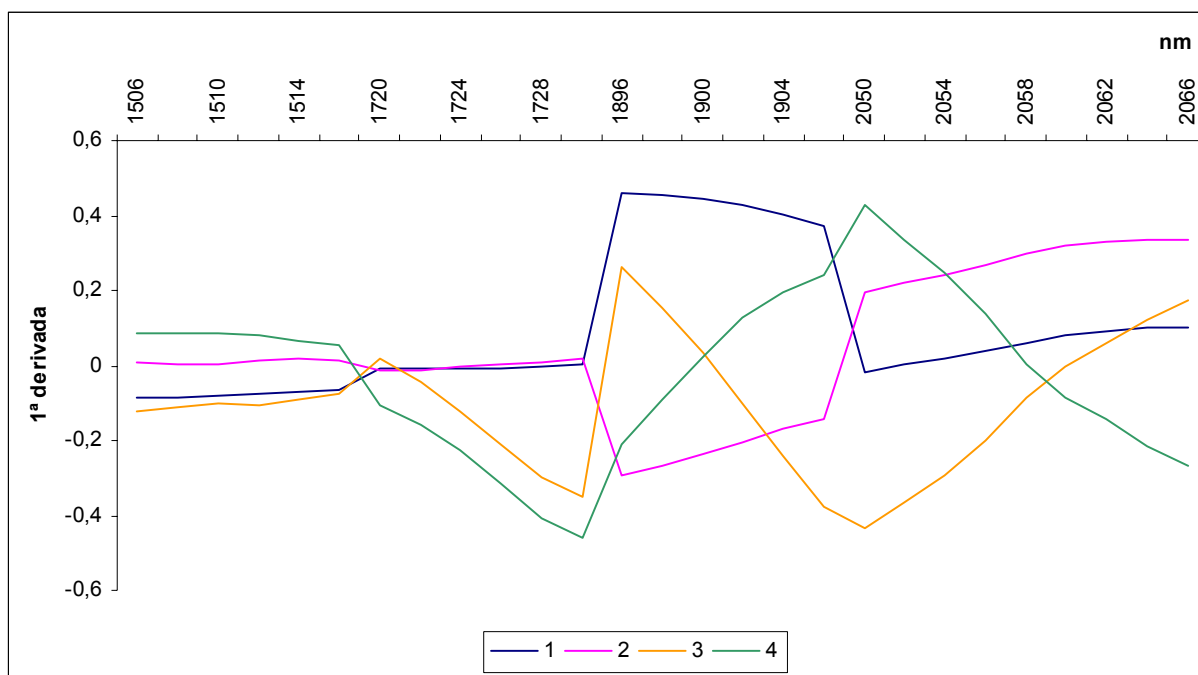


Gráfico 18: Pesos do modelo PLS para arginina em farelo de soja.

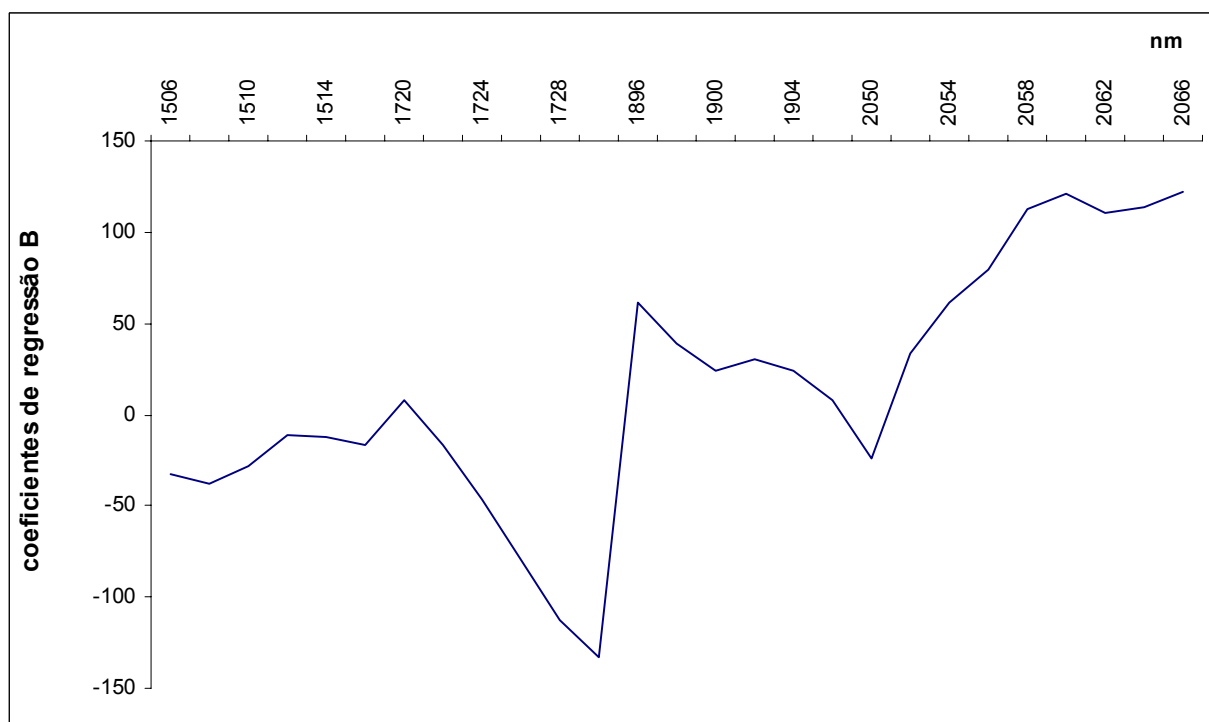


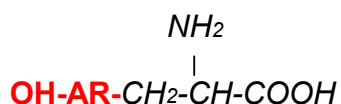
Gráfico 19: Coeficientes de regressão B do modelo PLS para arginina em farelo de soja.

Tabela 16: Valores e estatísticas do conjunto de validação externa PLS para ARG.

nº	Real (%)	Previsto (%)	CV%	dif (real -prev.)
1	3,350	3,342	0,17	0,008
2	3,393	3,424	0,64	-0,031
3	3,487	3,566	1,58	-0,079
4	3,413	3,409	0,08	0,004
5	3,458	3,416	0,86	0,042
6	3,423	3,448	0,51	-0,025
7	3,523	3,501	0,44	0,022
8	3,473	3,490	0,35	-0,017
9	3,515	3,472	0,87	0,043
10	3,435	3,434	0,02	0,001
11	3,466	3,460	0,12	0,006
12	3,450	3,432	0,37	0,018
13	3,440	3,454	0,29	-0,014
14	3,386	3,412	0,54	-0,026
15	3,402	3,424	0,46	-0,022
16	3,454	3,403	1,05	0,051
17	3,450	3,382	1,41	0,068
18	3,328	3,352	0,51	-0,024
19	3,520	3,486	0,69	0,034
20	3,365	3,401	0,75	-0,036
21	3,460	3,423	0,76	0,037
22	3,466	3,452	0,29	0,014
23	3,457	3,476	0,39	-0,019
24	3,460	3,418	0,86	0,042
25	3,437	3,450	0,27	-0,013
26	3,460	3,525	1,32	-0,065
27	3,464	3,472	0,16	-0,008
			MIN-MAX	3,237 - 3,683
SD Método de Referência (%) 0,071			dif. Ideal	0,3550
CV Método de Referência (%) 1,830			Dif. Real	0,4462

4.3.1.5 - Aminoácido Tirosina

Fórmula molecular:



A tirosina é um aminoácido aromático. Foram escolhidas 4 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 107 amostras para o conjunto de calibração e 28 amostras para o conjunto de validação externa, sendo necessárias 5 variáveis latentes para descrever o modelo com 27 variáveis

selecionadas. Os coeficientes de correlação estão entre 0,72 à 0,80 com erros de previsão iguais ou inferiores à 0,019 (tabela 17).

Os Pesos (gráfico 20) e os coeficientes de regressão B (gráfico 21) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,035 e 1,68 , sendo que a faixa de concentração utilizada está 4,41 vezes maior que o desvio padrão da análise (0,059 %), como mostra a tabela 18.

Tabela 17: Dados do modelo PLS para tirosina em farelo de soja.

Dado	Valor
Conjunto de calibração	107
Conjunto de validação	28
Número de Variáveis Latentes	6
Erro Padrão de Calibração (%)	0,0190
Coeficiente de Correlação do Conjunto de Calibração	0,8010
Coeficiente de Correlação da Validação Cruzada	0,7515
Erro Padrão de Previsão (%)	0,0290
Slope	0,8960
Bias	-0,003
Coeficiente de Correlação do Conjunto de Validação	0,7501
Número de variáveis	27
1418-1442,2 nm	OH-AR
1760-1768,2 nm	CH / CH ₂
1896-1900,2 nm	C=O /COOH
2086-2096,2 nm	CONH ₂ / OH
Teor mínimo (%)	1,8375
Teor máximo (%)	2,0977
Diferença entre os teores (%)	0,2602

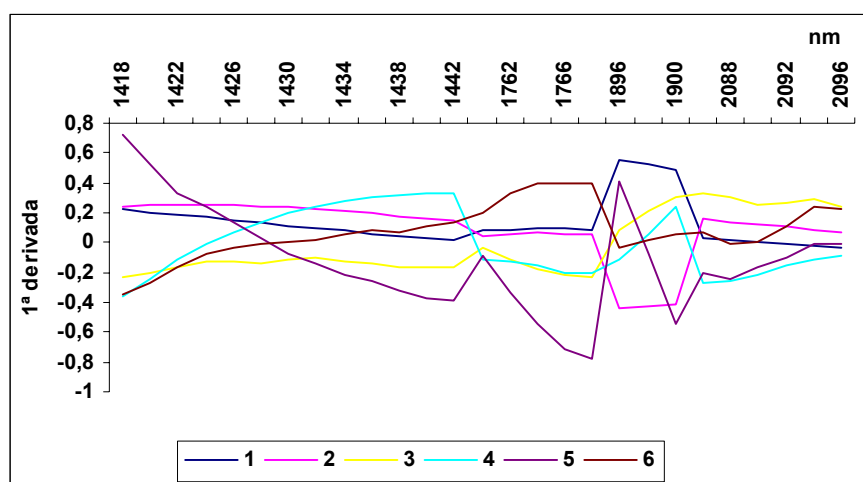


Gráfico 20: Pesos do modelo PLS para tirosina em farelo de soja.

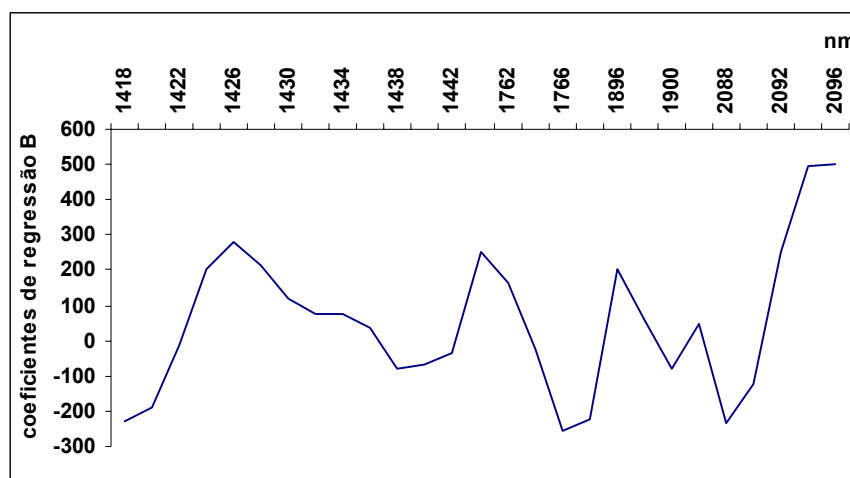


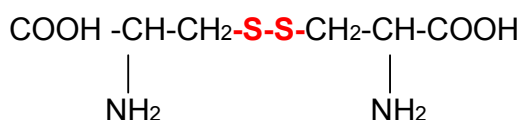
Gráfico 21: Coeficientes de regressão B do modelo PLS para tirosina em farelo de soja.

Tabela 18: Valores e estatísticas do conjunto de validação externa PLS para TYR.

nº	Real (%)	Previsto (%)	CV%	dif (real- prev.)
1	1,910	1,902	0,297	0,008
2	1,865	1,889	0,904	-0,024
3	1,927	1,935	0,293	-0,008
4	1,970	2,013	1,527	-0,043
5	1,938	1,958	0,726	-0,020
6	1,945	1,961	0,579	-0,016
7	2,042	2,075	1,134	-0,033
8	1,944	1,952	0,290	-0,008
9	2,000	2,001	0,035	-0,001
10	1,972	1,994	0,784	-0,022
11	1,971	1,989	0,643	-0,018
12	1,951	1,979	1,008	-0,028
13	1,948	1,976	1,009	-0,028
14	1,966	1,967	0,036	-0,001
15	1,961	1,958	0,108	0,003
16	1,923	1,937	0,513	-0,014
17	1,961	1,947	0,507	0,014
18	1,946	1,931	0,547	0,015
19	1,890	1,904	0,522	-0,014
20	1,968	1,981	0,466	-0,013
21	1,911	1,949	1,392	-0,038
22	1,939	1,923	0,586	0,016
23	1,964	1,939	0,906	0,025
24	1,963	1,978	0,538	-0,015
25	1,937	1,964	0,979	-0,027
26	1,952	1,943	0,327	0,009
27	1,957	2,004	1,678	-0,047
28	1,967	1,976	0,323	-0,009
			MIN-MAX	1,838 à 2,098
SD Método de Referência (%) 0,059			dif. Ideal	0,177 à 0,295
CV Método de Referência (%) 3,28			dif. Real	0,260

4.3.1.6- Aminoácido Cistina

Fórmula molecular:



A cistina é um aminoácido sulfurado, produto em alguns casos da oxidação da cisteína. Foram escolhidas 4 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 107 amostras para o conjunto de calibração e 24 amostras para o conjunto de validação externa, sendo necessárias 5 variáveis latentes para descrever o modelo com 22 variáveis selecionadas. Os coeficientes de correlação estão entre 0,79 à 0,82 com erros de previsão iguais ou inferiores à 0,0060 (tabela 19).

Os pesos (gráfico 22) e os coeficientes de regressão B (gráfico 23) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,00 e 1,27%. A faixa de concentração é 3,6 vezes maior que o desvio padrão da análise (0,026 %), como mostra a tabela 24.

Tabela 19: Dados do modelo PLS para cistina em farelo de soja.

Dado	Valor
Conjunto de calibração	107
Conjunto de validação	24
Número de Variáveis Latentes	3
Erro Padrão de Calibração (%)	0,0068
Coeficiente de Correlação do Conjunto de Calibração	0,8260
Coeficiente de Correlação da Validação Cruzada	0,7956
Erro Padrão de Previsão (%)	0,0060
Slope	1,0000
Bias	-0,000
Coeficiente de Correlação do Conjunto de Validação	0,8000
Número de variáveis	22
1736-1746 nm	SH/ S-S
1760-1768 nm	CH-CH ₂
1896-1900 nm	C=O / COOH
2046-2064 nm	CONH ₂
Teor mínimo (%)	0,6120
Teor máximo (%)	0,7045
Diferença entre os teores (%)	0,0925



Gráfico 22: Pesos do modelo PLS para cistina em farelo de soja.

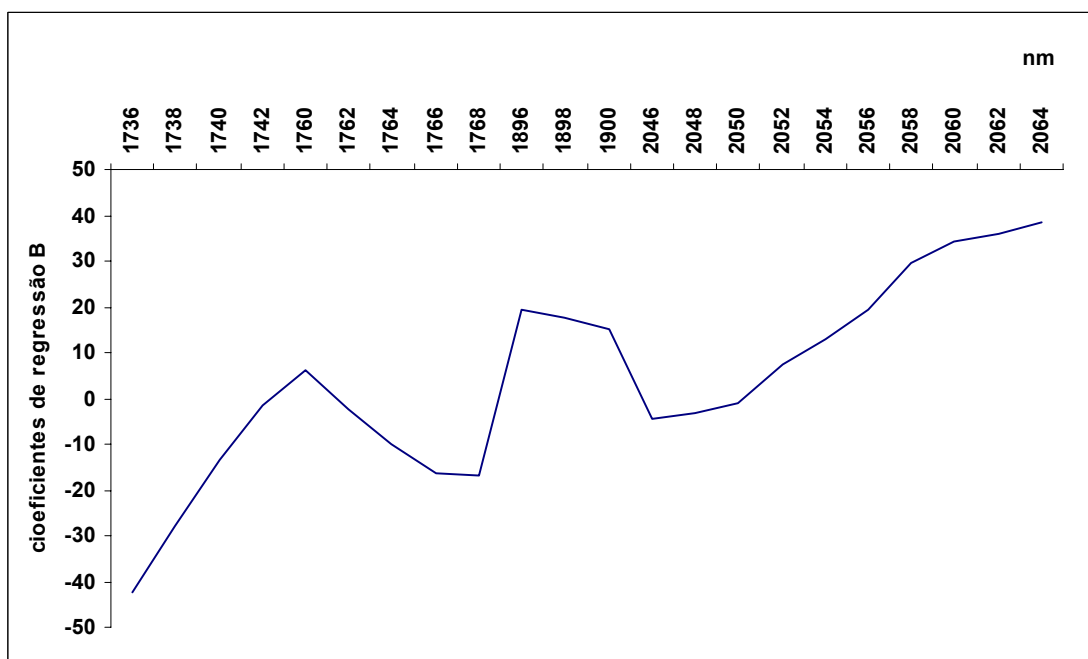


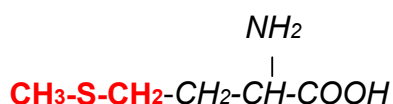
Gráfico 23: Coeficientes de regressão B do modelo PLS para cistina em farelo de soja.

Tabela19 : Valores e estatísticas do conjunto de validação externa PLS para CYS.

nº	Real (%)	Previsto (%)	CV%	dif (real - prev.)
1	0,638	0,632	0,67	0,006
2	0,644	0,655	1,20	-0,011
3	0,666	0,679	1,37	-0,013
4	0,648	0,650	0,22	-0,002
5	0,660	0,650	1,08	0,010
6	0,650	0,655	0,54	-0,005
7	0,668	0,666	0,21	0,002
8	0,659	0,665	0,64	-0,006
9	0,659	0,660	0,11	-0,001
10	0,652	0,654	0,22	-0,002
11	0,658	0,657	0,11	0,001
12	0,644	0,654	1,09	-0,010
13	0,655	0,656	0,11	-0,001
14	0,643	0,646	0,33	-0,003
15	0,661	0,645	1,73	0,016
16	0,670	0,662	0,85	0,008
17	0,638	0,649	1,21	-0,011
18	0,648	0,649	0,11	-0,001
19	0,668	0,657	1,17	0,011
20	0,656	0,662	0,64	-0,006
21	0,647	0,651	0,44	-0,004
22	0,652	0,655	0,32	-0,003
23	0,655	0,672	1,81	-0,017
24	0,657	0,660	0,32	-0,003
			MIN-MAX	0,612 à 0,705
SD Método de Referência (%) 0,026			dif. Ideal	0,0780 à 0,130
CV Método de Referência (%) 4,06			dif. Real	0,0930

4.3.1.7 - Aminoácido Metionina

Fórmula molecular:



A metionina é um aminoácido sulfurado. Foram escolhidas 5 faixas espectrais de absorção para descrever sua estrutura molecular, como mostra a tabela 20 . Utilizou-se 107 amostras para o conjunto de calibração e 26 amostras para o conjunto de validação externa, sendo necessárias 3 variáveis latentes para descrever o modelo com 31

variáveis selecionadas. Os coeficientes de correlação estão entre 0,72 à 0,79 com erros de previsão iguais ou inferiores à 0,0070%. Os pesos (gráfico 24) e os coeficientes de regressão B (gráfico 25) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,00 e 1,27%. A faixa de concentração utilizada é 3,6 vezes maior que o desvio padrão da análise (0,024 %), como mostra a tabela 21.

Tabela 21: Dados do modelo PLS para metionina em farelo de soja.

Dado	Valor
Conjunto de calibração	107
Conjunto de validação	26
Número de Variáveis Latentes	3
Erro Padrão de Calibração (%)	0,0075
Coeficiente de Correlação do Conjunto de Calibração	0,7910
Coeficiente de Correlação da Validação Cruzada	0,7358
Erro Padrão de Previsão (%)	0,0070
Slope	1,0140
Bias	0,0050
Coeficiente de Correlação do Conjunto de Validação	0,7230
Número de variáveis	31
1690-1702 nm	CH ₃
1736-1742 nm	SH/S-S
1760-1768 nm	CH-CH ₂
1896-1900 nm	C=O / COOH
2046-2064 nm	CONH ₂
Teor mínimo (%)	0,6034
Teor máximo (%)	0,6930
Diferença entre os teores (%)	0,0896

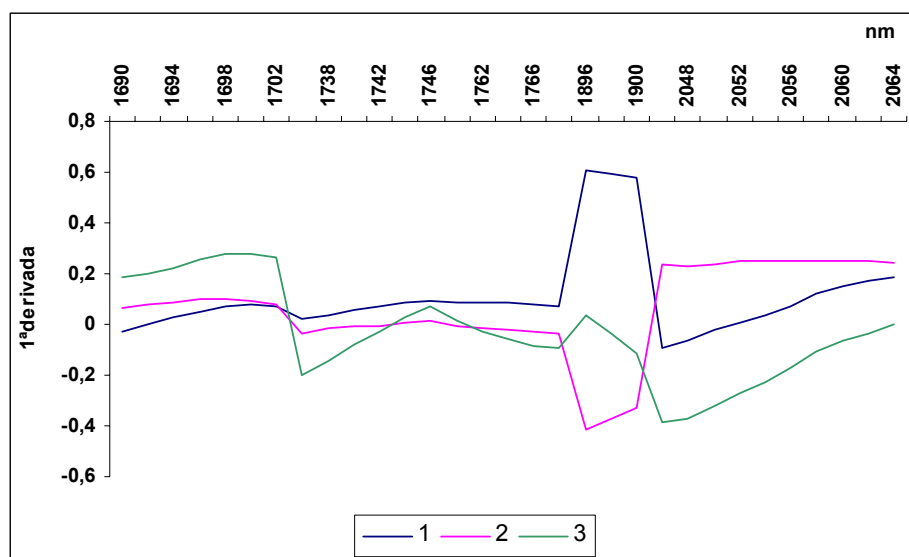


Gráfico 24: Pesos do modelo PLS para metionina em farelo de soja.

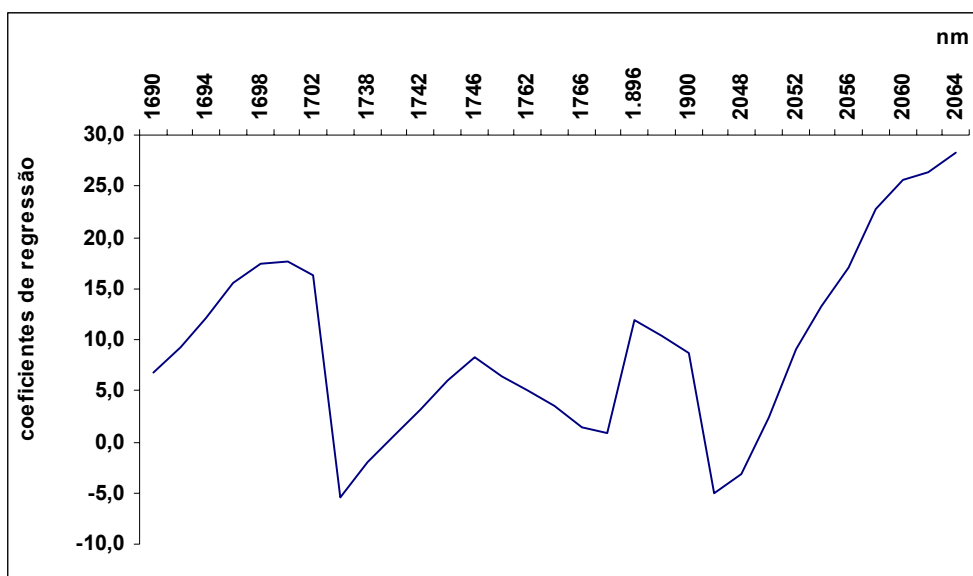


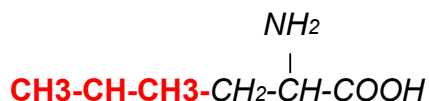
Gráfico 25: Coeficientes de regressão B do modelo PLS para metionina em farelo de soja.

Tabela 22: Valores e estatísticas do conjunto de validação externa PLS para MET.

nº	Real (%)	Previsto (%)	CV%	dif (real -prev.)
1	0,625	0,626	0,11	-0,001
2	0,635	0,642	0,78	-0,007
3	0,65	0,663	1,40	-0,013
4	0,638	0,635	0,33	0,003
5	0,64	0,64	0,00	0,000
6	0,64	0,642	0,22	-0,002
8	0,659	0,653	0,65	0,006
9	0,649	0,653	0,43	-0,004
10	0,649	0,649	0,00	0,000
11	0,642	0,645	0,33	-0,003
12	0,648	0,644	0,44	0,004
13	0,635	0,641	0,66	-0,006
14	0,646	0,645	0,11	0,001
15	0,633	0,635	0,22	-0,002
16	0,644	0,636	0,88	0,008
17	0,646	0,632	1,55	0,014
18	0,622	0,619	0,34	0,003
19	0,648	0,653	0,54	-0,005
20	0,629	0,638	1,00	-0,009
21	0,639	0,637	0,22	0,002
22	0,648	0,648	0,00	0,000
23	0,646	0,652	0,65	-0,006
24	0,652	0,642	1,09	0,010
25	0,643	0,642	0,11	0,001
26	0,646	0,663	1,84	-0,017
			MIN-MAX	0,604 à 0,693
SD Método de Referência (%) 0,024			dif. Ideal	0,072 à 0,120
CV Método de Referência (%) 6,54			dif. Real	0,0896

4.3.1.8 - Aminoácido Leucina

Fórmula molecular:



A leucina é um aminoácido alifático ou de cadeia aberta. Foram escolhidas 3 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 108 amostras para o conjunto de calibração e 26 amostras para o conjunto de validação externa, sendo necessárias 4 variáveis latentes para descrever o modelo com 48 variáveis selecionadas. Os coeficientes de correlação estão entre 0,71 à 0,78 com erros de previsão iguais ou inferiores à 0,050% (tabela 22).

Os pesos (gráfico 26) e os coeficientes de regressão B (gráfico 27) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,04 e 1,71%. A faixa de concentração é 3,4 vezes maior que o desvio padrão da análise (0,142 %), como mostra a tabela 23.

Tabela 23: Dados do modelo PLS para leucina em farelo de soja.

Dado	Valor
Conjunto de calibração	108
Conjunto de validação	26
Número de Variáveis Latentes	6
Erro Padrão de Calibração (%)	0,0343
Coeficiente de Correlação do Conjunto de Calibração	0,7816
Coeficiente de Correlação da Validação Cruzada	0,7643
Erro Padrão de Previsão (%)	0,0500
Slope	0,8100
Bias	0,0050
Coeficiente de Correlação do Conjunto de Validação	0,7100
Número de variáveis	48
1506-1536,nm	NH-PB
1720-1770,nm	CH-CH ₂
1896-1906,nm	C=O/ COOH
Teor mínimo (%)	3,4281
Teor máximo (%)	3,9112
Diferença entre os teores (%)	0,4831

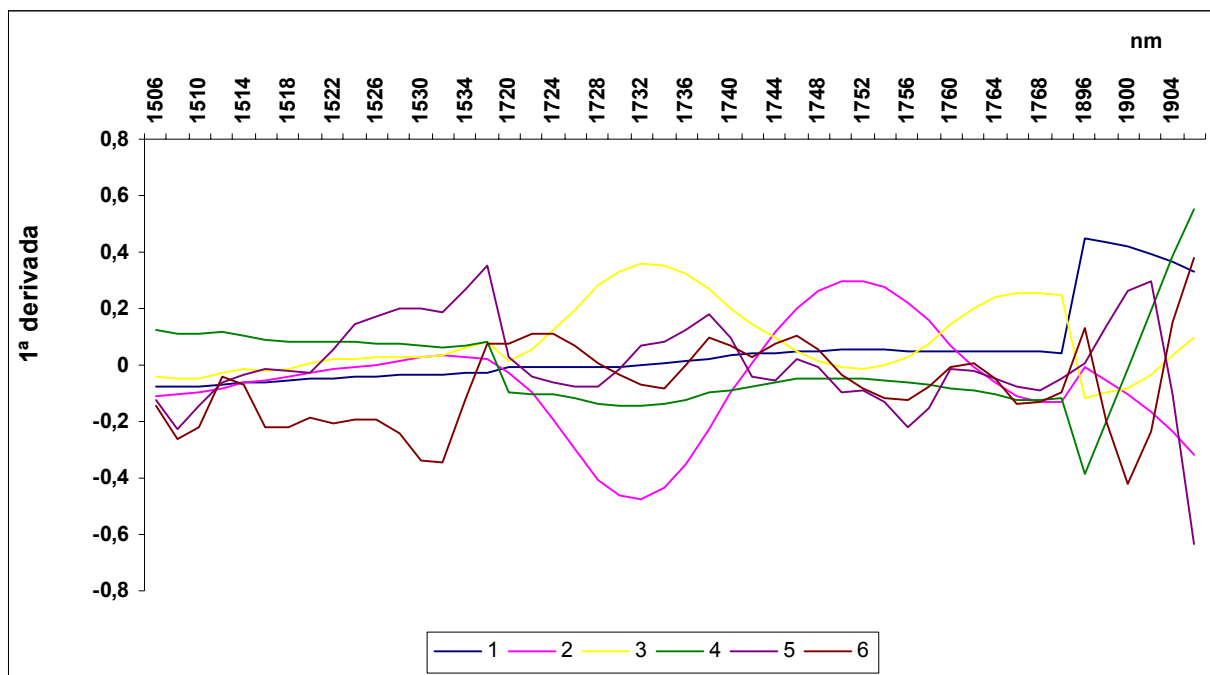


Gráfico 26: Pesos do modelo PLS para leucina em farelo de soja.

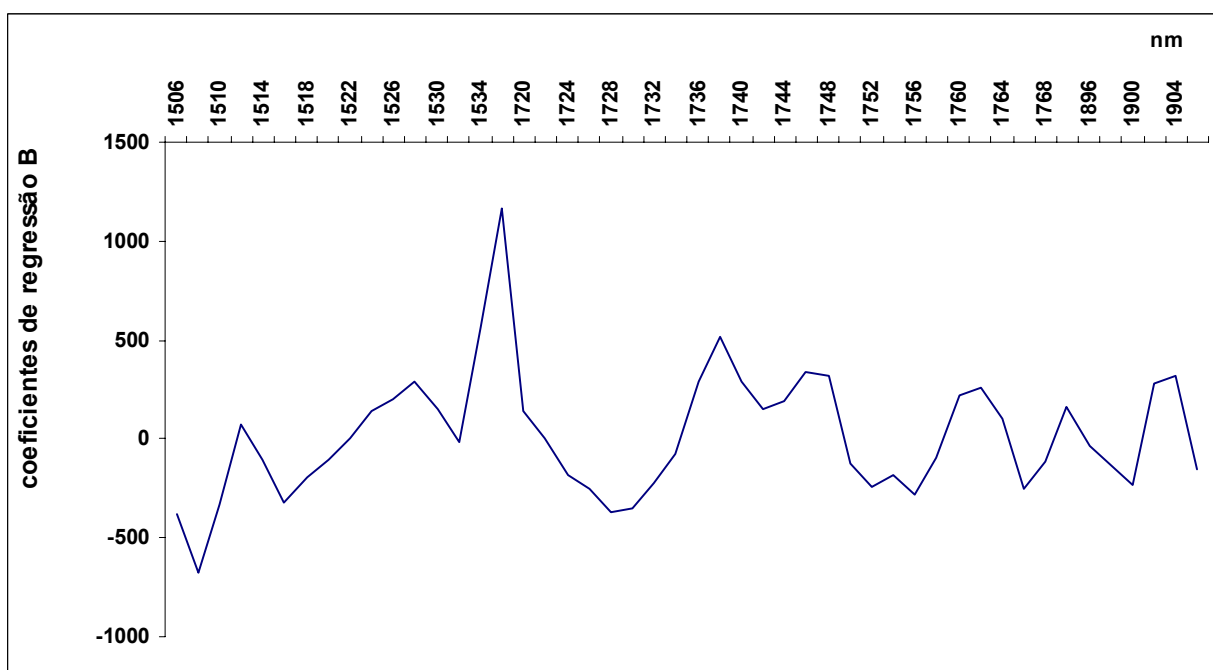


Gráfico 27 : Coeficientes de regressão B do modelo PLS para leucina em farelo de soja.

Tabela 24: Valores e estatísticas do conjunto de validação externa PLS para LEU.

nº	Real (%)	Previsto (%)	CV%	dif (real - prev.)
1	3,485	3,551	1,33	-0,066
2	3,639	3,612	0,53	0,027
3	3,621	3,608	0,25	0,013
4	3,650	3,563	1,71	0,087
5	3,632	3,652	0,39	-0,020
6	3,737	3,726	0,21	0,011
7	3,684	3,671	0,25	0,013
8	3,734	3,666	1,30	0,068
9	3,645	3,647	0,04	-0,002
10	3,677	3,668	0,17	0,009
11	3,678	3,637	0,79	0,041
12	3,663	3,654	0,17	0,009
13	3,593	3,635	0,82	-0,042
14	3,609	3,664	1,07	-0,055
15	3,662	3,589	1,42	0,073
16	3,680	3,594	1,67	0,086
17	3,644	3,579	1,27	0,065
18	3,677	3,679	0,04	-0,002
19	3,570	3,621	1,00	-0,051
20	3,623	3,660	0,72	-0,037
21	3,700	3,640	1,16	0,060
22	3,667	3,708	0,79	-0,041
23	3,710	3,673	0,71	0,037
24	3,646	3,653	0,14	-0,007
25	3,664	3,752	1,68	-0,088
26	3,634	3,695	1,18	-0,061
			MIN-MAX	3,428 à 3,911
SD Método de Referência (%) 0,142			dif. Ideal	0,426 à 0,710
CV Método de Referência (%) 3,28			dif. Real	0,4831

4.3.1.9 - Aminoácido Triptofano

Fórmula molecular:



O triptofano é um aminoácido aromático. Foram escolhidas 6 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 110 amostras para o conjunto de calibração e 21 amostras para o conjunto de validação externa, sendo necessárias 7 variáveis latentes para descrever o modelo com 62 variáveis selecionadas. Os coeficientes de correlação estão entre 0,73 à 0,79 com erros de

previsão iguais ou inferiores à 0,040 (tabela 24). Os pesos (gráfico 28) e os coeficientes de regressão B (gráfico 29) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,00 e 1,27%. A faixa de concentração utilizada é 3,2 vezes maior que o desvio padrão da análise (0,025 %), como mostra a tabela 25.

Tabela 25 : Dados do modelo PLS para triptofano em farelo de soja.

Dado	Valor
Conjunto de calibração	110
Conjunto de validação	21
Número de Variáveis Latentes	7
Erro Padrão de Calibração (%)	0,0430
Coeficiente de Correlação do Conjunto de Calibração	0,9100
Coeficiente de Correlação da Validação Cruzada	0,8685
Erro Padrão de Previsão (%)	0,0400
Slope	1,000
Bias	0,0010
Coeficiente de Correlação do Conjunto de Validação	0,8203
Número de variáveis	62
1418-1442,2 nm	AR
1496-1516,2 nm	NH/NH
1506-1536,2 nm	NH ₂
1896-1906,2 nm	C=O/COOH
2040-2070,2 nm	NH ₂ /PB
2176-2186,2 nm	NH ₂ /PB
Teor mínimo (%)	0,6631
Teor máximo (%)	0,5839
Diferença entre os teores	0,0792

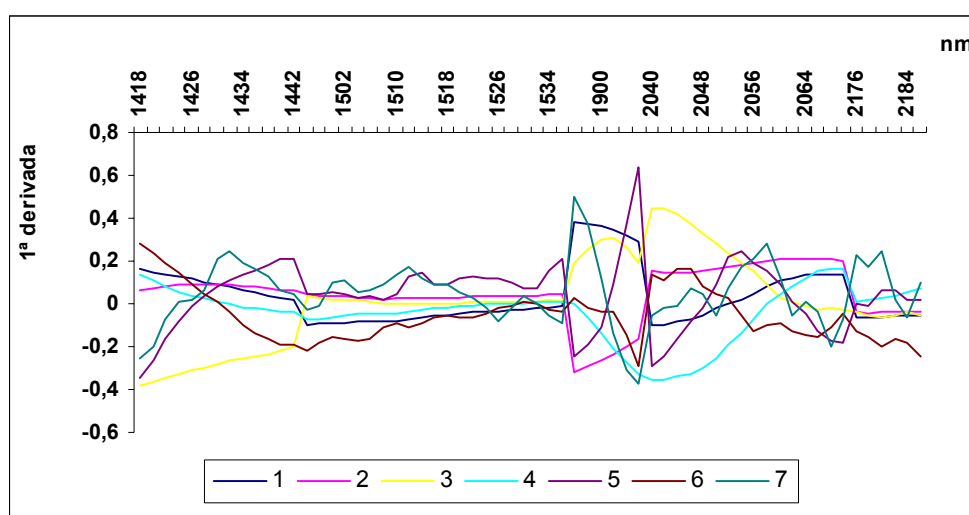


Gráfico28: Pesos do modelo PLS para triptofano em farelo de soja.

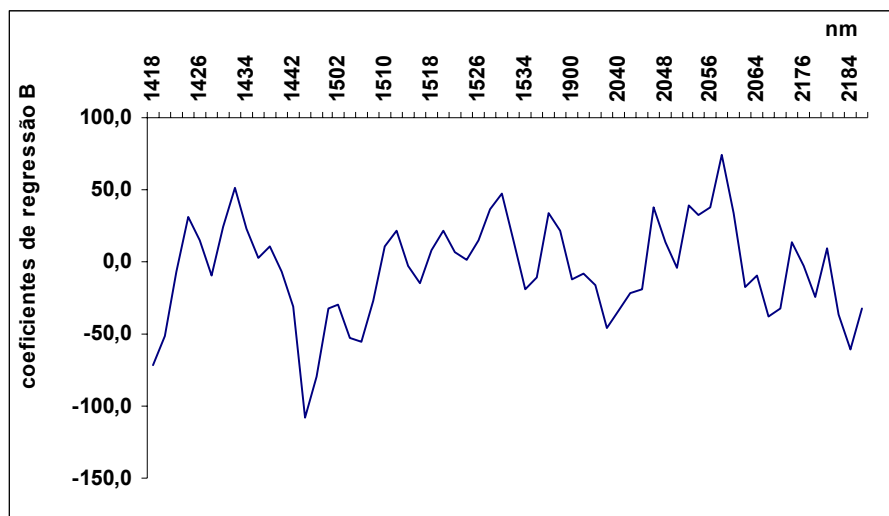


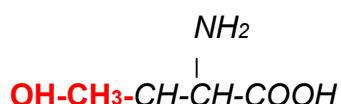
Gráfico 29: Coeficientes de regressão B do modelo PLS para triptofano em farelo de soja.

Tabela 26: Valores e estatísticas do conjunto de validação externa PLS para TRP.

nº	Real (%)	Previsto (%)	CV%	dif (real - prev.)
1	0,603	0,597	0,707	0,006
2	0,657	0,655	0,216	0,002
3	0,619	0,614	0,573	0,005
4	0,615	0,614	0,115	0,001
5	0,620	0,610	1,150	0,010
6	0,616	0,620	0,458	-0,004
7	0,625	0,630	0,563	-0,005
8	0,625	0,625	0,000	0,000
9	0,617	0,620	0,343	-0,003
10	0,626	0,623	0,340	0,003
11	0,622	0,618	0,456	0,004
12	0,610	0,613	0,347	-0,003
13	0,619	0,608	1,268	0,011
14	0,623	0,612	1,260	0,011
15	0,618	0,609	1,037	0,009
16	0,606	0,616	1,157	-0,010
17	0,615	0,617	0,230	-0,002
18	0,622	0,625	0,340	-0,003
19	0,626	0,616	1,139	0,010
20	0,619	0,624	0,569	-0,005
21	0,616	0,623	0,799	-0,007
			MIN-MAX	0,584 à 0,663
SD Método de Referência (%) 0,025			dif. Ideal	0,075 à 0,125
CV Método de Referência % 6,700			dif. Real	0,0792

4.3.1.10 - Aminoácido Treonina

Fórmula molecular:



A treonina é um aminoácido hidroxilado. Foram escolhidas 3 faixas espectrais de absorção para descrever a estrutura molecular. Utilizou-se 108 amostras para o conjunto de calibração e 27 amostras para o conjunto de validação externa, sendo necessárias 6 variáveis latentes para descrever o modelo com 48 variáveis selecionadas. Os coeficientes de correlação estão entre 0,70 à 0,79 com erros de previsão iguais ou inferiores à 0,017%, como mostra a tabela 26.

Os pesos (gráfico 30) e os coeficientes de regressão B (gráfico 31) para a treonina são demonstrados a seguir. O modelo proposto apresenta coeficientes de variação entre 0,12 e 1,93 %. A faixa de concentração corresponde à 3 vezes o valor do desvio padrão do método de referência (tabela 27).

Tabela 27 : Dados do modelo PLS para treonina em farelo de soja.

Dado	Valor
Conjunto de calibração	108
Conjunto de validação	27
Número de Variáveis Latentes	6
Erro Padrão de Calibração (%)	0,0174
Coeficiente de Correlação do Conjunto de Calibração	0,7910
Coeficiente de Correlação da Validação Cruzada	0,7538
Erro Padrão de Previsão (%)	0,0260
Slope	0,8790
Bias	-0,0040
Coeficiente de Correlação do Conjunto de Validação	0,7010
Número de variáveis	48
1506-1536,nm	NH-PB
1720-1770,nm	CH/CH ₂
1896-1906,nm	C=O/COOH
Teor mínimo (%)	1,6978
Teor máximo (%)	1,9367
Diferença entre os teores (%)	0,2389

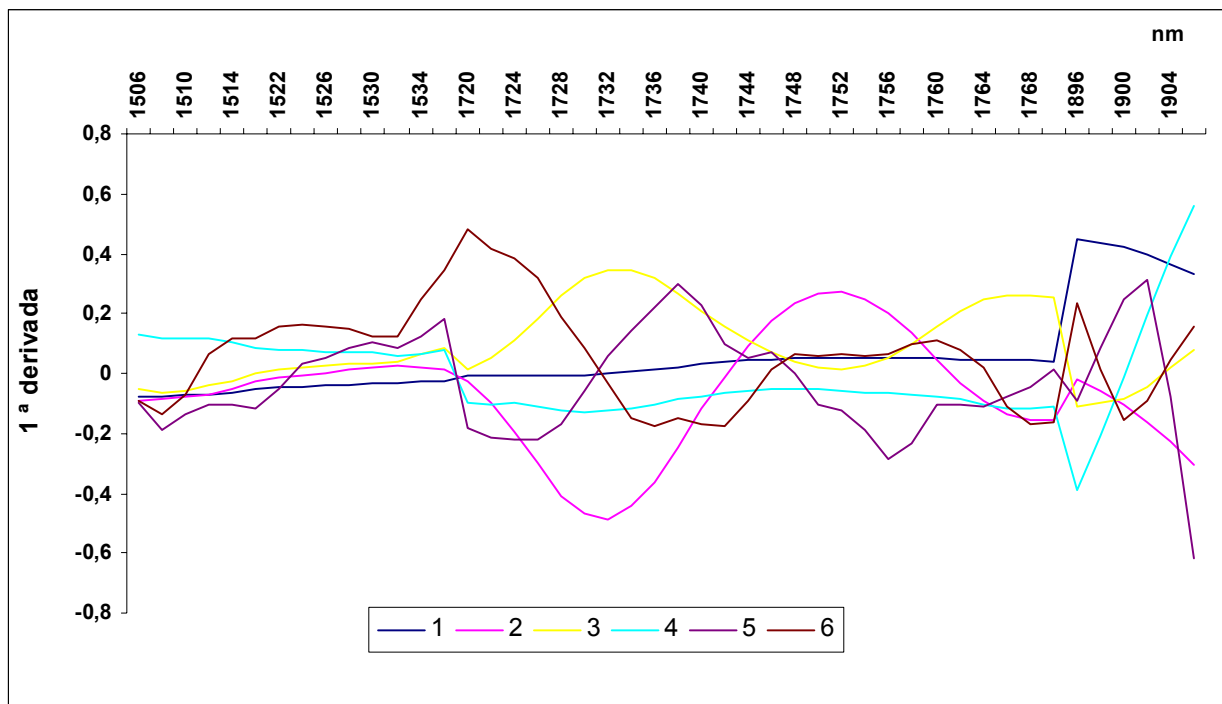


Gráfico 30: Pesos do modelo PLS para treonina em farelo de soja.

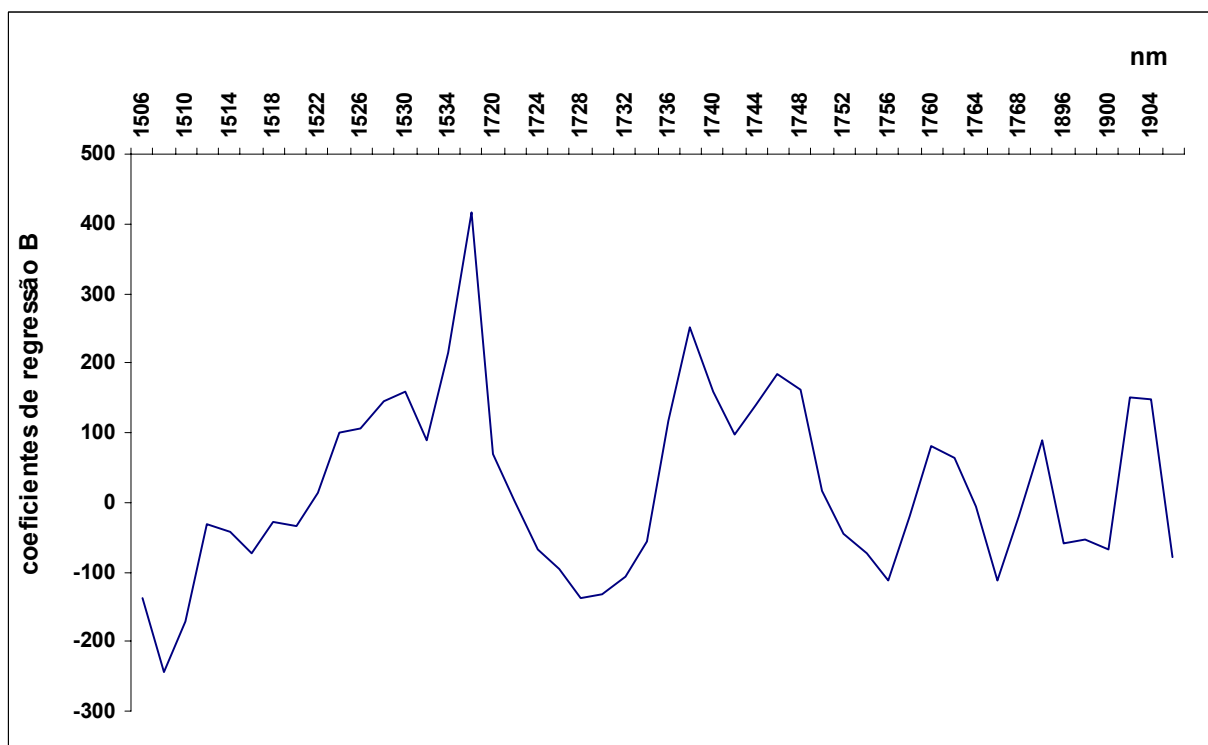


Gráfico 31: Coeficientes de regressão B para treonina em farelo de soja.

Tabela 28 : Valores e estatísticas do conjunto de validação externa PLS para THR.

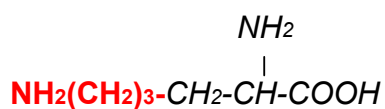
nº	Real (%)	Previsto (%)	CV%	dif (real - prev.)
1	1,724	1,720	0,16	0,004
2	1,781	1,788	0,28	-0,007
3	1,838	1,874	1,37	-0,036
4	1,792	1,786	0,24	0,006
5	1,808	1,766	1,66	0,042
6	1,797	1,806	0,35	-0,009
7	1,849	1,846	0,11	0,003
8	1,823	1,815	0,31	0,008
9	1,849	1,817	1,23	0,032
10	1,803	1,809	0,23	-0,006
11	1,819	1,816	0,12	0,003
12	1,815	1,863	1,85	-0,048
13	1,819	1,799	0,78	0,020
14	1,813	1,806	0,27	0,007
15	1,778	1,798	0,79	-0,020
16	1,786	1,815	1,14	-0,029
17	1,812	1,774	1,50	0,038
18	1,824	1,775	1,93	0,049
19	1,747	1,771	0,96	-0,024
20	1,819	1,824	0,19	-0,005
21	1,766	1,793	1,07	-0,027
22	1,793	1,808	0,59	-0,015
23	1,834	1,809	0,97	0,025
24	1,815	1,840	0,97	-0,025
25	1,790	1,819	1,14	-0,029
26	1,804	1,808	0,16	-0,004
27	1,797	1,830	1,29	-0,033
			MIN-MAX	1,698 à 1,937
SD Método de Referência (%) 0,077			dif. Ideal	0,231 à 0,385
CV Método de Referência (%) 2,84			dif. Real	0,239

4.3.2- Aminoácidos Não Determinados:

4.3.2.1 – Aminoácidos cuja faixa de concentração precisa ser ampliada

4.3.2.1.1 Aminoácido Lisina

Fórmula molecular:



Analisando a estrutura molecular é possível notar que este é um aminoácido básico. Foram escolhidas 5 faixas espectrais de absorção para tentar descrever a estrutura molecular de todas as regiões possíveis no espectro.

Foram utilizadas 111 amostras para o conjunto de calibração e 26 amostras para o conjunto de validação externa, sendo necessárias 4 variáveis latentes para descrever o modelo com 50 variáveis selecionadas. Os coeficientes de correlação estão entre 0,80 à 0,83 com erros de previsão iguais ou inferiores à 0,025%, (tabela 28).

Os pesos (gráfico 32) e os coeficientes de regressão B são demonstrados à seguir. Os coeficientes de variação estão entre 0,12 e 1,55%, o problema é que a variação da faixa de concentração das amostras está apenas 2,5 vezes maior do que o desvio padrão do método de referência (tabela 29), o que poderia afetar na qualidade das medidas apresentadas por esse modelo. Como o ideal seria um valor no mínimo 3 vezes maior, será necessário fazer uma atualização na curva aumentando essa faixa que vai de 2,79 à 3,17%, pois a variação desse tipo de amostra pode ser de 2,37 à 3,24 % em conteúdo de lisina.

Tabela 29: Dados do modelo PLS para lisina em farelo de soja.

Dado	Valor
Conjunto de calibração	111
Conjunto de validação	26
Número de Variáveis Latentes	4
Erro Padrão de Calibração (%)	0,0252
Coeficiente de Correlação do Conjunto de Calibração	0,8336
Coeficiente de Correlação da Validação Cruzada	0,8152
Erro Padrão de Previsão (%)	0,0350
Slope	0,7010
Bias	-0,004
Coeficiente de Correlação do Conjunto de Validação	0,8014
Número de variáveis	50
1356-1366,nm	CH-CH ₂
1720-1770,nm	CH-CH ₂
1896-1906,nm	COOH
2166-2178,nm	NH-PB
2296-2304,nm	NH-PB
Teor mínimo (%)	2,7956
Teor máximo (%)	3,1665
Diferença entre os teores (%)	0,3709

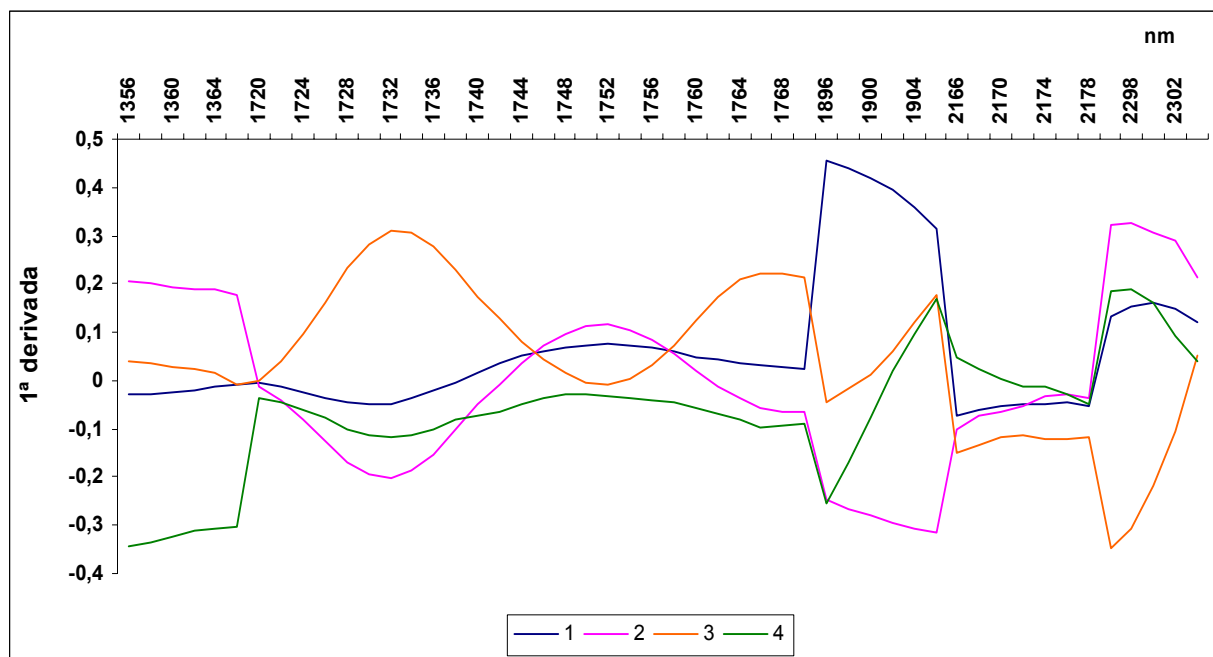


Gráfico 32: Pesos do modelo PLS para lisina em farelo de soja.

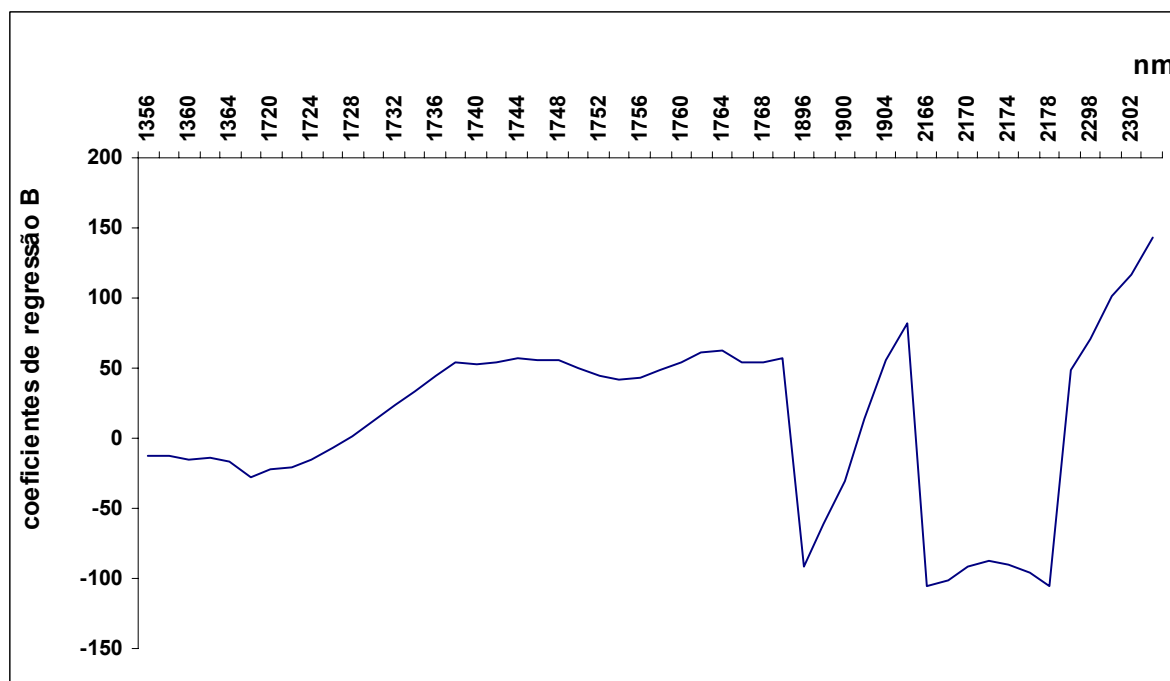


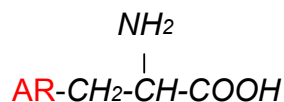
Gráfico 33: Coeficientes de regressão B do modelo PLS para lisina em farelo de soja.

Tabela 30 : Valores e estatísticas do conjunto de validação externa PLS para LYS.

nº	Real (%)	Previsto (%)	CV%	dif (real - prev.)
1	2,826	2,820	0,15	0,006
2	2,970	2,926	1,06	0,044
3	2,936	2,930	0,14	0,006
4	2,958	2,919	0,94	0,039
5	3,093	3,122	0,66	-0,029
6	2,944	2,976	0,76	-0,032
7	3,030	3,037	0,16	-0,007
8	2,987	3,007	0,47	-0,020
9	2,985	2,994	0,21	-0,009
10	3,000	2,971	0,69	0,029
11	2,981	2,997	0,38	-0,016
12	2,986	2,969	0,40	0,017
13	2,970	2,982	0,29	-0,012
14	2,912	2,941	0,70	-0,029
15	3,017	2,955	1,47	0,062
16	2,992	2,951	0,98	0,041
17	2,970	2,913	1,37	0,057
18	2,960	2,911	1,18	0,049
19	2,981	3,047	1,55	-0,066
20	2,894	2,928	0,83	-0,034
21	2,937	2,932	0,12	0,005
22	2,980	3,010	0,71	-0,030
23	2,973	2,999	0,62	-0,026
24	2,934	2,963	0,70	-0,029
25	2,956	2,969	0,31	-0,013
26	2,944	2,998	1,29	-0,054
			MINI-MAX	2,796 à 3,167
SD Método de Referência (%) 0,142			dif. Ideal	0,426 à 0,710
CV Método de Referência (%) 3,490			dif. Real	0,371

4.3.2.1.2 - Aminoácido Fenilalanina

Fórmula molecular:



A fenilalanina é um aminoácido aromático. Foram escolhidas 4 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 109 amostras para o conjunto de calibração e 28 amostras para o conjunto de validação externa, sendo

necessárias 7 variáveis latentes para descrever o modelo com 46 variáveis selecionadas. Os coeficientes de correlação estão entre 0,74 à 0,80, com erros de previsão iguais ou inferiores à 0,025 (tabela 30). Os pesos (gráfico 34) e os coeficientes de regressão B (gráfico 35) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,059 e 2,18. A faixa de concentração é 2,7 vezes maior que o desvio padrão da análise (0,128%), o ideal seria ampliar essa faixa para obter um modelo com maior precisão estatística.

Tabela 31: Dados do modelo PLS para fenilalanina em farelo de soja.

Dado	Valor
Conjunto de calibração	109
Conjunto de validação	28
Número de Variáveis Latentes	7
Erro Padrão de Calibração (%)	0,0264
Coeficiente de Correlação do Conjunto de Calibração	0,8080
Coeficiente de Correlação da Validação Cruzada	0,7436
Erro Padrão de Previsão (%)	0,0250
Slope	1,0000
Bias	0,0000
Coeficiente de Correlação do Conjunto de Validação	0,7612
Número de variáveis	46
1416-1448,2 nm	AR
1506-1524,2 nm	NH / RNH ₂
1760-1768,2 nm	CH/CH ₂
1896-1900,2 nm	C=O /COOH
2040-2070,2 nm	CONH ₂
Teor mínimo (%)	2,3138
Teor máximo (%)	2,6632
Diferença entre os teores (%)	0,3494

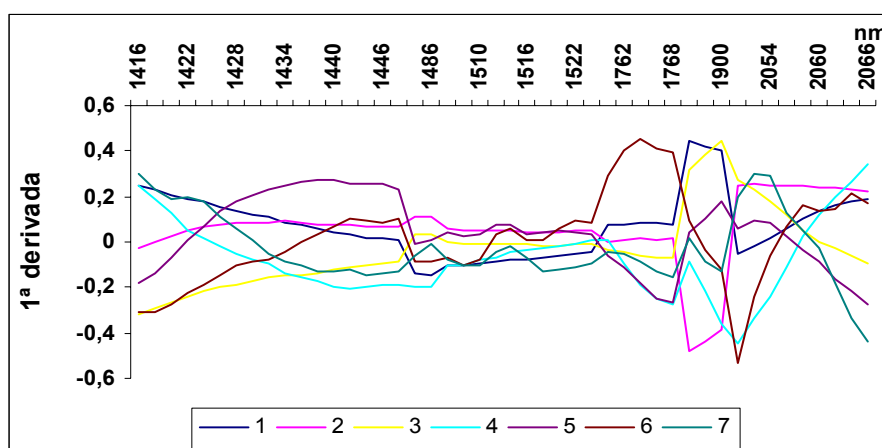


Gráfico 34 : Pesos do modelo PLS para fenilalanina em farelo de soja.

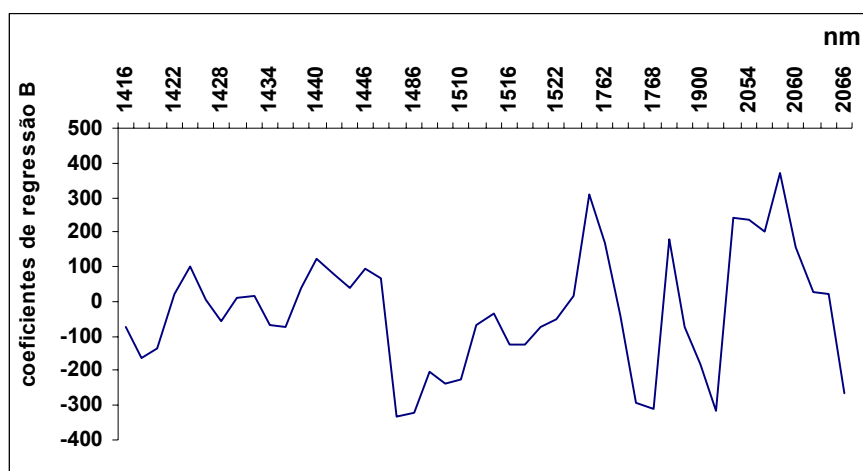


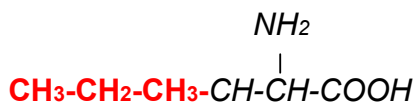
Gráfico 35 : Coeficientes de regressão B do modelo PLS para fenilalanina em farelo de soja.

Tabela 32: Valores e estatísticas do conjunto de validação externa PLS para PHE.

nº	Real (%)	Previsto (%)	CV%	dif (real - prev.)
1	2,410	2,337	2,175	0,073
2	2,350	2,401	1,518	-0,051
3	2,444	2,463	0,548	-0,019
4	2,490	2,562	2,016	-0,072
5	2,418	2,453	1,016	-0,035
6	2,428	2,443	0,435	-0,015
7	2,590	2,622	0,868	-0,032
8	2,430	2,487	1,639	-0,057
9	2,501	2,522	0,591	-0,021
10	2,500	2,495	0,142	0,005
11	2,474	2,477	0,086	-0,003
12	2,458	2,488	0,858	-0,030
13	2,444	2,480	1,034	-0,036
14	2,487	2,478	0,256	0,009
15	2,439	2,455	0,462	-0,016
16	2,510	2,471	1,107	0,039
17	2,427	2,443	0,465	-0,016
18	2,480	2,447	0,947	0,033
19	2,397	2,413	0,470	-0,016
20	2,496	2,521	0,705	-0,025
21	2,423	2,466	1,244	-0,043
22	2,460	2,462	0,057	-0,002
23	2,493	2,482	0,313	0,011
24	2,490	2,511	0,594	-0,021
25	2,510	2,465	1,279	0,045
26	2,475	2,482	0,200	-0,007
27	2,492	2,545	1,488	-0,053
28	2,457	2,505	1,368	-0,048
			MIN-MAX	2,314 à 2,663
SD Método de Referência (%) 0,128			dif. Ideal	0,384 à 0,640
CV Método de Referência (%) 4,29			dif. Real	0,349

4.3.2.1.3 - Aminoácido Isoleucina

Fórmula molecular:



A isoleucina é um aminoácido alifático ou de cadeia aberta. Foram escolhidas 4 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 110 amostras para o conjunto de calibração e 26 amostras para o conjunto de validação externa, sendo necessárias 4 variáveis latentes para descrever o modelo com 34 variáveis selecionadas. Os coeficientes de correlação estão entre 0,70 à 0,74 com erros de previsão iguais ou inferiores à 0,024 (tabela 32). Os pesos (gráfico 36) e os coeficientes B (gráfico 37) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,00 e 2,00%. A faixa de variação da concentração utilizada é 2,5 vezes maior que o desvio padrão da análise (0,12%), como mostra a tabela 33. Existem amostras com 6,4 vezes esse valor, por isso esse modelo precisa ser atualizado com outros valores de isoleucina que variam de 1,7 à 2,4 %, para ampliar sua qualidade de previsão.

Tabela 33: dados do modelo PLS para isoleucina em farelo de soja.

Dado	Valor
Conjunto de calibração	110
Conjunto de validação	26
Número de Variáveis Latentes	4
Erro Padrão de Calibração (%)	0,0249
Coeficiente de Correlação do Conjunto de Calibração	0,7434
Coeficiente de Correlação da Validação Cruzada	0,7344
Erro Padrão de Previsão (%)	0,0320
Slope	0,7980
Bias	0,0020
Coeficiente de Correlação do Conjunto de Validação	0,7080
Número de variáveis	34
1506-1536,nm	PB-NH ₂
1690-1700,nm	CH-CH ₃
1760-1770,nm	CH-CH ₂
1896-1906,nm	C=O/COOH
Teor mínimo (%)	2,0474
Teor máximo (%)	2,3419
Diferença entre os teores (%)	0,2945

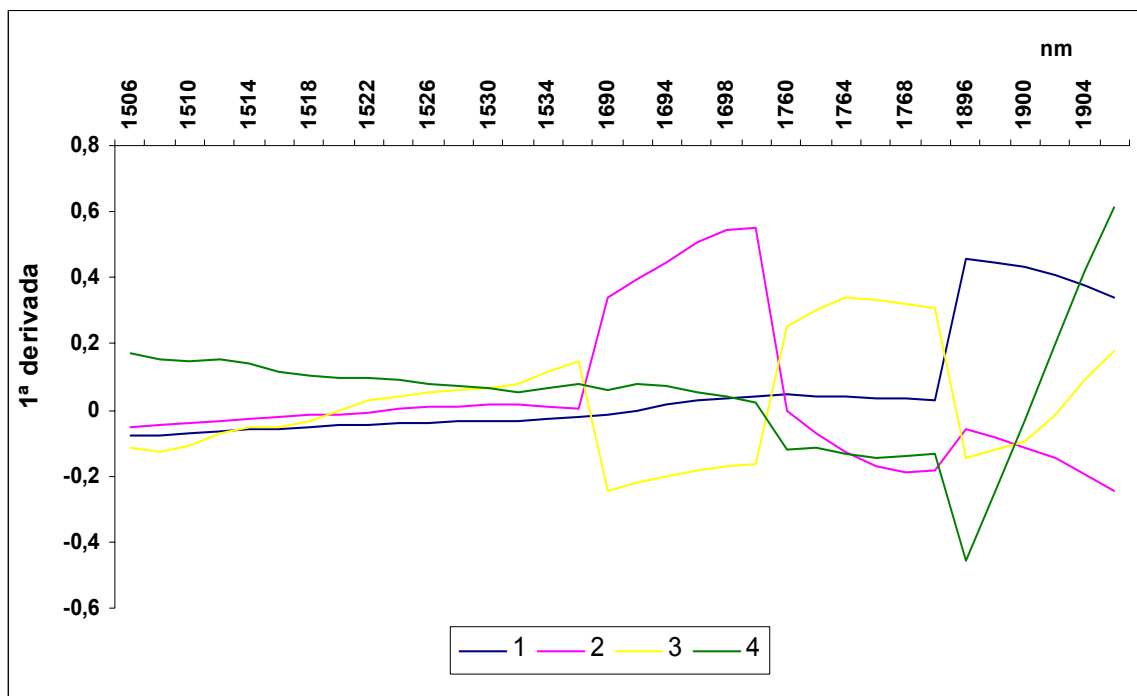


Gráfico 36: Pesos do modelo PLS para isoleucina em farelo de soja.

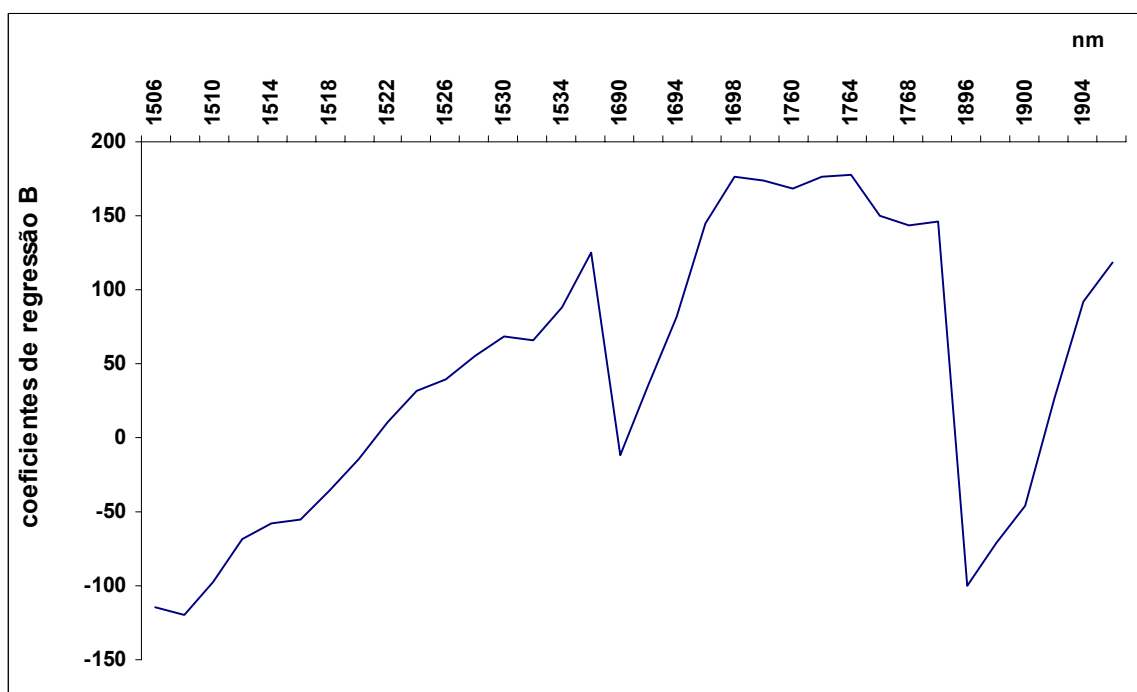


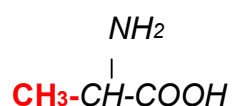
Gráfico 37: Coeficientes de regressão B do modelo PLS para isoleucina em farelo de soja.

Tabela 34: Valores e estatísticas do conjunto de validação externa PLS para ILE.

nº	Real (%)	Previsto (%)	CV%	dif (real-prev.)
1	2,130	2,115	0,50	0,015
2	2,153	2,164	0,36	-0,011
3	2,225	2,247	0,70	-0,022
4	2,165	2,156	0,29	0,009
5	2,200	2,144	1,82	0,056
6	2,171	2,185	0,45	-0,014
7	2,235	2,234	0,03	0,001
8	2,203	2,208	0,16	-0,005
9	2,244	2,208	1,14	0,036
10	2,250	2,203	1,49	0,047
11	2,198	2,206	0,26	-0,008
12	2,153	2,192	1,27	-0,039
13	2,190	2,198	0,26	-0,008
14	2,179	2,179	0,00	0,000
15	2,158	2,184	0,85	-0,026
16	2,212	2,165	1,52	0,047
17	2,200	2,148	1,69	0,052
18	2,280	2,232	1,50	0,048
19	2,134	2,174	1,31	-0,040
20	2,166	2,166	0,00	0,000
21	2,230	2,189	1,31	0,041
22	2,193	2,213	0,64	-0,020
23	2,164	2,198	1,10	-0,034
24	2,180	2,184	0,13	-0,004
25	2,195	2,258	2,00	-0,063
26	2,197	2,207	0,32	-0,010
			MIN-MAX	2,047 à 2,342
SD Método de Referência (%) 0,120			dif. Ideal	0,360 à 0,600
CV Método de Referência (%) 3,350			dif. Real	0,2945

4.3.2.1.4 - Aminoácido Alanina

Fórmula molecular:



A alanina é um aminoácido alifático ou de cadeia aberta. Foram escolhidas 3 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 109 amostras para o conjunto de calibração e 26 amostras para o conjunto de validação

externa, sendo necessárias 5 variáveis latentes para descrever o modelo com 40 variáveis selecionadas. Os coeficientes de correlação estão entre 0,70 à 0,82 com erros de previsão iguais ou inferiores à 0,018 (tabela 34). Os pesos (gráfico 38) e os coeficientes de regressão B (gráfico 39) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,00 e 1,57%. A faixa de concentração utilizada não está muito representativa, sendo 2,0 vezes maior que o desvio padrão da análise (0,134 %). O ideal é aumentar essa faixa que vai de 1,7 à 2,4 % de alanina no farelo de soja, essa quantidade é 5,3 vezes maior que o desvio analítico.

Tabela 35: Dados do modelo PLS para alanina em farelo de soja.

Dado	Valor
Conjunto de calibração	109
Conjunto de validação	26
Número de Variáveis Latentes	5
Erro Padrão de Calibração (%)	0,0181
Coeficiente de Correlação do Conjunto de Calibração	0,8204
Coeficiente de Correlação da Validação Cruzada	0,8110
Erro Padrão de Previsão (%)	0,0191
Slope	1,001
Bias	-0,002
Coeficiente de Correlação do Conjunto de Validação	0,7020
Número de variáveis	40
1390-1444, nm	CH/CH ₂
1690-1700, nm	CH-CH ₃
1896-1906, nm	C=O/COOH
Teor mínimo (%)	1,9818
Teor máximo (%)	2,2464
Diferença entre os teores (%)	0,2646

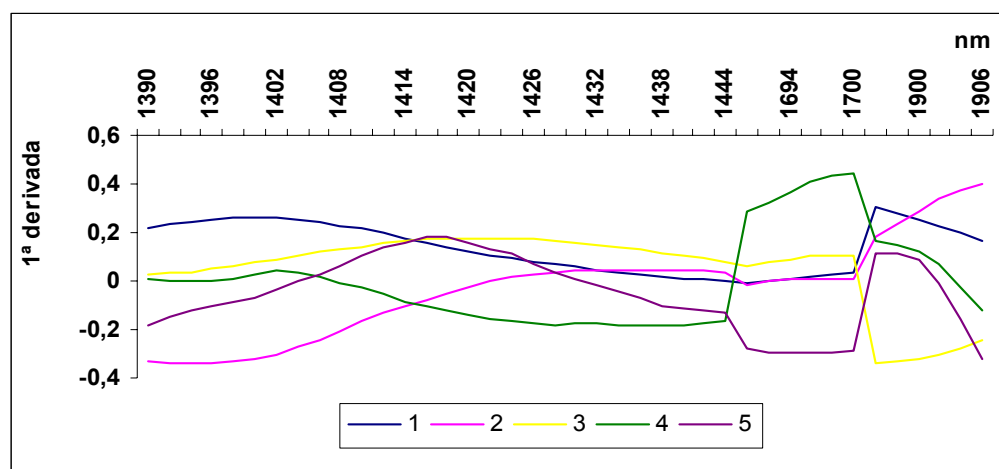


Gráfico 38: Pesos do modelo PLS para alanina no farelo de soja.

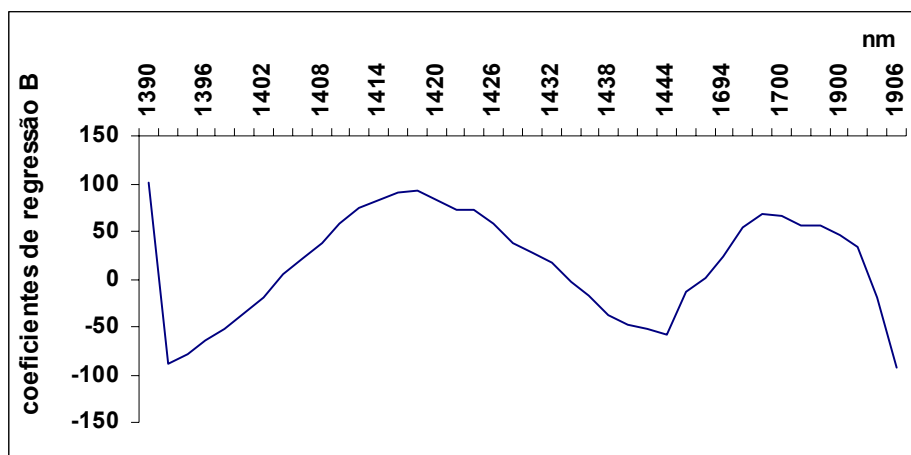


Gráfico 39: Coeficientes de regressão B do modelo PLS para alanina no farelo de soja.

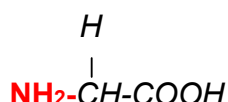
Tabela 36: Valores e estatísticas do conjunto de validação externa PLS para ALA.

nº	Real (%)	Previsto (%)	CV%	dif (real -prev.)
1	2,002	2,047	1,57	-0,045
2	2,223	2,226	0,10	-0,003
3	2,112	2,091	0,71	0,021
4	2,139	2,163	0,79	-0,024
5	2,080	2,081	0,03	-0,001
6	2,055	2,097	1,43	-0,042
7	2,122	2,112	0,33	0,010
8	2,100	2,132	1,07	-0,032
9	2,116	2,134	0,60	-0,018
10	2,159	2,138	0,69	0,021
11	2,112	2,110	0,07	0,002
12	2,068	2,103	1,19	-0,035
13	2,126	2,109	0,57	0,017
14	2,064	2,093	0,99	-0,029
15	2,073	2,077	0,14	-0,004
16	2,110	2,089	0,71	0,021
17	2,028	2,033	0,17	-0,005
18	2,112	2,112	0,00	0,000
19	2,050	2,083	1,13	-0,033
20	2,135	2,088	1,57	0,047
21	2,136	2,094	1,40	0,042
22	2,107	2,110	0,10	-0,003
23	2,079	2,090	0,37	-0,011
24	2,094	2,096	0,07	-0,002
25	2,102	2,127	0,84	-0,025
26	2,111	2,116	0,17	-0,005
			MIN-MAX	1,982 à 2,246
SD Método de Referência (%) 0,134			dif. Ideal	0,402 à 0,670
CV Método de Referência (%) 6,460			dif. Real	0,265

4.3.2.2 – Aminoácidos cujo método de referência precisa ser modificado

4.3.2.2.1- Aminoácido Glicina

Fórmula molecular:



A glicina é um aminoácido alifático ou de cadeia aberta. Foram escolhidas 3 faixas espectrais de absorção para descrever sua estrutura molecular. Utilizou-se 112 amostras para o conjunto de calibração e 27 amostras para o conjunto de validação externa, sendo necessárias 5 variáveis latentes para descrever o modelo com 26 variáveis selecionadas. Os coeficientes de correlação estão entre 0,73 à 0,79 com erros de previsão iguais ou inferiores à 0,020 (tabela 36).

Os pesos (gráfico 40) e os coeficientes de regressão B (gráfico 41) são demonstrados a seguir. Foram obtidos coeficientes de variação entre 0,00 e 1,69%. Contudo, a faixa de concentração utilizada não está muito representativa, sendo 1,8 vezes maior que o desvio padrão da análise (0,154 %), como esse desvio é alto e não existe na natureza a diferença que abranja esses valores (pois a faixa vai de 1,50 à 2,16 % de glicina em amostras de farelo de soja), o ideal é substituir esse modelo por outro com dados produzidos por derivatização pré-coluna, pois o modelo atual é tendencioso e não reflete a margem de variação natural desse aminoácido no farelo de soja. Como os dados não forneceram um bom modelo não será mostrada a validação externa, pois apesar de estatisticamente os valores apresentarem certa concordância, quimicamente não apresentam validade.

Tabela 37: Dados do modelo PLS para glicina em farelo de soja.

Dado	Valor
Conjunto de calibração	112
Conjunto de validação	27
Número de Variáveis Latentes	5
Erro Padrão de Calibração (%)	0,0209
Coeficiente de Correlação do Conjunto de Calibração	0,7984
Coeficiente de Correlação da Validação Cruzada	0,7701
Erro Padrão de Previsão (%)	0,0270
Slope	0,9150
Bias	-0,002
Coeficiente de Correlação do Conjunto de Validação	0,7320
Número de variáveis	26
1496-1516, nm	NH-PB
1896-1906, nm	C=O/COOH
2050-2066, nm	NH/PB
Teor mínimo (%)	1,8690
Máximo (%)	2,1486
Diferença entre os teores (%)	0,2796

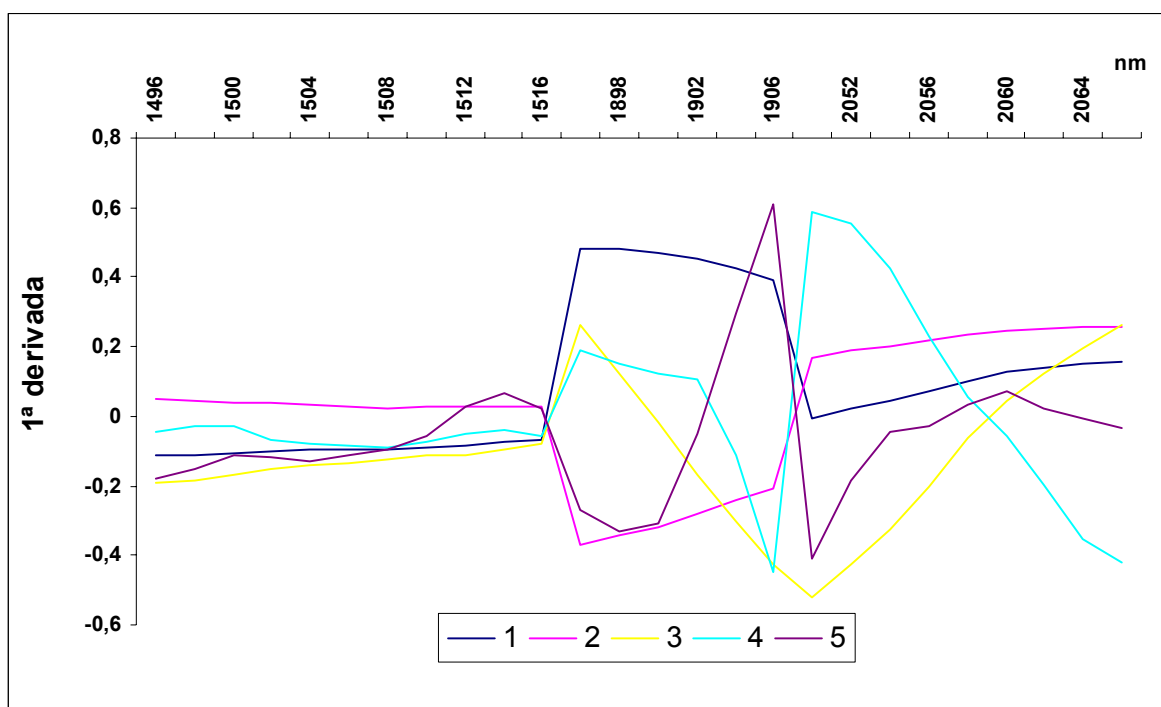


Gráfico 40 : Pesos do modelo PLS para glicina em farelo de soja.

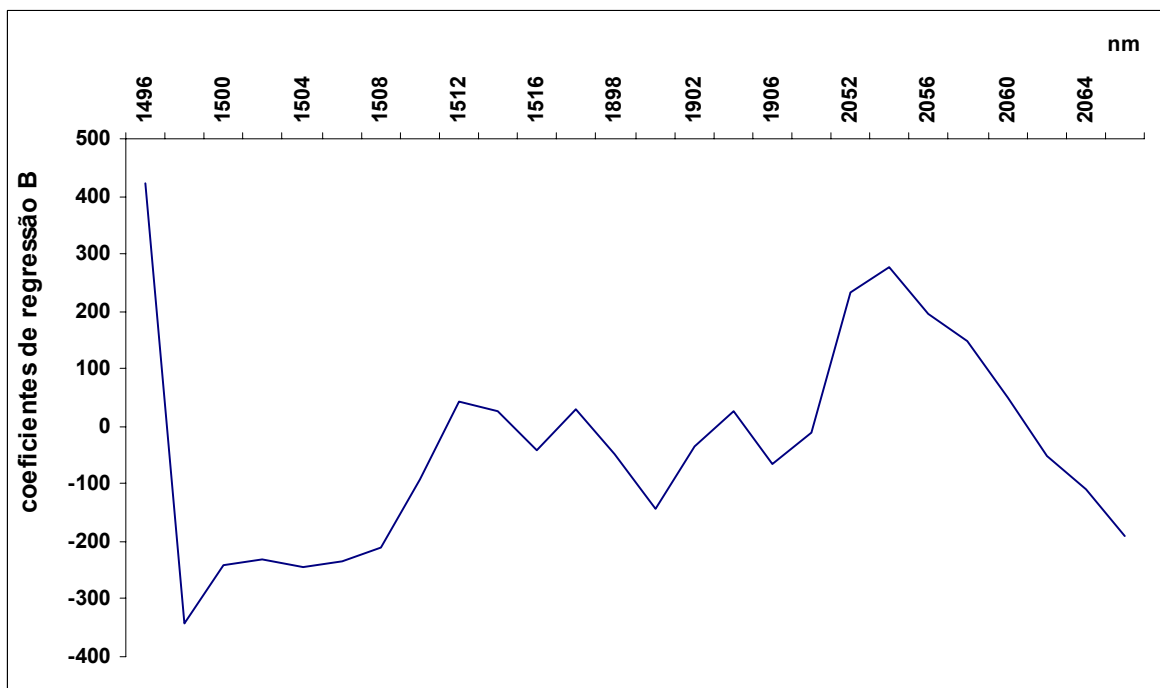
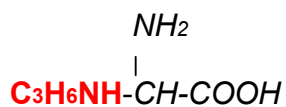


Gráfico 41: Coeficientes B de regressão do modelo PLS para glicina em farelo de soja.

4.3.2.2.2- Aminoácido Prolina

Fórmula molecular:



A prolina é um iminoácido. Foram escolhidas 3 faixas espectrais de absorção para descrever a estrutura molecular. Utilizou-se 108 amostras para o conjunto de calibração e 24 amostras para o conjunto de validação externa, sendo necessárias 3 variáveis latentes para descrever o modelo 27 variáveis selecionadas. Os coeficientes de correlação estão entre 0,80 à 0,87 com erros de previsão iguais ou inferiores à 0,018% (tabela 38).

Os pesos (gráfico 42) e os coeficientes de regressão B (gráfico 43) são mostrados a seguir. Os coeficientes de variação estão entre 0,029 e 1,18%. Uma vez que tanto o coeficiente de variação (9,4%) quanto o desvio padrão (0,201%) são altos esses resultados estatísticos, embora bons com certeza apresentam problemas por serem tendenciosos e não refletem a realidade analítica deste aminoácido. A variação da faixa de concentração das amostras não é satisfatória, pois está apenas 1,51 vezes maior que o desvio padrão do método de referência, o que poderia não só afetar na qualidade das medidas apresentadas por esse modelo, como também é muito restrito perto do universo de 1,9 à 2,7 % de prolina que pode conter um farelo de soja.

Não existe na natureza esta diferença de 5 vezes o desvio padrão da análise. Como a prolina é um aminoácido de comportamento extremamente instável, o ideal seria trabalhar com derivatização pré-coluna que produz resultados mais robustos com menores desvios analíticos. O modelo aqui apresentado é meramente ilustrativo sendo útil para nortear futuros trabalhos para determinar o aminoácido prolina, mas que embora estatisticamente correto deixa a desejar do ponto de vista químico, por claramente estar erros analíticos, assim a validação externa não será apresentada.

Tabela 38: Dados do modelo PLS para prolina em farelo de soja.

Dado	Valor
Conjunto de calibração	108
Conjunto de validação	24
Número de Variáveis Latentes	3
Erro Padrão de Calibração (%)	0,0181
Coeficiente de Correlação do Conjunto de Calibração	0,8788
Coeficiente de Correlação da Validação Cruzada	0,8662
Erro Padrão de Previsão (%)	0,0181
Slope	1,023
Bias	-0,006
Coeficiente de Correlação do Conjunto de Validação	0,8010
Número de variáveis	27
1896-1906,2 nm	C=O / COOH
2050-2080,2 nm	NH ₂ / PB
2296-2304,2 nm	NH ₂ / PB
Teor mínimo (%)	2,3002
Teor máximo (%)	2,6046
Diferença entre os teores (%)	0,3044

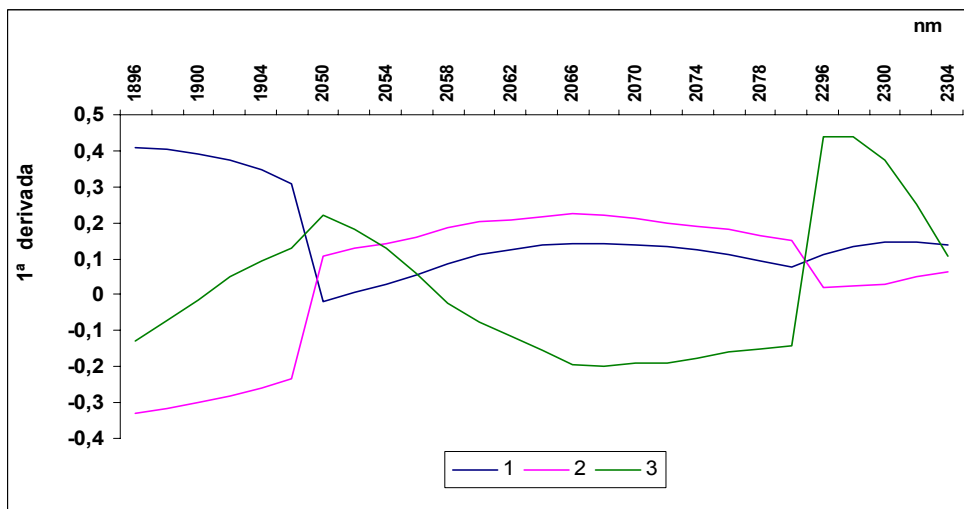


Gráfico 42: Pesos do modelo PLS para prolina em farelo de soja.

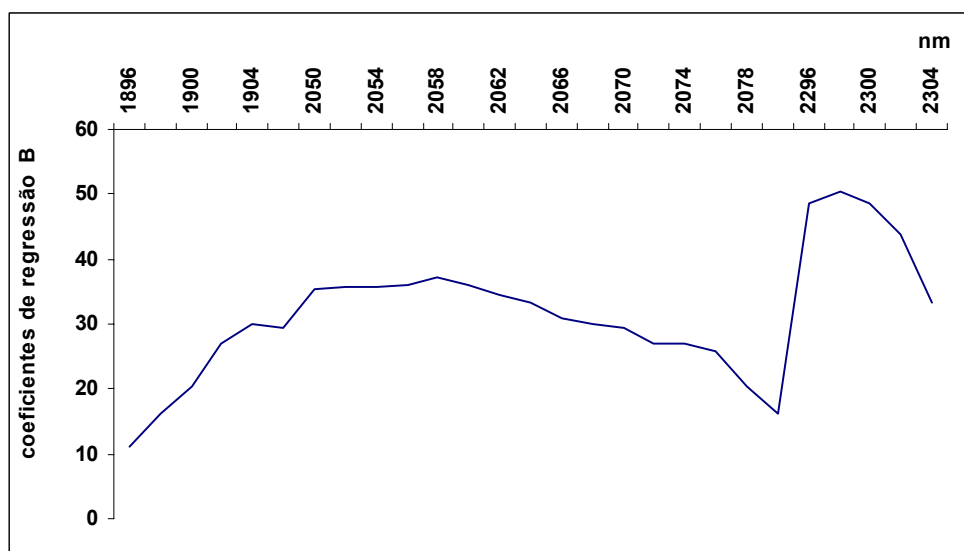
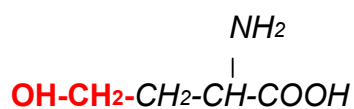


Gráfico 43: Coeficientes de regressão B do modelo PLS para a prolina em farelo de soja.

3.3.2.2.3 - Aminoácido Serina

Fórmula molecular:



A serina é um aminoácido hidroxilado. Foram escolhidas 3 faixas espectrais de absorção para descrever a estrutura molecular. Utilizou-se 111 amostras para o conjunto de calibração e 23 amostras para o conjunto de validação externa, sendo necessárias 7 variáveis latentes para descrever o modelo com 60 variáveis selecionadas. Os coeficientes de correlação estão entre 0,73 à 0,80 com erros de previsão iguais ou inferiores à 0,010 (tabela 39).

Os pesos (gráfico 44) e os coeficientes de regressão B (gráfico 45) são demonstrados a seguir. Os resultados podem ser considerados tendenciosos uma vez que é alto o desvio padrão (0,208%) e o coeficiente de variação (8,52%) do método de referência. A faixa de concentração apresentada está 1,3 vezes maior que o desvio padrão da análise, sendo que a diferença necessária (1,04%) ultrapassa a faixa possível de ser encontrada (0,90%) nos farelos de soja do mercado. Assim, esse modelo também não apresenta confiabilidade analítica e deve ser refeito utilizando dados gerados por derivatização pré-coluna que apresenta desvios analíticos mais baixos, o que leva a não ser apresentados os dados de validação externa.

Tabela 39: Dados do modelo PLS para serina em farelo de soja.

Dado	Valor
Conjunto de calibração	111
Conjunto de validação	23
Número de Variáveis Latentes	7
Erro Padrão de Calibração (%)	0,0108
Coeficiente de Correlação do Conjunto de Calibração	0,8020
Coeficiente de Correlação da Validação Cruzada	0,7838
Erro Padrão de Previsão (%)	0,0220
Slope	0,8650
Bias	-0,005
Coeficiente de Correlação do Conjunto de Validação	0,7370
Número de variáveis	60
1390-1444,2 nm	CH/CH ₂
1896-1096,2 nm	C=O/ COOH
2050-2100,2 nm	NH ₂ /PB
Teor mínimo (%)	2,5498
Teor máximo (%)	2,2741
Diferença entre os teores (%)	0,2754

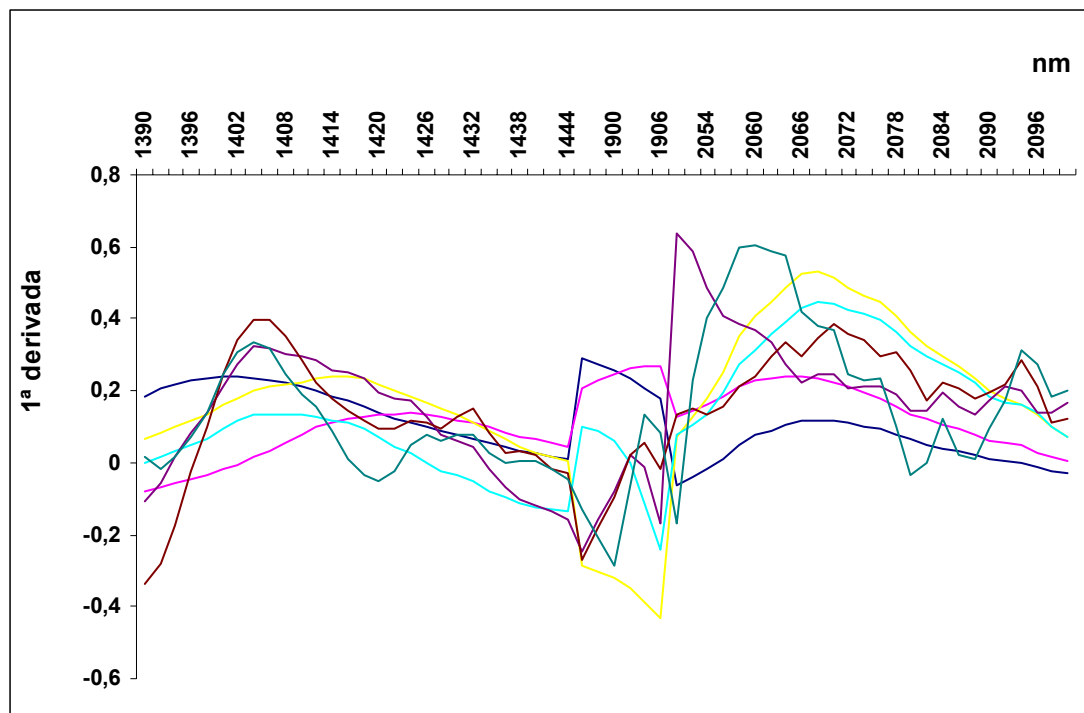


Gráfico 44: Pesos do modelo PLS para serina em farelo de soja.

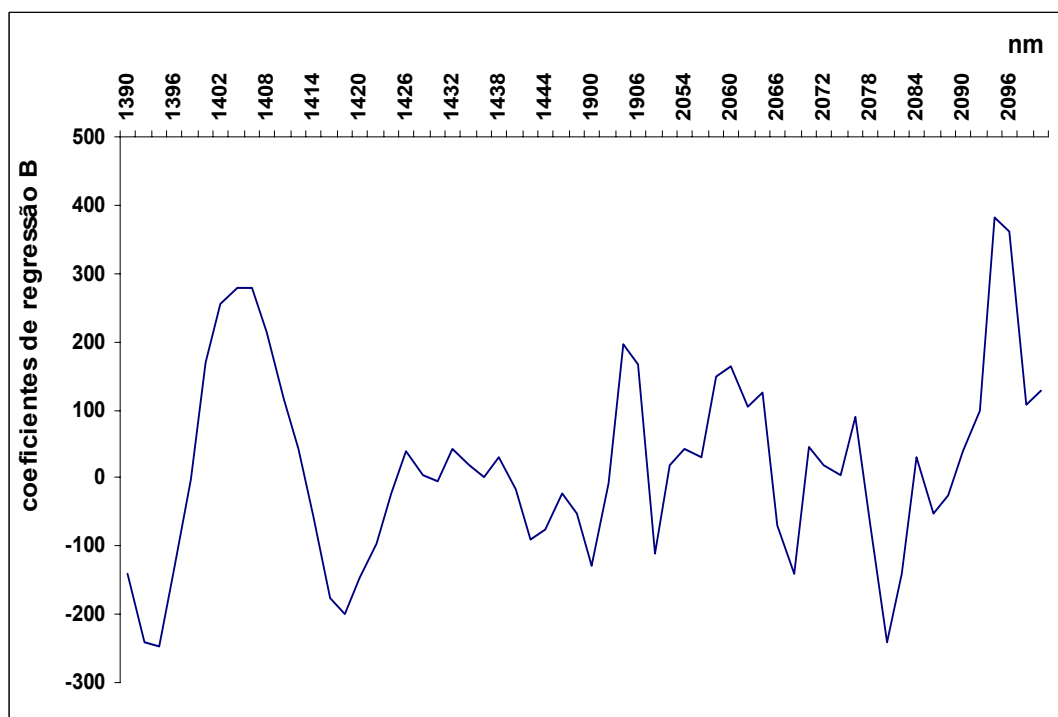
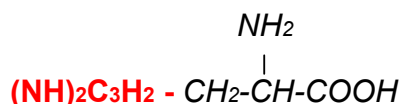


Gráfico 45: Coeficientes de regressão B do modelo PLS para serina em farelo de soja.

3.3.2.2.4 - Aminoácido Histidina

Fórmula molecular:



A histidina é um aminoácido básico. Foram escolhidas 4 faixas espectrais de absorção para descrever a estrutura molecular. Utilizou-se 113 amostras para o conjunto de calibração e 27 amostras para o conjunto de validação externa, sendo necessárias 3 variáveis latentes para descrever o modelo com 28 variáveis selecionadas. Os coeficientes de correlação estão entre 0,71 à 0,74 com erros de previsão iguais ou inferiores à 0,012 (tabela 40).

Os pesos (gráfico 46) e os coeficientes de regressão B (gráfico 47) são mostrados a seguir. Os coeficientes de variação estão entre 0,12 e 1,81%, mas esse resultados embora bons com certeza apresentam problemas por serem tendenciosos uma vez que tanto o coeficiente de variação (8%) quanto o desvio padrão (0,43%) são altos e os resultados não refletem a realidade analítica deste aminoácido. A faixa de variação das amostras não é satisfatória, pois está apenas 0,34 vezes maior que o desvio padrão do método de referência, o que compromete totalmente o modelo já que o farelo de soja pode conter de 0,860 à 1,350% em histidina.

Como a histidina é um aminoácido de cadeia fechada básica, precisa de muitos cuidados no processo de hidrólise. O ideal seria trabalhar com derivatização pré-coluna que produz resultados mais robustos com menores desvios analíticos. O modelo aqui apresentado é meramente ilustrativo sendo útil para nortear futuros trabalhos para determinar o aminoácido histidina, mas que embora estatisticamente correto deixa a desejar do ponto de vista da lógica analítica, por claramente está modelando erros analíticos, assim também não serão apresentados os dados de validação externa.

Tabela 40: Dados do modelo PLS para histidina em farelo de soja.

Dado	Valor
Conjunto de calibração	113
Conjunto de validação	27
Número de Variáveis Latentes	3
Erro Padrão de Calibração (%)	0,0125
Coeficiente de Correlação do Conjunto de Calibração	0,7430
Coeficiente de Correlação da Validação Cruzada	0,7334
Erro Padrão de Previsão (%)	0,0150
Slope	0,8010
Bias	0,004
Coeficiente de Correlação do Conjunto de Validação	0,7130
Número de variáveis	28
1496-1516,nm	NH-PB
1616-1626,nm	CH=CH ₂
1896-1906,nm	COOH
2166-2174,nm	HC=CH
Teor mínimo (%)	1,0621
Teor máximo (%)	1,2096
Diferença entre os teores (%)	0,1475

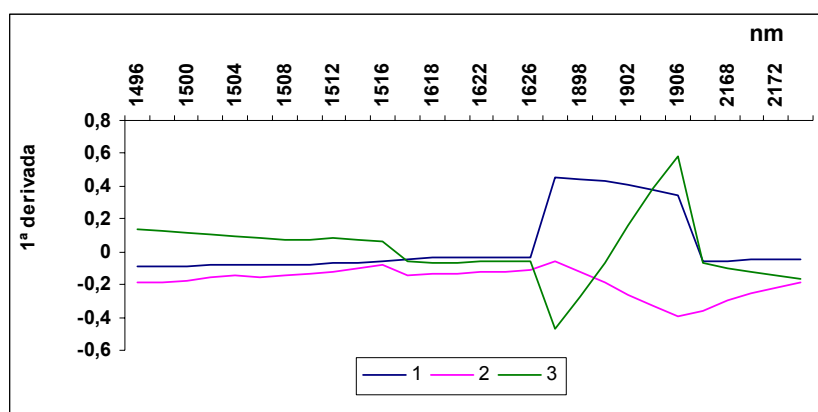


Gráfico 46: Pesos do modelo PLS para histidina em farelo de soja.

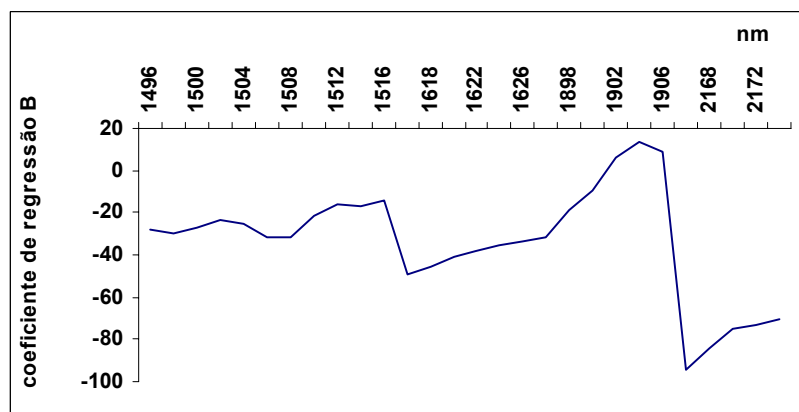


Gráfico 47: coeficientes de regressão B do modelo PLS para histidina.

4.4- Vantagens do uso de modelos desenvolvido via espectroscopia no infravermelho próximo sobre as tabelas nutricionais de aminoácidos

O uso de tabelas com o teor aproximado de aminoácidos em uma determinada matéria-prima é prática comum no meio de nutrição animal ^(52,53). O grande problema é que o valor de proteína bruta é dado pelo nitrogênio total da amostra que nem sempre é proveniente de fontes de origem protéica, a fraude mais comum no mercado é adicionar uréia para elevar o teor de nitrogênio da amostra e transformar farelos pobres em proteína e aminoácidos em altas fontes protéicas. Outro problema muito comum é o superaquecimento do farelo de soja que leva a elevação do teor de proteína e consequentemente de aminoácidos, mas que por mexer na estrutura dos mesmos, torna seu valor nutricional totalmente comprometido ^(54,55,56,57).

Existem análises para verificação de possíveis contaminações com uréia e desnaturação protéica por aquecimento, que são o índice de nitrogênio não protéico (NNP) e a solubilidade protéica em KOH 0,2 % ⁽⁵⁸⁾, mas são métodos demorados que costumam ser solicitados geralmente por nutricionistas mais experientes que têm conhecimento do aspecto visual de um bom farelo de soja e que podem desconfiar de uma possível fraude ou mal processamento do produto.

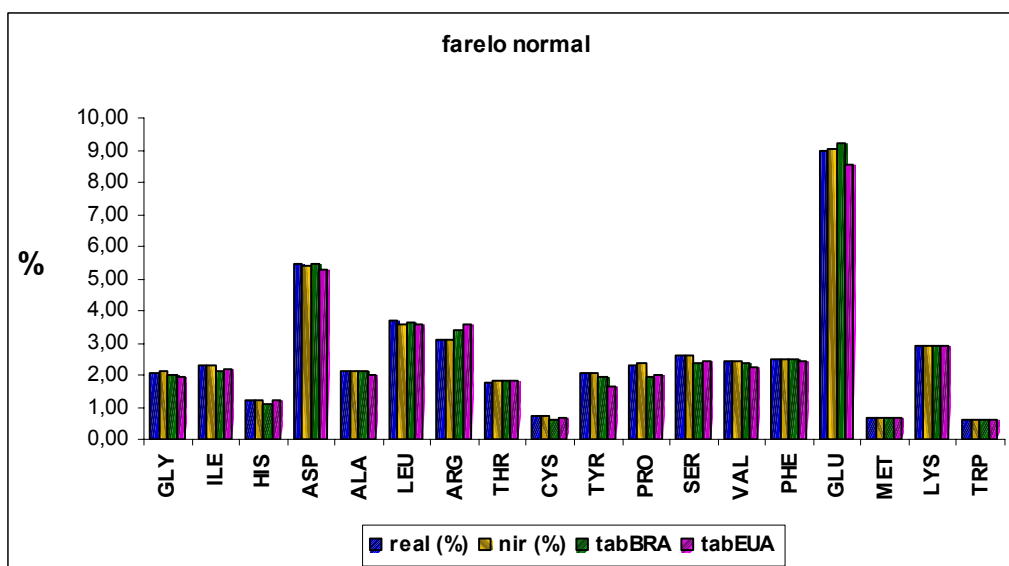


Gráfico 48 : Comparativo do teor de aminoácidos para um farelo de soja normal.

As tabelas existentes no mercado apresentam uma grande variação de uma para a outra e não são constantemente atualizadas, o que pode gerar erro e desperdício econômico em uma formulação. O gráfico 48, mostra um comparativo feito entre resultados via infravermelho próximo, método de referência para aminoácidos, valores de uma tabela brasileira ⁽⁵²⁾ e valores de uma tabela americana ⁽⁵³⁾ para um farelo de soja normal. É possível notar uma variação analítica considerável para aminoácidos de fácil determinação como ácido glutâmico, ácido aspártico e arginina. Obviamente, essa tendência é flutuante dependendo da tabela que é utilizada, e o objetivo é apenas chamar a atenção para essa característica que existe e mostrar que quando bem definido o método baseado no infravermelho próximo apresenta boa repetibilidade em relação ao método convencional.

Essa mesma amostra foi contaminada com uréia e seus resultados são mostrados no gráfico 49.

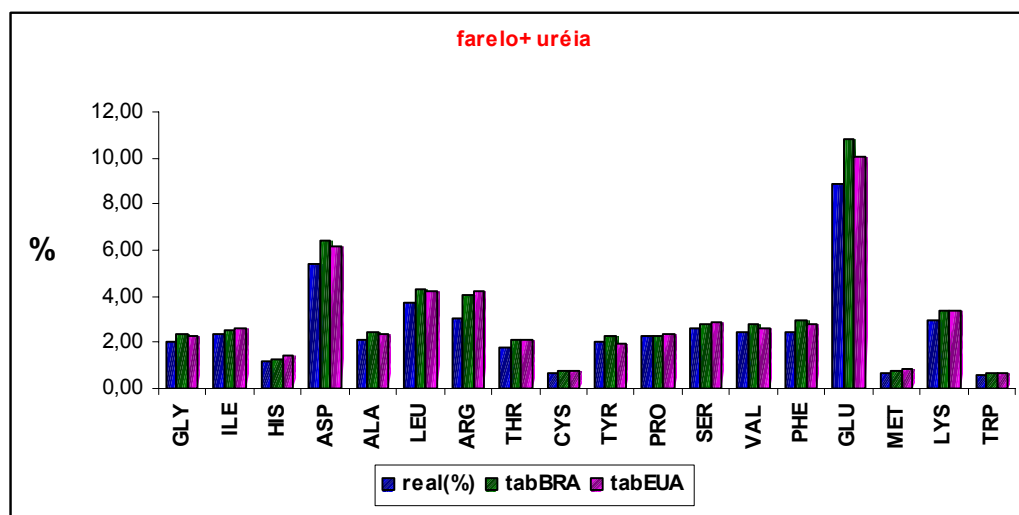


Gráfico 49: Comparativo do teor de aminoácidos para um farelo de soja contaminado com uréia

Analisando os valores é possível notar que existe alteração dos teores de todos os aminoácidos e pelo menos 70% dos mesmos são bem significativos. Não é possível fazer a comparação via espectroscopia no infravermelho próximo, pois o modelo quimiométrico desenvolvido consegue mostrar a diferença espectral devido à

contaminação da matriz com uréia e define a amostra como incompatível ao modelo de calibração.

A terceira amostra sofreu um processo de super aquecimento e os valores são demonstrados no gráfico 50.

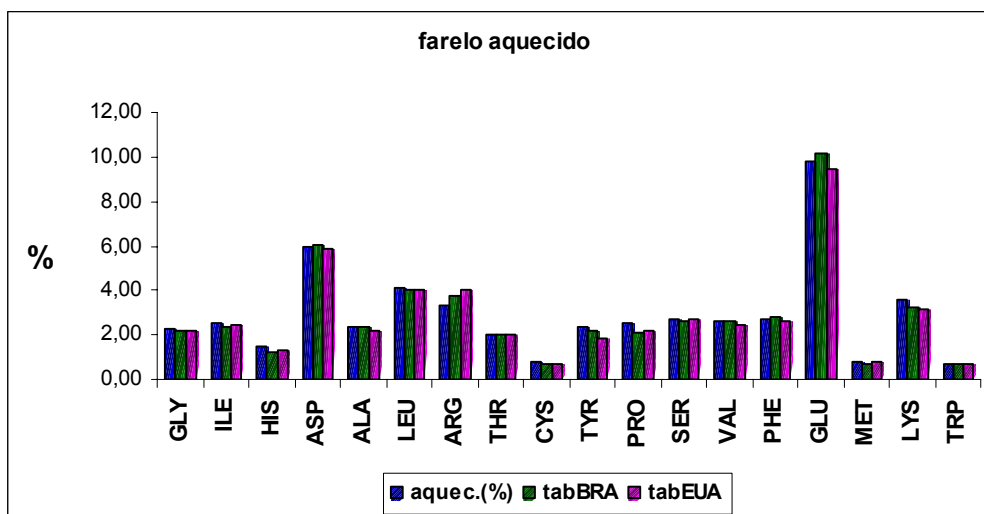


Gráfico 50: Comparativo do teor de aminoácidos para um farelo de soja superaquecido.

O comportamento é semelhante à amostra contaminada com uréia, contudo a variação é menor, pois nem o método de referência consegue identificar uma amostra super aquecida. Contudo, a análise de proteína solúvel em KOH 0,2% ou o espectro da amostra via infravermelho próximo pode facilmente apontar os problemas com a qualidade dos aminoácidos da amostra, pois o espectro obtido tem uma característica totalmente diferente de um farelo normal, como é possível notar no gráfico 51. A amostra contaminada com uréia tem alteração espectral mas não é tão marcante quanto um farelo aquecido, assim o sistema de infravermelho próximo, quando calibrado com amostras de boa procedência, respeitando toda a faixa de representatividade amostral para cada aminoácido e desde que se trabalhe com um método de referência com desvios e coeficientes de variação baixos, como atualmente é possível no método de referência pela derivatização pré-coluna pode fornecer valores extremamente precisos, rápidos e a um custo bem mais baixo que um aminograma normal.

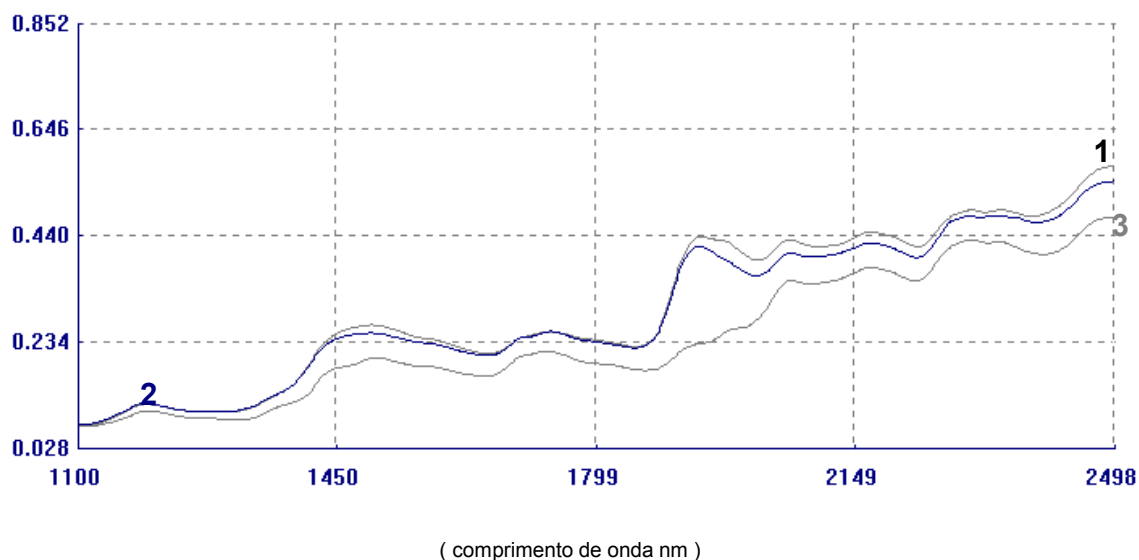


Gráfico 51: Espectros dos 3 farelos de soja utilizados nos comparativos. (a amostra 1 e 2 correspondem a amostra normal e contaminada com uréia respectivamente. A amostra 3 corresponde ao farelo super aquecido.

A nutrição feita com maior precisão numa fábrica de ração representa uma economia de custos e a melhora na performance animal. Apesar das análises via infravermelho próximo apresentarem um desvio de 3 à 5 % para a previsão de aminoácidos, as tabelas desatualizadas podem representar um erro de até 30%. Como cada aminoácido tem uma faixa de concentração e os valores não podem ser de modo algum fixados como sugerem algumas tabelas, isso pode significar a diminuição de erro em uma formulação de até 50%. Portanto, o sistema NIRS pode ser utilizado para a própria correção das tabelas, melhorando assim o banco de dados dos nutricionistas a um custo 6 vezes menor que o método HPLC.

CONCLUSÕES

Os resultados obtidos mostraram-se satisfatórios para onze modelos que atendem perfeitamente todas as exigências do ponto de vista químico e estatístico das metodologias secundárias de calibração por espectroscopia no infravermelho próximo. Segundo ASTM E1655-00 ⁽⁵⁰⁾, a faixa utilizada na construção dos modelos de calibração deve ser no mínimo de 3 à 5 vezes maior do que o desvio padrão do método de referência, tal condição foi conseguida para proteína bruta, ácido aspártico, ácido glutâmico, arginina, valina, treonina, leucina, tirosina, triptofano, cistina e metionina. Quatro modelos podem ser utilizados com cautela, mas precisam ser ampliados para aumentar a faixa de previsão e diminuir uma possível margem de incerteza estatística gerada pelos desvios padrões do método de referência pós-coluna (lisina, alanina, isoleucina e fenilalanina) e apenas quatro são considerados inviáveis por apresentarem alto erro analítico que foi modelado e apresenta resultados considerados bons estatisticamente, contudo tendenciosos, que não refletem a realidade do farelo de soja comercial (histidina, serina, glicina e prolina). Esses modelos teriam que ser refeitos utilizando outro método de referência como a derivatização pré-coluna ⁽⁵¹⁾ que produz baixos desvios analíticos, sendo possível obter dados melhores, possíveis de serem calibrados.

Rotineiramente, as indústrias utilizam todo o espectro na construção de seus modelos de calibração, o que acaba gerando os mesmos coeficientes de regressão B e os pesos para todos os modelos de aminoácidos e proteína bruta, ou seja, todos os modelos são iguais e a única grandeza comparada é o valor de proteína bruta que tem relação com os aminoácidos e são previstos por associação. A técnica é bem aceita pela indústria, é funcional, mas deixa a desejar do ponto de vista científico (embora ainda seja melhor do que valores de tabela por estar relacionado a medida primária com o espectro da amostra). Essa questão foi levantada no final deste trabalho e notou-se que é possível determinar cada aminoácido separadamente pois existem bandas de absorção características para cada tipo de grupo funcional e ligações das moléculas de aminoácidos como é o caso dos aminoácidos sulfurados que constituem apenas 1% do teor de proteína do farelo de soja mas que possui 70% de informação na banda de absorção de 1736-1746 nm que compreendem as ligações do tipo S-H e S-S.

Aqui foram apenas sugeridos algumas regiões espectrais definidas com o auxílio de informações teóricas e de um recurso do equipamento da FOSS⁽⁴³⁾ que aponta o tipo de ligação encontrada ao longo do espectro de uma amostra.

Os modelos desenvolvidos via NIR para aminoácidos em farelo de soja podem ser utilizados pelos laboratórios de controle de qualidade, fornecendo resultados mais precisos que os das tabelas nutricionais que estão à disposição dos zootecnistas que fazem formulações para a área de nutrição animal. Sabe-se que o recurso das tabelas é utilizado devido ao alto custo da análise de aminoácidos e o tempo de espera para a obtenção de um aminograma completo em uma matriz como o farelo de soja. Entretanto, estas tabelas são genéricas e não possibilitam o conhecimento completo da matéria-prima que pode diminuir as incertezas de uma formulação.

Assim, a vantagem de ter análises que relacionam a medida espectral com a medida de análises clássicas, desde que se respeite o intervalo de valores utilizados na construção dos modelos, é poder reduzir o custo e o tempo de análise (de 48 horas para 2 minutos), custo de reagentes, além de não gerar resíduos que é uma contribuição à conservação do meio ambiente. Conhecer o valor total de proteína bruta é uma informação importante, porém como a proteína nada mais é do que a reunião de vários aminoácidos com funções diferentes e que atendem a exigências de metabolismo diferentes dependendo da espécie animal que se pretende alimentar, cada dia mais se faz necessário obter resultados rápidos, confiáveis e baratos do perfil de aminoácidos para melhorar a saúde e o desenvolvimento animal e consequentemente aumentar os lucros do ramo de negócios agropecuários.

A escolha da matriz farelo de soja foi feita devido à grande importância da soja e seus subprodutos na economia brasileira. Portanto, o desenvolvimento de métodos analíticos que possam melhorar a utilização do “complexo soja” faz das pesquisas científicas uma importante ferramenta para as indústrias de nutrição e para a economia do Brasil.

PERSPECTIVAS FUTURAS

Este é um trabalho que pode ser ainda mais detalhado num outro momento, através de estudos que possibilitem isolar melhor com evidências científicas, as melhores bandas de absorção para cada aminoácido. O estudo pode ser feito utilizando seleção de variáveis para continuar trabalhando com infravermelho próximo ou com MID (infravermelho médio), que é mais utilizado para caracterizar substâncias.

Outra questão é a sensibilidade do método de referência utilizado para obtenção dos resultados analíticos para aminoácidos, sabe-se que a melhor opção hoje é a análise de aminoácidos com derivatização pré-coluna ⁽⁵¹⁾, que produz cromatogramas com maior eficiência na separação e coeficientes de variação menores que a técnica pós-coluna (derivatização com ninidrina), utilizada no início deste estudo e aceita pela Comunidade Européia. Como são análises de custo extremamente alto, o ideal seria iniciar um novo estudo já com o método que produz um menor desvio padrão entre os resultados analíticos.

É importante lembrar que a quimiometria pode auxiliar muito na interpretação de dados químicos, quanto maior for a confiabilidade da medida melhor será a reprodução da mesma. Tabelas e cálculos de conversão são arriscados quando não se conhece o material com que se está trabalhando. A Espectroscopia no Infravermelho Próximo é sem dúvida uma das melhores técnica para análise de componentes orgânicos por modelos matemáticos utilizadas atualmente.

A técnica de construção do modelo quer PLS, PCR ou qualquer outra escolhida pode ser adequada de acordo com o resultado que se pretende observar, mas o principal é conhecer a região onde está a informação que se pretende modelar e relacionar com as variáveis medidas, pois o espectro é o todo, mas cada constituinte químico é uma parte que deve ser isolada e estudada. Assim, validada essa técnica é possível partir para o estudo de aminoácidos digestíveis que são extremamente informativos nutricionalmente.

REFERÊNCIAS BIBLIOGRÁFICAS

1. **ASSOCIAÇÃO NACIONAL DOS FABRICANTES DE ALIMENTOS PARA ANIMAIS : ANFAL, 2005** . Guia de métodos analíticos, proteína bruta. In: Compêndio de nutrição animal. São Paulo : Sindirações., 2ª ed., p. 19-25.
2. **AMERICAN ASSOCIATION OF CEREAL CHEMISTS – AACC,1994**. Official methods of Analysis. Crude protein - combustion method. First approval 9.23.92, reviewed 10.26.94. Test nº 46-30.
3. **ANDRIGUETTO, J.M., PERLY, L., MINARDI, I., FLEMMING, J.S., GEMAEL, A., SOUZA, G.A., FILHO, A.B., 1986**. Considerações sobre o uso dos diferentes alimentos nas rações de aves e suínos. In: Nutrição animal aplicada. V.2 . São Paulo: Nobel , 4ª ed. , p. 15-79.
4. **PENZ JR., A.M., BRUGALLI,I., 2001**. Soja e seus derivados na alimentação de aves. In: Simpósio sobre Ingredientes na Alimentação Animal, Campinas, SP. Anais... Campinas: CBNA, p.85-108.
5. **GOLDFLUS, F., 2001**. Ingredientes derivados do processamento da soja aplicados na nutrição animal. In: simpósio sobre o manejo e nutrição de aves e suínos e tecnologia de produção de rações, Campinas, SP. Anais,...Campinas: CBNA, p. 97-188.
6. **ARAUJO, M.F., GOMES, J.C., SANT'ANNA, R., SILVA, D.J., 1986**. Caracterização química de um isolado protéico de soja produzido no Brasil. Rev. Ceres, v. 33, nº 188, p. 320-329.
7. **CLARKE, J.H., WISEMAN, J. , 1999**. Nutritional value of soya products for non ruminant and farm animals. Singapura: American Soybean Association - ASA, nº 060/10/98, p.183.

8. **ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEO VEGETAL, 2005.** Estatísticas do complexo soja no Brasil. Disponível na word wide web : < [http: \\ www.abiove.com.br](http://www.abiove.com.br) > . [acesso em 20 março de 2005].São Paulo: ABIOVE.
9. **SAKOMURA, N.K., 1996.** Estudo do valor nutricional das sojas integrais processadas e de sua utilização na alimentação de frangos e poedeiras. Jaboticabal: Tese para obtenção do Título de Livre Docente – UNESP.
10. **PARSONS, C.M., HASHIMOTO, K., WEDEKIND, K.J., BAKER, D.H., 1992.** Effect of overprocessing on availability of amino acids and energy in soybean meal. Poultry Sci., v. 71, p. 133-140.
11. **CHEFTEL, J.C., LORIENT, D., 1990.** Las proteínas de soja. In: Proteínas alimentares, Bioquímica, propriedades Funcionales e Valor Nutricional. Zaragoza: Acribia, p. 257-275.
12. **SGARBIERI, V.C., 1996.** Fontes de proteínas na alimentação. In: Proteínas em alimentos protéicos – propriedades, degradações e modificações. São Paulo: Livraria Varela, p.227-230.
13. **SIMÕES, M. G., 2003.** Avaliação da qualidade protéica pelo escore dos aminoácidos corrigidos pela digestibilidade da proteína calculada com o conteúdo fecal e ileal. Campinas: tese de mestrado – UNICAMP.
14. **CAMPOS, S.D.S., BALDINI, V.L.S., SREBERNICH, S.M., 1983.** O papel dos aminoácidos peptídios e proteínas no sabor dos alimentos. Boletim ITAL, Campinas, v.20, nº 1, p.1-18.

15. **CAFÉ , M.B., 1993.** Determinação da disponibilidade biológica dos aminoácidos e da energia metabolizável da soja integral processada para aves. In: Conferência Apinco de Ciência e Tecnologia Avícolas, Santos, SP. Anais...Campinas: Facta , p.13.
16. **LEE, H., GARLICH, J.D., 1992.** Effect of overcooked meal on chicken performace and amino acid availability. Poultry Sci., v. 71, p.499-508.
17. **PARSONS, C.M., HASHIMOTO, K., WEDEKIND, K. J., BAKER, D.H., 1991.** Soybean Protein solubility in KOH: an in vitro test of in vivo protein quality. J. Anim. Sci. , nº 69, p. 2918-2924.
18. **MARTINS, J.B., 1995.** Efeito do tratamento térmico sobre a qualidade nutricional do grão de soja no desempenho e na composição da carcaça de frangos de corte. Porto Alegre: Tese de mestrado em zootecnia – UFRGS.
19. **LIMA, G. J.M.M., 1999.** Importância da qualidade nutricional da soja e seus subprodutos no mercado de rações: situação atual e perspectivas futuras. In: Congresso Brasileiro de Soja. Londrina, PR. Anais... Londrina: Emprapa Soja. p.165-175.
20. **ALMEIDA, M.E.W.A., SIMÃO, A. M., 1972.** Aminoácidos de proteína de soja e dosagem de ácido glutâmico. Rev. Inst. Adolfo Lutz, nº 32, p. 105-112.
21. **ACASIO, U.A., 1997.** Handling and store of soybeans and soybean meal. Technical Bulletin. Singapura: American Soybean Assotiation - ASA, nº 044/11/96. p. 16.
22. **BAKER, D.H., 2000.** Nutritional Constraints to use of soy products by animals . In: Soy in Animal Nutricion. Savoy: Federation of Animal Science Societies, p.1-12.

23. **SWICK R.A., 1994.** Soybean meal quality . Technical Bulletin. Singapura: American Soybean Association - ASA , nº 071/12/93, p.13.
24. **KIRBY, S.R., PESTI, G.M., DORFMAN,J.H., 1993.** An investigation of the distribution of the protein content of samples of corn, meat and bone meal, and soybean meal. Poultry Sci., v. 72, p. 2294-2298.
25. **FOOD AGRICULTURE ORGANIZATION, 1976.** Amino acids content of foods. Roma: FAO, nº 24, p.190.
26. **WILLIAMS, P., NORRIS, K., 1990.** Near Infrared Technology in the Agricultural and Food Industries. Minnesota: AACC.
27. **BURNS, D.A., CIURCZACK, E.W., 2001.** Handbook of Near Infrared Analysis. New York: Marcel Dekker, 2nd.
28. **OTTO, M. ,1999.** Chemometrics: statistics and computer application in analytical chemistry. Chichester: Wiley-VCH.
29. **WOLD, S., MARTENS,H., 1983.** The multivariate calibration: problem in Chemistry solved by the PLS method. Lect. Notes Math. , p.286.
30. **HOWARD, A., RORRES, C., 2001.** Álgebra linear com aplicações. Porto alegre: Bookman , 8ª ed.
31. **MASSART, B. G.M., VANDEGISTE, S., DEMING, S.N., 1988.** Chemometrics: a text book. Amsterdã: Elsevier.
32. **MILLER, J.N., MILLER,J.C., 2000.** Statistics and Chemometrics for Analytical Chemistry. Chichester: Prentice Hall, 4nd.

33. **VANDEGNISTE, B.G.M., MASSAT, D.L., BUYDENS, L.M.C., JONG, S., LEWI, P.J., VERBEQUE, J.S., 1998.** Handbook of chemometrics and qualimetrics: part B. Amsterdã: Elsevier.
34. **DRAPER, N., SMITH, H., 1981.** Applied Regression Analysis. New York: Willey .
35. **CUSTÓDIO, R., ANDRADE, J.C., AUGUSTO, F., 1997.** O ajuste de funções matemáticas à dados experimentais, quím. nova, v.20, P.219-405.
36. **BRERETON, R.G., 2000.** Introducion to multiderivate calibration in analytical chemistry. Analyst, p. 2125.
37. **SJOSTROM, M., 1983.** A multiderivate calibration problem in analytical chemistry solved by partial least squares models in latent variables. Anal. Chim. Acta 150, p.61-70.
38. **GELADI, P., KOWALSKI, B.R., 1986.** Partial Least Square Regression: a tutorial, Anal. Chim. Acta 185, p.1-17.
39. **SHENK, J. S., WESTERHAUS, M.O., 1993.** Near Infrared Reflectance Analysis with single and multiproduct calibration. Crop Sci. , v. 33, p. 582-584.
40. **WOLD, S., 1978.** Cross validatory estimation of the number of componentes in factor and principal componentes models. Technometrics, v. 20, nº 4, p.397-405.
41. **SHENK, J. S., WESTERHAUS, M.O., 1991.** New standartization and calibration procedure for NIRS Analytical Systems. Crop Sci. , v. 31, p. 1694-1696.
42. **SHENK, J.S., WESTERHAUS, M.O., 1995.** Monograph analysis of agriculture and food products by Near Infrared Reflectance Spectroscopy. Port Maltilda: ISI.

43. **SHENK, J. S., WESTERHAUS, M.O., 1998.** Calibration System for Spectrographic Analyzing Instruments . UNITED STATE PATENT : 5,798,526. Appl.78259, filed jan. 24,1997. Assignee: Infracore International LLC. Disponível na word wide web <<http://www.patft.uspto.gov/netacgi/nph-parser?sect1=pto§2=hitoff&d=.../5,798,526>> [acesso em 30/12/2004].
44. **BLOSSER, T.H., REEVES, J.B., 1986.** Near Infrared Reflectance Spectroscopy for predicting chemical composition of undried feedstuff. J. Dairy Sci. v. 69 , p. 136 (abstr.).
45. **ABRAMS, S.M, SHENK, J.S., 1987.** Determination of forage quality by Near Infrared Reflectance Spectroscopy : Efficacy of broad based calibration equations. J. Dairy Sci. v. 70 , p. 806-813.
46. **LLAMES, C.R., FONTAINE, J., 1994.** Determination of amino acids in feeds: collaborative study. J. AOAC Int., v.77, nº 6, p. 1362-1402.
47. **ROSTAGNO, H.S., 2000.** Tabelas Brasileiras para aves e suínos: composição de alimentos e exigências nutricionais. Viçosa: Imprensa Universitária. UFV . p.141.
48. **FONTAINE, J., HÖRR, J., 2001.** Near Infrared Reflectance Spectroscopy, enables the fast and accurate prediction of essencial amino acids contents in soy , rapessed meal , sunflower meal , peas , fismel, meat meal products and poultry meal . J. Agric. Food. Chem. v. 49, p.57-66.
49. **ANDRIGUETTO, J.M., PERLY, L., MINARDI, I., FLEMMING, J.S., GEMAEL, A., SOUZA, G.A., FILHO, A.B., 1985.** As proteínas na nutrição animal. In: Alimentos e os princípios nutritivos vol. I . São Paulo: Nobel , 3ª ed., p.110-131.

50. **THE AMERICAN SOCIETY FOR TESTING AND MATERIALS (ASTM), 2000.** Pratices E 1655-00, ASTM Annual Book of Standards. Pennsylvania (USA): West Conshohocken, v.03.06.ASTM, p. 573-600.
51. **HANGEN, S.R., FROST, B., AUGUSTIN, J., 1989.** Precolumn phenylisothiocyanate derivatization and liquid-chromatography of amino-acids in food. J AOAC, v. 72, nº6, 912-916.
52. **EUROLYSINE, 1993.** Composição em aminoácidos de farelo de soja e milho no Brasil. São Paulo: Biolatina. p. 1-12.
53. **UNITED STATES- CANADIAN, 1990.** Table of feed amino acids composition. New York: Feed and food composition, 3nd., p. 1-12.
54. **PERILLA, N.S., CRUZ, M.P., DIAZ.G.J., 1997.** Effects of temperature of extrusion on the nutritional value of full fat soybeans for broiler chickens. Poutry Sci., v.38, p. 412-416.
55. **RHEE, K.C., 1998.** Effects of processing on nutrient content of soybean meal. Technical Bulletin. Bangkok: American Soybean Association - ASA, n.096/11/97, p.6.
56. **KESSHAVARZ, K., JAKSON, M.E., 1992.** Performace of growing pullets and laying beans feed low- protein, amino acids and suplements diets. Poutry Sci. , v.71, p. 905-918.
57. **VAN KENPEN, T.A.T.G., KIM, I.B., JANSMAN, A.J.M., 2002.** Regional and processor variation in ileal digestible amino acid content of soybean meals mesured in growing swine. J. Anim. Sci., v.80, p. 429-439.
58. **ROLIM, H.M.V., MAFFIA, L.M., COELHO, D.T. VIEIRA M., 1979.** Avaliação nutricional da proteína de soja texturizada por extrusão. Rev. Ceres, v. 26, nº 145, p. 275-290.