



DANIELA SOUZA FERREIRA

**“APLICAÇÃO DE ESPECTROSCOPIA NO INFRAVERMELHO E
ANÁLISE MULTIVARIADA PARA PREVISÃO DE PARÂMETROS DE
QUALIDADE EM SOJA E QUINOA”**

***“APPLICATION OF INFRARED SPECTROSCOPY AND
MULTIVARIATE ANALYSIS TO PREDICT QUALITY PARAMETERS
IN SOYBEAN AND QUINOA”***

CAMPINAS

2013



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA DE ALIMENTOS

DANIELA SOUZA FERREIRA

**APLICAÇÃO DE ESPECTROSCOPIA NO INFRAVERMELHO E
ANÁLISE MULTIVARIADA PARA PREVISÃO DE PARÂMETROS DE
QUALIDADE EM SOJA E QUINOA**

Orientadora: Profa. Dra. Juliana Azevedo Lima Pallone

Coorientador: Prof. Dr. Ronei Jesus Poppi

***APPLICATION OF INFRARED SPECTROSCOPY AND
MULTIVARIATE ANALYSIS TO PREDICT QUALITY PARAMETERS
IN SOYBEAN AND QUINOA***

Tese de doutorado apresentada ao Programa de Pós Graduação em
Ciência de Alimentos da Faculdade de Engenharia de Alimentos da Universidade
Estadual de Campinas para obtenção do título de Doutora em Ciência de Alimentos.

*Doctorate thesis presented to the Food Sciences Postgraduate
Program of the School of Food Engineering of the University of Campinas to obtain the
Ph.D. grade in Food Science.*

**ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE
DEFENDIDA PELA ALUNA DANIELA SOUZA FERREIRA E
ORIENTADA PELA PROFA. DRA. JULIANA AZEVEDO LIMA
PALLONE**

Assinatura da Orientadora

CAMPINAS

2013

FICHA CATALOGRÁFICA ELABORADA POR
CLAUDIA AP. ROMANO DE SOUZA – CRB8/5816 - BIBLIOTECA DA FACULDADE DE
ENGENHARIA DE ALIMENTOS – UNICAMP

F413a Ferreira, Daniela Souza, 1978-
Aplicação de espectroscopia no infravermelho e análise multivariada para previsão de parâmetros de qualidade em soja e quinoa / Daniela Souza Ferreira. -- Campinas, SP: [s.n.], 2013.

Orientador: Juliana Azevedo Lima Pallone.
Coorientador: Ronei Jesus Poppi.
Tese (doutorado) - Universidade Estadual de Campinas, Faculdade de Engenharia de Alimentos.

1. Análise de dados multivariados. 2. Espectroscopia na região do infravermelho próximo. 3. Composição centesimal. 4. Soja. 5. Quinoa. I. Pallone, Juliana Azevedo Lima. II. Poppi, Ronei Jesus. III. Universidade Estadual de Campinas. Faculdade de Engenharia de Alimentos. IV. Título.

Informações para Biblioteca Digital

Título em inglês: Application of infrared spectroscopy and multivariate analysis to predict quality parameters in soybean and quinoa

Palavras-chave em Inglês:

Multivariate data analysis

Near-infrared spectroscopy

Composition analysis

Soybean

Quinoa

Área de concentração: Ciência de Alimentos

Titulação: Doutora em Ciência de Alimentos

Banca examinadora:

Juliana Azevedo Lima Pallone

Caroline Joy Steel

Roger Wagner

Patrícia Valderrama

Alessandra Borin

Data da defesa: 27-03-2013

Programa de Pós Graduação: Ciência de Alimentos

BANCA EXAMINADORA

Profa. Dra. Juliana Azevedo Lima Pallone - Orientadora
Universidade Estadual de Campinas (FEA / UNICAMP)

Profa. Dra. Caroline Joy Steel
Universidade Estadual de Campinas (FEA / UNICAMP)

Prof. Dr. Roger Wagner
Universidade Federal de Santa Maria (UFSM)

Profa. Dra. Alessandra Borin
Pontifícia Universidade Católica de Campinas (Química / PUC-Campinas)

Prof. Dra. Patrícia Valderrama
Universidade Tecnológica Federal do Paraná (UTFPR)

Profa. Dra. Helena Teixeira Godoy
Universidade Estadual de Campinas (FEA / UNICAMP)

Prof. Dr. Célio Pasquini
Universidade Estadual de Campinas (IQ / UNICAMP)

Prof. Dr. Edenir Rodrigues Pereira Filho
Universidade Federal de São Carlos (DQ / UFSCAR)

Dedico este trabalho aos meus pais Antonio e Irany, exemplo constante em minha vida de dedicação, trabalho, incentivo e amor incondicional e ao meu irmão Eduardo, minha cunhada Ana Flávia e ao meu sobrinho Guilherme.

AGRADECIMENTOS

A Deus por estar presente diariamente em minha vida me dando força, saúde e proteção.

À minha orientadora, Profa. Dra. Juliana Pallone, pela paciência, amizade e confiança dada a mim e à minha capacidade de realização deste trabalho.

Ao Prof. Dr. Ronei Poppi pela orientação, confiança e grande oportunidade de crescimento científico.

Aos membros, titulares e suplentes, da banca da minha tese de doutorado pelas valiosas contribuições, críticas e sugestões que certamente elevaram a qualidade do trabalho.

Às minhas queridas técnicas/estagiárias, Gabriela, Suélen, Cristiane e Bruna pela colaboração prática do trabalho e por me ajudarem quando eu mais precisava.

À Prof. Dra. Helena Godoy, pelas contribuições ao longo do doutorado e por abrir a sua casa para os momentos de descontração.

Aos colegas do Laboratório de Análise de Alimentos por tornarem o ambiente de trabalho agradável e familiar.

A todos que de certa forma participaram do meu esforço e me apoiaram para a conclusão deste trabalho.

À Faculdade de Engenharia de Alimentos da UNICAMP e ao Departamento de Ciência de Alimentos que, juntamente com o programa de pós-graduação em Ciência de Alimentos, forneceram suporte estrutural para o desenvolvimento da minha tese.

Por fim, agradeço ao CNPq, CAPES e à FAPESP (Processo nº 10/50418-3) pelo apoio financeiro.

Muito obrigada!

"O conhecimento é o processo de acumular dados; a sabedoria reside na sua simplificação."

Martin Henry Fischer

"Qualquer coisa que você possa fazer, ou sonha fazer, comece a fazê-la. A ousadia tem em si genialidade, força e magia."

Johann Goethe

SUMÁRIO GERAL

SUMÁRIO	xiii
LISTA DE TABELAS.....	xvii
LISTA DE FIGURAS.....	xix
LISTA DE ABREVIATURAS E SIGLAS	xxi
RESUMO GERAL.....	xxiii
SUMMARY	xxv
INTRODUÇÃO GERAL.....	1
OBJETIVOS	7
CONCLUSÃO GERAL.....	119

SUMÁRIO

CAPÍTULO I

REVISÃO BIBLIOGRÁFICA

1. SOJA.....	11
2. QUINOA.....	12
3. COMPOSIÇÃO CENTESIMAL	13
4. ESPECTROSCOPIA NA REGIÃO DO INFRAVERMELHO – PRÓXIMO (NIR) E MÉDIO (MIR).....	14
4.1. Reflectância Difusa	19
4.2. Quimiometria.....	22
4.2.1. Pré-tratamento dos dados	23
4.2.1.1. Correção do espalhamento multiplicativo	24
4.2.1.2. Normalização e transformação padrão normal de variação	24
4.2.1.3. Diferenciação Savitzky-Golay	25

4.2.2. Análise de Componentes Principais	25
4.2.3. Regressão por Mínimos Quadrados Parciais	27
4.3. Aplicações das Técnicas NIR e MIR.....	29
Referências bibliográficas	32

CAPÍTULO II

COMPARAÇÃO E APLICAÇÃO DE ESPECTROSCOPIA NA REGIÃO DO INFRAVERMELHO PRÓXIMO (NIR) E MÉDIO (MIR) PARA DETERMINAÇÃO DE PARÂMETROS DE QUALIDADE EM AMOSTRAS DE SOJA

RESUMO	39
ABSTRACT	40
1 INTRODUÇÃO.....	41
2 MATERIAIS E MÉTODOS.....	43
2.1 Amostras	43
2.2 Espectroscopia na região do infravermelho próximo e médio	44
2.3 Análises Quimiométricas.....	44
3 RESULTADOS E DISCUSSÃO	45
4 CONCLUSÕES.....	53
Agradecimentos.....	53
Referências bibliográficas	53

CAPÍTULO III

FOURIER TRANSFORM NEAR-INFRARED SPECTROSCOPY (FT-NIRS) APPLICATION TO ESTIMATE BRAZILIAN SOYBEAN [GLYCINE MAX (L.) MERRIL] COMPOSITION

ABSTRACT	59
1. INTRODUCTION	60
2. MATERIALS AND METHODS	62
2.1. Samples	62
2.2. Sample Preparation.....	62

2.4. Near-infrared Analysis	62
2.5. Chemometrics Analysis.....	63
2.5.1. Data modeling	63
2.5.2. Sample choice for calibration and prediction	63
2.5.3. Pretreatment of spectra.....	63
2.5.4. Model evaluation	64
3. RESULTS AND DISCUSSION.....	66
3.1. Composition analysis in soybean.....	66
3.2. Development of calibration models and validation	66
4. CONCLUSION.....	73
Acknowledgements.....	73
References	74
APPENDIX	77

CAPÍTULO IV

COMBINATION OF NEAR-INFRARED SPECTROSCOPY AND CHEMOMETRICS TO PREDICT CHEMICAL CHARACTERISTICS OF CHENOPODIUM QUINOA FLOUR AND GRAINS

ABSTRACT	81
1. INTRODUCTION	82
2. MATERIALS AND METHODS	84
2.1. Samples	84
2.2. Reference Analysis	84
2.3. Near-infrared spectroscopy.....	85
2.4. Data processing	85
3. RESULTS AND DISCUSSION.....	86
3.1. Exploratory analysis	86
3.2. Chemical analysis	88

3.3. Calibration modeling and validation	89
4. CONCLUSION.....	95
ACKNOWLEDGEMENTS	95
REFERENCES	95
APPENDIX	99

CAPÍTULO V

*PREDIÇÃO DO CONTEÚDO DE FIBRAS ALIMENTARES EM SOJA (*Glycine max*) PELO MÉTODO NÃO DESTRUTIVO FT-NIRS*

RESUMO	103
ABSTRACT	104
INTRODUÇÃO.....	105
MATERIAIS E MÉTODOS.....	108
AMOSTRAS	108
MÉTODO ENZIMÁTICO-GRAVIMÉTRICO.....	108
ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO.....	109
ANÁLISES QUIMIOMÉTRICAS	109
RESULTADOS E DISCUSSÃO.....	110
ANÁLISE ENZIMÁTICA-GRAVIMÉTRICA.....	110
CALIBRAÇÃO MULTIVARIADA	111
CONCLUSÃO	115
AGRADECIMENTOS.....	116
REFERÊNCIAS BIBLIOGRÁFICAS.....	116

LISTA DE TABELAS

CAPÍTULO I

Tabela 1. Regiões Espectrais no infravermelho	14
---	----

CAPÍTULO II

Tabela 1. Parâmetros instrumentais para obtenção dos espectros de farinha de soja em infravermelho próximo (NIR) e infravermelho médio (MIR)	44
--	----

Tabela 2. Faixa de valores obtidos pelos métodos de referência para amostras de soja de diferentes regiões de cultivos.....	47
---	----

Tabela 3. Resultados da análise multivariada (PLS) dos parâmetros de soja com técnicas espectroscópicas NIR e MIR.....	48
--	----

Tabela 4. Comparação entre o valor medido e o valor previsto pelo modelo de calibração aplicado para os espectros NIR e MIR para proteínas e lipídios	51
---	----

Tabela 5. Comparação entre o valor medido e o valor previsto pelo modelo de calibração aplicado para os espectros NIR e MIR para cinzas e umidade	52
---	----

CAPÍTULO III

Table 1. Descriptive statistics of soybean chemical parameters used in the calibration and validation	66
---	----

Table 2. Wavenumber region and treatments used in calibration and validation	67
--	----

Table 3. Chemometrics results of PLS models by FT-NIRS of the evaluated parameters of soybeans	68
--	----

CAPÍTULO IV

Table 1. Chemical composition of the calibration and validation sample set (dry weight) .	89
---	----

Table 2. Prediction results of the moisture, ash, lipid, protein and carbohydrate content from NIRS spectra collected from quinoa grains and flour	91
--	----

Supplemental Table A. Optimal parameters for NIRS calibrations of moisture, ash, lipid, protein and carbohydrate content on quinoa flour and grain.....	99
---	----

CAPÍTULO V

Tabela 1. Resultados de fibras alimentares totais para soja e outros grãos.....	111
Tabela 2. Tentativa de associar bandas para celulose e lignina	113
Tabela 3. Parâmetros quimiométricos do modelo de calibração para fibras totais	114
Tabela 4. Comparação entre o valor medido e o valor previsto (grupo de validação) pelo modelo de calibração para o conteúdo de FD.....	115

LISTA DE FIGURAS

CAPÍTULO I

Figura 1. Ilustração do método de reflectância difusa com esfera de integração 21

Figura 2. Representação esquemática da decomposição por PCA..... 26

CAPÍTULO II

Figura 1. Espectros MIR para farinha de soja, obtidos na faixa entre 4000 a 400 cm^{-1} , sem pré-tratamento 45

Figura 2. Espectros NIR para farinha de soja, obtidos na faixa entre 10000 a 4000 cm^{-1} , sem pré-tratamento 46

Figura 3. Relação entre os valores reais e os valores preditos pelos modelos obtidos por NIR para os teores de proteínas (A), lipídios (B), cinzas (C) e umidade (D)..... 49

Figura 4. Relação entre os valores reais e os valores preditos pelos modelos obtidos por MIR para os teores de proteínas (E), lipídios (F), cinzas (G) e umidade (H) 50

CAPÍTULO III

Fig. 1. NIR spectra of soybean samples obtained in the range of 10000 to 4000 cm^{-1} 67

Fig. 2. Plot of the predicted values by NIR against the values measured by standard methods for moisture content..... 70

Fig. 3. Plot of the predicted values by NIR against the values measured by standard methods for ash content 70

Fig. 4. Plot of the predicted values by NIR against the values measured by standard methods for lipid content..... 71

Fig. 5. Plot of the predicted values by NIR against the values measured by standard methods for protein content 71

Fig. 6. Plot of the predicted values by NIR against the values measured by standard methods for carbohydrate content..... 71

Fig 7. Absolute errors: evaluated parameters measured minus the predicted values of the soybean validation group	72
Supplemental Figure A. Absolute error: evaluated value minus predicted value of the soybean validation group (I) moisture, (II) ash and (III) lipid contents.....	77
Supplemental Figure B. Absolute error: evaluated value minus predicted value of the soybean validation group (IV) protein and (V) carbohydrate contents	78
Supplemental Figure C. First derivative of the spectra obtained from ground soybean grains.....	78

CAPÍTULO IV

Figure 1. Score plot of the first and second principal components from PCA of the NIRS of quinoa grains	87
Figure 2. PCA loadings plot of the quinoa grains. (A) PC 1 (B) PC 2 vs. variables	88
Figure 3. Measured parameters vs. predicted values from NIRS of quinoa flour	93
Figure 4. Measured parameters vs. predicted values from NIRS of quinoa grain	94
Supplemental Figure A. Histogram of quinoa grain on NIRS mean centered	100
Supplemental Figure B. NIRS spectra of quinoa grain (a) and quinoa flour (b) without preprocessing	100

CAPÍTULO V

Figura 1. Espectros NIR sem pré-tratamento obtidos da soja moída	112
Figura 2. Coeficientes de regressão (“loadings”) para o modelo de calibração construído com o conteúdo de fibras alimentares centrado na média	112
Figura 3. Relação entre os valores previstos através do espectro NIR e os valores medidos em laboratório para fibra alimentar (FA) de soja.....	114

LISTA DE ABREVIATURAS E SIGLAS

FT IR	Infravermelho com Transformada de Fourier (<i>Fourier transform infrared spectroscopy</i>)
NIR	Espectroscopia na região do infravermelho próximo (<i>Near-infrared spectroscopy - reflectance</i>)
MIR	Espectroscopia na região do infravermelho médio (<i>Medium-infrared spectroscopy - reflectance</i>)
FIR	Espectroscopia na região do infravermelho distante (<i>Far-infrared spectroscopy - reflectance</i>)
PCA	Análise de Componentes Principais (<i>Principal component analysis</i>)
PLS	Mínimos Quadrados Parciais (<i>Partial least squares calibration</i>)
PC	Componente Principal (<i>Principal component</i>)
SNV	Transformação padrão de variação (<i>Standard normal variate</i>)
MSC	Correção de espalhamento multiplicativo (<i>Multiplicative scatter correction</i>)
NIPALS ou SIMPLS	<i>Straight Foward Implementation of Statistically Inspired Modification of PLS</i>
R ²	Coefficiente de determinação (<i>Coefficient of determination</i>)
RMSE	Raiz quadrada do erro médio quadrático (calibração, previsão ou validação)
RMSEC	<i>Root mean square error of calibration</i>
RMSEP	<i>Root mean square error of prediction</i>
RMSECV	<i>Root mean square error of cross validation</i>
RPD	<i>Residual predictive deviation</i>
SNR	Relação sinal ruído (<i>Signal-to-noise ratio</i>)
ATR	Reflectância total atenuada (<i>Attenuated total reflectance</i>)
UVE	Eliminação de variáveis não informativas (<i>Uninformative variables elimination</i>)
AACC	<i>American Association of Cereal Chemists</i>
AOAC	<i>Association of Official Agricultural Chemists</i>
FDA	<i>Food and Drug Administration</i>
FD	Fibras dietéticas totais
E%	Erro relativo

RESUMO GERAL

A avaliação da qualidade nutricional de alimentos é realizada principalmente por meio da determinação dos componentes majoritários, conhecida como composição centesimal (umidade, proteína, cinza, lipídio, carboidrato e fibra). No entanto, os métodos tradicionais de análise são demorados e utilizam materiais, equipamentos e diversos reagentes químicos, que além de oferecerem risco ao analista, geram resíduos tóxicos. Diante disto, uma alternativa para a análise química de grãos, rápida, de baixo custo e sem uso de reagentes químicos é a espectroscopia na região do infravermelho. Visando atender a demanda do Brasil por pesquisas empregando espectroscopia no infravermelho para análise de alimentos, o objetivo desse trabalho foi avaliar a possibilidade de utilização das técnicas espectroscopia no infravermelho próximo (NIR), principalmente, e médio MIR, associadas à quimiometria, para previsão de parâmetros de qualidade da soja brasileira e quinoa da América do Sul. Para comparar a aplicação de NIR e MIR, amostras de soja provenientes do Paraná foram analisadas pelas duas técnicas para previsão da composição centesimal. Os erros relativos (E%) entre os valores de referência e os valores previstos pelos modelos de calibração PLS, foram pequenos tanto para o NIR como para o MIR, no entanto, os resultados sugerem o uso de NIR para previsão de lipídios (0,2 a 9,2%) e o uso de MIR para proteínas (0,2 a 5,6%), cinzas (0 a 5,0%) e umidade (0,1 a 2,0%). Posteriormente, foram construídos modelos de calibração PLS com NIR para previsão dos parâmetros de qualidade em soja moída e para a quinoa, grão inteiro e moído. Os melhores modelos de calibração para soja encontrados neste estudo foram para o conteúdo de proteína e umidade, com melhores coeficientes de determinação e raiz quadrada do erro médio quadrático de calibração ($R^2 = 0,81$, RMSEC = 0,58% e $R^2 = 0,80$, RMSEC = 0,28%, respectivamente), contudo, a técnica mostrou capacidade adequada de predição para todos os parâmetros, incluindo lipídios, cinzas, carboidratos e fibras. Para amostras de quinoa, os espectros NIR foram inicialmente submetidos a uma análise de componentes principais (PCA) para tentar separá-las em grupos, de acordo com a origem geográfica destes grãos, os

quais eram provenientes do Brasil, Bolívia e Peru. Duas componentes principais explicaram 98,2% do total da variância e três grupos foram observados na separação por PCA de acordo com o país de origem. A técnica de calibração por PLS produziu modelos adequados, que permitiu a quantificação da composição majoritária tanto para o grão inteiro como farinha de quinoa, mostrando boa correlação entre o valor previsto e o valor real, com $R^2 > 0,65$ e $RMSEC < 1,70\%$. Portanto, este estudo demonstra que a técnica de NIR é potencialmente útil como um método analítico não destrutivo para determinações rápidas e simples de constituintes alimentares, além de não necessitar nenhum tipo de preparo de amostra, já que os espectros dos grãos inteiros de quinoa forneceram bons resultados para previsão dos parâmetros estudados.

SUMMARY

Evaluation of nutritional quality of food has been mainly performed by determination of major compounds, which is known as centesimal composition (moisture, protein, ash, lipid, carbohydrate and fiber). However, the traditional methods of analysis are time-consuming, use many materials and equipment, and also toxic reagents, that generate waste and are a risk for the analyst. Thus, infrared spectroscopy is an alternative to chemical analysis of grains, as it is a rapid, low cost technique and it does not use toxic reagents. In coming years, Brazilian researches using infrared for food analysis should increase, thus the objective of this work was to evaluate the possibility of application mainly of near-infrared (NIR) and mid-infrared (MIR) spectroscopy techniques coupled with chemometrics to predict quality parameters in Brazilian soybean and South America quinoa. In order to compare NIR and MIR techniques, the soybean group from Paraná (Brazil) was analyzed using both techniques to predict centesimal composition. The related errors (E%) between reference values and predicted values by partial least square (PLS) were low for both the NIR and the MIR. However, the results propose the use of NIR to predict lipid (E% of 0.2 to 9.2) content and the use of MIR to predict protein (E% of 0.2 to 5.6), ash (E% of 0 to 5.0), and moisture (E% of 0.1 to 2.0) contents. Subsequently, PLS regression models were constructed using NIR to predict quality parameters in ground soybean and quinoa, grain and ground. The best calibration models to soybean found in this study were the ones used to determine protein and moisture content ($R^2 = 0.81$, RMSEP = 1.61% and $R^2 = 0.80$, RMSEC = 1.55%, respectively). However, the technique shows high predictability for all parameters, including lipids, ash, carbohydrates and fibers, RMSECV of 0.40 to 2.30% and RMSEP 0.38 to 3.71%. For quinoa samples NIR spectra were obtained and principal component analysis (PCA) was applied to try to identify the geographic origin of quinoa samples, from Brazil, Peru and Bolivia. Two principal components explained 98.3% of the total variance and three groups were observed using PCA. The PLS models developed for the chemical composition showed that the proposed methodology

produced adequate results, as whole grain as ground quinoa, with the graph of the real and predicted concentration having a coefficient of determination (R^2) > 0.65 and RMSEC < 1.70%. The viability of the NIR technique with no waste generation, low cost, reduced time and no kind of sample preparation for replacing laborious methods of analysis was demonstrated because the results for grains were satisfactory.

INTRODUÇÃO GERAL

A espectroscopia na região do infravermelho próximo (NIR) ou médio (MIR) é uma técnica rápida, não destrutiva, capaz de fornecer informações para análise de substâncias orgânicas em minutos. Este método apresenta grande aplicação na indústria de alimentos e na agricultura desde a década de 60 até os dias atuais. Em 1910 William Coblentz (1912) aplicava espectroscopia no infravermelho para elucidação de grupos funcionais em compostos orgânicos e em 1960 Karl Norris aplicou infravermelho próximo em produtos da agricultura. Desde então, um grande número de metodologias analíticas desenvolvidas com base em espectros obtidos na região do infravermelho próximo e médio demonstra a habilidade desta técnica na determinação do teor de proteínas, umidade, carboidratos e gordura com rapidez e eficiência, em diferentes tipos de cereais (Hart; Norris; Golumbic, 1962; Settle, 1997; Williams; Norris, 2001).

Durante a avaliação simultânea de diversos parâmetros de qualidade, métodos de calibração multivariada são empregados para geração de modelos matemáticos para a quantificação de espécies químicas a partir dos espectros NIR ou MIR. Os métodos de calibração multivariada, que fazem parte da disciplina denominada quimiometria, necessitam de um grande número de resultados provenientes de ensaios analíticos, e espectros, para o desenvolvimento de um modelo matemático, ou seja, um significativo número de amostras. Além disso, uma vez construído o modelo, este deve ser validado com um conjunto de dados que não fizeram parte da calibração. Após estes procedimentos, para qualquer amostra analisada pelo espectro no infravermelho, obtém-se o resultado esperado em pouco tempo aplicando o modelo de calibração previamente construído (Agelet; Hurburgh, 2010; Benito; Ojeda; Rojas, 2008; Martens; Naes, 1992; Workman, 1996).

Conforme Huang *et al.* (2008), a técnica de NIR e a calibração multivariada já foram empregadas para a determinação de proteínas e umidade em trigo, onde uma estação de análise foi instalada na colheita; assim como realizaram

Maertens; Reyns e Baerdemaeker (2004) para avaliar parâmetros de qualidade como proteínas e umidade, também em trigo. MIR também tem sido reportado para detectar adulteração de óleo de oliva com vários óleos vegetais (óleo de girassol, óleo de soja, óleo de gergelim e óleo de milho) e monitorar o processo de oxidação (Vlachos et al., 2006). Mais recentemente, Szigedi *et al.* (2011) quantificaram proteína por FT-NIR (Infravermelho próximo com transformada de Fourier) em diversos vegetais e Agelet *et al.*, (2012) acompanharam danos causados em grãos de milho e soja através de NIR. Além desses trabalhos, verifica-se que diversos produtos da agricultura podem ser monitorados pela técnica, como: cevada (Cozzolino; Roumeliotis; Eglinton, 2013); algodão (Huang *et al.*, 2013), soja (Lee; Choung, 2011) e farinha de trigo durante a fermentação (Vigni; Cocchi, 2013)

A análise da composição centesimal em alimentos visa garantir a qualidade de produtos, atender exigências de órgãos oficiais de fiscalização e regulamentação, além de contribuir para a proteção da saúde do consumidor. Estas análises químicas, que incluem a determinação de umidade, proteínas, cinzas, lipídios e carboidratos, tradicionalmente utilizam métodos laboriosos, que consomem muito tempo e geram muitos resíduos químicos. Por outro lado, os métodos que empregam a técnica NIR e MIR fornecem resultados rápidos e precisos, com mínimo ou sem nenhum preparo de amostra e sem geração de resíduos; sendo uma alternativa para essas determinações (Balabin; Smirnov, 2011; Cen; He, 2007; Delwiche *et al.*, 1998; Hong *et al.*, 1996).

Paralelamente, observa-se uma maior preocupação da população brasileira com a alimentação e com isso, o aumento do interesse por alimentos funcionais. Dentro desta classe estão os grãos, conhecidos desde a antiguidade, como fonte de calorias, além de sua importância como preventivos de doenças cardiovasculares e degenerativas, já revelada por diversos estudos. Nestes aspectos, a soja e a quinoa têm se destacado. A soja é uma leguminosa conhecida por sua riqueza em proteínas e óleo, sendo um produto de exportação importante para o Brasil. A quinoa é um pseudocereal, superior aos cereais em

qualidade e proporção de proteínas, lipídios, vitaminas e minerais. Estes grãos estão sendo incluídos na dieta humana, principalmente na formulação de pães, massas, barras de cereais, cereais matinais, entre outros, valendo-se da ausência de glúten em sua composição (Borges *et al.*, 2010; Nagata; Yamada, 2009).

Portanto, técnicas de avaliação da qualidade de grãos, de forma rápida e não destrutiva, podem ser aplicadas para seleção dos mesmos, principalmente por se tratar de uma matriz alimentícia cuja composição apresenta alta variabilidade, sendo influenciada pela variedade ou cultivar, pelas condições climáticas e solo.

REFERÊNCIAS BIBLIOGRÁFICAS

AGELET, L. E. *et al.* Feasibility of near infrared spectroscopy for analyzing corn kernel damage and viability of soybean and corn kernels. **Journal of Cereal Science**, v. 55, n. 2, p. 160-165, mar. 2012.

AGELET, L. E.; HURBURGH, C. R. A Tutorial on Near Infrared Spectroscopy and Its Calibration. **Critical Reviews in Analytical Chemistry**, v. 40, n. 4, p. 246-260, 29 out. 2010.

BALABIN, R. M.; SMIRNOV, S. V. Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. **Talanta**, v. 85, n. 1, p. 562-8, 15 jul. 2011.

BENITO, M. T. J.; OJEDA, C. .; ROJAS, F. S. Process Analytical Chemistry: Applications of Near Infrared Spectrometry in Environmental and Food Analysis: An Overview. **Applied Spectroscopy Reviews**, v. 43, n. 5, p. 452-484, 24 jul. 2008.

BORGES, J. T. *et al.* Características físico-químicas, nutricionais e formas de consumo da quinoa (*Chenopodium quinoa* Willd.). **Temas Agrários**, v. 15, n. 1, p. 9-23, 2010.

CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**, v. 18, n. 2, p. 72-83, fev. 2007.

COBLENTZ, W. W. The diffuse reflecting power of various substances. **Journal of the Franklin Institute**, v. 174, n. 5, p. 549-582, nov. 1912.

COZZOLINO, D.; ROUMELIOTIS, S.; EGLINTON, J. Monitoring water uptake in whole barley (*Hordeum vulgare* L.) grain during steeping using near infrared reflectance spectroscopy. **Journal of Food Engineering**, v. 114, n. 4, p. 545-549, fev. 2013.

DELWICHE, S. R. *et al.* Protein content of wheat by near-infrared spectroscopy of whole grain : Collaborative study. **Journal of AOAC International Y**, v. 81, n. 3, p. 587-603, 1998.

HART, J. R.; NORRIS, K. H.; GOLUMBIC, C. Determination of the moisture content of seeds by near-infrared spectrophotometry of the methanol extracts. **Cereal Chemistry**, v. 39, p. 94-99, 1962.

HONG, J. *et al.* Near-infrared diffuse reflectance spectroscopic analysis of the amounts of moisture, protein, starch, amylose, and tannin in buckwheat flours. **Journal of Nutritional Science and Vitaminology**, v. 42, n. 4, p. 359-366, 1996.

HUANG, H. *et al.* Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. **Journal of Food Engineering**, v. 87, p. 303-313, 8 jan. 2008.

HUANG, Z. *et al.* Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed. **Industrial Crops and Products**, v. 43, p. 654-660, maio. 2013.

LEE, J. H.; CHOUNG, M.-G. Nondestructive determination of herbicide-resistant genetically modified soybean seeds using near-infrared reflectance spectroscopy. **Food Chemistry**, v. 126, n. 1, p. 368-373, maio. 2011.

MAERTENS, K.; REYNS, P.; BAERDEMAEKER, J. DE. On-line measurement of grain quality with NIR technology. **American Society of Agricultural Engineers**, v. 47, n. 4, p. 1135-1140, 2004.

MARTENS, H.; NAES, T. **Multivariate Calibration**. first ed. London: John Wiley & Sons, 1992. p. 438

NAGATA, J.; YAMADA, K. Effects of Simultaneous Intake of Soybean Protein and Diacylglycerol on Lipid Profiles and Body Fat Accumulation in Rats. **Bioscience, Biotechnology, and Biochemistry**, v. 73, n. 6, p. 1328-1332, 2009.

SETTLE, F. A. **Handbook of Instrumental Techniques for Analytical Chemistry**. 1^a ed. New Jersey: Prentice Hall, 1997. p. 543

SZIGEDI, T. *et al.* Protein content determination in Brassica oleracea species using FT-NIR technique and PLS regression. **International Journal of Food Science & Technology**, v. 47, n. 2, p. 436-440, 22 fev. 2011.

VIGNI, M. L.; COCCHI, M. Near infrared spectroscopy and multivariate analysis to evaluate wheat flour doughs leavening and bread properties. **Analytica chimica acta**, v. 764, p. 17-23, 18 fev. 2013.

VLACHOS, N. *et al.* Applications of Fourier transform-infrared spectroscopy to edible oils. **Analytica Chimica Acta**, v. 573-574, p. 459-65, 28 jul. 2006.

WILLIAMS, P.; NORRIS, K. H. **Near-infrared technology in the agricultural and food industries**. 2nd. ed. Universidade de Wisconsin - Madison: American Association of Cereal Chemists, 2001. p. 296

WORKMAN, J. J. J. Interpretive spectroscopy for near infrared. **Applied Spectroscopy Reviews**, v. 31, n. 3, p. 251-320, 1996.

OBJETIVOS

O objetivo principal do trabalho foi a aplicação de espectroscopia na região do infravermelho próximo, associada à calibração multivariada, para avaliação de parâmetros de qualidade de grãos funcionais como soja e quinoa que apresentam importância econômica e nutricional.

O desenvolvimento da tese consistiu das seguintes etapas:

- 1) Avaliar os parâmetros de qualidade, que incluem umidade, proteínas, cinzas, lipídios e carboidratos em soja e quinoa, por meio de métodos analíticos de referência.
- 2) Avaliar o teor de fibras dietéticas de soja através do método enzimático-gravimétrico.
- 3) Obter espectros no infravermelho próximo (NIR) e médio (MIR) por reflectância difusa de soja e espectros no infravermelho próximo por reflectância difusa de quinoa.
- 4) Desenvolver modelos de calibração multivariada através dos espectros obtidos por NIR, principalmente, e MIR juntamente com os resultados de métodos analíticos de referência.
- 5) Prever o conteúdo de umidade, proteínas, cinzas, lipídios, carboidratos e fibras e validar com grupos que não participaram da calibração.

CAPÍTULO I

REVISÃO BIBLIOGRÁFICA

1. Soja

A soja [*Glycine max* (L.) Merrill] equivale a aproximadamente 70 % de todas as sementes oleaginosas produzidas no mundo. De acordo com a Companhia Nacional de Abastecimento (CONAB, 2012), no Brasil, o ano-safra 2012/2013 prevê uma área de cultivo entre 26,42 e 27,33 milhões de hectares e produção entre 80,06 e 82,82 milhões de toneladas. No Centro Oeste se concentra a maior produção de soja, sendo Mato Grosso e Goiás os principais estados, com áreas plantadas de 7,8 milhões de hectares (colheita de 3,1 t/ha) e 2,8 milhões de hectares (colheita de 3,1 t/ha), respectivamente. O estado do Paraná é o segundo maior produtor do Brasil, com área plantada de 4,7 milhões de hectares (colheita de 3,2 t/ha) (CONAB, 2012).

Cultivada inicialmente na China, a soja foi introduzida no Brasil, no estado do Rio Grande do Sul, por volta de 1900 (Ruedell, 2003). Em 2003 o Brasil já figurava como o segundo maior produtor mundial de soja, ultrapassado apenas pelos Estados Unidos, participando com aproximadamente 26,8 % da produção mundial (EMBRAPA, 2012). A exportação brasileira de soja em grãos no ano-safra de 2010/2011 foi de 36,25 milhões de toneladas (CONAB, 2012).

Os atributos mais notáveis da soja são proteínas e lipídios, no entanto ela também é uma importante fonte de fibras dietéticas e isoflavonas (Vieira; Cabral; Paula, 1999). Diversos estudos apontam o potencial bioativo e a capacidade antioxidante de peptídeos provenientes de proteínas de soja, além de seu efeito hipocolesterolêmico (Beermann et al., 2009 e Koba et al., 1993 e Wang; Mejia, 2005).

A soja é destinada para produção de óleo de cozinha, na culinária oriental como missô, shoyu e tofú, como substituto da carne, em sucos, pães, saladas, massas, barras de cereais, cereais matinais e alimentos infantis (Caperuto; Amaya-Farfan; Camargo, 2000). Todos estes produtos possuem o apelo funcional, ou seja, a proposta de uma alimentação saudável.

2. Quinoa

O cultivo da quinoa (*Chenopodium quinoa* Willd.), um pseudocereal da família Chenopodiaceae, originário da região dos Andes a 4.000 m de altitude, tem despertado interesse em pesquisa e desenvolvimento nos últimos 15 anos, principalmente devido ao seu alto valor nutricional. Atualmente é cultivada na Bolívia, sendo o maior produtor com 20.000 ha de área plantada ao ano, seguido do Peru, Colômbia, Equador, Chile e Argentina (FAO Food and Agricultural Organization, 2012). Alguns estudos demonstram seu potencial de cultivo no Brasil, principalmente no plantio em sucessão de outras culturas, levando em consideração seu elevado rendimento de grãos e de biomassa. (Spehar, 2006).

A quinoa apresenta melhor qualidade de proteína e mais equilíbrio na distribuição de aminoácidos essenciais em relação a outros cereais e assemelha-se à caseína – fração proteica do leite (Gross *et al.*, 1989; Schlick; Bubenheim, 1996). Outra característica importante é que a quinoa não contém glúten em sua composição, além de seu elevado teor de fibras e vitaminas do complexo B (Ascheri; Nascimento; Spehar, 2002; Valencia-Chamorro, 2004), possui ácidos graxos similares aos do óleo de soja, sendo boa fonte de ácidos graxos essenciais como linoleico e linolênico (Spehar; Santos, 2002).

Por suas características nutricionais, a farinha de quinoa destaca-se como ingrediente alimentar altamente desejável para consumo como alimento de subsistência (base alimentar) ou para o enriquecimento da dieta de muitas comunidades em países subdesenvolvidos ou em desenvolvimento. O valor biológico de sua proteína faz com que seu grão seja aplicável na fortificação de farinhas de trigo, milho e tubérculos, enquanto a ausência de gliadinas (presentes no trigo) e frações proteicas correspondentes às gliadinas (encontradas na cevada, centeio e malte), a torna adequada para a elaboração de produtos farináceos popularmente referidos como “isentos de glúten”; aspectos importantes que possibilitam uma maior variedade e oferta de produtos alimentícios mais nutritivos e adequados aos portadores da doença celíaca (Hager *et al.*, 2012).

3. Composição centesimal

Para consumidores e cientistas que trabalham com tecnologia de grãos e cereais a qualidade destes alimentos e seus derivados é considerada, basicamente, como sinônimo de funcionalidade, ou seja, o grão é selecionado em conformidade com o produto desejado, proporcionando características organolépticas adequadas (sabor, aroma, cor), a preço acessível e/ou características diferenciadas.

Dados sobre a composição de alimentos consumidos nas diferentes regiões do Brasil fornecem elementos básicos para ações de orientação nutricional, baseadas no desenvolvimento local e diversificação da alimentação.

Em muitos países, o preço do grão é estipulado em função do seu conteúdo de proteína, de amido e pela dureza. Diversos estudos mostram que os parâmetros de qualidade podem ser significativamente variáveis, mesmo para grãos plantados na mesma região e na mesma época (Bramble *et al.*, 2002; Reyns; Spaepen; Baerdemaeker, 2000).

Para análises químicas de grãos é necessária uma amostragem representativa e a determinação dos nutrientes é realizada em laboratório de análises de alimentos, segundo critérios internacionais. A composição centesimal inclui a determinação do teor de umidade, proteínas, lipídios totais, carboidratos totais, fibra alimentar total e cinzas (NEPA/UNICAMP, 2012).

A umidade do grão está relacionada à sua qualidade frente à estocagem e o processamento. A cinza é o resíduo inorgânico que permanece após a queima da matéria orgânica, seu teor é aceito como índice de refinação em farinhas. Os carboidratos são os principais componentes dos grãos (Nielsen, 2010).

As proteínas exercem funções biológicas associadas às atividades vitais em toda célula viva, portanto, são importantes componentes na alimentação humana e caracterizam a qualidade do grão. Os lipídios são definidos como componentes

insolúveis em água e solúveis em solventes orgânicos, tais como éter de petróleo, éter etílico, acetona, clorofórmio e benzeno (Spector, 1999).

As fibras dietéticas ou alimentares consistem de partes comestíveis de plantas, polissacarídeos, ligninas e substâncias associadas, resistentes à digestão pelas enzimas alimentares humanas. São divididas entre solúveis e insolúveis. Produtos com mais de 6 g de fibras por 100 g de alimento são considerados ricos em fibras e fazem parte do grupo de alimentos funcionais (ANVISA Agência Nacional de Vigilância Sanitária, 2003).

A técnica de espectroscopia na região do infravermelho (NIR) é uma alternativa para determinar estas espécies químicas em alimentos e pode substituir os métodos analíticos tradicionais (Osborne, 2000).

4. Espectroscopia na região do infravermelho – Próximo (NIR) e Médio (MIR)

A região espectral do infravermelho é dividida em três partes (sub-regiões), sendo estas: a região do infravermelho próximo (NIR- *Near Infrared*), a região do infravermelho médio (MIR- *Middle Infrared*) e a região do infravermelho distante (FIR-*Far Infrared*), conforme mostrado na Tabela 1 (Osborne; Fearn; Hindle, 1993; Smith, 2011).

Tabela 1. Regiões Espectrais no infravermelho

Região no infravermelho	Intervalos de número de onda (ν) – (cm^{-1})	Intervalos de comprimento de onda (λ) – (nm)
Próximo (NIR)	14000 a 4000	750 a 2500
Médio (MIR)	4000 a 400	2500 a 25000
Distante (FIR)	400 a 100	25×10^3 a 100×10^3

A absorção da radiação infravermelha ocorre para espécies moleculares que possuem pequenas diferenças de energia entre diversos estados vibracionais e rotacionais.

Estas transições vibracionais ou rotacionais das moléculas são provocadas por variações de energia quando há incidência de radiação infravermelha. A resposta são espectros de absorção, emissão e reflexão. Uma ilustração da vibração atômica de estiramento seriam duas massas conectadas por uma mola. Se uma força é aplicada ao longo do eixo da mola, resulta em uma vibração chamada movimento harmônico simples. A energia potencial deste oscilador harmônico é proporcional ao trabalho requerido para deslocar a massa, e sua curva é uma parábola. A energia é máxima quando a mola está comprimida ou esticada em sua amplitude máxima. O sistema de oscilação é dependente da constante de força da mola e da massa do corpo, resultando na frequência natural do oscilador mecânico, mas esta frequência independe da energia fornecida ao sistema. Variações de energia resultam em uma variação na amplitude de vibração (Atkins; Paula, 2006; Hsu, 1997).

Deste modo, a frequência da vibração molecular (ν_m) é calculada a partir da Equação 1, após substituir m_1 e m_2 pela massa dos dois átomos.

$$\nu_m = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} = \frac{1}{2\pi} \sqrt{\frac{k(m_1 + m_2)}{m_1 m_2}} \quad \text{Equação 1}$$

A grandeza k é a constante de força da ligação química, a qual é uma medida de sua rigidez, μ é a massa reduzida das duas massas m_1 e m_2 conectadas por uma mola.

Para descrever o comportamento de partículas de dimensões atômicas, podemos considerar o conceito de oscilador harmônico para descrever as equações de onda da mecânica quântica, como a Equação 2:

$$E_v = \left(n + \frac{1}{2} \right) \frac{h}{2\pi} \sqrt{\frac{k}{\mu}}$$

Equação 2

As transações dos níveis de energia vibracional podem ocorrer devido à absorção da radiação e a diferença de energia ocorre tanto entre os estados quânticos vibracionais como pela variação do momento de dipolo. A partir das Equações 1 e 2, encontramos a Equação 3:

$$E_v = \left(n + \frac{1}{2} \right) h\nu_m,$$

Equação 3

Onde E_v é a energia vibracional, n é o número quântico vibracional, h é a constante de Planck ($6,63 \times 10^{-34}$ Joule-seg.) e ν_m é a frequência vibracional (Pasquini, 2003; Williams; Norris, 2001).

Em temperatura ambiente, as moléculas se encontram em seu nível fundamental de energia vibracional ($n=0$). A função fundamental de energia vibracional de uma molécula corresponde a uma transição do primeiro nível de energia $n=0$, para o primeiro estado excitado $n=1$, conforme descrito nas Equações 4 e 5.

$$E_0 = \left(\frac{1}{2} \right) h\nu_m$$

Equação 4

$$E_1 = \left(\frac{3}{2} \right) h\nu_m$$

Equação 5

A frequência da radiação ν que produzirá esta variação de energia é idêntica à frequência vibracional clássica, como apresentado na Equação 6:

$$\nu = \nu_m = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}$$

Equação 6

Se desejarmos expressar a radiação em números de onda, rearranjamos:

$$\bar{\nu} = \frac{1}{2\pi c} \sqrt{\frac{k}{\mu}} = 5,3 \times 10^{-12} \sqrt{\frac{k}{\mu}}$$

Equação 7

Onde $\bar{\nu}$ é o número de onda de máxima absorção (cm^{-1}), k é a constante de força da ligação em newtons por metros (N/m), c é a velocidade da luz (cm/s) e μ é a massa reduzida (kg). Esta equação pode ser usada para estimar o número de onda da banda de absorção fundamental ou a absorção devido à transição do estado fundamental ao primeiro estado excitado para uma variedade de tipos de ligação (Hsu, 1997).

A região do MIR (*mid-infrared*) é a principal região da espectroscopia vibracional. Esta região está relacionada principalmente com as transições vibracionais fundamentais, que envolvem o estado fundamental e o primeiro estado vibracional excitado. Os espectros no MIR permitem a identificação de moléculas orgânicas e a caracterização de grupos funcionais pertencentes às proteínas, polissacarídeos e lipídios (Bellon; Vigneau; Sévila, 1994).

A região MIR começou a ser aplicada no final dos anos 50 para análise qualitativa de compostos orgânicos, desde então, sua aplicação estava relacionada principalmente à identificação de ligações moleculares. Até meados da década de 70, a maioria dos instrumentos era do tipo dispersivo baseado em redes de difração. Com o surgimento dos espectrômetros com Transformada de Fourier (FT), relativamente baratos, aumentou notavelmente o número e o tipo de aplicações da espectroscopia no infravermelho. Dentre as vantagens dos instrumentos que empregam a Transformada de Fourier estão: a melhora da relação sinal-ruído, a significativa redução no tempo de varredura, melhora na resolução espectral e reprodutibilidade no comprimento de onda. A partir daí, MIR passou a ser usado também para análise quantitativa de amostras complexas por absorção (Hsu, 1997; Sun, 2009).

As bandas de absorção na região NIR são resultado de sobretons (transições mais fracas, com ΔE menor e $\Delta n = \pm 2$ e ± 3) ou combinações de modos vibracionais fundamentais (bandas de combinação) observados na faixa do MIR. Existe correlação para mostrar onde certos grupos funcionais absorvem radiação na região espectral do infravermelho próximo. Considerando um exemplo, a ligação O-H absorve radiação na frequência 3600 cm^{-1} . O primeiro, segundo e terceiro sobretons deste modo fundamental de vibração podem ser observados no espectro de NIR nas regiões 7200 , 8900 e 13800 cm^{-1} , respectivamente (Cen; He, 2007).

O primeiro, segundo e terceiro sobretons correspondem a transições do estado fundamental diretamente para os estados $n=2$, $n=3$ e $n=4$, respectivamente. Se as vibrações das ligações moleculares se comportassem como perfeitas oscilações harmônicas, então os níveis de energia seriam igualmente espaçados. Mas na verdade as oscilações são anarmônicas e os níveis de energia não são perfeitamente espaçados. Em função desta não harmonicidade, as transições de sobreton podem ser observadas, embora estas transições sejam 10 a 100 vezes mais fracas que a transição fundamental. O terceiro sobreton somente pode ser observado por bandas fundamentais muito fortes e que resultem de modos vibracionais com alta anarmonicidade. Geralmente o quarto sobreton é muito fraco para ser observado (Burns; Ciurczak, 2007; Kalivas; Gemperline, 2006).

Bandas de combinações correspondem a transições simultâneas em dois modos. Por exemplo, a molécula que possui um grupo funcional carbonila ($\nu_1 = 1750 \text{ cm}^{-1}$) e hidroxila ($\nu_2 = 3600 \text{ cm}^{-1}$) próximos pode mostrar a combinação da banda $\nu_1 + \nu_2 = 5350 \text{ cm}^{-1}$ (Burns; Ciurczak, 2007).

Na região do infravermelho, várias vibrações moleculares fundamentais podem ser detectadas, tais como C-H, N-H, O-H e C=O ou outros grupos funcionais (Shenk; Workman; Westerhaus, 2007; Sun, 2009). Quando um alimento é irradiado com radiação infravermelha, ele interage (absorve) energia de acordo

com sua característica funcional e outras frequências serão transmitidas ou refletidas. Por exemplo, concentração de proteína é detectada em espectros de amostras de trigo através de ligações N-H, bem como ligações O-H da água são detectadas e empregadas como base para determinação de umidade. A radiação também é afetada pelas propriedades físicas, como tamanho das partículas, reflectividade da superfície e densidade. O NIR é capaz de detectar matéria orgânica em quantidade acima de 0,1 % na amostra (Kalivas; Gemperline, 2006; Workman, 1996).

Dentre as vantagens da análise por espectroscopia no infravermelho estão: (i) a rapidez do método, pois com o instrumento devidamente calibrado, detector adequado, é possível obter resultados de uma análise em minutos; (ii) simplicidade da operação, pois, após a calibração do instrumento, a análise da amostra torna-se rotineira; (iii) redução da necessidade de uso de equipamentos e reagentes químicos. O processo de calibração utiliza métodos laboratoriais convencionais, no entanto, após o procedimento, é necessária apenas a utilização de computador equipado com software adequado; (iv) análise simultânea de multicomponentes, sendo que mais de uma calibração pode ser efetuada para o instrumento (Dendy, 2001; Shenk; Workman; Westerhaus, 2007; Sun, 2009).

A aplicação de espectroscopia na região do infravermelho vai de encontro com os princípios da química verde, onde a tendência da química analítica é a redução de tempo de análise, maior eficiência para obtenção de respostas, além da diminuição do consumo de reagentes e, conseqüentemente, da geração de substâncias nocivas.

Reflectância Difusa

Existem diversas técnicas de aquisição de dados espectrais de amostras para NIR e MIR. Eles são dependentes da finalidade de aplicação da técnica de infravermelho. Em transmitância, a radiação é detectada após atravessar a

amostra. Na reflectância, a radiação medida é aquela refletida da amostra e a relação sinal ruído (SNR) é muito menor. A medida de reflectância pode ser realizada por: a) reflectância total atenuada (ATR), onde a amostra é aplicada em um cristal altamente refrativo; b) reflectância especular, onde o ângulo de incidência de energia na amostra é exatamente o mesmo refletido; e c) reflectância difusa, onde há interação da radiação com a amostra e ocorre absorção e espalhamento (Coblentz, 1912; Hsu, 1997; Sun, 2009).

A reflectância difusa é uma importante forma de medida, capaz de determinar os constituintes de amostras na forma de pequenas partículas sólidas (inteiras ou em pó). Medidas por reflectância são obtidas a partir da radiação espalhada da superfície da amostra (Williams; Norris, 2001). A radiação alcança certo grau de penetração na amostra, é absorvida pela mesma e retorna para a superfície (Agelet; Hurburgh, 2010).

A medida absoluta da reflectância é o logaritmo da reflectância relativa, ou seja, a razão entre a intensidade de radiação refletida da amostra e a reflectância de um padrão de reflexão (referência), tais como brometo de potássio (KBr), sulfato de bário (BaSO_4), óxido de magnésio e o mais utilizado atualmente, o *Spectralon*, com aproximadamente 100% de reflectância. No entanto, o caminho percorrido pela radiação não pode ser controlado e varia de acordo com o tamanho das partículas e o índice de refração. A radiação que retorna de forma difusa de uma amostra é geralmente muito inferior que a radiação incidente. Em função disto, os acessórios de reflectância difusa apresentam esquemas óticos para convergir a radiação ao sistema de detecção do instrumento. A reflectância pode ser relacionada com a absorbância conforme a Equação 8, como uma forma de aproximação proposital para linearizar a relação entre a reflectância medida ($R_{relativa}$) e a concentração da espécie absorvida (“ASTM,” 2011):

$$Absorbância = -\log(R_{relativa}) = \log\left(\frac{1}{R_{relativo}}\right) = \log\left(\frac{R_{referência}}{R_{amostra}}\right) = \log\left(\frac{1}{R_{amostra}}\right) + \log(R_{referência})$$

Equação 8

Onde a reflectância relativa ($R_{relativa}$) é medida como a razão entre a reflectância medida da amostra ($R_{amostra}$) e o material altamente refletivo ($R_{referência}$) (Agelet; Hurburgh, 2010; Kortum, 1996).

Um acessório de reflectância difusa que contribui para a qualidade da análise é a esfera de integração. Trata-se de uma esfera altamente refletiva que inclui um conjunto de espelhos, sendo um deles reflete e direciona o feixe de energia diretamente para a amostra, que está posicionada em um recipiente transparente. A esfera faz uma integralização da energia refletida da amostra várias vezes antes de chegar ao detector (Figura 1) (Hanssen; Snail, 2001).

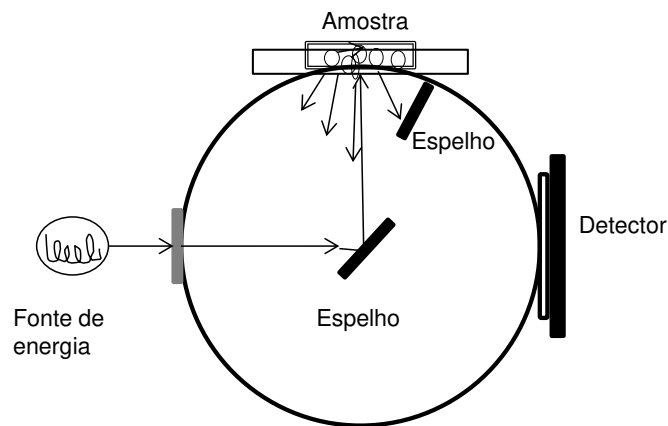


Figura 1. Ilustração do método de reflectância difusa com esfera de integração

As medidas da reflectância dependem das condições experimentais, do ângulo de incidência e de reflexão, da temperatura, do tamanho das partículas da amostra, entre outros. Portanto, para atingir qualidade analítica é importante o controle desses parâmetros. Focando esta qualidade de análise, no início de 1980 medidas com NIR através de reflectância tornaram-se métodos oficiais de determinação de proteína em trigo pela *American Association of Cereal Chemists* ("AACC International," 2012).

Quimiometria

O uso de ferramentas estatísticas, matemáticas e gráficas para resolver sistemas químicos deu origem a uma área da química conhecida como quimiometria. A quimiometria pode ser útil tanto no planejamento e otimização de experimentos quanto na análise de resultados para obtenção da maior quantidade de informação possível (Munck *et al.*, 2010).

Geladi e Kowalski (1986) afirmam que o surgimento da quimiometria se deve a dois fatores: i) a introdução de equipamentos que fornecem respostas contendo enorme quantidade de variáveis por amostra e ii) ao aumento do acesso aos computadores.

Em geral, para obtenção de informações qualitativas e quantitativas a partir de espectros complexos na região do infravermelho, é imprescindível a utilização de métodos quimiométricos multivariados. Tais métodos são desenvolvidos e disponibilizados em programas computacionais e são, juntamente com os avanços tecnológicos dos instrumentos, os responsáveis pela popularização do uso de espectroscopia vibracional. Os softwares mais utilizados são: The Unscrambler®X (CAMO Software AS), MATLAB (Mathworks, Inc.), Pirouette (Infometrix, Inc.), Statistica (Statsoft) e Octave (www.octave.org).

Uma importante ferramenta estatística aplicada para análise de dados multivariados quantitativos é a regressão por mínimos quadrados parciais (*partial least squares* – PLS), aplicada desde a década de 80 com vantagens sobre os métodos mais antigos de calibração, como CLS (*classical least squares*), ILS (*inverse least squares*) e PCR (*principal component regression*). O PLS, assim como análise de componentes principais (PCA), decompõe a matriz dos espectros no infravermelho para um menor número de dados em um novo sistema de coordenadas com vetores *loading*. O PLS possui a etapa de calibração que tenta relacionar as informações espectrais com os componentes da amostra (Haaland; Thomas, 1988; Martens; Naes, 1992).

Não há um número fixo de amostras necessárias para o desenvolvimento da calibração, mas alguns autores apontam números entre 20 e 30 amostras para viabilizar o estudo e iniciar uma calibração (Williams; Norris, 2001). A ASTM E1665-5 (“ASTM,” 2011) preconiza um número de amostras suficiente para definir a relação entre as variáveis espectrais e a concentração do componente de interesse. Se a amostra possui um grande número de componentes com alta variabilidade, então é necessário um grande número de amostras para o desenvolvimento do modelo, e somente após o desenvolvimento do modelo é possível definir o número de variáveis espectrais necessárias para a calibração.

Muitas vezes, em modelos de calibração multivariada, há um número muito grande de variáveis (número de onda) em relação ao número de amostras, portanto, a redução do número destas variáveis é aplicada para melhorar o desempenho do modelo e aumentar a capacidade preditiva. A seleção pode extrair variáveis irrelevantes, ruído e variáveis indesejáveis que elevam a complexidade do modelo. A PCA pode ser utilizada para seleção de variáveis, além de outros recursos como *forward selection*, *backward elimination*, *stepwise*, algoritmo genético (GA), eliminação de variáveis não informativas (*uninformative variables elimination* - UVE) e o comumente utilizado iPLS, que seleciona faixas espectrais relevantes. Além disso, alguns instrumentos de alta resolução possuem um filtro com esta finalidade (Andersen; Bro, 2010; Kalivas; Gemperline, 2006).

4.1.1. Pré-tratamento dos dados

Muitas vezes os dados a serem modelados são expressos em grandezas diferentes, apresentam muitos ruídos ou variações de linha de base que podem prejudicar o desempenho do modelo. Portanto, uma etapa importante na análise exploratória de dados, classificação ou na construção de um bom e robusto modelo de calibração, consiste no pré-tratamento dos dados espectrais.

Assumindo uma relação linear entre a concentração do analito e os dados espectrais, conforme a Equação 9:

$$c = Rb + e$$

Equação 9

onde c ($n \times 1$) é o vetor da concentração do analito, R ($n \times p$) é a matriz das respostas espectrais, b ($p \times 1$) é o vetor de regressão e e é o vetor residual. n é o número de amostras de calibração e p é o número de comprimentos de onda.

Muitos fenômenos físicos e químicos provocam um desvio desta relação linear entre a concentração e a absorvância, como espalhamento de luz das partículas, diferentes tamanhos de partículas, interferentes, interações moleculares, mudanças no índice refrativo em altas concentrações (Bjorsvik; Martens, 2007). Os pré-tratamentos são designados para compensar estes desvios e assim melhorar a relação linear entre o sinal espectral e a concentração do analito (Brown; Vega-Montoto; Wentzell, 2000; Sun, 2009).

4.1.1.1. Correção do espalhamento multiplicativo

O método de correção de espalhamento multiplicativo (MSC – *multiplicative scatter correction*) é comumente aplicado em espectroscopia para correção do efeito de espalhamento de luz presente nos espectros obtidos por reflexão. Este espalhamento é provocado pela não homogeneidade da amostra, diferenças de granulometria, empacotamento e orientação das partículas. Este método tenta remover o efeito do espalhamento pela linearização de cada espectro em função de um espectro de referência, usualmente, considera-se o espectro médio do grupo de amostras para o qual se deseja realizar a correção de linha de base. É aplicada uma regressão linear para calcular o coeficiente angular e linear do gráfico entre o espectro ideal e o espectro que será corrigido. O espectro corrigido é calculado subtraindo cada ponto do espectro pelo valor do coeficiente linear dividindo este valor pelo coeficiente angular (Geladi; MacDougall; Martens, 1985).

4.1.1.2. Normalização e transformação padrão normal de variação

Transformação padrão de variação (SNV – *standard normal variate transformation*) também é uma técnica de correção de espalhamento multiplicativo e tamanho de partícula, semelhante ao MSC. Mas o SNV não necessita de um

espectro ideal para fazer a correção dos espectros. A correção é realizada pela normalização de cada espectro para o seu próprio desvio padrão.

A normalização é utilizada para remover a variação sistemática da resposta espectral de cada variável. Cada uma das variáveis (de uma dada amostra) é dividida por um fator de normalização. Como resultado, todas as amostras estarão numa mesma escala (Barnes; Dhanoa; Lister, 1989; Gemperline, 2006).

4.1.1.3. Diferenciação Savitzky-Golay

A rotina de Savitzky-Golay (Brown; Vega-Montoto; Wentzell, 2000) calcula derivadas e pode ser aplicada para redução de ruído e para destacar sinais espectrais de interesse. Os dados são filtrados através de uma janela geralmente simétrica, esta janela contém $2g + 1$ pontos (onde g é o número de pontos de cada lado), que irá estimar o ponto central. Os $2g + 1$ pontos são ajustados por um polinômio de determinada ordem, e os coeficientes encontrados com este ajuste são usados para estimar a nova frequência. Em alguns casos, a quantidade de ruído pode ser acentuada na seleção de uma janela muito ampla, o que pode ocasionar perda de informações importantes na escala espectral. Normalmente 7 a 11 pontos para suavizar o espectro, além de segundo ou quarto grau polinomial para o procedimento de ajuste, é suficiente para alta resolução do espectro. E quanto mais alto grau de derivada, como a segunda derivada, maior o número de pontos na janela para a redução da razão sinal ruído (Brown; Vega-Montoto; Wentzell, 2000; Sun, 2009).

4.1.2. Análise de Componentes Principais

A análise de componentes principais (PCA – *principal component analysis*) é um tipo de análise exploratória de dados que visa extrair o máximo das informações de uma tabela de dados convertendo-a em gráficos que mostram a relação entre amostras (linhas de uma matriz) e as variáveis (colunas de uma matriz) (Beebe; Kowalski, 1987).

A PCA produz uma aproximação da tabela de dados, ou seja, uma matriz X em termos da soma de várias matrizes M_i de posto, na qual posto significa um número que expressa a verdadeira dimensionalidade da matriz, como mostra a Figura 2.

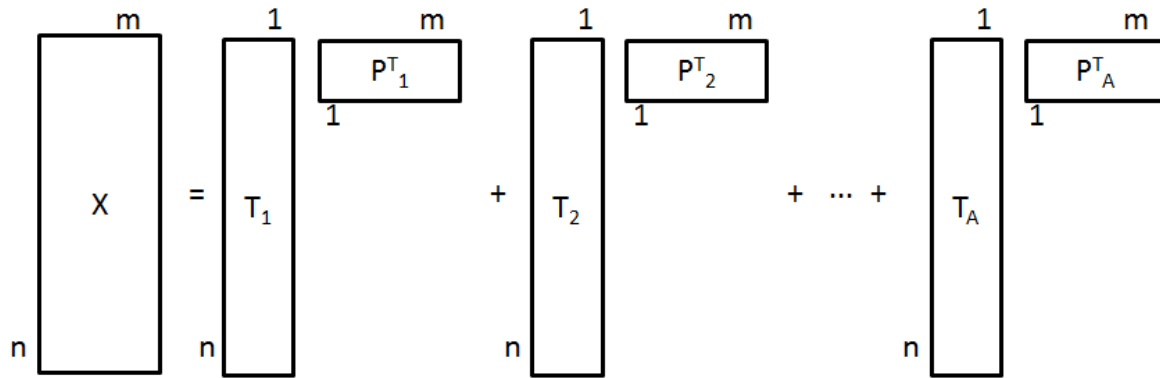


Figura 2. Representação esquemática da decomposição por PCA. Fonte: Geladi; Kowalski (1986)

Essa matriz pode ser escrita como produto de dois vetores, escore (*scores*) t_h e peso (*loadings*) p_h .

$$X = t_1 p_1 + t_2 p_2 + \dots + t_a p_a, \quad \text{Equação 10}$$

para um número igual a “a” de componentes principais.

Outra maneira de escrever a equação anterior é:

$$X = TP^T \quad \text{Equação 11}$$

Onde X é o conjunto original dos dados com n linhas (amostras) e p colunas (variáveis); T é a matriz dos escores com n linhas e d colunas (número de componentes principais, novas coordenadas no novo sistema de eixos), P é a matriz dos pesos (informação do peso de cada variável original na formação dos novos eixos) com d linhas e p colunas.

O novo conjunto de variáveis (componentes principais) resulta da combinação linear das variáveis originais. Os novos eixos são ortogonais entre si e se ordenam de forma decrescente em relação à quantidade de variância dos dados originais que descrevem. Assim, podemos dizer que a PCA tenta agrupar aquelas variáveis que estão altamente correlacionadas numa nova variável chamada componente principal (Gemperline, 2006).

Como usualmente a maior fração da variância é descrita pelas primeiras componentes principais, é possível visualizarmos dados pelo gráfico dos escores de uma componente contra a outra.

4.1.3. Regressão por Mínimos Quadrados Parciais

O método de calibração multivariada mais amplamente utilizado é a Regressão por Mínimos Quadrados Parciais (PLS – *partial least squares*) desenvolvido na década de 70 por Herman Wold e baseado em uma relação linear entre as variáveis instrumentais (X) e as variáveis de interesse (Y). As informações da matriz X e da matriz Y são usadas ao mesmo tempo na fase de calibração. A matriz dos espectros é decomposta em matrizes de variações dos espectros (*loadings* ou pesos) e a posição das amostras (escores). Os espectros originais podem ser considerados como combinações lineares dos espectros (pesos) onde os escores representam suas contribuições (Geladi; Kowalski, 1986).

As matrizes X e Y podem ser representadas pela PCA:

$$X = TP^T + E \quad \text{Equação 12}$$

$$Y = UQ^T + F \quad \text{Equação 13}$$

onde T e U são as matrizes escores de X e Y, respectivamente; P e Q são as matrizes peso de X e Y; E e F são os resíduos.

Uma relação linear entre os dois blocos pode ser realizada correlacionando os escores para cada componente de cada vez, utilizando o modelo linear.

$$U_h = b_h T_h$$

Equação 14

onde “h” é o número de variáveis latentes.

Para que a covariância de T e U seja maximizada deve-se buscar um modelo onde as matrizes dos resíduos E e F sejam as menores possíveis e, ao mesmo tempo, conseguir uma relação linear ótima entre T e U.

No PLS isto é obtido por uma leve mudança nos valores dos escores, de forma a produzir a melhor relação possível. Nesta etapa, as componentes principais deixam de possuir este nome sendo chamadas então de variáveis latentes, pois elas não mais descrevem a máxima variância dos dados.

O algoritmo utilizado para a estimativa dos vetores coeficientes de regressão b é o NIPALS ou SIMPLS (do inglês, *Straight Foward Implementation of Statistically Inspired Modification of PLS*). Chamando A_{max} o número máximo de fatores a serem computados pelo algoritmo, executa-se os seguintes passos para cada um dos fatores $a=1,2,...A_{max}$.

- 1) Encontrar vetor peso W_a para maximização da covariância entre a combinação linear $X_{a-1}W_a$ e y , com condição que $w_a^T w_a = 1$. Isto corresponde encontrar o vetor unitário w_a que maximiza $w_a^T X_{a-1}^T y_{a-1}$, ou seja, a variância escalada entre X_{a-1} e y_{a-1} .
- 2) Encontrar os escores, t_a como projeção de X_{a-1} em w_a , isto é, $t_a = X_{a-1} W_a$;
- 3) Realizar a regressão de X_{a-1} em t_a para encontrar os vetores pesos

$$p_a^T = X_{a-1}^T t_a / t_a^T t_a;$$

- 4) Realizar regressão de y_{a-1} em t_a para encontrar vetores pesos

$$q_a = y_{a-1} t_a / t_a^T t_a;$$

- 5) Subtrair $t_a p_a^T$ de X_{a-1} e chamar esta nova matriz de X_a e subtrair $t_a q_a$ de y_{a-1} e chamar esta nova matriz de y_a ;

6) Otimizar o número de fatores por validação cruzada, e considerar as seguintes matrizes:

$$W=\{w_a\};$$

$$P=\{p_a\};$$

Para a número de fatores otimizados;

7) Calculam-se, então, os coeficientes de regressão b através da seguinte relação:

$$b=W(P^TW)^{-1}q$$

em que W é a matriz de pesos do PLS.

Em seguida, é feita a validação do modelo com novas amostras. A validação cruzada é baseada na avaliação da grandeza dos erros de previsão comparando com 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 calibração. Na validação cruzada “um por vez” (*leave one out*), o PLS separa uma das amostras de calibração e a usa como amostra de validação. Isto é feito até que todas as amostras de calibração sejam usadas como amostra de validação. Em paralelo, é realizada 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 desempenho muito semelhante em todos os parâmetros de avaliação dos modelos de calibração.

Os modelos de calibração construídos para NIR e MIR devem seguir os procedimentos práticos da norma padrão ASTM E1655-05 (“ASTM,” 2011).

Aplicações das Técnicas NIR e MIR

Vários estudos têm reportado a habilidade da espectroscopia no infravermelho em caracterizar produtos em termos de composição química,

características físicas e sensoriais, bem como em propostas de classificação. As técnicas MIR e NIR têm sido aplicadas em processos de autenticação, com o intuito de avaliar se o produto está estritamente em conformidade com o rótulo descrito ou obedece a legislação vigente (Hourant *et al.*, 2000; Reid *et al.*, 2005; Vlachos *et al.*, 2006).

No campo de análise de óleos, MIR tem sido efetivamente usado para determinar o conteúdo de *-cis* e *-trans*, iodeto, índice de saponificação, peróxido e ácidos graxos livres em amostras de óleos e gorduras (Innawong *et al.*, 2004; Voort; Sedman; Russin, 2001). MIR também tem sido reportada para detectar adulteração de óleo de oliva com vários óleos vegetais (óleo de girassol, óleo de soja, óleo de gergelim e óleo de milho) e monitorar o processo de oxidação (Vlachos *et al.*, 2006). Mas, NIR e espectroscopia Raman também já foram empregadas para detectar adulteração em óleo, usado em ração animal (Haughey *et al.*, 2012).

Desde o início da década de 1960, a tecnologia de NIR tem sido associada a análises de grãos e produtos derivados (Hart; Norris; Golumbic, 1962). Em 1975, NIR foi adotada pela Comissão Canadense de Grãos como um método oficial de análise de proteínas em trigo (“Canadian Grain Commission,” 2012). Em 1980, a técnica foi adotada pelo Serviço de Inspeção Federal de Grãos dos Estados Unidos, e o teste de proteína de trigo por NIR foi finalmente aprovado pela *American Association of Cereal Chemists* (AACC) em 1982 (Osborne, 2000). Atualmente, os instrumentos comerciais permitem análises sem preparo das amostras. Na área de grãos e sementes oleaginosas é possível aplicar reflectância difusa para determinação de umidade e lipídios (Kandala; Sundaram; Puppala, 2012).

Sorensen (2009) utilizou medidas de reflectância e a espectroscopia no infravermelho próximo para determinar proteínas, gorduras, fibras dietéticas, açúcar, cinzas, ácidos graxos saturados, ácidos graxos monoinsaturados, ácidos graxos insaturados e sódio presentes em diversas variedades de pães constituídos de trigo e centeio. Os modelos de calibração foram construídos

utilizando PLS e os espectros médios foram corrigidos ou não por MSC. As análises de referência foram realizadas por métodos convencionais em laboratório. O autor verificou melhores resultados, com menores erros na validação cruzada, para reflectância na região de 1100-2500 nm do que transmitância na faixa de 900-1100 nm. Os resultados de RMSECV (raiz quadrada do erro médio quadrático de validação cruzada) foram baixos para todas as determinações, indicando que o NIR foi eficiente para as análises realizadas em pão, com exceção de sódio, que apresentou R^2 de 0,76, uma vez que analitos inorgânicos geralmente não podem ser determinados diretamente por NIR.

Para a calibração multivariada de espectros obtidos por reflectância difusa em NIR para amostras de trigo, foi aplicado *Genetic inverse least squares* (GILS), método baseado em algoritmo genético (GA). Os parâmetros previstos pelo modelo foram: umidade, proteína, peso seco, proteína em base seca e dureza do grão (Ozdemir, 2006).

A área agrícola constitui o maior campo de aplicação de NIR. O emprego da técnica para análise de amadurecimento de grão de trigo é realizado na Hungria, já que a técnica não provoca destruição na amostra. Com o uso do NIR é possível observar a mudança de absorção de água em duas bandas de combinação, além de carboidratos e proteínas (Salgó; Gergely, 2012).

A breve avaliação da importância da espectroscopia no infravermelho para a indústria e a pesquisa foi realizada através de uma busca com ajuda do *ISI Web of Knowledge (Web of Science)* através dos termos “calibração” e “NIR”. Nos últimos 5 anos foi possível encontrar cerca de 2000 artigos na área, sendo apenas 6 do Brasil. O Brasil ainda apresenta uma pequena contribuição, embora crescente, associada ao desenvolvimento e aplicação da tecnologia NIR na área agrícola.

Referências bibliográficas

AACC International. Disponível em: <<http://www.aaccnet.org/Pages/default.aspx>>.

AGELET, L. E.; HURBURGH, C. R. A Tutorial on Near Infrared Spectroscopy and Its Calibration. **Critical Reviews in Analytical Chemistry**, v. 40, n. 4, p. 246-260, 29 out. 2010.

ANDERSEN, C. M.; BRO, R. Variable selection in regression-a tutorial. **Journal of Chemometrics**, v. 24, n. 11-12, p. 728-737, 10 nov. 2010.

ANVISA AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA. **Resolução RDC nº 360 de 23/12/2003.** Disponível em: <http://www.anvisa.gov.br/legis/resol/2003/rdc/360_03rdc.htm>. Acesso em: 15 nov. 2012.

ASCHERI, J. L. R.; NASCIMENTO, R. E.; SPEHAR, C. R. **Composição Química Comparativa de Farinha Instantânea de Quinoa, Arroz e Milho.** Rio de Janeiro, RJ.

ASTM. E1655-05. Standard Practices for Infrared Multivariate Quantitative Analysis. **Anais...** West Conshohocken, PA: 2011 Disponível em: <<http://www.astm.org>>

ATKINS, P. W.; PAULA, J. DE. **Atkins' Physical Chemistry.** 8th. ed. Oxford: W.H. Freeman & Company, 2006. p. 1072

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. **Applied Spectroscopy**, v. 43, n. 5, p. 772-777, 1 jul. 1989.

BEEBE, R. K.; KOWALSKI, B. R. An Introduction to Multivariate Calibration and Analysis. **Analytical Chemistry**, v. 59, n. 17, p. 1007-1017, 1987.

BEERMANN, C. *et al.* Anti-oxidative capacity of enzymatically released peptides from soybean protein isolate. **European Food Research and Technology**, v. 229, n. 4, p. 637-644, 2 jun. 2009.

BELLON, V.; VIGNEAU, J. L.; SÉVILA, F. Infrared and near-infrared technology for the food industry and agricultural uses: on-line applications. **Food Control**, v. 5, n. 1, p. 21-27, jan. 1994.

BJORSVIK, H.-R.; MARTENS, H. Calibration of NIR Instruments by PLS Regression. *In*: CIURCZAK, E. W.; BURNS, D. A. (Eds.). **Handbook of Near-Infrared Analysis.** 3. ed. New York, USA: CRC Press Taylor & Francis Group, 2007. p. 189-205.

BRAMBLE, T. *et al.* Single Kernel Protein Variance Structure in Commercial Wheat Fields in Western Kansas. **Crop Science**, v. 42, p. 1488-1492, 2002.

BROWN, C. D.; VEGA-MONTOTO, L.; WENTZELL, P. D. Derivative Preprocessing and Optimal Corrections for Baseline Drift in Multivariate Calibration. **Applied Spectroscopy**, v. 54, n. 7, p. 1055-1068, 2000.

BURNS, D. A.; CIURCZAK, E. W. (EDS.). **Handbook of Near-Infrared Analysis** . Third ed. Boca Raton: CRC Press, 2007.

Canadian Grain Commission. Disponível em: <www.grainscanada.gc.ca>. Acesso em: 10 jul. 2012.

CAPERUTO, L. C.; AMAYA-FARFAN, J.; CAMARGO, C. R. O. Performance of quinoa (*Chenopodium quinoa* Willd) flour in the manufacture of gluten-free spaghetti. **Journal of Science of Food and Agriculture**, v. 81, n. 1, p. 95-101, 2000.

CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. **Trends in Food Science & Technology**, v. 18, n. 2, p. 72-83, fev. 2007.

COBLENTZ, W. W. The diffuse reflecting power of various substances. **Journal of the Franklin Institute**, v. 174, n. 5, p. 549-582, nov. 1912.

CONAB - Companhia Nacional de Abastecimento. Disponível em: <<http://www.conab.gov.br>>. Acesso em: 8 maio. 2012.

DENDY, D. A. V. **Cereals and cereal products : chemistry and technology / David A.V. Dendy, Bogdan J. Dobraszczyk. - Version details - Trove**. Gaithersbrug, Maryland: Aspen Publishers, 2001. p. 429

EMBRAPA. **Empresa Brasileira de Pesquisa Agropecuária**. Disponível em: <<http://www.cnpso.embrapa.br/>>. Acesso em: 9 jun. 2012.

FAO FOOD AND AGRICULTURAL ORGANIZATION. **World Information and Early Warning System (WIEWS)**. Disponível em: <<http://www.fao.org/docrep/>>. Acesso em: 9 jun. 2012.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, p. 1-17, 1986.

GELADI, P.; MACDOUGALL, D.; MARTENS, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. **Applied Spectroscopy**, v. 39, n. 3, p. 491-500, 1 maio. 1985.

GEMPERLINE, P. J. Principal component analysis PCA. *In*: GEMPERLINE, P. J. (Ed.). **Practical Guide to Chemometrics**. Second ed. [S.I.] CRC Press, 2006. p. 69-104.

GROSS, R. *et al.* Chemical composition and protein quality of some local Andean food sources. **Food Chemistry**, v. 34, n. 1, p. 25-34, jan. 1989.

HAALAND, D. M.; THOMAS, E. V. Partial Least-Squares Methods for Spectral Analyses . 1 . Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. **Analytical chemistry**, v. 60, n. 11, p. 1193-1202, 1988.

HAGER, A.-S. *et al.* Nutritional properties and ultra-structure of commercial gluten free flours from different botanical sources compared to wheat flours. **Journal of Cereal Science**, v. 56, n. 2, p. 239-247, set. 2012.

HANSEN, L. M.; SNAIL, K. A. Integrating Spheres for Mid- and Near-infrared Reflection Spectroscopy Integrating Spheres for Mid- and Near-infrared Reflection Spectroscopy. *In*: CHALMERS, J. M. *et al.* (Eds.). **Handbook of Vibrational Spectroscopy**. New York, USA: Wiley & Sons, 2001. p. 1175-1191.

HART, J. R.; NORRIS, K. H.; GOLUBIC, C. Determination of the moisture content of seeds by near-infrared spectrophotometry of the methanol extracts. **Cereal Chemistry**, v. 39, p. 94-99, 1962.

HAUGHEY, S. A. *et al.* The application of Near-Infrared Reflectance Spectroscopy (NIRS) to detect melamine adulteration of soya bean meal. **Food Chemistry**, v. In Press, p. 1-5, jan. 2012.

HOURANT, P. *et al.* Oil and Fat Classification by Selected Bands of Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, n. 8, 2000.

HSU, C. P. S. Infrared Spectroscopy. *In*: SETTLE, F. A. (Ed.). **Handbook of Instrumental Techniques for Analytical Chemistry**. First ed. Michigan: Prentice Hall PTR, 1997. p. 257-283.

INNAWONG, B. *et al.* The determination of frying oil quality using Fourier transform infrared attenuated total reflectance. **LWT - Food Science and Technology**, v. 37, n. 1, p. 23-28, fev. 2004.

KALIVAS, J. H.; GEMPERLINE, P. J. Calibration. *In*: GEMPERLINE, P. (Ed.). **Practical Guide to Chemometrics**. Second ed. [S.l.] CRC Press Taylor & Francis Group, 2006. p. 105-166.

KANDALA, C. V; SUNDARAM, J.; PUPPALA, N. Analysis of Moisture Content , Total Oil and Fatty Acid Composition by NIR Reflectance Spectroscopy : A Review. **Lecture Notes in Electrical Engineering**, v. 146, p. 59-80, 2012.

KOBA, K.; HORROBIN, D. F.; SUGANO, M. Interrelationship between dietary protein, cholesterol and n-6 polyunsaturated fatty acid metabolism. **Progress in Lipid Research**, v. 32, n. 2, p. 123-137, jan. 1993.

KORTUM, G. **Reflectance Spectroscopy: Principles, Methods, Applications**. 1st. ed. New York, USA: Springer, 1996. p. 366

MARTENS, H.; NAES, T. **Multivariate Calibration**. first ed. London: John Wiley & Sons, 1992. p. 438

MUNCK, L. *et al.* A physiochemical theory on the applicability of soft mathematical models-experimentally interpreted. **Journal of Chemometrics**, v. 24, n. 7-8, p. 481-495, 8 jul. 2010.

NEPA/UNICAMP. **NEPA Núcleo de Estudos e Pesquisas em Alimentação**.

NIELSEN, S. S. **Food Analysis**. 4th. ed. New York, USA: Springer, 2010. p. 271

OSBORNE, B. Near-infrared spectroscopy in food analysis. *In*: MEYERS, R. A. (Ed.). **Encyclopedia of Analytical Chemistry**. Chichester: John Wiley & Sons Ltd, 2000. p. 1-14.

OSBORNE, B. G.; FEARN, T.; HINDLE, P. H. **Practical NIR spectroscopy with applications in food and beverage analysis**. 2nd. ed. New York, USA: Longman Scientific & Technical, 1993. p. 227

OZDEMIR, D. Genetic multivariate calibration for near infrared spectroscopic determination of protein, moisture, dry mass, hardness and other residues of wheat. **International Journal of Food Science and Technology**, v. 41, n. s2, p. 12-21, dez. 2006.

PASQUINI, C. Near Infrared Spectroscopy: Fundamentals , Practical Aspects and Analytical Applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198-219, 2003.

REID, L. M. *et al.* Differentiation of apple juice samples on the basis of heat treatment and variety using chemometric analysis of MIR and NIR data. **Food Research International**, v. 38, n. 10, p. 1109-1115, dez. 2005.

REYNS, P.; SPAEPEN, P.; BAERDEMAEKER, J. DE. Site-Specific Relationship Between Grain Quality and Yield. **Precision Agriculture**, v. 2, n. 3, p. 231-246, 2000.

RUEDELL, J. **Cultura da Soja: a verdade sobre a transgenia** Passo Fundo FUNDACEP, , 2003.

SALGÓ, A.; GERGELY, S. Analysis of wheat grain development using NIR spectroscopy. **Journal of Cereal Science**, v. 56, n. 1, p. 31-38, jul. 2012.

SCHLICK, G.; BUBENHEIM, D. L. Quinoa: Candidate Crop for NASA's Controlled Ecological Life Support Systems. *In*: JANICK, J. (Ed.). **Progress in new crops**. Arlington, VA: ASHS PRes, 1996. p. 632-640.

SHENK, J. S.; WORKMAN, J. J.; WESTERHAUS, M. O. Application of NIR Spectroscopy to Agricultural Product. *In*: CIURCZAK, E. W.; BURNS, D. A. (Eds.). **Handbook of Near-Infrared Analysis**. Third ed. [S.l.] CRC Press, 2007. p. 347-386.

SMITH, B. C. **Fundamentals of Fourier Transform Infrared Spectroscopy**. Second ed. New York, USA: CRC Press Taylor & Francis Group, 2011.

SORENSEN, L. K. Application of reflectance near infrared spectroscopy for bread analyses. **Food Chemistry**, v. 113, n. 4, p. 1318-1322, abr. 2009.

SPECTOR, A. A. Essentiality of fatty acids. **Lipids**, v. 34 Suppl, p. S1-3, jan. 1999.

SPEHAR, C. R. Adaptation of quinoa (*Chenopodium quinoa* Willd .) to increase the agricultural and alimentary diversity in Brazil. **Cadernos de Ciência & Tecnologia**, v. 23, n. 1, p. 41-62, 2006.

SPEHAR, C. R.; SANTOS, R. L. DE B. Quinoa BRS Piabiru : alternativa para diversificar os sistemas de produção de grãos. **Pesquisa Agropecuária Brasileira**, v. 37, n. 6, p. 889-893, 2002.

SUN, D.-W. **Infrared Spectroscopy for Food Quality Analysis and Control**. First ed. New York, USA: Elsevier, 2009. p. 424

VALENCIA-CHAMORRO, S. A. Quinoa. *In*: WRIGLEY, C.; CORKE, H.; WALKER, C. E. (Eds.). **Encyclopedia of Grain Science**. Elsevier L ed. [S.l: s.n.]. p. 1-8.

VIEIRA, C. R.; CABRAL, L. C.; PAULA, A. C. O. Composição centesimal e conteúdo de aminoácidos, ácidos graxos e minerais de seis cultivares de soja destinadas à alimentação humana 1. **Pesquisa Agropecuária Brasileira**, v. 34, n. 7, p. 1277-1283, 1999.

VLACHOS, N. *et al.* Applications of Fourier transform-infrared spectroscopy to edible oils. **Analytica Chimica Acta**, v. 573-574, p. 459-65, 28 jul. 2006.

VOORT, F. R. VAN DE; SEDMAN, J.; RUSSIN, T. Lipid analysis by vibrational spectroscopy. **European Journal of Lipid Science and Technology**, v. 103, n. 12, p. 815-826, 2001.

WANG, W.; MEJIA, E. G. DE. in Soy Bioactive Peptides that May Prevent. **Science**, v. 4, 2005.

WILLIAMS, P.; NORRIS, K. H. **Near-infrared Technology in the Agricultural and Food Industries**. 2nd. ed. University of Wisconsin - Madison: American Association of Cereal Chemists, 2001. p. 296

WORKMAN, J. J. J. Interpretive spectroscopy for near infrared. **Applied Spectroscopy Reviews**, v. 31, n. 3, p. 251-320, 1996.

CAPÍTULO II

COMPARAÇÃO E APLICAÇÃO DE ESPECTROSCOPIA NA REGIÃO DO INFRAVERMELHO PRÓXIMO (NIR) E MÉDIO (MIR) PARA DETERMINAÇÃO DE PARÂMETROS DE QUALIDADE EM AMOSTRAS DE SOJA

Manuscrito em preparação a ser submetido para publicação no periódico *Food Control*

RESUMO

A composição química de grãos está relacionada diretamente à sua qualidade. As análises químicas tradicionais são, em geral, laboriosas, consomem muito tempo e geram muitos resíduos químicos. Isto justifica o desenvolvimento de métodos alternativos que possibilitem um controle rápido e eficiente de produtos. As técnicas de espectroscopia na região do infravermelho próximo (NIR) e médio (MIR) associadas a ferramentas quimiométricas destacam-se por fornecer resultados rápidos e precisos com diversas aplicações em análises químicas de produtos agrícolas. O objetivo deste estudo foi desenvolver e comparar métodos para o estabelecimento dos parâmetros de qualidade utilizando espectroscopia na região do infravermelho próximo e médio, por reflectância difusa, associada a métodos de calibração multivariada. Para isto, foi avaliado o conteúdo de umidade, proteínas, lipídios e cinzas, de 20 variedades de grãos de soja, cultivadas em Ponta Grossa e Londrina, totalizando 40 amostras. Os dados espectrais foram processados por meio de PLS (*partial least squares*) e os modelos de calibração obtidos pelas técnicas de NIR e MIR foram comparados. Os coeficientes de determinação (R^2) entre os valores previstos pelo modelo e os valores analisados para umidade, lipídios, proteínas e cinzas foram: 0,72, 0,73, 0,88 e 0,81 para o NIR e 0,63, 0,87, 0,81 e 0,67 para MIR, respectivamente, com RMSECV (*root mean square error of cross validation*) < 2,09 % e RMSEP (*root mean square error of prediction*) < 1,32 % para ambas as técnicas. Os resultados comprovaram a capacidade preditiva de ambas as técnicas.

Palavras-chave: Espectroscopia no infravermelho; soja; *partial least square*; qualidade dos alimentos.

ABSTRACT

Grain composition is directly related to quality. The chemical composition has been determined using traditional and laborious methods, which are time-consuming and generate much chemical waste. This justifies the development of alternative methodologies capable of fast and accurate product control. Near-infrared (NIR) and mid-infrared spectroscopy (MIR) techniques associated to chemometrics tools have been applied to provide fast and accurate results with many applications in chemical analysis of agricultural products. The aim of this study is to develop and to compare methods to determine quality parameters, such as moisture, protein, lipid and ash content in 20 varieties of soybean, grown in Ponta Grossa and Londrina (Brazil), totalizing 40 samples. Near-infrared and mid-infrared spectroscopy, with diffuse reflectance, associated with multivariate calibration methods were used. The spectral data were analyzed by PLS (partial least squares) and the calibration models for NIR and MIR were compared. The determination coefficients (R^2) for moisture, protein, lipid and ash content were 0.72, 0.73, 0.88 and 0.81 for NIR and 0.63, 0.87, 0.81 e 0.67 for MIR, respectively, having a RMSECV (standard error of cross validation) < 2.09 % and RMSEP (root mean square error of cross prediction) < 1.32 % for both techniques. The results show that both infrared (NIR and MIR) techniques have predictive abilities.

Keywords: Infrared spectroscopy; soybean; partial least square; food quality

1 INTRODUÇÃO

A qualidade dos grãos é um importante parâmetro do cultivo e também do valor econômico destas commodities. Em muitos países, o preço do grão é determinado pelo conteúdo de proteína, amido e pela dureza. Existem algumas técnicas alternativas à aplicação das tradicionais para avaliar estes parâmetros de qualidade, tais como a composição centesimal através da tecnologia de infravermelho, que permitem avaliar diferentes constituintes em grãos, inclusive em sistemas “on-line” na própria plantação (Maertens, Reyns, & Baerdemaeker, 2004).

A espectroscopia no infravermelho é apontada como um meio rápido e confiável de investigação da qualidade e segurança dos alimentos (Downey, 1998). Vários estudos têm reportado a capacidade dessa técnica em caracterizar produtos em termos químicos, físicos e sensoriais, bem como em realizar classificações visando, por exemplo, estabelecer a origem de um determinado produto. As técnicas no infravermelho médio (MIR) e próximo (NIR) com transformada de Fourier têm sido aplicadas em processos de autenticação, com o intuito de avaliar se o produto está estritamente em conformidade com o rótulo descrito ou obedece à legislação vigente (Borin, Ferrão, Mello, Maretto, & Poppi, 2006; Hourant, Baeten, Orales, & Meurens, 2000; Reid, Woodcock, O'Donnell, Kelly, & Downey, 2005; Vlachos et al., 2006).

Desde o início da década de 1960, a tecnologia de NIR tem sido associada a análises de grãos e produtos derivados, sendo Karl Norris o pioneiro no desenvolvimento de métodos analíticos baseados em espectroscopia NIR. Sua equipe, do Departamento de Agricultura dos Estados Unidos, foi capaz de determinar o conteúdo de umidade de sementes usando espectros NIR associados à calibração multivariada (Hart, Norris, & Golumbic, 1962). Em 1975, a espectroscopia NIR foi adotada pela Comissão Canadense de Grãos como um método oficial de análise de proteínas em trigo (Williams, 1975). Em 1980, a

técnica foi adotada pelo Serviço de Inspeção Federal de Grãos dos Estados Unidos, e o teste de proteína de trigo por NIR foi finalmente aprovado pela *American Association of Cereal Chemists (AACC)* em 1989 (Osborne, 2000).

Diversos autores reportaram a possibilidade de medidas diretas de proteína, umidade, peso seco e amido de diversos grãos por NIR (Engel, Long, Carlson, & Reed, 1997; Maertens et al., 2004; Montes et al., 2006). Para isso, foram construídos modelos de calibração empregando-se os espectros NIR e os parâmetros de qualidade com o uso do algoritmo PLS, validados através de validação cruzada, com erro padrão da validação cruzada (SECV) em torno de 0,57 a 0,31 % para proteína e umidade em trigo, respectivamente (Maertens et al., 2004).

No campo de análise de óleos, MIR tem sido efetivamente usada para determinar o conteúdo de ácidos graxos *cis* e *trans*, índice de iodo, índice de saponificação, peróxido e ácidos graxos livres em amostras de óleos e gorduras (Innawong, Mallikarjunan, Irudayaraj, & Marcy, 2004; Voort, Sedman, & Russin, 2001). MIR também tem sido reportado para detectar adulteração de óleo de oliva com vários óleos vegetais (óleo de girassol, óleo de soja, óleo de gergelim e óleo de milho) e monitorar o processo de oxidação (Vlachos et al., 2006).

A espectroscopia MIR permite a identificação de compostos a partir de frequências vibracionais fundamentais, compreende a faixa espectral de 2500 a 50000 nm, enquanto os espectros NIR, faixa do espectro eletromagnético de 780 a 2500 nm, contêm informações de sobretons e combinações destas vibrações, sendo mais difíceis de interpretar (Reid et al., 2005). A interpretação destes espectros no infravermelho geralmente envolve a combinação de múltiplas amostras, onde cada uma possui um grande número de características correlacionadas. Assim, uma variedade de algoritmos deve ser aplicada para reduzir a complexidade dos dados, objetivando identificar as variáveis de interesse. A calibração multivariada é largamente aplicada para extrair a informação relevante do espectro obtido visando prever a composição química da

amostra. Entretanto, o principal problema associado a estes métodos é a não linearidade dos dados (Balabin, Safieva, & Lomakina, 2007).

Poucos estudos comparam as técnicas de NIR e MIR para aplicações específicas, incluindo discriminação de óleos comestíveis e gorduras (Yang, Irudayaraj, & Paradkar, 2005); quantificação de óleos essenciais (Schulz, Schrader, Quilitzsch, Pfeffer, & Krüger, 2003) e avaliação de diferenças químicas em vinhos tintos e brancos aplicando diferentes cepas de *Oenococcus oeni* para fermentação (Cozzolino, McCarthy, & Bartowsky, 2012). Tarr, Diepeveen e Appels (2012) aplicaram FTIR (MIR) e Raman como complemento das informações obtidas por NIR para monitorar o grão de cevada durante o processo de fabricação do malte. Os autores reportaram que através da construção de um modelo de calibração PLS foi possível monitorar parâmetros de qualidade para o malte por NIR. O FTIR foi utilizado para captação de imagens da superfície do grão de cevada na região entre 4000 a 700 cm^{-1} .

A proposta deste trabalho não foi apenas comparar as duas regiões espectrais para avaliar parâmetros de qualidade, mas também avaliar a capacidade preditiva das técnicas de espectroscopia no infravermelho NIR e MIR, em associação a ferramentas quimiométricas, para o conteúdo de umidade, proteínas, lipídios e cinzas, em soja.

2 MATERIAIS E MÉTODOS

2.1 Amostras

As seguintes variedades de soja foram estudadas: BRS 48, 184, 213, 232, 233, 257, 258, 259, 260, 261, 262, 267, 268, 282, 242RR, 244RR, 245RR, 246RR, 255RR e 256RR. Cada variedade foi cultivada simultaneamente em Londrina e Ponta Grossa, PR (Brasil) pela Embrapa Soja, totalizando 40 amostras, que foram secas em estufa a 40°C por 24h e avaliadas quanto aos teores de referência para

umidade, lipídios, cinzas e proteínas (AOAC, 2006) na Universidade Estadual de Londrina (UEL, PR, Brasil).

2.2 Espectroscopia na região do infravermelho próximo e médio

Os espectros das 40 amostras de farinha de soja foram obtidos na região do infravermelho médio e próximo por reflectância difusa. Foram utilizados os equipamentos FT-NIR Spectrum 100N da Perkin Elmer (Estados Unidos) e FT-IR MB-100 da Bomem (Quebec, Canadá), ambos tipo FT (transformada de Fourier). Operando de acordo com os parâmetros descritos na Tabela 1.

Tabela 1. Parâmetros instrumentais para obtenção dos espectros de farinha de soja no infravermelho próximo (NIR) e médio (MIR).

	<i>Unidade</i>	<i>MIR</i>	<i>NIR</i>
Faixa espectral	cm ⁻¹	400-4000	4000-8000
	nm	25000-2500	2500-1250
Resolução	cm ⁻¹	4	4
Número de scans (espectro e referência)		64	32
Referência		KBr	BaSO ₄

2.3 Análises Quimiométricas

A calibração foi realizada por regressão por mínimos quadrados parciais (PLS). Os resultados obtidos por NIR e MIR foram tratados pelo programa Matlab com as ferramentas PLS Toolbox v. 5.8 (Eigenvector, Research, Inc., EUA) desenvolvido para ambiente Matlab 7.8 (MathWorks, EUA).

Os espectros foram pré-processados pela combinação de vários tratamentos matemáticos: primeira e segunda derivadas, transformação padrão normal de variação (SNV – *standard normal variate transformation*) e todos os dados espectrais foram centrados na média (MC) antes da calibração. O número de variáveis latentes do PLS foi determinado por validação cruzada (*leave-one-out*) (Geladi & Kowalski, 1986).

Do total de 40 amostras de soja, 30 foram utilizadas para a construção do modelo de calibração e 10 foram utilizadas para a validação externa (conjunto de previsão); a escolha das amostras para a calibração e previsão foi realizada pelo algoritmo de Kennard–Stone (Kennard & Stone, 1969) do pacote PLS-Toolbox versão 5.8.

A qualidade do modelo foi aferida através da determinação da concentração de cada uma das 10 amostras retiradas, além dos melhores resultados para raiz quadrada do erro médio quadrático de calibração (RMSEC), raiz quadrada do erro médio quadrático da validação cruzada (RMSECV), raiz quadrada do erro médio quadrático de previsão (RMSEP) e coeficiente de determinação (R^2) do gráfico dos valores previstos pelo modelo *versus* os valores de referência (Williams & Norris, 2001).

Os *outliers* foram identificados e removidos quando necessário através da análise dos gráficos Q_{residual} *versus* T^2 e *leverage* (Naes, Isaksson, Fearn, & Davies, 2002).

3 RESULTADOS E DISCUSSÃO

Os espectros no infravermelho médio e próximo (MIR e NIR) da farinha de soja estão apresentados nas Figuras 1 e 2, sem tratamento matemático.

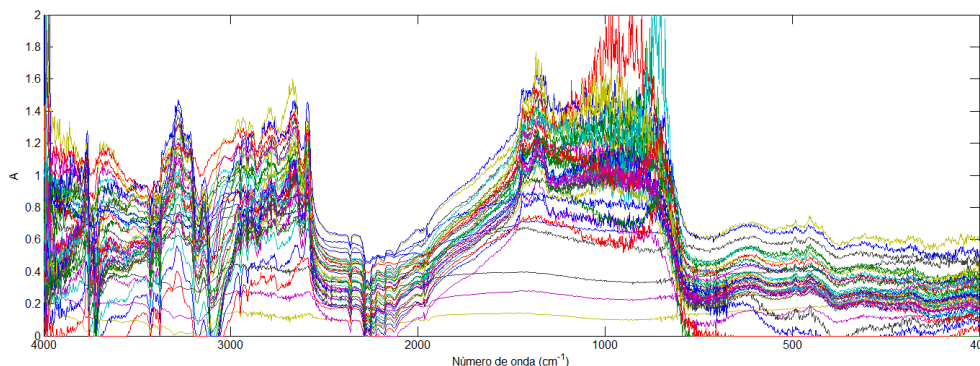


Figura 1. Espectros MIR para farinha de soja, obtidos na faixa entre 4000 a 400 cm^{-1} , sem pré-tratamento.

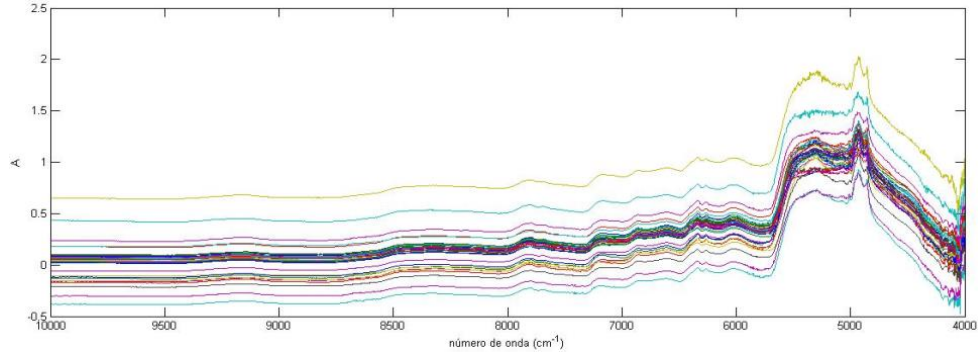


Figura 2. Espectros NIR para farinha de soja, obtidos na faixa entre 10000 a 4000 cm^{-1} , sem pré-tratamento.

As características dos espectros MIR e NIR para as amostras de soja (Figuras 1 e 2) representam grupos funcionais associados principalmente aos seus conteúdos de umidade, proteínas, lipídios e carboidratos. As absorções devido a transições energéticas vibracionais no NIR correspondem às transições vibracionais entre o nível de energia fundamental e os níveis de energia de ordem superior e a bandas de combinação. Nesta região, destaca-se a faixa compreendida entre 5200 e 5100 cm^{-1} o grupo funcional O-H (1º sobretom do modo de combinação), 5000 a 5600 cm^{-1} o grupo C-H (1º sobretom de CH_3 e $-\text{CH}=\text{CH}-$, de ácidos graxos), 4400-4033 cm^{-1} o grupo C-H e também 5000 a 4500 cm^{-1} para estiramento de N-H e C=O, características das proteínas. Já as absorções no MIR correspondem às transações vibracionais entre o nível de energia fundamental e o primeiro nível vibracional excitado, onde a faixa espectral 1200 a 900 cm^{-1} é conhecida como *fingerprint*, e pode ser usada para comparar, discriminar ou classificar amostras. Na região do MIR, identificam-se os grupos C-O, estiramento de C-C e C-O-H, além disso, destaca-se a região entre 3040 e 2850 cm^{-1} para os grupos C-H para estiramentos assimétricos e simétricos de CH_2 e CH_3 , caracterizando cadeias de ácidos graxos, e 1654 cm^{-1} para N-H estiramento vibracional, para as proteínas (Espinoza, Lucas, & Littlejohn, 1999; Hourant et al., 2000; Kays, Windham, & Barton, 1998; Liu, Cho, Sakuri, Miura, & Ozaki, 1994; Shenk, Workman, & Westerhaus, 2007).

Em quimiometria, as respostas espectrais e os valores de referência são tratados como matrizes. Na composição do vetor das variáveis dependentes (y) foram utilizados os valores de referência para teores de umidade, proteínas, lipídios e cinzas, obtidos através de métodos laboratoriais recomendados, conforme apresentados na Tabela 2. Desta forma, foi obtida uma matriz X composta pelas absorvâncias nos diferentes números de onda de cada amostra, sendo cada linha da matriz representada por uma amostra diferente e as colunas representam as respectivas absorvâncias para as amostras estudadas em cada número de onda (Beebe & Kowalski, 1987).

Tabela 2. Faixa de valores obtidos pelos métodos de referência para amostras de soja de diferentes regiões de cultivo

<i>Região de cultivo</i>	<i>Umidade (%)</i>	<i>Proteína (%)</i>	<i>Lipídios (%)</i>	<i>Cinzas (%)</i>
Londrina	6,70 - 8,50	32,91 – 39,41	18,16 – 23,86	5,65 – 6,55
Ponta Grossa	7,40 - 8,80	37,26 – 42,16	17,95 – 23,98	4,91 – 6,49

As técnicas de MIR e NIR permitiram o desenvolvimento de modelos de calibração para quantificação do conteúdo de umidade, proteínas, lipídios e cinzas, apresentados na Tabela 3. A validação cruzada foi aplicada em todos os modelos. A seleção do melhor número de variáveis latentes foi realizada através dos vetores *loading* e um número ótimo de variáveis latentes, de 5 a 9, foi escolhido para cada modelo sem incorporar muito ruído (Haaland & Thomas, 1988). O pré-tratamento SNV nos espectros MIR e NIR corrigiu os efeitos do espalhamento multiplicativo provocado pela distribuição do tamanho da partícula, comum na espectroscopia por reflectância, permitindo os melhores modelos de calibração (Rodriguez-Saona, Fry, & Calvey, 2000), enquanto a primeira (janela com 11 pontos, polinômio 2º grau, 1ª deriv.) e a segunda (janela com 11 pontos, polinômio 2º grau, 2ª deriv.) derivadas acentuaram as informações espectrais para as técnicas MIR e NIR, respectivamente. A segunda derivada aplicada para os espectros MIR não melhorou o modelo de calibração em função dos dados

conterem muito ruído (Figura 1), e este pré-tratamento aumenta ainda mais esta interferência.

De acordo com a Tabela 3, os erros de calibração, validação cruzada e previsão foram baixos e o coeficiente de determinação foi similar para NIR e MIR (R^2 de 0,63 a 0,91 para MIR e 0,72 a 0,88 para NIR).

Tabela 3. Resultados da análise multivariada (PLS) dos parâmetros de soja com técnicas espectroscópicas NIR e MIR.

	<i>Umidade</i>		<i>Cinzas</i>		<i>Proteínas</i>		<i>Lipídios</i>	
	NIR	MIR	NIR	MIR	NIR	MIR	NIR	MIR
R^2	0,72	0,63	0,73	0,87	0,88	0,91	0,88	0,67
RMSEC	0,12	0,13	0,11	0,06	0,33	0,39	0,38	0,14
RMSECV	0,52	0,47	0,44	0,41	1,76	1,74	2,09	1,44
RMSEP	0,35	0,41	0,36	0,25	1,32	1,06	1,08	1,87
VL	8	5	9	7	9	8	6	8
Pré-tratamento	2 ^a . deriv.- MC	1 ^a . deriv.- MC	MC	SNV- MC	SNV-MC	SNV- MC	2 ^a . deriv.- MC	SNV- MC

R^2 – coeficiente de determinação, RMSEC, RMSEP, RMSECV – raiz quadrada do erro médio quadrático de calibração, de previsão e de validação cruzada, VL – variáveis latentes, MC – dados centrados na média, SNV – “standard normal variate”.

Os gráficos de regressão (Figuras 3 e 4) para o conteúdo de proteínas, lipídios, cinzas e umidade mostraram boa correlação entre os valores estimados pela técnica de infravermelho e as análises de referência para ambas regiões espectrais MIR e NIR.

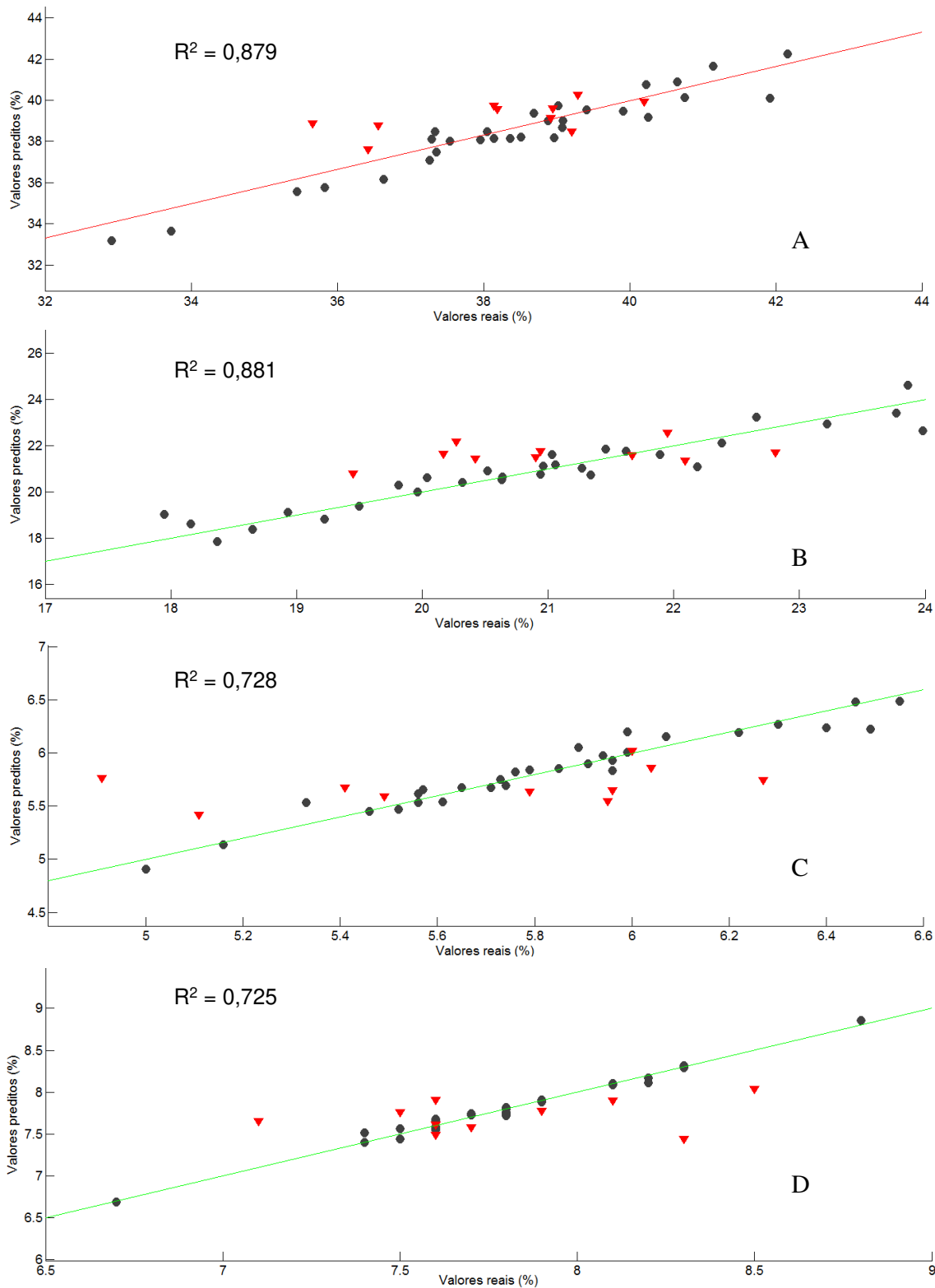


Figura 3. Relação entre os valores reais e os valores preditos pelos modelos obtidos por NIR para os teores de proteínas (A), lipídios (B), cinzas (C) e umidade (D). (▼) valores de predição, (●) valores de calibração.

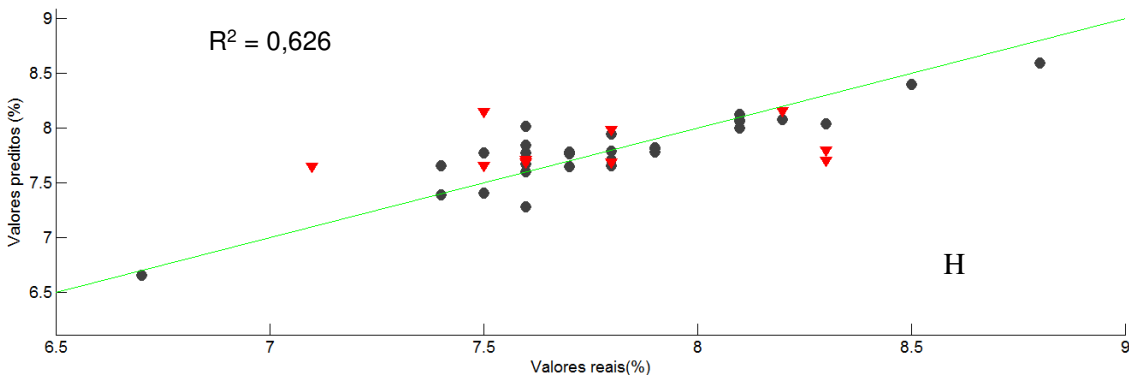
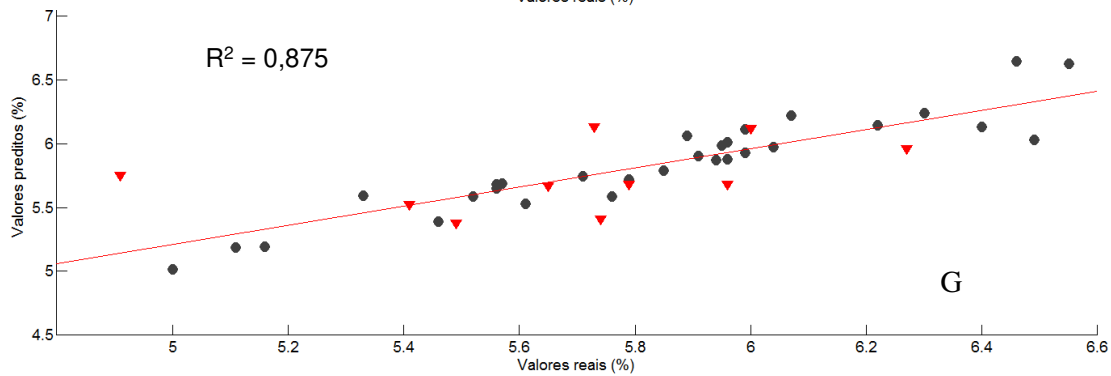
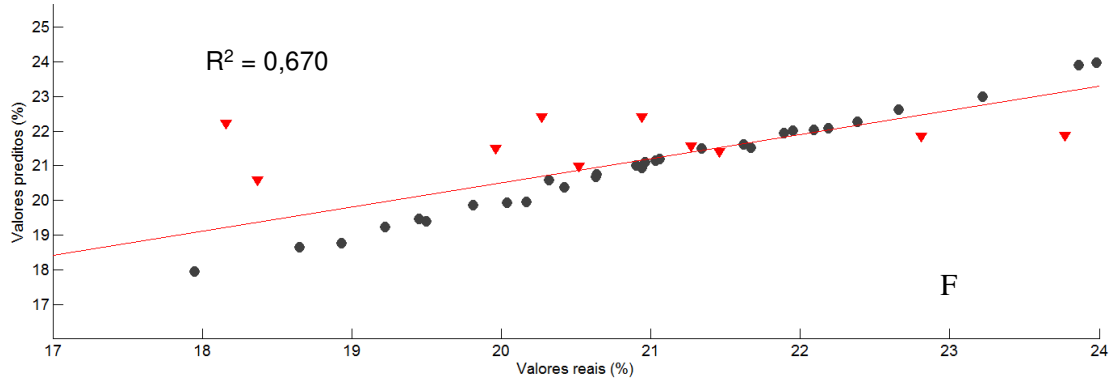
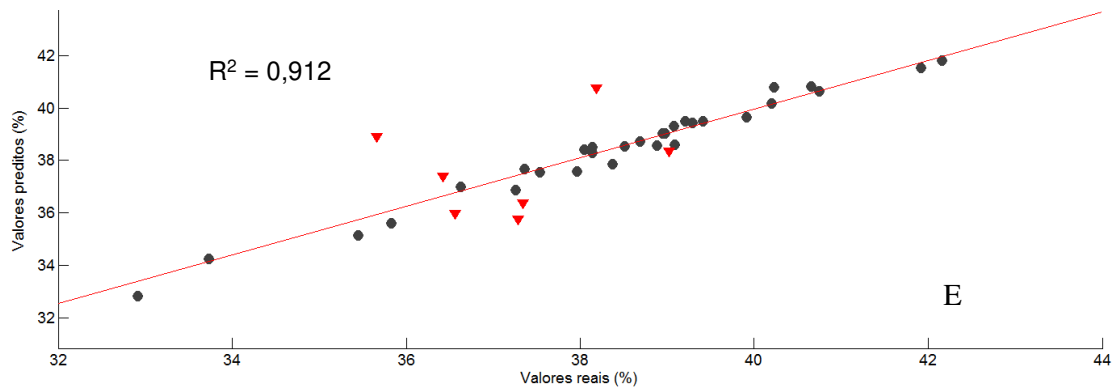


Figura 4. Relação entre os valores reais e os valores preditos pelos modelos obtidos por MIR para os teores de proteína (E), lipídio (F), cinza (G) e umidade (H). (▼) valores de predição, (●) valores de calibração.

Para a representação de uma boa relação entre os valores determinados experimentalmente e os valores previstos para o conjunto de 10 amostras de validação, foi aplicado o erro relativo, que consiste na relação, em porcentagem, entre o valor medido menos o previsto dividido pelo valor medido. Na Tabela 4, estão apresentados os valores previstos e reais para proteínas e lipídios e, na Tabela 5, para o conteúdo de cinzas e umidade. De um modo geral, o erro relativo foi baixo tanto para modelos gerados com o NIR como para o MIR, no entanto, os resultados sugerem o uso de NIR para previsão de lipídios (0,2 a 9,2 %) e o uso de MIR para proteínas (0,2 a 5,6 %), cinzas (0 a 5,0 %) e umidade (0,1 a 2,0 %).

Para o modelo de calibração com o MIR (Tabela 4) para teor de proteínas, alguns *outliers* foram retirados após análise do gráfico dos resíduos ($T^2 \times Q$ residual).

Tabela 4. Comparação entre o valor medido e o valor previsto pelo modelo de calibração aplicado para os espectros NIR e MIR para proteínas e lipídios.

Proteínas					Lipídios				
Valor medido	NIR	ER ^a (%)	MIR	ER ^a (%)	Valor medido	NIR	ER ^a (%)	MIR	ER ^a (%)
35,66	38,02	6,61	37,12	4,09	18,16	18,63	2,58	20,10	10,68
36,42	37,98	4,28	37,40	2,69	18,37	17,70	3,64	20,60	12,13
36,56	38,96	6,56	35,79	2,10	19,96	21,15	5,96	21,46	7,51
38,14	39,47	3,48	<i>b</i>	<i>b</i>	20,27	22,13	9,17	20,36	0,44
38,19	39,24	2,74	38,28	0,23	20,52	20,99	1,85	20,92	1,94
38,92	39,24	0,82	<i>b</i>	<i>b</i>	20,94	20,80	0,66	23,65	12,94
38,95	39,62	1,72	<i>b</i>	<i>b</i>	21,27	21,23	0,18	21,51	1,12
39,21	39,00	0,53	37,11	5,35	21,46	21,74	1,11	21,40	0,27
39,29	40,22	2,36	37,08	5,62	22,81	21,25	6,83	23,06	1,09
40,20	39,67	1,31	38,28	4,77	23,77	23,92	0,63	23,47	1,26

^a ER - Erro relativo: Calculado como a diferença entre o valor medido e o valor previsto versus o valor medido.

^b amostras retiradas para construção do modelo como *outliers*.

Tabela 5. Comparação entre o valor medido e o valor previsto pelo modelo de calibração aplicado para os espectros NIR e MIR para cinzas e umidade.

<i>Cinzas</i>					<i>Umidade</i>				
<i>Valor medido</i>	<i>NIR</i>	<i>ER^a (%)</i>	<i>MIR</i>	<i>ER^a (%)</i>	<i>Valor medido</i>	<i>NIR</i>	<i>ER^a (%)</i>	<i>MIR</i>	<i>ER^a (%)</i>
4,91	5,74	16,90	4,79	2,44	7,40	8,04	8,65	7,39	0,13
5,11	5,40	5,67	5,11	0	7,50	7,65	2,00	7,65	0,49
5,41	5,66	4,62	5,39	0,37	7,50	7,31	2,53	7,40	0,25
5,49	5,60	2,00	5,46	0,54	7,60	8,14	7,10	7,71	1,18
5,79	5,61	3,11	5,87	1,38	7,70	7,94	3,12	7,65	1,45
5,95	5,62	5,55	6,04	1,51	7,70	7,78	1,04	7,76	1,14
5,96	5,68	4,69	5,66	5,03	7,90	7,16	9,37	7,81	2,00
6,00	6,01	0,17	5,98	0,33	8,10	8,13	0,37	8,12	0,65
6,04	5,90	2,32	6,08	0,66	8,20	8,04	1,95	8,16	0,78
6,27	5,83	7,01	6,30	0,48	8,50	8,44	0,70	8,40	1,33

^a ER - Erro relativo: Calculado como a diferença entre o valor medido e o valor previsto versus o valor medido.

Diante dos resultados aqui apresentados, verificou-se que tanto a técnica de NIR, quanto de MIR, podem ser aplicadas para as amostras de soja brasileira. Guo, You, Prisecaru, Costescu e Nelson (2000) também concluíram que a técnica de NIR é rápida, precisa e de baixo custo para análise da composição de proteínas, lipídios, umidade, carboidratos e fibras em soja (EUA) e seus produtos derivados (leite e tofu), com $R^2 \cong 0,99$ e $RMSE < 1,0 \%$. Já Delwiche et al. (2007) não comprovaram a viabilidade da técnica de NIR para quantificação de proteínas e frações da mesma em grãos inteiros de soja do Tennessee (EUA), e obtiveram R^2 de 0,02 a 0,73 e $RMSEP$ de 0,15 a 11,9 %. Assim como este estudo, Balabin e Smirnov (2011) também comprovaram que as duas técnicas foram rápidas, robustas e de baixo custo, e detectaram melamina, um adulterante aplicado em produtos lácteos para elevar o teor de nitrogênio e superestimar o teor proteico do alimento. Porém, os autores analisaram amostras de leite, fórmulas infantis e leite em pó por MIR e NIR e relacionaram as duas técnicas como não lineares. Haughey, Graham, Cancouët e Elliott (2012) também detectaram melamina após aplicação como adulterante em ração animal elaborada com soja, e desenvolveram modelos de calibração para NIR com coeficientes de determinação (R^2) de 0,89 a 0,99. Não foram encontrados estudos que relacionam estas duas técnicas para amostras de soja brasileira.

4 CONCLUSÕES

Tanto a técnica de MIR como NIR geraram modelos PLS compatíveis e com capacidades preditivas. O tempo total requerido para a preparação e análise das amostras foi menos de 5 minutos, comparado a 10 a 16 horas requeridas para determinação do conteúdo de umidade, proteínas, lipídios e cinzas pelos métodos de referência. Diante dos resultados, ambas as técnicas de MIR e NIR podem ser aplicadas efetivamente na indústria ou beneficiadora de soja como uma ferramenta analítica para monitorar a qualidade da farinha de soja e simultaneamente avaliar a composição química.

Agradecimentos

Este trabalho foi financiado pela Fapesp (projeto nº 2010/50418-3) e contou com o auxílio do prof. Dr. Olívio F. Galão (UEL, PR, Brasil).

Referências bibliográficas

- AOAC. (2006). Official methods of analysis of AOAC International. In W. Horwitz & G. W. Latimer (Eds.), *AOAC International* (18th ed., Vol. 18th Ed). Gaithersbrug, Maryland.
- Balabin, R. M., Safieva, R. Z., & Lomakina, E. I. (2007). Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. *Chemometrics and Intelligent Laboratory Systems*, *88*(2), 183–188.
- Balabin, R. M., & Smirnov, S. V. (2011). Melamine detection by mid- and near-infrared (MIR/NIR) spectroscopy: a quick and sensitive method for dairy products analysis including liquid milk, infant formula, and milk powder. *Talanta*, *85*(1), 562–8.
- Beebe, R. K., & Kowalski, B. R. (1987). An Introduction to Multivariate Calibration and Analysis. *Analytical Chemistry*, *59*(17), 1007–1017.
- Borin, A., Ferrão, M. F., Mello, C., Maretto, D. A., & Poppi, R. J. (2006). Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk. *Analytica chimica acta*, *579*(1), 25–32.

- Cozzolino, D., McCarthy, J., & Bartowsky, E. (2012). Comparison of near infrared and mid infrared spectroscopy to discriminate between wines produced by different *Oenococcus Oeni* strains after malolactic fermentation: A feasibility study. *Food Control*, 26(1), 81–87.
- Delwiche, S. R., Pordesimo, L. O., Panthee, D. R., & Pantalone, V. R. (2007). Assessing Glycinin (11S) and β -Conglycinin (7S) Fractions of Soybean Storage Protein by Near-Infrared Spectroscopy. *Journal of the American Oil Chemists' Society*, 84(12), 1107–1115.
- Downey, G. (1998). Food and food ingredient authentication by mid-infrared spectroscopy and chemometrics. *TrAC Trends in Analytical Chemistry*, 17(7), 418–424.
- Engel, R., Long, D., Carlson, G., & Reed, J. F. (1997). On-the-go grain protein sensing is near. Does it have a future in precision nitrogen. Management for wheat? *Better Crops With Plant Food*, 81(4), 20–24.
- Espinoza, L. H., Lucas, D., & Littlejohn, D. (1999). Total Organic Carbon Content in Aqueous Samples Determined by Near-IR Spectroscopy. *Applied Spectroscopy*, 53(1), 103–107.
- Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185(1), 1–17.
- Guo, J., You, T., Prisecaru, V., Costescu, D., & Nelson, R. L. (2000). NIR Calibrations for Soybean Seeds and Soy Food Composition Analysis: Total Carbohydrates, Oil, Proteins and Water Contents. *Proteins*, (Aacc 1995).
- Haaland, D. M., & Thomas, E. V. (1988). Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Analytical chemistry*, 60(11), 1193–1202.
- Hart, J. R., Norris, K. H., & Golumbic, C. (1962). Determination of the moisture content of seeds by near-infrared spectrophotometry of the methanol extracts. *Cereal Chemistry*, 39, 94–99.
- Haughey, S. a., Graham, S. F., Cancouët, E., & Elliott, C. T. (2012). The application of Near-Infrared Reflectance Spectroscopy (NIRS) to detect melamine adulteration of soya bean meal. *Food Chemistry, In Press*, 1–5.
- Hourant, P., Baeten, V., Orales, M. T. M., & Meurens, M. A. R. C. (2000). Oil and Fat Classification by Selected Bands of Near-Infrared Spectroscopy. *Applied Spectroscopy*, 54(8).
- Innawong, B., Mallikarjunan, P., Irudayaraj, J., & Marcy, J. E. (2004). The determination of frying oil quality using Fourier transform infrared attenuated total reflectance. *LWT - Food Science and Technology*, 37(1), 23–28.

- Kays, S. E., Windham, W. R., & Barton, F. E. (1998). Prediction of Total Dietary Fiber by Near-Infrared Reflectance Spectroscopy in High-Fat- and High-Sugar-Containing Cereal Products. *Methods*, 8561(1997), 854–861.
- Kennard, R. W., & Stone, L. A. (1969). Computer Aided Design of Experiments. *Technometrics*, 11(1), 137–148.
- Liu, Y., Cho, R., Sakuri, K., Miura, T., & Ozaki, Y. (1994). Studies on Spectra/Structure Correlations in Near-Infrared Spectra of Proteins and Polypeptides. Part I: A Marker Band for Hydrogen Bonds. *Applied Spectroscopy*, 48(10), 1249–1254.
- Maertens, K., Reyns, P., & Baerdemaeker, J. De. (2004). On-line measurement of grain quality with NIR technology. *American Society of Agricultural Engineers*, 47(4), 1135–1140.
- Montes, J. M., Utz, H. F., Schipprack, W., Kusterer, B., Muminovic, J., Paul, C., & Melchinger, a. E. (2006). Near-infrared spectroscopy on combine harvesters to measure maize grain dry matter content and quality parameters. *Plant Breeding*, 125(6), 591–595.
- Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). Outlier detection. *A User-Friendly Guide to Multivariate Calibration and Classification* (pp. 177–189). Chichester, UK: NIR Publications.
- Osborne, B. (2000). Near-infrared spectroscopy in food analysis. In R. A. Meyers (Ed.), *Encyclopedia of Analytical Chemistry* (pp. 1–14). Chichester: John Wiley & Sons Ltd.
- Reid, L. M., Woodcock, T., O'Donnell, C. P., Kelly, J. D., & Downey, G. (2005). Differentiation of apple juice samples on the basis of heat treatment and variety using chemometric analysis of MIR and NIR data. *Food Research International*, 38(10), 1109–1115.
- Rodriguez-Saona, L. E., Fry, F. S., & Calvey, E. M. (2000). Use of fourier transform near-infrared reflectance spectroscopy for rapid quantification of castor bean meal in a selection of flour-based products. *Journal of agricultural and food chemistry*, 48(11), 5169–77.
- Schulz, H., Schrader, B., Quilitzsch, R., Pfeffer, S., & Krüger, H. (2003). Rapid classification of basil chemotypes by various vibrational spectroscopy methods. *Journal of agricultural and food chemistry*, 51(9), 2475–81.
- Shenk, J. S., Workman, J. J., & Westerhaus, M. O. (2007). Application of NIR Spectroscopy to Agricultural Product. In E. W. Ciurczak & D. A. Burns (Eds.), *Handbook of Near-Infrared Analysis* (Third., pp. 347–386). CRC Press.
- Tarr, a., Diepeveen, D., & Appels, R. (2012). Spectroscopic and chemical fingerprints in malted barley. *Journal of Cereal Science*, 56(2), 268–275.

- Vlachos, N., Skopelitis, Y., Psaroudaki, M., Konstantinidou, V., Chatzilazarou, A., & Tegou, E. (2006). Applications of Fourier transform-infrared spectroscopy to edible oils. *Analytica Chimica Acta*, 573-574, 459–65.
- Voort, F. R. Van De, Sedman, J., & Russin, T. (2001). Lipid analysis by vibrational spectroscopy. *European Journal of Lipid Science and Technology*, 103(12), 815–826.
- Williams, P. C. (1975). Application of near infrared reflectance spectroscopy to analysis of cereal grains and oilseeds. *Cereal Chemistry*, 52(4), 561–576.
- Williams, P., & Norris, K. H. (2001). *Near-infrared Technology in the Agricultural and Food Industries* (2nd ed., p. 296). University of Wisconsin - Madison: American Association of Cereal Chemists.
- Yang, H., Irudayaraj, J., & Paradkar, M. (2005). Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. *Food Chemistry*, 93(1), 25–32.

CAPÍTULO III

FOURIER TRANSFORM NEAR-INFRARED SPECTROSCOPY (FT-NIRS) APPLICATION TO ESTIMATE BRAZILIAN SOYBEAN [GLYCINE MAX (L.) MERRIL] COMPOSITION

Manuscript published in Food Research International:

Daniela Souza Ferreira, Juliana Azevedo Lima Pallone, Ronei Jesus Poppi, Fourier transform near-infrared spectroscopy (FT-NIRS) application to estimate Brazilian soybean [Glycine max (L.) Merrill] composition, Food Research International, Volume 51, Issue 1, April 2013, Pages 53-58, ISSN 0963-9969, 10.1016/j.foodres.2012.09.015.

(<http://www.sciencedirect.com/science/article/pii/S0963996912003742>)

Keywords: Multivariate data analysis; Composition analysis; Near-infrared reflectance spectroscopy; Soybean

Abstract

This study examined the ability of near-infrared reflectance spectroscopy method (FT-NIRS) and multivariate calibration to estimate the concentration of moisture, protein, lipid, ash and carbohydrate of Brazilian soybean. The spectra obtained in a range of 4000 to 10,000 cm^{-1} were preprocessed by several combinations of mathematical treatments: MSC (multiplicative scatter correction), SNV (standard normal variate) or first and second derivative and all data were mean centered before the calibration, for which was used the PLS method (partial least squares). The best calibration models found in this study were the ones used to determine protein and moisture contents ($R^2 = 0.81$, RMSEP = 1.61% and $R^2 = 0.80$, RMSEC = 1.55%, respectively). However, the technique shows high predictability for all parameters, including lipids, ashes and carbohydrates, with RMSECV of 0.40 to 2.30% and RMSEP of 0.38 to 3.71%. This result shows the viability of using NIR in controlling the quality parameters of soybeans.

Keywords: Multivariate data analysis, Composition analysis, Near-infrared reflectance spectroscopy, soybean.

1. INTRODUCTION

Fourier transform near-infrared spectroscopy (FT-NIRS) is a non-destructive fast technique capable of analyzing organic substances in a reliable way. Such analytical method has presented a great impact in the food industry and agriculture. According to Williams (2001) work, a huge number of analytical methodologies based on near-infrared spectroscopy have already demonstrated the ability of such technique in dealing safely and quickly with so diverse problems as determining proteins, moisture, carbohydrates and fat content in different types of food.

The centesimal food content analysis aims to guarantee the product quality, to follow official supervision and regulation committee requirements, as well as, to contribute to the consumer health. These chemical analysis, which includes the determination of moisture, protein, ash, fat and carbohydrate contents, traditionally has been performed by laborious methods which are time demanding and generate much chemical waste. However, the methods that use the FT-NIRS technique provide precise and fast results, with minimal or null sample preparation and waste generation, being able to be an alternative to such determinations (Kandala, Sundaram, & Puppala, 2012; Salgó & Gergely, 2012; Zhou, Yang, Huang, & Han, 2012).

Nevertheless, due to spectral features in the NIR region and in order to simultaneously evaluate diverse parameters, multivariate calibration methods must be used to obtain quantitative information. Multivariate calibration methods are part of chemometrics, which needs a huge number of results from analytical trials for the development of a mathematical model. Furthermore, once a model is built, it has to be validated by a group of data that were not part of the calibration (Agelet & Hurburgh, 2010; Jimaré Benito, Bosch Ojeda, & Sanchez Rojas, 2008). Toward this Sun, Han, Yan, Yang and Sato (2008) estimated the fatty acid concentrations in one hundred and eight Chinese soybean varieties.

The FT-NIRS technique and the multivariate calibration have already been used to determine some major contents in food, like Haughey et al.(in press) who described the evaluation of melamine content in soybean, Kuligowski, Carrión, Quintás, Garrigues, and de la Guardia (2012) who evaluated the NIR method to determination of polymerized triacylglyceride in vegetable oils and Szigedi, Lénárt, Dernovics, Turza and Fodor (2011) who determined protein in different vegetables. Zhu et al. (2011) also measured soybean seeds, nevertheless by near-infrared hyperspectral imaging system, for the determination of protein and fat contents of samples.

The wide range of results of the developed models, demonstrates that with only one spectroscopic technique, the FT-NIRS, it is possible to simultaneously analyze many parameters: moisture, protein, ash, lipid and carbohydrate. While Munck et al. (2010) analyzed starch, protein and lipid contents of barley varieties by NIR only, while an application of NMR, FT-Raman, FT-IR and NIR was necessary for Holve, Larsen, Hansen, and Engelsen (2011) in order to characterize carbohydrates, proteins, dietary fibers and unsaturated fatty acids in leguminous plants from Africa.

Brazil is the second country in world with the largest production of soybean (CONAB, 2012), being it a very important food in Brazil and an inexpensive source of oil and protein. According to the data of the Brazilian Agricultural Research Corporation (EMBRAPA, 2012), many varieties are produced and the evaluation of the quality of the beans depend on the centesimal content. Hence the aim of this work is to offer a fast and efficient method to predict the chemical characteristics of soybean with the application of FT-NIRS and multivariate calibration to several Brazilian varieties.

2. MATERIALS AND METHODS

2.1. Samples

A total of 100 soybean [*Glycine max* (L.) Merrill] varieties from the south and central parts of Brazil were donated by EMBRAPA Soybean. The landmark analyses were taken between 2009 and 2010.

2.2. Sample Preparation

The 250 g received from each bean were crushed in a grinding mill (A11 model, IKA, Germany) and the particle gradation standardized to 0.6 mm in a 28 mesh strainer. While a small portion (about 1 g) of such sample was immediately analyzed by near-infrared spectroscopy by diffuse reflectance, another portion was used for chemical analyses by standard methods.

2.3. Landmark Analysis

The total content of moisture, protein and ash were determined by 925.09, 979.09, 923.03 reference methods, respectively, as described by AOAC (2006). The total lipid content was determined by Bligh and Dyer (1959) and the total carbohydrate content was calculated by the formula ($\% \text{ carbohydrate} = 100 - \% \text{ moisture} - \% \text{ lipid} - \% \text{ protein} - \% \text{ ash}$). All the analyses were carried out in triplicate and right after the grinding in order to avoid any sort of alteration with the environment.

2.4. Near-infrared Analysis

The ground beans were directly analyzed by diffuse reflectance, using a NIR Infrared Reflectance accessory - NIRA (Perkin Elmer, United States), and a Spectrum 100N equipment (Perkin Elmer, United States), with a source of quartz halogen bulb. The software used for NIRA is Spectrum – V. 6.3.1.0132 (Perkin

Elmer), the region measured between 10,000 to 4000 cm^{-1} with 4 cm^{-1} of resolution and 64 scans and the analysis was run in duplicate.

2.5. *Chemometrics Analysis*

2.5.1. *Data modeling*

The multivariate calibration model was carried out by partial least squares (PLS). The NIR spectra were treated by the PLS Toolbox v. 5.8 (Eigenvector Research Inc., United States) program developed for Matlab 7.8 (MathWorks, United States).

2.5.2. *Sample choice for calibration and prediction*

From the total of 100 soybean samples, 70 were used for the construction of the calibration model and 30 were used for the external validation (prediction set). The sample choice for the calibration and prediction was carried out by the Kennard-Stone algorithm (Kennard & Stone, 1969). After defining the number of samples to calibrate, this algorithm initiates the selection of two spectra: the closest one to the central area and the other with the longer Euclidian distance in relation to the first. Afterward, the other chosen sample is the one with the longest distance in relation to this last one and thus successively until the chosen number of samples to calibrate is reached. The rest of the samples are the prediction set. Finally, the quality of the model was measured by determining the chemical composition of each one of the 30 samples.

2.5.3. *Pretreatment of spectra*

The spectra in the infrared region usually present baseline problems due to light scattering from the non-homogeneity of the matrix particle distribution (Bjorsvik & Martens, 2007). In order to minimize the effect of the change in the

baseline, the spectra were preprocessed by the combination of many mathematical treatments: MSC (multiplicative scatter correction) (Geladi, MacDougall, & Martens, 1985), SNV (standard normal variate transformation) (Barnes, Dhanoa, & Lister, 1989) or first and second derivative using Savitzky-Golay method (Brown, Vega-Montoto, & Wentzell, 2000), and all spectra data were mean centered before calibration. The number of PLS factors was determined by cross-validation (leave-one-out) procedure (Geladi & Kowalski, 1986).

2.5.4. Model evaluation

The performance of the final PLS model was evaluated according to: root mean square error calibration (RMSEC), root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP), coefficient of determination (R^2) of the plot between the predicted values by FT-NIRS and the reference values, and bias (Eq. (1)), which consists of the difference between the mean value predicted by FT-NIRS and the mean value of the reference model and residual predictive deviation (RPD) (Williams, 2001; Williams & Sobering, 1995; Windham, Mertens, & Barton, 1989). High results of R^2 and low RMSE indicate the best model to predict the chemical content.

$$bias \equiv \frac{1}{n} \sum_{i=1}^n (x_i - y_i) \quad (1)$$

where n is the number of samples of the calibration, x_i is the result obtained by FT-NIRS and y_i is the result obtained by the reference method of the sample i .

In order to evaluate the RMSECV, a spectrum of a sample is removed from the calibration set and the PLS model is constructed with the remaining spectra. The removed sample was predicted with such model and the process was repeated with the removal of each sample. The RMSECV was then calculated as Eq. (2).

While the number of the PLS factors included in the model was chosen according to the lowest value of RMSECV, the RMSEP, which expresses the degree of accordance between the estimated values by a model previously constructed and the value considered real or of reference, was also calculated as in Eq. (2).

RMSEC, which provides information about the adjustment of the model to the calibration data, is calculated according to Eq. (2).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (2)$$

where n is the number of samples (of the calibration set or prediction set), y_i is the result of the reference to the sample i and x_i is the estimated result by the NIR model.

Furthermore, the RPD represents the division between the reference data standard deviation and the mean error of the prediction (Williams & Sobering, 1995).

$$RPD = \frac{DP}{RMSEP} \quad (3)$$

The coefficient of determination between the predicted value by NIR and the reference value was calculated for the prediction sets.

The outliers were identified and removed when necessary by $Q_{\text{residual}} \times T^2$ graphics and of leverage. Hotelling's T^2 means the square of the distance of the sample from the model mean that can be explained by normal variations within the group. The Q residuals mean the distance of the sample from the model that can be explained by random error (Naes, Isaksson, Fearn, & Davies, 2002).

3. RESULTS AND DISCUSSION

3.1. *Composition analysis in soybean*

The results obtained in the reference analyses of the moisture, protein, ash and carbohydrate contents of the samples that are used to predict and validate the model are presented in Table 1. By analyzing the values of the mean and standard deviation, the huge variability among the samples, especially considering lipid, protein and carbohydrate contents, is observed.

Table 1. Descriptive statistics of soybean chemical parameters used in the calibration and validation.

<i>Parameter</i>	<i>Calibration</i>					<i>Validation</i>					<i>SE ±</i>
	<i>Min. (%)</i>	<i>Max. (%)</i>	<i>Mean (%)</i>	<i>SD ±</i>	<i>No.</i>	<i>Min. (%)</i>	<i>Max. (%)</i>	<i>Mean (%)</i>	<i>SD ±</i>	<i>No.</i>	
Moisture	8.16	18.10	13.70	2.14	70	10.40	16.56	13.61	1.60	30	0.086
Ashes	4.32	6.14	5.11	0.36	70	4.44	5.69	5.04	0.32	30	0.069
Lipids	12.55	26.96	21.91	2.20	70	12.93	24.27	21.61	1.94	30	0.162
Proteins	31.52	43.48	38.91	2.34	70	34.12	43.46	39.32	2.29	30	0.274
Carbohydrates	13.34	26.96	20.51	3.06	69	14.04	27.50	20.02	3.12	29	-

Min. – minimum, Max. – maximum, SD – standard deviation, SE- standard error of analysis, No. – number of samples

The variability of the results, represented by the standard deviation, is placed in the prediction errors and the cross-validated results.

3.2. *Development of calibration models and validation*

In Fig. 1 all the spectra are presented, without any pretreatment, in the region of the near-infrared, obtained from the 100 soybean samples, in the range of 10,000 to 4000 cm^{-1} .

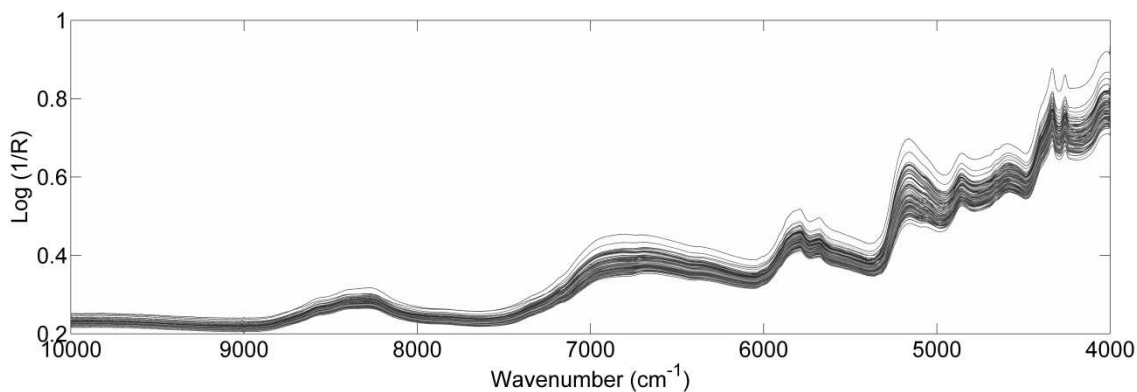


Fig. 1. NIR spectra of soybean samples obtained in the range of 10000 to 4000 cm^{-1} .

Among the pretreatments applied before calibration, a procedure that presented a good performance in the elimination of the multiplicative light scattering effect was the first derivative. Furthermore, such data were mean centered in order to reduce matrix effects from the samples on the model.

Table 2 shows the wavenumber regions and final pretreatments used for each model. Examination of the loading vector indicated which bands of the spectrum were associated with the relevant chemical information (Haaland & Thomas, 1988).

Table 2. Wavenumber region and treatments used in calibration and validation.

<i>Parameter</i>	<i>Wavenumber region</i>	<i>Derivative treatment^a</i>	<i>Scatter correction</i>	<i>N^o of outliers^b</i>
Moisture	10000-3500	X,11,2,1		
Ashes	10000-4900	X,51,3,2		
Lipids	10000-4000		SNV	
Proteins	7500-4900	X,51,3,2		
Carbohydrates	7500-4500	X,15,2,1		2

^a Savitzky-Golay employs the terms (*y*, *width*, *order*, *deriv*): *y*, the matrix of row vectors to be smoothed; *width*, number of points in the filter; *order*, order of the polynomial; and *deriv*, derivative.

^b Number of samples eliminated as outliers during calibration

The results used to evaluate the efficiency of predicting moisture, lipid, protein, ash and carbohydrate contents of the Brazilian soybean samples are presented in Table 3. The best models to predict the soybean chemical composition were chosen based on the lowest results of RMSECV and the highest results of R^2 (Windham et al., 1989).

Table 3. Chemometrics results of PLS models by FT-NIRS of the evaluated parameters of soybeans.

<i>Parameters</i>	<i>LV</i>	<i>R²</i>	<i>RMSEC (%)</i>	<i>RMSECV (%)</i>	<i>RMSEP (%)</i>	<i>bias</i>	<i>RPD</i>
Moisture	5	0.80	0.28	2.30	1.55	-0.049	1.38
Ash	5	0.63	0.07	0.40	0.38	-0.080	0.95
Lipids	6	0.71	1.13	1.51	1.20	-0.023	1.83
Proteins	6	0.81	0.58	1.14	1.61	-0.020	1.45
Carbohydrates	8	0.50	1.11	1.62	3.71	-0.196	0.83

LV – latent variables, R^2 – determination coefficient, RMSEC, RMSEP, RMSECV – root mean square error of the calibration, prediction and cross-validation, respectively, bias – systematic calibration error (the mean difference between the concentration predicted by NIR and the reference data), RPD (residual predictive deviation = standard deviation of the reference data/root mean square error of the prediction).

All the models presented good correlation between the reference values and the NIR predicted ones. All the regression errors were low, demonstrating the capacity to predict the soybean chemical characteristics by FT-NIRS. Considering the expectations, the errors of the group of prediction (RMSEP) were higher than the errors of the calibration group (RMSEC).

Low calibration and prediction errors of the moisture and ash contents (RMSEC 0.28 and 0.07%, RMSEP 1.55 and 0.28%, respectively) were obtained due to the few variables to describe the models (5 LV for each one). However, the determination coefficient was better to moisture content when compared to ash content (R^2 0.80 and 0.63) due to the difficulty in predicting the compound present in lower content and inorganic substance.

The wide dispersion of the results for the carbohydrate content in the reference analysis (Table 1, DP = 3.07) had as a consequence a lower correlation (Table 2, $R^2 = 0.50$) between the reference value and the predicted one by the model, indicating a lower accuracy to the prediction of such result, even then higher the latent variables numbers (Table 2, 8 LV). This behavior can be explained by the difficulty to obtain a model capable of appropriately describing the value range to the carbohydrate content in the matrix.

The calibration model obtained in this study presented high R^2 (0.81) and low prediction error (RMSEP = 1.61%) for the protein content with 6 latent variables, which is in conformity with other references that evaluated proteins of vegetables and soybeans (Kovalenko, Rippke, & Hurburgh, 2006; Szigedi et al., 2011).

The prediction errors (RMSEP) and bias obtained in this study were low and presented the same value of the results presented in the study of bread samples by Sorensen (2009), who also obtained a higher RMSEP for carbohydrate content in relation to other errors.

According to Williams (2001) a value of RPD > 2.4 is desirable to an appropriate model, while Williams & Sobering (1995) indicate that the value of 3 or more is the recommended value. This result depends on the errors of the prediction (RMSEP) which must be lower than the standard deviation of the data group. Therefore, the result of the RPD used to determine lipid content (RPD 1.83) was better, once the other results were low (RPD from 0.83 to 1.45), especially of carbohydrate content which presented higher RMSEP than the standard deviation (RMSEP = 3.71%, DP = 3.08). However, evaluating only RPD does not determine the predictable capacity of the models, since the other RMSEPs were low when compared to the DP, indicating low prediction errors. Thus, a wide range of results were chosen in group in order to evaluate the predictable ability of FT-NIRS.

The mean difference between the values predicted by the model and the reference values (bias) were not representative, the other parameters being higher

for carbohydrates only (bias = -0.196). Despite such good result, the proximity between the reference values and the ones predicted by the model was only noted due to the analysis in the group with the other results of the model.

The ability of the PLS models to predict soybean chemical contents was also evaluated by the plot that relates the value measured by the reference method and the one predicted by FT-NIRS, presented in Figs. 2 to Fig. 6.

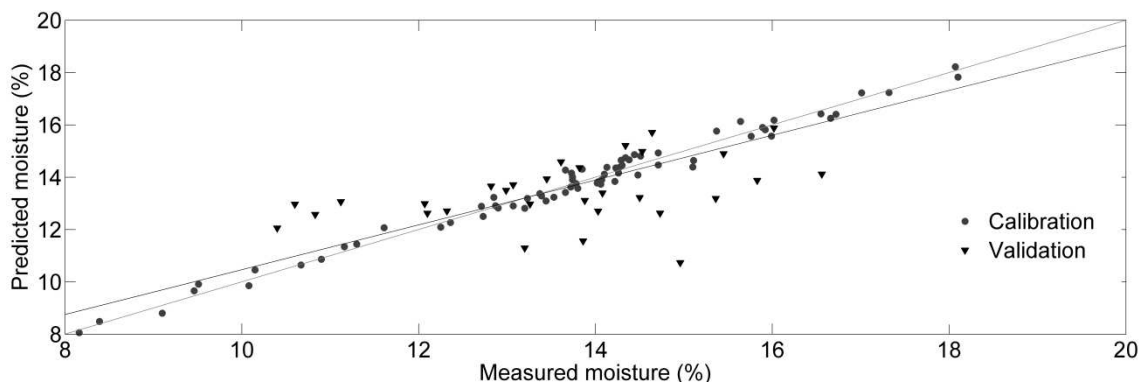


Fig. 2. Plot of the predicted values by NIR against the values measured by standard methods for moisture content.

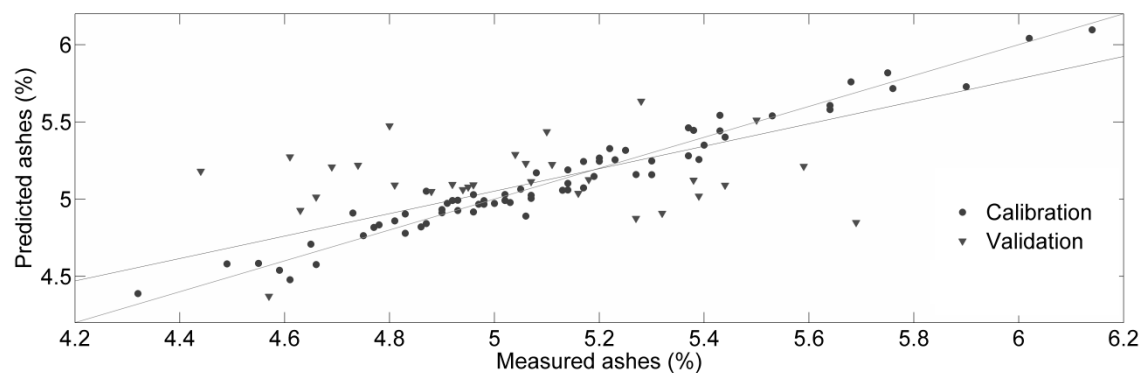


Fig. 3. Plot of the predicted values by NIR against the values measured by standard methods for ash content.

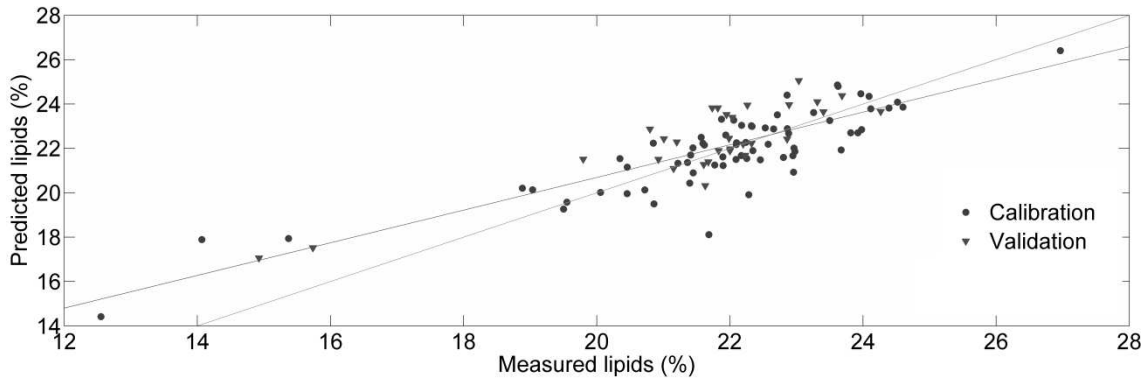


Fig. 4. Plot of the predicted values by NIR against the values measured by standard methods for lipid content.

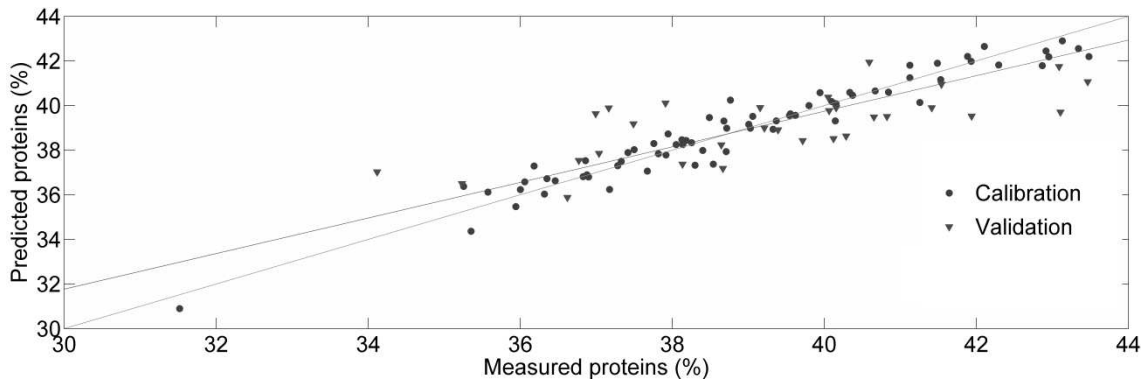


Fig. 5. Plot of the predicted values by NIR against the values measured by standard methods for protein content.

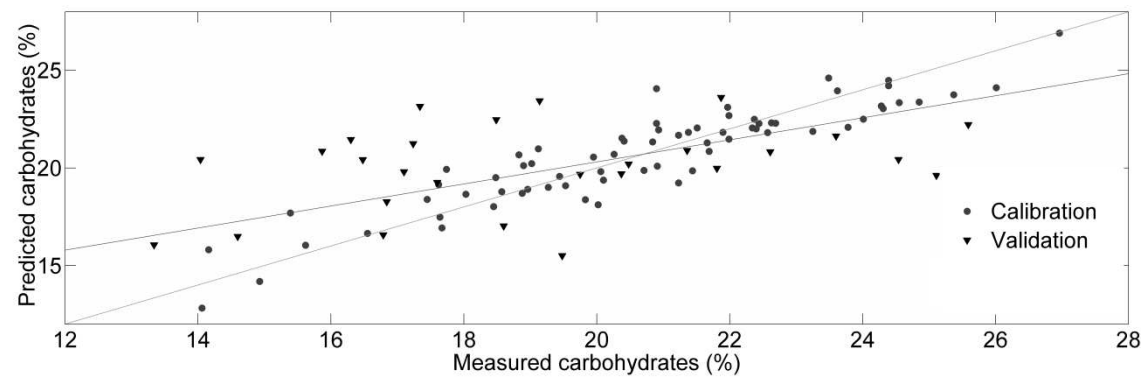


Fig. 6. Plot of the predicted values by NIR against the values measured by standard methods for carbohydrate content.

In all the figures the curve of predicted values versus the real ones in the 30 samples of the predictable group presents an appropriate adjustment to the model. It is noted that the inclination of the line is close to 1 and the intercept next to zero, two parameters that prove the absence of systematic errors.

It is observed that the parameters presented an appropriate relation between the evaluated and the predicted values, the models being closer to the real ones, confirming once the best results for moisture and protein contents (Figs. 2 and 5 - R^2 of 0.80 and 0.81, respectively).

In the next figure, the absolute errors from the predictions obtained by the models of moisture, ash, lipid, protein and carbohydrate contents of the 30 samples of the prediction set (Fig. 7) are presented.

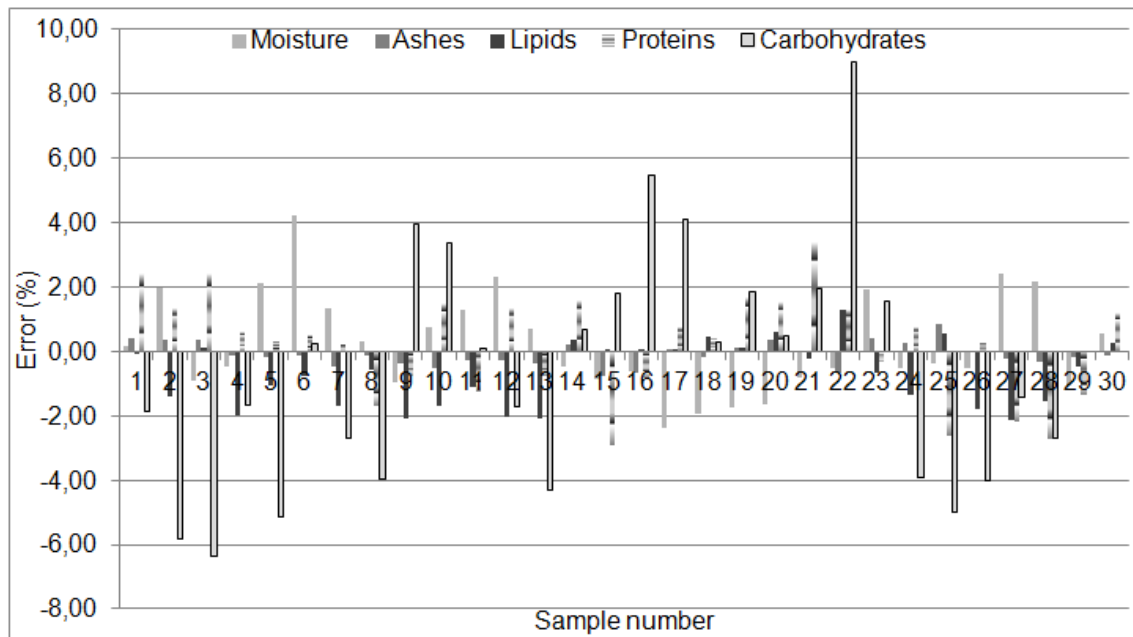


Fig 7. Absolute errors: evaluated parameters measured minus the predicted values of the soybean validation group.

Considering that the value of the bias is not enough to emphasize the difference between the mean values obtained by the conventional methods and the

predicted values obtained by FT-NIRS technique, the absolute error demonstrated uniformity for all the parameters with the exception of the carbohydrate values, which presented higher differences, and the lipid content, which despite the lower unities, the values estimated by the model were higher than the content obtained by the reference method (Fig. 7). Nevertheless, low values to bias and RMSEP suggest that the lipid content can be predicted by NIR, but not so accurately as the other parameters.

4. CONCLUSION

This study demonstrated that it is possible to predict the parameters of the quality of moisture, ashes, proteins, lipids and carbohydrates in soybeans by associating the near-infrared spectroscopy with multivariate calibration.

It can be concluded that despite the existence of a wide range of studies on the FT-NIRS technique, the methodology proposed in this work to develop a NIR model for a grain, controlling the quality of the soybean, has a high potential for implanting and developing calibration models in industries. The offer of the technique to inspection authorities can be a solution to faster and more efficiently control of the quality of a wide range of food in Brazil, enabling the replacement of laborious methods.

Acknowledgements

The authors thank EMBRAPA Soybean for providing the soybean samples and FAPESP (2010/50418-5) and CNPq for the financial support.

References

- Agelet, L. E., & Hurburgh, C. R. (2010). A Tutorial on Near Infrared Spectroscopy and Its Calibration. *Critical Reviews in Analytical Chemistry*, 40(4), 246–260.
- AOAC. (2006). Official methods of analysis of AOAC International. In W. Horwitz & G. W. Latimer (Eds.), (18th ed.). Gaithersbrug, Maryland: AOAC International
- Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5), 772–777.
- Bjorsvik, H.-R., & Martens, H. (2007). Calibration of NIR Instruments by PLS Regression. In E. W. Ciurczak & D. A. Burns (Eds.), *Handbook of Near-Infrared Analysis* (pp. 189–205). (3rd ed). New York, USA: CRC Press Taylor & Francis Group.
- Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, 37(8), 911–917.
- Brown, C. D., Vega-Montoto, L., & Wentzell, P. D. (2000). Derivative Preprocessing and Optimal Corrections for Baseline Drift in Multivariate Calibration. *Applied Spectroscopy*, 54(7), 1055–1068.
- CONAB - Companhia Nacional de Abastecimento. (2012). Retrieved May 8, 2012, from <http://www.conab.gov.br>
- EMBRAPA. (2012). Brazilian Agricultural Research Corporation. Retrieved July 9, 2012, from <http://www.cnpso.embrapa.br/>
- Geladi, Paul, & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185, 1–17.
- Geladi, P, MacDougall, D., & Martens, H. (1985). Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy*, 39(3), 491–500.
- Haaland, D. M., & Thomas, E. V. (1988). Partial Least-Squares Methods for Spectral Analyses . 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information. *Analytical chemistry*, 60(11), 1193–1202.
- Haughey, S. a., Graham, S. F., Cancouët, E., & Elliott, C. T. (2012). The application of Near-Infrared Reflectance Spectroscopy (NIRS) to detect melamine adulteration of soya bean meal. *Food Chemistry, In Press*, 1–5. doi:10.1016/j.foodchem.2012.01.068
- Holse, M., Larsen, F. H., Hansen, Å., & Engelsen, S. B. (2011). Characterization of marama bean (*Tylosema esculentum*) by comparative spectroscopy: NMR, FT-Raman, FT-IR and NIR. *Food Research International*, 44(1), 373–384.

- Jimaré Benito, M. T., Bosch Ojeda, C., & Sanchez Rojas, F. (2008). Process Analytical Chemistry: Applications of Near Infrared Spectrometry in Environmental and Food Analysis: An Overview. *Applied Spectroscopy Reviews*, 43(5), 452–484.
- Kandala, C. V., Sundaram, J., & Puppala, N. (2012). Analysis of Moisture Content , Total Oil and Fatty Acid Composition by NIR Reflectance Spectroscopy : A Review. *Lecture Notes in Electrical Engineering*, 146, 59–80.
- Kennard, R. W., & Stone, L. A. (1969). Computer Aided Design of Experiments. *Technometrics*, 11(1), 137–148.
- Kovalenko, I. V., Rippke, G. R., & Hurburgh, C. R. (2006). Determination of amino acid composition of soybeans (*Glycine max*) by near-infrared spectroscopy. *Journal of agricultural and food chemistry*, 54(10), 3485–91.
- Kuligowski, J., Carrión, D., Quintás, G., Garrigues, S., & de la Guardia, M. (2012). Direct determination of polymerised triacylglycerides in deep-frying vegetable oil by near infrared spectroscopy using Partial Least Squares regression. *Food Chemistry*, 131(1), 353–359.
- Munck, L., Jespersen, B. M., Rinnan, Å., Seefeldt, H. F., Engelsen, M. M., Nørgaard, L., & Engelsen, S. B. (2010). A physiochemical theory on the applicability of soft mathematical models-experimentally interpreted. *Journal of Chemometrics*, 24(7-8), 481–495.
- Naes, T., Isaksson, T., Fearn, T., & Davies, T. (2002). Outlier detection. *A User-Friendly Guide to Multivariate Calibration and Classification*. Chichester, UK: NIR Publications.
- Salgó, A., & Gergely, S. (2012). Analysis of wheat grain development using NIR spectroscopy. *Journal of Cereal Science*, 56(1), 31–38.
- Sorensen, L. K. (2009). Application of reflectance near infrared spectroscopy for bread analyses. *Food Chemistry*, 113(4), 1318–1322.
- Sun, J. M., Han, F. X., Yan, S. R., Yang, H., & Sato, T. (2008). Rapid determination of fatty acids in soybeans [*Glycine max* (L.) Merr.] by FT-near-infrared reflectance spectroscopy. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis*, 28(6), 1290–1295.
- Szigedi, T., Lénárt, J., Dernovics, M., Turza, S., & Fodor, M. (2011). Protein content determination in Brassica oleracea species using FT-NIR technique and PLS regression. *International Journal of Food Science & Technology*, 47(2), 436–440.
- Williams, P. (2001). Near-Infrared Technology: In the Agricultural and Food Industries. In P. Williams & K. Norris, (Eds.) (2nd ed) *American Association of Cereal Chemists*. (pp. 296). St. Paul, Minn., USA: American Association of Cereal Chemists.
- Williams, P. C., & Sobering, D. (1995). How do we do it: a brief summary of the methods we use in developing near infrared calibrations. In A. M. C. Daves & P. C. Williams

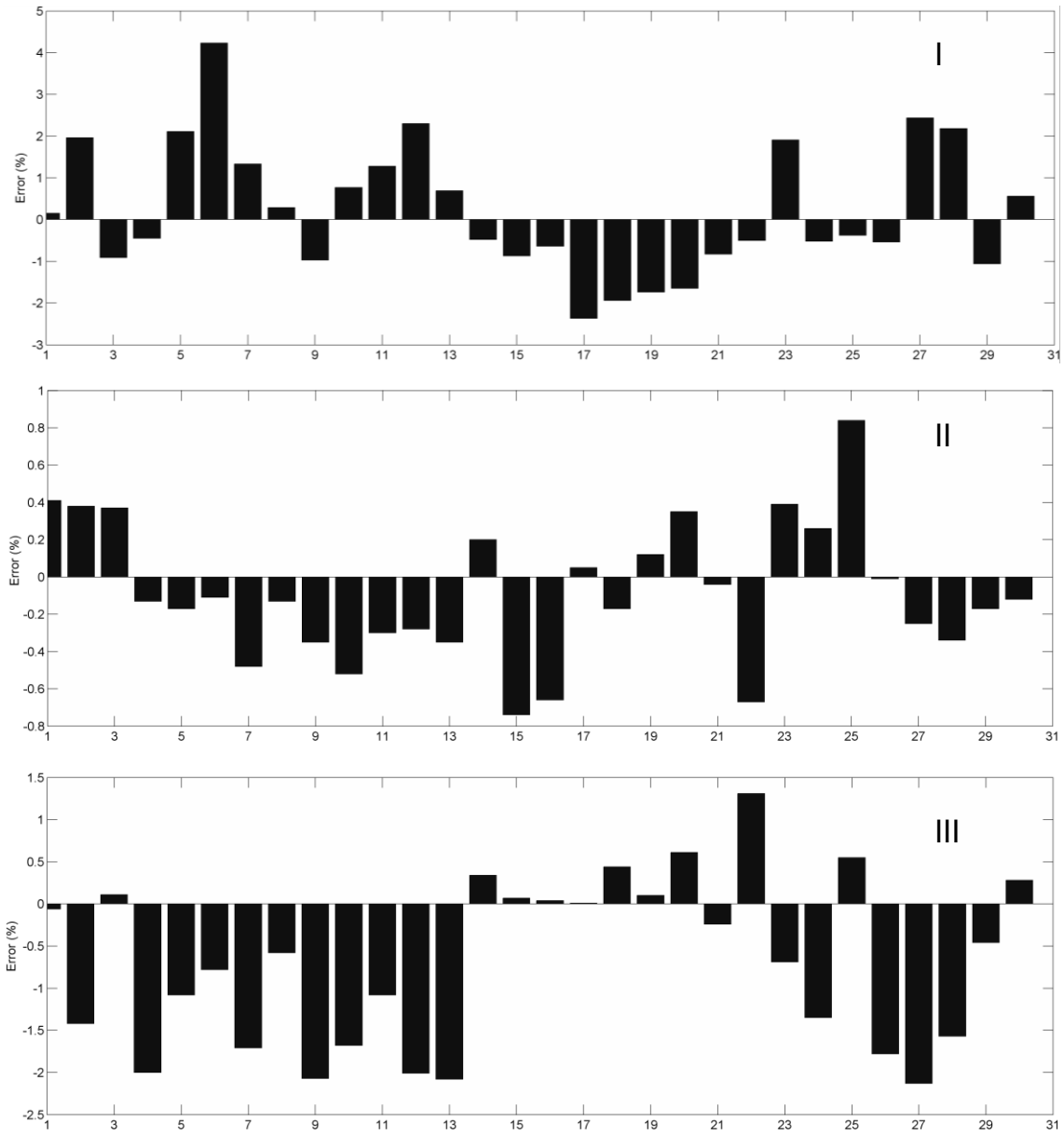
(Eds.), *Near Infrared Spectroscopy: The Future Waves* (pp. 185–188). Chichester, UK: NIR Publications.

Windham, W. R., Mertens, D. R., & Barton II, F. E. (1989). Protocol for NIRS calibration: sample selection and equation development and validation. In C. G. Marten, J. S. Shenk, & F. E. Barton (Eds.), *Near Infrared Reflectance Spectroscopy (NIRS): analysis of forage quality* (pp. 96–103) (643rd ed). Washington, DC: USDA-ARS: Agricultural Handbook.

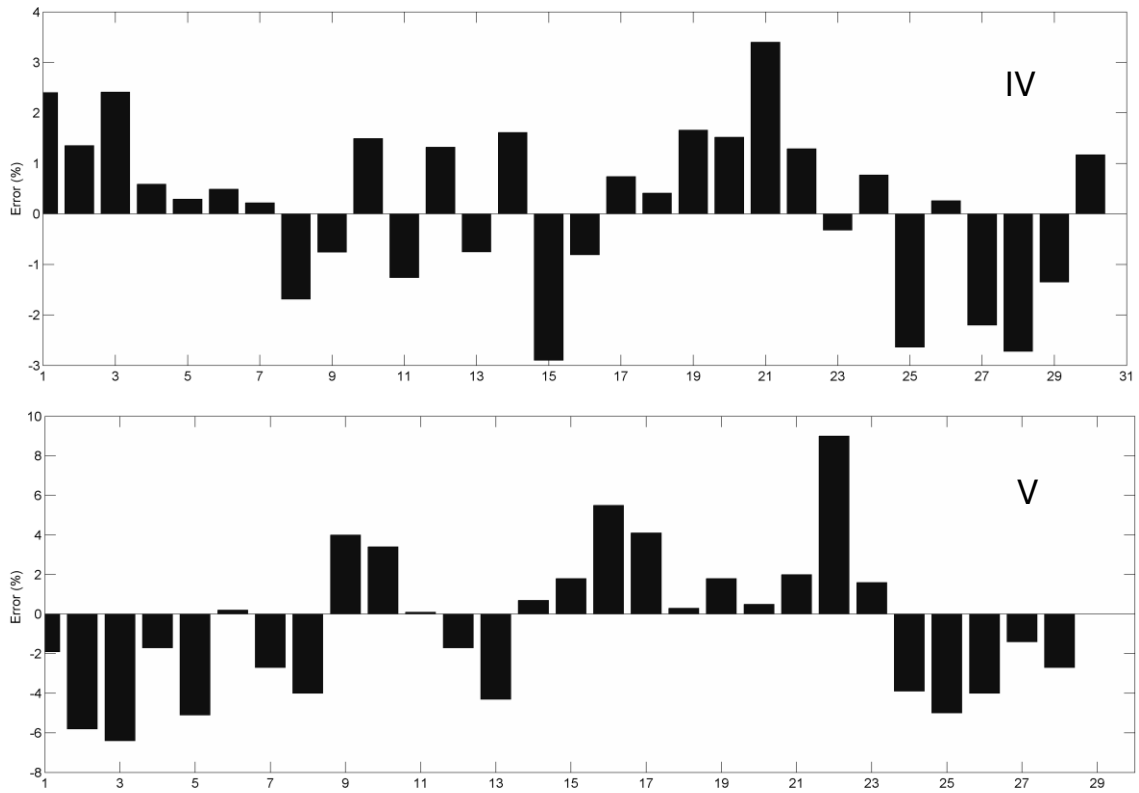
Zhou, X., Yang, Z., Huang, G., & Han, L. (2012). Non-invasive detection of protein content in corn distillers dried grains with solubles: method for selecting spectral variables to construct high-performance calibration model using near infrared reflectance spectroscopy. *Journal of Near Infrared Spectroscopy*, 20(3), 407.

Zhu, D., Wang, K., Zhang, D., Huang, W., Yang, G., Ma, Z., & Wang, C. (2011). Quality Assessment of Crop Seeds by Near-Infrared Hyperspectral Imaging. *Sensor Letters*, 9(3), 1144–1150.

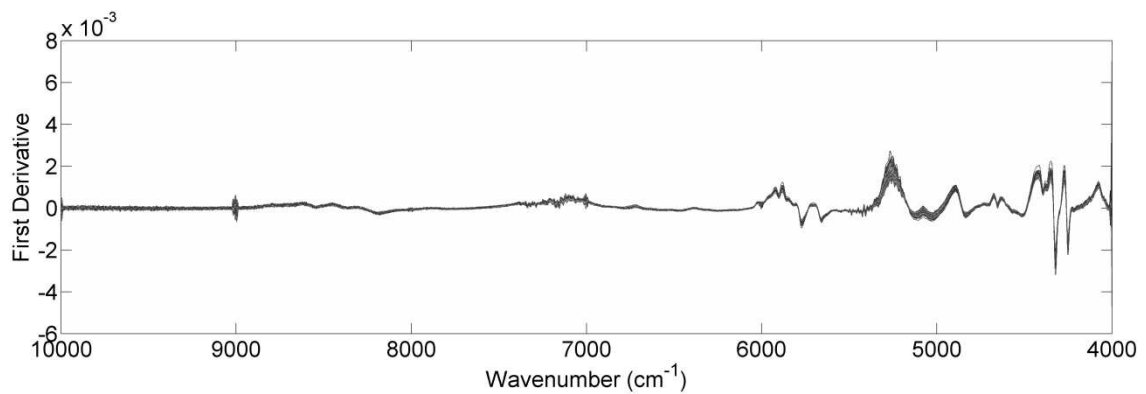
APPENDIX



Supplemental Figure A. Absolute error: evaluated value minus predicted value of the soybean validation group. (I) moisture, (II) ash and (III) lipid contents.



Supplemental Figure B. Absolute error: evaluated value minus predicted value of the soybean validation group. (IV) protein and (V) carbohydrate contents.



Supplemental Figure C. First derivative of the spectra obtained from ground soybean grains.

CAPÍTULO IV

COMBINATION OF NEAR-INFRARED SPECTROSCOPY AND CHEMOMETRICS TO PREDICT CHEMICAL CHARACTERISTICS OF CHENOPODIUM QUINOA FLOUR AND GRAINS

Manuscript submitted for publication in the journal *Food Chemistry* (FOODCHEM-D-12-04095)

ABSTRACT

Interest in the research and development of quinoa, an indigenous Andean crop, has increased in the last 15 years, mainly because of its nutritional value and high seed yield. Near-infrared spectroscopy (NIRS) is a rapid non-destructive technique, which is useful for studying the origin and chemical properties of these crops. Considering these advantages, the objectives of this research were the following: (i) to develop NIRS calibrations suitable for the routine determination of dietary constituents in 98 varieties of quinoa using partial least squares (PLS) and (ii) to identify the geographic origin of quinoa samples using principal component analysis (PCA). Two principal components explained 98.3% of the total variance. The PLS models developed for the quantification of moisture, ash, lipid, protein and carbohydrate content showed that the proposed methodology produced suitable results, with the graph of the real and predicted concentrations having a coefficient of determination (R^2) > 0.68 and RMSEP < 4.56%.

Keywords

Near-infrared spectroscopy; *Chenopodium quinoa*; Partial least squares; Principal components analysis.

1. Introduction

Quinoa (*Chenopodium quinoa* Willd. from the Chenopodiaceae family) is a dicotyledonous indigenous plant of the Andes region above 4000 m in altitude. Quinoa has received considerable attention in the last 15 years as an alternative crop throughout the world, primarily because of its high nutritional value. This pseudocereal is still widely cultivated in South America, including Bolivia, the largest producer with 20,000 ha per year, followed by Peru, Colombia, Ecuador, Chile and Argentina (FAO, 2012). Commercial quinoa production in the United States has been successful, with the grain targeted to the health food sector. Recently, quinoa has been introduced in Europe, North America, Asia and Africa. Many European countries are members of the “Quinoa - A multipurpose crop for EC’s agricultural diversification” project, which was approved in 1993 (FAO, 2011; Jacobsen, Mujica & Ortiz, 2003). Some Brazilian tests on quinoa have yielded good results and demonstrate the potential of quinoa for succession cropping because of its high grain and biomass yield, soil protection and ability for mechanical harvesting (Spehar, 2006).

This seed is also considered to be a potential crop for the National Aeronautics and Space Administration (NASA) Controlled Ecological Life Support System (CELSS) (Schlick & Bubenheim, 1996). Moreover, the year 2013 has been declared “The International Year of Quinoa” by the United Nations General Assembly, FAO and the Government of Bolivia, with the support of Argentina, Azerbaijan, Ecuador, Georgia, Honduras, Nicaragua, Paraguay, Peru and Uruguay. The objective of this project is to focus world attention on the role that quinoa's biodiversity and nutritional value play in providing food security and high quality food and in eradicating poverty (FAO, 2012).

Quinoa has high protein content. In particular, the essential amino acid composition of the quinoa seed protein is close to the ideal protein balance recommended by FAO and similar to milk protein (Comai, Bertazzo, Bailoni, Zancato, Costa & Allegri, 2007; Gross, Koch, Malaga, Miranda, Schoeneberger &

Trugo, 1989; Spehar & Souza, 1993). Quinoa seed lipids appear to be a high quality edible vegetable oil, similar in fatty acid composition to soybean oil, and quinoa is a good source of linoleic acid and linolenic acid (Valencia-Chamorro, 2004). Because of its potential, quinoa can be considered an alternative food, with high nutritional value and low cholesterol. Moreover, quinoa is gluten-free and has high fiber content (Caperuto, Amaya-Farfan & Camargo, 2000). Hirose, Fujita, Ishii and Ueno (2010) concluded that quinoa seeds cultivated in Japan are the most effective functional foodstuff among cereals, pseudocereals, and plant foods approved for consumption, in terms of being a source of antioxidative and bioactive flavonoids.

The nutritional value of grains has been determined using laborious methods of analysis, such as the analysis of lipid content (Bligh & Dyer, 1959) and protein content (Kjeldahl, AOAC, 2006), which are time-consuming and generate much chemical waste. Near-infrared spectroscopy (NIRS) is a rapid and non-destructive technique frequently applied to evaluate the quality of grains for food and agriculture (Shenk, Workman, & Westerhaus, 2007). NIRS studies have been used to analyze grains. Examples include estimating the moisture, ash, protein, lipid and carbohydrate composition of Brazilian soybean seeds; estimating the protein, dietary fiber and fatty acid content in marama beans, cultivated in Botswana, Namibia and South Africa; and evaluating the germinal capacity of maize, in addition to the protein and moisture content and hardness of wheat (Ferreira, Pallone, & Poppi, 2012; Holse, Larsen, Hansen, & Engelsen, 2011; Ozdemir, 2006). However, the predictive ability of this non-destructive technique to determine the nutritional potential of quinoa has not been studied.

To further the recognition of quinoa grains, the aim of this work is to evaluate the flour and grain of quinoa based on chemical parameters, such as moisture, protein, lipid, ash and carbohydrate content, using near-infrared spectroscopy and chemometrics tools. Moreover, the feasibility of using the NIRS technique without sample preparation will be demonstrated. For this purpose, only whole grains will be used, replacing traditional methods, which are time-consuming

and require toxic reagents. With this technique, the chemical characteristics of samples will be identified and classified according to origin.

2. Materials and methods

2.1. Samples

Quinoa grains (*Chenopodium quinoa* Willd.) harvested in Brazil in different seasons of the year (BRAZIL1 and BRAZIL2), between 2009 and 2011 (87 samples), and samples from Peru (3 samples) and Bolivia (8 samples) were evaluated.

NIR spectra were acquired with the grains directly in the near infrared reflectance accessory (NIRA) equipment and with samples that were milled using a TE-620 knife mill grinder in a 100 mesh strainer (Tecnal, São Paulo, Brazil).

The milled samples were used in the chemical analysis with standard methods.

2.2. Reference Analysis

The total contents of moisture, protein and ash were determined using the 925.09, 979.09 and 923.03 reference methods, respectively, described by the AOAC (2006). The total lipid content was determined using the Bligh and Dyer (1959) method and the total carbohydrate content was calculated with the formula ($\% \text{ carbohydrate} = 100 - \% \text{ moisture} - \% \text{ lipid} - \% \text{ protein} - \% \text{ ash}$). All the analyses were performed in triplicate and immediately after grinding to avoid any changes from contact with the environment.

2.3. Near-infrared spectroscopy

The grains and the ground grains were directly analyzed using diffuse reflectance, with the Fourier transform near-infrared reflectance (FT-NIRS) accessory NIRA (Perkin Elmer, United States), equipped with a Spinning Sample Module, on a Spectrum 100N (Perkin Elmer, United States). The NIRA software is Spectrum v.6.3.1.0132 (Perkin Elmer) and the measured region was between 10000 and 4000 cm^{-1} with 4 cm^{-1} of resolution, 64 scans. Measurements were performed in duplicate.

2.4. Data processing

The NIRS spectra acquired from grains were evaluated using principal component analysis (PCA) to visualize the maximum variability of the groups, identify patterns between the grains, classify the grains according to origin and detect outliers. The outliers were identified from the scores plot of PCA versus T2 (Hotelling) (Esbensen, 2002; Wold, Esbensen & Geladi, 1987). However, the outliers were not clear and none were removed until the construction of the calibration models.

The multivariate calibration model was created using partial least squares (PLS). The NIRS spectra were treated using the PLS Toolbox v. 5.8 (Eigenvector, Research, Inc., United States) program developed for Matlab 7.8 (MathWorks, United States).

From a total of 98 quinoa samples, 68 were used to construct the calibration model, and 30 were used for the external validation (prediction set). The samples for calibration and prediction were chosen using the Kennard-Stone algorithm (Kennard & Stone, 1969).

To minimize the effect of changes in the baseline, the spectra were preprocessed using a combination of many mathematical treatments: multiplicative scatter correction (MSC) (Geladi, MacDougall & Martens, 1985), smoothing,

standard normal variate transformation (SNV) (Barnes, Dhanoa, & Lister, 1989), or first and second derivatives using the Savitzky-Golay method (Brown, Vega-Montoto & Wentzell, 2000). Additionally, all spectral data were mean-centered before calibration. The number of PLS factors was determined using a cross-validation (leave-one-out) procedure (Geladi & Kowalski, 1986).

The best combination of statistics used to evaluate the calibration models included the following: root mean square error calibration (RMSEC), root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP), coefficient of determination (R^2) of the plot between the values predicted from NIRS and the reference values and residual predictive deviation (RPD), as described in previous work (Ferreira et al., 2012; Williams & Sobering, 1995; Windham, Mertens & Barton II, 1989).

3. Results and Discussion

3.1. Exploratory analysis

Mean-centered preprocessing has been applied to the spectra obtained from the quinoa grains to classify the samples according to origin. The new matrix was used for the decomposition from PCA.

PCA was utilized to show the variety of quinoa groups involved in one calibration. Fig. 1 shows the relocation of samples into the dimensions composed of principal components PC 1 and PC 2, where the new structure can explain the maximum variance (98.34%) of the original data.

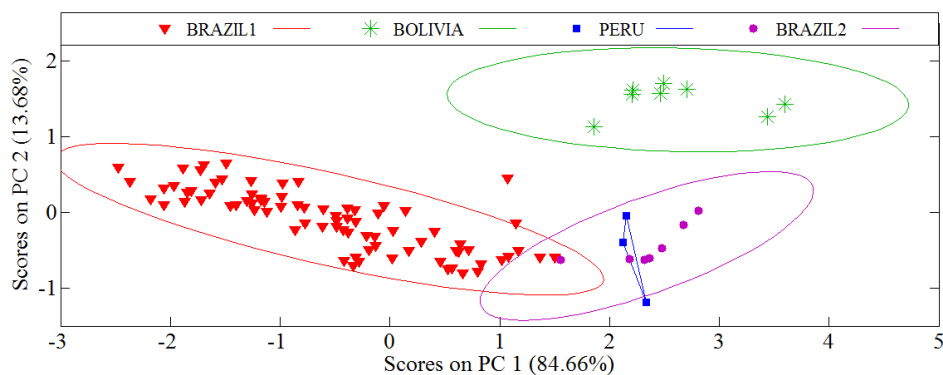


Figure 1. Score plot of the first and second principal components from PCA of the NIRS of quinoa grains.

PCA of the NIRS spectra of the quinoa grain resulted in three groups. Two groups were considered Brazilian samples, BRAZIL1 and BRAZIL2, corresponding to different varieties cultivated in Brazil during different harvest seasons. The BOLIVIA samples were clearly separated from the Brazilian samples (BRAZIL1 and 2) by PC 1 and PC 2. The PERU samples are chemically equal to the BRAZIL2 samples. The loadings of PC 1 (Fig. 2A), which separated the BRAZIL1 samples, were examined. Using this data, the absorption peaks from 5000 to 4000 cm^{-1} in the negative loadings can be attributed to the stretching vibrations of $-\text{COOR}$ and $-\text{HC}=\text{CH}-$, which are from carbohydrates, protein and fat. Similar results were found for the positive loadings of PC 2, which separated the BOLIVIA samples. Because of the importance of the PC 2 loadings, the bands from 5300 to 4400 cm^{-1} can be attributed to $\text{C}=\text{O}$ and $\text{C}-\text{H}$ groups (Fig. 2B). Thus, the PCA results justify combining all varieties to develop a broad-based model.

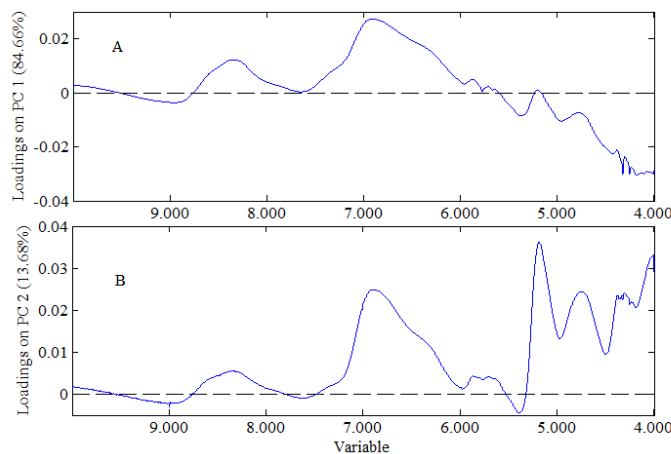


Figure 2. PCA loadings plot of the quinoa grains. (A) PC 1 (B) PC 2 vs. variables.

3.2. Chemical analysis

Table 1 presents, in dry weight, the range and mean of the ash, lipid, protein and carbohydrate content found using reference methods, in addition to the moisture content calculated for these grains. The standard deviation shows the high variability of the data set used for calibration.

The values reported in this study for quinoa from many regions of South America are equal to those previously reported by Wright, Pike, Fairbanks and Huber (2002), who evaluated samples from the Benson Agriculture and Food Institute of Bolivia with 2.83 to 3.40% ash, 5.77 to 6.26% lipid, 16.12 to 17.25% protein and 73.08 to 75.27% carbohydrate. The data reported by Ogungbenle (2003) for quinoa from Canada are similar to values presented in Table 1 for protein (15.20%), lipid (7.09%) and carbohydrate (65.65%), but the ash content of the quinoa (1.35%) is lower than the value found in this study (5.49%). This difference is related to the fact that the samples are from different localities and, in this study, the analyses were conducted on grains, with the bran, which increases the ash content.

Table 1. Chemical composition of the calibration and validation sample set (dry weight).

<i>Parameter</i>	<i>Calibration set</i>			<i>Validation set</i>		
	<i>Minimum (%)</i>	<i>Maximum (%)</i>	<i>Mean ± SD(%)</i>	<i>Minimum (%)</i>	<i>Maximum (%)</i>	<i>Mean ± SD(%)</i>
Moisture ^a	25.66	33.16	29.03 ± 1.77	27.04	32.93	29.16 ± 1.59
Ash	3.07	9.15	5.49 ± 1.39	3.64	7.93	5.87 ± 0.75
Lipid	6.19	15.52	8.29 ± 2.15	6.52	14.06	7.96 ± 1.39
Protein	11.40	36.10	23.94 ± 5.04	16.30	33.64	24.00 ± 3.05
Carbohydrate	43.64	76.37	62.27 ± 6.75	46.41	73.80	61.86 ± 5.21

SD = standard deviation

^a humid weight

A high moisture value is found in these quinoa grains (29.03%), which is explained by the analysis being performed without previously drying the seeds, as usually occurs when a company is interested in storing seeds. Moreover, the seeds easily lose and absorb water and the accumulation of moisture is favorable for drought tolerance, especially when quinoa is adapted to Brazilian soil (Spehar, 2006).

3.3. Calibration modeling and validation

The optimal conditions were selected for each calibration and were adopted for both internal (cross-validation, leave-one-out) and external (validation set) validations of the moisture, ash, lipid, protein and carbohydrate content of the quinoa flour and grain. Moreover, spectral regions were selected by taking into account the characteristic absorption bands of the analyzed parameters. Selecting the entire range of the spectra (10000 to 4000 cm^{-1}) provided the maximum information for most calibration models. This information is particularly important for the milled quinoa analysis because the spectra of these samples have a lower intensity than the grain spectra. The range between 600 to 3977 cm^{-1} was only chosen to measure the carbohydrate content of quinoa flour and the range between 6000 to 4000 cm^{-1} was chosen to measure the ash and protein content of quinoa grains.

All spectra were transformed to their first derivatives, using widths of 15 (number of points in filter) and second order polynomials. SNV was also employed for the moisture and lipid content of quinoa flour and the protein content of quinoa grains, while the NIR spectra was also transformed using MSC for the ash content of quinoa grains.

The NIRS spectra for quinoa grains and quinoa flour are very similar. However, when a pre-treatment, such as a derivative, SNV or MSC is applied, the differences can generally be observed in the same regions in which the reconstructed signals from the selected coefficients wavelet show significant contributions. The main chemical constituents, especially those that contribute to these differences, are proteins, lipids and moisture, which was also observed by Cocchi, Durante, Foca, Marchetti, Tassi and Ulrici (2006) for durum wheat and wheat flour. Applying the first derivative emphasized the peaks below and above the baseline, providing the best resolution for the expected signals (Shenk et al., 2007). SNV and MSC pre-treatments corrected the multiplicative scattering effect and uneven particle size, as in the milled samples and the grain.

Table 2 shows the calibration and prediction results of quinoa flour and grains. The value of R^2 ranges from 0.72 to 0.77 for quinoa flour and from 0.68 to 0.85 for grain, while the root mean square error of prediction (RMSEP) ranges from 0.86 to 4.56% and from 0.62 to 4.27% for quinoa flour and grains, respectively. The results were similar between flour and grains with a low prediction error for the models constructed for ash content (RMSEP from 0.62 to 0.86%). The root mean square errors of calibration (RMSEC) were low, ranging from 0.08 to 0.85% for quinoa flour and from 0.12 to 1.54% for quinoa grains. These results were higher for the grains, which were expected because the reference analyses were performed after milling. Moreover, energy is scattered because of the distribution of grain inside the sample door. The same increase is observed for the root mean square errors of cross-validation (RMSECV) from 0.82 to 4.51% for flour and 0.79 to 5.01% for grains.

Table 2. Prediction results of the moisture, ash, lipid, protein and carbohydrate content from NIRS spectra collected from quinoa grains and flour.

Parameters	Quinoa flour						Quinoa grain					
	R ²	LV	RMSEC %	RMSEP %	RMSECV %	RPD	R ²	LV	RMSEC %	RMSEP %	RMSECV %	RPD
Moisture	0.77	7	0.43	1.42	1.56	1.24	0.74	9	0.41	1.52	1.65	1.16
Ash	0.74	9	0.08	0.86	0.82	1.62	0.85	8	0.15	0.62	0.79	2.24
Lipid	0.72	9	0.21	1.37	1.58	1.57	0.75	5	0.12	1.30	1.61	2.15
Protein	0.75	9	0.49	3.02	3.50	1.67	0.68	9	0.75	3.37	3.49	1.49
Carbohydrate	0.77	10	0.85	4.56	4.51	1.48	0.76	8	1.54	4.27	5.01	1.58

R² = coefficient of determination; LV = latent variables; RMSEC = root mean square error of calibration; RMSEP = root mean square error of prediction; RMSECV = root mean square error of cross-validation; RPD = residual predictive deviation calculated as SD/RMSEP.

The best results for the relation between RMSECV and standard deviation (RPD) were for the ash and lipid content of grains (2.24 and 2.15, respectively), showing the good performance of the technique in predicting these chemical contents of grains. However, based on the determination coefficients and the low calibration and prediction errors, it can be concluded that this technique also has the ability to predict moisture, ash, lipid, protein and carbohydrate content of flour and grains.

In a previous study, Haughey, Graham, Cancouët and Elliott (2012) used the same number of latent variables (7 to 10) and the same pre-treatment as this study, but they obtained greater accuracy of calibration for soybeans, with R² from 0.89 to 0.99, and low RMSEC and RMSEP (0.27% and 0.36%, respectively). Cocchi et al. (2006) achieved low calibration and prediction errors (\cong 0.5%) for wheat flour, associated with the official Italian method, of 1%. The calibration errors for quinoa obtained in this study were lower than 1%, except for the carbohydrate content of the grains (1.54%) because the reference value for carbohydrates that was used for the calibration was obtained by difference.

Fig. 3 (quinoa flour) and Fig. 4 (quinoa grain) present the relations between the values measured using reference methods and the values predicted from the NIRS calibration. Good agreement between the measured values and those predicted from models was observed, mainly for the ash content of quinoa grain and the moisture content of quinoa flour. This finding confirmed that all the samples were appropriately modeled. These plots show that quinoa without previous preparation can be analyzed using near-infrared diffuse reflectance.

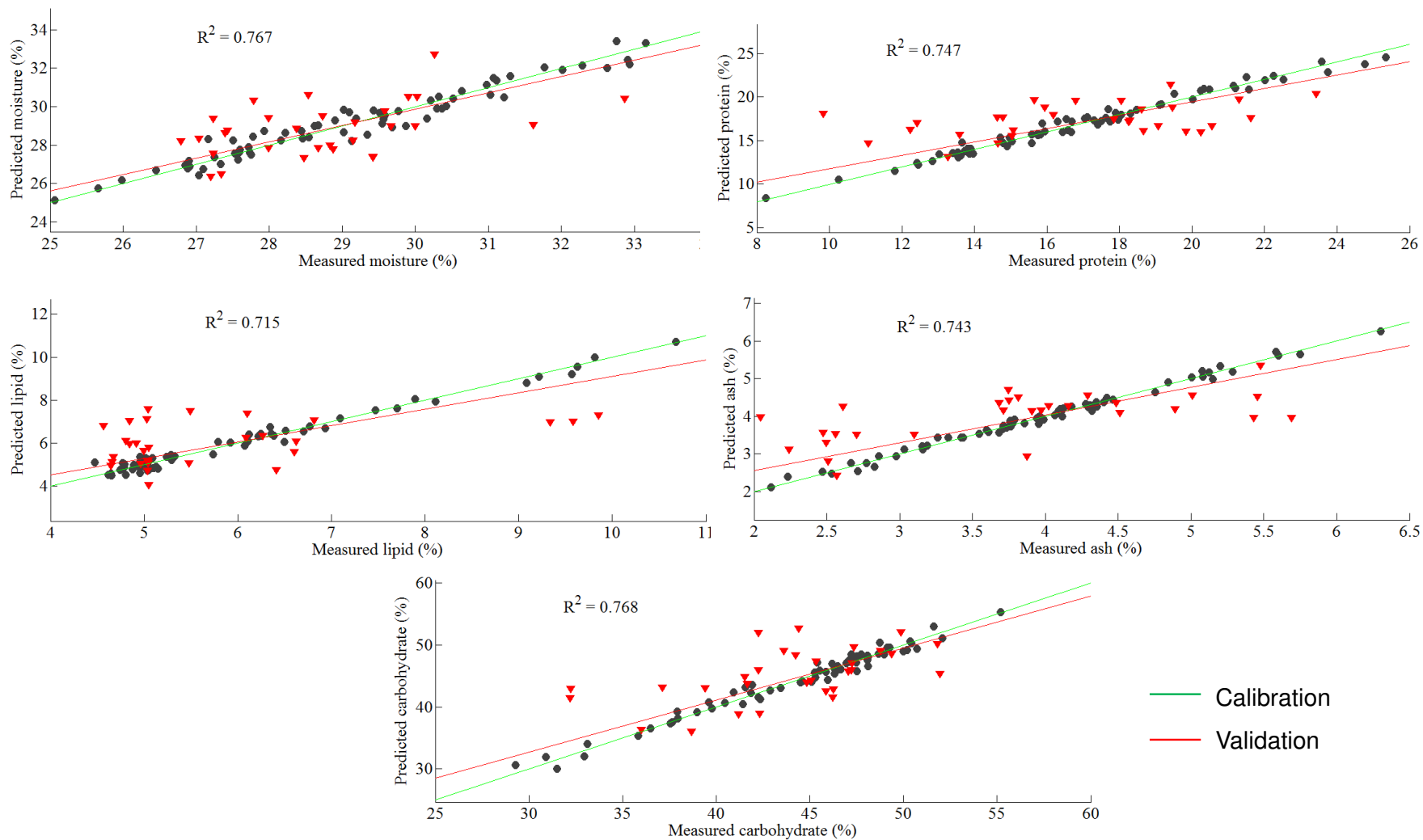


Figure 3. Measured parameters vs. predicted values from NIRS of quinoa flour. (●) calibration values, (▼) predicted values.

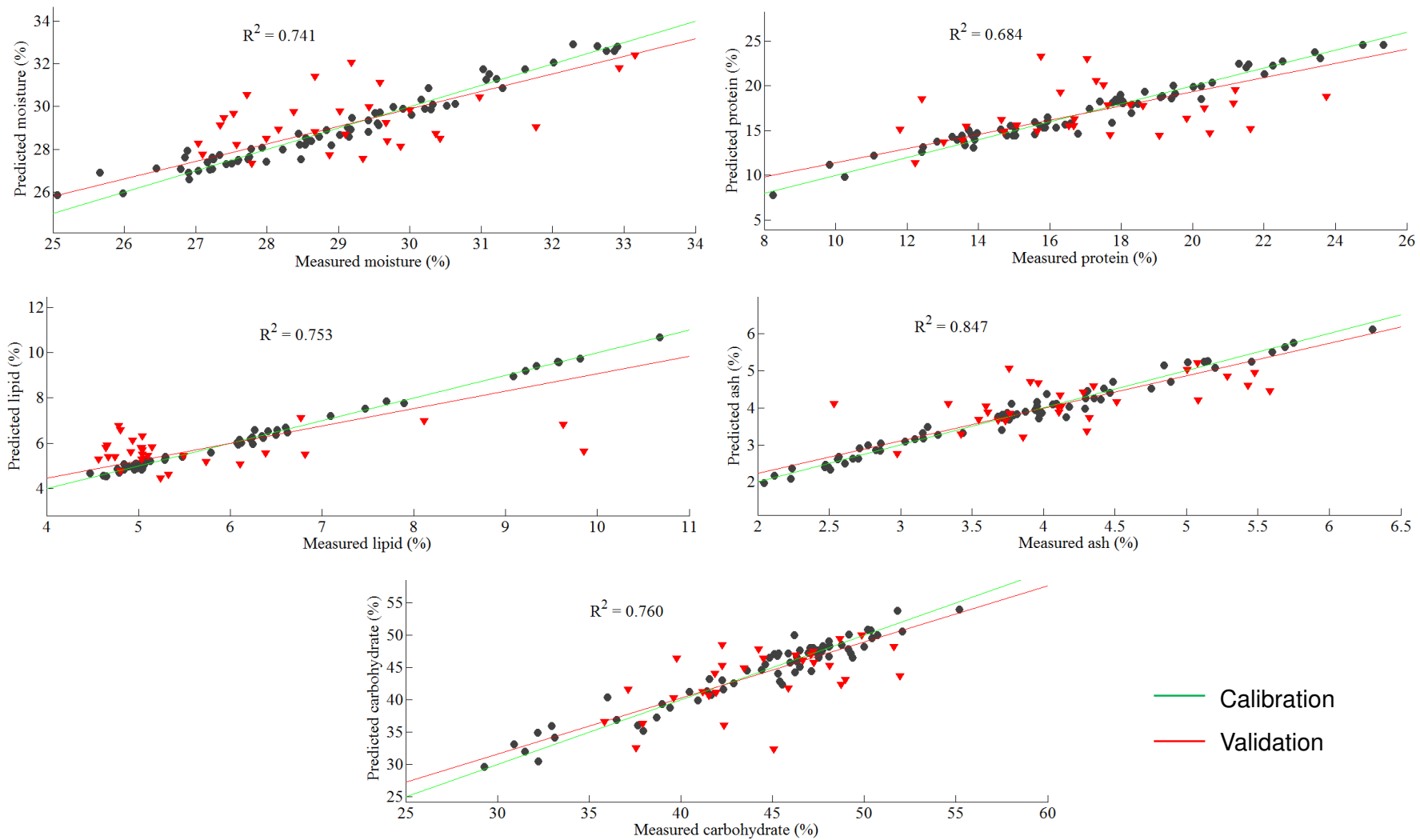


Figure 4. Measured parameters vs. predicted values from NIRS of quinoa grain. . (●) calibration values, (▼) predicted values.

4. Conclusion

Based on the results presented in this study, the method based on near-infrared spectroscopy associated with chemometrics analysis was suitable for predicting the moisture, protein, lipid, ash and carbohydrate content, both for quinoa flour and quinoa grains.

In addition, the NIRS spectrum of grains allows the mapping of grain origins by PCA according to chemical characteristics.

The viability of the NIRS technique with no waste generation, low cost, reduced time and no sample preparation for replacing laborious methods of analysis was demonstrated because the results for grain were satisfactory.

Acknowledgements

The authors thank the FAPESP (2010/50418-3) and CNPq for financial support and Embrapa Cerrados for supplying samples.

References

- AOAC. (2006). Official methods of analysis of AOAC International. In W. Horwitz & G. W. Latimer (Eds.), *AOAC International* (18th ed., Vol. 18th Ed). Gaithersbrug, Maryland.
- Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, *43*(5), 772–777.
- Bligh, E. G., & Dyer, W. J. (1959). A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, *37*(8), 911–917.
- Brown, C. D., Vega-Montoto, L., & Wentzell, P. D. (2000). Derivative Preprocessing and Optimal Corrections for Baseline Drift in Multivariate Calibration. *Applied Spectroscopy*, *54*(7), 1055–1068.
- Caperuto, L. C., Amaya-Farfan, J., & Camargo, C. R. O. (2000). Performance of quinoa (*Chenopodium quinoa* Willd) flour in the manufacture of gluten-free spaghetti. *Journal of Science of Food and Agriculture*, *81*(1), 95–101.

- Cocchi, M., Durante, C., Foca, G., Marchetti, A., Tassi, L., & Ulrici, A. (2006). Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. *Talanta*, *68*(5), 1505–11.
- Comai, S., Bertazzo, A., Bailoni, L., Zancato, M., Costa, C. V. L., & Allegri, G. (2007). The content of proteic and nonproteic (free and protein-bound) tryptophan in quinoa and cereal flours. *Food Chemistry*, *100*(4), 1350–1355.
- Esbensen, K. H. (2002). *Multivariate Data Analysis - in Practice: An Introduction to Multivariate Data Analysis and Experimental Design*. (D. Guyot, F. Westad, & L. P. Houmøller, Eds.) (5th ed., p. 598). Aalborg University, Esbjerg: Multivariate Data Analysis.
- FAO. Food and Agriculture Organization of the United Nations. (n.d.). International Year of the Quinoa IYQ-2013. Retrieved October 7, 2012, from <http://www.rlc.fao.org/en/about-fao/iyq-2012/>
- FAO. Food and Agriculture Organization of the United Nations. (2011). Quinoa: An ancient crop to contribute to world food security. *Regional Office for Latin America and the Caribbean*. FAO. Retrieved July 10, 2012, from http://www.fao.org/alc/file/media/pubs/2011/cultivo_quinoa_en_lr.pdf
- Ferreira, D. S., Pallone, J. A. L., & Poppi, R. J. (2012). Fourier transform near-infrared spectroscopy (FT-NIRS) application to estimate Brazilian soybean [*Glycine max* (L.) Merrill] composition. *Food Research International*, *51*(1 April 2013), 53–58.
- Geladi, P., MacDougall, D., & Martens, H. (1985). Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. *Applied Spectroscopy*, *39*(3), 491–500.
- Geladi, Paul, & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, *185*, 1–17.
- Gross, R., Koch, F., Malaga, I., De Miranda, A. F., Schoeneberger, H., & Trugo, L. C. (1989). Chemical composition and protein quality of some local Andean food sources. *Food Chemistry*, *34*(1), 25–34.
- Haughey, S. a., Graham, S. F., Cancouët, E., & Elliott, C. T. (2012). The application of Near-Infrared Reflectance Spectroscopy (NIRS) to detect melamine adulteration of soya bean meal. *Food Chemistry*, *In Press*, 1–5.
- Hirose, Y., Fujita, T., Ishii, T., & Ueno, N. (2010). Antioxidative properties and flavonoid composition of Chenopodium quinoa seeds cultivated in Japan. *Food Chemistry*, *119*(4), 1300–1306.
- Holse, M., Larsen, F. H., Hansen, Å., & Engelsen, S. B. (2011). Characterization of marama bean (*Tylosema esculentum*) by comparative spectroscopy: NMR, FT-Raman, FT-IR and NIR. *Food Research International*, *44*(1), 373–384.
- Jacobsen, S. E., Mujica, A., & Ortiz, R. (2003). The Global Potential for Quinoa and Other Andean Crops. *Food Reviews International*, *19*(1-2), 139–148.

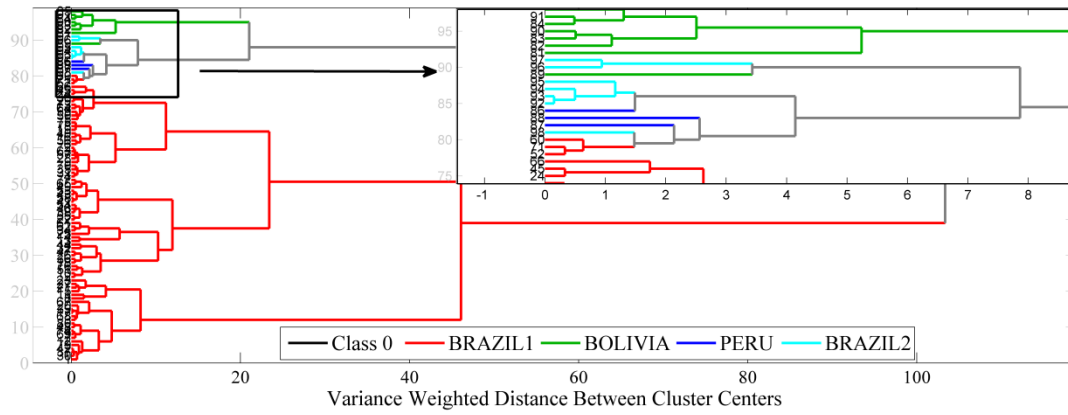
- Kennard, R. W., & Stone, L. A. (1969). Computer Aided Design of Experiments. *Technometrics*, 11(1), 137–148.
- Ogungbenle, H. N. (2003). Nutritional evaluation and functional properties of quinoa (*Chenopodium quinoa*) flour. *International journal of food sciences and nutrition*, 54(2), 153–8.
- Ozdemir, D. (2006). Genetic multivariate calibration for near infrared spectroscopic determination of protein, moisture, dry mass, hardness and other residues of wheat. *International Journal of Food Science and Technology*, 41(s2), 12–21.
- Schlick, G., & Bubenheim, D. L. (1996). Quinoa: Candidate Crop for NASA's Controlled Ecological Life Support Systems. In J. Janick (Ed.), *Progress in new crops* (pp. 632–640). Arlington, VA: ASHS PRes.
- Shenk, J. S., Workman, J. J. J., & Westerhaus, M. O. (2007). Application of NIR spectroscopy to agricultural products. In D. A. Burns & emil W. Ciurczak (Eds.), *Handbook of near-infrared analysis* (3rd ed., pp. 347–386). CRC Press.
- Spehar, C. R. (2006). Adaptation of quinoa (*Chenopodium quinoa* Willd .) to increase the agricultural and alimentary diversity in Brazil. *Cadernos de Ciência & Tecnologia*, 23(1), 41–62.
- Spehar, C. R., & Souza, P. I. de M. (1993). Adapting the quinoa (*Chenopodium quinoa* Willd) to cultivation in the brazilian highlands: preliminary results. *Pesquisa Agropecuária Brasileira*, 28(5), 635–639.
- Valencia-Chamorro, S. A. (2004). Quinoa. In C. Wrigley, H. Corke, & C. E. Walker (Eds.), *Encyclopedia of Grain Science* (Elsevier L., pp. 1–8).
- Williams, P. C., & Sobering, D. (1995). How do we do it: a brief summary of the methods we use in developing near infrared calibrations. In A. M. C. Daves & P. C. Williams (Eds.), *Near Infrared Spectroscopy: The Future Waves* (pp. 185–188). Chichester, UK: NIR Publications.
- Windham, W. R., Mertens, D. R., & Barton II, F. E. (1989). Protocol for NIRS calibration: sample selection and equation development and validation. In C. G. Marten, J. S. Shenk, & F. E. Barton (Eds.), *Near Infrared Reflectance Spectroscopy (NIRS): analysis of forage quality* (643rd ed., pp. 96–103). Washington, DC: USDA-ARS: Agricultural Handbook.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1-3), 37–52.
- Wright, K. H., Pike, O. A., Fairbanks, D. J., & Huber, C. S. (2002). Composition of Atriplex hortensis , Sweet and Bitter Chenopodium quinoa Seeds. *JFS: Food Chemistry and Toxicology*, 67(4), 1383–1385.

APPENDIX

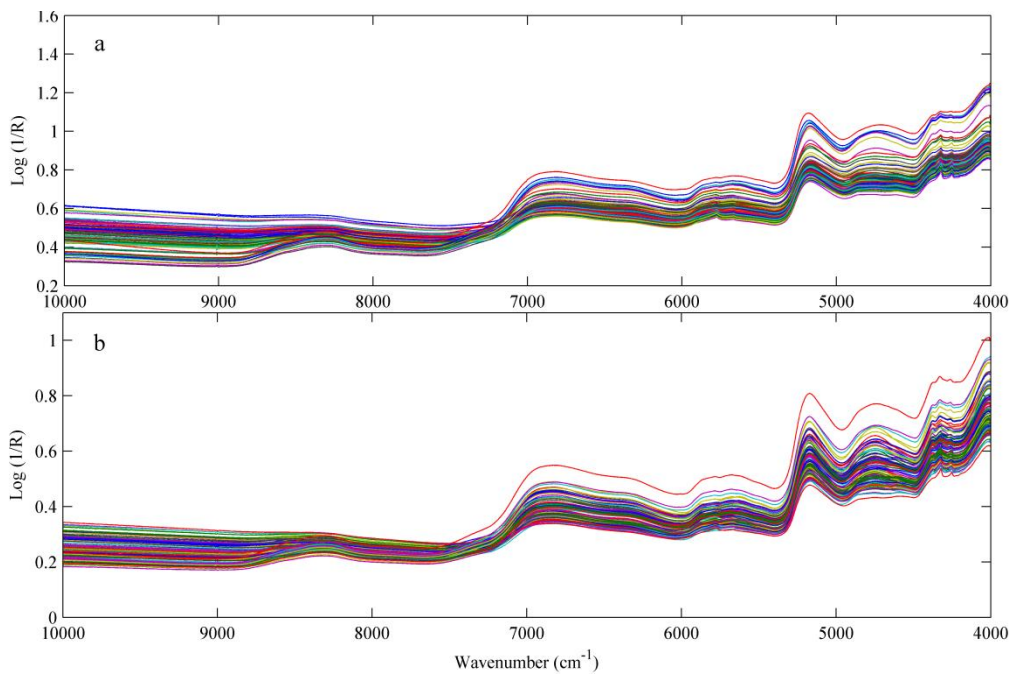
Supplemental Table A. Optimal parameters for NIRS calibrations of moisture, ash, lipid, protein and carbohydrate content on quinoa flour and grain.

<i>Matrix</i>		<i>Moisture</i>	<i>Ash</i>	<i>Lipid</i>	<i>Protein</i>	<i>Carbohydrate</i>
1						
	Wavenumber range (cm ⁻¹)	10000 - 4000	10000- 4000	10000 - 4000	10000- 4000	6000 - 3977
Quinoa flour	Pretreatment	1st derivative ^a + SNV+ mean centered	1st derivative ^a + mean centered	1st derivative ^a + SNV+ mean centered	1st derivative + mean centered	1st derivative ^a + mean centered
2						
	Wavenumber range (cm ⁻¹)	10000 - 4000	6000 - 4000	10000 - 4000	6000 - 4000	10000 - 4000
Quinoa grain	Pretreatment	1st derivative ^a + mean centered	1st derivative ^a + MSC	1st derivative ^a + mean centered	1st derivative ^a + SNV	1st derivative ^a + mean centered

^a Savitzky-Golay employs the terms (y, width, order, deriv): y, the matrix of row vectors to be smoothed; width, number of points in filter; order, order of the polynomial; and deriv, derivative. It was employed (y,15,2,1).



Supplemental Figure A. Histogram of quinoa grain on NIRS mean centered



Supplemental Figure B. NIRS spectra of quinoa grain (a) and quinoa flour (b) without preprocessing.

CAPÍTULO V

PREDIÇÃO DO CONTEÚDO DE FIBRAS ALIMENTARES EM SOJA (Glycine max) PELO MÉTODO NÃO DESTRUTIVO FT-NIRS

Manuscrito em preparação a ser submetido para publicação no periódico
Journal of Agricultural and Food Chemistry

RESUMO

O conteúdo de fibras alimentares ou dietéticas em alimentos é um importante parâmetro de qualidade para a rotulagem e informação quanto à funcionalidade do alimento. As fibras não são digeríveis pelas secreções endógenas do trato gastrointestinal e participam de diversas funções na digestão humana. Seu conteúdo em produtos vegetais é determinado como fibras dietéticas totais (FD), o qual, por definição, inclui polissacarídeos não amiláceos de plantas, oligossacarídeos, frutooligosacarídeos, amido resistente e lignina. Entretanto, a desvantagem do método oficial é a demora na sua realização em razão da demanda de muitas etapas. O procedimento é o enzimático-gravimétrico e leva 5 dias para ser desenvolvido. Neste trabalho foi aplicada a técnica de NIR associada à quimiometria como alternativa para substituição ao método oficial de determinação de fibras em soja brasileira. O método enzimático-gravimétrico apresenta valores de 10,63 a 19,50% para 80 variedades de soja. O modelo PLS desenvolvido para prever fibras dietéticas totais, usando reflectância difusa, apresentou RMSEP de 1,79% e R^2 de 0,67, indicando adequação do modelo na predição de FD em soja.

Palavras-chave: fibras dietéticas, NIR, análise multivariada de dados.

ABSTRACT

Dietary fiber content is an important quality parameter for labeling and information about food function. Fibers are resistant to digestion by the endogenous secretion of gastrointestinal tract (enzymes) and have many functions in human digestion. The fiber content of plant material is measured as total dietary fiber (DF), which, by definition, includes plant nonstarch polysaccharides, oligosaccharides, fructooligosaccharides, resistant starch and lignin. But the official dietary fiber method is extremely time consuming and labor intensive. The procedure is the enzymatic-gravimetric method and takes 5 days to perform. The NIR technique associated to chemometric was applied as alternative to official methods of dietary fiber determination in Brazilian soybean. Enzymatic-gravimetric method in 80 varieties of soybean presents values range 10.63 to 19.50%. PLS models developed for predicting total dietary fiber, using reflectance spectra of ground soybean, showed RMSEP of 1.79% and R^2 of 0.67, indicating accuracy of the model in the prediction of total dietary fiber in soybean.

Keywords: Dietary fiber, NIR, multivariate data analysis.

INTRODUÇÃO

As fibras dietéticas ou alimentares (FA) são compostas principalmente de polissacarídeos, lignina e substâncias associadas, tais como proteínas de parede celular, compostos fenólicos, fitatos e oxalatos, resistentes à hidrólise (digestão) pelas enzimas alimentares humanas. Estas macromoléculas compreendem dois grupos químicos: com estrutura de polissacarídeos vegetais representados pela celulose, hemicelulose e pectinas, e aqueles sem a referida estrutura, tais como a lignina, as gomas e mucilagens. Podem ser classificadas em relação ao seu grau de solubilidade em água: solúveis e insolúveis ¹. E segundo o Codex Alimentarius ² estes carboidratos possuem grau de polimerização não inferior a 3.

A relação das fibras com a saúde começou na década de 1970, com pesquisas demonstrando que, em comunidades nas quais se consumiam poucos alimentos processados industrialmente, eram raras as doenças como hipertensão, infarto do miocárdio, obesidade e distúrbios gastrointestinais. As fibras insolúveis fornecem a massa necessária para a ação peristáltica do intestino, enquanto que as fibras solúveis têm a propriedade de se ligarem à água, formando um gel que reduz a absorção de lipídios e açúcares, tornando-se substrato para a formação de rica flora bacteriana ¹. De acordo com o FDA ³, os alimentos ricos em fibras foram incluídos na categoria de alimentos funcionais, pois a sua utilização dentro de uma dieta equilibrada pode reduzir o risco de algumas doenças, como as coronarianas e alguns tipos de câncer. O reconhecimento das fibras dietéticas como importantes constituintes da alimentação deve ser considerado para a maioria dos alimentos, pois além de serem relevantes devido às funções já descritas anteriormente, fazem parte dos itens exigidos para a rotulagem nutricional obrigatória de alimentos embalados descritos na Resolução 360 de 23/12/2003 da Agência Nacional de Vigilância Sanitária (ANVISA) ⁴. Além disso, quando presentes nos alimentos, sua concentração é utilizada para o cálculo dos carboidratos e das calorias totais.

O método de determinação de FD, método oficial, é diferente das determinações por fibra bruta, fibra detergente ácido e fibra detergente neutro, que não quantificam todos os componentes citados como fibras. Algumas tabelas nutricionais apresentam dados obtidos por métodos que quantificam fibra bruta, cujo resultado gerado é muito distante do valor real quantificado por métodos enzimáticos gravimétricos. Como já definido anteriormente, fibra dietética é resistente à hidrólise (digestão) pelas enzimas alimentares humanas, mas parcialmente ou totalmente fermentadas no intestino grosso. Com base nesta definição, metodologias apropriadas foram desenvolvidas e validadas através de estudos laboratoriais internacionais, e hoje se tornaram métodos padrões da AOAC Internacional (Métodos oficiais 985.29, 991.43 e 994.13) e AACC Internacional (Métodos 32-07, 32-21 e 32-05) ⁵. Com o avanço das pesquisas em fibras dietéticas, foi reconhecido que alguns componentes são subestimados ou não são quantificados, por este motivo foram criados métodos para quantificação específica de amido resistente (AOAC 2002.02), oligossacarídeos não digestíveis (NDO), o que inclui oligossacarídeos tipo inulina (fruto-oligossacarídeos, FOS) (AOAC 997.08, 999.03), galacto-oligossacarídeos (GOS) (AOAC 2001.02), polidextrose (PDX) (AOAC 2000.11) e maltodextrina resistente (RMD) (AOAC 2001.03) ⁶.

A determinação do conteúdo de fibras alimentares é realizada pelo método oficial gravimétrico-enzimático ⁶, que simula a digestão humana submetendo a amostra às três enzimas digestivas, α -amilase, amiloglicosidase e protease, em suas respectivas temperaturas e pHs ideais. Após a digestão, é realizada a análise do teor de proteínas e cinzas para subtração no cálculo de fibras resultantes do experimento. Diante disto, a completa realização da análise requer muitas etapas, o que gera uma série de erros sistemáticos e uso de reagentes químicos, totalizando 5 dias.

Uma alternativa aos métodos laboriosos de determinação de fibras é a técnica de espectroscopia de infravermelho próximo por reflectância difusa (NIRS).

Técnica de NIRS associada à quimiometria tem sido aplicada para medidas diretas de diferentes constituintes em grãos, sendo uma ótima alternativa aos métodos tradicionais de análises, pois não requer tempo, reagentes químicos e custo de análise, além de pouco ou nenhum preparo de amostra. Diversos estudos apontam sua viabilidade para o desenvolvimento de modelos de calibração para quantificação de amido, proteína, umidade e fibras alimentares em grãos e produtos da agricultura ^{7,8}. Hollung et al. ⁹ avaliaram 30 genótipos de soja da área de Novi Sad, Vojvodina, Sérvia e Montenegro por NIRS para predição do conteúdo de polissacarídeos não amiláceos e oligossacarídeos e reportaram a viabilidade desta técnica não destrutiva. Kays et al. ¹⁰ comprovaram a precisão, o baixo custo por amostra e a rapidez do método por NIR para determinar fibras dietéticas em cevada. Os autores obtiveram melhores resultados na cevada moída em função da disponibilidade de maior informação na região espectral. Haughey et al. ¹¹ obtiveram excelentes modelos de calibração e predição por NIRS para detectar adulteração de melamina em ração animal produzida com soja (Irlanda do Norte).

A soja é reconhecida pelo seu conteúdo de proteína e óleo, porém é também boa fonte de fibras dietéticas e isoflavonas. Suas fibras possuem boas propriedades nutricionais e funcionais. Possuem sabor neutro e ausência de cor, o que as torna interessantes para incorporação em alimentos sem qualquer alteração na qualidade, ao contrário das fibras de trigo. Redondo-Cuenca et al. ¹² encontraram o teor de 24,37g/100 g de matéria seca de soja da Espanha, pelo método enzimático-gravimétrico. Para aumentar o conteúdo de fibras e componentes bioativos na dieta, a aplicação de grãos inteiros em alimentos é capaz de contribuir para o acréscimo do teor de fibras.

Diante da importância das fibras para a dieta e rotulagem de alimentos, e ainda considerando o tempo despendido com o método enzimático-gravimétrico de análise, o presente estudo objetivou a aplicação da técnica rápida de NIRS para previsão do conteúdo de fibras dietéticas em 80 variedades de soja brasileira.

MATERIAIS E MÉTODOS

AMOSTRAS

Foram analisadas 80 variedades de soja [*Glycine max* (L.) Merrill], doadas pela Embrapa Soja (Brasil).

Os grãos de soja foram moídos em moinho portátil (modelo A11, IKA, Alemanha) e a granulometria padronizada a 0,6 mm em peneira mesh 28.

MÉTODO ENZIMÁTICO-GRAVIMÉTRICO

O método utilizado para determinação de fibras totais nas amostras de soja foi o enzimático-gravimétrico, adaptado dos procedimentos fornecidos pela Sigma Aldrich, que se baseia na AOAC ¹³, além de Instituto Adolfo Lutz ¹⁴ e o método 991.43 da AOAC ⁶ que se refere ao método de Prosky et al. ¹⁵. A metodologia consiste na digestão enzimática das amostras previamente secas e desengorduradas em quadruplicata. Os cadinhos tipo Gooch, com placa sinterizada, utilizados para a análise, são submetidos ao detergente Extran® 2%, solução de HCl 0,5 M, enxaguados em abundância e mantidos em mufla a 550°C por 5 horas. Esta vidraria é utilizada na etapa de filtração do resíduo obtido após digestão. Para a execução da etapa de digestão, fez-se uso de três enzimas: α -amilase, protease e amiloglicosidase, utilizadas sequencialmente por 30 minutos cada, em suas respectivas temperaturas e pHs ideais: 100°C pH 6,0, 60°C pH 7,5 e 60°C pH 4,5. Os polissacarídeos remanescentes da digestão foram precipitados com Etanol 95% por uma noite e o precipitado foi filtrado nos cadinhos envolvidos por uma cama de celite. O material foi seco em estufa e posteriormente, foram determinadas cinzas e proteínas. Ao final da análise, com os valores de massa de amostra seca, teor de cinzas, de proteínas e de lipídios, encontrou-se a porcentagem de fibras totais.

ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO

Os grãos moídos de soja foram analisados diretamente por reflectância difusa, acessório NIRA (Perkin Elmer, Estados Unidos), na radiação do infravermelho próximo por transformada de Fourier no equipamento FT-NIR Spectrum 100N (Perkin Elmer, Estados Unidos), região entre 10000 a 4000 cm^{-1} com 4 cm^{-1} de resolução e 64 varreduras. Os espectros foram obtidos em duplicata.

ANÁLISES QUIMIOMÉTRICAS

Os espectros obtidos por NIR foram tratados pelo programa PLS Toolbox V. 5.8 (Eigenvector Research Inc., Estados Unidos) desenvolvido para ambiente Matlab 7.8 (MathWorks). Os dados foram pré-tratados utilizando MSC (correção do espalhamento multiplicativo), SNV (transformação padrão normal de variação), ou primeira e segunda derivadas (método Savitzky-Golay) e todos os dados espectrais centrados na média ¹⁶. A calibração foi realizada por regressão por mínimos quadrados parciais (PLS) ¹⁷.

O número de variáveis latentes PLS foi determinado por validação cruzada, sendo utilizado o modelo que produziu o menor RMSECV (raiz quadrada do erro médio quadrático da validação cruzada).

Os *outliers* foram identificados e retirados quando necessário através dos gráficos $Q_{\text{residual}} \times T^2$ e de *leverage*. Os resíduos correspondem aos desvios entre os dados de referência e os estimados pelo modelo. O valor de *leverage* pequeno indica que as amostras se encontram próximas do centro dos dados e conseqüentemente tem pouca influência na construção do modelo.

Para todos os modelos foi escolhida a *cross validation* e a validação externa, com um grupo de valores que não participaram da calibração. O grupo de amostras de calibração e o grupo de amostras da validação foram escolhidos pelo algoritmo Kennard-Stone. O desempenho do modelo final foi avaliado de acordo com os menores resultados e RMSEC, RMSEP, RMSECV (raiz quadrada do erro médio quadrático de calibração, predição e validação cruzada), do maior valor de R^2 (coeficiente de determinação) ao plotar os valores previstos pelo modelo PLS e os valores de referência. Além de *bias*, que indica a presença de erros sistemáticos¹⁶⁻¹⁹.

RESULTADOS E DISCUSSÃO

ANÁLISE ENZIMÁTICA-GRAVIMÉTRICA

A composição de fibras totais (FA) da soja estabelecida neste trabalho e de outros grãos, obtidos na literatura para comparação, estão apresentados na Tabela 1. Pertencente ao grupo de fibras dietéticas está o amido resistente, não absorvido no intestino delgado, e quando presente em elevadas concentrações em cereais e leguminosas é capaz de reduzir o índice glicêmico²⁰.

A faixa de FA encontrada neste estudo para soja foi de 10,63 a 19,50%. Foram reportadas faixas de 5,76 a 19,71% para sorgo e de 8,5 a 11,9% para algumas variedades de trigo. O sorgo, o arroz e a soja apresentaram o maior conteúdo de fibras totais e o trigo apresenta o menor conteúdo devido à retirada do farelo ou outras frações do grão no preparo da farinha. Estudos comprovam que o uso de grão inteiro no preparo de alimentos eleva o conteúdo de fibras totais²¹. No resultado apresentado por Escarnot et al.²² utiliza-se o trigo inteiro, por este motivo o conteúdo de fibras foi elevado em relação ao resultado de Ragaee et al.²¹. O mesmo ocorreu com as amostras de soja avaliadas neste trabalho. O valor aqui reportado se assemelha ao conteúdo encontrado no trabalho de Redondo e Cuenca¹².

Tabela 1. Resultados de fibras alimentares totais para soja e outros grãos

<i>Fibras em grãos</i>	<i>Média ± DP (g/100g)</i>
Soja ^a	15,42 ± 1,57
Soja ^b	~ 16,38
Cevada ^c	9,47 ± 2,59
Trigo ^d	4,59 ± 0,51
Painço ^d	14,95 ± 0,41
Arroz ^d	17,77 ± 0,53
Sorgo ^d	21,01 ± 0,41
Trigo ^e	10,3 ± 1,40

DP = desvio padrão; ^a Resultados obtidos neste estudo; ^b Redondo e Cuenca ¹² em base seca (8,2 % de umidade); ^c Kays et al. ^{10,d}; Ragae et al. ²¹; ^e Escarnot et al. ²²

A soja analisada neste estudo possui elevado conteúdo de fibras dietéticas (15,42%), as quais foram quantificadas dentro do conteúdo de carboidratos apresentados em trabalho anterior (13,34 a 27,5%), uma vez que nesta composição estão presentes os açúcares, as ligninas, β-glucanas, celulosas etc. Além disso, esta leguminosa é considerada fonte de proteínas (em torno de 40%) e óleo (em torno de 20%) ²³.

No Brasil, a Portaria nº 27 de 1998, da Secretaria Nacional de Vigilância Sanitária (ANVISA) ²⁴, estabelece que um alimento pode ser considerado fonte de fibras alimentares (FA) quando apresentar 3 g FA/100g produto e rico em fibras com teores acima de 6 g FA/100g de produto. Deste modo, estes cereais e leguminosas são de grande importância nutricional. Vários autores, baseados em estudos epidemiológicos, ratificam a recomendação de consumo destes alimentos na prevenção de doenças crônicas não transmissíveis, incluindo câncer de trato gastrointestinal ^{1,25}.

CALIBRAÇÃO MULTIVARIADA

Na Figura 1 estão apresentados os espectros obtidos para as 80 amostras de soja analisadas por NIR. Na Figura 2 estão os coeficientes de regressão para o modelo de calibração, centrado na média, construído com os dados obtidos para o teor de fibras alimentares.

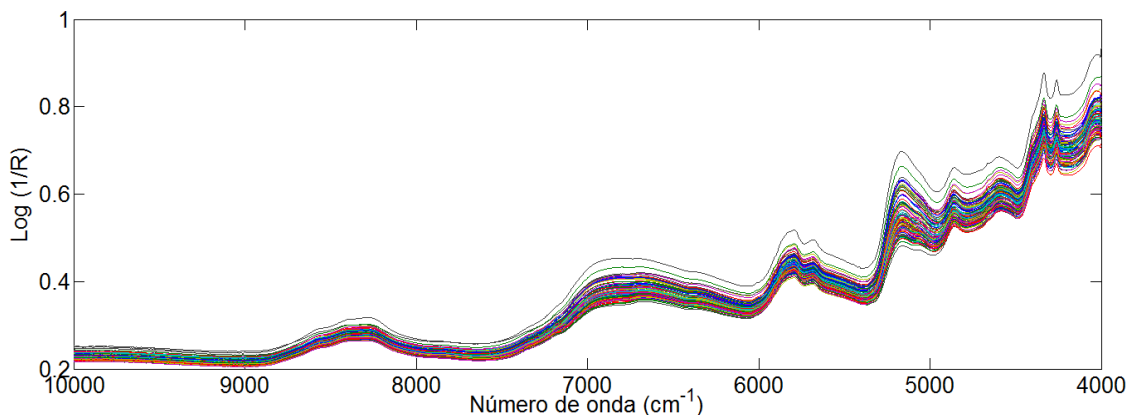


Figura 1. Espectros NIR sem pré-tratamento obtidos da soja moída.

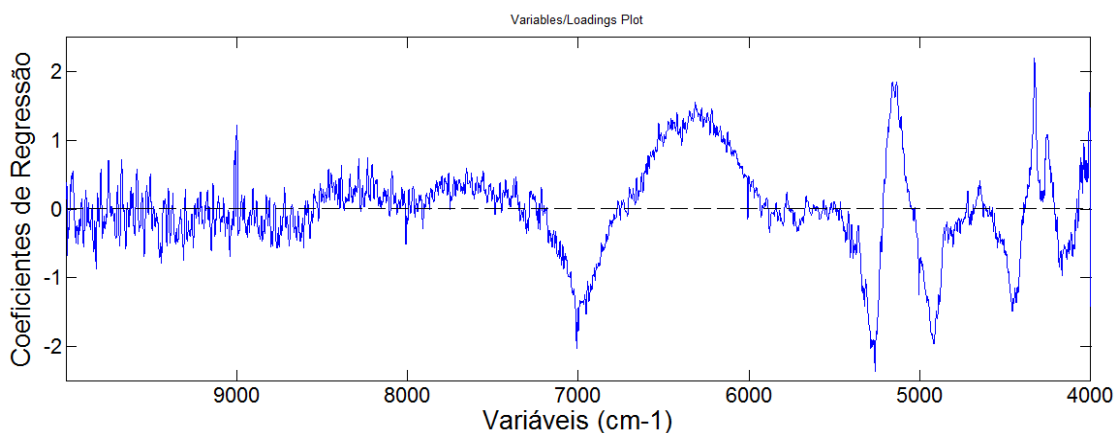


Figura 2. Coeficientes de regressão (“loadings”) para o modelo de calibração construído com o conteúdo de fibras alimentares centrado na média.

Inicialmente foi construído um modelo com a utilização de toda a faixa espectral. Entretanto, a partir da Figura 2 foi possível observar o peso (“loading”) das variáveis (números de onda) no modelo de calibração construído para fibras alimentares. Até 7000 cm^{-1} há muito ruído no espectro e nas faixas $7000\text{-}6000$, $5200\text{-}4900$, $4500\text{-}4000\text{ cm}^{-1}$ observa-se sinais de maior importância sobre o modelo. Portanto, foram selecionadas as variáveis 7200 a 6000 e 5500 a 4000 cm^{-1} para o novo modelo de calibração.

Kays et al.¹⁰ também encontraram nas regiões entre 6000 a 4000 cm^{-1} as maiores variações de absorção, portanto, essa região foi selecionada por conter as informações mais importantes para o desenvolvimento de modelo de calibração, para fibras dietéticas totais, em cevada por reflectância difusa (NIR) e

obtiveram erro padrão de validação cruzada de 10,4 g/kg e R^2 de 0,82, indicando suficiente confiança para rejeição e seleção de cultivares com elevado conteúdo de fibras.

Os espectros NIR têm sido calibrados para a maioria dos componentes fibrosos com sucesso e tentativas de associação de bandas espectrais às fibras são referentes às estruturas de celulose, pectina e lignina, como pode ser observado na Tabela 2.

Tabela 2. Tentativa de associar bandas para celulose e lignina

Tipo de material	Bandas que podem ser associadas	Comprimento de onda (cm^{-1})
Celulose	1º sobretom do estiramento O-H	6711 (1490 nm)
	1º sobretom do estiramento C-H	5618 (1780 nm)
	Estiramento O-H e 2º sobretom do estiramento C-O	5495 (1820 nm)
	Estiramento C-H e deformação C-H	4283 (2335 nm)
	Estiramento simétrico CH_2 e deformação $=\text{CH}_2$	4261 (2347 nm)
	2º sobretom da deformação CH_2	4252 (2352 nm)
	Combinação do estiramento C-H/estiramento C-C	4019 (2488 nm)
	Lignina	2º sobretom do estiramento C-H
1º sobretom do estiramento O-H		7092 (1410 nm)
Combinação do estiramento C-H		7057 (1417 nm)
1º sobretom do estiramento O-H		7042 (1420 nm)
Combinação do estiramento C-H		6944 (1440 nm)
1º sobretom do estiramento C-H		5935 (1685 nm)

Fonte: Shenk et al.²⁶.

Os melhores parâmetros quimiométricos para a construção do modelo de calibração para fibras dietéticas em soja estão apresentados na Tabela 3.

Os valores de RMSEC, RMSEP e RMSECV (Tabela 3) estão abaixo de 2%, portanto, foi possível a predição do conteúdo de fibras em soja. Porém, comparando os valores previstos pelo modelo construído por NIR e os valores

determinados em laboratório, estes apresentaram relação de 0,67. Isto se deve à metodologia tradicional de análise que possui uma série de erros sistemáticos, por se tratar de um método complexo, principalmente com muitas etapas envolvidas.

Tabela 3. Parâmetros quimiométricos do modelo de calibração para fibras totais

Parâmetro	pré-tratamento	VL	RMSEC	RMSEP	RMSECV	bias	R ²
Fibra alimentar	Primeira derivada e SNV	7	0,20	1,79	1,86	-1,77	0,67

VL = variáveis latentes; RMSE = raiz quadrada do erro médio quadrático de calibração, predição e validação cruzada; R² = coeficiente de determinação.

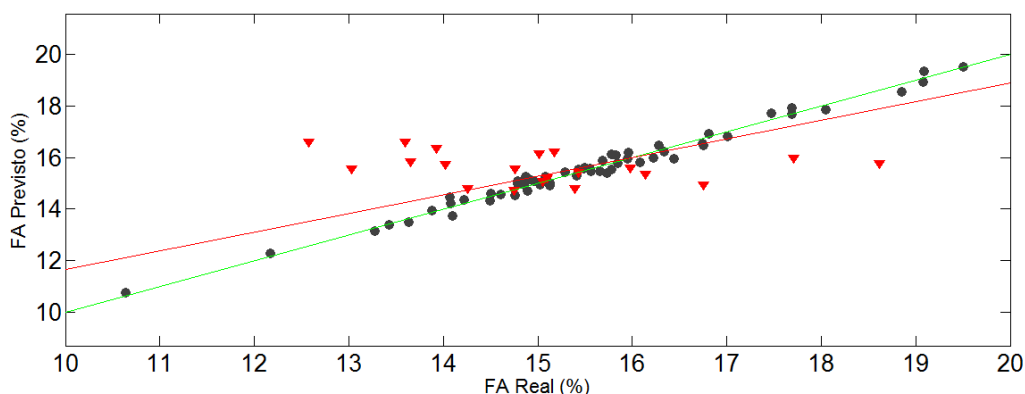


Figura 3. Relação entre os valores previstos através do espectro NIR e os valores medidos em laboratório para fibra alimentar (FA) de soja. Calibração (●), validação (▼)

De acordo com a Figura 3, houve proximidade entre os valores calibrados e previstos pelo método de NIR. Entretanto, o grupo de validação apresentou maior dispersão em relação aos valores medidos em laboratório (Tabela 4).

Dessa forma, foi constatada a potencialidade da técnica de NIR para a previsão do conteúdo de fibras em soja. Vale ressaltar que a leitura requer apenas 2 minutos e mais alguns minutos para entrega dos resultados, uma vez que o

modelo de calibração foi previamente construído por um quimiometrista. Mais estudos são necessários para comprovar a robustez do método.

Tabela 4. Comparação entre o valor medido e o valor previsto (grupo de validação) pelo modelo de calibração para o conteúdo de FD.

<i>Amostra</i>	<i>Valor medido</i>	<i>Valor previsto</i>	<i>Erro relativo (%)</i>
1	13,03	15,58	-19,53
2	18,62	15,84	14,90
3	15,18	16,25	-7,08
4	13,92	16,49	-18,46
5	15,98	15,60	2,34
6	14,75	14,77	-0,10
7	15,39	14,88	3,30
8	17,71	16,04	9,46
9	13,60	16,55	-21,72
10	12,58	16,63	-32,20
11	13,65	15,87	-16,26
12	16,14	15,49	4,03
13	15,43	15,53	-0,66
14	16,75	15,09	9,92
15	14,02	15,73	-12,17
16	15,02	16,20	-7,88
17	15,11	15,09	0,07
18	14,25	14,94	-4,81
19	15,05	15,03	0,14
20	14,76	15,46	-4,80

Erro relativo é calculado como a diferença entre o valor medido e o valor previsto pelo modelo versus o valor medido.

CONCLUSÃO

O método oficial enzimático-gravimétrico utilizado para determinação de fibras dietéticas requer dias para a conclusão da análise e apresentação do resultado final. Ficou comprovado neste estudo que a técnica de infravermelho próximo associada à quimiometria é capaz de prever o teor de fibras dietéticas em soja em menor tempo e sem reagentes químicos.

AGRADECIMENTOS

Os autores agradecem à FAPESP (processo nº 2010/50418-3) pelo auxílio financeiro.

REFERÊNCIAS BIBLIOGRÁFICAS

- (1) Gonçalves, M. C. R.; Costa, M. J. C. C.; Ascitti, L. S. R.; Diniz, M. F. F. M. *Revista Brasileira de Nutrição Clínica* **2007**, 22, 167–173.
- (2) FAO Food and Agricultural Organization *Codex Alimentarius commission*; 2009; p. 90.
- (3) Administration, F. U. S. F. and D. Health Claims Meeting Significant Scientific Agreement (SSA) - Federal Register Final Rule 73 FR 47828 August 15, 2008: Food Labeling: Health Claims; Soluble Fiber From Certain Foods and Risk of Coronary Heart Disease <http://www.fda.gov/Food/LabelingNutrition/LabelClaims/HealthClaimsMeetingSignificantScientificAgreementSSA/ucm074245.htm> (accessed Nov 15, 2012).
- (4) ANVISA Agência Nacional de Vigilância Sanitária Resolução RDC nº 360 de 23/12/2003 http://www.anvisa.gov.br/legis/resol/2003/rdc/360_03rdc.htm (accessed Nov 15, 2012).
- (5) International, A. AACC International: St. Paul, MN, USA, 1999.
- (6) AOAC In *AOAC International*; Horwitz, W.; Latimer, G. W., Eds.; Gaithersbrug, Maryland, 2006; Vol. 18th Ed.
- (7) Huang, H.; Yu, H.; Xu, H.; Ying, Y. *Journal of Food Engineering* **2008**, 87, 303–313.
- (8) Salgó, A.; Gergely, S.; Gebruers, K.; Shewry, P. R.; Ward, J. L. In *HEALTHGRAIN methods: analysis of bioactive components in small grain cereals*; Shewry, P. R.; Ward, J. L., Eds.; American Association of Cereal Chemists, Inc (AACC), 2009; pp. 247–261.
- (9) Hollung, K.; Øverland, M.; Hrustić, M.; Sekulić, P.; Miladinović, J.; Martens, H.; Narum, B.; Sahlstrøm, S.; Sørensen, M.; Storebakken, T.; Skrede, A. *Journal of agricultural and food chemistry* **2005**, 53, 9112–21.
- (10) Kays, S. E.; Shimizu, N.; Barton, F. E.; Ohtsubo, K. *Crop Science* **2005**, 45, 2307.

- (11) Haughey, S. a.; Graham, S. F.; Cancouët, E.; Elliott, C. T. *Food Chemistry* **2012**, *In Press*, 1–5.
- (12) Redondo-Cuenca, A.; Villanueva-Suárez, M. J.; Mateos-Aparicio, I. *Food Chemistry* **2008**, 108, 1099–1105.
- (13) In *AOAC International*; Horwitz, W.; Latimer, G. W., Eds.; AOAC International: Arlington, VA, 1990; Sec. 985.29 1105.
- (14) Instituto Adolfo Lutz *Métodos físico-químicos para análise de alimentos*; Zenebon, O.; Pascuet, N. S., Eds.; 4^a ed.; São Paulo, 2008; p. 1020.
- (15) Prosky, L.; Asp, N. G.; Furda, I.; DeVries, J. W.; Schweizer, T. F.; Harland, B. F. *Journal - Association of Official Analytical Chemists* **1985**, 68, 677–9.
- (16) Agelet, L. E.; Hurburgh, C. R. *Critical Reviews in Analytical Chemistry* **2010**, 40, 246–260.
- (17) Geladi, P.; Kowalski, B. R. *Analytica Chimica Acta* **1986**, 185, 1–17.
- (18) Naes, T.; Isaksson, T.; Fearn, T.; Davies, T. In *A User-Friendly Guide to Multivariate Calibration and Classification*; NIR Publications: Chichester, UK, 2002; pp. 177–189.
- (19) Brereton, R. G. *The Analyst* **2000**, 125, 2125–2154.
- (20) Englyst, K. N.; Vinoy, S.; Englyst, H. N.; Lang, V. *The British journal of nutrition* **2003**, 89, 329–40.
- (21) Ragaei, S.; Abdelaal, E.; Noaman, M. *Food Chemistry* **2006**, 98, 32–38.
- (22) Escarnot, E.; Agneessens, R.; Wathélet, B.; Paquot, M. *Food Chemistry* **2010**, 122, 857–863.
- (23) Ferreira, D. S.; Pallone, J. A. L.; Poppi, R. J. *Food Research International* **2012**.
- (24) ANVISA Agência Nacional de Vigilância Sanitária Portaria nº 27 http://www.anvisa.gov.br/legis/portarias/27_98.htm (accessed Nov 15, 2012).
- (25) Dreher, M. L.; Cho, S. S. *Handbook of Dietary Fiber*; CRC Press, 2001.
- (26) Shenk, J. S.; Workman, J. J. J.; Westerhaus, M. O. In *Handbook of near-infrared analysis*; Burns, D. A.; Ciurczak, emil W., Eds.; CRC Press, 2007; pp. 347–386.

CONCLUSÃO GERAL

A espectroscopia no infravermelho é uma técnica analítica que não requer preparo de amostra, basta uma pequena quantidade após amostragem e não provoca destruição da mesma. Os resultados espectrais são obtidos em menos de 3 minutos e, tanto a espectroscopia no infravermelho próximo como médio, mostraram-se eficientes para previsão da composição centesimal em soja. Porém, concluiu-se que o infravermelho próximo apresentou resultados mais robustos quando foi aplicado um número grande de amostras e com o emprego do acessório de reflectância difusa.

Os dados obtidos com amostras de soja brasileira, proveniente de um banco de germoplasma, demonstraram alta variabilidade em sua composição, o que permitiu a geração de modelos de calibração mais precisos. Além disso, um controle de qualidade rápido permite a destinação adequada desta leguminosa para determinadas finalidades alimentícias, tais como o aproveitamento do elevado conteúdo de óleo e produtos ricos em proteína de soja.

A quinoa possui elevado valor nutricional. Através da espectroscopia no infravermelho próximo, aliado à quimiometria, é possível prever o conteúdo da composição centesimal de diversas variedades. A construção do modelo de calibração que inclui variedades de diversos países permite um controle da qualidade deste rico grão sem a necessidade de aplicação de modelos diferentes. Além disso, o grão inteiro pode ser avaliado, sem a necessidade de tratamento da amostra e no local de beneficiamento, já que o grão não é danificado.

A determinação do conteúdo de fibras alimentares por infravermelho próximo é uma forma de substituição dos métodos tradicionais, além da economia de tempo e solventes orgânicos.

O Infravermelho próximo pode ser implantado em diversas indústrias para controle de qualidade de alimentos de forma rápida, sem uso de solventes químicos e com elevada precisão.