



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

NIR spectral techniques and chemometrics applied to food processing

Author:

AMANDA TEIXEIRA BADARÓ

Supervisors:

Professor Nuria Aleixos Borràs (Universitat Politècnica de València)

Professor Douglas Fernandes Barbin (University of Campinas)

November - 2021

AMANDA TEIXEIRA BADARÓ

Técnicas espectrais de infravermelho próximo (NIR) e quimiometria aplicadas ao processamento de alimentos

NIR spectral techniques and chemometrics applied to food processing

Técnicas espectrales de infrarrojo cercano (NIR) y quimiometría aplicadas al procesamiento de alimentos

Tese de doutorado apresentada ao Programa de Pós-Graduação em Engenharia de Alimentos da Faculdade de Engenharia de Alimentos - Universidade Estadual de Campinas e à Escola de Doutorado em Ciência, Tecnologia e Gestão de Alimentos da Universidade Politécnica de Valência, como um dos requisitos para a obtenção do título de Doutora em Engenharia de Alimentos e em Ciência, Tecnologia e Gestão de Alimentos. Tese produzida no âmbito de um acordo de Cotutela firmado entre a Unicamp e a Universidade Politécnica de Valência.

Thesis presented to the Postgraduate Program in the School of Food Engineering of the University of Campinas and in the Doctorate School of Food Science, Technology and Management in partial fulfilment of the requirements for the degree of Doctor in Food Engineering and Food Science, Technology and Management. Thesis produced under a Cotutelle Agreement signed between Unicamp and the Polytechnic University of Valencia.

Tesis de doctorado presentada al Programa de Postgrado en Ingeniería de Alimentos de la Facultad de Ingeniería de Alimentos - Universidad Estadual de Campinas y a la Escuela de Doctorado en Ciencia, Tecnología y Gestión Alimentaria de la Universitat Politècnica de València, como uno de los requisitos para la obtención del título de Doctora en Ingeniería de Alimentos y en Ciencia, Tecnología y Gestión Alimentaria. Tesis producida en virtud de un acuerdo de cotutela firmado entre la Unicamp y la Universitat Politècnica de València.

Supervisors/Orientadores:

Professor Douglas Fernandes Barbin and Professor Nuria Aleixos Borràs

Este trabalho corresponde à versão final da tese defendida pela aluna Amanda Teixeira Badaró, e orientada pelo professor Dr. Douglas Fernandes Barbin e a professora Dra. Nuria Aleixos Borràs. / This copy corresponds to the final version of the thesis defended by the student Amanda Teixeira Badaró and supervised by Professor Douglas Fernandes Barbin and Professor Nuria Aleixos Borràs.

November - 2021

FICHA CATALOGRÁFICA

COMMITTEE

Professor Douglas Fernandes Barbin

SUPERVISOR– DEA/FEA/UNICAMP

Professor Nuria Aleixos Borràs

SUPERVISOR– SCHOOL OF INDUSTRIAL ENGINEERING/DEPARTMENT OF GRAPHIC ENGINEERING/UPV

Professor María del Carmen Alamar Gavidia

TITULAR MEMBER - PLANT SCIENCE LABORATORY/CRANFIELD UNIVERSITY

Professor José Javier Benedito Fort

TITULAR MEMBER - DEPARTMENT OF FOOD TECHNOLOGY/UPV

Professor Angel Antonio Carbonell Barrachina

TITULAR MEMBER - DEPARTMENT OF AGRO-FOOD TECHNOLOGY/CONSELLERIA DE INNOVACIÓN, UNIVERSIDADES, CIENCIA Y SOCIEDAD DIGITAL

Professor Wouter Saeys

TITULAR MEMBER - DEPARTMENT OF BIOSYSTEMS/ KU LEUVEN

Professor Paul James Williams

ALTERNATE MEMBER - DEPARTMENT OF FOOD SCIENCE/STELLENBOSCH UNIVERSITY

Professor Pilar Barreiro Elorza

ALTERNATE MEMBER - DEPARTMENT OF AGROFORESTRY ENGINEERING/ POLYTECHNIC UNIVERSITY OF MADRID

Professor Delia Rita Tapia Blácido

ALTERNATE MEMBER - DEPARTMENT OF CHEMISTRY/FACULTY OF PHILOSOPHY, SCIENCE AND LETTERS/USP-RIBEIRÃO PRETO

A Ata de Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade. / The minutes of defense with the members' respective signatures are on the students' academic report.

*Dedico esta tese aos meus pais, meu porto seguro, e ao meu
noivo, minha calma.*

ACKNOWLEDGMENT

À Deus sobre todas as coisas, por sempre me abrir caminhos e me proporcionar tudo que tenho vivido.

Aos meus pais por me amarem incondicionalmente, por sempre me apoiarem, incentivarem, acreditarem e investirem em mim. Obrigada por sempre me proporcionarem tudo de melhor que eu podia ter. Nunca terei palavras suficientes para descrever minha gratidão por ser sua filha. Tudo que sou devo a vocês! Obrigada, eu amo vocês!

Ao meu noivo, parceiro e melhor amigo, Josi, quem sempre confiou no meu potencial e me incentivou a correr atrás dos meus sonhos, mesmo que isso me levasse para longe dele. Obrigada por tudo meu amor, muito do que eu vivi durante minha vida acadêmica devo a você. Obrigada por ter sido minha calma em muitos momentos de tormento.

Ao professor Douglas Fernandes Barbin, pela orientação, conhecimentos compartilhados e pela amizade. Sou extremamente grata por todas as oportunidades que me proporcionou. Em um período de tanta pressão na vida de um aluno de doutorado, só consigo pensar que tive a sorte de ter encontrado, além de um mentor, um exemplo de dedicação e profissionalismo. Obrigada por sempre me incentivar. O mundo acadêmico precisa de mais professores/orientadores como você!

Ao LINA e todos seus integrantes pelo convívio. Em especial às minhas amigas Bianca, Juliana, Fernanda, Bruna e Maria, por todo carinho, conselhos e conversas. Obrigada pelos lanches, pirulitos e risadas que compartilhamos todo esse tempo, vocês fizeram esse convívio muito mais leve. Em especial também ao meu amigo Pier, pelas infinitas conversas por áudios de 30 minutos no whatsapp, nas quais eu sempre falava e era sempre humilhada, hahahaha. À minha aluna de iniciação científica, Fernanda Morimitsu, por me ajudar com meus experimentos e me permitir lhe ensinar um pouco do que aprendi no doutorado.

A todos os meus amigos, que sempre me mostraram que eu sou capaz, sempre me apoiaram, deram conselhos e puxões de orelha. Em especial à minha amiga Ellen, que sempre me socorria quando o código do matlab não rodava ou quando eu ainda era a manquinha de Taubaté e precisava de alguém para me levar para a fisioterapia. Ao meu amigo Manoel, que sempre me ajudou com os problemas de TI hahahaha. À minha amiga e professora, Marcela que sempre vibrou as minhas vitórias. À minha amiga Ana Gabriela, que sempre teve; uma palavra de carinho ou informações de um amigo que teve bolsa FAPESP para me oferecer. À minha amiga Débora, por todo carinho e ajuda em um momento tão delicado da minha vida. À minha amiga Tati, por toda paciência e palavras de conforto em momentos difíceis e por toda diversão e comilança também.

À Faculdade de Engenharia de Alimentos (FEA) e a Universidade Estadual de Campinas (UNICAMP) pela oportunidade e por fornecer toda a estrutura para a realização do doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). À Fundação de Amparo à

Pesquisa do Estado de São Paulo (FAPESP) (Projetos números 2017/17628-3 e 2019/06842-0 pelo apoio financeiro. À Nutrassim pelo fornecimento de parte das amostras.

Ao Laboratório de Cereais, Raízes e Tubérculos, à professora Maria Teresa Pedrosa Silva Clerici e à Amanda Rios Ferreira, por disponibilizarem a estrutura e o conhecimento no desenvolvimento das amostras utilizadas nesta tese, e por acreditar no projeto.

Ao professor Célio Pasquini e à Cristine Vidal, por disponibilizarem a estrutura e o conhecimento em parte dos experimentos necessários ao desenvolvimento desta tese.

Ao professor Sylvio Barbon Junior e seu grupo de pesquisa, por todo conhecimento compartilhado.

Agradezco al profesor José Manuel Amigo, por todo conocimiento compartido. Eres un gran profesional y soy muy feliz por haber tenido la oportunidad de trabajar contigo.

Al Instituto Valenciano de Investigaciones Agrarias y a todos sus miembros, por ser tan amables y haberme recibido y acogido tan bien cuando estuve en España. En especial al profesor Jose Blasco, por la supervisión y amistad. A profesora Nuria Aleixos, por la supervisión, la amistad y los cafés online. A Sergio y Henrique por toda la ayuda con la estructura de mis pruebas. A mis amigos Andrés, Carlos 1, Carlos 2 y Diego por los cafés de todas las mañanas, las cervezas de jueves por la noche y la cocción de macarrones de todos los días jajajajaja. A Manolo por el cariño y las granadas que siempre me traía. Gracias à todos por los almuerzos, los master chefs, las horchatas con fartones y los cacahuets.

To the members of the examining board for their attention, availability and suggestions for improving this thesis.

*“O período de maior ganho em conhecimento e experiência é
o período mais difícil da vida de alguém.”*

*“The period of greatest gain in knowledge and experience is
the most difficult period in one’s life.”*

Dalai Lama

RESUMO

Técnicas rápidas, não destrutivas e livres de produtos químicos estão em crescente demanda em muitas áreas da indústria. As técnicas de espectroscopia no infravermelho próximo (NIRS) e de imagem hiperespectral no infravermelho próximo (NIR-HSI) têm demonstrado grande potencial na determinação de parâmetros de qualidade de alimentos, autenticação de produtos alimentícios, detecção de fraudes alimentares, entre muitas outras aplicações. Na espectroscopia NIR, as medições são feitas em pontos específicos da amostra, detectando apenas uma pequena porção; enquanto na imagem hiperespectral, as informações espectrais e espaciais são combinadas, sendo uma escolha adequada para muitos produtos alimentícios, uma vez que são matrizes muito heterogêneas. Portanto, este estudo teve como objetivo revisar todas as aplicações de NIRS (dispersivos), Transformada de Fourier (FT) NIR e HSI na avaliação de parâmetros de qualidade da farinha de trigo e de produtos à base de trigo, bem como para a autenticação e determinação da composição desses produtos. Além disso, este trabalho teve como objetivo identificar e classificar diferentes tipos de amostras de fibras adicionadas à semolina e massas produzidas por formulações de fibras e semolina, e monitorar o processo de cozimento dessa massa enriquecida com fibras por técnicas espectrais. Além disso, este trabalho teve como objetivo a aplicação de HSI em outro produto em pó, sendo utilizada para quantificar o teor de pectina em cascas de laranja. Primeiramente, espectros NIR foram adquiridos para comparar a precisão na classificação de amostras enriquecidas com fibras, para quantificar a quantidade dessas fibras e verificar sua distribuição das fibras adicionadas à semolina. A Análise de Componentes Principais (PCA) e a *Soft Independent Modelling of Class Analogy* (SIMCA) foram usadas para a classificação. Modelos de regressão de mínimos quadrados parciais (PLSR) aplicados aos espectros NIR-HSI mostraram R^2P entre 0,85 e 0,98, e RMSEP entre 0,5 e 1% do teor de fibra, e os modelos foram usados para construir os mapas químicos das amostras. Além disso, NIR-HIS, em conjunto com *Multivariate Curve Resolution-Alternating Least Squares* (MCR-ALS), foi testado para investigar a capacidade de avaliação, resolução e quantificação da distribuição de fibras em massas alimentícias enriquecidas. Os resultados mostraram coeficiente de determinação de validação (R^2V) entre 0,28 e 0,89, % de falta de ajuste (LOF) <6%, variância explicada acima de 99% e similaridade entre espectros puros e recuperados acima de 96% e 98% em modelos usando farinha pura e macarrão controle como estimativas iniciais, respectivamente. Além disso, o VIS/NIR-HSI no modo de transmissão foi testado como uma alternativa objetiva para a classificar amostras de acordo com o tempo de cocção de massas alimentícias como forma de automatizar a determinação de

atributos de massas alimentícias. A Análise Discriminante Linear (LDA) mostrou valores de sensibilidade e especificidade entre 0,14 - 1,00 e 0,51 - 1,00, respectivamente, e taxa de não erro (NER) acima de 0,62. A Análise Discriminante por Mínimos Quadrados Parciais (PLSDA) apresentou valores de sensibilidade e especificidade entre 0,67 - 1,00 e 0,10 - 1,00, respectivamente, e NER acima de 0,80. Os resultados da primeira parte deste trabalho mostraram que a técnica NIR-HSI pode ser utilizada para a identificação e quantificação da fibra adicionada à semolina. Além disso, NIR-HSI e MCR-ALS, juntos, são capazes de identificar fibras em massas. A imagem hiperespectral no modo de transmissão demonstrou ser uma técnica adequada como alternativa objetiva para classificar amostras de macarrão de acordo tempo de cocção, como forma de automatizar a determinação dos atributos das massas. A determinação do teor de pectina em cascas de laranja foi investigada usando NIR-HSI. LDA mostrou melhores resultados de discriminação considerando três grupos: baixo (0–5%), intermediário (10–40%) e alto (50–100%) conteúdo de pectina. Modelos PLSR baseados em espectros completos mostraram maior precisão ($R^2 > 0,93$, RMSEP entre 6,50 e 9,16% da pectina) do que aqueles baseados em comprimentos de onda selecionados (R^2 entre 0,92 e 0,94, RMSEP entre 8,03 e 9,73% de pectina). Os resultados demonstram o potencial do NIR-HSI para quantificar o conteúdo de pectina em cascas de laranja, fornecendo uma técnica valiosa para produtores de laranja e indústrias de processamento.

Palavras-chave: Espectroscopia no infravermelho próximo, Imagem hiperespectral, Semolina, Fibra, Farinha de trigo, Tempo de cozimento, Pectina, Resíduos de laranja, Casca de laranja.

ABSTRACT

Fast, non-destructive and chemical-free techniques are in increasing demand in many fields of the industry. Near-infrared spectroscopy (NIRS) and NIR hyperspectral imaging (NIR-HSI) techniques have shown great potential in determining food quality parameters, authenticating food products, detecting food fraud, among many other applications. While in near infrared spectroscopy, the measurements are taken at specific points on the sample, detecting only a small portion; in hyperspectral imaging, spectral and spatial information are combined, making it a suitable choice for many food products, since they are very heterogeneous matrices. Therefore, this study aimed to review all the application of (dispersive) NIRS, Fourier Transform (FT) NIR, and HSI in assessing wheat flour and wheat-based products quality parameters, as well for the authentication and determination of composition of these products. Moreover, this work aimed to identify and classify different types of fibre samples added to the semolina and pasta produced by semolina-fibre formulations, and to monitor the cooking process of this fibre-enriched pasta by spectral techniques. In addition, this work had the aim of applying HSI to other powdered product, so the pectin content in orange peels was quantified. First, NIR spectra were acquired to compare the accuracy in the classification of fibre-enriched samples, to quantify the amount of these fibres and verify their distribution on semolina samples. Principal Component Analysis (PCA) and Soft Independent Modelling of Class Analogy (SIMCA) were used for classification. Partial Least Squares Regression (PLSR) models applied to NIR-HSI spectra showed R^2_P between 0.85 and 0.98, and RMSEP between 0.5 and 1% of fibre content, and the models were used to construct the chemical maps to check the fibre distribution on the samples surface. Moreover, NIR-HSI together with Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), was tested to investigate the ability for the evaluation, resolution and quantification of fibre distribution in enriched pasta. Results showed coefficient of determination of validation (R^2_V) between 0.28 and 0.89, % of lack of fit (LOF) <6%, variance explained over 99%, and similarity between pure and recovered spectra over 96% and 98% in models using pure flour and control as initial estimates, respectively. In addition, VIS/NIR-HSI in the transmission mode was tested as an objective alternative for the classification of pasta samples according to cooking time as way of automating the determination of pasta attributes. Linear Discriminant Analysis (LDA) showed values of sensitivity and specificity between 0.14 – 1.00 and 0.51 – 1.00, respectively, and non-error rate (NER) over 0.62. Partial Least Square Discriminant Analysis (PLSDA) showed values of sensitivity and specificity between 0.67 – 1.00 and 0.10 – 1.00, respectively, and NER over

0.80. The results of the first part of this work showed that NIR-HSI technique can be used for the identification and quantification of fibre added to semolina. Additionally, NIR-HSI and MCR-ALS are able to identify fibre in pasta. Hyperspectral imaging in the transmission mode demonstrated to be a suitable technique as an objective alternative for the classification of pasta samples according to the cooking time as a way of automating the determination of pasta attributes. Determination of pectin content in orange peels was investigated using NIR-HSI. LDA showed better discrimination results considering three groups: low (0–5%), intermediate (10–40%) and high (50–100%) pectin content. PLSR models based on full spectra showed higher precision ($R^2 > 0.93$, RMSEP between 6.50 and 9.16% of pectin) than those based on few selected wavelengths (R^2 between 0.92 and 0.94, RMSEP between 8.03 and 9.73% of pectin). The results demonstrate the potential of NIR-HSI to quantify pectin content in orange peels, providing a valuable technique for orange producers and processing industries.

Keywords: Near infrared spectroscopy, Hyperspectral imaging, Semolina, Fibre, Wheat flour, Cooking time, Pectin, Orange waste, Orange peel.

RESUMEN

Las técnicas rápidas, no destructivas y libres de químicos tienen una demanda creciente en muchos campos de la industria. Las técnicas de espectroscopia de infrarrojo cercano (NIRS) y de imágenes hiperespectrales NIR (NIR-HSI) han mostrado un gran potencial para determinar los parámetros de calidad de los alimentos, autenticar los productos alimenticios, detectar el fraude alimentario, entre muchas otras aplicaciones. Mientras que en la espectroscopia de infrarrojo cercano, las medidas se toman en puntos específicos de la muestra, detectando solo una pequeña porción; en la imagen hiperespectral, la información espectral y espacial se combinan, lo que la convierte en una opción adecuada para muchos productos alimenticios, ya que son matrices muy heterogéneas. Por lo tanto, este estudio tuvo como objetivo revisar toda la aplicación de NIRS (dispersivos), NIR de Transformada de Fourier (FT) y HSI en la evaluación de los parámetros de calidad de la harina de trigo y los productos a base de trigo, así como para la autenticación y determinación de la composición de estos productos. Además, este trabajo tuvo como objetivo identificar y clasificar diferentes tipos de muestras de fibra agregadas a la semolina y pasta producidas por formulaciones de fibra y semolina, y monitorear el proceso de cocción de esta pasta enriquecida en fibra mediante técnicas espectrales. Además, este trabajo tuvo como objetivo aplicar HSI a otro producto en polvo, por lo que se cuantificó el contenido de pectina en las cáscaras de naranja. Primero, se adquirieron espectros NIR para comparar la precisión en la clasificación de muestras enriquecidas con fibra, para cuantificar la cantidad de estas fibras y verificar su distribución en muestras de sémola. Para la clasificación se utilizaron el Análisis de Componentes Principales (PCA) y el *Soft Independent Modelling of Class Analogy* (SIMCA). Los modelos de regresión de mínimos cuadrados parciales (PLSR) aplicados a espectros NIR-HSI mostraron R^2P entre 0,85 y 0,98 y RMSEP entre 0,5 y 1% de contenido de fibra, y los modelos se utilizaron para construir los mapas químicos para verificar la distribución de fibra en las superficies de las muestras. Además, se probó el NIR-HSI junto con los *Multivariate Curve Resolution-Alternating Least Squares* (MCR-ALS) para investigar la capacidad de evaluación, resolución y cuantificación de la distribución de fibra en la pasta enriquecida. Los resultados mostraron coeficiente de determinación de validación (R^2V) entre 0.28 y 0.89, % de falta de ajuste (LOF) <6%, varianza explicada sobre 99% y similitud entre espectros puros y recuperados sobre 96% y 98% en modelos que usan harina pura y pasta control como estimaciones iniciales, respectivamente. Además, se probó VIS / NIR-HSI en el modo de transmisión como una alternativa objetiva para la clasificación de muestras de pasta según el tiempo de cocción como forma de automatizar la determinación de los atributos de la

pasta. El análisis discriminante lineal (LDA) mostró valores de sensibilidad y especificidad entre 0,14 - 1,00 y 0,51 - 1,00, respectivamente, y una tasa de ausencia de error (NER) superior a 0,62. El análisis discriminante de mínimos cuadrados parciales (PLSDA) mostró valores de sensibilidad y especificidad entre 0,67 - 1,00 y 0,10 - 1,00, respectivamente, y NER superiores a 0,80. Los resultados de la primera parte de este trabajo mostraron que la técnica NIR-HSI se puede utilizar para la identificación y cuantificación de la fibra agregada a la semolina. Además, NIR-HSI y MCR-ALS pueden identificar la fibra en la pasta. La imagen hiperespectral en el modo de transmisión demostró ser una técnica adecuada como alternativa objetiva para la clasificación de muestras de pasta según el tiempo de cocción como una forma de automatizar la determinación de los atributos de la pasta. La determinación del contenido de pectina en cáscaras de naranja se investigó usando NIR-HSI. LDA mostró mejores resultados de discriminación considerando tres grupos: contenido de pectina bajo (0–5%), intermedio (10–40%) y alto (50–100%). Los modelos PLSR basados en espectros completos mostraron mayor precisión ($R^2 > 0,93$, RMSEP entre 6,50 y 9,16% de pectina) que los basados en pocas longitudes de onda seleccionadas (R^2 entre 0,92 y 0,94, RMSEP entre 8,03 y 9,73% de pectina). Los resultados demuestran el potencial de NIR-HSI para cuantificar el contenido de pectina en las cáscaras de naranja, proporcionando una técnica valiosa para los productores de naranja y las industrias de procesamiento.

Palabras clave: Espectroscopia de infrarrojo cercano, Imagen hiperespectral, Sémola, Fibra, Harina de trigo, Tiempo de cocción, Pectina, Residuos de naranja, Cáscara de naranja.

RESUM

Les tècniques ràpides, no destructives i lliures de químics tenen una demanda creixent en molts camps de la indústria. Les tècniques d'espectroscopia d'infraroig proper (NIRS) i d'imatges hiperespectrals NIR (NIR-HIS) han demostrat tindre un gran potencial per a determinar paràmetres de qualitat d'aliments, autenticar productes alimentaris, detectar frau alimentari entre moltes altres aplicacions. Mentre que en l'espectroscopia d'infraroig proper les mesures es prenen en punts específics de la mostra i es detecta una porció menuda, en la imatge hiperespectral es combina informació espectral i espacial de tal manera que és una opció adient per a molts tipus de productes alimentaris, ja que són matrius molt heterogènies. Per tant, este estudi va tindre com objectiu revisar tota l'aplicació de NIRS (dispersius), NIR de Transformada de Fourier (FT) i HSI en l'avaluació dels paràmetres de qualitat de la farina de blat i els productes a base de blat, així com per a l'autenticació i determinació de la composició d'estos productes. A més a més, este estudi va tindre com objectiu identificar i classificar diferents tipus de mostres de fibra afegides a la semolina i pasta produïdes per formulació de fibra i semolina, i monitorar mitjançant tècniques espectrals el procés de cocció d'aquesta pasta enriquida amb fibra. A més, este treball va tindre com objectiu aplicar HSI a un altre producte en pols, de tal manera que es va quantificar el contingut de pectina en les corfes de taronja. Primer, es van adquirir espectres NIR per comparar la precisió en la classificació de mostres enriquides amb fibra, per quantificar estes fibres i verificar la seua distribució en mostres de sèmola. Per a la classificació es van emprar l'Anàlisi de Components Principals (PCA) i el SIMCA (Soft Independent Modelling of Class Analogy). Els models de regressió de mínims quadrats parcials (PLSR) aplicats a espectres NIR-HSI mostraren R²P entre 0,85 i 0,98 i RMSEP entre 0,5 i 1% de contingut de fibra, i els models s'utilitzaren per construir els mapes químics per verificar la distribució de fibra en les superfícies de les mostres. Així mateix, es va provar NIR-HSI amb Multivariate Curve Resolution-Alternating Least Square (MCR-ALS) per a investigar la capacitat d'avaluació, resolució i quantificació de la distribució de fibra en la pasta enriquida. Els resultats mostraren un coeficient de determinació de validació (R²V) entre 0,28 i 0,89%, lack of fit (LOF) <6%, variància explicada sobre 99% i similitud entre espectres purs i recuperats sobre 96% i 98% en models que empraren farina pura i pasta control com a estimacions inicials respectivament. D'altra part, es va provar VIS / NIR-HSI en el mode de transmissió com una alternativa objectiva per a la classificació de mostres de pasta segons el temps de cocció com a forma d'automatitzar la determinació dels atributs de la pasta. L'anàlisi discriminant lineal (LDA) va mostrar valors de sensibilitat i especificitat entre 0,14 – 1,00 i

0,51 – 1,00 respectivament, i una taxa d'absència d'error (NER) superior a 0,62. L'anàlisi discriminant de mínims quadrats parcials (PLSDA) va mostrar valors de sensibilitat i especificitat entre 0,67 – 1,00 i 0,10 – 1,00 respectivament, i NER superiors a 0,80. Els resultats de la primera part d'este treball mostraren que la tècnica NIR-HSI es pot emprar per a la identificació i quantificació de la fibra afegida a la semolina. A més a més, NIR-HSI i MCR-ALS poden identificar la fibra en la pasta. La imatge hiperespectral en mode de transmissió va demostrar ser una tècnica adient com a alternativa objectiva per a la classificació de mostres de pasta segons el temps de cocció com forma d'automatitzar la determinació dels atributs de la pasta. La determinació del contingut de pectina en corfa de taronja es va investigar emprant NIR-HSI. LDA va mostrar millors resultats de discriminació considerant tres grups: contingut de pectina baix (0–5%), intermedi (10–40%) i alt (50–100%). Els models PLSR basats en espectres complets van mostrar major precisió ($R^2 > 0,93$, RMSEP entre 6,50 i 9,16% de pectina) que els basats en longituds d'ona seleccionades (R^2 entre 0,92 i 0,94, RMSEP entre 8,03 i 9,73% de pectina). Els resultats demostren el potencial de NIR-HSI per a quantificar el contingut de pectina en corfa de taronja i proporcionen una tècnica valuosa per als productors de taronja i les indústries de processament.

Paraules clau: Espectroscopia d'infraroig proper, imatge hiperespectral, sèmola, fibra, farina de blat, temps de cocció, pectina, residus de taronja, corfa de taronja.

LIST OF ILLUSTRATIONS

Figure 1.1. (A) Approaches for spectra acquisition, Scheme of the operation of (B) a NIR instrument, (C) a FT-NIR instrument, and (D) a NIR-HSI instrument.....	38
Figure 2.1. Sample preparation steps.....	86
Figure 2.2. Average raw (a and b) and pre-processed spectra after SNV (c and d) and 1st Der (e and f) of each pure semolina and fiber.....	86
Figure 2.3. Score plots (PC1 vs PC2) of PCA performed on spectra after Savitzky–Golay 1st Der pre-processing (a and b) and their correspondent Loadings plots (c and d).....	87
Figure 2.4. Calculated class responses for SIMCA models. Calibration models: (a), (c), (e), (g) and (i) and validation models: (b), (d), (f), (h) and (j).....	89
Figure 2.5. Predicted vs. measured plot for the PLSR models obtained on the raw spectra. (a) Fiber 1, (b) Fiber 2, (c) Fiber 3, (d) Fiber 4 and (e) Fiber 5 added to semolina.....	90
Figure 2.6. Prediction maps for visual identification of fiber added to semolina using PLSR models. The dark blue image represents pure semolina, otherwise fiber.....	91
Figure 3.1. NIR spectra of the control, pure flour, and pure fiber samples.....	95
Figure 3.2. Concentration maps some of the samples studied in this work. The color represents the percentage of fiber in each pixel of the image.....	96
Figure 3.3. Pure and recovered spectra from fiber and control sample (Data preprocessed with Standard Normal Variate – SNV).....	97
Figure 3.4. Identification of fiber granules in Sample 3, comparing the 3 thresholds. A) Sample with 2% of fiber, B) Sample with 3.5% of fiber, C) Sample with 5% of fiber, and D) Sample with 7% of fiber.....	98
Figure 3.5. Expected versus Predicted fiber content after MCR-ALS analysis (Models highlighted in dark grey in Table 3.1 for the control sample as the initial estimate).....	100
Figure 4.1. Scheme of the equipment used.....	109
Figure 4.2. Average raw spectra of NIR (A to G) and VIS (H to N) of pasta samples.....	112
Figure 4.3. Pasta samples in each cooking time: Pasta 1 (A – M) and Pasta 7 (N – Z), from time 0 to 18 minutes, respectively.....	114
Figure 4.4. Score plots (PC1 vs PC2) of PCA performed on NIR spectra of Pasta 1 after Savitzky–Golay 1st derivative and the correspondent loadings plot.....	115
Figure 4.5. Score plots (PC1 vs PC2) of PCA performed on NIR spectra of all samples after Savitzky–Golay 1st and 2nd derivative and the correspondent loadings plot.....	117

Figure 5.1. Spectra of orange peels and pectin: a) Smoothing, b) Smoothing + SNV, c) Smoothing + 1st derivative, d) Smoothing + 2nd derivative, e) Smoothing + MSC, f) Smoothing + MSC + 2nd derivative, g) Smoothing + SNV + 1st derivative.....	126
Figure 5.2. PCA scores and loadings of orange peels and pectin: (a and e) Smoothing + 2nd derivative, (b and f) Smoothing + MSC + 2nd derivative, (c and g) Smoothing + SNV + 1st derivative, (d and h) Smoothing + 1st derivative.....	127
Figure 5.3. PCA scores of orange peels and pectin after variable selection: a) Smoothing + 2nd derivative, b) Smoothing + MSC + 2nd derivative, c) Smoothing + SNV + 1st derivative, d) Smoothing + 1st derivative.....	128
Figure 5.4. Predicted vs. reference plot for the PLSR models obtained with full spectra: a) Smoothing, b) Smoothing + SNV, c) Smoothing + 1st derivative, d) Smoothing + 2nd derivative, e) Smoothing + MSC, f) Smoothing + MSC + 2nd derivative, g) Smoothing + SNV + 1st derivative.....	129
Figure 5.5. Regression coefficients for the PLSR models obtained with full spectra: a) Smoothing, b) Smoothing + SNV, c) Smoothing + 1st derivative, d) Smoothing + 2nd derivative, e) Smoothing + MSC, f) Smoothing + MSC + 2nd derivative, g) Smoothing + SNV + 1st derivative.....	130

LIST OF TABLES

Table 1.1. Equations for models performance calculation.....	46
Table 1.2. Literature related to the adulteration, contamination, and addition of ingredients in wheat flour and wheat-based products.....	54
Table 1.3. Literature related to the determination of quality parameters in wheat flour and wheat-based products.....	61
Table 1.4. Literature related to other applications of NIR, FT-NIR, and NIR-HSI in wheat flour and wheat-based products.....	67
Table 1.5. Literature related to use of data fusion in the analysis of wheat flour and wheat-based products with NIR and MIR.....	72
Table 2.1. Samples properties.....	85
Table 2.2. SIMCA models performance for NIRS and NIR-HSI applied for discrimination of pure semolina and fiber-added semolina.....	88
Table 2.3. Model performance of PLSR based on characteristic wavelengths for quantitative analysis of fiber added to semolina.....	89
Table 3.1. Parameters of OLS regression models.....	99
Table 4.1. Properties of fibres used in pasta formulation.....	107
Table 4.2. Results of LDA models.....	119
Table 4.3. Results of PLSDA models.....	120
Table 5.1. LDA models performance for qualitative analysis of pectin in orange peels.....	128
Table 5.2. PLSR model performance for quantitative analysis of pectin in orange peels.....	128

LIST OF CONTENTS

<i>Introduction, Objectives and Thesis Structure</i>	24
1. INTRODUCTION.....	25
2. OBJECTIVES.....	27
2.1. General Objective.....	27
2.2. Specific Objectives.....	27
3. THESIS STRUCTURE.....	28
CHAPTER 1	31
<i>Near-infrared spectral techniques for wheat flour and wheat-based products evaluation: A review</i>	31
1.1. Introduction.....	35
1.2. Spectral Techniques.....	36
1.3. Chemometrics.....	40
1.3.1. Data pre-processing.....	41
1.3.2. Exploratory analysis.....	41
1.3.3. Quantitative analyses.....	42
1.3.4. Discriminant and classification analysis.....	43
1.3.5. Curve resolution/spectral unmixing methods.....	44
1.4. Application in wheat flour and wheat-based products.....	47
1.4.1. Authentication.....	47
1.4.2. Determination of quality parameters.....	57
1.4.3. Other applications.....	63
1.4.4. Data fusion.....	70
1.5. Conclusions and future trends.....	73
Acknowledgments.....	74
Declaration of competing interest.....	74
Author contributions.....	74
References.....	74
CHAPTER 2	83
<i>Identification of fibre added to semolina by Near Infrared (NIR) spectral techniques</i>	83
2.1. Introduction.....	84
2.2. Material and Methods.....	85
2.2.1. Sample Preparation.....	85

2.2.2.	NIR spectra and hyperspectral image acquisition.....	85
2.2.3.	Pre-processing.....	85
2.2.4.	Spectral pre-processing.....	85
2.2.5.	Data analysis.....	86
2.2.6.	Discriminant analysis.....	87
2.2.7.	Development of the prediction model and distribution map.....	87
2.3.	Results and discussion.....	87
2.3.1.	Raw spectra of samples.....	87
2.3.2.	Principal components analysis (PCA).....	89
2.3.3.	Sample classification.....	90
2.3.4.	Prediction models.....	90
2.3.5.	Distribution maps.....	91
2.4.	Conclusion.....	91
	Acknowledgments.....	91
	Conflict of interest disclosure.....	92
	References.....	92
	CHAPTER 3.....	93
	<i>Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fibre distribution in enriched pasta.....</i>	93
3.1.	Introduction.....	94
3.2.	Material and Methods.....	97
3.2.1.	Sample Preparation.....	97
3.2.2.	Image acquisition.....	97
3.2.3.	Data analysis: MCR-ALS.....	97
3.3.	Results and discussion.....	99
3.3.1.	Control sample and pure flour spectra as initial estimates.....	99
3.3.2.	MCR-ALS in pasta samples.....	99
3.3.2.1.	Fibre distribution.....	99
3.3.2.2.	MCR-ALS models.....	99
3.3.2.3.	Distribution of fibre content in the surface of samples	101
3.4.	Conclusion.....	101
	CRedit authorship contribution statement.....	101
	Declaration of Competing Interest.....	101

Acknowledgments.....	101
Appendix A. Supplementary data.....	101
References.....	101
CHAPTER 4.....	103
<i>Study of changes in pasta during cooking by hyperspectral imaging.....</i>	103
4.1. Introduction.....	106
4.2. Materials and methods.....	107
4.2.1. Sample preparation.....	107
4.2.2. Cooking process.....	108
4.2.3. Acquisition hardware.....	108
4.2.4. Data processing.....	109
4.2.5. Multivariate data analysis.....	110
4.3. Results and discussion.....	111
4.3.1. Raw spectra of samples.....	111
4.3.2. Principal Component Analysis (PCA)	115
4.3.3. Classification of pasta samples - Linear Discriminant Analysis (LDA)	118
4.3.4. Classification of pasta samples - Partial Least Squares Discriminant Analysis (PLSDA)	119
4.4. Conclusion.....	120
Acknowledgment.....	121
Conflict of interest disclosure.....	121
References.....	121
CHAPTER 5.....	124
<i>Determination of pectin content in orange peels by near infrared hyperspectral imaging.....</i>	124
5.1. Introduction.....	125
5.2. Material and Methods.....	127
5.2.1. Pectin extraction from raw material.....	127
5.2.2. Sample preparation and image acquisition.....	128
5.2.3. Spectra extraction and multivariate analysis.....	129
5.2.4. Principal components analysis (PCA)	129
5.2.5. Linear discriminant analysis (LDA)	130
5.2.6. Partial least Square regression (PLSR)	130

5.3.	Results and discussion.....	131
5.3.1.	Spectral analysis.....	131
5.3.2.	Principal components analysis (PCA)	131
5.3.3.	Linear discriminant analysis (LDA)	131
5.3.4.	Partial least square regression (PLSR)	131
5.4.	Conclusion	132
	CRedit authorship contribution statement	132
	Declaration of Competing Interest.....	132
	Acknowledgments.....	132
	References.....	132
	GENERAL DISCUSSION.....	134
	GENERAL CONCLUSION.....	138
	STUDENT STORY.....	141
	REFERENCES.....	145
	FINAL CONSIDERATIONS.....	148
	APPENDIX I.....	151
	<i>Food Quality and NIR Spectroscopy in the Omics Era.....</i>	<i>151</i>
	APPENDIX II.....	168
	<i>Automated method for determination of cheese meltability by computer vision.....</i>	<i>168</i>
	APPENDIX III.....	187
	<i>Identification of turkey meat and processed products using near infrared spectroscopy...187</i>	<i>187</i>
	APPENDIX IV.....	198
	<i>Supplementary material of Chapter 2.....</i>	<i>198</i>
	APPENDIX V.....	200
	<i>Supplementary material of Chapter 3.....</i>	<i>200</i>
	APPENDIX VI.....	202
	<i>Supplementary material of Chapter 5.....</i>	<i>202</i>

1. INTRODUCTION

Food quality and safety have become global issues, since consumers are increasingly concerned about food contamination, adulteration and authenticity. Food contamination can be due to physical, chemical, and biological factors. A very clear example of this is the production of different products in the same facility, which can lead to cross-contamination. The use of a common, or even close environment for milling peanuts and other food materials, such as wheat, cocoa beans and soybeans, can enable the contamination of powdered food products with an allergenic ingredient (Mishra et al., 2015).

Food adulteration usually involves the replacement or dilution of high-cost ingredients with cheap low-quality products without the consumer's knowledge. Fraudulent substitution in ground meat products, for example, involves not only economy, quality and safety issues, but also religion (Barbin et al., 2020). This is a very important issue since it represents, not only quality and economic effects, but also health threats to consumers (Verdú et al., 2016). If the product label does not indicate all the ingredients, the consumer can be deceived about what they are really buying. The intentional addition of bread wheat flour in durum flour for pasta production, for instance, is an adulteration that leads to a product with a lower resistance to cooking and therefore lower quality (Cocchi et al., 2006).

In the same way, the enrichment of products with other compounds, such as fibres, for example, might be controlled. The addition of an alternative ingredient in the formulation of products increases its nutritional value, and can affect its technological and sensorial characteristics (Badaró et al., 2019). In this context, the authentication of food products has become crucial to obtain a safe product with high quality.

Moreover, the determination of quality parameters in food products is a key step in the industry. The variability in the raw materials leads to highly variable products, which must be controlled. However, analytical techniques used in food industry to determine physicochemical and quality parameters as pH, colour, protein, fat, moisture and ash content are destructive, expensive, slow, require sample preparation, besides the amount of chemicals used, which involves environmental impact (Porep et al., 2015). Apart from that, some of these assessments have a subjective nature, being subject to variation because of human error that can be due to fatigue or experience of the assessor (Teye et al., 2015).

Methods based on optical properties, as near infrared spectroscopy (NIR) and NIR hyperspectral imaging (NIR-HSI), have been widely studied as an alternative to this situation. Spectroscopy techniques have showed their potential to replace or at least complement classical

methods. Although these techniques require reference values for the development of calibration models, once the calibrations have been set, they are easy to use (Üçüncüoğlu et al., 2013).

NIRS and NIR-HSI are two of many spectroscopy techniques used to detect food adulteration and contamination, and to authenticate food products. NIRS is an optical technology based on the interaction between incident light and molecules in matters. The resultant spectrum records the absorbed energy by the different molecules in the product, specifically the overtones and combinations of fundamental vibrations of C-H, N-H, O-H and S-H bond (Manley, 2014). However, the measurements are taken in specific points of the sample, detecting only a small portion; therefore, the spectra might not be representative for the whole sample, especially when the product is a food matrix. In order to obtain spatial information, another technology, as NIS-HSI can be used. It integrates both spectral and spatial information in a system, being a suitable choice for food products, since they are heterogeneous matrices. Moreover, this technique can also be used to develop chemical maps, which allow the visualization of how components of interest are distributed over the sample surface (Feng and Sun, 2012). On the other hand, these techniques provide a big amount of information; thus, chemometric methods are necessary to explore, classify samples or quantify components.

Considering the need for fast and non-destructive food analysis techniques, this work suggests the use of spectral techniques, as NIR and NIR-HSI, to study the distribution of fibre added to semolina and pasta. In addition, the study aims to verify the efficiency of such techniques in the quantification of these fibres. Although widely used, the application of these techniques in the identification, classification or quantification of fibres in products, such as pasta, has not yet been reported. Thus, the justification for this work is the use of fast and non-invasive techniques for control of the type and quantity of alternative ingredients, in this case, fibres, added to the formulation of semolina and pasta, as an alternative of food authentication of enriched products. This study also aims to apply two different techniques as NIR-HIS in the transmission mode to monitor the cooking process of fibre-enriched pasta, as a first step to determine parameters in cooked pasta. Moreover, this work aims to study the potential of NIR-HSI for identification of pectin content in orange peels as an alternative for pectin quantification in orange waste without the previous pectin extraction.

2. OBJECTIVES

2.1. General Objective

This study aimed to identify and classify different types of fibre samples added to the semolina and pasta produced by semolina-fibre formulations, and to study and to monitor the cooking process of this fibre-enriched pasta by spectral techniques. Moreover, this work aimed to quantify pectin content in orange peels using the spectral information of samples instead of the extraction of pectin for quantification.

2.2. Specific Objectives

- Acquisition of spectral information of samples of fibre-enriched semolina by two different techniques (portable near infrared spectrometry and hyperspectral imaging);
- Study of the spectral information of samples, identification of peaks and the molecules correspondent to each peak. Application of different pre-processing as an alternative to enhance the information in the spectra. Development of discrimination models to discriminate fibre from semolina samples and optimization of classification models according to the type of fibre in each sample;
- Development of regression models to determine the fibre content in semolina;
- Application of the regression models to the hyperspectral images to visualize the distribution of the fibres in the samples;
- Acquisition of spectral and spatial information of samples of fibre-enriched pasta hyperspectral imaging;
- Study of the spectral information of samples, identification of peaks and the molecules correspondent to each peak. Application of different pre-processing as an alternative to enhance the information in the spectra. Development of a multivariate curve resolution method to identify fibre distribution in pasta;
- Development of regression models to correlate the amount of fibre estimated by the curve resolution model and the real content;
- Acquisition of spectral and spatial information of samples of fibre-enriched pasta during cooking by NIR hyperspectral imaging in the transmission mode;

- Study of the spectral information of samples, identification of peaks and the molecules correspondent to each peak. Application of different pre-processing as an alternative to enhance the information in the spectra. Development of classification and discrimination models to classify and discriminate samples of pasta regarding their cooking time;
- Acquisition of spectral and spatial information of samples of orange peels by NIR hyperspectral imaging;
- Study of the spectral information of samples, identification of peaks and the molecules correspondent to each peak. Application of different pre-processing as an alternative to enhance the information in the spectra. Development of discrimination models to discriminate orange peel samples regarding the pectin content;
- Development of regression models based on full spectra and optimum wavelengths to quantify pectin content in orange peels.

3. THESIS STRUCTURE

The development stages of this research project are presented in 7 chapters. First, in the *Introduction, Objectives and Thesis Structure* are presented the main idea of this study, the objectives and the stages involved for its accomplishment.

In *Chapter 1 - Near-infrared spectral techniques for wheat flour and wheat-based products evaluation: A review*, we reviewed the specific applications of Near-Infrared Spectroscopy, Fourier Transform Near-infrared Spectroscopy, and Hyperspectral Imaging in assessing wheat flour and wheat-based products quality parameters, as well for the authentication and determination of composition of these products. This work provided the information existent in the literature and the gaps in this field, which contribute to the development of the thesis and the following works.

In *Chapter 2 - Identification of fibre added to semolina by Near Infrared (NIR) spectral techniques*, we compare the efficiency of the near infrared and hyperspectral imaging techniques in the classification of samples of different types of fibres added to semolina, quantification, and distribution of the fibre content. Three sets of 140, 220 and 351 samples were prepared, NIR spectra and hyperspectral images were acquired and interpreted using chemometric techniques. Principal Component Analysis (PCA) of pure samples from both techniques showed their separation into different groups, especially HSI. Soft Independent Modelling of Class Analogy (SIMCA), performed in raw and pre-treated spectra, showed high

values of sensitivity and specificity for HSI data, and the efficiency of the model was confirmed by the Coomans plot. In turn, NIR spectra showed a very low efficiency in the discrimination and classification of semolina added of fibre. Regression models were developed with HSI spectra by Partial Least Squares Regression (PLSR), showing R^2_P between 0.85 and 0.98, RMSECV and RMSEP in the range of 0.5 and 1%. In addition, the models were applied to each pixel of the hyperspectral images, allowing a visual distribution of the fibres in the samples.

In *Chapter 3 - Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fibre distribution in enriched pasta*, we propose the identification of fibre in pasta by NIR-HSI imaging, and the study of multivariate analysis tools for fibre identification in pasta. Seven types of fibre were used in pasta formulation in 2 percentages (3.5 and 7%). HSI images were acquired and Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) was used for the evaluation and quantification of fibre distribution in these pasta samples. Results showed R^2_V between 0.28 and 0.89, %LOF <6%, variance explained over 99%, and similarity between pure and recovered spectra over 96% and 98% in models using pure flour and control as initial estimates, respectively, demonstrating the applicability of NIR-HIS and MCR-ALS in the identification of fibre in pasta.

In *Chapter 4 - Study of changes in pasta during cooking by hyperspectral imaging*, we suggest the use VIS/NIR-HSI in the transmission mode for the study of changes in pasta during cooking and the determination of optimal cooking time of pasta as an alternative to automate the determination of pasta attributes. The *fettuccine* samples were produced replacing flour by seven types of fibre in two percentages (3.5 % and 7 %). For each type and percentage of fibre, ten portions of 25 grams of pasta were weighted, totalling 140 experiments (7 types of fibre x 2 percentages x 10 repetitions). Two units of pasta were removed from cooking water each 90 seconds. The cooking process was carried out for 18 minutes, totalling 13 times (0, 1.5, 3, 4.5, 6, 7.5, 9, 10.5, 12, 13.5, 15, 16.5 and 18 minutes). VIS/NIR hyperspectral images were acquired in the transmission mode. The two spectral range were separately evaluated, but VIS data did not show any relationship among samples. Therefore, further multivariate data analysis was carried out only on NIR data. PCA was performed and the PCA scores that better grouped the samples were used as variables to perform a Linear Discriminant Analysis (LDA). In comparison, Partial Least Square Discriminant Analysis (PLSDA) was performed to test the ability of the technique to predict optimum cooking time of pasta samples. LDA had values of sensitivity and specificity between 0.14 – 1.00 and 0.51 – 1.00, respectively, and non-error rate (NER) over 0.62. PLSDA had values of sensitivity and specificity between 0.67 – 1.00 and 0.10

– 1.00, respectively, and NER over 0.80. This study suggested that hyperspectral image in the transmission mode have good potential as an objective method to optimum cooking time determination, meeting this industry need.

In *Chapter 5 - Determination of pectin content in orange peels by Near Infrared Hyperspectral Imaging*, we quantify pectin content in orange residues using NIR-HSI. Hyperspectral images from orange peel (140 samples) with different amounts of pectin were acquired and the spectra was used for calibration models using multivariate statistical analyses. LDA showed better discrimination results considering three groups: low (0–5%), intermediate (10–40%) and high (50–100%) pectin content. PLSR models based on full spectra showed higher precision ($R^2 > 0.93$) than those based on few selected wavelengths (R^2 between 0.92 and 0.94). The results showed that this technique holds potential as an alternative to the carbazole colorimetry method to quantify pectin in orange peels, and to categorize orange peel samples into groups of different pectin concentration, and can be used to justify investments in the waste processing and extraction methods.

General discussion presents a general discussion of the results obtained in this thesis. The *General conclusion* presents a general conclusion of the results obtained in this thesis linking all the knowledge obtained in response to the general objective of the thesis. Finally, there is a briefly story regarding the student journey during her PhD in *Student Story*. All extra documents regarding other works developed by the student and the permission from the journals to use the published works in this thesis are in the *APPENDIX*.

CHAPTER 1

Near-infrared spectral techniques for wheat flour and wheat-based products evaluation: A review

**Near-infrared spectral techniques for wheat flour and wheat-based products evaluation:
A review**

Authors: Amanda Teixeira Badaró^{a,b,c*}, João Pedro Hebling e Tavares^a, Jose Blasco-Ivars^c, Nuria Aleixos Borrás^b, Douglas Fernandes Barbin^a

^aDepartment of Food Engineering, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, 13083-862, Campinas – São Paulo, Brazil

^bDepartamento de Ingeniería Gráfica, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain

^cCentro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), Ctra. CV-315, km. 10,7, 46113, Moncada, Valencia, Spain

* Corresponding author; tel: +55 19 98917 3064, E-mail: tbadaro.amanda@gmail.com

Declarations of interest: none

Paper submitted

Abstract

Background: Wheat flour is a food ingredient that provides compounds as fibre, carbohydrates, and proteins. It is used in different products, including pasta, cake, bread, among others. Therefore, the authentication and assurance of good quality are of great importance. The traditional techniques used for quality parameters determinations are laborious, destructive, and involve chemical reagents. So, it is necessary the development of techniques capable to overcome these disadvantages. The spectral techniques are rapid, non-destructive, and chemical-free.

Scope and Approach: This review approaches the applications of Near Infrared (NIR) Spectroscopy, Fourier Transform Near-Infrared (FT-NIR) Spectroscopy, and Hyperspectral Imaging (HSI) in the wheat flour and wheat-based products authentication and assessment of quality parameters, composition, and contamination.

Key Findings and Conclusions: The techniques here approached are effective in the determination of parameters in wheat flour and wheat-based products, as protein, gluten, fatty acids and are also non-destructive. Moreover, the detection and quantification of allergens and insect infestation, and the authentication of flour added of other ingredients showed promising results using these techniques. Future studies can explore the in/on-line applications of these techniques for industrial process lines and compare the use of handheld and benchtop spectrometers in these applications. There is a challenge in the near-infrared capability of quantifying parameters in low levels; the use of data fusion can be a way to resolve this challenge.

Keywords: Near-Infrared Spectroscopy, Fourier Transform, Hyperspectral Imaging, flour, bread, pasta.

Abbreviations			
ALS	Alternating least squares	PCR	Principal Component Regression
ASCA	ANOVA–Simultaneous Component Analysis	P.G.I.	Protected Geographical Indication
CV	Canonical Variables	PLS-DA	Partial Least Square Discriminant Analysis
ELM	Extreme Learning Machine	PLSR	Partial Least Square Regression
ELISA	Enzyme-linked immunosorbent assay	RER	Ratio of the Range in the Prediction
FT	Fourier Transform	RMSEC	Root Mean Square Error in Calibration
HFN	Hagberg Falling Number	RMSECV	Root Mean Square Error in Cross-Validation
HCA	Hierarchical Cluster Analysis	RMSEP	Root Mean Square Error in Prediction
HSI	Hyperspectral Imaging	RPD	Ratio of Prediction to Deviation
IR	Infrared	R_p	Coefficient of correlation of prediction
LDA	Linear Discriminant Analysis	R²_c	Coefficient of determination of calibration
%LOF	Percentage of lack of fit	R²_{cv}	Coefficient of determination of cross-validation
LV	Latent Variables	R²_p	Coefficient of determination of prediction
LWR-PLS	Locally Weighted PLS regression	SEE	Standard Error of Estimation
MCR	Multivariate Curve Resolution	SEP	Standard Error of Prediction
MEMS	MicroElectroMechanical System	SG	Savitzky-Golay
MIR	Mid Infrared	SIMCA	Soft Independent Modeling of Class Analogy
MLR	Multiple Linear Regression	SNV	Standard Normal Variate
MPLS	Modified Partial Least Squares	SVM	Support Vector Machine
MSC	Multiplicative Scatter Correction	SVR	Support Vector Regression
MSPC	Multivariate Statistical Process Control	SW	Short Wave
NIR	Near Infrared	UV	Ultra-Violet
OLS	Ordinary Least Squares	VCPA	Variable Combination Population Analysis
PAT	Process Analytical Technology	VIS	Visible
PC	Principal Components	WILMA	Wavelet Interface to Linear Modelling Analysis
PCA	Principal Components Analysis		

1.1.Introduction

Whole-wheat flour is a staple food widely consumed around the world, as it is the main ingredient for many products, including bread, cake, pasta, cookies, among others. Its consumption provides many nutrients in the human diet, especially fibre, carbohydrates, and proteins (Dave & Modi, 2018; Hussain et al., 2019). Moreover, it is usually cheap and accessible, being a good alternative to be implemented as a food vehicle as a strategy for food fortification (Akhtar et al., 2008).

The quality of wheat-based products is highly dependent on the grain and flour quality, so, the evaluation of quality parameters is very important in the food industry (Caporaso et al., 2018). Analytical methods traditionally used to determine quality parameters in wheat flour and wheat-based products are centred on laborious and/or destructive techniques. The gluten content in wheat or other cereals, for example, is determined using an immunological-based method named enzyme-linked immunosorbent assay (ELISA), which is destructive and uses reagents for extraction (Haraszi et al., 2011). Moreover, wet and dry gluten, moisture, protein, and ash contents of wheat flour and rheological properties of wheat dough, as alveograph and farinograph parameters, are determined by AACC methods (ACCC, 2000). The determination of quality parameters in pasta, as optimal cooking time, swelling index, cooking losses, among others, are also determined by time-consuming and destructive techniques (Bustos et al., 2015).

Although most of these techniques are accurate and are extensively used for flour parameters determination, they present disadvantages, including chemical reagents, which are used in most analytical methods. Moreover, some of these techniques, for example, some of the classical methods used in the industry for gluten determination, rely on forming the dough balls, washing them out and weighting the remained (gluten). Some other methods as enzyme linked immunosorbent assay (ELISA), polymerase chain reaction (PCR) and liquid chromatography coupled with mass spectrometry are also used, but they are expensive and require appropriate sample preparation (Czaja et al., 2018). Therefore, there is a need in developing rapid, non-destructive, and chemical-free methods, as spectral techniques, that can predict quantitative and/or qualitative parameters better or at least equal to the traditional methods (Verdú et al., 2016).

Many reports can be found in the literature regarding the spectral techniques, more specifically Near Infrared and Hyperspectral Imaging for food assessment. These techniques,

more specifically NIR spectroscopy, are reported to be one of the best adapted methods to evaluate wheat products, since it is a straightforward method of analysis for a fast, accurate and reliable determination of wheat and its derived products. Moreover, it can be used as a screening method in plant-breeding and as an in-line tool to monitor the changes that may occur during processing (Aït-Kaddour & Cuq, 2011). In the cereal field, the applicability of NIR spectroscopy as a tool to monitor wheat product processing and chemical changes during processing has been reviewed by Aït-Kaddour & Cuq (2011). Recently, quality and safety parameters assessed by NIR hyperspectral imaging were reviewed by Sendin et al. (2018), and quantitative and qualitative analysis of chemical composition in grains using NIR and HSI techniques were reviewed by Caporaso et al. (2018). In other recent works, Levasseur-Garcia (2018) introduced an updated overview of Infrared Spectroscopy methods for detecting mycotoxins on cereals, while Femenias et al. (2020) specifically approached the improvements in the assessment of *Fusarium* and deoxynivalenol (a mycotoxin) contamination in cereals using HSI. Hussain, Sun & Pu (2019) reviewed the classical and emerging technologies for assessment of safety and quality parameters of cereals. However, no review was found showing specifically the applicability of spectral techniques for wheat flour and its products. Therefore, this review brings the specific applications of (dispersive) NIR, FT-NIR, and HSI in assessing wheat flour and wheat-based products quality parameters, as well for the authentication and determination of composition of these products.

1.2. Spectral Techniques

Different spectral techniques have been constantly used in the food, pharmaceutical, and petrochemical fields. These techniques operate in different wavelength ranges of the electromagnetic field (from 100 nm to 1000 μm), from ultra-violet (UV) to far-infrared region, going through visible (VIS), near (NIR), and mid (MIR) infrared, according to the absorbed energy (Porep et al., 2015). In addition, spectrometers can be designed as portable or benchtop, dispersive or Fourier Transform devices, and their use will depend on the samples to be evaluated (Lohumi et al., 2015).

Near-infrared spectroscopy is a handy tool in process analytical technology (PAT) and quality control that has found widespread application in various fields, especially in qualitative and quantitative analysis of food products (Cortes et al., 2019a; Walsh et al., 2020). The NIR region covers wavelengths from 780 nm up to 2500 nm, and the resulting spectra of

such techniques are associated with absorbed energy of organic molecules, specifically the overtones and combinations of fundamental vibrations of X–H chemical bonds, as C–H, O–H and N–H bonds. Consequently, as water is the major food constituent and absorbs NIR radiation, it strongly affects the spectra of food products (Manley, 2014).

The interaction of light and the constituents of food products can happen in many ways. According to these diverse interactions, in order to determine external and internal sample properties, this vibrational technique can be used in different spectral modes, as reflectance (specular and diffuse), transmission, interactance, and transreflectance. The spectral mode choice is made depending on the type, physical properties, and characteristics of the samples. Reflectance measures the light that reflects or scatters from a sample surface, so, generally, it is used for measuring the spectra of solids. If the reflection is from a smooth surface, it is called specular reflectance; whereas, if the reflection is from a rough surface, as in food samples, it is called diffuse reflectance. In diffuse reflectance, the amount of scattering through the sample length is so high that the light that could cross the sample thickness is reflected instead of transmitted (Huang et al., 2008).

In transmission mode, the light passes through the sample and carries information regarding its internal properties; this mode can be used for analysing solid, liquid, and gaseous samples. Transreflectance and interactance modes combine reflectance and transmittance modes. In these cases, the transmitted light goes through a sample and it is reflected by either a reflector, in case of transreflectance, or by the own sample, in case of interactance. Therefore, transreflectance is usually applied for measuring the spectra of thin or clear samples; while interactance is commonly used when the transmission is not enough to obtain information regarding a solid sample (Lohumi et al., 2015). Figure 1A illustrates the sample presentation techniques described.

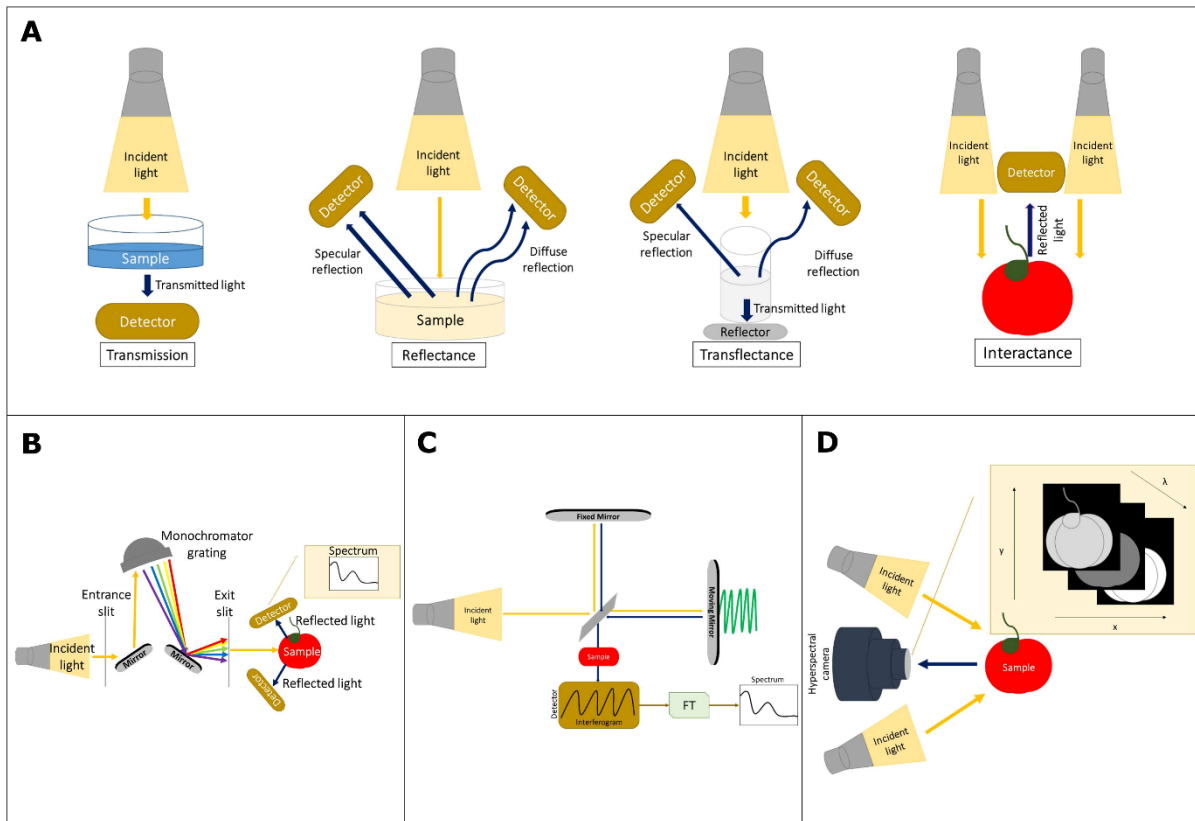


Figure 1.1. (A) Approaches for spectra acquisition, Scheme of the operation of (B) a dispersive NIR instrument, (C) a FT-NIR instrument, and (D) a NIR-HSI instrument.

Regarding the NIR instruments available, it is possible to find many models and brands, which include portable devices, benchtop, hyperspectral devices, Fourier Transform, among others. The technology behind these instruments is also different. In dispersive infrared instruments (Figure 1B), the measurements are made at one region of the spectrum at a time, so each wavelength is measured individually, lowering the scanning speed wavelength. In contrast, FT-NIR instruments scan the entire infrared spectrum, and individual scans can be combined to better represent the absorbance of the sample. The detector in these instruments measures the light intensity, generating an interferogram, which is converted into a spectrum by a Fourier Transform algorithm. Consequently, there is an increase in the scanning speed and resolution compared to dispersive equipment (Manley et al., 2002). Figure 1C illustrates the operation of an FT-NIR instrument.

The application of NIRS in the literature is generally based on benchtop devices (Casian et al., 2021). However, in the last years, the use of portable devices grew with its development and marketing (Pasquini, 2018). The most important feature of these devices is the miniaturization of benchtop equipment. This feature has a great impact in the industrial

environment as it makes it more feasible to apply NIR technology in for in-line routine monitoring along the production chain; additionally (Cortes et al., 2019b), the application of NIR spectroscopy in delivery and supply chains is becoming more practical, allowing the verification of the compliance in the processing line and the final products. (González-Martín et al., 2021; Mayr et al., 2021).

In the literature, many studies compare the application of benchtop and portable devices in various applications (Ayvaz et al., 2015; Bizerra Brito et al., 2017; Casian et al., 2021; Cecchini et al., 2021; González-Martín et al., 2021; Kirchler et al., 2017; Malegori et al., 2017; Mayr et al., 2021; Rukundo et al., 2021). Although few poor performances, overall, the portable devices showed good results when compared to benchtop equipment. In the field of portable NIR devices related to wheat flour and wheat products there are also studies (Badaró et al., 2019; Chen et al., 2021; Grassi et al., 2018; Jiang et al., 2020; Kumagai et al., 2004; Stellacci et al., 2012). Overall, the results found in these works were very satisfactory for portable devices, with exception of the work developed by Badaró et al. (2019), since the authors compared the results of a portable NIR spectrometer and a hyperspectral camera.

An important disadvantage of portable devices is the small working area, so that to obtain representative information of the sample in case of heterogeneous samples, it is necessary many measurements to cover it. Moreover, the construction of miniaturized equipment also can result in low signal-to-noise ratios (Mayr et al., 2021; Pasquini, 2018). The advances in MEMS (MicroElectroMechanical System) technology resulted in the production of tiny devices that, when together, build a complete handheld device (Pasquini, 2018). These handheld devices are called “micro” instruments because they are compact and have low weight (around 100g). Nevertheless, they feature internal radiation sources and wireless technologies as Bluetooth and Wi-Fi for data transmission. The operation of these instruments is not the same. As an example, a MicroNIR manufactured by Viavi Solutions contains two small tungsten filament radiation sources, the wavelength selection is based on a LVF (linearly variable interference filter) placed on a chip with an array of 124 InGaAs sensors and the integration time is controlled by a software (Friedrich, 2014). There are several micro instruments available in the market, which Pasquini (2018) brings a list in his review. The author detailed the operation of a representative number of micro instruments. As mentioned by the author, these miniaturized instruments allowed NIRS technology to be used in the field or at a point of production, for example, however, they require careful evaluation.

The traditional spectrometers only detect one point or a small portion of the samples, in case of many measurements of the sample, so when it comes to food products,

which are very heterogeneous matrices, the spectra may not be representative of the whole sample (Feng & Sun, 2012). An alternative to suppress this situation is the application of hyperspectral imaging technique (HSI) for simultaneous detection of both spectral and spatial information.

Near-infrared hyperspectral imaging (NIR-HSI) has demonstrated capable of exploring both the chemical composition and its spatial distribution within a sample. In comparison to traditional NIR spectroscopy, HSI applies spectroscopy coupled with imaging and the data is provided as a “hypercube” due to the three-dimensional structure, with two spatial dimensions (x , y) and one spectral dimension (λ) (Figure 1D) (Amigo et al., 2015). Therefore, hyperspectral imaging allows the visualization of the distribution of the substance of interest in samples.

This technique has been used for different purposes, including assessment of the presence of deoxynivalenol and ergosterol in wheat samples (Femenias et al., 2021); detection of peanuts in wheat flour (Lohumi et al., 2018); detection and quantification of particles of ergot bodies in cereal flour (Vermeulen et al., 2017); determination of bulk density and particle size of wheat flour (Zhu et al., 2017); classification of Italian wheat durum spaghetti (Menesatti & Bucarelli, 2007), among others.

Even though spectroscopy presents good performance for the described applications, it provides a large amount of data, which should be processed to extract useful chemical information and avoid redundant information. Thus, the association of vibrational techniques to chemometrics becomes necessary (Caporaso et al., 2017).

1.3. Chemometrics

Chemometrics is a research area, which uses mathematical and statistical methods to extract the relevant information in any type of data, including the chemical information of spectra and correlate with quality parameters or physical properties of a sample. This technique involves unsupervised methods of analysis, such as exploratory analysis, or supervised methods, as classification or prediction of quantitative properties, which are based on reference chemical or physical measurements (Caporaso et al., 2017).

1.3.1. Data pre-processing

Prior to the application of multivariate methods of analysis, the data is pre-processed. The spectrum or the image can be affected by a large number of factors, e.g., the effects of light scattering as baseline shifts and non-linearities. Moreover, the pixels measured in hyperspectral images carry all the information from the sample surface, which includes mixed information from the components, spectral noise, and irrelevant information, spatial interferences. Therefore, spectra and image pre-processing must be applied in order to reduce undesirable interference without losing valuable information. Among the most commonly used pre-processing techniques, there are Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV), normalization, which reduce scattering effects; and spectral derivatives, as Savitzky-Golay (SG) polynomial derivative filters, which smooth spectra by reducing baseline variations (Lohumi et al., 2015). The pre-processing improves the spectral information by improving the noise-ratio signal and enhancing peaks that overlap in the raw spectrum. The parameters used for the pre-processing, such as the selection of the window size and the polynomial order in a 1st Derivative, for example, has a large influence in smoothing the spectrum (Su & Sun, 2016). In imaging pre-processing, there is also dead pixels, which are missing data in the images, and background information, for example, from the surface where the samples is placed, that can influence the models, so they must be removed (Amigo, 2020). After pre-processing, data is available for exploratory analysis and the development of regression, classification or curve resolution models.

1.3.2. Exploratory analysis

Exploratory analysis involves unsupervised methods, such as Principal Components Analysis (PCA) and Hierarchical Cluster Analysis (HCA). PCA is the most common method used prior to classification or regression to investigate relationships between samples and describe the variation among them. This method works by decomposing the data into new variables, which are linear combination of the original data, called Principal Components (PC's). Score plots of the new variables are constructed, allowing the visualization of the relationship between samples and identifying anomalous samples, called outliers. PCA can be applied with no prior information about the sample chemical composition. By

investigating the PCA scores and loadings plots, samples and variables can be selected to create optimal calibration models and validate them (Manley, 2014).

1.3.3. Quantitative analysis

One of the most powerful methods for quantitative analysis is PLSR, which is applied to the linear regression modelling of the multivariate to several independent variables. The PLSR algorithm uses spectral information (X-block) and their respective parameter (Y variables) to predict unknown samples. The Y variable, also called the dependent variable, is a chemical or physical parameter (Su & Sun, 2016). In PLSR, the data are reduced by relating linearly the spectral information and the parameter of interest. The resulting combinations of the variables are named Latent Variables (LV's) and are used for the calibration. Hence, it is essential that the first few latent variables contain as much information of predictive value as possible. However, an inadequate number of LV's can lead to an under- or overfitted model and, consequently, poor predictions result of unknown samples. In order to determine an optimum number of factors, apart from a small number of LV's, a low prediction error must be considered. Moreover, to test the models reliability, the prediction of unknown samples must be performed (Porep et al., 2015).

Thus, the evaluation of these models' performance is based on the parameters of coefficient of determination of prediction (R^2_p); the root means square error of prediction (RMSEP); and the ratio of prediction to deviation (RPD). The coefficient of determination of calibration (R^2_c) and cross-validation (R^2_{cv}) and the root means square error of calibration (RMSEC), cross-validation (RMSECV) are also considered, but the prediction with independent external data, which are not included in the models, is necessary. The equations for the calculation of these parameters are presented in Table 1.1.

Among many applications, PLSR has been used in the prediction of contamination of organic wheat flour with common wheat flour, cassava flour, and corn flour (Su & Sun, 2017); of wheat flour with peanut and walnut (Zhao et al., 2018); and bulk density prediction in wheat flour (Zhu et al., 2017).

1.3.4. Discriminant and classification analysis

Among all discriminant and classification analysis methods, the most used are Partial Least Square Discriminant Analysis (PLS-DA), Soft Independent Modelling of Class Analogy (SIMCA), and Linear Discriminant Analysis (LDA). Each one with its particularity, these supervised methods are used as qualitative multivariate data analysis techniques or pattern recognition methods by developing specific algorithms with the spectral data matrix (X-block) and the correlated category of the samples (Y variables) (Su & Sun, 2016).

For a discriminant analysis, PLS-DA, for example, is used to raise the separation between groups based on Partial Least Square Regression (PLSR) models, where the category is coded as 0 and 1, with 1 corresponding to the category of samples of interest and 0 to all the other samples. The classes are delimited adopting a threshold value between 0 and 1 using probabilistic density functions and the Bayesian theory (Ferreira, 2015). PLS-DA has been employed in qualitative analysis, as the assessment of rye flour, organic wheat flour, spelt flour and organic spelt flour (Su & Sun, 2016), and the discrimination of kernels and flours from bread wheat, spelt, durum, emmer, and einkorn (Ziegler et al., 2016).

Such as PCA, LDA is a classification method used for dimensionality reduction. However, it searches the Canonical Variables (CV) that maximize the separation between multiple classes and minimize the variance within the same classes. The first CV represents the maximum ratio between inter- and intra-class variances. Another particularity of such a method is the number of samples, which must be greater than the number of variables for calculating covariance matrices. In addition, if the chosen variables are highly correlated, the model will be overfitted. Therefore, relevant variables must be selected and this step can be automatically performed by algorithms as “jack-knife” (Westad & Martens, 2000) or manually, by selecting the important wavelengths (peaks and valleys) in the loadings plot of PCA (Kumar & Chandrakant Karne, 2017). This technique showed promising results discriminating between wheat kernels with high or acceptable Hagberg Falling Number (HFN) and those with excessively low HFN, which is a parameter used to determine sprouting problems in wheat (Caporaso et al., 2017).

On the other hand, SIMCA is not only a discriminant method, but a classification method based on principal component analysis (PCA). Therefore, a PCA is developed for each class analysed, and the relevant number of components is selected based on the minimum cross-validation error in order to calculate Q-statistic and Hotelling T² and determine the boundaries

of each class. SIMCA calculates the Euclidean distance of each class using the principal component models, and the results can be visualized in the Coomans plot, which shows the distances for each sample from the model for class 1 against class 2. Different from the others, SIMCA can classify an unknown sample as belonging to one of the classes of interest or not (Rodríguez et al., 2019). This method was used to detect quinoa flour adulteration (Rodríguez et al., 2019); and qualification of Italian pasta produced by traditional or industrial production parameters (Menesatti et al., 2014).

The evaluation of these models' performance is based on the parameters of specificity and sensitivity. The capacity of a sample that belongs to the class of interest to be classified into this class is called sensitivity, while the capacity of a sample that does not belong to the interest class to be classified into the uninterested class is called specificity. Based on these parameters, the best models should obtain higher sensitivity and specificity (Ziegler et al., 2016). The sensitivity and specificity can be described by Equations 1 and 2 (Table 1.1). Moreover, as PLS-DA is based on PLSR, the calibration, cross-validation, and prediction errors, are also used to evaluate the performance of the models.

1.3.5. Curve resolution/spectral unmixing methods

As already described, a hyperspectral image consists of a three-dimensional cube, providing both chemical (spectrum) and spatial (pixels) information of a sample. Each pixel carries the physicochemical information of mixtures of constituents of a sample. Then, when describing powerful methods, curve resolution or spectral unmixing methods are important in hyperspectral image analysis. Multivariate Curve Resolution (MCR) is one of these methods. MCR is a bilinear model that can relate the pure spectra with the concentrations of each compound in a mixture. Initially, a matrix containing spatial information (XY) by λ (spectral data) called D is developed reshaping the hyperspectral cube (Forchetti & Poppi, 2017). Then, MCR is applied to this matrix; a matrix called C is generated containing the concentrations of the compounds in each pixel, and a matrix called S is also developed with the pure spectra of the compounds. This model can be described as:

$$D = CS^T + E \quad (1)$$

Where E is the matrix that expresses the experimental error or variance unexplained by the bilinear model (Amigo & Ravn, 2009; De Juan et al., 2014).

The MCR aims to solve the matrix C and the matrix S under constraints and initial estimates using an algorithm. One of the most common is based on alternating least squares (ALS) (Forchetti & Poppi, 2017). The choice of constraints and initial estimates is the first step in the MCR-ALS method and the acquisition of reliable results depends on this choice (Badaró et al., 2021; De Juan et al., 2014). Constraints function as mathematical conditions that improve the achievement of final solutions by MCR-ALS; as an example, one constraint can be non-negativity that forces the profiles to have positive values. Initial estimates must obey the selected constraints and are spectra or concentration profiles; if there is previous knowledge (e.g., pure spectra), the initial estimates choice can be based on it.

Considering the approach of MCR, this method is suitable for the formulation of a mixture analysis model (de Juan & Tauler, 2021). The MCR methods can produce, as already seen, pure spectra and the concentration distribution map of each compound in a mixture. It is essential to feature that MCR-ALS on its own is not a quantitative method (Badaró et al., 2021). MCR methods have been used to evaluate, resolve, and quantify fibre in enriched pasts (Badaró et al., 2021) and in the resolution of chemical compounds spectra to improve the interpretability of the white wheat bread crumb staling process (Amigo et al., 2021).

Table 1.1. Equations for models' performance calculation.

Equation number	Parameter	Equation	Description	Range
1	Sensitivity	$Sensitivity (\%) = \frac{TP}{TP + FN} \times 100$	TP: number of true positives FN: number of false negatives	0-1 or 0-100%; being 0 no accuracy of the model and 1 or 100% good accuracy of the model
2	Selectivity	$Sensitivity (\%) = \frac{TP}{TP + FN} \times 100$	TN: number of true negatives FP: number of false positives	
3	Determination coefficient of calibration (R^2_c)		\hat{y}_i : predicted values y_i : reference values	0-1 or 0-100%; being 0 no correlation and 1 or 100% linear correlation between predicted and reference values
4	Determination coefficient of cross-validation (R^2_{cv})	$R_x^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$	\bar{y}_i : average of reference values n : number of samples	
5	Determination coefficient of prediction (R^2_p)		x : C – calibration; CV – cross-validation; P – prediction	
6	Root mean square error in calibration (RMSEC)	$RMSEC = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n - A - 1}}$	\hat{y}_i : predicted values y_i : reference values	
7	Root mean square error in cross-validation (RMSECV)	$RMSECV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$	n : number of samples A : number of factors used in the model	Lower values indicate a better prediction ability (must be as low as possible to obtain acceptable models)
8	Root mean square error in prediction (RMSEP)	$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{val}} (\hat{y}_i - y_i)^2}{n_{val}}}$	n_{val} : number of samples for external validation	
9	Ratio of prediction to deviation (RPD)	$RPD = \frac{SD}{RMSEP \text{ or } RMSECV}$	SD : standard deviation $RMSEP$: root mean square error in prediction $RMSECV$: root mean square error in cross-validation	Higher values indicate a better prediction ability

1.4. Application in wheat flour and wheat-based products

Spectroscopy has been used for the qualification and quantification of many parameters in wheat flour and wheat-based products. The main techniques used for these determinations are NIR, FT-NIR, and NIR-HSI. The works found in the literature were organised and split into sections according to the application. First, the works which spectral techniques were applied in authentication of wheat flour and wheat-flour products. After that, the application of these techniques to the determination of quality parameters in those products, followed by composition determination. Then, other applications of these techniques as the discrimination of different flours or thermal treatments were also cited. Finally, those works in which data fusion was used.

1.4.1. Authentication

In recent years, an increase in the demand for healthy food products arose. Thereby, many fortified products have been developed as pasta enriched with fermented quinoa flour (Lorusso et al., 2017) or whole wheat cracker fortified with tuna bone bio-calcium (Benjakul & Karnjanapratum, 2018). The addition of new ingredients in the product formulation can improve its nutritional characteristics; however, it can affect its structure and sensorial attributes, generating undesirable modifications in the product (Lorusso et al., 2017). Therefore, rapid and efficient methods for food authentication must be developed; thus, the use of spectral techniques is an excellent alternative for food authentication. Some application of spectral techniques studied in this work for wheat flour and wheat-flour products are summarised in Table 1.2.

As mentioned before, this technique is based on the chemical information of the analysed sample, which is directly related to the product used to fortify. Wang et al. (2018) investigated the practicality of using NIR spectroscopy as a rapid method to quantify the amount of potato flour used to formulate Chinese steamed bread. Wheat flour was replaced by potato flour in a range of 0-42% in 2% intervals during bread preparation. The authors reported that, even with the challenge of having the same major constituents as starch, the NIR spectroscopy had satisfactory sensitivity and repeatability to quantify samples with potato flour contents greater than 20% ($R^2_p=0.8865$, $RPD=3.07$). In a similar context, Huang et al. (2018) also explored the NIR spectroscopy ability to determine the potato flour content added to wheat flour in the formulation of staple food. In this work, wheat flour was mixed with potato flour

from 0% to 100% (1% increment) for further NIR analysis. The results showed good performance, with a coefficient of determination of 0.9995, a standard deviation of 0.69 %, and an averaged repeatability standard deviation of 0.246, and a coefficient of variation of 0.967.

When it comes to food, which is a very complex matrix, some techniques are more suitable than others. For example, a recent work compared near-infrared spectroscopy and hyperspectral imaging techniques in the quantification of fibre added to semolina and its distribution (Badaró et al., 2019). The authors added to semolina different types of fibre, which differed in source and particle size, in different percentages and compared the accuracy of both techniques to classify, quantify and identify the fibre distribution in flour samples. The data were analysed with SIMCA and PLSR. In this work, the portable NIR did not show good results in the classification of samples with different percentages of fibre, showing sensitivity and specificity close or equal to zero. On the other hand, the HSI-NIR technique resulted in sensitivity and specificity values close or equal to one, showing the good performance of the models. Moreover, the authors used the HSI-NIR spectra to build prediction models, obtaining R^2_p over 0.85 and RMSECV and RMSEP between 0.44 and 1.09%. Additionally, as an advantage of this technique, distribution maps were developed, and the spatial distribution of the fibre in the sample was successfully identified.

The NIR-HSI evaluation of a wheat product enriched with fibre was also studied in other work (Badaró et al., 2021). This work aimed to evaluate and quantify fibre in dry fettuccine-type pasta since fibre enrichment affects the technological quality of pasta. The pasta flour was replaced with fibre in seven different types and in four different proportions, 2%, 3.5%, 5%, and 7% (w/w), and controls samples without fibre were also produced. The proposed analysis technique was MCR-ALS. As initial estimates, the spectra of pure fibre, pure flour, and control sample were chosen; non-negativity for the concentration profiles and normalization of the spectral profiles were used as constraints. As already mentioned, MCR-ALS on its own is not a quantitative method; thus, Ordinary Least Squares (OLS) Regression was performed using the proportion of fibre obtained by thresholding the MCR-ALS concentration profiles and the real proportion of fibre. The similarity between pure spectra and recovered spectra in models using pure flour as the initial estimate was above 96% and using the control sample was 98%. The percentage of lack of fit (%LOF) was lower than 6%, and more than 99% of the total variance was explained by MCR-ALS models for all samples. Differences between real fibre content and the content predicted by the regression were observed due to the lack of homogeneity as a result of fibre particle size and characteristics.

Therefore, the authors declared that NIR-HSI coupled to MCR-ALS could be a tool for the identification and quantification of fibre added to pasta.

The addition of ingredients in food products formulation does not always intend to promote health and well-being to consumers. The replacement of original ingredients with cheaper or low-cost materials characterizes an unscrupulous adulteration and food fraud, requiring rapid and efficient methods to identify and quantify these adulterants. Therefore, methods based on spectral techniques have been studied for this purpose, as in the study of Verdú et al. (2016), where the authors verified the capability of short wave (SW) NIR hyperspectral image technique to identify adulterations in wheat flour and bread with cheap grains. The authors prepared different binary flour mixtures of wheat flour and sorghum, oats, and corn flour in the percentages of 2.5, 5, 7.5, and 10% (w/w). Then, these mixtures were used in the bread preparation. The SW-NIR hyperspectral analysis was carried out in both flour and bread crumb samples and data analysed based on Multivariate Statistical Process Control (MSPC). The results obtained in this work were shown in terms of Q-residuals. The samples of both datasets (flour and bread crumb) were beyond the 95% confidence limit of the pure samples, and showed similar behaviour, with a slight difference between flour and crumbs in the distance of sample groups. However, they concluded that the technique reached 2.5% of adulteration, showing some limitations of the method regarding small concentrations.

The demand for organic food has increased over the years; however, due to the lower yield and difficult qualification process, the prices of organic products are high. Therefore, aiming for high profits, the adulteration of organic products has become recurrent, and the authentication of these products, consequently, became more necessary. For this objective, Su & Sun (2016) investigated the adulteration of organic spelt flour with a hyperspectral imaging system. The authors proposed an algorithm to select optimum wavelengths in order to evaluate the addition of rye flour, organic wheat flour and spelt flour in organic spelt flour. Samples were adulterated from 3 to 75% (w/w) in a 3% interval. PLS-DA was applied for the qualitative analysis of the different flours; meanwhile, the quantification of adulteration was determined by PLSR and MLR. The model's performance was evaluated in terms of the R^2 and RMSE of calibration, cross-validation, and prediction. In addition, the performance of PLS-DA was also assessed in terms of specificity and sensitivity of cross-validation. The coefficient of determination of prediction of PLSR models (based on the full spectrum and optimum wavelengths) was between 0.922 and 0.995, and the errors were lower than 0.061. With a threshold around 0.5, both the specificity and sensitivity in calibration and cross validation reached 1.

These authors also investigated the effectiveness of the same technique to quantitatively detect Irish organic wheat flour adulterated with common wheat flour, cassava flour, and corn flour (Su & Sun, 2017). Samples were adulterated from 3 to 75% (w/w) in a 3% interval, and the quantification of the adulteration was analysed in terms of PCR and PLSR. The performance of the models was evaluated according to the coefficient of determination, and root mean square error of prediction, which were very satisfactory (R^2_p : 0.948-0.976 and RMSEP: 0.035-0.050; R^2_p : 0.968-0.986, RMSEP: 0.026-0.039; R^2_p : 0.951-0.988, RMSEP: 0.051-0.144, for wheat, corn, and cassava flour, respectively). Additionally, the authors developed reduced models with optimum wavelengths with high accuracy (R^2_p =0.971, 0.973, and 0.986 for wheat, corn, and cassava flour, respectively). Based on these models, distribution maps were predicted to evaluate the adulteration distribution in organic wheat flour. Considering these satisfactory results, hyperspectral imaging coupled to chemometrics could be used as a tool for the authentication of organic flour in the industry.

Rachmawati et al. (2017) combined near-infrared spectroscopy and chemometrics for authentication of few samples of taro flour from wheat and sago flour, which are cheaper than taro flour and, consequently, potential adulterants for this type of products. The experiment was carried out using a mixture of 5%, 25%, and 50% of adulterated taro flour from wheat and sago flour, and the authors used principal component analysis and discriminant analysis to build an authentication model. The discriminant models built with the principal components showed good accuracy in authenticating taro flour with 90.48% and 85% of the samples mixed with wheat and sago flour correctly identified. The authors concluded that the near-infrared spectroscopy could be used for authentication of taro flour from wheat and sago flour.

Azodicarbonamide is a powder chemical used as a dough conditioner to fortify gluten in flour. However, its use in flour products has concerned some researchers. In addition to the indirect changes, it can make in the flour structure and the compounds it generates in the environment of food processing, Azodicarbonamide can cause severe harm after consumption for an extended time. Therefore, Che et al. (2017) studied the possibility of using Visible/Near-Infrared Spectroscopy to detect Azodicarbonamide in wheat flour in a very low concentration. In this work, samples concentration ranged from 0 to 300 mg/kg, in a concentration gradient of 3 mg/kg. Due to low concentrations, a stepwise dilution method was used in order to reduce the error. The authors used PLS, Back Propagation Neural Network, and Radial Basis functions to quantify the samples adulteration and the outcomes were very promising. The best result was achieved by the Radial Basis model, with R^2_p =0.99996, RMSEP=0.5467, and RPD=116.5858,

showing the feasibility of using Visible/Near Infrared Spectroscopy coupled to chemometrics to detect Azodicarbonamide adulteration in wheat flour at very low concentrations.

Spectral techniques have also been applied in the authentication of food products to reduce or avoid the risk of food contamination, either by allergens or other extraneous impurities. Peanut, for example, is a popular ingredient in commercial food products, and several reactions can occur by its ingestion for those who are allergic. Considering this, Mishra et al. (2015) explored the ability of NIR hyperspectral imaging for the detection and quantification of peanut traces in wheat flour, which is a very common product. Hyperspectral images of pure flour, pure peanut, samples with a known position of peanut on the surface, and eight homogeneously mixed samples of flour and peanut in an adulteration level ranging from 10 to 0.01% were acquired. PCA was performed on the spectral data, and the loadings of the first principal component were used to detect the pixels related to peanuts and flour. The results obtained for samples with less than 0.10% of adulteration had high errors and showed to be not reliable. However, a correlation of 0.946 between actual and predicted values of peanuts adulteration was achieved. Therefore, the authors concluded that this technique could be used to detect peanut traces in flour in this range of adulteration, facilitating quality control on process lines.

In the same context, Zhao et al. (2018) also found promising results in the evaluation of NIR-HSI for the detection of peanut and walnut powders in whole wheat flour. Hyperspectral images were acquired from samples contaminated with 0.01%, 0.05%, 0.1%, 0.5%, 1%, 3%, 5%, and 10% (w/w) of the adulterants. PLSR models were developed based on full and reduced spectra. The authors also faced the challenge of trying to predict the adulterants at extremely low concentrations. However, models developed for both contaminants showed good accuracy, with a coefficient of determination ranging from 0.985 to 0.994 and 0.996 to 0.998 for peanut and walnut, respectively. The root mean square error of prediction was also satisfactory, ranging from 0.251 to 0.398 % and 0.153 to 0.198 % for peanut and walnut. In addition, reduced models were developed and had excellent performance (R^2_P : 0.981-0.988, 0.990-0.997, 0.960-0.987 and RMSEP: 0.348-0.465 %, 0.170-0.324 %, 0.373-0.645 %, for peanut, walnut, and peanut + walnut, respectively) and visualization maps were well predicted. Therefore, this work showed the feasibility of such technique for quantitative analysis of peanut and walnut contamination in wheat flour.

Moreover, there is a concern about the presence of toxic and undesirable contaminants in cereals used as ingredients in the food industry. For instance, the presence of ergot bodies in cereal flour can generate a risk of poisoning due to their toxic components.

Therefore, the European Commission (EC) has established a limit of 500 mg.kg⁻¹ of ergot bodies in cereals for human consumption. In this context, Vermeulen et al. (2017) carried out their studies on the online detection and quantification of particles of ergot bodies in wheat flour using near-infrared hyperspectral imaging. The authors prepared two sets of experiments, one containing wheat flour contaminated with ergot from 0% to 2.6%, and one containing wheat flour contaminated with 50%, 20%, 10%, 5%, and 1% of ergot. A PLS-DA model was developed for all the samples; meanwhile, another PLS-DA model was developed for the first set of experiments. The PLS-DA models were built by selecting around 800 pixels of 6 images containing ergot bodies and 4 images with wheat particles. After the models were created and evaluated in terms of sensitivity and specificity of cross-validation, they were applied in images of unknown samples to estimate the amount of ergot bodies in it. Then, the correlation between the reference and predict values for ergot bodies (mg/kg) was calculated determining the ratio between the number of pixels detected as ergot bodies and the sum of the pixels detected as wheat and ergot together. The results were satisfactory in both cases, with R²_P of 0.99 and 0.87, respectively. In addition, the results showed that, even with false positives and false negatives, ergot bodies were detected in all the contaminated samples and no ergot bodies were detected in pure wheat flour samples. However, although these promising results, the PLS-DA models did not discriminate samples with a low concentration of ergot bodies.

The contamination by insect fragments was also studied by Bhuvanewari et al. (2011). In this study, the authors used near-infrared (NIR) hyperspectral imaging to develop models for the prediction of insect fragments in semolina. Samples were prepared in a ratio of 0, 50, 75, 150, 300 insect fragments per 50 g of semolina. The PLSR model developed achieved a coefficient of determination of 0.99, showing the effectiveness of NIR hyperspectral imaging to detect inset fragments in semolina compared to the traditional method (R² = 0.639-0.767). With the same objective, Mishra et al. (2018) used FT-NIR to determine insect infestation in wheat grain by analysing the change in uric acid, protein, final moisture content, thousand kernel weight, and hardness. The insects affect the grain quality by producing holes over the surface, consuming nutrients; therefore, lowering the kernel weight, and producing uric acid. For the experiment, two types of insects infested the wheat grain in the following ratios: 0, 5, 10, and 15 insects per 100 g; the containers with wheat grain had as initial moisture content 10, 12, 14, and 16% (wb). The containers were also maintained in an incubator set for 45, 90, 135, and 180 days. The spectral data were analysed using PLSR. The developed models for the five quality parameters had R²_P varying from 0.895 to 0.938 and RPD from 3.034 to 3.971, showing a good performance for each parameter. Additionally, the author declared that the model

predicted the infestation duration efficiently with no significant difference between the reference analysis and the FT-NIR data ($P > 0.05$); however, the protein content had a significant difference ($P < 0.05$), but, since the Pearson's correlation coefficient was 0.993, the model is accurate to predict the protein content.

Table 1.2. Literature related to the adulteration, contamination, and addition of ingredients in wheat flour and wheat-based products

Category	Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Adulteration	Flour/Bread	VIS/SW-NIR-HSI	400–1000 nm	Bench top	Detection of adulterations in wheat flour and bread with cheap grains, and comparison of hyperspectral information with physicochemical alterations in the properties of products	ANOVA, MSPC	$\alpha = 0.01$	(Verdú et al., 2016)
	Flour	NIR-HSI	900–1700 nm	Bench top	Determination of the fidelity of organic spelt flour from three categories of adulterants including rye flour, organic wheat flour and spelt flour	PLSDA, PLSR, MLR	Accuracy: 100%, R^2_p : 0.928-0.998; RMSEP < 0.116	(Su & Sun, 2016)
	Flour	NIR-HSI	900–1700 nm	Bench top	Detection of Irish organic wheat flour (OWF) adulterated with common wheat flour (WF), cassava flour (CaF) and corn flour (CoF)	PLSR, PCR	Full spectra: (WF): R^2_p : 0.948-0.976 RMSEP: 0.035-0.050; (CoF): R^2_p : 0.968-0.986, RMSEP: 0.026-0.039, (CaF): R^2_p : 0.951-0.988, RMSEP: 0.051-0.144; Optimum wavelengths: (WF): R^2_p =0.971, RMSEP=0.038; (CoF): R^2_p =0.973, RMSEP=0.036; (CaF): R^2_p =0.986, RMSEP=0.026	(Su & Sun, 2017)
	Flour	NIR	1000-2500 nm	Bench top	Identification and authentication of taro flour from wheat (WF) and sago flour (SF).	PCA, DA	Accuracy: (WF): 90.48% (SF): 85%	(Rachmawati et al., 2017)
	Flour	VIS/NIR	400-2500 nm	Bench top	Prediction of predict Azodicarbonamide concentrate in wheat flour	PLSR, Back Propagation, Neural Network (BP), Radial Basis (RBF)	(PLS): R^2_p =0.99621, RMSEP=5.5411 (BP): R^2_p =0.99937, RMSEP=2.3965 (RFB): R^2_p =0.99996, RMSEP=0.5467	(Che et al., 2017)

Table 1.2. (continued)

Category	Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Contamination	Flour	NIR-HSI	1000-2500 nm	Bench top	Detection and quantification of peanut traces in whole wheat flour	PCA	R ² _p =0.946	(Mishra et al., 2015)
	Flour	NIR-HSI	936-1720 nm	Bench top	Detection and prediction of peanut and walnut powders in whole wheat flour	PLSR	<p>Full spectra: (Peanut): R²_p: 0.985-0.994, RMSEP: 0.251-0.398; (Walnut): R²_p: 0.996-0.998, RMSEP: 0.153-0.198;</p> <p>Optimum wavelengths: (Peanut): R²_p: 0.981-0.988, RMSEP: 0.348-0.465; (Walnut): R²_p: 0.990-0.997, RMSEP: 0.170-0.324; (Peanut+Walnut): R²_p: 0.960-0.987, RMSEP: 0.373-0.645</p>	(Zhao et al., 2018)
	Flour	NIR-HSI	1118-2425 nm	Bench top	Detection and quantification of particles of ergot bodies in cereal flour	PLSDA	R ² _p : 0.99 and R ² _p : 0.87	(Vermeulen et al., 2017)
	Grain	FT-NIR	1100-2500 nm	Bench top	Determination of insect infestation analysing the change in uric acid, protein, final moisture content, thousand kernel weight, and hardness	PLSR	<p>(Moisture): R²_p=0.901 RMSECV=0.485 RPD=3.108</p> <p>(Protein): R²_p=0.938 RMSECV=0.248 RPD=3.971</p> <p>(Uric acid): R²_p=0.895 RMSECV=2.58 RPD=3.034</p> <p>(1000 kernel weight): R²_p=0.907 RMSECV=0.576 RPD=3.170</p> <p>(Hardness): R²_p=0.912 RMSECV=0.762 RPD=3.290</p>	(Mishra et al., 2018)

Table 1.2. (continued)

Category	Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Addition of ingredients	Bread	NIR	1000–2500 nm	Bench top	Quantification of potato flour added to Chinese steamed bread formulation	PLSR	$R^2_p=0.8865$, RPD=3.07	(Wang et al., 2018)
	Flour	NIR/ NIR-HSI	900–1700 nm/ 900–2500 nm	Portable/ Bench top	Classification, quantification and distribution of different types of fiber added to semolina	SIMCA, PLSR	Sensitivity: 0.6-1.0, Specificity: 0.9-1.0, R^2_p : 0.85-0.98, RMSEP: 0.5-1.0 R^2_p : 0.28 – 0.90 %LOF < 6%	(Badaró et al., 2019)
	Pasta	NIR-HSI	928 – 2524 nm	Bench top	Evaluation and quantification of fibre in enriched pasta	MCR-ALS	Similarity (pure spectra and recovered): 96% – 98%	(Badaró et al., 2021)
	Flour	NIR	850–1100 nm	Bench top	Determination of or the potato flour content in potato-wheat blended powders	PLSR	$R^2_p=0.9995$, SEP=0.69, SDr= 0.264, CVr=0.967	(Huang et al., 2018)

1.4.2. Determination of quality parameters

Spectral techniques have also shown their applicability in the determination of wheat flour and products quality parameters, which include protein, moisture, wet and dry gluten, sedimentation, enzyme activity, farinograph, and alveograph (Table 1.3) (Başlar & Ertugay, 2011; Miralbés, 2003). These parameters are directly connected to the final product. Therefore, to develop a high-quality product, it is imperative to determine these parameters to properly select the appropriate flour for each product (Chen et al., 2017). In this context, Başlar and Ertugay (2011) used a NIR spectrometer to analyse protein, wet and dry gluten contents and Zeleny sedimentation of wheat flours. The authors developed PLS and MLR models with NIR spectra of 120 varieties of bread wheat from different regions of Turkey. The models resulted in a very satisfactory prediction, with coefficients of determination of 0.985, 0.976, 0.953, and 0.924 for protein, wet gluten, dry gluten, and Zeleny sedimentation. The authors concluded that NIR spectroscopy was able to well predict quality parameters and associated the good performance of wet and dry gluten models to the protein content, which was well predicted by the models.

In a recent work, Chen et al. (2017) also applied NIR spectroscopy in the rapid determination of total protein and wet gluten in commercial wheat flour. In addition, the authors tested a synergy interval algorithm trying to improve the obtained models. Data pre-treatment was also evaluated to determine the best models. The models with better performance in determining total protein in wheat flour were Support Vector Regression (SVR) with data pre-treated with SNV+2nd derivative ($R^2_p = 0.888$ and $RMSEP = 0.512$) and PLS with data pre-treated with SNV ($R^2_p = 0.834$ and $RMSEP = 0.520$). In case of wet gluten determination, the best models were SVR with data pre-treated with SNV+1st derivative ($R^2_p = 0.781$ and $RMSEP = 1.303$) and PLS with raw data ($R^2_p = 0.673$ and $SEP = 1.503$). After variable selection, the model's performance improved for both SVR and PLS in both parameters' determination. However, SVR showed the best performance, with R^2_p of 0.906 and 0.862 and RMSEP of 0.425 and 1.122 % for total protein and wet gluten prediction, respectively. The authors concluded that using the full spectra or a selected interval, the protein content was better predicted than wet gluten. However, the synergy intervals were able to improve both predictions, indicating good applicability of NIR spectroscopy and chemometrics to rapidly predict these quality parameters in wheat flour.

Aiming to compare the performance of a hyperspectral imaging instrument with traditional NIR instruments for the prediction of food composition, Morales-Sillero et al. (2018) determined the protein content of wheat flour as an example to evaluate the accuracy of both techniques. This work includes the use of two NIR instruments, one NIR System DS2500 (NIR-DS) and one 5000 Autocup DVP6BX NIR instrument (NIR-Perstop), and one NIR hyperspectral line-scan camera (NIR-HSI). In addition, the authors tested two strategies to select calibration and validation sets. The first strategy split randomly 75% and 25% of samples in calibration and validation sets, respectively. The second strategy built the calibration model with the first 60 samples analysed, and the 19 remaining samples were used as the external validation set to check stability of the instruments over time. Moreover, the analysis of all the samples were carried out with over three consecutive days for all the three instruments. PCA was performed to have an overview of the datasets, and PLS models were developed with full spectra range and, also, with the common range among them (1120 – 2424 nm). The Hotelling's T^2 resulted in average values of 7.62, 6.87, and 15.38 for NIR-DS, NIR-Perstop, and NIR-HSI, respectively. In other words, the two NIR instruments were closer to the multivariate mean than HSI, indicating more variability between HSI than NIR data, probably due to the larger amount of information obtained by the HSI instrument. PLS models based on full wavelengths showed that NIR-DS had the best performance ($R^2_p=0.99$, RMSEP=0.14, and RPD=8.55), followed by NIR-Pestop ($R^2_p=0.99$, RMSEP=0.16 %, and RPD=7.56) and NIR-HSI ($R^2_p=0.99$, RMSEP=0.17 %, and RPD=7.07 %). By selecting the common wavelength range, the first strategy had the best performance, with the best prediction of NIR-DS ($R^2_p=0.99$, RMSEP=0.15 %, and RPD=8.08), followed by NIR-HSI ($R^2_p=0.99$, RMSEP=0.15 %, and RPD=7.92) and NIR-Pestop ($R^2_p=0.99$, RMSEP=0.16 %, and RPD=7.56). Although with the second strategy was still possible to obtain reliable models (R^2_p over 0.94, RMSEP above 0.23 %, and RPD above 5.15), prediction decreased compared to the other two procedures. This is an interesting approach in this work because it is associated with the instrument's stability over time. Therefore, it can be inferred that these instruments must be calibrated more than once if used for routine analysis. Moreover, this rises an important point insight about the external validation, which must be performed to test the robustness of the models. Overall, hyperspectral imaging proved the ability to quantify chemical parameters in wheat flour, with good potential to be applied in the industry.

The comparison between equipment was also made by Cecchini et al. (2021). A sensor with short-wavelength NIR (SW-NIR) range between 700 and 1100 nm and a handheld NIR spectrometer (1600-2400 nm) were compared as their performance to evaluate durum

wheat semolina quality. The quality parameters used for this analysis were protein content, gluten content, gluten index, sedimentation value, alveograph parameters (W and P/L), and GlutoPeak parameters (maximum consistency and total energy). As chemometric techniques, PLS Regression was used to predict the quality parameters using the data obtained by the two equipment and SIMCA was applied to divide the semolina samples into quality classes (best, good, and sufficient). The protein content was satisfactorily predicted using SW-NIR ($R_p = 0.9561$; RMSEC = 0.2903; SEP = 0.4899) and NIR ($R_p = 0.9788$; RMSEC = 0.2028; SEP = 0.3263) data but NIR had a better performance. The other quality parameters did not have a good performance: SW-NIR alveograph P/L parameter had R_p equal to 0.6672, and NIR alveograph W parameter had R_p equal to 0.7249 as the worst results. The authors stated that a not large data set could be the reason for the bad robustness of the models. The SIMCA had a classification performance worst in the NIR data (sensitivity = 60%) than in the SW-NIR data (sensitivity = 70%).

All quality parameters are fundamental to process control, developing equipment and products, determining operation efficiency, and so on. Other fundamental parameters of food powder, as particle size and bulk density, have also been evaluated by spectral techniques. Zhu et al. (2017) compared VIS/SW-NIR spectroscopy and hyperspectral scattering for the quantitative determination of bulk density and qualitative analysis of particle size in wheat flour samples. A total of 474 samples, including different varieties and size fractions, were analysed, and PLSR and PLS-DA models were developed to determine bulk density and classify particle size, respectively. PLSR results for NIR and HSI presented R^2_p of 0.55 and 0.87 and RMSEP of 57.80 and 30.18 mg mL⁻¹, respectively. Moreover, RPD values for NIR and HSI were 1.47 and 2.79, respectively. Based on these results, the authors concluded hyperspectral scattering was better than NIR for predicting bulk density of wheat flour. PLS-DA results showed good accuracy in classifying wheat flour samples, with 96.8 and 98.2% for NIR and HSI, respectively. Additionally, based on the confusion matrix of wheat flour classification of both techniques, the authors noticed that NIR spectroscopy performed better than hyperspectral scattering for samples with bigger particle size and vice versa. In general, the work showed that hyperspectral scattering technique is able to determine bulk density and particle size, which are two important parameters in the characterization of wheat flour samples.

The determination of fatty acids is also an important quality parameter during the storage of wheat flour. During storage, some chemical reactions as oxidation of lipids and production of carbonyl compounds happen, increasing the fatty acids values. Then, their increase led to acidity and bitterness of the final product. In a study, Jiang et al. (2020)

developed a portable NIR spectrometer for this determination. The storage of wheat flour was made at a temperature of 25 °C; each month occurred a sampling of the wheat flour bag (samples with 20 g) maintained in storage and the spectra acquisition of the samples. Firstly, relevant wavelength variables with high correlation with fatty acid values were optimized by variable combination population analysis (VCPA), as this technique randomly combine different variable sets; extreme learning machine (ELM) used these different combinations to develop a quantitative model to predict the fatty acid value in wheat flour at different storage duration. Considering four cases with different relevant wavelengths, R^2_p was above 0.96 in all models and RMSEP between 0.9375 and 1.0677 mg KOH/100g. The authors highlighted the importance of using VCPA for the elimination of useless information from NIR spectra and declared this capacity reduces the quantitative model complexity.

Other parameter in flour and its products composition is also an important factor in food quality that directly affects the final characteristic of the final product. The moisture content in the process of pasta-making, for instance, varies considerably from the beginning to the end of the process and, consequently, has a great influence on product stability. Therefore, it is important to know the product parameters to have more rigorous process control and a quality product. In this context, Czaja et al. (2018) measured and compared the moisture content in dough and pasta samples by different techniques, including FT-IR, FT-NIR, and FT-Raman. Samples were analysed at different stages of production: dough (30% of moisture), after pasta production (22% of moisture), and after pasta drying (9% of moisture), and PLSR models were developed for each dataset. FT-IR had the best performance for moisture determination ($R^2_p=0.998$, $RSE_{test}=2.50$), followed by FT-NIR ($R^2_p=0.997$, $RSE_{test}=3.41$) and FT-Raman ($R^2_p=0.991$, $RSE_{test}=5.18$). All three techniques showed potential for moisture quantification in pasta at different stages of production. Moreover, the work suggested that the lower quality of Raman compared to the other two techniques may be associated with the radiation absorption by water, which is weaker in Raman spectrum than in NIR or IR.

Table 1.3. Literature related to the determination of quality parameters in wheat flour and wheat-based products.

Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Flour	NIR	1100–2500 nm	Bench top	Determination of protein, wet and dry gluten contents and Zeleny sedimentation of flours	PLS, MLR	(Protein-PLS): $R^2_p=0.985$, SEP=0.377 (Wet gluten-PLS): $R^2_p=0.976$, SEP=1.36 (Dry gluten-PLS): $R^2_p=0.953$, SEP=0.635 (Zeleny sed-MLR): $R^2_p=0.924$, SEP=3.74 Full spectra: (Protein): (SVR): $R^2_p=0.888$, RMSEP=0.512; (PLS): $R^2_p=0.834$, RMSEP=0.520; (Wet gluten): (SVR): $R^2_p=0.781$, RMSEP=1.303; (PLS): $R^2_p=0.673$, RMSEP=1.503	(Başlar & Ertugay, 2011)
Flour	NIR	800–3030 nm	Bench top	Rapid determination of total protein and wet gluten in commercial wheat flour	SVR, PLS	Optimum wavelengths: (Protein): (SVR): $R^2_p=0.906$, RMSEP=0.425; (PLS): $R^2_p=0.887$, RMSEP=0.445; (Wet gluten): (SVR): $R^2_p=0.862$, RMSEP=1.233; (PLS): $R^2_p=0.779$, RMSEP=1.283 (NIR1): $R^2_p=0.98-0.99$, RMSEP=0.14-0.16% (NIR2): $R^2_p=0.98-0.99$, RMSEP=0.14-0.16% (HSI): $R^2_p=0.94-0.99$, RMSEP=0.15-0.23%	(Chen et al., 2017)
Flour	NIR/ NIR-HSI	400–2498 nm/ 1100–2498 nm/ 1128–2425 nm	Bench top	Quantification of protein in wheat using near infrared hyperspectral imaging in comparison with conventional near infrared spectroscopy	PLS	(NIR1): $R^2_p=0.98-0.99$, RMSEP=0.14-0.16% (NIR2): $R^2_p=0.98-0.99$, RMSEP=0.14-0.16% (HSI): $R^2_p=0.94-0.99$, RMSEP=0.15-0.23%	(Morales-Sillero et al., 2018)
Flour	NIR	899.22–1724 nm	Portable	Determination of fatty acids in wheat flour at different storage periods using NIR	VCPA, ELM	$R^2_p=0.9627-0.9675$ RMSEP=0.9375-1.0677 mg KOH/100g	(Jiang et al., 2020)

Table 1.3. (continued)

Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Semolina	SW-NIR/ NIR	700 – 1100 nm/ 1600 – 2400 nm	Portable	Comparison of a SW-NIR sensor and a NIR portable spectrometer to predict quality parameters of durum wheat semolina and to classify according to quality classes	PLSR, SIMCA	(Protein): SW-NIR $R^2_p=0.9462$, NIR $R^2_p=0.9692$; (Sedimentation value): SW-NIR $R^2_p=0.7268$, NIR $R^2_p=0.6599$; (W): SW-NIR $R^2_p=0.6555$, NIR $R^2_p=0.6712$; (P/L): SW-NIR $R^2_p=0.6178$, NIR $R^2_p=0.7070$; (BEM): SW-NIR $R^2_p=0.8062$, NIR $R^2_p=0.6392$; (Total energy): SW-NIR $R^2_p=0.6552$ (NIR): $R^2_p=0.55$, RMSEP=57.13 mg.mL ⁻¹ , RPD=1.47, Accuracy: 96.8% (HSI): $R^2_p=0.87$, RMSEP=30.02 mg.mL ⁻¹ , RPD=2.79, Accuracy: 98.2% (FT-IR): $R^2_p=0.998$, RSEPCalibration=2.54, RSEPValidation=2.15, RSEPTest=2.50; (FT-NIR): $R^2_p=0.997$, RSEPCalibration=3.16, RSEPValidation=3.32, RSEPTest=3.41; (FT-Raman): $R^2_p=0.991$, RSEPCalibration=5.56, RSEPValidation=5.67, RSEPTest=5.18.	(Cecchini et al., 2021)
Flour	VIS/SW-NIR/ NIR-HSI	400–1000 nm/ 500–1000 nm/	Portable/ Bench top	Bulk density determination and particle size classification of wheat flour	PLS,PLS DA	(HSI): $R^2_p=0.87$, RMSEP=30.02 mg.mL ⁻¹ , RPD=2.79, Accuracy: 98.2% (FT-IR): $R^2_p=0.998$, RSEPCalibration=2.54, RSEPValidation=2.15, RSEPTest=2.50; (FT-NIR): $R^2_p=0.997$, RSEPCalibration=3.16, RSEPValidation=3.32, RSEPTest=3.41; (FT-Raman): $R^2_p=0.991$, RSEPCalibration=5.56, RSEPValidation=5.67, RSEPTest=5.18.	(Zhu et al., 2017)
Pasta	FT-IR/ FT-NIR/ FT-Raman	2500–25000 nm/ 1100–2500 nm/ 2700–100000 nm	Bench top	Determination of moisture content in pasta	PLS	(FT-IR): $R^2_p=0.998$, RSEPCalibration=2.54, RSEPValidation=2.15, RSEPTest=2.50; (FT-NIR): $R^2_p=0.997$, RSEPCalibration=3.16, RSEPValidation=3.32, RSEPTest=3.41; (FT-Raman): $R^2_p=0.991$, RSEPCalibration=5.56, RSEPValidation=5.67, RSEPTest=5.18.	(Czaja et al., 2018)

1.4.3. Other applications

Spectral techniques have also been used for discrimination of different flours, products, or process types (Table 1.4). The thermal treatment used in pasta, for example, is performed to reduce mould spores and spoilage microorganisms; however, it has effects on the product quality due to interactions between proteins and starch. Aiming to evaluate heat treatment in flour by a VIS/SW-NIR hyperspectral imaging technique, Verdú et al. (2016) combined different temperatures and times to treat flour samples for cake production to characterize the heat treatment process by the imaging technique. In addition, cakes produced with these treated flours were analysed in terms of final height, mass loss during the baking process, texture profile analysis, and average area of bubble in crumb. The relationship between these cake parameters and hyperspectral data of flour, was evaluated using SVM. PCA was carried out on VIS/SW-NIR data, and results gave an overview of samples behaviour based on time and temperature of the heat treatment in flour. Moreover, a PCA was recalculated with important wavelengths, and a correlation between the PC1 scores and time and temperature were constructed, enabling to evaluate how the PC1 evolved regarding these two factors. According to these results, a non-linear regression model was developed based on the SVM technique, using the cake properties data as the independent variables and the PCA scores as the dependent variables. The coefficient of determination in the cross-validation ($R^2_{CV}=0.985$) suggested a dependency between the spectral information of flour treated by different heat treatments and cake properties. Thus, hyperspectral information can be an excellent alternative control and improve heat treatments of flour in the processing line.

These authors also evaluated the effect of wheat flour substitution and heat treatments of oat flour on bread properties by the same spectral technique. The spectral information of treated oat flour also showed a pattern regarding the temperature used for heat treatment. Non-linear regression models based on SVM were built according to the PC1 and PC2 scores and the two level of flour substitution (10 and 20%). The models had very satisfactory performance, with R^2_{CV} of 0.95 and 0.91, for PC1 and PC2, respectively, in a 10% level of flour substitution, and R^2_{CV} of 0.98 and 0.97, for PC1 and PC2, respectively, in a 20% level of flour substitution. Additionally, the spectral information of treated oat flour was tested regarding the bread properties. The R^2_{CV} ranged from 0.81 to 0.97 for a 10% of substitution and from 0.63 to 0.98 for a 20% of substitution. Therefore, hyperspectral imaging demonstrated the

applicability also for a non-destructive control tool to monitor heat treatment in oat flour to be replaced in bread formulation (Verdú et al., 2017).

Ziegler et al. (2016) studied dispersive and FT near-infrared reflectance spectroscopy for the rapid discrimination of kernels and flours of different wheat species, including bread wheat, spelt, durum, emmer, and einkorn. Due to the better performance of the industrial NIR spectrometer (dispersive spectroscopy) for kernel samples, spectra of flour samples were recorded with this instrument, and data were also analysed by PLS-DA. The model specificity and sensitivity ranged from 0.91-1.00 and 0.50-1.00, respectively, considering all five classes. In order to improve the discrimination power of the models, two classes (bread wheat and spelt) were considered, and the model recalculated, resulting in specificity values of 0.90 and 0.94, and sensitivity values of 0.94 and 0.90, for bread wheat and spelt, respectively. Moreover, the authors also investigated the adulteration of these flours, by mixing them in different ratios. A PLSR model was developed with the NIR spectral data, but the RMSECV was 22.6%, probably due to the small number of samples. Therefore, this work showed good discrimination of flour using NIR spectroscopy, with a small limitation on the number of classes. Additionally, it suggests an alternative technique for flour adulteration, corroborating with other works mentioned in this review.

Recently, Wadood et al. (2019) analysed wheat kernels and flour with NIR reflectance spectroscopy to better classify wheat according to the origin, production year, and genotypes. Samples of three distinct genotypes across 4 years from three different geographical areas of China were evaluated and PCA was carried out prior to multivariate ANOVA and LDA. ANOVA showed that all the factors affected the NIR spectra of kernel and flour samples and a significant difference among them was observed. LDA aimed to classify samples according to production year, genotypes, and geographical origin in each year. The accuracy of the model was 69% for production year, 69% for genotypes, and ranged from 72.2 to 100% for geographical origin for flour samples. This work showed that NIR spectroscopy could be a successful tool to determine geographical origin, production year, and genotype of flour samples, which is a valuable contribution to flour traceability.

Firmani et al. (2020) applied NIR spectroscopy to authenticate *Gragnano Pasta*, a P.G.I. product that is the eponym of the town in southern Italy where it is produced. In this study, 949 samples were used and two classes were considered: *Gragnano* and non-*Gragnano*. The proposed chemometric methods were PLS-DA and SIMCA. The models' performance was very satisfactory: PLS-DA was able to classify correctly 100% of *Gragnano* samples and 98.10% of the non-*Gragnano* samples using mean centring and 1st derivative as pre-treatments,

whereas SIMCA had sensitivity equal to 96.57% and specificity equal to 100% using mean centring and SNV as pre-treatments. Considering this performance, NIR spectroscopic coupled with PLS-DA and SIMCA can be a useful tool for the authentication of *Gragnano* pasta.

Shelf life is also an important factor to be considered in the food industry, since some problems may occur during this period, as spoilage or changes in texture or taste, concern both the industry and the consumers. Therefore, the use of technologies to improve quality control and, consequently, extended the shelf life of the product, are very important. In this context, NIR Spectroscopy was studied by Cevoli et al. (2015) as a fast and non-destructive tool to discriminate samples of *Piadina*, which is an Italian bread, with different formulations, and to estimate the storage time. PCA was carried out with three out of five formulations and the scores plot showed a clear separation between the clusters. Moreover, PLSR showed a good prediction of storage time (R^2_P ranging from 0.959 to 0.969 and RMSEP from 1.87 to 1.98), suggesting NIR spectroscopy as a good alternative for prediction of storage time of *Piadina* and the possibility of using such technique for other types of bread.

The use of edible coatings in food products has gained much interest as an alternative for protection regarding external factors, and for preventing or reducing microbial or oxidant activity. One of the important steps in applying an edible coating is the drying time. Therefore, studies were carried out to rapidly monitor the drying process of edible coating on bread surfaces (Chakravartula et al., 2019). Samples were tested with one layer, two layers and two different drying temperatures (25 and 60 °C). A FT-NIR spectrometer was used to provide spectral information of samples and PCA and PLSR were carried out to evaluate data. PCA was carried out as an exploratory analysis to identify and discriminate samples submitted to different drying times. In general, a clear separation was observed among samples regarding the drying time, and the loadings plot suggested that this separation may be due to moisture and protein content. Therefore, PLSR was performed to predict the moisture content in the coated bread. The results showed a successful performance of the model, with R^2_P ranging from 0.902 to 0.963 and RMSEP from 2.51 to 3.15%. Based on these results, this work suggests the applicability of NIR spectroscopy in moisture prediction and as a potential technique for drying time determination.

In a different approach, Neves et al. (2019) investigated the applicability of a handheld NIR spectrometer for rapid determination of nutritional parameters of pasta/sauce blends. The authors combined five different kinds of pasta and five different sauces in five blend ratios, totalling 125 samples. The nutritional value of each mixture was determined according to the information provided in the package. Spectral data were then, recorded and

PLSR models were developed for each nutritional parameter. The prediction showed reliable values for all the parameters, with R^2_p of 0.86, 0.85, 0.89, 0.90, 0.86 and 0.88, RMSEP of 10.64, 3.59, 0.95, 1.11, 1.39 and 0.61, and RPD of 2.02, 2.54, 2.77, 2.45, 2.26, 2.19, for energy, carbohydrate, fat, fibre, protein, and sugar, respectively. As many other works mentioned, this work shows that NIR spectroscopy can be a very powerful tool for quality control, even using a handheld instrument to predict parameters established in the package.

Table 1.4. Literature related to other applications of NIR, FT-NIR, and NIR-HSI in wheat flour and wheat-based products.

Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Flour/ Cake	VIS/SW- NIR-HSI	400–1000 nm	Bench top	Characterization of heat treatment process of cake wheat flour	PCA, SVM	R ² _{CV} =0.985	(Verdú et al., 2016)
Flour/ Bread	VIS/SW- NIR-HSI	400–1000 nm	Bench top	Characterization of heat treatment in oat flour and evaluation of hyperspectral data and bread	PCA, PLS	<p>10% of substitution: R²_{CV}: 0.81-0.97 RMSECV: 0.00-4.03</p> <p>20% of substitution: R²_{CV}: 0.63-0.98 RMSECV: 0.00-6.00</p> <p>Five classes: <i>FT-NIR spectroscopy:</i> (Kernel): Sensitivity: 0.35-0.95, Specificity: 0.86-0.94, Accuracy: 81.0-94.0%; <i>Dispersive spectroscopy:</i> (Kernel): Sensitivity: 0.65-0.95, Specificity: 0.90-1.00, Accuracy: 88.0-99.0%; (Flour): Sensitivity: 0.50-1.00, Specificity: 0.88-1.00, Accuracy: 85.1-100%;</p> <p>Two classes: (Bread wheat flour): Sensitivity: 0.94, Specificity: 0.90, Accuracy: 92.1%; (Spelt flour): Sensitivity: 0.90, Specificity: 0.94, Accuracy: 92.1%;</p> <p>(PLSR): R²_C=0.966, RMSEC=5.2% RMSEC=22.6%</p>	(Verdú et al., 2017)
Kernel/ Flour	NIR/ FT-NIR	1200–2400 nm/ 650–2500 nm	Bench top	Rapid discrimination of kernels and flours of different wheat species	PCA, PLSDA, PLS	<p>Sensitivity: 0.50-1.00, Specificity: 0.88-1.00, Accuracy: 85.1-100%;</p> <p>Two classes: (Bread wheat flour): Sensitivity: 0.94, Specificity: 0.90, Accuracy: 92.1%; (Spelt flour): Sensitivity: 0.90, Specificity: 0.94, Accuracy: 92.1%;</p> <p>(PLSR): R²_C=0.966, RMSEC=5.2% RMSEC=22.6%</p>	(Ziegler et al., 2016)

Table 1.4. (continued).

Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Kernel/ Flour	NIR	950–1650 nm	Bench top	Evaluation of wheat according to the origin, production year and genotypes	PCA, LDA	Flour: (Origin): Accuracy: 72.2-100% (Year): Accuracy: 69% Accuracy: 69% Kernel: (Origin): Accuracy: 61.1-88.9% (Year): Accuracy: 55.17% (Genotypes): Accuracy: 75.86% (PLS-DA):	(Wadood et al., 2019)
Pasta	FT-NIR	1000-2500 nm	Bench top	Differentiation between P.G.I. <i>Gragnano</i> pasta and non- <i>Gragnano</i> pasta	PLS-DA, SIMCA	Sensitivity=100%, Specificity=98.10% (SIMCA): Sensitivity=96.57%, Specificity=100%	(Firmani et al., 2020)
Bread	NIR	800–2500 nm	Bench top	Discrimination of Italian bread samples as a function of formulation and prediction of storage time	PCA, PLS	R ² _p : 0.959-0.969, RMSEP: 1.87-1.98	(Cevoli et al., 2015)
Bread	FT-NIR	800–2500 nm	Bench top	Monitoring the drying process of edible coating on bread surfaces.	PCA, PLS	R ² _p : 0.902-0.963, RMSEP: 2.51-3.15	(Chakravartula et al., 2019)

Table 1.4. (continued).

Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Pasta	NIR	908–1676 nm	Handheld	Determination of nutritional parameters of pasta/sauce blends	PLS	(Energy): $R^2_p=0.86$, RMSEP=10.64, RPD=2.02; (Carbohydrate): $R^2_p=0.85$, RMSEP=3.59, RPD=2.54; (Fat): $R^2_p=0.89$, RMSEP=0.95, RPD=2.77; (Fibre): $R^2_p=0.90$, RMSEP=1.11, RPD=2.45; (Protein): $R^2_p=0.86$, RMSEP=1.39, RPD=2.26; (Energy): $R^2_p=0.88$, RMSEP=0.61, RPD=2.19	(Neves et al., 2019)

1.4.4. Data fusion

The individual study of spectral techniques sometimes is not enough to achieve the desired results. In this point of view, there are advantages in using data fusion combined with chemometrics, which enable the use of more accurate information about samples and, consequently, provides more solid results. This method integrates the information from different techniques using three strategies: low-, mid-, and high-level data fusion (Chen et al., 2017).

In low-level data fusion, the data is concatenated before multivariate pre-processing and analysis. In other words, the data is fused in its raw form (Mishra et al., 2019). The most relevant information is selected in mid-level data fusion and extracted for later fusion and analysis (Chen et al., 2017). On the other hand, in high-level data fusion, the outputs obtained by the models developed for each data set individually are fused to improve the final result (Mishra et al., 2019). Only a few studies were found using data fusion in wheat flour and its products determinations. Chen et al. (2017) carried out studies on the rapid determination of farinograph parameters of wheat flour combining NIR and MIR spectral regions. First, the spectra of each technique were analysed individually by PLSR models but did not achieve good results. To solve this issue, both data were low and mid-level fused. In low-level data fusion, NIR and MIR data were combined and, then, analysed. In mid-level data fusion, an algorithm called forward interval data fusion PLS (fi-DF-PLS) was implemented in the study to select representative variables in the MIR and NIR regions and improve the performance of the model. Low-level data fusion performed worse than each method individually, probably due to the high amount of irrelevant information added to the model. On the other hand, mid-level data fusion improved the performance of the models. The models parameters with better prediction power for water absorption, dough development time, dough stability, and degree of softening were R (correlation coefficient) = 0.96, 0.94, 0.95 and 0.94, RMSEP = 0.521 %, 0.514 (min), 0.640 (BU) and 15.014 (BU), and RPD = 3.812, 2.814, 2.566 and 2.647. This study shows the potential of data fusion and spectral techniques as NIR and MIR for prediction of flour parameters.

In another approach, Ringsted et al. (2017) developed two-dimensional correlation spectroscopy on near and mid-infrared spectra of wheat bread crumb during aging. First, important regions in MIR and NIR infrared were identified and correlated according to bread hardness. An important absorption band at 1047 cm^{-1} was observed in the MIR region regarding amylopectin retrogradation, positively correlated to bread hardness. This band had a very

satisfactory correlation with three bands in the NIR region, including a band at 910 nm ($R^2=0.88$ to hardness), at 1688 nm ($R^2=0.97$ to hardness), and at 2288 nm ($R^2=0.97$ to hardness). Another important band in the MIR region was found at 1022 cm^{-1} that presented a high negative correlation to bread crumb hardness. This band seemed to be related to different levels of hydrogen bonds, and it was correlated with NIR absorbance bands at 974 nm ($R^2=0.96$ to hardness), 1412 nm ($R^2=0.94$ to hardness), and 2258 nm ($R^2=0.92$ to hardness). Moreover, PCA was performed on NIR and MIR data individually, and the first principal components from both techniques were correlated, with an $R^2=0.98$. The results showed that bread staling processes could be evaluated by near and mid-infrared spectroscopy and high-level data fusion. The works found in the literature regarding the use of data fusion on the wheat flour and wheat-flour products were summarised on Table 1.5.

Table 1.5. Literature related to use of data fusion in the analysis of wheat flour and wheat-based products with NIR and MIR.

Sample	Technique	Spectral Range	Type of equipment	Objective	Statistical Method	Performance	Reference
Flour	MIR/ NIR	2500–20000 nm/ 800–3000 nm	Bench top	Investigation of the potential of combining NIR and mid-infrared (MIR) spectral regions to predict wheat flour farinograph quality properties: water absorption (WA), dough development time (DDT), dough stability (DS), and degree of softening (DOS)	PLSR	(WA): $R_p=0.96$; RMSEP=0.521; RPD=3.812 (DDT): $R_p=0.94$; RMSEP=0.514; RPD =2.814 (DS): $R_p=0.95$; RMSEP=0.640; RPD =2.566 (DOS): $R_p=0.94$; RMSEP=15.014; RPD =2.647	(Chen et al., 2017)
Bread	NIR/MIR	400–2500 nm/ 2500–20000 nm	Bench top	Development of a two-dimensional correlation spectroscopy on near- and mid-infrared spectra of wheat bread crumb during aging	PCA	$R^2: 0.98$	(Ringsted et al., 2017)

1.5. Conclusions and future trends

This review summarized spectral techniques for wheat flour and wheat-based products, more specifically near-infrared, Fourier-transform near-infrared and hyperspectral imaging. Overall, the techniques have been used in different range within the near infrared spectrum, with different instruments. The number of samples, acquisition mode and statistical methods performed varied according to the study aim. Another important detail is the condition of analysis, which was performed in the laboratory in most of the studies, and there is not a lot of information about the application in the industry.

These techniques showed to be more related to determine composition, authentication, and quality parameters. However, future works could explore other important parameters in the cereal and bakery industry as to defects of pasta or bread. In addition, assuming good performance of quantitative and qualitative models for parameters in wheat flour and wheat-based product, studies can be carried out to determine other parameters based on these easily detected. An example of this methodology is the work of Mishra et al. (2018) here reviewed that used quality parameters of wheat grain to determine insect infestation.

Only a few works have applied the spectral technique in-line, suggesting a deep study in order to apply these techniques in the process line so that the industry would have a more controlled process. The combination of off-line and in-line measurements could be a first step on the use of NIR in in-line analysis. Also, handheld spectrometers could be used in the process line; therefore, the literature comparing the application of benchtop and handheld equipment in different situations is needed for the application of these equipment to develop new equipment.

The application of the techniques here approached for the determination of wheat based products geographical origin and production process was demonstrated by Wadood et al. (2019) and Firmani et al. (2020). The adulteration of these products can occur and are not easily noticed; for consumer protection, these techniques can be more studied. Additionally, the use of data fusion for the determination of quality parameters in wheat flour and wheat-based products is still scarce. One of the biggest challenges of using near infrared is the development of models to quantify parameters at a low level. Therefore, data fusion could allow the association of techniques that, together, could be more sensitive to small concentrations, enabling the development of more reliable and accurate models for quantitative determinations.

Acknowledgments

Funding: This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) [Finance Code 001]; and São Paulo Research Foundation (FAPESP) [grant numbers 2015/24351-2, 2017/17628-3, 2019/06842-0, 2020/09653-0]. This work was also partially funded through projects AEI PID2019-107347RR-C31 and PID2019-107347RR-C32, and the European Union through the European Regional Development Fund (ERDF) of the Generalitat Valenciana 2014-2020.

Declaration of competing interest

Authors declare no conflict of interest.

Author contributions

Amanda Teixeira Badaró: Conceptualization, Writing - Original Draft, Writing - Review & Editing. **João Pedro Hebling e Tavares:** Writing - Original Draft, Writing - Review & Editing. **Jose Blasco-Ivars:** Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration. **Nuria Aleixos Borrás:** Writing - Review & Editing, Supervision, Project administration. **Douglas Fernandes Barbin:** Conceptualization, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision, Project administration, Funding acquisition.

References

- ACCC. (2000). *Approved methods of the American Association of Cereal Chemists*. (10th ed.). St. Paul, Minnesota: The Association.
- Ait-Kaddour, A., & Cuq, B. (2011). Dynamic NIR Spectroscopy to Monitor Wheat Product Processing: A Short Review. *American Journal of Food Technology*, 6(3), 186–196. <https://doi.org/10.3923/ajft.2011.186.196>
- Akhtar, S., Anjum, F. M., Rehman, S. U., Sheikh, M. A., & Farzana, K. (2008). Effect of fortification on physico-chemical and microbiological stability of whole wheat flour. *Food Chemistry*, 110(1), 113–119. <https://doi.org/10.1016/j.foodchem.2008.01.065>
- Amigo, J.M., & Ravn, C. (2009). Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging. *European Journal of Pharmaceutical Sciences*, 37(2), 76–82. <https://doi.org/10.1016/j.ejps.2009.01.001>.

- Amigo, J. M. (2020). Hyperspectral and multispectral imaging: setting the scene. In José Manuel Amigo (Eds.), *Data Handling in Science and Technology* (pp. 3–16). Elsevier.
- Amigo, J. M., Olmo, A. del, Engelsens, M. M., Lundkvist, H., & Engelsens, S. B. (2021). Staling of white wheat bread crumb and effect of maltogenic α -amylases. Part 3: Spatial evolution of bread staling with time by near infrared hyperspectral imaging. *Food Chemistry*, 353, 129478. <https://doi.org/10.1016/j.foodchem.2021.129478>
- Ayvaz, H., Plans, M., Towers, B. N., Auer, A., & Rodriguez-Saona, L. E. (2015). The use of infrared spectrometers to predict quality parameters of cornmeal (corn grits) and differentiate between organic and conventional practices. *Journal of Cereal Science*, 62, 22–30. <https://doi.org/10.1016/j.jcs.2014.12.004>
- Badaró, A. T., Amigo, J. M., Blasco, J., Aleixos, N., Ferreira, A. R., Clerici, M. T. P. S., & Barbin, D. F. (2021). Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fiber distribution in enriched pasta. *Food Chemistry*, 343, 128517. <https://doi.org/10.1016/j.foodchem.2020.128517>
- Badaró, A. T., Morimitsu, F. L., Ferreira, A. R., Clerici, M. T. P. S., & Fernandes Barbin, D. (2019). Identification of fiber added to semolina by near infrared (NIR) spectral techniques. *Food Chemistry*, 289, 195–203. <https://doi.org/10.1016/j.foodchem.2019.03.057>
- Başlar, M., & Ertugay, M. F. (2011). Determination of protein and gluten quality-related parameters of wheat flour using near-infrared reflectance spectroscopy (NIRS). *Turkish Journal of Agriculture and Forestry*, 35(2), 139–144. <https://doi.org/10.3906/tar-0912-507>
- Benjakul, S., & Karnjanapratum, S. (2018). Characteristics and nutritional value of whole wheat cracker fortified with tuna bone bio-calcium powder. *Food Chemistry*, 259, 181–187. <https://doi.org/10.1016/j.foodchem.2018.03.124>
- Bhuvaneshwari, K., Fields, P. G., White, N. D. G., Sarkar, A. K., Singh, C. B., & Jayas, D. S. (2011). Image analysis for detecting insect fragments in semolina. *Journal of Stored Products Research*, 47(1), 20–24. <https://doi.org/10.1016/j.jspr.2010.08.003>
- Bizerra Brito, A. L., Pereira Santos, A. V., Tavares Melo Milanez, K. D., José Coelho Pontes, M., & Bezerra Lira Pontes, L. F. (2017). Calibration transfer of flour NIR spectra between benchtop and portable instruments. *Analytical Methods*, 9(21), 3184–3190. <https://doi.org/10.1039/c7ay00391a>
- Bustos, M. C., Perez, G. T., & Leon, A. E. (2015). Structure and quality of pasta enriched with functional ingredients. *RSC Advances*, 5(39), 30780–30792. <https://doi.org/10.1039/c4ra11857j>
- Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2017). Application of calibrations to

- hyperspectral images of food grains: example for wheat falling number. *Journal of Spectral Imaging*, 6, 1–15. <https://doi.org/10.1255/jsi.2017.a4>
- Caporaso, N., Whitworth, M. B., & Fisk, I. D. (2018). Near-Infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. In *Applied Spectroscopy Reviews*, 53 (8), 667–687. <https://doi.org/10.1080/05704928.2018.1425214>
- Casian, T., Gavan, A., Iurian, S., Porfire, A., Toma, V., Stiufluic, R., & Tomuta, I. (2021). Testing the Limits of a Portable NIR Spectrometer: Content Uniformity of Complex Powder Mixtures Followed by Calibration Transfer for In-Line Blend Monitoring. *Molecules*, 26(4), 1129. <https://doi.org/10.3390/molecules26041129>
- Cecchini, C., Antonucci, F., Costa, C., Marti, A., & Menesatti, P. (2021). Application of near-infrared handheld spectrometers to predict semolina quality. *Journal of the Science of Food and Agriculture*, 101(1), 151–157. <https://doi.org/10.1002/jsfa.10625>
- Cevoli, C., Gianotti, A., Troncoso, R., & Fabbri, A. (2015). Quality evaluation by physical tests of a traditional Italian flat bread Piadina during storage and shelf-life improvement with sourdough and enzymes. In *European Food Research and Technology*, 240 (6), 1081–1089. <https://doi.org/10.1007/s00217-015-2429-7>
- Chakravartula, S. S. N., Cevoli, C., Balestra, F., Fabbri, A., & Dalla Rosa, M. (2019). Evaluation of drying of edible coating on bread using NIR spectroscopy. *Journal of Food Engineering*, 240, 29–37. <https://doi.org/10.1016/j.jfoodeng.2018.07.009>
- Che, W., Sun, L., Zhang, Q., Zhang, D., Ye, D., Tan, W., Wang, L., & Dai, C. (2017). Application of Visible/Near-Infrared Spectroscopy in the Prediction of Azodicarbonamide in Wheat Flour. *Journal of Food Science*, 82(10), 2516–2525. <https://doi.org/10.1111/1750-3841.13859>
- Chen, J., Ye, F., & Zhao, G. (2017). Rapid determination of farinograph parameters of wheat flour using data fusion and a forward interval variable selection algorithm. *Analytical Methods*, 9(45), 6341–6348. <https://doi.org/10.1039/c7ay02065a>
- Chen, J., Zhu, S., & Zhao, G. (2017). Rapid determination of total protein and wet gluten in commercial wheat flour using siSVR-NIR. *Food Chemistry*, 221, 1939–1946. <https://doi.org/10.1016/j.foodchem.2016.11.155>
- Chen, X., Siesler, H. W., & Yan, H. (2021). Rapid analysis of wheat flour by different handheld near-infrared spectrometers: A discussion of calibration model maintenance and performance comparison. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy*, 252, 119504. <https://doi.org/10.1016/j.saa.2021.119504>

- Cortés, V., Blasco, J., Aleixos, N., Cubero, S., Talens, P. (2019a) Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends in Food Science & Technology*, 85, 138–148. <https://doi.org/10.1016/j.tifs.2019.01.015>.
- Cortés, V., Cubero, S., Blasco, J., Aleixos, N., Talens, P. (2019b). In-line application of visible and near-infrared diffuse reflectance spectroscopy to identify apple varieties. *Food and Bioprocess Technology*, 12, 1021-1030. <https://doi.org/10.1007/s11947-019-02268-0>
- Czaja, T., Kuzawińska, E., Sobota, A., & Szostak, R. (2018). Determining moisture content in pasta by vibrational spectroscopy. *Talanta*, 178, 294–298. <https://doi.org/10.1016/j.talanta.2017.09.050>
- Dave, G., & Modi, H. (2018). FT-IR method for estimation of phytic acid content during bread-making process. *Journal of Food Measurement and Characterization*, 12(3), 2202–2208. <https://doi.org/10.1007/s11694-018-9836-y>
- De Juan, A., Jaumot, J., & Tauler, R. (2014). Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Analytical Methods*, 6(14), 4964–4976. <https://doi.org/10.1039/c4ay00571f>
- de Juan, A., & Tauler, R. (2021). Multivariate Curve Resolution: 50 years addressing the mixture analysis problem – A review. *Analytica Chimica Acta*, 1145, 59–78. <https://doi.org/10.1016/j.aca.2020.10.051>
- Femenias, A., Gatiús, F., Ramos, A. J., Sanchis, V., & Marín, S. (2020). Use of hyperspectral imaging as a tool for Fusarium and deoxynivalenol risk management in cereals: A review. *Food Control*, 108, 106819. <https://doi.org/10.1016/j.foodcont.2019.106819>
- Femenias, A., Gatiús, F., Ramos, A. J., Sanchis, V., & Marín, S. (2021). Near-infrared hyperspectral imaging for deoxynivalenol and ergosterol estimation in wheat samples. *Food Chemistry*, 341, 128206. <https://doi.org/10.1016/j.foodchem.2020.128206>
- Feng, Y.-Z., & Sun, D.-W. (2012). Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review. *Critical Reviews in Food Science and Nutrition*, 52(11), 1039–1058. <https://doi.org/10.1080/10408398.2011.651542>
- Ferreira, M. M. C. (2015). *Quimiometria- Conceitos, Métodos e Aplicações*. Editora Unicamp.
- Firmani, P., La Piscopia, G., Bucci, R., Marini, F., & Biancolillo, A. (2020). Authentication of P.G.I. Gragnano pasta by near infrared (NIR) spectroscopy and chemometrics. *Microchemical Journal*, 152, 104339. <https://doi.org/10.1016/j.microc.2019.104339>
- Forchetti, D. A. P., & Poppi, R. J. (2017). Use of NIR hyperspectral imaging and multivariate curve resolution (MCR) for detection and quantification of adulterants in milk powder. *LWT -*

- Food Science and Technology*, 76, 337–343. <https://doi.org/10.1016/j.lwt.2016.06.046>
- González-Martín, M. I., Escuredo, O., Hernández-Jiménez, M., Revilla, I., Vivar-Quintana, A. M., Martínez-Martín, I., & Hernández-Ramos, P. (2021). Prediction of stable isotopes and fatty acids in subcutaneous fat of Iberian pigs by means of NIR: A comparison between benchtop and portable systems. *Talanta*, 224, 121817. <https://doi.org/10.1016/j.talanta.2020.121817>
- Grassi, S., Cardone, G., Bigagnoli, D., & Marti, A. (2018). Monitoring the sprouting process of wheat by non-conventional approaches. *Journal of Cereal Science*, 83, 180–187. <https://doi.org/10.1016/j.jcs.2018.08.007>
- Haraszi, R., Chassaing, H., Maquet, A., & Ulberth, F. (2011). Analytical Methods for Detection of Gluten in Food—Method Developments in Support of Food Labeling Legislation. *Journal of AOAC International*, 94(4), 1006–1025. <https://doi.org/10.1093/jaoac/94.4.1006>
- Huang, H., Yu, H., Xu, H., & Ying, Y. (2008). Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: A review. *Journal of Food Engineering*, 87 (3), 303–313. <https://doi.org/10.1016/j.jfoodeng.2007.12.022>
- Huang, Y., Xu, F., Hu, H., Dai, X., & Zhang, H. (2018). Development of a predictive model to determine potato flour content in potato-wheat blended powders using near-infrared spectroscopy. *International Journal of Food Properties*, 21(1), 2030–2036. <https://doi.org/10.1080/10942912.2018.1502199>
- Hussain, N., Sun, D.-W., & Pu, H. (2019). Classical and emerging non-destructive technologies for safety and quality evaluation of cereals: A review of recent applications. *Trends in Food Science & Technology*, 91, 598–608. <https://doi.org/10.1016/j.tifs.2019.07.018>
- Jiang, H., Liu, T., & Chen, Q. (2020). Quantitative detection of fatty acid value during storage of wheat flour based on a portable near-infrared (NIR) spectroscopy system. *Infrared Physics & Technology*, 109, 103423. <https://doi.org/10.1016/j.infrared.2020.103423>
- Kirchler, C. G., Pezzei, C. K., Beć, K. B., Mayr, S., Ishigaki, M., Ozaki, Y., & Huck, C. W. (2017). Critical evaluation of spectral information of benchtop vs. portable near-infrared spectrometers: Quantum chemistry and two-dimensional correlation spectroscopy for a better understanding of PLS regression models of the rosmarinic acid content in Rosmarin. *Analyst*, 142(3), 455–464. <https://doi.org/10.1039/c6an02439d>
- Kumagai, M., Matsuura, N., Li, H., Ohisa, N., Amano, T., & Ogawa, N. (2004). Application of a Portable near Infrared Spectrometer for the Manufacturing of Noodle Products. *Journal of Near Infrared Spectroscopy*, 12(2), 127–131. <https://doi.org/10.1255/jnirs.417>
- Kumar, Y., & Chandrakant Karne, S. (2017). Spectral analysis: A rapid tool for species detection in meat products. *Trends in Food Science & Technology*, 62, 59–67.

<https://doi.org/10.1016/j.tifs.2017.02.008>

Levasseur-Garcia, C. (2018). Updated overview of infrared spectroscopy methods for detecting mycotoxins on cereals (corn, wheat, and barley). *Toxins*, *10*(1).

<https://doi.org/10.3390/toxins10010038>

Lohumi, S., Lee, H., Kim, M. S., Qin, J., Kandpal, L. M., Bae, H., Rahman, A., & Cho, B. K. (2018). Calibration and testing of a Raman hyperspectral imaging system to reveal powdered food adulteration. *PLoS ONE*, *13*(4), 1–16. <https://doi.org/10.1371/journal.pone.0195253>

Lohumi, S., Lee, S., Lee, H., & Cho, B.-K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. In *Trends in Food Science & Technology journal*, *46*(1), 85-98. <https://doi.org/10.1016/j.tifs.2015.08.003>

Lorusso, A., Verni, M., Montemurro, M., Coda, R., Gobbetti, M., & Rizzello, C. G. (2017). Use of fermented quinoa flour for pasta making and evaluation of the technological and nutritional features. *LWT - Food Science and Technology*, *78*, 215–221. <https://doi.org/10.1016/j.lwt.2016.12.046>

Malegori, C., Nascimento Marques, E. J., de Freitas, S. T., Pimentel, M. F., Pasquini, C., & Casiraghi, E. (2017). Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. *Talanta*, *165*, 112–116. <https://doi.org/10.1016/j.talanta.2016.12.035>

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chem. Soc. Rev.*, *43*(24), 8200–8214. <https://doi.org/10.1039/C4CS00062E>

Manley, M., Van Zyl, L., & Osborne, B. G. (2002). Using fourier transform near infrared spectroscopy in determining kernel hardness, protein and moisture content of whole wheat flour. *Journal of Near Infrared Spectroscopy*, *10*(1), 71–76. <https://doi.org/10.1255/jnirs.323>

Mayr, S., Beć, K. B., Grabska, J., Wiedemair, V., Pürgy, V., Popp, M. A., Bonn, G. K., & Huck, C. W. (2021). Challenging handheld NIR spectrometers with moisture analysis in plant matrices: Performance of PLSR vs. GPR vs. ANN modelling. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *249*, 119342. <https://doi.org/10.1016/j.saa.2020.119342>

Mayr, S., Schmelzer, J., Kirchler, C. G., Pezzei, C. K., Beć, K. B., Grabska, J., & Huck, C. W. (2021). Theae nigrae folium: Comparing the analytical performance of benchtop and handheld near-infrared spectrometers. *Talanta*, *221*, 121165. <https://doi.org/10.1016/j.talanta.2020.121165>

Menesatti, P., Antonucci, F., Pallottino, F., Bucarelli, F. M., & Costa, C. (2014).

- Spectrophotometric Qualification of Italian Pasta Produced by Traditional or Industrial Production Parameters. *Food and Bioprocess Technology*, 7(5), 1364–1370. <https://doi.org/10.1007/s11947-013-1138-0>
- Menesatti, P., & Bucarelli, A. (2007). Quality classification of Italian wheat durum spaghetti by means of different spectrophotometric techniques. *Optics for Natural Resources, Agriculture, and Foods II*, 6761, 67610D. <https://doi.org/10.1117/12.735553>
- Miralbés, C. (2003). Prediction chemical composition and alveograph parameters on wheat by near-infrared transmittance spectroscopy. *Journal of Agricultural and Food Chemistry*, 51(21), 6335–6339. <https://doi.org/10.1021/jf034235g>
- Mishra, G., Srivastava, S., Panda, B. K., & Mishra, H. N. (2018). Rapid Assessment of Quality Change and Insect Infestation in Stored Wheat Grain Using FT-NIR Spectroscopy and Chemometrics. *Food Analytical Methods*, 11(4), 1189–1198. <https://doi.org/10.1007/s12161-017-1094-9>
- Mishra, P., Herrero-Langreo, A., Barreiro, P., Roger, J. M., Diezma, B., Gorretta, N., & Lleó, L. (2015). Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis. *Journal of Near Infrared Spectroscopy*, 23(1), 15–22. <https://doi.org/10.1255/jnirs.1141>
- Mishra, P., Nordon, A., Mohd Asaari, M. S., Lian, G., & Redfern, S. (2019). Fusing spectral and textural information in near-infrared hyperspectral imaging to improve green tea classification modelling. *Journal of Food Engineering*, 249, 40–47. <https://doi.org/10.1016/j.jfoodeng.2019.01.009>
- Morales-Sillero, A., Pierna, J. A. F., Sinnaeve, G., Dardenne, P., & Baeten, V. (2018). Quantification of protein in wheat using near infrared hyperspectral imaging: Performance comparison with conventional near infrared spectroscopy. *Journal of Near Infrared Spectroscopy*, 26(3), 186–195. <https://doi.org/10.1177/0967033518780506>
- Neves, M. D. G., Poppi, R. J., & Siesler, H. W. (2019). Rapid determination of nutritional parameters of pasta/sauce blends by handheld near-infrared spectroscopy. *Molecules*, 24(11). <https://doi.org/10.3390/molecules24112029>
- Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*, 1026, 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>
- Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology*, 46(2), 211–230. <https://doi.org/10.1016/j.tifs.2015.10.002>

- Rachmawati, Rohaeti, E., & Rafi, M. (2017). Combination of near infrared spectroscopy and chemometrics for authentication of taro flour from wheat and sago flour. *Journal of Physics: Conference Series*, 835(1). <https://doi.org/10.1088/1742-6596/835/1/012011>
- Ringsted, T., Siesler, H. W., & Engelsen, S. B. (2017). Monitoring the staling of wheat bread using 2D MIR-NIR correlation spectroscopy. *Journal of Cereal Science*, 75, 92–99. <https://doi.org/10.1016/j.jcs.2017.03.006>
- Rodríguez, S. D., Rolandelli, G., & Buera, M. P. (2019). Detection of quinoa flour adulteration by means of FT-MIR spectroscopy combined with chemometric methods. *Food Chemistry*, 274, 392–401. <https://doi.org/10.1016/j.foodchem.2018.08.140>
- Rukundo, I. R., Danao, M.-G. C., Mitchell, R. B., Masterson, S. D., & Weller, C. L. (2021). Comparing the Use of Handheld and Benchtop NIR Spectrometers in Predicting Nutritional Value of Forage. *Applied Engineering in Agriculture*, 37(1), 171–181. <https://doi.org/10.13031/aea.14157>
- Sendin, K., Williams, P. J., & Manley, M. (2018). Near infrared hyperspectral imaging in quality and safety evaluation of cereals. *Critical Reviews in Food Science and Nutrition*, 58(4), 575–590. <https://doi.org/10.1080/10408398.2016.1205548>
- Stellacci, A. M., Castrignanò, A., Diacono, M., Troccoli, A., Ciccarese, A., Armenise, E., Gallo, A., De Vita, P., Lonigro, A., Mastro, M. A., & Rubino, P. (2012). Combined approach based on principal component analysis and canonical discriminant analysis for investigating hyperspectral plant response. *Italian Journal of Agronomy*, 7(3), 34. <https://doi.org/10.4081/ija.2012.e34>
- Su, W.-H., & Sun, D.-W. (2016). Facilitated wavelength selection and model development for rapid determination of the purity of organic spelt (*Triticum spelta* L.) flour using spectral imaging. *Talanta*, 155, 347–357. <https://doi.org/10.1016/j.talanta.2016.04.041>
- Su, W.-H., & Sun, D.-W. (2017). Evaluation of spectral imaging for inspection of adulterants in terms of common wheat flour, cassava flour and corn flour in organic Avatar wheat (*Triticum spp.*) flour. *Journal of Food Engineering*, 200, 59–69. <https://doi.org/10.1016/j.jfoodeng.2016.12.014>
- Verdú, S., Ivorra, E., Sánchez, A. J., Barat, J. M., & Grau, R. (2016). Spectral study of heat treatment process of wheat flour by VIS/SW-NIR image system. *Journal of Cereal Science*, 71, 99–107. <https://doi.org/10.1016/j.jcs.2016.08.008>
- Verdú, S., Vásquez, F., Grau, R., Ivorra, E., Sánchez, A. J., & Barat, J. M. (2016). Detection of adulterations with different grains in wheat products based on the hyperspectral image technique: The specific cases of flour and bread. *Food Control*, 62, 373–380.

<https://doi.org/10.1016/j.foodcont.2015.11.002>

Verdú, S., Vásquez, F., Ivorra, E., Sánchez, A. J., Barat, J. M., & Grau, R. (2017). Hyperspectral image control of the heat-treatment process of oat flour to model composite bread properties. *Journal of Food Engineering*, *192*, 45–52. <https://doi.org/10.1016/j.jfoodeng.2016.07.017>

Vermeulen, P., Ebene, M. B., Orlando, B., Fernández Pierna, J. A., & Baeten, V. (2017). Online detection and quantification of particles of ergot bodies in cereal flour using near-infrared hyperspectral imaging. *Food Additives and Contaminants - Part A Chemistry, Analysis, Control, Exposure and Risk Assessment*, *34*(8), 1312–1319. <https://doi.org/10.1080/19440049.2017.1336798>

Wadood, S. A., Guo, B., Zhang, X., & Wei, Y. (2019). Geographical origin discrimination of wheat kernel and white flour using near-infrared reflectance spectroscopy fingerprinting coupled with chemometrics. *International Journal of Food Science and Technology*, *54*(6), 2045-2054. <https://doi.org/10.1111/ijfs.14105>

Walsh, K., Blasco, J., Zude-Sasse, M., Sun, X. (2020). Visible-NIR ‘point’ spectroscopy in postharvest fruit and vegetable assessment. *Postharvest Biology and Technology*, *168*, 111246. <https://doi.org/10.1016/j.postharvbio.2020.111246>.

Wang, H., Lv, D., Dong, N., Wang, S., & Liu, J. (2018). Application of near-infrared spectroscopy for screening the potato flour content in Chinese steamed bread. *Food Science and Biotechnology*, *28*, 55–963. <https://doi.org/10.1007/s10068-018-00552-x>

Westad, F., & Martens, H. (2000). Variable selection in near infrared spectroscopy based on significance testing in partial least squares regression. *Journal of Near Infrared Spectroscopy*, *8*(2), 117–124. <https://doi.org/10.1255/jnirs.271>

Zhao, X., Wang, W., Ni, X., Chu, X., Li, Y.-F., & Sun, C. (2018). Evaluation of Near-Infrared Hyperspectral Imaging for Detection of Peanut and Walnut Powders in Whole Wheat Flour. *Applied Sciences*, *8*(7), 1076. <https://doi.org/10.3390/app8071076>

Zhu, Q., Xing, Y., Lu, R., Huang, M., & Ng, P. K. W. (2017). Visible/shortwave near infrared spectroscopy and hyperspectral scattering for determining bulk density and particle size of wheat flour. *Journal of Near Infrared Spectroscopy*, *25*(2), 116–126. <https://doi.org/10.1177/0967033517704081>

Ziegler, J. U., Leitenberger, M., Longin, C. F. H., Würschum, T., Carle, R., & Schweiggert, R. M. (2016). Near-infrared reflectance spectroscopy for the rapid discrimination of kernels and flours of different wheat species. *Journal of Food Composition and Analysis*, *51*, 30–36. <https://doi.org/10.1016/j.jfca.2016.06.005>

CHAPTER 2

Identification of fibre added to semolina by Near Infrared (NIR) spectral techniques

Food Chemistry 289 (2019) 195–203



Contents lists available at ScienceDirect

Food Chemistry

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

Research Article

Identification of fiber added to semolina by near infrared (NIR) spectral techniques



Amanda Teixeira Badaró, Fernanda Lie Morimitsu, Amanda Rios Ferreira, Maria Teresa Pedrosa Silva Clerici, Douglas Fernandes Barbin*

Department of Food Engineering, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, Campinas, SP 13083-862, Brazil

ARTICLE INFO

Keywords:

Near infrared spectroscopy
Hyperspectral imaging

Semolina

Fiber

ABSTRACT

Ingredients added in food products can increase the nutritional value, but also affect their functional properties. After processing, determination of added ingredients is difficult, thus it is important to develop rapid techniques for quantification of food ingredients. In the current work, near infrared spectroscopy (NIRS) and hyperspectral imaging (NIR-HSI) were investigated to quantify the amount of fiber added to semolina and its distribution. NIR spectra were acquired to compare the accuracy in the classification, quantification and distribution of fibers added to semolina. Principal Component Analyses (PCA) and Soft Independent Modeling of Class Analogy (SIMCA) were used for classification. Partial Least Squares Regression (PLSR) models applied to NIR-HSI spectra showed R^2 between 0.85 and 0.98, and RMSEP between 0.5 and 1%, and were used for prediction map of the samples. These results showed that NIR-HSI technique can be used for the identification and quantification of fiber added to semolina.

1. Introduction

Wheat flour is widely used as the main ingredient for many products, such as bread, cookies, cakes and pasta, due to its source of proteins, fats and vitamins (FAO, 2017). Hence, it is an appropriate food vehicle, since it can add nutrients such as vitamins, proteins and fiber (Akhtar, Anjum, Rehman, Sheikh, & Farzana, 2008).

Fiber-fortification of food products has been associated to many health benefits, including the reduction of chronic diseases. However, the addition of alternative ingredients in the formulation of some products can affect their technological and sensory properties. Thus, the nutrient should be properly distributed in the product, and in the right amount it is informed in the package. Brazilian food legislation, for instance, considers a product “source of fiber” if it has at least 2.5 g of fiber per 100 g and “high fiber” if it has a minimum of 5 g of fiber per 100 g of product (ANVISA, 2012). In Europe, the specifications are at least 3 g per 100 g portion of product to be considered “source of fiber” and 6 g per 100 g of product to be considered “high fiber” (European Commission, 2012).

Rapid and non-destructive vibrational techniques have been developed to evaluate food quality, being used in laboratory and industrial scale. Near infrared spectroscopy (NIRS) and hyperspectral imaging (HSI) have shown promising results in several fields, such as

agricultural and pharmaceutical, among others (Porep, Kammerer, & Carle, 2015).

NIRS is a vibrational technique based on the interaction of radiation with the sample, being influenced by overtones, combinations and resonances of fundamental vibrations of C–H, N–H, O–H and S–H bond in organic molecules (Pasquini, 2018). Therefore, this technique has been widely used in the food field to analyze constituents such as fat, sugars, protein, and moisture. However, the potential of NIRS for analysis of technological or sensory attributes of food products is complex in case of heterogeneous matrices, once spectroscopic equipment offer spectral information of a small portion of the sample and do not provide information on the spatial distribution.

In order to overcome this challenge and obtain spatial information of the sample, computer vision techniques are required. Also, systems that work on the visible range of the electromagnetic spectrum are not usually enough for detecting chemical and biological parameters. Therefore, a technique that combines both spectroscopy and computer vision, such as NIR-HIS, becomes a wise alternative (Feng & Sun, 2012). NIR-HSI integrates spectra and images of a sample, so spectral and spatial information are obtained. The final data acquired by NIR-HSI systems, called “hypercube” or hyperspectral image, contains three dimensions (x, y, λ), being two for spatial coordinates (x and y) and one for spectral information (λ). Due to the large amount of information,

* Corresponding author.

E-mail addresses: fernanda@morimitsu.com.br (F.L. Morimitsu), mcclerici@unicamp.br (M.T.P.S. Clerici), dobarbin@unicamp.br (D. Fernandes Barbin).

<https://doi.org/10.1016/j.foodchem.2019.03.057>

Received 13 December 2018; Received in revised form 7 March 2019; Accepted 11 March 2019

Available online 12 March 2019

0308-8146/© 2019 Elsevier Ltd. All rights reserved.

more difficulties are found in data processing (Feng & Sun, 2012).

To facilitate data processing, statistical and mathematical methods are combined to extract only relevant information from hyperspectral images. Chemometrics uses mathematics and statistics to interpret the data by using multivariate analysis. Cluster analysis, such as Principal Component Analysis (PCA), and classification analysis, such as Soft Independent Modeling of Class Analogy (SIMCA), are used as multivariate classifications for qualitative determinations of parameters of interest. Partial Least Square Regression (PLSR), on the other hand, is applied in the prediction of desired parameters in samples for quantitative determinations (Porep et al., 2015).

NIR-HSI technique has shown good performance on the quantitative detection and visualization of contamination of peanut and walnut powders in whole wheat flour (Zhao et al., 2018). In previous works, Mishra et al. (2015) demonstrated the potential of NIR-HSI in detecting and quantifying peanut traces in wheat flour products; Vermeulen, Ebene, Orlando, Fernández Pierna, and Baeten (2017) showed the applicability and efficiency of the technique in discriminating ergot body particles from cereal flour. NIR-HIS demonstrated its aptitude for inspection of adulterants in terms of common wheat flour, cassava flour and corn flour in organic Avatar wheat (*Triticum* spp.) flour (Su & Sun, 2017) and for determination of purity of organic spelt flour from three categories of adulterants, including rye flour, organic wheat flour and spelt flour (Su & Sun, 2016). Zhu, Xing, Lu, Huang, and Ng (2017) compared visible/short-wave near infrared (Vis/SW-NIR) spectroscopy and hyperspectral scattering for bulk density determination and particle size classification of wheat flour. Recently, Rachmawati, Rohaeti, and Rafi (2017) combined near infrared spectroscopy and chemometrics for authentication of few samples of taro flour from wheat and sago flour. Although all these works provided encouraging results on the application of NIR technique for flour attributes assessment, the use of such technique for analysis of semolina and fiber has not been previously reported.

Therefore, the purpose of this study is to compare NIRS and NIR-HSI techniques for classification of different types of fiber added to semolina. In addition, it aims to quantify the amount of fiber from different sources and particle sizes added to semolina and show the spatial distribution using chemical maps.

2. Material and methods

2.1. Sample preparation

Three different brands of semolina were purchased in a local market and five types of fibers, which differ on source, particle size and solubility (Table 1), were obtained from a food ingredients company. The study was divided into three sets of experiments (Fig. 1). In the first set, 140 samples (20 samples of semolina from 3 different brands + 20 samples \times 3 types of fiber + 10 samples \times 2 types of fiber) were weighed in a total of 20 g each.

In the second set of experiments, only one brand of semolina was used, and three groups of samples were prepared: pure semolina, and samples prepared with semolina with two percentages of fiber, 3.5%

and 7% by weight. Samples were vigorously mixed to get a total sample weight of 20 g each time at every mixture of semolina and fiber. Then, 220 samples (20 samples of semolina with 3.5% of fiber and 20 samples of semolina with 7% of fiber \times 5 types of fiber and 20 samples of pure semolina) were obtained. During spectra acquisition, samples were placed on vials with no absorbance on the NIR range and for image acquisition samples were placed in Petri dishes with an exposed sample surface area of 19.63 cm².

In the third set, brand of semolina was used, and six groups of samples were prepared: pure semolina, and samples prepared with semolina with five percentages of fiber, 2%, 3.5%, 5%, 7% and 8% by weight. Samples were vigorously mixed to get a total sample weight of 10 g at every type of semolina and fiber. Then, 351 samples (5 samples of semolina with 2% of fiber, 25 samples with 3.5% of fiber, 5 samples with 5% of fiber, 25 samples with 7% of fiber, 5 samples with 8% of fiber \times 5 types of fiber and 26 samples of pure semolina) were prepared. The samples were placed in Petri dishes for image acquisition.

2.2. NIR spectra and hyperspectral image acquisition

NIR spectra were recorded from 900 to 1700 nm at 3.51 nm intervals using a portable spectrometer (DLP NIRscan Nano Evaluation Module (EVM), Texas Instruments, Dallas, TX), in absorbance mode. The spectrometer is equipped with a reflective sampling module with 2 tungsten lamps, and the process was controlled by DLP NIRscan Nano Reference Software, with a diffuse reflectance standard as white reference.

A SisuCHEM SWIR hyperspectral camera (Specim Spectral Imaging Ltd, FIN-90571 Oulu, Finland) was used to acquire the NIR hyperspectral images. The spectral range was 928–2524 nm at 6.23 nm intervals, 10 nm spectral resolution and 320 pixels per line scan. The camera is equipped with a tungsten-halogen source and two-dimensional array detectors with 256 wavelength channels. The samples were placed on the translational base and moved to reach the field of view of the equipment and the spectra were acquired with an exposure time of 2.1 ms using a 50 mm lens, and the process was controlled by Evince software (Umbio AB, Sweden). The program automatically subtracted the white (~99% reflectance) and dark (0% reflectance) references from subsequently acquired images.

In the first and second set of experiments, a spectrum and an image were acquired from each sample, using the portable NIR equipment and the HSI system, respectively. In the third set, an image was acquired from each of the 351 samples.

2.3. Preprocessing of hyperspectral data

A binary mask was constructed by subtracting a low-reflectance band image (162 nm) from a high-reflectance band image (58 nm) in the same hyperspectral image, for segmentation of the regions of interest (ROI). After removing the background from the image, morphological operations of erosion were performed to remove edges of Petri dishes, creating a final mask containing only the region of interest (ROI). The ROI was used to acquire an average spectrum for each sample. The procedure was repeated for all hyperspectral images acquired. Image segmentation and spectrum extraction were performed using Matlab R2016a software (The Mathworks Inc., Natick, MA, USA).

2.4. Spectral pre-processing

The entire set of spectral data acquired by NIRS and NIR-HSI were mean centered prior to any analysis. Pre-processing techniques, such as first derivative (1st Der) (Savitsky-Golay smoothing, 15 points window, 2 order polynomial), Standard Normal Variate (SNV) and a combination of both, were used. The multivariate data analysis was performed using The Unscrambler X 10.4 software.

Table 1
Samples properties.

Sample	Source	Particle size (μm)
Semolina 1	–	320
Semolina 2	–	80
Semolina 3	–	160
Fiber 1	20% psyllium 80% bamboo	160
Fiber 2	Bamboo	60
Fiber 3	Wheat	60
Fiber 4	50% psyllium 50% cellulose	160
Fiber 5	Cellulose	115

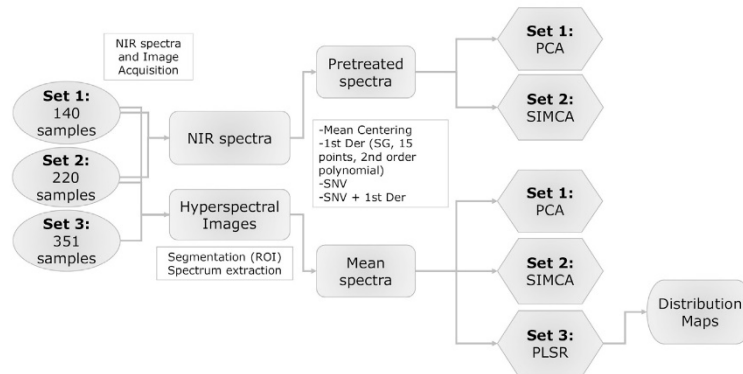


Fig. 1. Sample preparation steps.

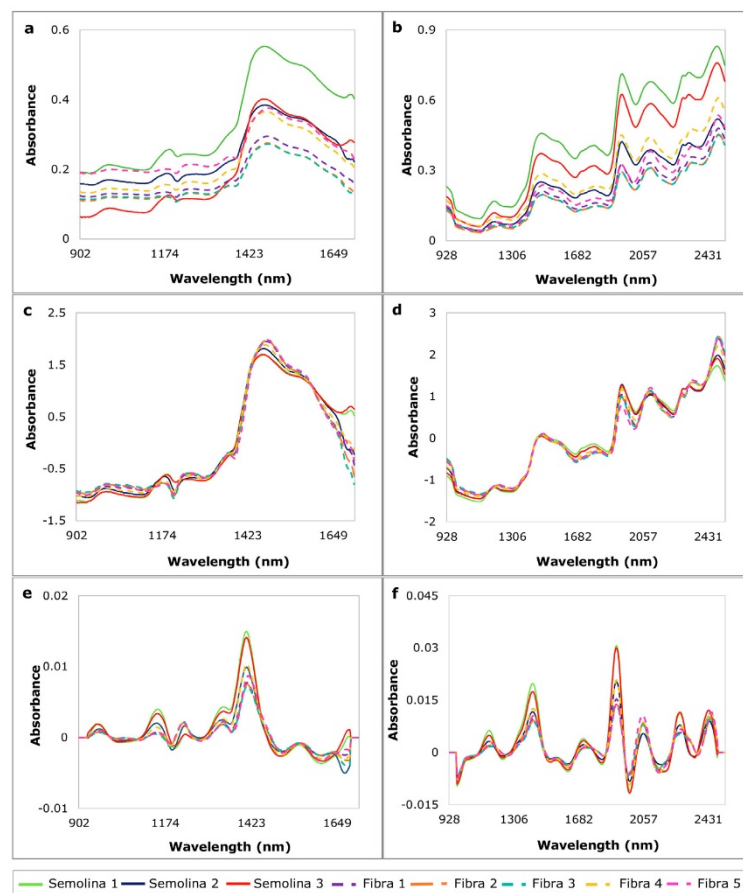


Fig. 2. Average raw (a and b) and pre-processed spectra after SNV (c and d) and 1st Der (e and f) of each pure semolina and fiber.

2.5. Data analysis

Principal Component Analysis (PCA) were applied on both NIRS and NIR-HSI raw and pre-processed data, as exploratory data analysis, to

obtain an overview of spectral variation among samples. PCA is an unsupervised method that linearly combines the variables into principal components (PCs), which explain the greatest variance in the dataset and keep only the relevant information of the spectra (Rachmawati

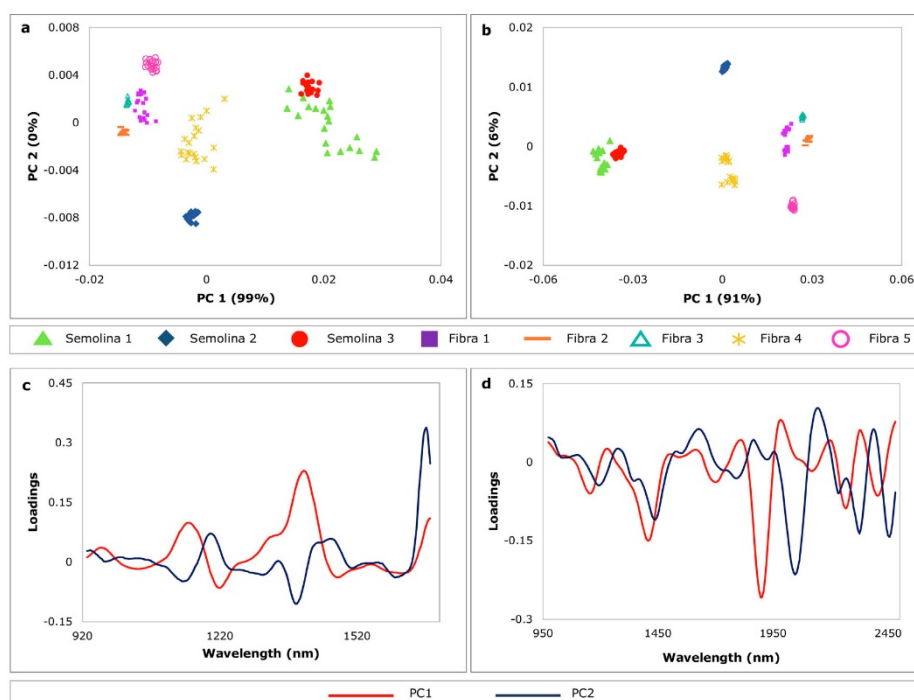


Fig. 3. Score plots (PC1 vs PC2) of PCA performed on spectra after Savitzky–Golay 1st Der pre-processing (a and b) and their correspondent Loadings plots (c and d) for NIR and HSI data, respectively.

et al., 2017).

2.6. Discriminant analysis

Based on raw and pre-processed data, a Soft Independent Modeling of Class Analogy (SIMCA) was performed on the second dataset, to discriminate pure and fiber-added semolina. SIMCA is a supervised chemometric method for pattern recognition based on principal component analysis. A PCA is developed for each class and analyzed individually, so each method has its relevant number of components. F-statistic and Hotelling T^2 were calculated and the models were validated in the new dataset (not included in the calibration set). SIMCA determines the class distance using Euclidean distance from the principal component models and its results are plotted in the Coomans plot, which shows the distances of each sample from two different classes in the same time. In this plot, there are four zones with a confidence level of, usually, 95%. One zone represents samples of class 1, another represents samples of class 2, one for both classes and one for samples which do not belong to any classes. This method can assign the samples into one class, more than one class or to any of them (Rodríguez, Rolandelli, & Buera, 2019).

In order to test the model in a new set of data, seventy percent of the samples were assigned to the calibration set and the remaining 30% of samples constituted the validation set (independent test set). The calibration models were evaluated based on a full cross validation (leave-one out) and in the test set validation. The evaluation of these models performance was based on the parameters of sensitivity (a measure of how often the model correctly identified a positive sample as positive) and specificity (a measure of how accurate the model was against false-positives) and the best models are plotted in the Coomans Plots.

2.7. Development of the prediction model and distribution map

In the third dataset, Partial Least Square Regressions (PLSR) were performed to test the ability of NIR-HSI to quantify the different percentages of fiber added to semolina. Similarly to SIMCA, seventy percent of the samples were assigned to the calibration set and the remaining 30% of samples constituted the validation set (independent test set). The calibration models were applied on raw spectra and evaluated based on a full cross validation (leave-one out) and in the test set validation. Moreover, latent variables (LV) and loadings resulting from PLSR were used to select calibration samples (using Kennard-Stone algorithm) and optimal wavelengths, respectively. Once calibration set and optimal wavelengths were identified, a new PLSR model was recalculated using the spectral data at these wavelengths. The performance of these models was evaluated by coefficient of determination and root mean square error of cross-validation (R_{CV}^2 , RMSECV) and prediction (R_p^2 , RMSEP). The new PLSR model was, then, used to predict the fiber content in each pixel of the hyperspectral image, creating a chemical map of the samples.

3. Results and discussion

3.1. Raw spectra of samples

The average NIRS and NIR-HSI raw and pre-processed spectra of pure semolina and fiber are shown in Fig. 2. All spectra in each technique had a very similar shape but differed in the intensity of absorbance across the spectral region. Differences on baseline between samples can be observed on raw spectra (Fig. 2a and b) and corresponds to modification in the scattering properties of the samples, probably associated to fiber granulometry. Semolina 1, which has the biggest

Table 2
SIMCA models performance for NIRS and NIR-HSI applied for discrimination of pure semolina and fiber-added semolina.

		NIR spectroscopy												
Fiber Type	Pre-processing	Number of Samples	Sensitivity						Specificity					
			Calibration			Validation			Calibration			Validation		
			C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
Fiber 1	None	60	0.13	0.08	0.07	0.00	0.25	0.00	0.14	0.10	0.63	0.62	0.00	0.54
	SNV	60	0.00	0.00	0.00	0.00	0.13	0.00	0.48	0.00	0.56	0.62	0.00	0.38
	1st derivative	60	0.00	0.08	0.20	0.00	0.25	0.00	0.63	0.01	0.70	0.77	0.00	0.69
	SNV+1st derivative	60	0.07	0.00	0.47	0.00	0.00	0.40	0.67	0.27	0.67	0.77	0.20	0.54
Fiber 2	None	60	0.38	0.08	0.15	0.00	0.00	0.14	0.38	0.31	0.48	0.50	0.09	0.18
	SNV	60	0.19	0.00	0.08	0.50	0.00	0.14	0.81	0.14	0.34	1.00	0.27	0.36
	1st derivative	60	0.56	0.38	0.15	0.75	0.14	0.00	0.88	0.38	0.79	1.00	0.36	0.45
	SNV+1st derivative	60	0.50	0.15	0.23	0.50	0.14	0.14	0.92	0.38	0.62	1.00	0.27	0.45
Fiber 3	None	60	0.18	0.08	0.15	0.00	0.00	0.14	0.60	0.20	0.17	0.40	0.20	0.00
	SNV	60	0.41	0.00	0.00	1.00	0.00	0.00	0.00	0.50	0.24	0.00	0.40	0.27
	1st derivative	60	0.00	0.00	0.08	0.00	0.00	0.00	0.80	0.37	0.00	0.53	0.30	0.00
	SNV+1 st derivative	60	0.24	0.00	0.15	0.33	0.00	0.14	0.88	0.60	0.14	0.73	0.40	0.09
Fiber 4	None	60	0.38	0.00	0.08	0.50	0.00	0.00	0.46	0.31	0.23	0.50	0.29	0.30
	SNV	60	1.00	0.00	0.08	1.00	0.00	0.00	0.93	0.58	0.47	0.92	0.43	0.60
	1st derivative	60	1.00	0.00	0.08	1.00	0.50	0.00	1.00	0.58	0.47	1.00	0.43	0.80
	SNV+1st derivative	60	1.00	0.00	0.08	1.00	0.25	0.00	1.00	0.58	0.47	1.00	0.43	0.70
Fiber 5	None	60	0.56	0.00	0.00	0.00	0.00	0.00	0.65	0.59	0.29	0.57	0.38	0.00
	SNV	60	0.63	0.47	0.09	0.50	0.60	0.22	0.85	0.78	0.55	0.57	0.77	0.56
	1st derivative	60	0.81	0.67	0.36	0.75	0.80	0.22	0.81	0.89	0.74	0.64	0.85	0.78
	SNV+1st derivative	60	1.00	0.93	0.45	1.00	0.80	0.44	0.88	0.89	0.97	0.79	0.85	0.89

		NIR-HSI												
Fiber Type	Pre-processing	Number of Samples	Sensitivity						Specificity					
			Calibration			Validation			Calibration			Validation		
			C1	C2	C3	C1	C2	C3	C1	C2	C3	C1	C2	C3
Fiber 1	None	60	1.00	0.87	0.92	0.83	1.00	0.86	1.00	0.96	0.93	1.00	0.92	1.00
	SNV	60	0.93	0.07	0.46	0.83	0.2	0.29	1.00	0.7	0.52	1.00	0.62	0.64
	1st derivative	60	1.00	1.00	1.00	1.00	0.6	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SNV+1st derivative	60	1.00	0.93	0.62	0.83	1.00	0.71	1.00	0.81	0.97	1.00	0.85	1.00
Fiber 2	None	60	1.00	0.93	0.93	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00
	SNV	60	0.92	0.73	0.64	0.86	1.00	0.83	1.00	0.78	0.86	1.00	0.92	1.00
	1st derivative	60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SNV+1st derivative	60	1.00	1.00	1.00	0.86	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Fiber 3	None	60	1.00	1.00	1.00	0.8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SNV	60	1.00	0.3	0.79	0.8	0.71	0.83	1.00	0.9	0.68	1.00	0.91	0.83
	1st derivative	60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SNV+1st derivative	60	1.00	1.00	1.00	0.8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Fiber 4	None	60	0.6	0.23	0.43	0.2	0.71	0.17	0.74	0.52	0.86	0.77	0.18	0.92
	SNV	60	0.6	0.23	0.57	0.8	0.57	0.33	0.78	0.59	0.86	0.77	0.55	0.92
	1st derivative	60	0.80	1.00	1.00	1.00	1.00	0.83	1.00	0.90	1.00	1.00	0.91	1.00
	SNV+1st derivative	60	0.8	1.00	1.00	1.00	1.00	0.5	1.00	0.89	1.00	1.00	0.72	1.00
Fiber 5	None	60	1.00	0.84	0.92	1.00	1.00	1.00	0.96	0.97	0.97	1.00	1.00	1.00
	SNV	60	1.00	0.62	0.67	1.00	0.43	1.00	1.00	0.86	0.83	1.00	1.00	0.6
	1st derivative	60	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SNV+1st derivative	60	1.00	1.00	0.92	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00	1.00

C1 = Semolina, C2 = Semolina + 3.5% of fiber, C3 = Semolina + 7% of fiber.

C1 = Semolina, C2 = Semolina + 3.5% of fiber, C3 = Semolina + 7% of fiber.

particle size, had high absorbance values, followed by Semolina 3. Fibers 2 and 3, which have the same particle size had very similar spectra on both techniques. Fibers 1 and 4, on the other hand, although having the same particle size, presented different behaviour. This could be due to the fibers composition, since they are not from the same source, but rather a mixture of soluble and insoluble fibers. Even though some samples presented higher absorbance, caused by larger particle size (Mishra et al., 2015), this was not observed for all samples. This may be explained by the fact that all samples had an average particle size, so the number of particles with smaller size in each sample is unknown.

Analyzing the raw NIRS spectra, four absorption peaks were observed at 1190 nm, 1215 nm, 1388 nm and 1450 nm, which correspond to stretching vibration of C–H in the second overtone and O–H in the first overtone. In NIR-HSI, the same peaks were observed, in addition to peaks at 1206 nm, 1664 nm and 2300 nm, corresponding to C–H stretching vibration in the first and second overtone and at 1450 nm

and 1940 nm corresponding to O–H stretching vibration or O–H–O combination of deformation. A peak at 1820 nm corresponds to O–H or C–O stretching vibration, characteristic of cellulose. Peaks at 2000 nm denote O–H and C–O stretching vibration or a combination of deformation, and at 2242 nm corresponds to N–H stretching vibration (Osborn, Fearn, & Hindle, 1993).

In order to reduce scattering effects, SNV and 1st Der were applied to NIRS and NIR-HSI raw spectra (Fig. 2). Regarding SNV pre-processed spectra, it could be observed peaks around 970 and 1580 nm, which correspond to O–H stretching vibration in the second and first overtones, associated to water and starch, respectively. After 1st Der pre-processing, NIRS spectra showed peaks around 970 and 1528 nm, which correspond to O–H stretching vibration in the second and first overtone, also characteristic of water and starch, respectively. A peak at 1150 nm corresponding to C–H stretching vibration in the second overtone was also observed. After 1st Der, HSI spectra showed peaks at

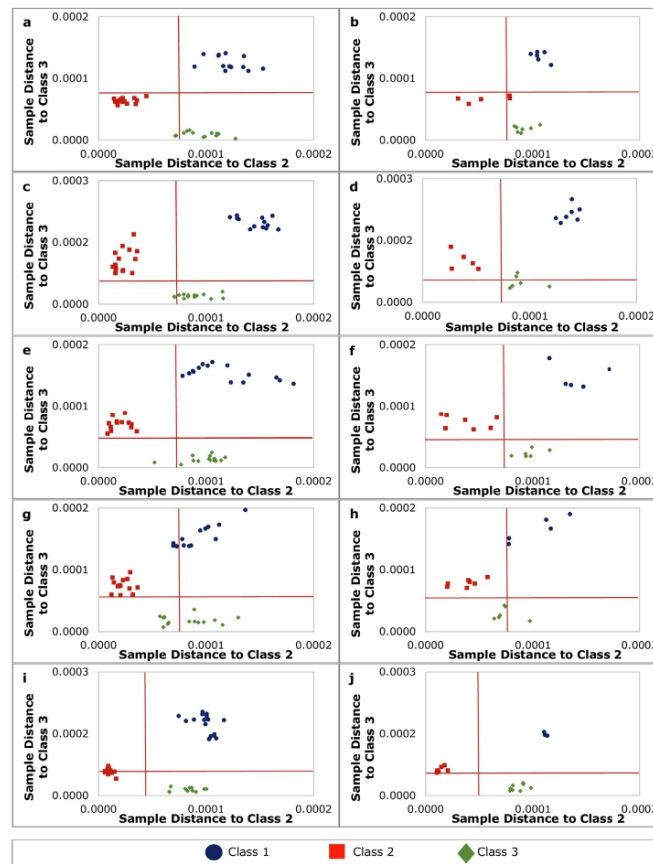


Fig. 4. Calculated class responses for SIMCA models. Calibration models: (a), (c), (e), (g) and (i) and validation models: (b), (d), (f), (h) and (j).

Table 3

Model performance of PLSR based on characteristic wavelengths for quantitative analysis of fiber added to semolina.

Models	Optimal wavelengths (nm)	LV	R _v ²	R _p ²	RMSECV	RMSEP
Fiber 1	1206, 1450, 1664, 1820, 1940, 2000, 2242, 2300	2	0.92	0.95	0.83	0.62
Fiber 2	1206, 1450, 1664, 1820, 1940, 2000, 2242, 2300	2	0.93	0.93	0.79	0.68
Fiber 3	1206, 1450, 1664, 1820, 1940, 2000, 2242, 2300	5	0.96	0.98	0.59	0.44
Fiber 4	1206, 1450, 1664, 1820, 1940, 2242, 2300	6	0.88	0.85	1.02	1.09
Fiber 5	1206, 1450, 1664, 1820, 1940, 2000, 2242, 2300	5	0.96	0.95	0.58	0.52

1620 and 1700 nm, associated to C–H stretching vibration in the first overtone, and a peak at 2100 nm, related to combination of O–H deformation and C–O stretching vibration, associated to starch.

3.2. Principal components analysis (PCA)

Principal component analysis was performed on both NIRS and NIR-HSI spectra of the first dataset, in order to visualize any variation among samples. In order to obtain a better sample separation, 1st Der, SNV, and a combination of both were applied. Both data showed better

separation when 1st Der (Savitsky–Golay smoothing, 15 points window, 2 order polynomial) was applied. PCA scores of pre-processed spectra with 1st Der and their correspondent loadings are illustrated in Fig. 3.

The first two principal components on both techniques explained more than 90% of the total variance of the samples. This variability observed on the scores could be due to chemical bonds or to the larger particle size of the samples, which could affect the spectral signal (Manley, 2014). Thus, by observing the loadings it is possible to identify wavelengths that contributed in the sample separation, with higher absolute values. The observed wavelengths are similar to those

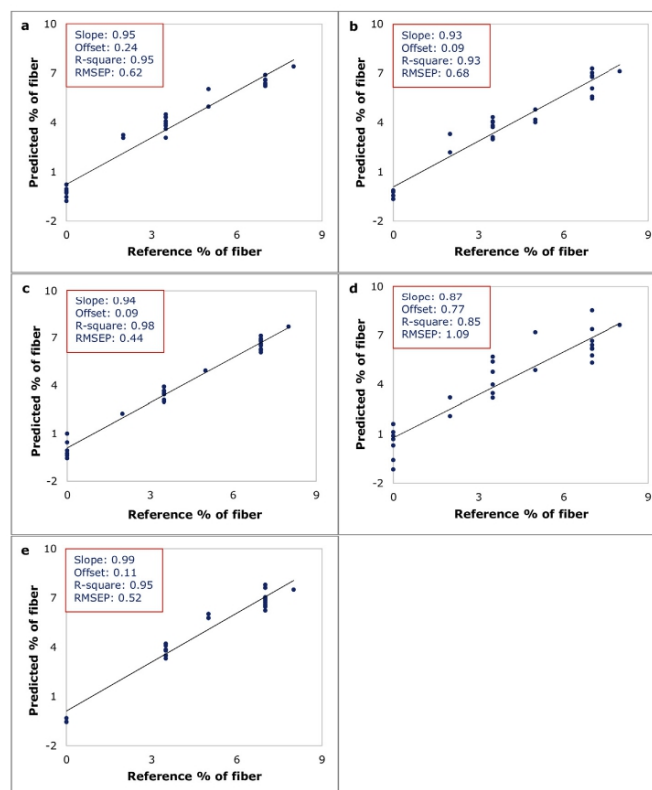


Fig. 5. Predicted vs. measured plot for the PLSR models obtained on the raw spectra. (a) Fiber 1, (b) Fiber 2, (c) Fiber 3, (d) Fiber 4 and (e) Fiber 5 added to semolina.

observed in pre-processed spectra, most predominantly related to C–H and O–H bonds, associated to water, starch and cellulose.

Even though it is noticeable that all semolinas and fibers formed different clusters, NIR-HSI showed better performance when grouping samples of the same set. Still, PCA was able to distinguish among semolina and fiber samples for both techniques.

3.3. Sample classification

SIMCA models applied on NIRS data (Table 2) presented sensitivity and specificity close or equal to zero, showing very low efficiency on discriminating and classifying the samples. The heterogeneity of the samples may be the cause of the poor performance of NIRS. As NIR spectra were acquired only in one single point of the sample, the spectrum could not be representative of the whole sample. Even though NIRS has been widely used to analyze different matrices, it is still a challenge when used in food products, once they are very heterogeneous (Porep et al., 2015). On the other hand, SIMCA models applied on raw and pre-processed NIR-HSI data (Table 2) generally obtained high values of sensitivity and specificity for calibration and validation. Spectral information pre-processed by Savitzky–Golay 1st Der provided increased accuracy for classification and differentiation between samples. Hyperspectral images have been a powerful tool to characterize

complex matrices as food, according to Lohumi, Lee, Lee, and Cho (2015). In this work, the mean spectra of each sample were enough to show the efficiency of NIR-HSI technique to distinguish and classify among pure semolina and samples containing 3.5 and 7% of fiber.

In addition, Fig. 4 illustrates the class responses for the calibration (Fig. 4a, c, e, g and i) and prediction steps (Fig. 4b, d, f, h and j) for the models with the best performance (highlighted in dark grey on Table 2). These figures show overall classification (significance level of 5%) for each class. Within the confidence limit of 95%, calibration models show misclassified samples for Fiber 4, and for validation step, the models developed for fibers 1 and 4 presented some misclassified samples. However, this did not undermine the model efficiency. Fiber 2, even though it had two samples in the validation set which exceeded the confidence limit, had no misclassified samples. This may have happened due to the distance between classes, which was closer to the real class (Class 3) than to Class 2. The visual representation in Fig. 4, for calibration and prediction, confirms the results presented on Table 2 for the three classes.

3.4. Prediction models

As in the discriminant analysis, the average spectrum of each sample showed promising results to quantify the amount of fiber added

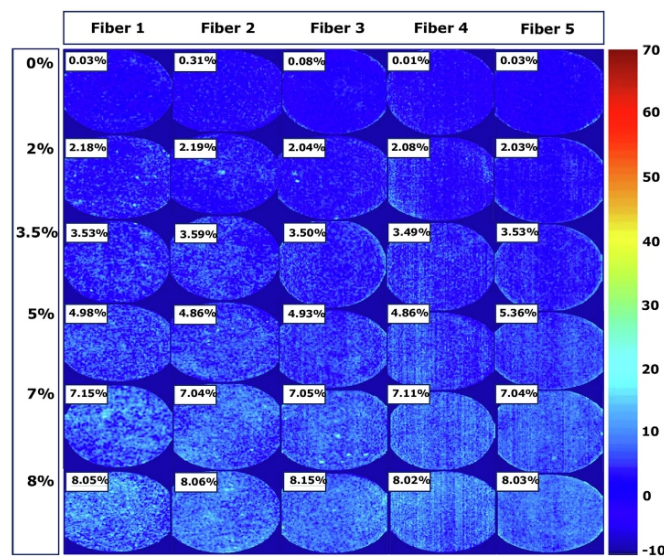


Fig. 6. Prediction maps for visual identification of fiber added to semolina using PLSR models. The dark blue image represents pure semolina, otherwise fiber. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in semolina. After selection of the optimal wavelengths, new models were developed with raw spectra and applied to an independent validation set, being able to predict the amount of fiber added to semolina with R^2 of 0.95, 0.93, 0.98, 0.85 and 0.95 for fibers 1, 2, 3, 4 and 5, respectively. These results agree with those found by Su and Sun (2016) for rapid determination of the purity of organic spelt (*Triticum spelta* L.) flour using spectral imaging. The authors were able to quantify the adulteration level between 3 and 75%, with R^2 over 0.93. Moreover, values of root mean squares errors of cross-validation (RMSECV) and prediction (RMSEP) are comparable, within a range between 0.5 and 1%. Table 3 reports the results of the regression models calculated on the raw spectra by PLS models.

Models for fibers 1, 2, 3 and 5, developed with eight wavelengths, and fiber 4, with seven wavelengths, showed that a simpler model is able to predict the amount of fiber added to semolina. Fiber 3 had the best accuracy as it had the highest R^2 and lowest RMSEP, followed by fiber 5, 1, 2 and 4, respectively. One of the main reasons for this difference is probably the variety among fiber particle size. The predicted values from each PLS models are displayed in Fig. 5.

According to these results, the developed calibration models are representative and can be used in the development of distribution maps, which allow to visualize the content of fiber in the spatial dimension from hyperspectral images.

3.5. Distribution maps

NIR-HSI technique has the advantage, not only to predict the percentage of fiber added to semolina, but to identify the spatial distributions of the fibers in the sample. The dot product between the optimal regression coefficients and their correspondent wavelengths is applied in each pixel of the image for spatial visualization.

Fig. 6 presents the prediction maps of some samples (from 0 to 8% of fiber). These maps were generated using a colour scale from low (dark blue) to high (red) concentration, where dark blue represents

semolina and light blue, fibers. Su and Sun (2017) showed the efficiency of spectral imaging associated to PLSR models for inspection of adulterants in organic wheat flour (OWF) adulterated in terms of common wheat flour (WF), cassava flour (CaF) and corn flour (CoF). The authors successfully obtained reasonably accurate prediction maps with concentration in a range of 3–75%.

The mixture of fiber and semolina is not homogeneous, for this reason, the colour varied not only among samples, but also within different locations of the same sample. Poor homogenization of these samples, for example, can generate major problems in other products which have these ingredients in their formulation. In an industrial scale it is not possible to visually identify these problems, so these discriminations can be carried out by hyperspectral imaging.

4. Conclusion

NIR-HSI can be used for the identification and quantification of different types of fiber added to semolina. Moreover, the technique showed a great potential on the visualization of different percentage of fiber in the spatial dimension based on spectral characteristics. This work shows that, by using this methodology, several fibers added to semolina can be identified and quantified.

Near infrared spectroscopy, on the other hand, even though is widely used in the food field, did not provide good results in identifying and discriminating the five different fibers added to semolina. It was easier to identify the different clusters by observing the PCA scores, when pure samples were scanned with the portable near infrared spectrometer. However, when the fiber samples were added to semolina and generated a heterogeneous sample, the spectrum of a single point was not representative of the whole sample.

Acknowledgments

This study was financed in part by the Coordenação de

A.T. Badaró, et al.

Food Chemistry 289 (2019) 195–203

Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001; and São Paulo Research Foundation (FAPESP) (project number 2017/17628-3, 2015/24351-2). The authors would like to thank Nutrassim Food Ingredients company for the donation of the fiber samples.

Conflict of interest disclosure

The authors declare no competing financial interest.

Appendix A. Supplementary data

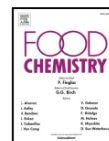
Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2019.03.057>.

References

- Akhtar, S., Anjum, F. M., Rehman, S. U., Sheikh, M. A., & Farzana, K. (2008). Effect of fortification on physico-chemical and microbiological stability of whole wheat flour. *Food Chemistry*, 110(1), 113–119. <https://doi.org/10.1016/j.foodchem.2008.01.065>.
- ANVISA (2012). Resolução RDC No 54, DE 12 DE NOVEMBRO DE 2012. Retrieved from http://portal.anvisa.gov.br/documents/33880/2568070/rdc0054_12_11_2012.pdf/c5ac23fd-974e-4f2c-9fbc-48f7e0a31864.
- European Commission. (2012). Nutrition Claims. Retrieved from https://ec.europa.eu/food/safety/labelling/nutrition/claims/nutrition_claims_en.
- FAO. (2017). Food, Agriculture Organization of the United Nations. Retrieved from <http://www.fao.org/worldfoodsituation/csd/en/>.
- Feng, Y.-Z., & Sun, D.-W. (2012). Application of hyperspectral imaging in food safety inspection and control: A review. *Critical Reviews in Food Science and Nutrition*, 52(11), 1039–1058. <https://doi.org/10.1080/10408398.2011.651542>.
- Lohumi, S., Lee, S., Lee, H., & Cho, B. K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science and Technology*, 46(1), 85–98. <https://doi.org/10.1016/j.tifs.2015.08.003>.
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: Non-destructive analysis of biological materials. *Chemical Society Reviews*, 43, 8200–8214. <https://doi.org/10.1039/c4cs00062e>.
- Mishra, P., Herrero-Langreo, A., Barreiro, P., Roger, J. M., Diezma, B., Gorretta, N., & Lleó, L. (2015). Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis. *Journal of Near Infrared Spectroscopy*, 23(1), 15–22. <https://doi.org/10.1255/jnirs.1141>.
- Osborn, B. G., Fearn, T., & Hindle, P. H. (1993). Theory of near infrared spectroscopy. *Practical NIR spectroscopy with applications in food and beverage analysis* (pp. 13–35). (2nd ed.). London: Longman Singapore Publishers (Pte) Ltd.
- Pasquini, C. (2018). Analytica Chimica Acta Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*, 1026, 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>.
- Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology*, 46(2), 211–230. <https://doi.org/10.1016/j.tifs.2015.10.002>.
- Rachmawati, Rohaeti, E., & Rafi, M. (2017). Combination of near infrared spectroscopy and chemometrics for authentication of taro flour from wheat and sago flour. *Journal of Physics: Conference Series*, 835. <https://doi.org/10.1088/1742-6596/835/1/012011>.
- Rodríguez, S. D., Rolandelli, G., & Buera, M. P. (2019). Detection of quinoa flour adulteration by means of FT-MIR spectroscopy combined with chemometric methods. *Food Chemistry*, 274(2018), 392–401. <https://doi.org/10.1016/j.foodchem.2018.08.140>.
- Su, W.-H., & Sun, D.-W. (2016). Facilitated wavelength selection and model development for rapid determination of the purity of organic spelt (Triticum spelta L.) flour using spectral imaging. *Talanta*, 155, 347–357. <https://doi.org/10.1016/j.talanta.2016.04.041>.
- Su, W.-H., & Sun, D.-W. (2017). Evaluation of spectral imaging for inspection of adulterants in terms of common wheat flour, cassava flour and corn flour in organic Avatar wheat (Triticum spp.) flour. *Journal of Food Engineering*, 200, 59–69. <https://doi.org/10.1016/j.jfoodeng.2016.12.014>.
- Vermeulen, P., Ebene, M. B., Orlando, B., Fernández Pierna, J. A., & Baeten, V. (2017). Online detection and quantification of particles of ergot bodies in cereal flour using near-infrared hyperspectral imaging. *Food Additives and Contaminants – Part A: Chemistry, Analysis, Control, Exposure and Risk Assessment*, 34(8), 1312–1319. <https://doi.org/10.1080/19440049.2017.1336798>.
- Zhao, X., Wang, W., Ni, X., Chu, X., Li, Y.-F., & Sun, C. (2018). Evaluation of near-infrared hyperspectral imaging for detection of peanut and walnut powders in whole wheat flour. *Applied Sciences*, 8(7), 1076. <https://doi.org/10.3390/app8071076>.
- Zhu, Q., Xing, Y., Lu, R., Huang, M., & Ng, P. K. W. (2017). Visible/shortwave near infrared spectroscopy and hyperspectral scattering for determining bulk density and particle size of wheat flour. *Journal of Near Infrared Spectroscopy*, 25(2), 116–126. <https://doi.org/10.1177/0967033517704081>.

CHAPTER 3

*Near infrared hyperspectral imaging and spectral
unmixing methods for evaluation of fibre distribution
in enriched pasta*



Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fiber distribution in enriched pasta

Amanda Teixeira Badaró^{a,b,c,*}, José Manuel Amigo^{d,e}, Jose Blasco^c, Nuria Aleixos^f, Amanda Rios Ferreira^g, Maria Teresa Pedrosa Silva Clerici^g, Douglas Fernandes Barbin^a

^a Department of Food Engineering, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, 13083-862 Campinas, São Paulo, Brazil

^b Departamento de Tecnología de Alimentos, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

^c Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), Ctra. CV-315, km. 10.7, 46113 Moncada, Valencia, Spain

^d IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

^e Department of Analytical Chemistry, University of the Basque Country UPV/EHU, P.O. Box 644, 48080 Bilbao, Basque Country, Spain

^f Departamento de Ingeniería Gráfica, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain

^g Department of Food Technology, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, 13083-862 Campinas, São Paulo, Brazil

ARTICLE INFO

Keywords:

Pasta
Hyperspectral imaging
NIR
Spectral unmixing
Multivariate curve resolution

ABSTRACT

Pasta is mostly composed by wheat flour and water. Nevertheless, flour can be partially replaced by fibers to provide extra nutrients in the diet. However, fiber can affect the technological quality of pasta if not properly distributed. Usually, determinations of parameters in pasta are destructive and time-consuming. The use of Near Infrared-Hyperspectral Imaging (NIR-HSI), together with machine learning methods, is valuable to improve the efficiency in the assessment of pasta quality. This work aimed to investigate the ability of NIR-HSI and augmented Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) for the evaluation, resolution and quantification of fiber distribution in enriched pasta. Results showed R^2_V between 0.28 and 0.89, %LOF < 6%, variance explained over 99%, and similarity between pure and recovered spectra over 96% and 98% in models using pure flour and control as initial estimates, respectively, demonstrating the applicability of NIR-HSI and MCR-ALS in the identification of fiber in pasta.

1. Introduction

Pasta is a prevalent staple food around the world. Its main ingredients are wheat flour and water, although different components can partially replace flour as proteins, antioxidants, and fibers in order to provide an extra amount of nutrients in the human diet (Kowalczewski et al., 2015). Recently, fiber has been added to pasta as a nutrient for a healthy diet (Ferreira, 2018; Muneer et al., 2018; Rakhesh, Fellows, & Sissons, 2015). However, the fortified product has to agree with the food regulations. In the case of fiber fortification, the Brazilian food legislation considers a product "source of fiber" if it has at least 2.5 g/100 g and "high fiber" if it has a minimum of 5 g/100 g of product (ANVISA, 2012). In Europe, the food regulation considers a product "source of fiber" with 3 g/100 g portion of product and "high fiber" with a minimum of 6 g/100 g of product (European Commission, 2012).

Additionally, the quality parameters of pasta (e.g., cooking time,

cooking losses, water absorption, texture, among others) depend on the interaction between starch and proteins during processing, and the addition of fiber or other ingredients in the formulation can affect the structure, taste, and physicochemical properties of the cooked pasta (Bustos, Perez, & Leon, 2015). Moreover, the analysis of these properties is destructive and time-consuming. Pasta firmness and adhesiveness, for instance, are measured with a texture analyzer (TPA) after cooking. This technique is performed by a sequence of compressing and releasing operations on the sample to simulate the chewing process. At the end of the analysis, the remaining sample is discarded (Bustos et al., 2015). These methods do not usually provide any spatial information on the distribution of the elements in the surface of the pasta. Therefore, the use of faster and accurate technologies is required to improve the efficiency in the assessment of the abovementioned characteristics (Huang, Liu, & Ngadi, 2014).

Vibrational spectroscopy techniques, as near infrared (NIR)

* Corresponding author.

E-mail addresses: tbadaro.amanda@gmail.com (A.T. Badaró), josemanuel.amigo@ehu.es (J.M. Amigo), jblasco.ivia@gmail.com (J. Blasco), naleixos@dig.upv.es (N. Aleixos), Amanda_5904@hotmail.com (A.R. Ferreira), mclerici@unicamp.br (M.T.P.S. Clerici), dfbarbin@unicamp.br (D.F. Barbin).

<https://doi.org/10.1016/j.foodchem.2020.128517>

Received 7 April 2020; Received in revised form 26 October 2020; Accepted 27 October 2020

Available online 31 October 2020

0308-8146/© 2020 Elsevier Ltd. All rights reserved.

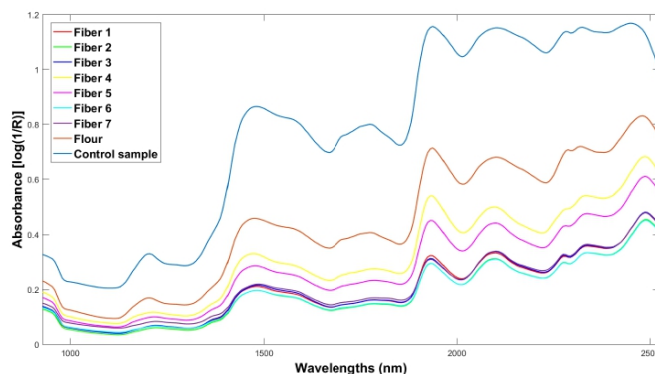


Fig. 1. NIR spectra of the control, pure flour, and pure fiber samples.

spectroscopy, are an alternative for rapid and non-destructive measurements. These measurements are usually taken from one or a few single points of the sample and are considered to represent the entire sample. Nevertheless, the composition of food products usually is heterogeneous, and much important information can be lost if the spatial distribution of the ingredients is not considered, affecting the accuracy and the representativeness of the collected data for decision-making during processing. Near infrared spectroscopy has a significant advantage in requiring several measurements for heterogeneous samples. In order to overcome this situation, near infrared spectroscopy can be integrated with hyperspectral imaging devices (HSI). NIR-HSI is an alternative for measurements of quality parameters in food products (Wu & Sun, 2013) considering the spatial distribution of the compounds in the measured surface of the product. NIR-HSI has already shown excellent results in qualitative and quantitative determinations of foodstuffs; however, only a few works can be found in the application of NIR-HSI in pasta, as the characterization of important parameters of Italian pasta produced by traditional or industrial production parameters (Menesatti, Antonucci, Pallottino, Bucarelli, & Costa, 2014) or in the discrimination between different production technology approaches (Menesatti & Bucarelli, 2007).

The great amount of information contained in the 3D cube makes the interpretation of the data a cumbersome task, requiring the association with advanced machine learning methodologies to extract relevant and useful information (Amigo, 2020; Ma, Sun, Pu, Cheng, & Wei, 2019; Wu & Sun, 2013). Machine learning (a.k.a. Multivariate analysis) compiles several algorithms with different purposes like spectral preprocessing, qualitative and quantitative data analysis, etc (Ferreira, 2015; Vidal & Amigo, 2012). When individual pixels are composed of a mixture of the signal of different compounds, the well-known spectral unmixing methodologies become a perfect choice in order to separate the weighted influence of the signal for each compound in the individual pixels (de Juan, 2020; Nascimento & Martín, 2020). Among them, Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) is one of the major algorithms used in hyperspectral image analysis in food (Forchetti & Poppi, 2017; Zhang, de Juan, & Tauler, 2015), pharma (Franch-Lage, Amigo, Skibsted, Maspocho, & Coello, 2011), chemical compounds in biodegradable active films (Terra et al., 2020), ibuprofen-nicotinamide cocrystal formation (Ishihara, Hattori, & Otsuka, 2019).

MCR-ALS is a bilinear model that aims at decomposing the hyperspectral image in the spectral profiles and relative distributions for each

compound in the sample (de Juan et al., 2004; Ghaffari, Hugelier, Duponchel, Abdollahi, & Ruckebusch, 2019). One of the major properties of MCR-ALS (or in any curve resolution method) is the second-order property. That is the ability to model the presence/absence of a particular component in the presence of unexpected interferences or when the mixture changes in composition. Nevertheless, the apparent simplicity of this model is hampered by the existence of ambiguities (rotational and intensity) that makes the model unstable towards a unique solution. The importance of the ambiguities strongly depends on the spectral differences of the components involved in the mixture. That is, the high degree of overlapping/similarity between the spectra of the components and the spectral variability between different fibers. Food is a complex matrix, and NIR is not a very specific method. This promotes that 1) a “pure component” can be, indeed, a mixture of several components or 2) there is not enough spectral information of minor components in the sample (rank deficiency). To solve or, better said, to minimize these two issues, several strategies are normally followed. The first one is the imposition of what is known as natural constraints, being non-negativity one of the most important ones (Amigo & Ravn, 2009). Another strategy, if possible, is to augment the images (i.e., analyze several images together, multiset arrangement). Augmenting the image can be done with other images of the same nature containing extra information about the pure compounds (Franch-Lage et al., 2011).

Moreover, MCR-ALS has the versatility of adding spectral and spatial information to the original data, and single or multiset arrangement can be used for model development (Franch-Lage et al., 2011; Piqueras, Burger, Tauler, & de Juan, 2012), in order to reduce the rank deficiency. In the food field, these techniques have been applied in the study of the distribution of different components in white and milk chocolate (Zhang et al., 2015), and in the detection and quantification of adulterants in milk powder (Forchetti & Poppi, 2017). However, no studies regarding pasta, especially fiber-enriched pasta, have been reported to our knowledge.

This work hypothesized that NIR-HSI, together with MCR-ALS, has the ability for the evaluation, resolution, and quantification of fiber in enriched pasta. The development of MCR-ALS models verified this hypothesis by studying the signal contributions and spatial distributions of fiber in the sample surface. Different strategies were tested in order to evaluate the feasibility of MCR-ALS in obtaining information in the pasta samples by using images of the raw ingredients (fiber and flour). Finally, the proposed methodology was tested as an alternative for rapid evaluation and authentication of fiber-enriched pasta.

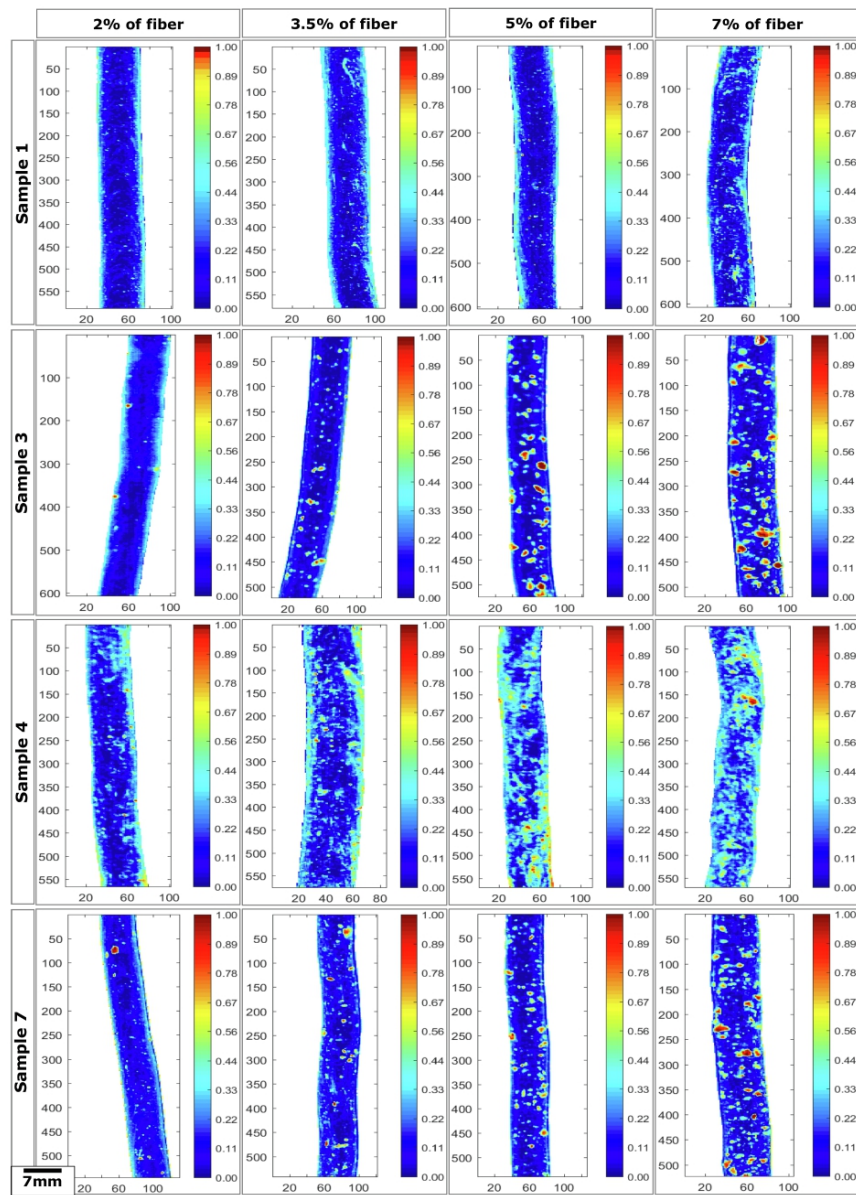


Fig. 2. Concentration maps some of the samples studied in this work. The color represents the percentage of fiber in each pixel of the image.

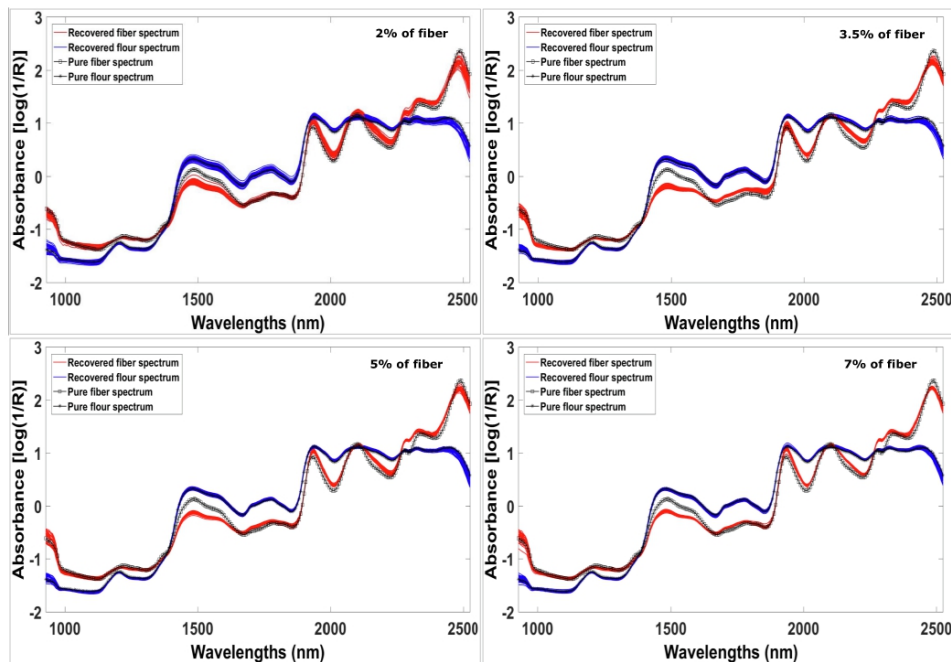


Fig. 3. Pure and recovered spectra from fiber and control sample (Data preprocessed with Standard Normal Variate – SNV).

2. Materials and methods

2.1. Sample preparation

Dry *fettuccine*-type pasta samples were produced in the Department of Food Technology at University of Campinas (Brazil) according to the methodology proposed by Ferreira (2018), replacing the flour by seven different types of fiber (Supplementary material – Fig. S1) in four percentages: 2%, 3.5%, 5% and 7% (w/w). For each percentage of fiber, 30 samples of *fettuccine* were produced, totaling 120 samples per type of fiber (Supplementary material – Table S1). In addition, 30 samples without fibers were produced (control samples).

2.2. Image acquisition

Images of samples were acquired using a SisuCHEMA SWIR hyperspectral camera (Specim Spectral Imaging Ltd, Finland). The system is composed of a high-speed push-broom camera that operates in the spectral range of 928–2524 nm with a spectral resolution of 6.3 nm and 320 spatial channels (156 μm pixel size). The camera is equipped with a tungsten-halogen source and two-dimensional array detectors with 256 wavelength channels. The spectra were acquired with an exposure time of 2.1 ms using a 50 mm lens and a scanning speed of 15.8 mm/s. The slit width was 30 μm . The measurement was controlled by Evince software (UmBio AB, Sweden). The software automatically subtracted the white (~99% reflectance measured with Spectralon) and dark (0% reflectance) references from subsequently acquired images. To optimize the process, one image included three samples, although they were individually analyzed. Images from pure fiber and wheat flour were also taken.

2.3. Data analysis: MCR-ALS

MCR-ALS assumes that the relative concentration and distribution of compounds in a sample of interest are determined according to the pure compounds based on a bilinear model, as showed in eq. (1).

$$D = CS^T + E \quad (1)$$

Where D is the 2D matrix built by unfolding the original hypercube, C is the concentration matrix provided by the algorithm, S^T is the matrix containing the pure spectra information, and E is the matrix expressing the error or variance unexplained by the bilinear model (Supplementary material – Fig. S2).

The original hypercubes are firstly unfolded into a matrix (D) of n dimensions (XY) by m wavelengths (λ). After that, all matrices are joined and preprocessed with standard normal variate (SNV) before MCR-ALS analysis to correct the effects of light scattering.

The first step in MCR-ALS is to determine the number of components, the initial estimates, and the correct constraints to be applied in the model development. In this work, two components were taken into account, since they were the major constituents in pasta samples. This was done by performing Principal Component Analysis (PCA) before MCR. The initial estimates and the correct constraint are the key to achieve reliable results (de Juan, Jaumot, & Tauler, 2014). The data analysis was carried out considering two different types of initial estimates: (a) pure fiber and pure flour spectral information, and (b) pure fiber and control samples spectral information. The constraints used here were non-negativity for the concentration profiles and normalization of the spectral profiles, so all components had the same relative importance in the model (Zhang & Tauler, 2013). Moreover, closure was

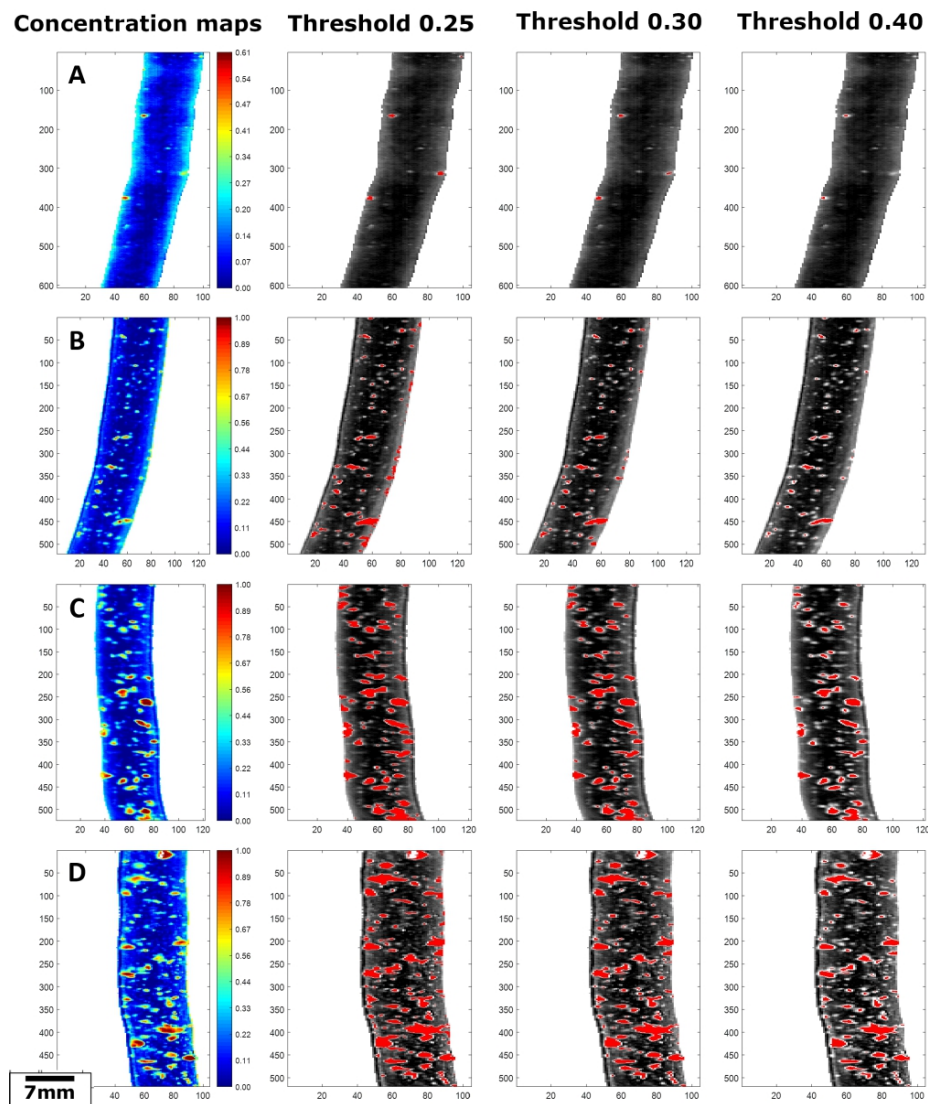


Fig. 4. Identification of fiber granules in Sample 3, comparing the 3 thresholds. A) Sample with 2% of fiber, B) Sample with 3.5% of fiber, C) Sample with 5% of fiber, and D) Sample with 7% of fiber.

applied to the concentration profiles. This assumption could be done in this case since 1) in NIR, the absorbance for each pixel in each wavelength can be assumed to be the sum of the individual absorbance of each compound, and 2) neglecting the minimal influence of artifacts, pasta is composed by two compounds. These constraints help to reduce the effects of rotational ambiguity and, consequently, increasing

spectral and spatial resolution, since they provide some more information about the S^T and C matrices (Firmani, Hugelier, Marini, & Ruckebusch, 2018). Then, the ALS optimization was performed until the convergence criterion was satisfied (0.1%).

A total of 14 augmented MCR-ALS models were developed (2 initial estimates \times 7 types of samples). The percentage of lack of fit (%LOF) and

A.T. Badaró et al.

Food Chemistry 343 (2021) 128517

Table 1
Parameters of OLS regression models.

Expected fiber content – Flour as initial estimate							
Sample	Threshold	biasC	biasV	R ² _C	R ² _V	RMSEC	RMSEV
Pasta 1	0.25	-0.03	-0.03	0.01	0.05	0.48	0.53
	0.30	0.00	-0.01	0.06	0.06	0.09	0.10
	0.40	0.00	0.00	0.04	0.06	0.01	0.01
Pasta 2	0.25	0.05	0.04	0.14	0.15	0.73	0.89
	0.30	0.00	0.00	0.01	0.01	0.05	0.04
	0.40	0.00	0.00	0.03	0.00	0.01	0.02
Pasta 3	0.25	0.93	1.21	0.85	0.84	3.71	3.84
	0.30	0.79	0.90	0.85	0.89	2.99	2.86
	0.40	0.57	0.59	0.84	0.84	2.10	2.00
Pasta 4	0.25	0.31	0.26	0.34	0.36	4.02	4.08
	0.30	0.16	-0.04	0.22	0.28	2.42	2.57
	0.40	0.04	0.05	0.30	0.38	0.33	0.31
Pasta 5	0.25	1.76	2.03	0.67	0.71	9.70	9.28
	0.30	1.65	1.75	0.62	0.66	8.20	8.19
	0.40	0.80	0.95	0.51	0.51	4.30	4.36
Pasta 6	0.25	0.39	0.58	0.41	0.36	2.33	2.41
	0.30	0.11	0.10	0.21	0.25	0.99	0.91
	0.40	0.00	0.00	0.35	0.41	0.02	0.03
Pasta 7	0.25	0.10	0.15	0.66	0.63	2.74	2.66
	0.30	0.22	-0.04	0.71	0.72	2.02	2.16
	0.40	0.27	0.31	0.78	0.80	1.26	1.19
Expected fiber content – Control sample as initial estimate							
Sample	Threshold	biasC	biasV	R ² _C	R ² _V	RMSEC	RMSEV
Pasta 1	0.25	-0.25	-0.45	0.01	0.02	3.11	3.40
	0.30	0.00	-0.03	0.07	0.06	0.25	0.31
	0.40	0.00	0.00	0.28	0.28	0.02	0.02
Pasta 2	0.25	2.56	2.72	0.33	0.37	16.55	14.17
	0.30	1.05	0.59	0.31	0.38	6.58	7.10
	0.40	0.01	0.04	0.15	0.12	0.13	0.09
Pasta 3	0.25	1.45	1.44	0.85	0.90	5.71	5.59
	0.30	1.22	1.66	0.91	0.89	4.01	4.08
	0.40	0.70	0.45	0.89	0.90	2.29	2.38
Pasta 4	0.25	0.71	1.62	0.35	0.35	10.26	9.68
	0.30	1.03	-0.21	0.40	0.53	6.88	7.86
	0.40	0.31	0.43	0.35	0.34	2.01	2.01
Pasta 5	0.25	4.77	4.15	0.71	0.73	22.78	22.95
	0.30	5.73	4.36	0.73	0.78	23.45	22.44
	0.40	3.90	4.39	0.60	0.64	18.34	18.67
Pasta 6	0.25	0.53	-0.01	0.22	0.26	6.27	7.66
	0.30	0.62	0.76	0.37	0.34	3.68	3.78
	0.40	0.01	0.03	0.19	0.21	0.13	0.10
Pasta 7	0.25	-1.95	-2.31	0.01	0.02	11.85	12.04
	0.30	-0.13	0.40	0.30	0.28	5.06	4.78
	0.40	0.27	0.15	0.73	0.75	1.65	1.70

the percentage of explained variance were considered to evaluate the model performance. Moreover, the histograms of the concentration maps generated by the MCR-ALS models were considered to evaluate the concentration value obtained by MCR-ALS in the sample surface, together with the Pearson correlation coefficient between the pure and the recovered spectra of the constituents (similarity). MCR-ALS is not, per se, a quantitative method. Therefore, Ordinary Least Squares (OLS) Regression was performed considering the real percentage of fiber and the one obtained by thresholding the fiber findings in the so-called concentration profiles obtained by MCR-ALS. The models were cross-validated by 10 iterations of random subsets (30% of the data in each iteration). The assessment of the OLS models was done by calculating the bias, R² and the root mean square error (RMSE) in calibration and validation. Data analysis was performed using Matlab software (R2016a, The Mathworks Inc., Natick, MA, USA) and the MCR-ALS 1.0 toolbox, freely available at www.mcrals.wordpress.com (last accessed July 2020).

3. Results and discussion

3.1. Control sample and pure flour spectra as initial estimates

The mean NIR spectra of the control, flour, and fiber samples are

shown in Fig. 1. The spectra of all fiber samples were strongly similar, with a slight variation on the baseline. Control samples had the highest absorbance values, followed by pure flour and, then, pure fiber samples. Additionally, it was possible to observe a peak around 1200 nm, which corresponds to the stretching vibration of the C—H bond in the second overtone. Two absorption bands of stretching vibration of the O—H bond in the first overtone of water were also observed around 1450 and 1940 nm. Peaks of absorption around 1780 nm and 2100 nm were also highlighted, corresponding to the stretching vibration of the C—H bond in the first overtone of cellulose and a combination of O—H deformation and C—O stretching vibration, characteristic of starch. Peaks at 2242 and 2294 nm were also noticed, corresponding to N—H stretching vibration. Finally, a combination of C—H and C—C stretching vibration was observed at 2480 nm (Osborn, Fearn, & Hindle, 1993). It is important to observe the high similarity between pasta/flour and fiber spectra. This had a negative impact on further MCR-ALS models, making necessary the constraints and additional information mentioned before to obtain reliable models.

3.2. MCR-ALS in pasta samples

3.2.1. Fiber distribution

MCR-ALS models were developed for each type of sample separately, using the spectra of pure fiber, pure flour, and control sample as initial estimates. Some of the final concentration profiles obtained can be seen in Fig. 2. The concentration maps revealed how fibers behaved in each sample. Some fibers showed to be better distributed in the samples, as Fiber 1. However, some of them agglomerated in some parts of the sample, as Fibers 3, 4, and 7. This may be explained by the particle size, composition, and/or percentage of fiber added to the samples. The particle size of the fiber shown on Supplementary material – Table S1 is the average of different size meshes used for granulometry determination. However, Fibers 3 and 7, for instance, presented bigger particles, as can be noticed on Supplementary material – Fig. S1. This aspect could be noticed in their respective concentration maps (Samples 3 and 7), where red spots (fiber) could be clearly identified.

Besides, the composition of each fiber was different. Fibers 2, 3, 6, and 7 are pure insoluble fibers, while 1, 4, and 5 are a mix of different soluble and insoluble fibers in a different ratio. However, even having soluble fibers in common, Fibers 1, 4, and 5 were differently distributed in their respective samples (Samples 1, 4, and 5). The amount of soluble fiber may explain this fact in each mix, which is 20%, 80%, and 50%, respectively. Soluble fibers generate a network around the granules of starch (Brennan & Tudorica, 2007), so the higher amount of psyllium in Fiber 4 may have led to the agglomerates observed in Sample 4. Similar behavior was observed on the concentration map of Sample 5 (data not shown).

The histograms of samples considering the spectral information of the control sample as the initial estimate are shown in Fig. S3 (Supplementary material). As observed, most of the pixels in the samples had fiber concentrations between 0 and 0.5. Apart from Sample 5, which had the most spread pixels values, all the samples showed a Gaussian distribution of the pixels. Moreover, Samples 3, 4, and 7 were slightly more disperse than Samples 1, 2, and 6. In addition, it is important to highlight that some samples presented a large number of pixels around 1, especially Samples 3 and 7, as seen in their respective concentration maps, indicating the presence of pure fiber in some of the pixels.

3.2.2. MCR-ALS models

The results of %LOF and similarity are reported on Supplementary material – Table S2. The LOF percentages of the MCR-ALS models for all samples were lower than 6%, regardless of the initial estimates. Moreover, all the models were able to explain more than 99% of the total variance in each sample, suggesting a good performance of the models. The similarity between the pure spectra and the spectra recovered by the augmented models was above 96% and 98% in models using pure flour

A.T. Badaró et al.

Food Chemistry 343 (2021) 128517

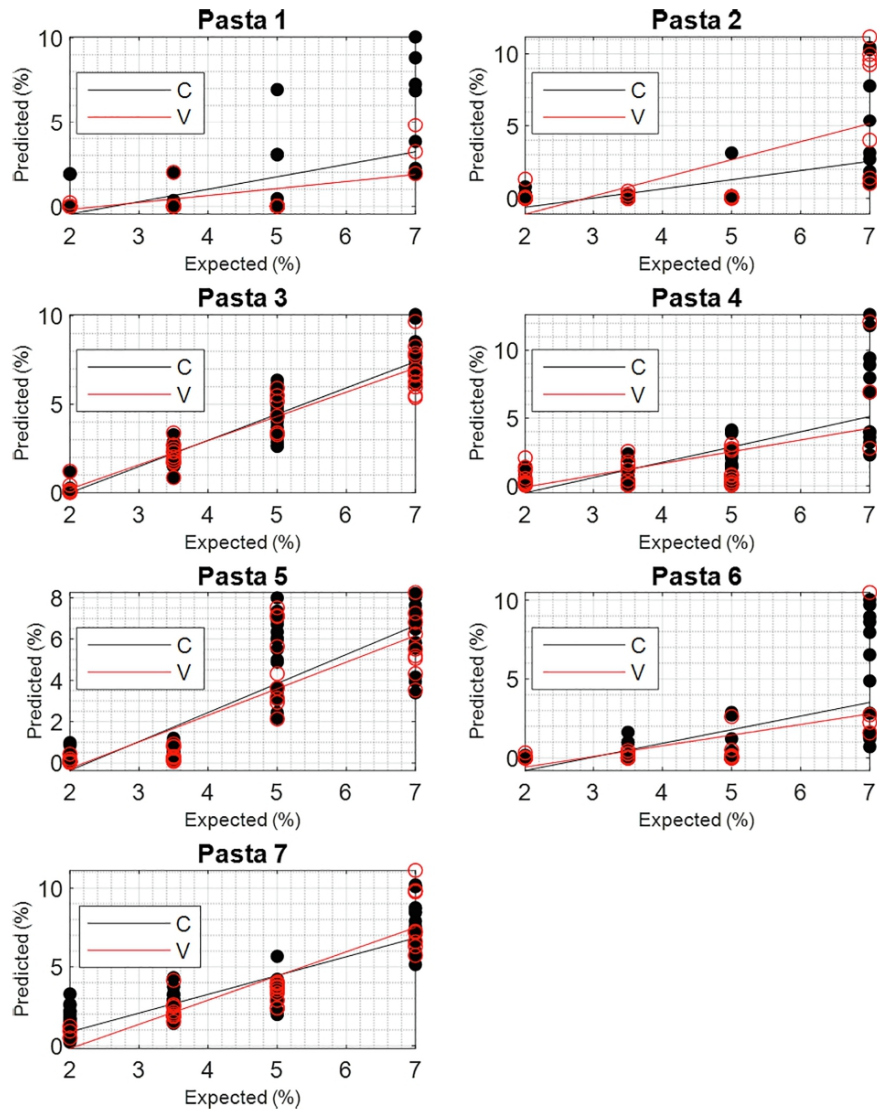


Fig. 5. Expected versus Predicted fiber content after MCR-ALS analysis (Models highlighted in dark grey in Table 1 for the control sample as the initial estimate).

and control sample as the initial estimates, respectively. Fig. 3 shows an overview of both pure and recovered spectra for Sample 3 using the control sample as the initial estimate. The results for the rest of the samples were of the same quality, assessing the validity of the MCR-ALS models performed and the fact that even having one pure spectrum as initial estimates, the actual composition of the fiber in the samples might be slightly different. An example of this is the fact that there is a

deviation between the obtained spectral profiles and the considered pure ones. Nevertheless, this deviation occurs around 1450 nm, associated with water (Osborn et al., 1993). This may be associated with the fact that pure fiber used as the initial estimate was in its powder form, and the prediction was performed on the pasta samples, which have passed through the drying process.

3.2.3. Distribution of fiber content in the surface of samples

The concentration maps resulted from MCR-ALS models do not give the exact amount of fiber added to the samples since they provide the signal contribution of each constituent (fiber and flour/control sample) that can be more related to the volume fraction than to the mass fraction (Franch-Lage et al., 2011). However, they provide the necessary information to establish a threshold and use the obtained value to make a regression model (with OLS) to ascertain the validity of the approach. Establishing a proper threshold is not an easy task. In our case, we tested three different values (0.25, 0.30 and 0.40) that we found significant in the concentration maps of enriched-pasta samples (Fig. 2) and the corresponding histograms (Supplementary material – Fig. S3). Pixels with values over the threshold were identified as plausible fiber indicators, and those below the threshold were identified as flour/control sample. As an example, the concentration maps of Sample 3 and its respective images for each threshold are shown in Fig. 4. The lower was the threshold value, the more pixels were identified as fiber; consequently, the higher was the predicted amount of fiber. Then, the amount of fiber was determined by the ratio in the sample of the area defined as fiber to the area defined as flour/control sample. This step provided a better overview of fiber distribution in the sample surface.

Bearing in mind that this is a semi-quantitative strategy, the percentage of fiber found by MCR-ALS was used as a parameter to perform regression with the real value. The results are displayed in Table 1. In this table, values of bias, R^2 and RMSE are reported for both calibration (C) and validation (V). Moreover, Fig. 5 shows the fiber content predicted versus expected using fiber and control sample as initial estimates. The differences observed between the actual values and the values predicted by regression were due to the fact that during sample production, the fiber characteristics and the particle size interfered in the complete dissolution of the powder in the mixture, being possible the presence of agglomerates of fiber, what may explain the lack of homogeneity in the sample surface. Moreover, from the amount of sample produced, only 30 samples of each type were randomly selected, which did not guarantee that each sample of pasta analyzed contained the exact amount of fiber indicated. This, in fact, emphasizes the importance of this work, since the distribution of fiber within the sample is not homogeneous, which may affect the quality and properties of the batch produced in the industry, that may not contain the minimum level necessary, while other parts of the batch may have a higher amount of fiber.

The models with the best performance according to each threshold for each sample are highlighted in dark grey. Overall, Samples 3, 5, and 7 showed the best correlation with the real content, unlike Samples 1, 2, 4, and 6. However, this may not indicate a bad performance of the models, but, as mentioned before, a problem with the homogeneity of the sample. Also, a great dispersion regarding the predicted fiber content among samples with the same percentage of fiber, which may have decreased the performance of the models.

4. Conclusions

The results found in this work confirmed the hypothesis that NIR-HSI, coupled to MCR-ALS, can identify and quantify fiber added to pasta samples. Homogeneity is a very important feature in the industry to guarantee the quality of the final product. In this work, the homogeneity of pasta samples showed to be a barrier in the quantification of the expected fiber content in the sample. However, the results showed low LOF, high total variance, and great similarity between pure and recovered spectral profiles, denoting good replicability of the models and ability to quantify the amount of fiber in the region of the sample analyzed.

This target could have been addressed by using other approaches (classification models, Classical Least Squares, etc.). Nevertheless, due to the complexity of the signals (high similarity of the fiber and flour/control sample spectra), we found that a step of resolution/unmixing

was needed to extract semi-quantitative results. The outputs showed that even using the raw ingredient spectral information as in pure fiber and flour, it is possible to develop reliable models to identify fiber in pasta. In addition, the concentration maps can be very useful in monitoring the homogeneity of the samples. The performance of the models also showed that it is possible to have a qualitative overview of the fiber distribution in pasta samples, and, in some cases, it is possible to quantify the amount of fiber in the sample surface. Thus, this tool presented a great possibility to apply such technique as a qualitative and quantitative method for authentication of fiber-enriched pasta.

CRedit authorship contribution statement

Amanda Teixeira Badaró: Conceptualization, Validation, Formal analysis, Investigation, Writing - original draft. **Jose Manuel Amigo:** Software, Validation, Formal analysis, Writing - review & editing. **Jose Blasco:** Writing - review & editing, Supervision. **Nuria Aleixos:** Writing - review & editing, Supervision. **Amanda Rios Ferreira:** Resources, Investigation. **Maria Teresa Pedrosa Silva Clerici:** Resources, Investigation. **Douglas Fernandes Barbin:** Conceptualization, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) [Finance Code 001]; São Paulo Research Foundation (FAPESP) [grant numbers 2008/57808-1, 2014/50951-4, 2015/24351-2, 2017/17628-3, 2019/06842-0]; and by GVA-IVIA and FEDER funds through project IVIA-51918. The authors would like to thank Nutrassim Food Ingredients company for the donation of the fiber samples, the support provided by Mrs. Cristiane Vidal during NIR-HSI system operation and data processing and Dr. Celio Pasquini for promptly receiving us in the laboratory that he coordinates (Grupo de instrumentação e automação em química analítica, Instituto de química, Universidade Estadual de Campinas, Campinas-SP, Brazil) to data acquisition.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2020.128517>.

References

- Amigo, J. M. (2020). *Hyperspectral Imaging* (1st ed.). Elsevier.
- Amigo, J. M., & Ravn, C. (2009). Direct quantification and distribution assessment of major and minor components in pharmaceutical tablets by NIR-chemical imaging. *European Journal of Pharmaceutical Sciences*, 37(2), 76–82. <https://doi.org/10.1016/j.ejps.2009.01.001>.
- ANVISA. Resolução RDC No 54, DE 12 DE NOVEMBRO DE 2012. (2012). http://portal.anvisa.gov.br/documents/33880/2568070/rdc0054_12_11_2012.pdf/c5ac23fd-974e-42c-9fbc-48f7e0a31864 Accessed 06 April 2020.
- Brennan, C. S., & Tudorica, C. M. (2007). Fresh pasta quality as affected by enrichment of nonstarch polysaccharides. *Journal of Food Science*, 72(9), 3659–3665. <https://doi.org/10.1111/j.1750-3841.2007.00541.x>.
- Bustos, M. C., Perez, G. T., & Leon, A. E. (2015). Structure and quality of pasta enriched with functional ingredients. *RSC Advances*, 5(39), 30780–30792. <https://doi.org/10.1039/C4RA11857J>.
- de Juan, A. (2020). Multivariate curve resolution for hyperspectral image analysis. In J. M. Amigo (Ed.), *Hyperspectral Imaging* (1st ed., pp. 115–150). Elsevier.
- de Juan, A., Jaumot, J., & Tauler, R. (2014). Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. *Analytical Methods*, 6(14), 4964–4976. <https://doi.org/10.1039/C4AY00571F>.

A.T. Badaró et al.

Food Chemistry 343 (2021) 128517

- Juan, A. d., Tauler, R., Dyson, R., Marcolli, C., Rault, M., & Maeder, M. (2004). Spectroscopic imaging and chemometrics: A powerful combination for global and local sample analysis. *TRAC Trends in Analytical Chemistry*, 23(1), 70–79. [https://doi.org/10.1016/S0165-9936\(04\)00101-3](https://doi.org/10.1016/S0165-9936(04)00101-3).
- European Commission. Nutrition Claims. (2012). https://ec.europa.eu/food/safety/labelling_nutrition/claims/nutrition_claims_en Accessed 06 April 2020.
- Ferreira, A. R. (2018). *Dry Fettuccine pasta with partial replacement of Triticum durum semolina by white fibers*. University of Campinas.
- Ferreira, M. M. C. (2015). *Quimiometria- Conceitos, Métodos e Aplicações*. Campinas: Editora Unicamp.
- Firmani, P., Hugelier, S., Marini, F., & Ruckebusch, C. (2018). MCR-ALS of hyperspectral images with spatio-spectral fuzzy clustering constraint. *Chemometrics and Intelligent Laboratory Systems*, 179, 85–91. <https://doi.org/10.1016/j.chemolab.2018.06.007>.
- Forchetti, D. A. P., & Poppi, R. J. (2017). Use of NIR hyperspectral imaging and multivariate curve resolution (MCR) for detection and quantification of adulterants in milk powder. *LWT - Food Science and Technology*, 76, 337–343. <https://doi.org/10.1016/j.lwt.2016.06.046>.
- Franch-Lage, P., Amigo, J. M., Skibsted, E., MasPOCH, S., & Coello, J. (2011). Fast assessment of the surface distribution of API and excipients in tablets using NIR-hyperspectral imaging. *International Journal of Pharmaceutics*, 411(1–2), 27–35. <https://doi.org/10.1016/j.ijpharm.2011.03.012>.
- Ghaffari, M., Hugelier, S., Duponchel, L., Abdollahi, H., & Ruckebusch, C. (2019). Effect of image processing constraints on the extent of rotational ambiguity in MCR-ALS of hyperspectral images. *Analytica Chimica Acta*, 1052, 27–36. <https://doi.org/10.1016/j.aca.2018.11.054>.
- Huang, H., Liu, L., & Ngadi, M. (2014). Recent developments in hyperspectral imaging for assessment of food quality and safety. *Sensors*, 14(4), 7248–7276. <https://doi.org/10.3390/s140407248>.
- Ishihara, S., Hattori, Y., & Otsuka, M. (2019). MCR-ALS analysis of IR spectroscopy and XRD for the investigation of ibuprofen - Nicotinamide cocrystal formation. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 221, 117142. <https://doi.org/10.1016/j.saa.2019.117142>.
- Kowalczewski, P., Lewandowicz, G., Makowska, A., Knoll, I., Błaszczyk, W., Białas, W., & Kubiak, P. (2015). Pasta fortified with potato juice: structure, quality, and consumer acceptance: Pasta fortified with potato juice. *Journal of Food Science*, 80(6), S1377–S1382. <https://doi.org/10.1111/1750-3841.12906>.
- Ma, J., Sun, D.-W., Pu, H., Cheng, J.-H., & Wei, Q. (2019). Advanced techniques for hyperspectral imaging in the food industry: Principles and recent applications. *Annual Review of Food Science and Technology*, 10, 197–220. <https://doi.org/10.1146/annurev-food-032818>.
- Menesatti, P., Antonucci, F., Pallottino, F., Bucarelli, F. M., & Costa, C. (2014). Spectrophotometric qualification of Italian pasta produced by traditional or industrial production parameters. *Food and Bioprocess Technology*, 7(5), 1364–1370. <https://doi.org/10.1007/s11947-013-1139-0>.
- Menesatti, P., & Bucarelli, A. (2007). Quality classification of Italian wheat durum spaghetti by means of different spectrophotometric techniques. In *Optics for natural resources, agriculture, and foods II* (Vol. 6761, p. 67610D). <https://doi.org/10.1117/12.735553>.
- Muneer, F., Johansson, E., Hedenqvist, M. S., Plivelic, T. S., Markedal, K. E., Petersen, I. L., Sørensen, J. C., & Kuktaitė, R. (2018). The impact of newly produced protein and dietary fiber rich fractions of yellow pea (*Pisum sativum* L.) on the structure and mechanical properties of pasta-like sheets. *Food Research International*, 106, 607–618. <https://doi.org/10.1016/j.foodres.2018.01.020>.
- Nascimento, J., & Martín, G. (2020). Nonlinear spectral unmixing. In J. M. Amigo (Ed.), *Hyperspectral imaging* (1st ed., pp. 151–166). Elsevier.
- Osborn, B. G., Fearn, T., & Hindle, P. H. (1993). *Theory of near infrared spectroscopy. In Practical NIR spectroscopy with applications in food and beverage analysis* (2nd ed., pp. 13–35). London: Longman Singapore Publishers (Pte) Ltd.
- Piqueras, S., Burger, J., Tauler, R., & de Juan, A. (2012). Relevant aspects of quantification and sample heterogeneity in hyperspectral image resolution. *Chemometrics and Intelligent Laboratory Systems*, 117, 169–182. <https://doi.org/10.1016/j.chemolab.2011.12.004>.
- Rakheh, N., Fellows, C. M., & Sissons, M. (2015). Evaluation of the technological and sensory properties of durum wheat spaghetti enriched with different dietary fibres: Enrichment of durum wheat spaghetti with dietary fibres. *Journal of the Science of Food and Agriculture*, 95(1), 2–11. <https://doi.org/10.1002/jsfa.6723>.
- Terra, L. R., Roque, J. V., Pola, C. C., Gonçalves, I. M., Soares, N. d. F. F., & Teófilo, R. F. (2020). Study of chemical compound spatial distribution in biodegradable active films using NIR hyperspectral imaging and multivariate curve resolution. *Journal of Chemometrics*, 34(1). <https://doi.org/10.1002/cem.3193>.
- Vidal, M., & Amigo, J. M. (2012). Pre-processing of hyperspectral images. Essential steps before image analysis. *Chemometrics and Intelligent Laboratory Systems*, 117, 138–148. <https://doi.org/10.1016/j.chemolab.2012.05.009>.
- Wu, D. I., & Sun, D.-W. (2013). Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review — Part I: Fundamentals. *Innovative Food Science & Emerging Technologies*, 19, 1–14. <https://doi.org/10.1016/j.ifset.2013.04.014>.
- Zhang, X., de Juan, A., & Tauler, R. (2015). Multivariate curve resolution applied to hyperspectral imaging analysis of chocolate samples. *Applied Spectroscopy*, 69(8), 993–1003. <https://doi.org/10.1366/14-07819>.
- Zhang, X., & Tauler, R. (2013). Application of multivariate curve resolution alternating least squares (MCR-ALS) to remote sensing hyperspectral imaging. *Analytica Chimica Acta*, 762, 25–38. <https://doi.org/10.1016/j.aca.2012.11.043>.

CHAPTER 4

Study of changes in pasta during cooking by hyperspectral imaging

Study of changes in pasta during cooking by hyperspectral imaging

Authors: Amanda Teixeira Badaró^{1,2,3,*}, Douglas Fernandes Barbin¹, Amanda Rios Ferreira⁴, Maria Teresa Pedrosa Silva Clerici⁴, Nuria Aleixos² and Jose Blasco³

¹Department of Food Engineering, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, 13083-862, Campinas – São Paulo, Brazil.

²Departamento de Ingeniería Gráfica, Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia, Spain.

³Centro de Agroingeniería, Instituto Valenciano de Investigaciones Agrarias (IVIA), Ctra. CV-315, km. 10,7, 46113, Moncada, Valencia, Spain.

⁴Department of Food Technology, University of Campinas, Rua Monteiro Lobato, 80, Cidade Universitária, 13083-862, Campinas – São Paulo, Brazil.

*Corresponding author; tel: +55 19 98917-3064, E-mail: tbadaro.amanda@gmail.com

Abstract

Pasta is a very important and popular food. The addition of fibre to pasta can affect its structure, since it modifies the protein-matrix. Among the most important pasta attributes, optimal cooking time, cooking and overcooking tolerance, can be named. Some of these assessments can be very subjective, so this study investigates the use VIS/NIR-HSI in the transmission mode for the study of changes in pasta during cooking and the determination of optimal cooking time of pasta as an alternative to automate the determination of pasta attributes. *Fettuccine* samples were produced with seven types of fibre in two percentages (3.5 % and 7 %). A total of 140 experiments (7 types of fibre x 2 percentages x 10 repetitions) were carried out and, for each one, two units of pasta were removed from cooking water each 90 seconds. The cooking process was carried out for 18 minutes, totalling 13 times. VIS/NIR hyperspectral images were acquired and the two spectral range were separately evaluated, but VIS data did not show any relationship among samples. Therefore, further multivariate data analysis was carried out only on NIR data. PCA was performed and the PCA scores that better grouped the samples were used as variables to perform a Linear Discriminant Analysis (LDA). In comparison, Partial Least Square Discriminant Analysis (PLSDA) was performed. LDA had values of sensitivity and specificity between 0.14 – 1.00 and 0.51 – 1.00, respectively, and non-error rate (NER) over 0.62. PLSDA had values of sensitivity and specificity between 0.67 – 1.00 and 0.10 – 1.00, respectively, and NER over 0.80. This study suggested that hyperspectral image in the transmission mode have good potential as an objective method to optimum cooking time determination, meeting this industry need.

Keywords: Cooking time, Enriched-pasta, Fibre, Multivariate data analysis.

4.1.Introduction

Pasta is a very important and popular food due to its ease of preparation, transportation and cooking, relatively low price, and nutritional value. It is traditionally obtained from the extrusion of durum wheat semolina and water, followed by drying. From the nutrition point of view, pasta has a low fat and salt content and a high content of carbohydrates, making it a suitable energy source (Majzoobi et al., 2011). Beyond these many pasta quality parameters, its long shelf life and stability, when compared to other flour products, make pasta a potential food for addition of alternative ingredients, as fibres. The incorporation of dietary fibres in pasta can considerably contribute to reducing its glycaemic index, for instance, which would result in a product with a higher nutritional value to the consumer, in comparison to conventional pasta (Bustos et al., 2011).

However, the incorporation of nutrients in traditional foods can affect their sensory and technological properties in a negative way. By adding functional ingredients, desirable properties of pasta, such as textural and cooking qualities, can change (Rakesh et al., 2015). The addition of fibre, for instance, can affect pasta structure, since it modifies the protein-matrix. Soluble fibres bind with protein, difficult the interactions between starch and protein difficult (Bustos et al., 2015). Consequently, the amount and type of ingredient used in pasta enrichment must be controlled in a way that the new product has satisfactory taste and nutritional value, but without many quality parameters changes. Among the most important pasta attributes, optimal cooking time, cooking and overcooking tolerance, colour and cooking losses can be named. However, for most of these attributes, the traditional methods of analysis can be time-consuming, expensive, and invasive, besides not allowing a control of how added nutrients can influence the food matrix. Moreover, some of these analyses used to estimate pasta-cooking quality can be very subjective, as sensory analysis, which is a challenge when many samples must be evaluated. In addition, optimal cooking time of pasta is determined by observing the time that the core of the pasta sample stays between two transparent glass slides when squeezing them (Del Nobile et al., 2005).

Therefore, many researchers have searched for more objective methods to evaluate pasta cooking quality parameters. Thus, techniques such as Near Infrared Hyperspectral Imaging (NIR-HSI) can be used to evaluate cooking time of fibre-enriched pasta. NIR-HSI combines spectral and spatial information of a sample, being able to provide both physical and

chemical characteristics at the same time. This technique has been widely used, including for the inspection of adulterants in organic wheat flour (Su and Sun, 2017) and for the differentiation between industrial and traditional pasta production (Menesatti et al., 2014). However, NIR-HSI systems measure the surface of the sample, entering only few millimetres in the sample, which is not always representative for thick and very heterogeneous samples. In this case, a transmission measurement is an option (Wold et al., 2016).

Thus, the aim of this study is to use VIS/NIR-HSI in the transmission mode for the study of changes in pasta during cooking and the determination of optimal cooking time of pasta as an alternative to automate the determination of pasta attributes.

4.2. Materials and methods

4.2.1. Sample preparation

Samples of dry *fettuccine*-type pasta were produced in the Department of Food Technology at University of Campinas (Brazil) according to the methodology proposed by Ferreira (2018). The *fettuccine* samples were produced replacing flour by seven types of fibre in two percentages (3.5 % and 7 % (w/w)). Fibres differed on source, particle size and solubility (Table 4.1).

Table 4.1. Properties of fibres used in pasta formulation.

Pasta sample	Fibre source	Average particle size (μm)
Pasta 1	Mix of 20 % soluble psyllium and 80 % insoluble bamboo	160
Pasta 2	Insoluble bamboo	60
Pasta 3	Insoluble bamboo	145
Pasta 4	Mix of 80 % soluble psyllium and 20 % insoluble cellulose	160
Pasta 5	Mix of 50 % soluble psyllium and 50 % insoluble cellulose	160
Pasta 6	Insoluble wheat	60
Pasta 7	Insoluble wheat	145

4.2.2. Cooking process

Pasta samples (approximately 1 cm wide x 0.2 cm thick) were cut into pieces approximately 5 cm long. The cooking process was carried out following the AACCI (2010) protocol with some modifications. For each type and percentage of fibre, ten portions of 25 grams of pasta were weighted, totalling 140 experiments (7 types of fibre x 2 percentages x 10 repetitions). The cooking process was carried out for 18 minutes, when two units of pasta were removed from cooking water each 90 seconds, totalling 13 times (0, 1.5, 3, 4.5, 6, 7.5, 9, 10.5, 12, 13.5, 15, 16.5 and 18 minutes). In total, there were 1820 samples analysed (140 experiments x 13 times).

4.2.3. Acquisition hardware

The pasta VIS/NIR-HSI images were acquired in a Hyperspectral Imaging System consisting of a monochrome camera (CoolSNAP ES, Photometrics, AZ, USA) with a high level of sensitivity between 320 nm and 1100 nm, coupled to two liquid-crystal tuneable filters (LCTF), one sensitive to the visible (Varispec VIS07, CRI Inc) and one sensitive to NIR (Varispec NIR07, CRI Inc). The samples were placed on a holder that fit in a hemispherical dome as shown in Figure 4.1 and illuminated by a halogen lamp (20W) from below, so that the VIS/NIR-HSI images were acquired in transmission mode. However, due to the thin thickness of pasta samples, a lot of light crossed the samples and the images appeared saturated, being necessary the correction of the time of the light exposition in each wavelength. Therefore, a calibration was carried out so that the integration time was decreased/increased as much as possible while ensuring that the maximum intensity (saturation) was not reached for any wavelength in any region of the image. The calibration was defined using uncooked pasta samples without fibre as the target reference, considering that this was the case that the maximum light would cross. As an alternative to ensure that any other sample would have saturated pixels, a percentage of 80% of maximum intensity (saturation) was considered to establish the integration time. Moreover, to avoid the low sensitivity of the sensors close to the edges of this range, the images were captured at every 10 nm in the working spectral range of 450–1020 nm. Samples were removed from water, and the excess of water was removed by

using paper towel. The two units of cooked pasta samples were placed manually on the holder in a way that only the light that was transmitted through the pasta reached the camera.

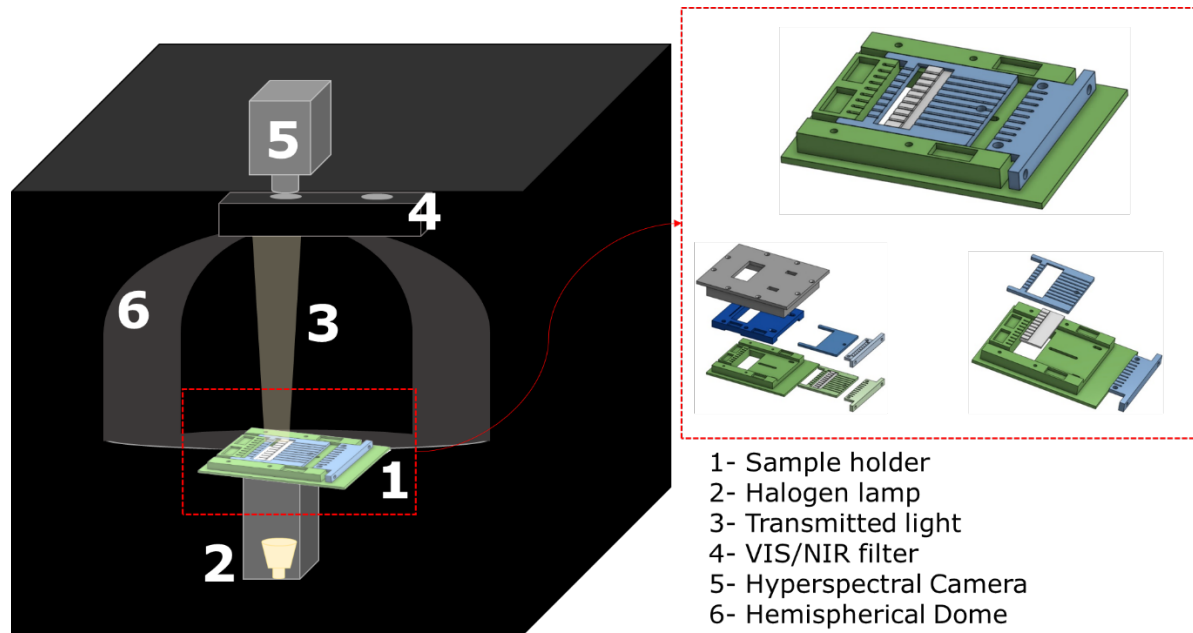


Figure 4.1. Scheme of the apparatus used for acquiring the images, the zoomed image shows the surface used to place the cooked pasta samples.

4.2.4. Data processing

The hyperspectral images were pre-processed for segmentation of the regions of interest (ROI). The ROIs were manually selected with help of a Matlab routine. After removing the background from the images and creating a final image containing only the region of interest (ROI), a mean spectrum that represent each sample was extracted (the mean spectrum of the two pasta units). In order to extract the actual response of the samples at each wavelength, while avoiding light-dependent intensities, a correction was performed using the image of a standard pasta reference, the same used to build the calibration of the system. The influence of the minimum dark current of the camera was also captured by switching off the lamps and placing a cap in the lens to prevent the light from getting inside the camera. After that, a mean spectrum of white and dark references were extracted. Then, the correction was performed using the Equation (1):

$$I = \frac{I_0 - I_{black}}{I_{white} - I_{black}} \quad (1)$$

where I_0 is the mean spectrum of the raw image of the pasta, I_{white} is the mean spectrum of the image of the standard pasta reference, and I_{black} is the mean spectrum of the image acquired while avoiding any light source. Image segmentation and spectrum extraction were executed using Matlab R2019b software (The Mathworks Inc., Natick, MA, USA). It is important to highlight that VIS and NIR spectral information were extracted separately.

4.2.5. Multivariate data analysis

Principal Component Analysis (PCA) was applied on VIS/NIR-HSI data, as exploratory data analysis, to obtain an overview of variation among the samples. PCA is an unsupervised method which combines linearly the variables into components known as principal components (PCs), which explain the greatest variance in the dataset and keep only the relevant information of the spectra (Rachmawati et al., 2017). Spectral pre-processing techniques were applied to investigate the effect of light scattering and other effects non-related to the sample composition in the spectral data. Hence, Standard Normal Variate (SNV), Multiplicative Scatter Correction (MSC), 1st and 2nd derivatives (order: 2, window: 7 points) were applied and compared to the PCA for raw data.

Linear discriminant analysis (LDA) is a very popular method for authentication, characterization and adulteration detection in foodstuffs (Esteki et al., 2018). In this study, LDA was applied in cooked pasta samples to verify the occurrence of a linear relationship in between the variables and to separate them into different classes. According to the literature, classification models based on small groups of classes perform better than multi-classes (Ziegler et al., 2016). Therefore, samples were split in 3 groups: 1 – Low cooking time (time 0 to 4.5 minutes); 2 – Intermediate cooking time (6.0 to 10.5 minutes); and 3 – High cooking time (12 to 18 minutes). These groups were established based on optimum cooking time of these type of pasta and the information found by Ferreira (2018).

Samples were also evaluated on PCA plots according to these groups. Then, PCA scores that better grouped the samples were used as variables to perform a Linear Discriminant Analysis (LDA). In comparison, Partial Least Square Discriminant Analysis (PLSDA) was performed to test the ability of the technique to predict optimum cooking time of pasta samples. The PLSDA assigned pasta samples into one of the three groups based on its spectral fingerprint. This technique is based on PLS regression technique and involves reduction of

dimensionality by projecting the data onto latent variables (LVs) subspace. Therefore, we chose an optimal number of LVs to constructed models based on the lowest prediction error in calibration and cross-validation (Erkinbaev et al., 2017). As both methods are supervised, and to test the model in an independent dataset Kennard-stone algorithm was used to split the data. Therefore, 70% of the samples were assigned to the calibration set and the remaining 30% constituted the validation set. The calibration models were developed based on cross-validation (random subsets with 10 splits and 5 iterations), and the classification model performance in both techniques was evaluated by calculating sensitivity, specificity and non-error rate (NER) in the validation. PCA and PLSDA were executed using Matlab R2019b software (The Mathworks Inc., Natick, MA, USA) and PLS Toolbox (Eigenvector Research, Inc., Manson, Washington, USA). LDA was executed using Classification toolbox for MATLAB - version 5.4 free available at www.michem.unimib.it (last accessed July 2021) (Ballabio & Consonni, 2013).

4.3. Results and discussion

4.3.1. Raw spectra of samples

The average VIS and NIR-HSI raw spectra of cooked pasta are shown in Figure 4.2. Samples were plotted according to the cooking time (0 - 18 minutes). Overall, NIR spectra had very similar shape, regardless the cooking time, but differed in the intensity of transmittance across the spectral region. Pasta samples 1, 2, 4, 5 and 7 (Figures 4.2H, 4.2I, 4.2K, 4.2L and 4.2N) had similar shape, with a slight valley around 710 nm corresponding to C-H stretching vibration in the fourth overtone and one around 960 nm, which correspond to stretching vibration of O-H in the second overtone. Except for Pasta 2 (Figure 4.2I), all other samples showed a slight peak around 850 nm related to C-H or C-C stretching vibration. Pasta 3 and 6 (Figures 4.2J and 4.2M) showed a peak around 950 nm followed by a valley at 960 nm. Differences on baseline between samples can be observed on all spectra. Samples removed from cooking at the beginning of the process had higher transmittance, confirming the assumption taken when performing the calibration that uncooked pasta samples had maximum light crossing them.

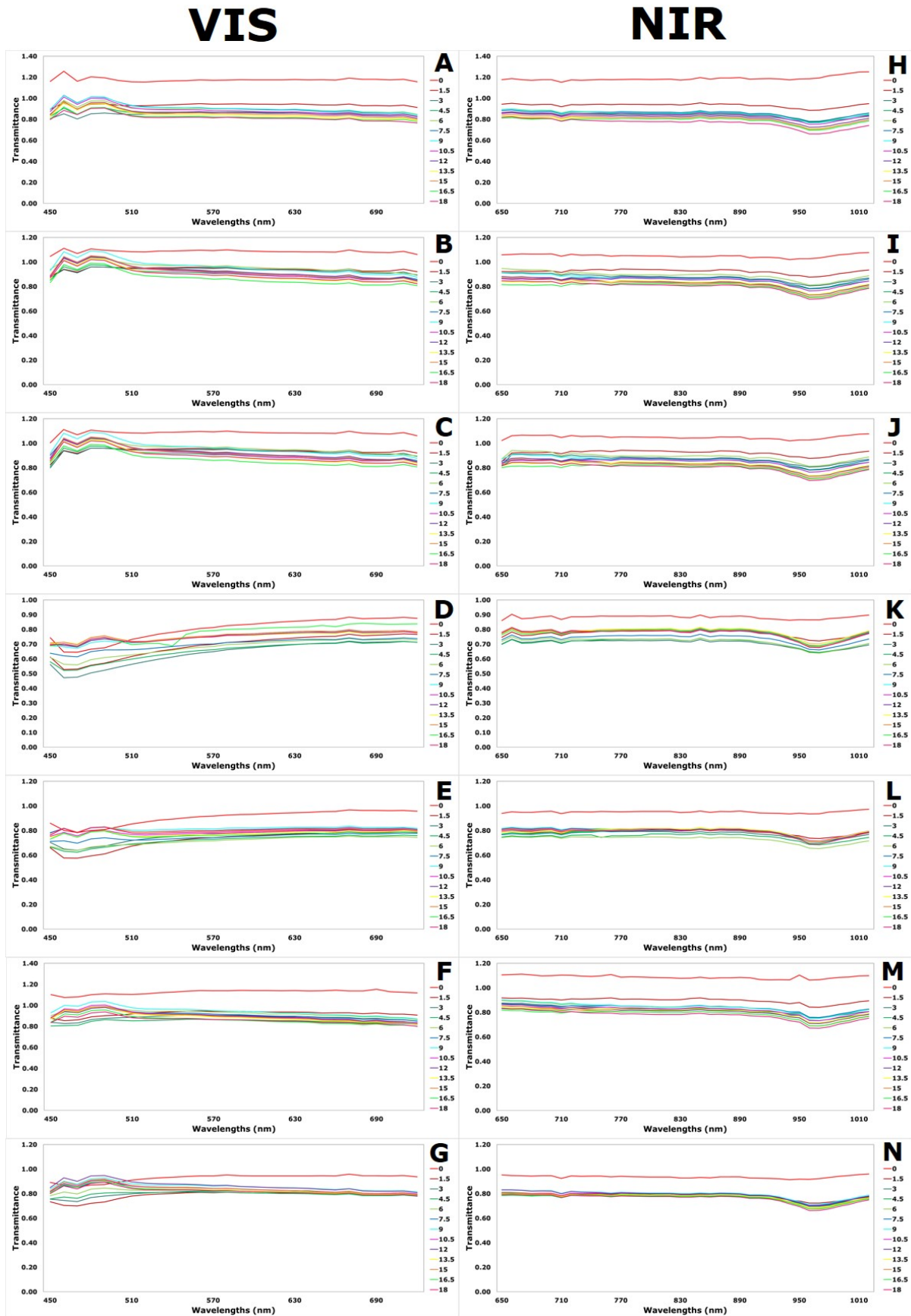


Figure 4.2. Average raw spectra of pastas samples in the NIR range (A to G) and in the VIS range (H to N).

On the other hand, VIS spectra did not show similar behaviour across the spectral region. The visible spectral range (400–720 nm) is related to the transition of electrons, which are associated to the determination of samples colour (Xiao et al., 2018). Figure 4.3 shows the visual changes in two types of pasta during cooking. A slight change in colour can be identified, at the beginning of cooking process, samples colour had dark yellow colour, and at the end light yellowish colour could be observed. The same was noticed in the other types of pasta (data not shown). A peak around 450 – 500 nm can be observed, associated to the yellow light (Xiao et al., 2018). Undercooked samples are more translucent, allowing that more light crosses the samples and is detected by the camera. This corroborates with what can be observed on Figure 4.2. Samples at time 0 had the highest transmittance values, while overcooked samples had lower transmittance values. However, it was expected a sequence of transmittance values from the highest to the lowest, from time 0 to time 18, respectively, as observed in the NIR spectra, but this did not happen. It is important to highlight that some of the fibre added to the pasta formulation were very insoluble and can be visually identified in the images displayed on Figure 4.3. These fibres may have blocked that transmission of light in the visible range, and as they are not equally distributed across the samples, this may have affected the transmittance detected by the camera. In fact, this was also observed in a previously work (Badaró et al., 2021) and emphasizes the importance of this work, since the distribution of fibre within the sample is heterogeneous and it may affect the properties of the final products, including the behaviour of samples during cooking. In order to reduce scattering effects, SNV, MSC, 1st and 2nd derivatives were applied to NIR and VIS spectra prior multivariate analysis. Data were also mean centred prior analysis. However, as expected, VIS spectra did not show any relationship among samples, so PCA, LDA and PLSDA analyses were carried out only with NIR data.

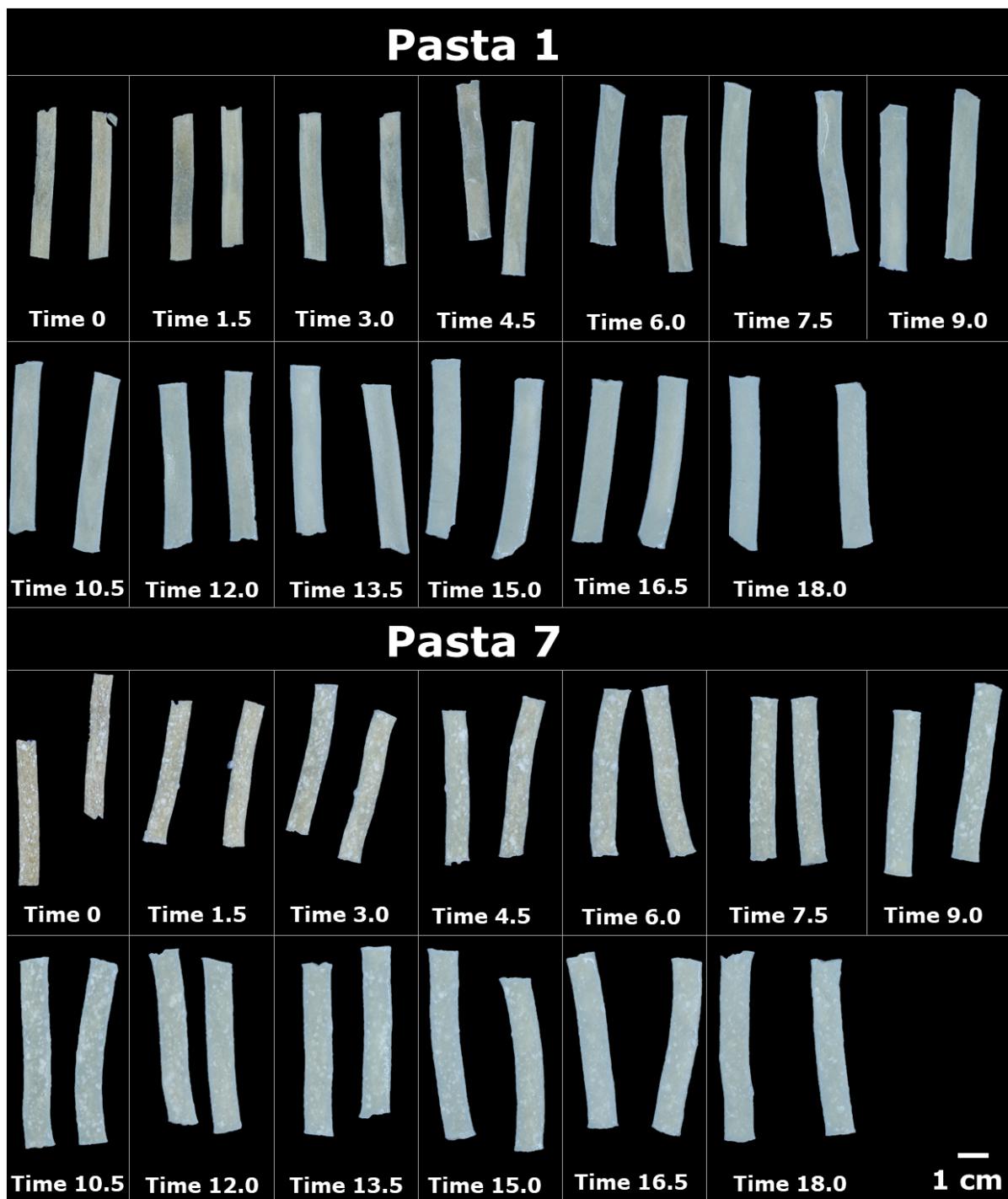


Figure 4.3. Images of pasta samples in each cooking time.

4.3.2. Principal Component Analysis (PCA)

Principal component analysis was performed on NIR spectra of all pasta samples, in order to visualize any variation among samples. In order to obtain a better sample separation, all pre-processes aforementioned were applied. All types of pasta showed better separation when 1st and 2nd derivatives (Savitzky–Golay smoothing, 7 points window, 2 order polynomial) were applied. Moreover, variable selection was manually performed by selecting the peaks in the loadings plot to improve the separation among samples. PCA scores of Pasta 1 and their correspondent loadings (after variable selection) are illustrated in Figure 4.4.

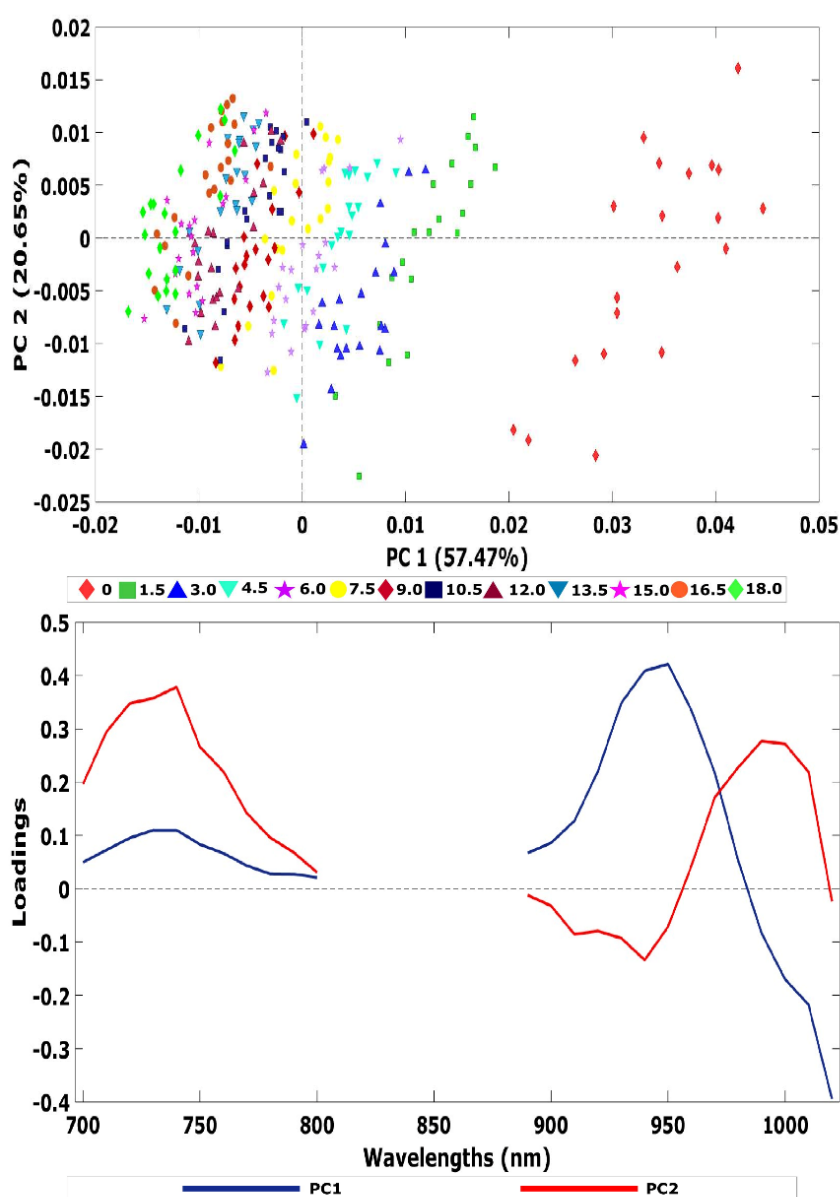


Figure 4.4. Score plots (PC1 vs PC2) of PCA performed on NIR spectra of Pasta 1 after Savitzky–Golay 1st derivative and the correspondent loadings plot.

The first two principal components explained around 78% of the total variance of the samples. This variability observed on the scores might be due to chemical changes that happen during cooking. Considering that the factors that can affect the characteristics of cooked pasta as quality and quantity of semolina protein, drying conditions, and the composition of cooking (Cunin et al., 1995) water were kept constant, the variation observed among pasta samples might be due to composition of fibre and cooking time. Commonly, at the beginning of cooking process starch granules are deeply in a protein matrix. However, the replacement of semolina for fibre during pasta production produces a less homogeneous product, which can affect gluten development (Aravind et al., 2012; Cunin et al., 1995). Figure 4.4 shows that samples at time 0 formed a cluster in the positive side of PC1. A high variance is observed among these samples and samples from next group (at time 1.5), suggesting a high chemical variation among these samples. Cunin et al. (1995) observed that starch granules were strongly swollen after immersing the pasta samples (*spaghetti* type) for 1 second in boiling water, but after 45 seconds, the surface became smoother, and some regions contained a filamentous network with rests of swollen granules. The time interval between these groups of samples were 90 seconds, which can explain the high lap between them in the score plot. Samples at 1.5 minutes were followed by samples at 3.0 minutes, and so on, until samples at 18 minutes, which formed a cluster in the negative side of PC1. However, visually analysing the PCA scores, we can state that samples around 12 – 13.5 minutes of cooking did not show a clear separation, suggesting that there was not a lot of chemical changes after this cooking time. Cunin et al. (1995) also reported that after 13 minutes of cooking process, it was not possible to distinguish between protein and starch, and, after 5 minutes in boiling water, the starch granules were swollen and had lost their shape.

The same behaviour was observed in the score plots of other pasta samples (data not shown). Taking this interpretation into account and considering the importance of identifying the cooking time, PCA was performed considering the three groups of samples previously established (low, intermediate and high cooking time). PCA score plots and their respective loadings plot, considering these three classes, are shown in Figure 4.5.

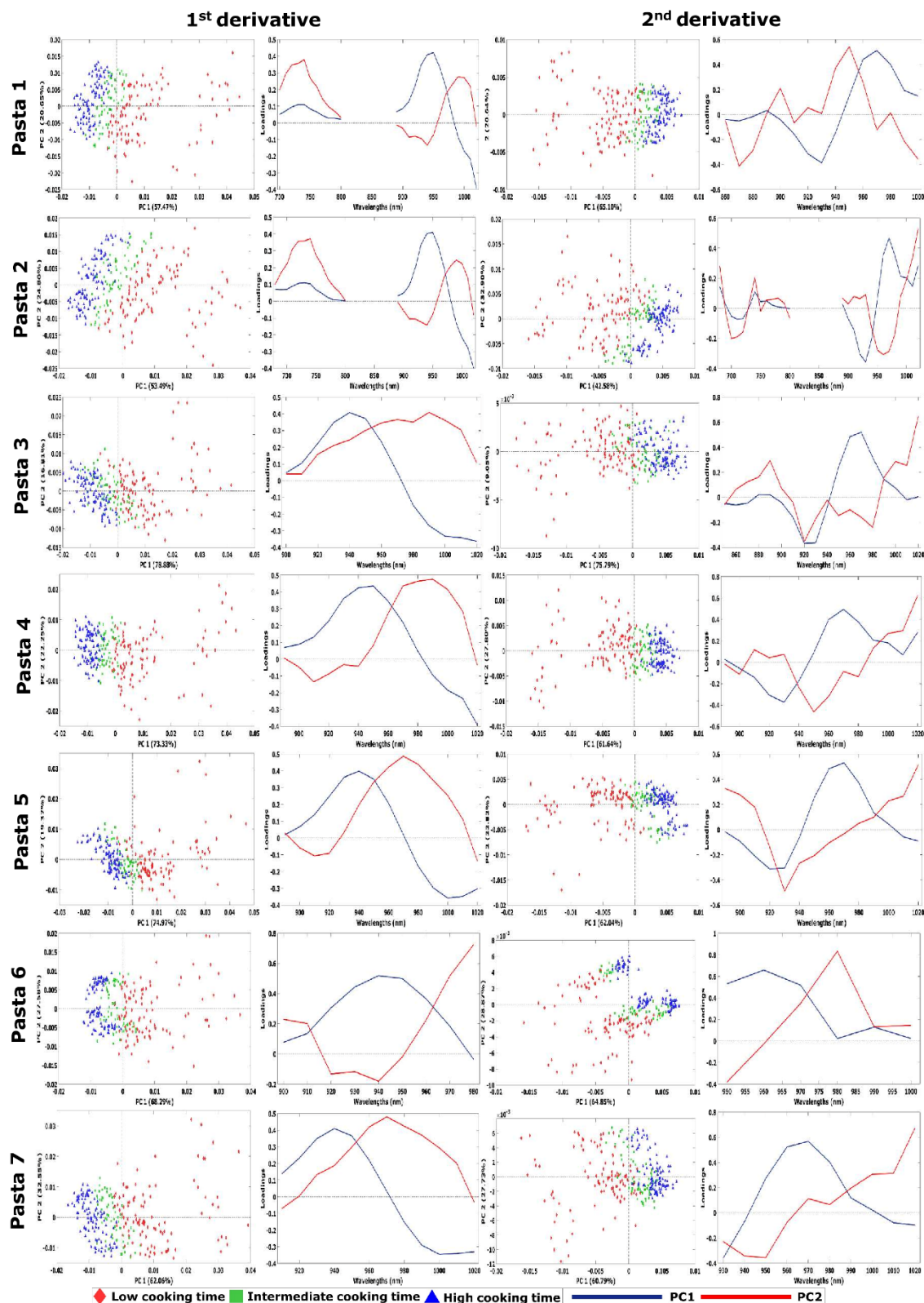


Figure 4.5. Score plots (PC1 vs PC2) of PCA performed on NIR spectra of all samples after Savitzky–Golay 1st and 2nd derivative and the correspondent loadings plot.

Overall, both pre-treatments showed good separation among the groups of samples. The variance among samples regarding cooking time is most explained by PC1. After pre-treated with 1st derivative, samples of low cooking time were displayed in the positive side of PC1 and samples of high cooking time were displayed in the negative side. Samples of intermediate cooking time is displayed in between the other two groups. For some of the pasta samples, there is no clear separation among the groups, with some overlapping of samples. Second derivative showed similar results, but samples of low cooking time displayed in the negative side of PC1, while samples of high cooking time were displayed in the positive side. In order to test the ability of classifying cooked pasta samples according to the cooking time, the PCA scores were used as input variables to developed discriminant classification models.

4.3.3. Classification of pasta samples - Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA) is a classification method used to discriminate samples into classes, i.e., low, intermediate and high cooking time. PCA scores were used as predictors for LDA models, and, as in PCA, samples were grouped into three classes defined as “Low cooking time” (0 – 6.0 minutes), “Intermediate cooking time” (7.5 – 10.5) and “High cooking time” (12 – 18 minutes). These classes were used to develop the classification models. Table 4.2 shows the models performance of LDA for qualitative analysis of cooking time of pasta. Data pre-treated with 1st and 2nd derivatives provided the calibration models with the best accuracies, and highest values of sensitivity, specificity and non-error rate (NER). However, Pastas 3 and 4 did not show good results for class 2, which is the intermediate cooking time. By observing the PCA scores of these two pasta samples, we can identify some overlapping. On the other hand, non-error rate (NER), or balanced accuracy, had satisfactory results, demonstrating good capacity of the models to discriminate pasta samples according to the cooking time. However, as an alternative to improve the discrimination of pasta samples, PLSDA models were constructed.

Table 4.2. Results of LDA models.

Sample	Pre-treatment	Number of PCs	Sensitivity			Specificity			Accuracy	NER
			C1	C2	C3	C1	C2	C3		
Pasta 1	1 st derivative	3	0.58	0.40	0.89	1.00	0.81	0.71	72%	0.62
	2 nd derivative	4	0.62	0.65	0.98	1.00	0.90	0.79	83%	0.75
Pasta 2	1 st derivative	3	0.93	0.38	0.98	1.00	0.96	0.62	79%	0.76
	2 nd derivative	3	0.86	0.46	0.98	1.00	0.95	0.66	80%	0.76
Pasta 3	1 st derivative	2	0.93	0.15	0.89	0.94	0.90	0.63	72%	0.66
	2 nd derivative	3	0.92	0.17	0.98	0.87	0.96	0.70	72%	0.69
Pasta 4	1 st derivative	3	0.88	0.57	0.83	0.98	0.84	0.83	79%	0.76
	2 nd derivative	4	0.85	0.18	1.00	0.98	0.96	0.51	74%	0.68
Pasta 5	1 st derivative	3	0.95	0.14	1.00	0.98	1.00	0.55	74%	0.70
	2 nd derivative	3	0.76	0.24	0.95	1.00	0.90	0.62	75%	0.65
Pasta 6	1 st derivative	5	0.70	0.80	0.88	0.99	0.85	0.92	84%	0.79
	2 nd derivative	4	0.69	0.67	1.00	1.00	0.93	0.81	87%	0.79
Pasta 7	1 st derivative	3	0.84	0.29	1.00	0.98	0.95	0.65	78%	0.71
	2 nd derivative	4	0.74	0.79	0.85	1.00	0.82	0.91	81%	0.79

C1 = Low cooking time

C2 = Intermediate cooking time

C3 = High cooking time

4.3.4. Classification of pasta samples - Partial Least Squares Discriminant Analysis (PLSDA)

In order to further explore the dataset and try to improve the classification ability, a supervised PLSDA model was built. Table 4.3 shows the results of PLSDA model for the validation set. As in PCA and LDA, 1st and 2nd derivatives had the best classification metrics. For most of the models two Latent Variables (LVs) were selected. It is important to highlight that PLSDA models showed better classification ability after variable selection, which was manually performed by selecting the peaks in the regression vectors created by the models. Although PLSDA involves reduction of dimensionality similarly to the principles of PCA, the models were not very accurate against false positives of class 2. The sensitivity and balanced accuracy were very satisfactory for all classes, though. Therefore, this suggests a good alternative in discriminating pasta samples according to the cooking time, regardless of content and source of the fibre added to the formulation, showing that the technique is suitable for different types of pasta.

Table 4.3. Results of PLSDA models.

Sample	Pre-treatment	Number of LVs	Sensitivity			Specificity			NER
			C1	C2	C3	C1	C2	C3	
Pasta 1	1 st derivative	3	0.82	0.95	0.96	0.97	0.16	0.65	0.91
	2 nd derivative	2	0.91	0.95	1.00	0.99	0.17	0.61	0.95
Pasta 2	1 st derivative	2	0.67	0.94	1.00	0.97	0.10	0.88	0.87
	2 nd derivative	2	0.73	0.89	0.96	0.99	0.11	0.82	0.86
Pasta 3	1 st derivative	2	0.71	0.95	1.00	0.98	0.18	0.73	0.89
	2 nd derivative	2	0.76	1.00	1.00	1.00	0.25	0.78	0.92
Pasta 4	1 st derivative	3	0.82	0.95	0.96	0.97	0.16	0.77	0.91
	2 nd derivative	2	0.91	0.95	1.00	0.97	0.17	0.58	0.95
Pasta 5	1 st derivative	4	1.00	0.91	1.00	0.99	0.18	0.55	0.97
	2 nd derivative	2	1.00	1.00	1.00	0.97	0.15	0.55	1.00
Pasta 6	1 st derivative	2	0.73	0.90	0.83	0.92	0.14	0.77	0.82
	2 nd derivative	2	0.73	0.90	0.98	0.92	0.16	0.77	0.87
Pasta 7	1 st derivative	2	0.88	0.76	0.91	0.89	0.23	0.67	0.85
	2 nd derivative	4	0.71	0.79	0.91	0.95	0.17	0.61	0.80

C1 = Low cooking time

C2 = Intermediate cooking time

C3 = High cooking time

4.4. Conclusion

In conclusion, this work showed that NIR-HSI in the transmission mode is a suitable technique as an objective alternative for the determination of cooking time of pasta as way of automating the determination of pasta attributes. Although some low values were found for one of the classes, in general, the models showed good ability in classifying pasta samples according to the cooking time. LDA models constructed based on PCA scores showed to be a good way of classifying samples with reduced dimensionality.

This work suggests hyperspectral image to have good potential as an objective method to cooking time determination, meeting this industry need. Although the spectrum range used in this work was small, it showed a great possibility to optimize food quality evaluation. Future studies should involve the use of such technique in a range that comprises more information about chemical changes that may occur during cooking.

Acknowledgment

This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) [Finance Code 001]; São Paulo Research Foundation (FAPESP) [grant numbers 2015/24351-2, 2017/17628-3, 2019/06842- 0]; and by projects AEI PID2019-107347RR-C31 and PID2019-107347RR-C32, and the European Union through the European Regional Development Fund (ERDF) of the Generalitat Valenciana 2014-2020. The authors would like to thank Nutrassim Food Ingredients company for the donation of the fibre samples, the support provided by Enrique Aguilar María, Carlos Alberto Velasquez Hernández, Diego Hernández Catalán, Carlos Ruiz Catalá and Andrés Estuardo Prieto López during system installation, experimental analysis and data acquisition.

Conflict of interest disclosure

The authors declare no competing financial interest.

References

- Aravind, N., Sissons, M., Egan, N., Fellows, C., 2012. Effect of insoluble dietary fibre addition on technological, sensory, and structural properties of durum wheat spaghetti. *Food Chem.* 130, 299–309. <https://doi.org/10.1016/j.foodchem.2011.07.042>
- Badaró, A.T., Amigo, J.M., Blasco, J., Aleixos, N., Ferreira, A.R., Clerici, M.T.P.S., Barbin, D.F., 2021. Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fiber distribution in enriched pasta. *Food Chem.* 343. <https://doi.org/10.1016/j.foodchem.2020.128517>
- Ballabio D., Consonni V., 2013. Classification tools in chemistry. Part 1: Linear models. *PLS-DA. Analytical methods*, 5, 3790-3798. <https://doi.org/10.1039/C3AY40582F>
- Bustos, M.C., Perez, G.T., Leon, A.E., 2015. Structure and quality of pasta enriched with functional ingredients. *RSC Adv.* 5, 30780–30792. <https://doi.org/10.1039/c4ra11857j>.

- Bustos, M.C., Perez, G.T., León, A.E., 2011. Sensory and nutritional attributes of fibre-enriched pasta. *LWT - Food Sci. Technol.* 44, 1429–1434.
<https://doi.org/10.1016/j.lwt.2011.02.002>
- Cunin, C., Handschin, S., Walther, P., Escher, F., 1995. Structural changes of starch during cooking of durum wheat pasta. *LWT - Food Sci. Technol.* 28, 323–328.
[https://doi.org/10.1016/S0023-6438\(95\)94552-0](https://doi.org/10.1016/S0023-6438(95)94552-0)
- Del Nobile, M.A., Baiano, A., Conte, A., Mocci, G., 2005. Influence of protein content on spaghetti cooking quality. *J. Cereal Sci.* 41, 347–356.
<https://doi.org/10.1016/j.jcs.2004.12.003>
- Erkinbaev, C., Henderson, K., Paliwal, J., 2017. Discrimination of gluten-free oats from contaminants using near infrared hyperspectral imaging technique. *Food Control* 80, 197–203. <https://doi.org/10.1016/j.foodcont.2017.04.036>
- Esteki, M., Shahsavari, Z., Simal-Gandara, J., 2018. Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. *Food Control*. <https://doi.org/10.1016/j.foodcont.2018.03.031>
- Majzoobi, M., Ostovan, R., Farahnaky, A., 2011. Effects of hydroxypropyl cellulose on the quality of wheat flour spaghetti. *J. Texture Stud.* 42, 20–30.
<https://doi.org/10.1111/j.1745-4603.2010.00264.x>
- Menesatti, P., Antonucci, F., Pallottino, F., Bucarelli, F.M., Costa, C., 2014. Spectrophotometric Qualification of Italian Pasta Produced by Traditional or Industrial Production Parameters. *Food Bioprocess Technol.* 7, 1364–1370.
<https://doi.org/10.1007/s11947-013-1138-0>
- Rachmawati, Rohaeti, E., Rafi, M., 2017. Combination of near infrared spectroscopy and

chemometrics for authentication of taro flour from wheat and sago flour, in: *Journal of Physics: Conference Series*. <https://doi.org/10.1088/1742-6596/835/1/012011>

Rakhesh, N., Fellows, C.M., Sissons, M., 2015. Evaluation of the technological and sensory properties of durum wheat spaghetti enriched with different dietary fibres. *J. Sci. Food Agric.* 95, 2–11. <https://doi.org/10.1002/jsfa.6723>.

Su, W.-H., Sun, D.-W., 2017. Evaluation of spectral imaging for inspection of adulterants in terms of common wheat flour, cassava flour and corn flour in organic Avatar wheat (*Triticum spp.*) flour. *J. Food Eng.* 200, 59–69. <https://doi.org/10.1016/j.jfoodeng.2016.12.014>

Wold, J.P., Kermit, M., Segtnan, V.H., 2016. Chemical Imaging of Heterogeneous Muscle Foods Using Near-Infrared Hyperspectral Imaging in Transmission Mode. *Appl. Spectrosc.* 70, 953–961. <https://doi.org/10.1177/0003702816641260>

Xiao, H., Li, A., Li, M., Sun, Y., Tu, K., Wang, S., Pan, L., 2018. Quality assessment and discrimination of intact white and red grapes from *Vitis vinifera* L. at five ripening stages by visible and near-infrared spectroscopy. *Sci. Hortic. (Amsterdam)*. 233, 99–107. <https://doi.org/10.1016/j.scienta.2018.01.041>

Ziegler, J.U., Leitenberger, M., Longin, C.F.H., Würschum, T., Carle, R., Schweiggert, R.M., 2016. Near-infrared reflectance spectroscopy for the rapid discrimination of kernels and flours of different wheat species. *J. Food Compos. Anal.* 51, 30–36. <https://doi.org/10.1016/j.jfca.2016.06.005>

CHAPTER 5

Determination of pectin content in orange peels by near infrared hyperspectral imaging



Determination of pectin content in orange peels by near infrared hyperspectral imaging



Amanda Teixeira Badaró^a, Juan Francisco Garcia-Martin^b, María del Carmen López-Barrera^b, Douglas Fernandes Barbin^{a,*}, Paloma Alvarez-Mateos^b

^a Department of Food Engineering, University of Campinas (UNICAMP), Rua Monteiro Lobato, 80, Cidade Universitária, Campinas-SP 13083-862, Brazil
^b Departamento de Ingeniería Química, Facultad de Química, Universidad de Sevilla, Sevilla 41012, Spain

ARTICLE INFO

Keywords:

Principal component analysis
Linear discriminant analysis
Partial least squares regression, near infrared spectra
Agriculture

ABSTRACT

Pectin has several purposes in the food and pharmaceutical industry making its quantification important for further extraction. Current techniques for pectin quantification require its extraction using chemicals and producing residues. Determination of pectin content in orange peels was investigated using near infrared hyperspectral imaging (NIR-HSI). Hyperspectral images from orange peel (140 samples) with different amounts of pectin were acquired in the range of 900–2500 nm, and the spectra was used for calibration models using multivariate statistical analyses. Principal component analysis (PCA) and linear discriminant analysis (LDA) showed better results considering three groups: low (0–5%), intermediate (10–40%) and high (50–100%) pectin content. Partial least squares regression (PLSR) models based on full spectra showed higher precision ($R^2 > 0.93$) than those based on few selected wavelengths (R^2 between 0.92 and 0.94). The results demonstrate the potential of NIR-HSI to quantify pectin content in orange peels, providing a valuable technique for orange producers and processing industries.

1. Introduction

Citrus are among the most cultivated crops around the world since they are beneficial for human consumption due to their nutritional and antioxidant properties. As a result, the citrus industry is responsible for millions of jobs (García-Martín, Olmo, & García, 2018; García, Olmo, & García, 2016). Spain is the first producing country of oranges and juices in the European Union and the fifth in the world. This agricultural activity is mainly located in Valencia, producing 3 million tons of citrus per year, of which, according to the Spanish Ministry of Agriculture, Fishing and Food (2017) (Ministerio de Agricultura, 2019), 1.5 million are oranges. The orange juice sector generates about 1.2 million tons waste per year in Spain. This waste is mainly used for animal feed (which is energy costly) or is ploughed into landfills. However, this organic waste has great potential for the food industry as source of many important and valuable compounds (Lessa, Gularte, Garcia, & Fajardo, 2017).

Orange waste can include peel, pulp, seeds, leaves and oranges without the quality requirements. This waste contains, as average, 16.9% (wt.) soluble sugars, 9.21% (wt.) cellulose, 10.5% (wt.)

hemicelluloses, and 42.5% (wt.) pectin, which is its most important component. The soluble sugars present in orange peel are glucose, fructose, and sucrose. The insoluble polysaccharides of the cellular wall of the orange peel are composed of pectin, cellulose, and hemicelluloses. Pectin and hemicelluloses are rich in galacturonic acid, arabinose, and galactose, and contain small amounts of xylose, rhamnose and glucose (Grohmann, Cameron, & Buslig, 1995; Rezzadori, Benedetti, & Amante, 2012). Pectin is composed of a main chain, which has a linear structure of α -D-galacturonic acids linked by α -(1,4)-glycosidic bonds, and a side chain that mainly contains neutral sugars (Liu, Fishman, Kost, & Hicks, 2003).

Pectin has several purposes in the food and pharmaceutical industry, as it is used as thickener, stabilizer, gelling agent, emulsifier and drug delivery vehicle (Wicker, Kim, Thirkield, Lin & Jung, 2014). Hence, the extraction of pectin from orange waste is very important and requires an economically feasible alternative. Among the different pectin extraction methods (dialysis, metal precipitation, membrane separation, etc.), alcohol precipitation is the most widely used methods despite its high energy and ethanol requirements (Ren et al., 2019). However, it is also important to quantify the amount of pectin in orange

* Corresponding author.

E-mail addresses: amandatbadaro@gmail.com (A.T. Badaró), jfgarmar@us.es (J.F. Garcia-Martin), mlopez91@us.es (M.d.C. López-Barrera), dfbarbin@unicamp.br (D.F. Barbin), palvarez@us.es (P. Alvarez-Mateos).

<https://doi.org/10.1016/j.foodchem.2020.126861>

Received 20 December 2019; Received in revised form 3 April 2020; Accepted 17 April 2020

Available online 18 April 2020

0308-8146/ © 2020 Elsevier Ltd. All rights reserved.

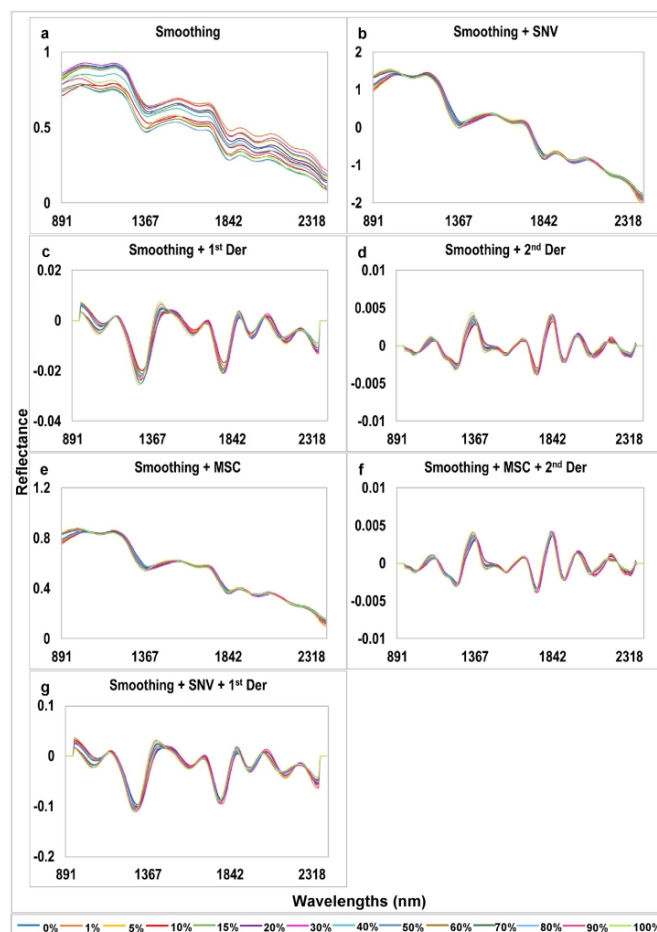


Fig. 1. Spectra of orange peels and pectin: a) Smoothing, b) Smoothing + SNV, c) Smoothing + 1st derivative, d) Smoothing + 2nd derivative, e) Smoothing + MSC, f) Smoothing + MSC + 2nd derivative, g) Smoothing + SNV + 1st derivative. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

residue, in order to justify investments in the waste processing and extraction methods. The current techniques for pectin quantification requires the previous pectin extraction. In this sense, the commonly-used sulphuric acid-carbazole colorimetric method for pectin content measuring in citrus involves pectin extraction, pectin depolymerisation, reaction of pectin with carbazole and absorbance measurement at approximately 525 nm (Wang, Chuang, & Hsu, 2008).

Some works have been reported using vibrational techniques, more specifically near infrared spectroscopy and Fourier transform near infrared spectroscopy, to determine some parameters in orange, including orange firmness, peel thickness and total pectin content (Bizzani, Flores, Colnago, & Ferreira, 2017; Chen et al., 2018). Although this techniques have achieved good results for many parameters, the fact that it only detects a single point of the sample at a time can be a barrier

for complex samples (Feng & Sun, 2012).

In this context, near infrared hyperspectral imaging (NIR-HSI), as an emerging technology, offers numerous advantages over conventional analysis methods and can overcome some problems faced by near infrared spectroscopy. Combined to chemometrics, NIR-HSI has achieved very precise results for many compounds and properties in food, as colour and pH in meat (Elmasry, Sun, & Allen, 2012), fibre in flour (Badaró, Morimitsu, Ferreira, Clerici, & Barbin, 2019) and melamine in milk powder (Forchetti & Poppi, 2017). As the hypercube data are massive, chemometrics is used to reduce the high dimensionality to the most meaningful dimension (data simplification) without compromising the information contained into the original image (Feng, Makino, Oshita, & Martín, 2018). Hence, HSI could provide a distribution map of pectin in the orange waste, as other authors have

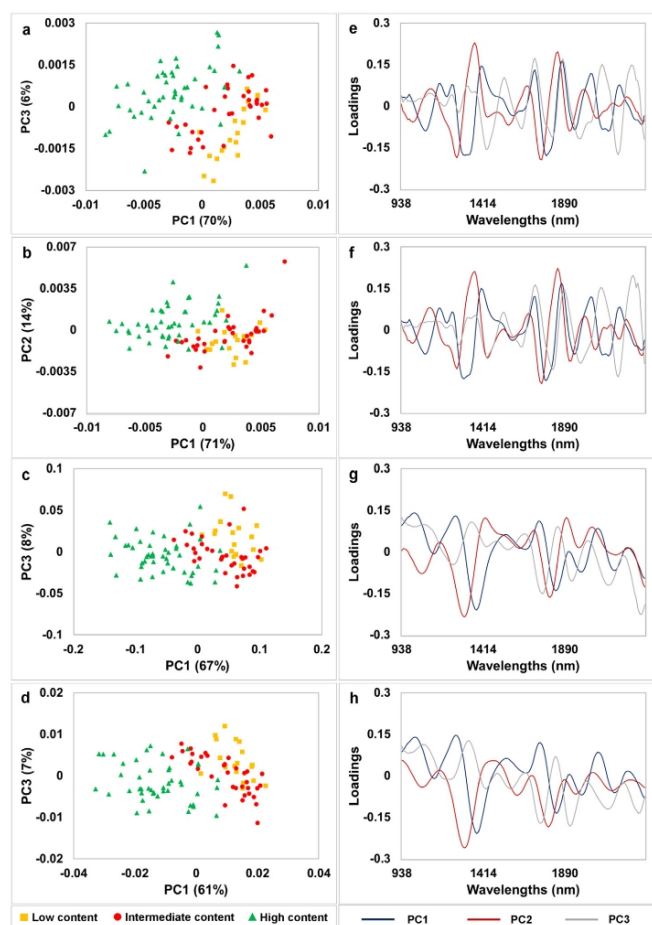


Fig. 2. PCA scores and loadings of orange peels and pectin: (a and e) Smoothing + 2nd derivative, (b and f) Smoothing + MSC + 2nd derivative, (c and g) Smoothing + SNV + 1st derivative, (d and h) Smoothing + 1st derivative. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

developed in other fruit such as peach (Zhu, Huang, Wu, Chen, & He, 2017). For quantitative assessment, partial least squares regression (PLSR), a type of linear analysis, is widely used to establish the correlation between spectral data of the hypercubes and real quantities or concentrations of the compound measured by ordinary laboratory assessments (Feng et al., 2018). NIR-HSI coupled with PLSR could be used in a preliminary analysis of orange waste to determine its constituents (such as pectin) in a quick way, to further decide whether to extract them or not based on the concentrations determined, being an alternative to improve extraction yield.

Thus, the hypothesis of this work is that NIR-HSI has a potential application for identification of pectin content in orange peels. This hypothesis will be verified by identification of the wavelengths that most contribute in the development of models with good precision and accuracy (defined as optimum wavelengths) and development of a

prediction model to quantify pectin content in orange peels.

2. Material and methods

2.1. Pectin extraction from raw material

Lanelate oranges harvested from January to July of 2019 were supplied by Export Orange S.L. The oranges were ground (1 cm × 1 cm × 1 cm) and the juice was separated from the peel using a strainer. The peel was washed with water at 100 °C, and washed again with water at 50 °C. This step was carried out to remove most of sugars. Then, the peel was strained and oven dried at 45 °C for 2 days. After, the peel was milled using a milling machine and sieved with an ASTM N°50 mesh (300 µm diameter). A sample of 5 g was dispersed in water at a solid-liquid ratio of 1:40 (w/v) and the pH of the mixture was

A.T. Badaró, et al.

Food Chemistry 323 (2020) 126861

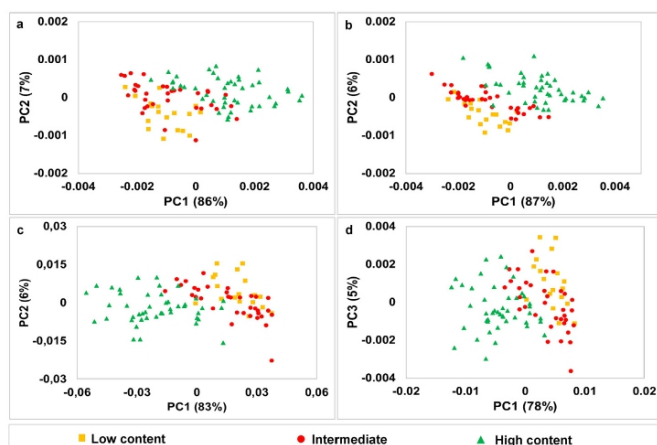


Fig. 3. PCA scores of orange peels and pectin after variable selection: a) Smoothing + 2nd derivative, b) Smoothing + MSC + 2nd derivative, c) Smoothing + SNV + 1st derivative, d) Smoothing + 1st derivative. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

LDA models performance for qualitative analysis of pectin in orange peels.

Pre-treatment	Wavelengths (nm)	Sensitivity (Validation)			Specificity (Validation)			Accuracy - Calibration Model (%)
		C1	C2	C3	C1	C2	C3	
Smoothing	1100, 1290, 1442, 1775, 1918, 2032, 2118, 2289	0.58	0.67	1.00	0.83	0.81	1.00	83.51
Smoothing + SNV	1318, 1442, 1775, 1871, 2032, 2213	0.67	0.40	0.93	0.70	0.81	1.00	84.54
Smoothing + 1st Derivative	1014, 1252, 1376, 1547, 1747, 1842, 1918, 1985, 2080, 2185, 2251	0.67	0.67	0.93	0.83	0.81	1.00	92.78
Smoothing + 2nd Derivative	1071, 1166, 1242, 1318, 1404, 1595, 1718, 1785, 1871, 1947, 2042, 2147, 2213	0.58	0.73	1.00	0.87	0.81	1.00	90.72
Smoothing + MSC	1318, 1452, 1775, 1871, 2051, 2213	0.67	0.47	0.93	0.73	0.81	1.00	85.57
Smoothing + MSC + 2nd Derivative	1081, 1166, 1233, 1309, 1404, 1595, 1718, 1785, 1871, 1947, 2042, 2137, 2213	0.58	0.67	1.00	0.83	0.81	1.00	88.66
Smoothing + SNV + 1st Derivative	1014, 1252, 1376, 1547, 1747, 1842, 1918, 1985, 2080, 2185, 2251	0.67	0.60	0.93	0.80	0.81	1.00	91.75

C1: low pectin content (0–5%); C2: intermediate pectin content (10–40%); C3: high pectin content (50–100%).

Table 2

PLSR model performance for quantitative analysis of pectin in orange peels.

Pre-treatment	Wavelengths (nm)	LV	R^2_{cv}	RMSECV	R^2_p	RMSEP	SEP	RPD	RER
Smoothing	Full spectra	6	0.93	8.54	0.93	9.16	8.27	4.11	12.09
	1271, 1366, 1452, 1690, 1766, 1861, 1947, 2261, 2365	6	0.93	8.61	0.93	8.84	8.04	4.22	12.44
Smoothing + SNV	Full spectra	6	0.95	7.43	0.94	7.99	7.15	4.75	13.98
	1271, 1328, 1452, 1690, 1766, 1871, 1947, 2261, 2346	5	0.95	7.57	0.94	8.42	7.59	4.47	13.17
Smoothing + 1st Derivative	Full spectra	3	0.93	8.71	0.95	7.94	7.66	4.43	13.06
	1223, 1395, 1652, 1728, 1813, 1899, 2166, 2318	2	0.94	8.37	0.92	9.73	9.20	3.69	10.87
Smoothing + 2nd Derivative	Full spectra	4	0.95	7.42	0.96	6.50	6.32	5.37	15.81
	1062, 1214, 1318, 1433, 1690, 1861, 2051, 2118, 2204, 2270	2	0.94	8.15	0.94	8.03	7.33	4.63	13.64
Smoothing + MSC	Full spectra	4	0.96	6.68	0.94	8.23	7.88	4.30	12.68
	1328, 1452, 1690, 1766, 1871, 1937, 2194, 2237	4	0.95	7.60	0.94	8.03	7.70	4.41	12.99
Smoothing + MSC + 2nd Derivative	Full spectra	4	0.95	7.30	0.96	6.94	6.67	5.09	14.99
	1033, 1318, 1423, 1690, 1861, 2118, 2204, 2270	2	0.94	7.85	0.94	8.32	7.85	4.32	12.74
Smoothing + SNV + 1st Derivative	Full spectra	5	0.95	7.13	0.95	7.96	7.51	4.52	13.32
	1290, 1395, 1499, 1652, 1728, 1899, 2166, 2299	3	0.95	7.54	0.94	8.44	7.90	4.29	12.66

adjusted to 1.5 with 0.1 M HCl. Subsequently, the mixture was placed in a water bath between 70 and 85 °C at 300 rpm for 70 min. After, the mixture was centrifuged at 5000 rpm for 20 min. The supernatant (solution of pectin extract) was removed and the liquid residue was oven dried at 45 °C for 2 days. Then, the dried residue was milled and sieved with the ASTM N°50 mesh.

2.2. Sample preparation and image acquisition

Standard pectin (Apple pectin powder – 100% purity, Solgar, Inc. Leonia, USA) was added to the dried residue of orange peel without pectin in the following percentages: 0, 1, 5, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 and 100% (being 0% residue of orange peel without pectin addition, and 100% pure pectin without residues of orange peel). Ten

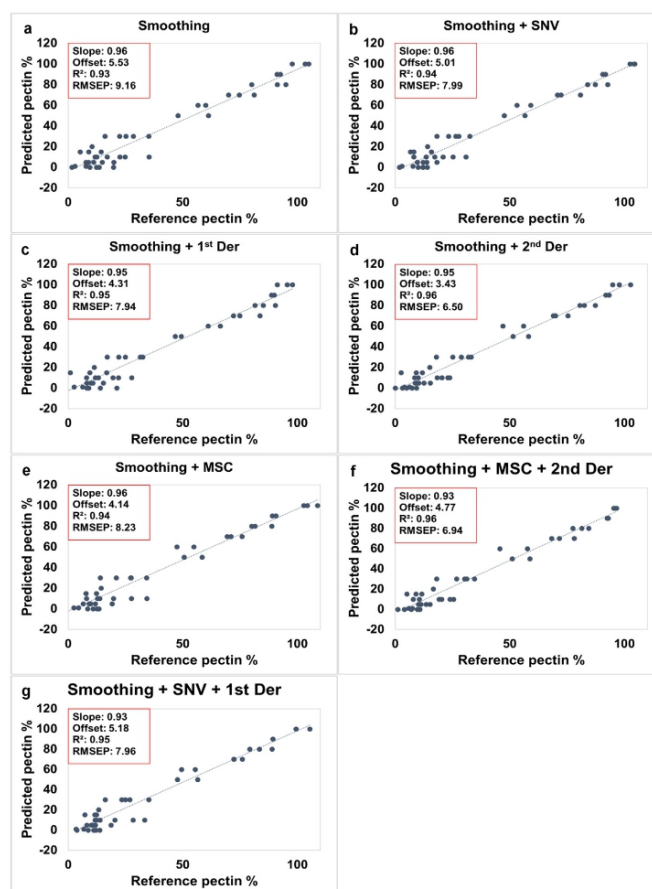


Fig. 4. Predicted vs. reference plot for the PLSR models obtained with full spectra: a) Smoothing, b) Smoothing + SNV, c) Smoothing + 1st derivative, d) Smoothing + 2nd derivative, e) Smoothing + MSC, f) Smoothing + MSC + 2nd derivative, g) Smoothing + SNV + 1st derivative.

samples of 5 g of each percentage were prepared, totaling 140 samples. Samples were placed in Petri dishes of 3 cm of diameter for image acquisition.

NIR hyperspectral images were acquired from each sample in the reflectance mode using a SWIR camera (Headwall Photonics SWIR M series, Massachusetts, USA), in the range of 900–2500 nm, with an illuminator of 75 W and a scanning speed of 14.7 mm/s. The program automatically subtracted the white (~99% reflectance) and dark (0% reflectance) references from subsequently acquired images.

2.3. Spectra extraction and multivariate analysis

Image segmentation and spectrum extraction were performed using a code developed by the research group using the open software Python (version 3.7.0; Python Software Foundation License). The averaged reflectance spectra were smoothed (Savitzky-Golay) and mean centred prior to multivariate data analysis. Data were pre-treated with Standard

Normal Variate (SNV), Multiplicative Scatter Correction (MSC), first derivative (1st Der) (Savitsky-Golay smoothing, 11 points window, 2nd order polynomial), second derivative (2nd Der) (Savitsky-Golay smoothing, 11 points window, 2nd order polynomial) and a combination of MSC + 2nd der, and SNV + 1st der. After that, qualitative and quantitative analysis such as Principal Component Analysis, Linear Discriminant Analysis and Partial Least Square Regression were carried out. The multivariate data analysis was performed using The Unscrambler X 10.4 software.

2.4. Principal components analysis (PCA)

PCA was performed as exploratory data analysis, in order to obtain an overview of the variation among samples, identify clusters and outliers. This step is applied to reduce the spectral information into principal components (PCs), which are a linear combination of the variables in the spectra data and contain most of the relevant

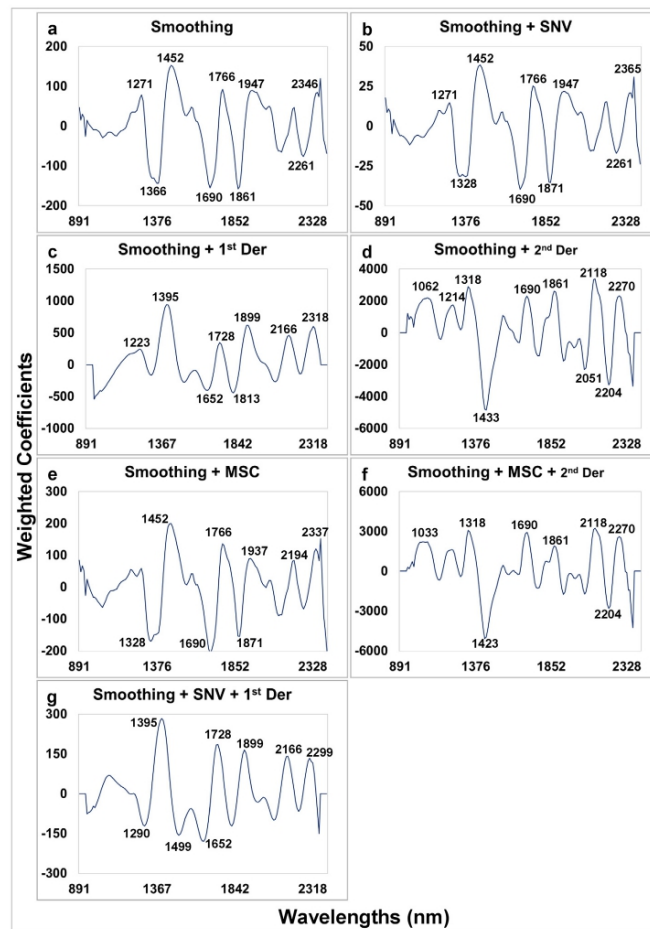


Fig. 5. Regression coefficients for the PLSR models obtained with full spectra: a) Smoothing, b) Smoothing + SNV, c) Smoothing + 1st derivative, d) Smoothing + 2nd derivative, e) Smoothing + MSC, f) Smoothing + MSC + 2nd derivative, g) Smoothing + SNV + 1st derivative.

information (Rachmawati, Rohaeti, & Rafi, 2017). Moreover, through PCA scores, data from both sets of experiments were split into calibration and validation sets (independent sets), using Kennard-Stone algorithm (Kennard & Stone, 1969), in a ratio of 70 and 30%, respectively. In addition, the loadings plot were used to manually select the wavelengths that contained the relevant information about the sample, defined as optimum wavelengths. These optimum wavelengths were chosen based on the most prominent peaks and valleys for further qualification quantification analysis (LDA).

2.5. Linear discriminant analysis (LDA)

Linear discriminant analysis (LDA) is a classification method used to differentiate samples into different classes. As a particularity of LDA method, the number of variables must be smaller than the number of

samples (Manley & Baeten, 2018). Therefore, optimum wavelengths selected in the loadings plot of PCA were used as the data input to perform LDA analysis in all pre-treated data. LDA was performed with a leave-one-out cross-validation method and an external validation was carried out with the independent dataset. Models performance was measured in terms of sensitivity, specificity (of validation set) and accuracy of calibration model. Values of sensitivity and selectivity close or equal to 1.00 and accuracy of 100% show good discriminative power.

2.6. Partial least Square regression (PLSR)

Partial Least Square Regressions (PLSR) was performed to test the ability of NIR-HSI data to quantify the different percentages of pectin added to the residues of orange peel. First, calibration models were developed using full spectra, and, then, the weighted regression

coefficients (B_w) resulting from these PLSR models were used for the development of reduced models. The wavelengths with the highest values (regardless of sign) were selected as optimum wavelengths. All the models were developed based on a full cross validation (leave-one out), and the models performance was evaluated in terms of coefficient of determination and root mean square error of cross-validation (R_{cv}^2 , RMSECV) and prediction (R_p^2 , RMSEP), the ratio performance to deviation (RPD) and the ratio of error range (RER), both based on the standard error of validation (SEP). It is expected coefficients of determination close to 1, errors close to 0, RPD and RER above 3 and 10, respectively (Malley, Yesmin, & Eilers, 2002).

3. Results and discussion

3.1. Spectral analysis

The average HSI reflectance spectra of orange peel and pectin in each percentage are presented in Fig. 1. The smoothed spectra (Fig. 1a) showed very similar shape for all samples, only differing in the intensity of reflectance across the spectral region. Overall, samples with lower amount of pectin had higher reflectance. This baseline difference was minimized after spectra pre-treatment with SNV and MSC (Fig. 1b and 1e, respectively). Peaks were observed around 1214 and 1728 nm, which correspond to stretching vibrations of C–H in the first and second overtone. A valley at 1306 nm indicated stretching vibration of C–H or a combination of deformation. The peak around 1918 nm is related to the stretching vibration of C=O in the second overtone of amides. At 2100 nm, there was a peak which denotes a combination of O–H deformation and C–O stretching vibration, associated to starch.

After 2nd derivatives and the combination of pre-treatments (MSC + 2nd der), other peaks and valleys were highlighted along the spectra (Fig. 1d and 1g). Most of these denote stretching vibrations of C–H in the first and second overtones or combination of deformation. In addition, the 2nd derivative highlighted in the spectra a peak around 2000 nm, which denote a combination of vibration of O–H and C–O deformation in starch, and a valley at 2080 nm corresponding to a combination of vibrations of O–H stretching and O–H deformation, associated to sucrose and starch (Osborn, Fearn, & Hindle, 1993).

3.2. Principal components analysis (PCA)

According to the literature, the performance of classification models based on multi-classes seems to be worse than classifications based on group of classes (Ziegler et al., 2016). Therefore, in order to achieve a better performance for models with different percentages of pectin, three classes were created and defined as “Low content” (samples with 0–5% of pectin), “Intermediate content” (samples with 10–40% of pectin) and “High content” (samples with 50–100% of pectin), based on the average pectin content in orange peels reported by other authors (El-Nawawi & Shehata, 1987; Rouse & Crandall, 1976), which is roughly 30% (wt.) on a dry matter basis. Then, principal component analysis was performed on reflectance spectra with all pre-treatments based on these classes. Those pre-treatments that better separated the groups of samples are displayed in Fig. 2. Fig. 2a, 2c and 2d show the scores plots of PC1 and PC3 for spectra pre-treated with 2nd derivative, a combination of SNV and 1st derivative, and 1st derivative only, respectively; whereas, Fig. 2b shows the scores plot of the first two principal components for spectra pre-treated with a combination of MSC and 2nd derivative. Overall, samples with high content of pectin are spread in the negative side of PC1, while samples with a low content of pectin are in the positive side. Accordingly, samples with an intermediate content of pectin are disposed between the other two groups.

The first three principal components explained over 75% of the variance among samples, and the loadings plot of these three PCs are represented in Fig. 2(e-h). Most of the peaks and valleys observed in these plots agree with the respective spectra, especially those associated

to cellulose, starch and sucrose. Consequently, these loadings plot were used in order to identify the wavelengths that most contributed to separate the groups of samples. Then, these optimum wavelengths were selected, and a new PCA was recalculated for each of the pre-treatments (Fig. 3 a-d). Fig. 3a and 3b showed that over 90% of the variance between samples were explained by the first two PCs, with samples of low pectin content displayed in the left side of PC1 and negative part of PC2, and samples of high pectin content in the right side of PC1 and positive part of PC2. Samples of intermediate pectin content remained in between the other two classes of samples. After variable selection, data pre-treated with SNV + 1st der and only 1st derivative showed similar behaviour on PCA plot as with full spectra, with PC1 and PC2 explaining 89% of the variance among samples.

3.3. Linear discriminant analysis (LDA)

The selected wavelengths from PCA loadings were used as predictors for LDA models, in order to evaluate the pre-treatments that better discriminated samples into their respective classes. Table 1 shows the models performance of LDA for qualitative analysis of pectin in orange peels. As in PCA, samples containing different percentages of pectin were grouped into three classes defined as “Low content” (samples with 0–5% of pectin), “Intermediate content” (samples with 10–40% of pectin) and “High content” (samples with 50–100% of pectin). These classes were used to develop the classification models of samples added of pectin. Data pre-treated with 1st derivative and a combination of SNV + 1st derivative provided the calibration models with the best accuracy (92.78 and 91.75%, respectively), and high values of sensitivity and specificity (0.60–1.00), thus demonstrating good capacity of the models to discriminate samples of residues of orange peel with different pectin content. Calibration models developed with pre-treated data with 2nd derivative and a combination of MSC + 2nd derivative had the second and third best accuracy (90.72 and 88.66%), respectively. However, the sensitivity and selectivity of these models were lower than those, ranging from 0.58 to 1.00. Data that was only smoothed, and SNV and MSC pre-treatments showed a suitable accuracy for calibration models (83.51, 84.54 and 85.57%, respectively). However, they did not improve the model discriminant power, since sensitivity and specificity of some classes were lower than 0.50. In addition, Class 3 had less misclassified samples when compared to Classes 1 and 2. This was also observed in the PCA score plots (Fig. 2 e-h), where, visibly, the distance between Classes 1 and 2 were closer than Class 3, so there was an overlap of samples from those classes.

3.4. Partial least square regression (PLSR)

The PLSR models performances for pectin content in orange peels are presented on Table 2. The predicted values of each PLSR model based on full spectra are displayed in Fig. 4. Overall, these models showed good precision, with high values of coefficient of determination (over 0.93), and errors slightly lower than those obtained by models based on optimum wavelengths, for both cross-validation and external validation.

The development of representative and accurate models with few wavelengths is more functional and interesting, especially for industrial applications. The weighted regression coefficients (B_w) provide information about the model quality, and those with large absolute value can be used as optimum wavelengths to develop reduced models (Kamruzzaman, Makino, Oshita, & Liu, 2015). Therefore, the important wavelengths were selected based on these weighted regression coefficients obtained by the full PLSR models, and the reduced models were developed. The results obtained by the reduced models also showed good coefficient of determination, which ranged from 0.92 to 0.94. Additionally, the errors of most models were slightly higher than those with full spectra, but still very close. Moreover, all the models based on full spectra or reduced models had RPD > 3 and RER > 10, which

A.T. Badaró, et al.

Food Chemistry 323 (2020) 126861

showed a good calibration performance of the models.

The weighted regression coefficients carry a lot of information regarding the variables in the dataset. While those of high absolute values have a large contribution in the model, those of small values have a little contribution in the model (Elmasry et al., 2012). The weighted regression coefficients of the PLSR models developed in this work (Fig. 5) showed strong peaks and valleys, which are positive and negatively related to the important variables for pectin quantification. Most of these peaks and valleys are the same to those observed in the pectin spectra, as show in Fig. 1. In addition, they had spectral characteristics and did not show a lot of noise, suggesting that the models can be considered reliable and this technique can be considered for further application.

4. Conclusion

Hyperspectral imaging technique showed a great potential to classify orange peel samples according to the pectin content. As PCA, LDA was able to separate the groups of samples containing different percentages of pectin into the same three classes (low, intermediate and high pectin content). Among the different combinations of data pre-treatments assayed, 1st derivative and a combination of SNV + 1st derivative achieved the best accuracy (92.78 and 91.75%, respectively), and the highest sensitivity and specificity (between 0.6 and 1.0). Additionally, PLSR models for pectin content quantification based on full spectra showed excellent precision, providing high coefficients of determination (over 0.93), whereas those of the PLSR models built with the most contributing wavelengths ranged between 0.92 and 0.94.

Hence, the study confirms the hypothesis that NIR-HSI can be used for quantification of pectin content in orange peels. The results showed that this technique holds potential as an alternative to the carbazole colorimetry method to quantify pectin in orange peels, and to categorize orange peel samples into groups of different pectin concentration, and can be used to justify investments in the waste processing and extraction methods.

CRediT authorship contribution statement

Amanda Teixeira Badaró: Writing - original draft, Methodology, Investigation, Validation, Software. **Juan Francisco Garcia Martin:** Conceptualization, Writing - original draft, Writing - review & editing, Investigation, Visualization, Funding acquisition. **María del Carmen López Barrera:** Formal analysis, Writing - review & editing, Software, Visualization. **Douglas Fernandes Barbin:** Conceptualization, Formal analysis, Supervision, Software, Writing - original draft, Writing - review & editing, Project administration, Funding acquisition. **Paloma Alvarez Mateos:** Conceptualization, Resources, Supervision, Methodology, Writing - review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Dr. Douglas Fernandes Barbin would like to acknowledge the University of Seville for the mobility grant (VIPIT-2019-1.3) awarded under the "VI Plan Propio de Investigación y Transferencia" of the University of Seville, and São Paulo Research Foundation (FAPESP), Young Researchers Award (Grant no. 2015/24351-2). Amanda Teixeira Badaró would like to thank São Paulo Research Foundation (FAPESP) for the financial support (project numbers 2017/17628-3 and 2019/

06842-0). This work was in part financed by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

- Badaró, A. T., Morimitsu, F. L., Ferreira, A. R., Clerici, M. T. P. S., & Barbin, D. F. (2019). Identification of fiber added to semolina by near infrared (NIR) spectral techniques. *Food Chemistry*, 289, 195–203. <https://doi.org/10.1016/j.foodchem.2019.03.057>.
- Bizzani, M., Flores, D. W. M., Colnago, L. A., & Ferreira, M. D. (2017). Non-invasive spectroscopic methods to estimate orange firmness, peel thickness, and total pectin content. *Microchemical Journal*, 133, 168–174. <https://doi.org/10.1016/j.micro.2017.03.039>.
- Chen, H., Xu, L., Jia, Z., Cai, K., Shi, K., & Gu, J. (2018). Determination of parameter uncertainty for quantitative analysis of shaddock peel pectin using linear and non-linear near-infrared spectroscopic models. *Analytical Letters*, 51(10), 1564–1577. <https://doi.org/10.1080/00032719.2017.1384479>.
- El-Nawawi, S. A., & Shehata, F. R. (1987). Extraction of pectin from Egyptian orange peel. Factors affecting the extraction. *Biological Wastes*, 20(4), 281–290. [https://doi.org/10.1016/0269-7483\(87\)90005-X](https://doi.org/10.1016/0269-7483(87)90005-X).
- Elmasry, G., Sun, D. W., & Allen, P. (2012). Near-infrared hyperspectral imaging for predicting colour, pH and tenderness of fresh beef. *Journal of Food Engineering*, 110(1), 127–140. <https://doi.org/10.1016/j.jfoodeng.2011.11.028>.
- Feng, C. H., Makino, Y., Oshita, S., & Martín, J. F. G. (2018). Hyperspectral imaging and multispectral imaging as the novel techniques for detecting defects in raw and processed meat products: Current state-of-the-art research advances. *Food Control*, 84, 165–176. <https://doi.org/10.1016/j.foodcont.2017.07.013>.
- Feng, Y.-Z., & Sun, D.-W. (2012). Application of hyperspectral imaging in food safety inspection and control: A review. *Critical Reviews in Food Science and Nutrition*, 52(11), 1039–1058. <https://doi.org/10.1080/10408398.2011.651542>.
- Forchetti, D. A. P., & Poppi, R. J. (2017). Use of NIR hyperspectral imaging and multivariate curve resolution (MCR) for detection and quantification of adulterants in milk powder. *LWT - Food Science and Technology*, 76, 337–343. <https://doi.org/10.1016/j.lwt.2016.06.046>.
- García-Martín, J. F., Olmo, M., & García, J. M. (2018). Effect of ozone treatment on postharvest disease and quality of different citrus varieties at laboratory and at industrial facility. *Postharvest Biology and Technology*, 137, 77–85. <https://doi.org/10.1016/j.postharvbio.2017.11.015>.
- García, J. F., Olmo, M., & García, J. M. (2016). Decay incidence and quality of different citrus varieties after postharvest heat treatment at laboratory and industrial scale. *Postharvest Biology and Technology*, 118, 96–102. <https://doi.org/10.1016/j.postharvbio.2016.03.019>.
- Grohmann, K., Cameron, R. G., & Buslig, B. S. (1995). Fractionation and pretreatment of orange peel by dilute acid hydrolysis. *Bioresource Technology*, 54(2), 129–141. [https://doi.org/10.1016/0960-8524\(95\)00121-2](https://doi.org/10.1016/0960-8524(95)00121-2).
- Kamruzzaman, M., Makino, Y., Oshita, S., & Liu, S. (2015). Assessment of visible near-infrared hyperspectral imaging as a tool for detection of horsemeat adulteration in minced beef. *Food and Bioprocess Technology*, 8(5), 1054–1062. <https://doi.org/10.1007/s11947-015-1470-7>.
- Kennard, R. W., & Stone, L. A. (1969). Computer-aided design of experiments. *Technometrics*, 11(1), 137–148. <https://doi.org/10.1080/00401706.1969.10490666>.
- Lessa, E. F., Gularte, M. S., Garcia, E. S., & Fajardo, A. R. (2017). Orange waste: A valuable carbohydrate source for the development of beads with enhanced adsorption properties for cationic dyes. *Carbohydrate Polymers*, 157, 660–668. <https://doi.org/10.1016/j.carbpol.2016.10.019>.
- Liu, L. S., Fishman, M. L., Kost, J., & Hicks, K. B. (2003). Pectin-based systems for colon-specific drug delivery via oral route. *Biomaterials*, 24(19), 3333–3343. [https://doi.org/10.1016/S0142-9612\(03\)00213-8](https://doi.org/10.1016/S0142-9612(03)00213-8).
- Malley, D. F., Yesmin, L., & Eilers, R. G. (2002). Rapid analysis of hog manure and manure-amended soils using near-infrared spectroscopy. *Soil Science Society of America Journal*, 66, 1677–1686. <https://doi.org/10.2136/sssaj2002.1677>.
- Manley, M., & Baeten, V. (2018). Spectroscopic technique: near infrared (NIR) spectroscopy. In D.-W. Sun (Ed.), *Modern techniques for food authentication* (pp. 51–102). (2nd ed.). Academic Press.
- Ministerio de Agricultura. Superficies y producciones anuales de cultivos. (2017). <https://www.mapa.gov.es/estadistica/temas/estadisticas-agricultura/superficies-producciones-anuales-cultivos/> Accessed 28 November 2019.
- Osborn, B. G., Fearn, T., & Hindle, P. H. (1993). *Theory of near infrared spectroscopy. Practical NIR spectroscopy with applications in food and beverage analysis* (pp. 13–35). (2nd ed.). London: Longman Singapore Publishers (Pte) Ltd.
- Rachmawati, Rohaeti, E., & Rafi, M. (2017). Combination of near infrared spectroscopy and chemometrics for authentication of taro flour from wheat and sago flour. *Journal of Physics: Conference Series*, 835. <https://doi.org/10.1088/1742-6596/835/1/012011>.
- Ren, J. N., Hou, Y. Y., Fan, G., Zhang, L. L., Li, X., Yin, K., & Pan, S. Y. (2019). Extraction of orange pectin based on the interaction between sodium caseinate and pectin. *Food Chemistry*, 283, 265–274. <https://doi.org/10.1016/j.foodchem.2019.01.046>.
- Rezzadori, K., Benedetti, S., & Amante, E. R. (2012). Proposals for the residues recovery: Orange waste as raw material for new products. *Food and Bioprocess Processing*, 90(4), 606–614. <https://doi.org/10.1016/j.fbp.2012.06.002>.
- Rouse, A. H., & Crandall, P. G. (1976). Nitric Acid Extraction of Pectin from citrus peel. *Proceedings of the Florida State Horticultural Society*, 89, 166–168.
- Wang, Y. C., Chuang, Y. C., & Hsu, H. W. (2008). The flavonoid, carotenoid and pectin content in peels of citrus cultivated in Taiwan. *Food Chemistry*, 106(1), 277–284.

A.T. Badaró, et al.

Food Chemistry 323 (2020) 126861

- <https://doi.org/10.1016/j.foodchem.2007.05.086>.
Wicker, L., Kim, Y., Kim, M. J., Thirkield, B., Lin, Z., & Jung, J. (2014). Pectin as a bioactive polysaccharide - Extracting tailored function from less. *Food Hydrocolloids*, 42(P2), 251–259. <https://doi.org/10.1016/j.foodhyd.2014.01.002>.
Zhu, N., Huang, W., Wu, D., Chen, K., & He, Y. (2017). Quantitative visualization of pectin distribution maps of peach fruits. *Scientific Reports*, 7(9275), <https://doi.org/10.1038/s41598-017-09817-7>.
Ziegler, J. U., Leitenberger, M., Longin, C. F. H., Würschum, T., Carle, R., & Schweiggert, R. M. (2016). Near-infrared reflectance spectroscopy for the rapid discrimination of kernels and flours of different wheat species. *Journal of Food Composition and Analysis*, 51, 30–36. <https://doi.org/10.1016/j.jfca.2016.06.005>.

GENERAL DISCUSSION

The need of fast, non-destructive and chemical-free techniques are in increasing demand in many fields of the industry. Near-infrared spectroscopy and hyperspectral imaging techniques have showed to have great potential in determining food quality parameters, authenticating food products, detecting food fraud, among many other applications. Because of this, in the recent years, many studies have emerged in the application of these techniques as an alternative for fast and automated determinations in the food industry. In the cereal field, spectroscopy techniques have been used for various applications, such as determination of grains quality parameters, assessment of mycotoxins in cereals, authentication of pure flour, determination of quality parameter in flour, among others. These applications were reported and discussed in Chapter 1.

Although many applications can be found in the literature for spectral techniques applied to cereals, the field is very wide and there is still a lack of some applications. In this context, this thesis was developed aiming to study the potential of NIR and NIR-HSI for identification, classification and quantification of different types of fibre samples added to the semolina and pasta produced by semolina-fibre formulations, and to monitor the cooking process to this fibre-enriched pasta by spectral techniques. The knowledge of these techniques allowed us to expand the initial objective of the thesis and applied NIR-HSI to quantify pectin content in orange peels as a faster and non-invasive method to justify investments in the waste processing and extraction methods.

Therefore, we started this thesis by investigating the potential of a near-infrared spectrometer and a hyperspectral imaging system in the assessment of five different types of fibre added in two different percentages to semolina (Chapter 2). In this part of the thesis, we dealt with many challenges, such as the heterogeneity of the samples, which showed us to be a barrier when working with single point's measurements. Soft Independent Modelling of Class Analogy (SIMCA) was performed on NIR data to discriminate pure and fibre-added semolina but presented sensitivity and specificity close or equal to zero, showing very low efficiency on discriminating and classifying the samples. The heterogeneity of the samples may be the cause of the poor performance of NIR. However, imaging techniques, such as NIR-HSI demonstrated to have a great potential in quantifying the amount of fibre in semolina as well as allowing the spatial distribution of fibre on the surface of semolina samples. SIMCA models applied on raw and pre-processed NIR-HSI data generally obtained high values of sensitivity and specificity for calibration and validation. Partial Least Square Regression (PLSR) was performed to test the ability of NIR-HSI to quantify the different percentages of fibre added to semolina. The

developed calibration models were representative and could be used to develop chemical maps, which allowed to visualize the content of fibre in the spatial dimension of hyperspectral images.

As aforementioned, these techniques have been investigated for the determination of many parameters in the cereal field. However, the use of such techniques for fibre determination in semolina and pasta was not found. Taking it into account, we followed the study by applying hyperspectral imaging in pasta samples in order to quantify fibre in pasta samples and identify its distribution (Chapter 3). In this part of the thesis, we applied Multivariate Curve Resolution with Alternating Least Squares (MCR-ALS) on the hyperspectral images to decompose them in the spectral profiles and relative distributions for each component in the sample. This algorithm showed its ability, together with NIR-HSI, for the evaluation, resolution, and quantification of fibre in enriched pasta. This part of the thesis allowed the study of the signal contributions and spatial distributions of fibre in the sample surface. Moreover, we were able to test different strategies to evaluate the feasibility of MCR-ALS in obtaining information in the pasta samples by using images of the raw ingredients (fibre and flour). The concentration maps constructed with the MCR-ALS results revealed how fibres behaved in each sample. Issues regarding the heterogeneity of the samples were also faced in this study. Some fibres showed to be better distributed, while some of them agglomerated in some parts of the sample. The results found by the MCR models as lack of fit (LOF), variance explained and similarity between the pure spectra and the spectra recovered by the augmented models were very satisfactory, suggesting a good performance of the models. Moreover, the concentration models provided us the necessary information to establish a threshold and use the obtained value to make a regression model (with OLS) to ascertain the validity of the approach. Half of the models showed good correlation with the real content, unlike the others. However, this may not indicate a bad performance of the models, but, as mentioned before, a problem with the homogeneity of the sample. Also, a great dispersion regarding the predicted fibre content among samples with the same percentage of fibre, which may have decreased the performance of the models.

At this point of the thesis, we saw the opportunity to enrich the content of this work. Therefore, we developed a study using hyperspectral images in the transmission mode for the determination of cooking time in pasta (Chapter 4). Optimal cooking time is a very subjective method, performed by manually pressing pasta samples between two glass slides. Then, this part of the thesis showed the good potential of HSI as an objective method to cooking time determination, as a first step to determine the optimum cooking time of pasta, meeting this

industry need. Linear Discriminant Analysis (LDA) models were built using the PCA scores and showed good ability in discriminating pasta samples in different times of cooking. Partial Least Squares Discriminant Analysis (PLSDA) models were also suitable to discriminate pasta samples. Moreover, the results demonstrate the possibility of developing qualitative models based on NIR-HSI transmission data.

Chapter 5 also presents good findings regarding hyperspectral imaging technique. In this part of the thesis, this technique was tested for the identification of pectin content in orange peels, in order to justify investments in the waste processing and extraction methods. Samples were separated in three groups that we named “Low content”, “Intermediate content” and “High content”, based on the average pectin content in orange peels reported by other authors. LDA models were developed in order to distinguish samples into one of these three classes, but, differently from previously chapter, the models were built with optimum wavelengths selected based on PCA loadings. The results for sensitivity and specificity were very satisfactory, demonstrating good capacity of the models to discriminate samples of residues of orange peel with different pectin content. Partial Least Squares Regression (PLSR) models for pectin content quantification showed excellent precision based on full or reduced spectra, providing high coefficients of determination.

GENERAL CONCLUSION

The review developed in Chapter 1 provided an overview on the application of spectral techniques for wheat flour and wheat-based products, more specifically near-infrared, Fourier-transform near-infrared and hyperspectral imaging. Moreover, this chapter provided information about the common wavelength ranges used for determination of this types of product, number of samples, acquisition mode and statistical methods performed used according to the study aim. In addition, we could identify the techniques more suitable to determine composition, authentication, and quality parameters. From that, we could link the initial idea of the project with the gaps found in the literature.

Chapter 2 revealed the potential of NIR-HSI for the identification and quantification of different types of fibre added to semolina. The results found in this chapter also showed the possibility of visualization of different percentage of fibre in the spatial dimension based on spectral characteristics. Moreover, this study showed that near infrared spectroscopy, although is widely used in the food field, did not provide good results in identifying and discriminating the different fibres added to semolina. Pure samples scanned with the portable near infrared spectrometer were easily identified by observing the PCA scores; however, when the fibre samples were added to semolina and generated a heterogeneous sample, the spectrum of a single point was not representative of the whole sample. Based on the results of this work, I methodology based on HSI could be developed and implemented in the wheat flour industry, to authenticate fibre enriched flour.

Chapter 3 showed that NIR-HSI, coupled to MCR-ALS, can identify and quantify fibre added to pasta samples. In this work, the homogeneity of pasta samples also showed to be a barrier, in the quantification of the expected fibre content in the sample. However, the results showed low LOF, high total variance, and great similarity between pure and recovered spectral profiles, denoting good replicability of the models and ability to quantify the amount of fibre in the region of the sample analysed. Although there is a high similarity of fibre and flour/control sample spectra, the outputs showed that even using the raw ingredient spectral information as in pure fibre and flour, it is possible to develop reliable models to identify fibre in pasta. In addition, the concentration maps can be very useful in monitoring the homogeneity of the samples. The performance of the models also showed that it is possible to have a qualitative overview of the fibre distribution in pasta samples, and, in some cases, it is possible to quantify the amount of fibre in the sample surface. Thus, this chapter presented a great possibility to apply such technique as a qualitative and quantitative method for authentication of fibre-enriched pasta.

Chapter 4 demonstrated that NIR-HSI in the transmission mode is a suitable technique as an objective alternative for the determination of optimal cooking time of pasta as way of automating the determination of pasta attributes. Although some low values were found for one of the classes, in general, the models showed good ability in classifying pasta samples according to the cooking time. LDA models constructed based on PCA scores showed to be a good way of classifying samples with reduced dimensionality. Therefore, this work suggested that hyperspectral imaging have good potential as an objective method to cooking time determination, meeting this industry need. Although the spectrum range used in this work was small, it showed a great possibility to optimize food quality evaluation. Based on the results of this work, we could develop a methodology to determine the optimum cooking time of pasta and other pasta attributes that are very difficult to measure with good precision. The determination of these parameters would contribute to determine these parameters in the pasta industry.

Finally, chapter 5 confirmed the potential of hyperspectral imaging technique to classify orange peel samples according to the pectin content. LDA models were able to separate the groups of samples containing different percentages of pectin into the same three classes (low, intermediate and high pectin content). Additionally, PLSR models for pectin content quantification based on full and reduced spectra showed excellent precision, providing high coefficients of determination. The results showed that this technique holds potential as an alternative to the carbazole colorimetry method to quantify pectin in orange peels, and to categorize orange peel samples into groups of different pectin concentration and can be used to justify investments in the waste processing and extraction methods. Although this work was tested by using the orange peels in the powder form, the results were very satisfactory and showed the possibility to apply this technique for pectin determination. Thus, future works could include the determination of pectin in orange peels in the fresh form.

STUDENT STORY

The Food Engineer Amanda Teixeira Badaró started as a Master's student at School of Food Engineering (FEA) at University of Campinas (UNICAMP), with a financial support of CNPq for 18 months (from august/2016 to February/2018). During this time, she fulfilled all the required credits related to the mandatory subjects of the program, as well as elective subjects, aiming to acquire more knowledge related to the project field. In all subjects attended, she reached maximum grade. Amanda also participated in the Training Internship Program (PED) to improve the student's education as a trainee in teaching experience. Due to her outstanding performance, she was advised by the board of lecturers that evaluated her after the first year to change for PhD before finishing her MSc degree. Hence, in fact, she spent her first year mainly taking lecturers and writing the project for a Master thesis, and she only changed to PhD in March 2018, with the financial support of FAPESP (grant number 2017/17628-3).

Once she started her PhD, the interest in a collaboration with Instituto Valenciano de Investigaciones Agrarias (IVIA) arose, and we contacted Professor Jose Blasco-Ivars to be a board member of her qualification exam. After, it was possible to obtain funding from FAPESP (grant number 2019/06842-0) for Amanda to stay in Valencia for one year to perform part of her PhD, under supervision of Professor Jose Blasco. After the student demonstrated competencies and skills to excel in the graduate program and securing funding approval for the student to stay abroad, she started as a PhD student in Food Science, Technology and Management at Universitat Politècnica de València (UPV) in a cotutelle agreement under the supervision of Professor Nuria Aleixo Borrás.

During her PhD, she contributed to experimental activities in other projects in the field, helping other students. In addition, she participated in events, international conferences and courses. Furthermore, she worked developing other projects in parallel, which generated co-authorship on other studies. All the studies that Amanda has participated in are listed below.

Published papers:

BADARÓ, A. T., DE MATOS, G. V., KARAZIACK, C. B., VIOTTO, W. H., BARBIN, D. F. (2021). Automated Method for Determination of Cheese Meltability by Computer Vision. *FOOD ANALYTICAL METHODS*, v.1, p.1 - . <https://doi.org/10.1007/s12161-021-02094-1>

BADARÓ, A. T.; AMIGO, J. M.; BLASCO, J.; ALEIXOS, N.; FERREIRA, A. R.; CLERICI, M. T. P. S.; BARBIN, D. F. (2021). Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fiber distribution in enriched pasta. *FOOD CHEMISTRY*, v. 343, p. 128517. <https://doi.org/10.1016/j.foodchem.2020.128517>

BADARÓ, A. T.; MARTIN, J. F. G.; BARRERA, M. C. L.; BARBIN, D. F.; MATEOS, P. A. (2020). Determination of pectin content in orange peels by Near Infrared Hyperspectral Imaging. *FOOD CHEMISTRY*, v. 323, p. 126861. <https://doi.org/10.1016/j.foodchem.2020.126861>

MARTÍN, J. F. G.; **BADARÓ, A. T.**; BARBIN, D. F.; MATEOS, P. A. (2020). Identification of Copper in Stems and Roots of *Jatropha curcas* L. by Hyperspectral Imaging. *PROCESSES*, v. 8, p. 823. <https://doi.org/10.3390/pr8070823>

CARVALHO, L. M.; MADRUGA, M. S.; ESTÉVEZ, MA.; **BADARÓ, A. T.**; BARBIN, D. F. (2020). Occurrence of wooden breast and white striping in Brazilian slaughtering plants and use of near-infrared spectroscopy and multivariate analysis to identify affected chicken breasts. *JOURNAL OF FOOD SCIENCE*, v. 85, p. 3102-3112. <https://doi.org/10.1111/1750-3841.15465>

BADARÓ, A. T.; MORIMITSU, F. L.; FERREIRA, A. R.; CLERICI, M. T. P. S.; BARBIN, D. F. (2019). Identification of fiber added to semolina by near infrared (NIR) spectral techniques. *FOOD CHEMISTRY*, v. 289, p. 195-203. <https://doi.org/10.1016/j.foodchem.2019.03.057>

PEREZ, I. M. N.; CRUZ-TIRADO, L. J. P.; **BADARÓ, A. T.**; DE OLIVEIRA, M. M.; BARBIN, D. F. (2019). Present and future of portable/handheld near-infrared spectroscopy in chicken meat industry. *NIR NEWS*, v. 30, p. 096033601986147-29. <https://doi.org/10.1177/0960336019861476>

BARBIN, D. F.; **BADARÓ, A. T.**; HONORATO, D. C. B.; IDA, E. Y.; SHIMOKOMAKI, M. (2019). Identification of turkey meat and processed products using near infrared spectroscopy. *FOOD CONTROL*, v. 107, p. 106816-9. <https://doi.org/10.1016/j.foodcont.2019.106816>

NOLASCO PEREZ, I. M.; **BADARÓ, A. T.**; BARBON, S.; BARBON, A. P. A. C.; POLLONIO, M. A. R.; BARBIN, D. F. (2018). Classification of Chicken Parts Using a Portable Near-Infrared (NIR) Spectrophotometer and Machine Learning. *APPLIED SPECTROSCOPY*, v. 1, p. 000370281878887, 2018. <https://doi.org/10.1177/0003702818788878>

Book chapters:

ZARPELÃO, B. B.; JUNIOR, S. B.; **BADARÓ, A. T.**; BARBIN, D. F. On the use of blockchain for agrifood traceability. *Food Authentication and Traceability*. 1st ed.: Elsevier, 2021, p. 279-302. <https://doi.org/10.1016/B978-0-12-821104-5.00001-5>

JUNIOR, S. B.; SANTANA, E. J.; **BADARÓ, A. T.**; BORRÁS, N. A.; BARBIN, D. F. Advantages of Multi-Target Modelling for Spectral Regression. *Spectroscopic Techniques & Artificial Intelligence for Food and Beverage Analysis*. 1st ed.: Springer Singapore, 2020 , p. 95-121. https://doi.org/10.1007/978-981-15-6495-6_5

BADARÓ, A. T.; PASQUINI, C.; BARBIN, D. F. Food Quality and NIR Spectroscopy in the Omics Era. *Reference Module in Food Science*. 1st ed.: Elsevier, 2020, p. 1-. <https://doi.org/10.1016/B978-0-08-100596-5.22849-7>

REFERENCES

- Badaró, A.T., Morimitsu, F.L., Ferreira, A.R., Clerici, M.T.P.S., Fernandes Barbin, D., 2019. Identification of fiber added to semolina by near infrared (NIR) spectral techniques. *Food Chem.* 289. <https://doi.org/10.1016/j.foodchem.2019.03.057>
- Barbin, D.F., Badaró, A.T., Honorato, D.C.B., Ida, E.Y., Shimokomaki, M., 2020. Identification of Turkey meat and processed products using near infrared spectroscopy. *Food Control* 107. <https://doi.org/10.1016/j.foodcont.2019.106816>
- Cocchi, M., Durante, C., Foca, G., Marchetti, A., Tassi, L., Ulrici, A., 2006. Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. *Talanta* 68, 1505–1511. <https://doi.org/10.1016/j.talanta.2005.08.005>
- Feng, Y.-Z., Sun, D.-W., 2012. Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review. *Crit. Rev. Food Sci. Nutr.* 52, 1039–58. <https://doi.org/10.1080/10408398.2011.651542>
- Manley, M., 2014. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chem. Soc. Rev.* 43, 8200–8214. <https://doi.org/10.1039/c4cs00062e>
- Mishra, P., Herrero-Langreo, A., Barreiro, P., Roger, J.M., Diezma, B., Gorretta, N., Lleó, L., 2015. Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis. *J. Near Infrared Spectrosc.* 23, 15–22. <https://doi.org/10.1255/jnirs.1141>
- Porep, J.U., Kammerer, D.R., Carle, R., 2015. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci. Technol.* 46, 211–230. <https://doi.org/10.1016/j.tifs.2015.10.002>
- Teye, E., Huang, X., Sam-Amoah, L.K., Takrama, J., Boison, D., Botchway, F., Kumi, F., 2015. Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food Chem.* 176, 403–410. <https://doi.org/10.1016/j.foodchem.2014.12.042>
- Üçüncüoğlu, D., Ilaslan, K., Boyaci, I.H., Özay, D.S., 2013. Rapid detection of fat adulteration in bakery products using Raman and near-infrared spectroscopies. *Eur. Food Res. Technol.* 237, 703–710. <https://doi.org/10.1007/s00217-013-2030-x>
- Verdú, S., Vázquez, F., Grau, R., Ivorra, E., Sánchez, A.J., Barat, J.M., 2016. Detection of adulterations with different grains in wheat products based on the hyperspectral image

technique: The specific cases of flour and bread. *Food Control* 62, 373–380.
<https://doi.org/10.1016/j.foodcont.2015.11.002>

FINAL CONSIDERATIONS

First of all, I would like to thank Douglas and Nuria for the supervision in this thesis. Secondly, I would like to thank the evaluators for taking the time to correct the thesis and give their suggestions so I could improve my work and acquire more knowledge. I corrected most of the comments, which I strongly agreed on. I am writing these final considerations to explain some points raised during the correction of the thesis.

I will start explaining the broader title of the thesis. My supervisors and I decided to go for a broader title because the thesis includes more than one topic. At first, I was going to work with NIRS and HSI to study flour and pasta, but during these five years of PhD we dealt with many drawbacks, which lead me to participate in other works. Then, we saw that my thesis was more about the techniques than the product. So, when I was doing my internship in Spain, Douglas and I had the opportunity to work with a group of researchers from Universidad de Sevilla, which was working with the application of HSI to quantify the amount of pectin orange peels. As I had worked with the technique in other powdered product, we decided that it would be suitable for me to work with them. Oranges are abundant in Spain, so it would be a great alternative to quantify the pectin in oranges peel prior extraction. That is why I also included a chapter about the pectin in orange peels.

Moreover, the initial idea of working with flour and pasta was not so applicable so we tried to fill the gaps we found in the literature. Some of the comments in this thesis was regarding the applicability of the work. The industry of pasta needs more objective and fast techniques to improve their assessments. In this context, we wanted to develop alternative techniques to authenticate enriched flour/pasta products and investigate the cooking process to, in the future, be able to determine the optimum cooking time by pressing pasta samples.

The fact that some errors were found to be higher than expected raised a question about either the work was applicably or not. In fact, there is a large gap for the application of these techniques in the food industry, and only a few works have applied spectral techniques in-line. Therefore, there is the need for further studies in order to apply these techniques in the process line so that the industry would have a more controlled process. The combination of off-line and in-line measurements could be a first step on the use of NIR in in-line analysis. Also, handheld spectrometers could be used in the process line.

In addition, another problem we faced during the development of the work was the heterogeneity of the samples. The fibre and flour did not mix well, so, by comparing a low cost and not so sensitive spectrometer with a robust hyperspectral camera, the results were quite

discrepant. Although HSI overcomes the problem with spatial information, some points regarding the heterogeneity in the depth of the samples were questioned. With flour samples, I believe that this could be solved by using a thin layer all the surface analysed. In contrast, pasta samples were approximately 2 mm thick, so we considered that all the information in the depth was acquired.

Finally, the reviewers raised some points regarding the theory of chemometrics, which, in fact, were not fully described in some specific parts. Hence, all changes were properly addressed when feasible.

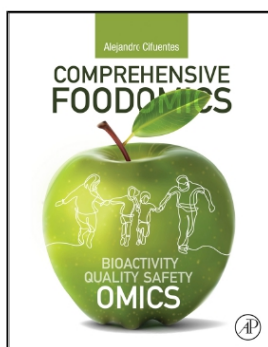
APPENDIX I

Food Quality and NIR Spectroscopy in the Omics Era

Author's personal copy

Provided for non-commercial research and educational use.
Not for reproduction, distribution or commercial use.

This article was originally published in *Comprehensive Foodomics*, published by Elsevier, and the attached copy is provided by Elsevier for the author's benefit and for the benefit of the author's institution, for non-commercial research and educational use including without limitation use in instruction at your institution, sending it to specific colleagues who you know, and providing a copy to your institution's administrator.



All other uses, reproduction and distribution, including without limitation commercial reprints, selling or licensing copies or access, or posting on open internet sites, your personal or institution's website or repository, are prohibited. For exceptions, permission may be sought for such use through Elsevier's permissions site at:

<https://www.elsevier.com/about/our-business/policies/copyright/permissions>

From Badaró, A.T., Pasquini, C., Barbin, D.F., 2021. Food Quality and NIR Spectroscopy in the Omics Era. In: Cifuentes, A. (Ed.), *Comprehensive Foodomics*, vol. 3. Elsevier, pp. 231–243.
ISBN: 9780128163955

Copyright © 2021 Elsevier Inc. All rights reserved.
Elsevier

Author's personal copy

3.20 Food Quality and NIR Spectroscopy in the Omics Era

Amanda Teixeira Badaró^{b,c}, Celio Pasquini^a, and Douglas Fernandes Barbin^b, ^aInstitute of Chemistry, University of Campinas, Campinas, Brazil; ^bSchool of Food Engineering, University of Campinas, Campinas, Brazil; and ^cUniversitat Politècnica de València, Valencia, Spain

© 2021 Elsevier Inc. All rights reserved.

3.20.1	Introduction	231
3.20.2	Recent Advances in Food Quality and NIR Spectroscopy	232
3.20.3	Vibrational Spectroscopy	232
3.20.4	Near Infrared Spectroscopy	233
3.20.5	Chemometrics	234
3.20.6	Spectral Pre-processing	236
3.20.7	Linear Methods	236
3.20.7.1	Principal Component Analysis (PCA)	236
3.20.7.2	Linear Discriminant Analysis (LDA)	236
3.20.7.3	Soft Independent Modeling by Class Analogy (SIMCA)	237
3.20.7.4	Partial Least Squares (PLS) and Partial Least Squares Discriminant Analysis (PLS-DA)	237
3.20.8	Non-linear Methods	238
3.20.8.1	Neural Networks (NN)	238
3.20.8.2	Deep Learning (DL)	238
3.20.9	Support Vector Machine (SVM)	239
3.20.10	Random Forests (RFs)	239
3.20.11	Data Fusion	240
References		240

3.20.1 Introduction

Nowadays there is a great concern regarding the authentication of food, from different points of view. It is highly demanded to produce safer food products, but also to determine their origin, precedence, and composition. Authentication involves procedures to determine whether a food sample complies with its label description and is in accordance with legal standards (Abbas et al., 2018).

Food fraud is generally performed for economic reasons and it is estimated that it causes a deficit of around US\$ 49 billion per year (McGrath et al., 2018). These losses are related mainly to an increase in different types of cheating and the difficulty to identify the different aspects of fraud (Bouzembrak and Marvin, 2016). These aspects involve several different traits of modification, such as adulteration, falsification, substitution or incorrect labeling of the product presented to consumer (Galvin-King et al., 2018; Lohumi et al., 2015; Spink and Moyer, 2011).

Fraudulent adulteration of foodstuffs may involve quality improvements only in appearance, but not real (such as adding colorants), reduction of costs by replacement of ingredients with low cost substituents, or using forbidden substances to extend product shelf-life (Barreto et al., 2018). In this sense, products of high economic value are usually the most susceptible to food frauds. These products include herbs and spices (Oliveira et al., 2019), olive oil (Azadmard-Damirchi, 2010), milk (Abermthy and Higgs, 2013; Kim et al., 2015), honey (Xue et al., 2013).

How food fraud started and have spread is still unknown. If the fraud seeks monetary profit, probably does not result in safety risks and mostly, there is no feedback regarding food safety by consumers (Johnson, 2014). Therefore, global regulations have been created in order to avoid fraud and mislabeling and to guarantee the authenticity of food products, and must be based on reliable techniques of analysis (Böhme et al., 2019). Regarding food composition and its quality attributes, there has been extensive search for development of analytical methods that may identify food composition with high accuracy, or that may determine its main attributes that differentiates from others (Lohumi et al., 2015).

Analytical methods of food authentication can be target or non-target. The methods most commonly used involve either identification of physical parameters such as color and texture measurements, as chemical methods such as high performance liquid chromatography (HPLC) and mass spectrometry (MS) (Lohumi et al., 2015). Recently, the called “-omics” science has gained much interest, and these technologies include proteomics, metabolomics and genomics (Ortea, O'Connor and Maquet, 2016).

Proteomics involves the knowledge of protein structure and function. This approach also includes the quantification of protein, how they interact between them or their mechanism of action (Ortea et al., 2016). The tools based on this approach can be very useful in food authentication, since proteins are present in food components and products. One of the methods for protein analysis is mass spectrometry (MS). This method is performed by characterizing and sequencing peptides and proteins, and, through it is also possible to obtain information about specific proteins (Böhme et al., 2019). In this context, untargeted proteomics does not measure a specific protein, it detects and compares patterns among different samples, which allows the detection of a wide

Author's personal copy

232 Food Quality and NIR Spectroscopy in the Omics Era

group of proteins simultaneously. On the other hand, targeted proteomics refers to a specific group of protein to be analyzed. However, targeted and untargeted proteomics can be used in a complementary way (Lieber and Zimmerman, 2013).

Metabolomics identify and quantify the metabolites of a biological system. This approach determines the protein activity, biochemical processes and the metabolites function (Van Emon, 2016). This study can be performed either comparing patterns from different metabolites groups (metabolite fingerprinting) or within a specific group (metabolic profiling). The most used tools for these analyses are MS and nuclear magnetic resonance (NMR). These techniques allow the high precision analysis of various components concomitantly, being first the primary polar compounds, followed by polar and nonpolar lipids, terpenoids, and sterols. Despite this great advantage and specificity, the amount of data generated by these analyses must be correctly interpreted, and only relevant information analyzed. Therefore, chemometrics have been used together with these techniques (Böhme et al., 2019; Ferri et al., 2015).

On the other hand, genomics involves the analysis of DNA structure and function. DNA-based techniques, as polymerase chain reaction (PCR), have been routinely used in the food field, since they allow the detection of low concentration of adulterants in food matrices. Recently, innovative genomics-based techniques as next-generation sequencing (NGS), high resolution melting (HRM), droplet digital PCR (ddPCR), loop mediated isothermal amplification (LAMP) and DNA-barcoding have been applied to identify food fraud as an alternative to improve DNA-based techniques in various terms (Böhme et al., 2019).

In food adulteration assessment, these targets can be used as markers for further determinations (Gallo and Ferranti, 2016). However, the methods of analysis performed in these techniques usually has a high detection cost, long period of analysis and use of reagents. Therefore, promising and innovative spectral techniques are an alternative in the -omics (Han et al., 2019).

Spectroscopic techniques have been investigated to evaluate a broad range of foods. These methods are minimally invasive, do not require chemical reagents (green chemistry), are adaptable to in-line inspection lines, and considering the number of samples it can analyze, less expensive (Porep et al., 2015).

Currently, the inspection of food in the supply chain to avoid food fraud represents large investments, since there are several stages during the production and distribution of food products. However, addressing these tasks to guarantee the product origin and prevent any type of modification during its production and distribution may help economic growth and increase the consumer confidence (van Ruth et al., 2017; Velásquez et al., 2017).

In this regard, the consumer must be aware of the tasks performed to prevent product alteration, as having knowledge of this management may enhance consumer assurance on the product or brand (de Jonge et al., 2004). This communication has become more important considering recent food fraud scandals that have affected consumer confidence.

3.20.2 Recent Advances in Food Quality and NIR Spectroscopy

Food authentication and fraud detection has faced great challenges as the number of different adulterants increase widely. Hence, identifying specific compounds has become a daunting task. Establishing targets for some specific elements might be easier to accomplish, but it must be strictly for the target compound. In this context, two approaches can be considered for food authentication: targeted and untargeted (or non-targeted) strategies (Pustjens et al., 2015).

Targeted analytical methods analyze a relatively small and specific number of compounds that are chemically characterized with established importance before data acquisition. In this sense, it is necessary to know how they behave (e.g., mass spectra, retention times) and the results expected. The main disadvantage is that other chemicals that could be present in the sample would not be identified. In addition, these methods usually are expensive and require a large amount of time, since quantification of metabolites is achieved using chemical standards and calibration curves for each specific metabolite, which limits their application to the laboratory (Esslinger et al., 2014).

On the other hand, untargeted approaches are developed in order to detect unpredicted variations in metabolite concentrations. The main objective is to maximize the number of components that could be detected and therefore allow for the identification of unexpected variations. A non-targeted method would allow detection of foreign substances, both known and unknown; thus, there is an increasing interest for applications of this methods in food and agricultural products (Esslinger et al., 2014).

The most commonly non-targeted methods include vibrational spectroscopy (Fourier Transform Infrared (FT-IR), near-infrared (NIR), hyperspectral imaging (HSI), multispectral imaging (MSI) and Raman imaging), isotopic techniques that include nuclear magnetic resonance (NMR) and electric spin resonance (ESR), and techniques based on spectrometry and chromatography (McGrath et al., 2018). The main advantage of vibrational spectroscopy is the capability of adaptation for online and infield applications (Velásquez et al., 2017). In contrast, the biggest challenge for non-targeted analysis is the data processing, since there is a large amount of data generated for each measurement. Therefore, these techniques rely on multivariate statistical analyses and chemometrics to analyze the results and efficiently achieve practical results (McGrath et al., 2018). Thus, major efforts have been done in the development of new statistical methods to attend this demand.

3.20.3 Vibrational Spectroscopy

Vibrational spectroscopy is a fast, reliable and future technique that has been increasingly applied to different fields, including the authentication and characterization food products. Vibrational techniques include Raman, Near Infrared Spectroscopy (NIR),

Author's personal copy

Fourier Transform Infrared Spectroscopy (FT-IR), and Hyperspectral Imaging (HIS). Even though these techniques are accurate, rapid, non-destructive, reagent free and allow the determination of different components using the same spectrum; each system has its particularities. The biggest differences are based on the dimension (spatial, spectral), sensitivity to minor constituents, amount of information provided and so on. However, these techniques have shown incredible potential in the authentication of food products (Lohumi et al., 2015; Oliveira et al., 2019).

Vibrational spectroscopy can be implemented using portable or benchtop devices. Usually, handheld devices are cheaper and easier to use, whereas benchtop can be more accurate, depending on what is being analyzed. Basically, a NIR spectrometer contains a radiation source, a sample holder, a device used for wavelength selection, a detector and a computer system. Moreover, the many platform and technology available suggest to use data fusion as a new strategy to improve the outcomes (Deidda et al., 2019; Oliveira et al., 2019; Porep et al., 2015).

3.20.4 Near Infrared Spectroscopy

Near Infrared (NIR) is a vibrational spectroscopy that covers wavelengths from 750 to 2500 nm (13,333 to 4000 cm^{-1}). This spectral region, considered of low energy when compared with the ultraviolet and visible regions, interacts with the chemical species by increasing their vibrational energy. The potential energy of a given vibration of a chemical bond between two atoms in a molecule describes a relationship in close agreement with the anharmonic behavior of the atomic oscillator, as shown in Fig. 1. The quantum anharmonic oscillator model restricts the vibrational energy to well defined levels, as also shown in Fig. 1.

Energy transitions between the quantum vibrational levels may result from the absorption of the energy transported by an electromagnetic wave of radiation. In this case, the radiation energy must match the difference of the energy of the vibrational levels. It means that the energy difference between two vibrational levels (ΔE) must be,

$$\Delta E = h\nu \tag{1}$$

where h is the Planck constant and ν is the frequency of the electromagnetic wave. In terms of vibrational levels transitions, is also true that the frequency of the radiation must match the vibrational frequency of the atomic oscillator, given by

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

where μ is the reduced mass of the oscillating system constituted of two bonded atoms, equals to

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \tag{3}$$

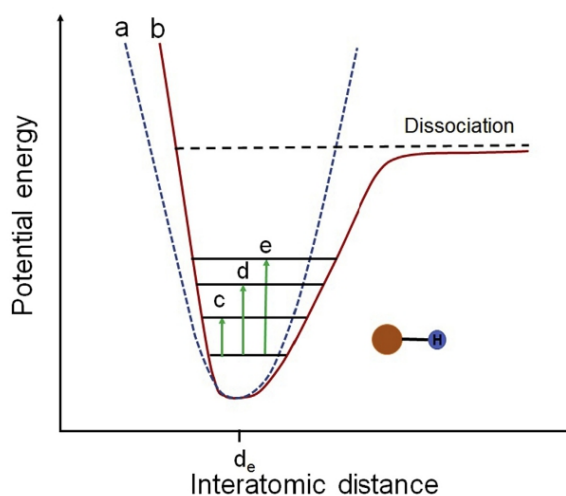


Figure 1 Potential energy behavior of a typical diatomic oscillator. d_e , equilibrium distance; a, harmonic potential; b, anharmonic potential; c, fundamental vibrational energy transition, d and e, first and second overtone transition.

Author's personal copy

234 Food Quality and NIR Spectroscopy in the Omics Era

In addition to the above requirements, the vibrational mode must produce a change in the dipole moment of the chemical bond in order the electromagnetic energy can be converted to the mechanical vibrational energy and thus absorbed. Those vibrational mode are referred as infrared active modes.

Though the most probably transition occurs between the fundamental level, where most of the molecules are found in ambient conditions, and the adjacent level, the energy correspondent to this transition are associated with the frequencies of the radiation covered by the mid infrared region. Nevertheless, the anharmonic behavior of the potential energy of real atomic oscillators permits transitions among the fundamental level and superior levels defined as overtones of the fundamental transition (as pointed in Fig. 1). The radiation associated with overtone energy transitions are of higher frequency, proportional to the energy differences between the fundamental level and the overtone level. Several chemical bonds are excited to the first, second and even third overtones requiring the radiation energy associated with higher radiation frequencies frequently to belong to the near-infrared spectral region.

In addition, the mechanical and electronic anharmonicity of a chemical bond allows for some vibrational modes to couple to be excited as a combination mode requiring the radiation energy that is the sum of the original excited levels. The energy associated with the excitation of a combination also may require the frequency of the radiation to belong to the near-infrared spectral region.

As a matter of fact, a typical NIR spectrum of a material results from the absorption of the radiation mostly employed to excite overtones and combinations of vibrational modes associated to the chemical bonds present in the chemical species present in that material. Secondly, a number of vibrational phenomena, such as Fermi and Darling-Davison resonance, may contribute to the complexity of the NIR absorption spectra.

Due to the requirements for an anharmonic vibrational behavior and the higher energy associated with the overtones and combinations, the interaction of the NIR radiation with matter occurs mainly due to excitation chemical groups containing atoms of carbon, nitrogen, oxygen or sulfur linked to hydrogen (Abbas et al., 2018; Pasquini 2003, 2018).

Though the limited types of chemical bond probed by NIR radiation are very limited, several additional effects occurring at secondary microscopic and macroscopic level contribute to increase the amount of information captured by an NIR absorption spectrum. In this way, changes in the chemical bond energy caused, for example, by substitution of hydrogen atoms in a C - H bond by a highly electronegative atom in a molecule, or change in the mass of the element isotopes result in shifts of the frequency of the absorbed radiation, according with Eq. (2). The inter-molecular hydrogen bonds reduce the bond energy between the donor and hydrogen atoms and decrease its degree of anharmonicity. Consequently, the absorption frequency of the donor-hydrogen vibration is shifted to low energies and the intensities of absorption reduced. The way a long hydrocarbon chain conforms in a solid sample state and the temperature of the samples are examples macroscopic level effects causing changes in the NIR absorption spectra.

Overall, the primary and secondary effects on the NIR spectrum of the chemical species adds a considerable amount of unique information about the chemical composition and physical characteristics, which can be used to identify, classify or quantify several samples properties.

In addition to the great amount of information captured in a NIR spectrum, the sample measurement can be easily performed, contributing to the dissemination of the use of this type of spectroscopy. In fact, most of the measurement in this spectral region can be performed directly in the intact sample or after it had been submitted to a little, fast and greener pre-treatment like grinding. In consequence, reflectance of grinding solids is the most used form of sample measurement in NIR. In this case, the radiation beam is impinged on the sample surface. Depending on the beam power, the radiation can penetrate up to 2–4 mm in the sample surface, interacting with the chemical species there present and return to be measured by the spectrophotometer detection system. Thus, the NIR spectrum is rapidly acquired, and NIR spectroscopy have been widely used in this way for food quality assessment (Table 1 and 2)

NIR spectroscopy provides a large amount of data, which must be correctly interpreted. In fact, NIR spectra usually have a great number of overlapping bands, reflecting the sample composition and several physical characteristics. Therefore, in order to obtain correct and reliable chemical information from sample spectrum, multivariate data analysis is required. This allows to extract relevant information from spectra without losing wanted information, and to develop suitable models for the characteristic to be studied (Kemsley et al., 2019; Manley, 2014).

3.20.5 Chemometrics

Chemometrics is a science that uses mathematics and statistics to select required and important information from a complex dataset, as NIR spectral data, and to develop models for further chemical investigation. Therefore, chemometrics in spectral data involves the reduction of the original data for better interpretation, performance of spectral correction (pre-processing), development of models according to the feature of interest (qualitative or quantitative analysis), and interpretation of results (Fernández Pierna et al., 2012).

In addition, the data analysis can be divided into two categories: unsupervised and supervised. Unsupervised methods are used when there is not a specific characteristic to be explored and seek to explore the behavior of the samples group. On the other hand, supervised methods relate the spectral data with an expected response, which can be a class (classification methods) or a value (regression methods) (Lohumi et al., 2015).

Author's personal copy

Table 1 Food quality assessment using NIR spectroscopy in the last 5 years

Food product	Task	Equipment	Spectral range	Statistical analyses	References
Beef, pork and chicken meat	Determination and quantification of pork adulteration in beef meatball	Benchtop	850–2000 nm	LDA, PLSR	Kuswandi et al. (2015)
	Assessment of plant and animal proteins as potential adulterants in minced beef and pork	Benchtop	400–1000 nm 900–1700 nm	LDA, decision trees, KNN, PLS-DA, FFNN, SVM, Naïve Bayes (NB).	Rady and Adedeji (2018b)
	Prediction of aging time of beef steaks	Portable	350–2500 nm	PCA, PLS-DA	Moran et al. (2018)
	Comparison of NIR spectroscopy with other techniques	Portable	900–1700 nm	PCA, LDA, PLSR	Nolasco-Perez et al. (2019)
Fish	Identification according to production method and geographical origin	Benchtop	1100–2500 nm	PCA, OPLS-DA	Ghidini et al. (2019)
Honey	Detection of honey origin and adulterants	Benchtop	400–2500 nm	PCA, PLSR	Kumaravelu and Gopal (2015)
	Quantification of honey adulterated with high fructose corn syrup	Benchtop	400–2500 nm	PCA, HCA, LDA, PLSR	Ferreiro-González et al. (2018)
	Identification of adulterants (rice syrup invert sugar, brown cane sugar, fructose syrup)	Benchtop	400–2500 nm	HCA, LDA, PLSR	Aliaño-González et al. (2019)
Vinegars	Typification of Spanish wine vinegars with protected designation of origin	Benchtop	12000–4000 cm ⁻¹	PCA, PLS-DA	Ríos-Reina et al. (2018)
	Identification of wine vinegar from protected denomination of origin	Benchtop	12000–4000 cm ⁻¹	(PCA, PARAFAC, MCR for decomposition) PLS-DA, NPLS-DA	Ríos-Reina et al. (2019)
Fruits	Identification of apple origin, cultivar and elevation	Benchtop	10000–4000 cm ⁻¹	PCA, QDA	Eisenstecken et al. (2019)
Rice	Classification into quality grades and identification of origin	Handheld, portable	740–1070 nm	PCA, k-NN, SVM	Teye et al. (2019)
	Discrimination between organic and conventional	Benchtop	830–2500 nm	PCA, PLS-DA	Xiao et al. (2019)
Cocoa powder	Detection of the adulteration in cocoa powders with carob powder	Benchtop	1100–2500 nm	PCA, PLS-DA, PLSR	Quelal-Vásconez et al. (2018)
Milk	Detection of adulteration of skim milk powder	Benchtop	400–2498 nm	PCA, SIMCA, PCR, PLSR	Capuano et al. (2015)
	Detection of adulteration in camel milk with goat milk	Benchtop	700–2500 nm	PCA, PLS-DA, PLSR	Mabood et al. (2017)
Coffee	Authentication of organic milk	Portable	908–1676 nm	PCA, PLS-DA	Liu et al. (2018)
	Detection of corn adulteration in coffee samples	Benchtop	400–2500 nm	PLSR	Winkler-Moser et al. (2015)
	Identification and quantification of Arabica coffee adulterations with Robusta coffee, corn, peels, and sticks	Portable	908–1676 nm	PCA, PLSR	Correia et al. (2018)
Nuts	Detection of hazelnut adulteration	Benchtop	12000–3650 cm ⁻¹	SIMCA	Márquez et al. (2016)
Starch	Classification and quantification of cheaper starches (corn and wheat starch) in ultrafine granular powder of Shanyao	Benchtop	1100–2300 nm	PCA, OPLS-DA, PLSR, (i, si)-PLSR	Ma et al. (2017)
Flour	Authentication of taro flour from wheat and sago flour	Benchtop	10000–4000 cm ⁻¹	PCA, DA	Rachmawati et al. (2017)
	Detection and quantification of unripe banana flour adulteration with wheat flour	Portable	447–1005 nm	PCA, PLSR	Faith et al. (2019)
Oil	Classification and quantification of palm oil adulteration	Portable	950–1650 nm	SIMCA, PLSR, CARS-PLSR	Basri et al. (2017)
	Detection of extra virgin olive oil adulteration	Benchtop	12000–4000 cm ⁻¹	PCA, SIMCA	Vanstone et al. (2018)
	Authentication of extra virgin olive oil origin	Benchtop	1100–2500 nm	PARAFAC, PCA, PLS1–DA, NPLS–DA	Jiménez-Carvelo et al. (2019b)
	Detection and quantification of argan oil adulteration with cheap vegetable oils	Portable	500–1000 nm 1000–1700 nm	PCA, PLSR	Farres et al. (2019)
Butter	Detection of fraud in oil capsules	Portable	900–1700 nm	PCA	Leme et al. (2019)
	Detection and quantification of tallow adulteration in clarified butter	Benchtop	10000–4000 cm ⁻¹	PCA, PLS-DA, PLSR	Mabood et al. (2018)
Spices	Detection of metanil yellow in turmeric powder	Portable	900–1700 nm	PCA, PCR, PLSR	Kar et al. (2018)

Author's personal copy

236 Food Quality and NIR Spectroscopy in the Omics Era

Table 2 Omics approach in food products

Field	Task	Technique	Statistical analyses	References
Metabolomics	Determination of perseitol (marker for the botanical origin of avocado honey)	HPLC, NIR	PCR, MPLS	Dvash et al. (2002)
	Discrimination between conventional and organic rice	UHPLC-Q-TOF MS analysis	PCA, HCA, PLS-DA, OPLS-DA	Xiao et al. (2018)
Genomics	Discrimination between grains from two gene groups	NIR	Classification and regression	Curzon et al. (2019)
Proteomics	Discrimination between meat and bone meal	FTIR	PCA, PLS-DA	Han et al. (2019)
	Discrimination between conventional and organic rice	Nano LC-MS/MS analysis	PCA, HCA	Xiao et al. (2019)
	Identification of N-Glycosites in chicken egg white proteins	MS/MS analysis	–	Geng et al. (2017)

In the context of omics science, more specifically in foodomics, data analysis usually aimed to explore the data, to identify patterns or to predict parameters (as in regression or classification models). Both unsupervised and supervised methods of analysis are very important in foodomics, and sometimes these methods are used together to achieve a better result (Bevilacqua et al., 2017).

Moreover, the methods of data analysis can still be classified as linear and nonlinear. As the name implies, linear methods are those applied when there is a linear relation between the dataset, whereas nonlinear methods are those used for nonlinear relationship (Rosipal, 2010). Linear methods include, among others, Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Partial Least Squares (PLS), Partial Least Squares Discriminant Analysis (PLS-DA) and Soft Independent modeling by class analogy (SIMCA). Nonlinear methods include Neural Networks (NN), Deep Learning (DL), Support Vector Machine (SVM) and Random Forest (RF). Linear methods continue to be the most used among researchers.

3.20.6 Spectral Pre-processing

Data pre-processing is usually the first step in chemometrics. This step is required in order to remove unwanted information, as light scattering and noise, and improve data interpretation. The spectral pre-processing methods often used are scattering correction and derivatives. The scattering correction can be performed by Multiplicative Scatter Correction (MSC), Standard Normal Variate (SNV) or normalization. Usually, this issue is faced when there is difference in particle size in the sample to be analyzed, and this effect can generate baseline variations. The derivatives are also able to reduce this effect by smoothing the spectra and; moreover, correct overlapping bands (Lohumi et al., 2015; Manley, 2014). Another commonly used method is the Mean Centring (MC), which averages the spectrum and subtract it from the original one. Consequently, these methods allow to minimize the sources of unwanted data variability, such as scattering, facilitating data interpretation (Porep et al., 2015).

3.20.7 Linear Methods

3.20.7.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is the exploratory method of analysis most used in chemometrics for pattern recognition. PCA is an unsupervised technique that allows the reduction of the original data into new variables, named Principal Components (PCs), which provide most of the information existent in the original data. Therefore, these new variables are a linear combination of the original data projected in a smaller dimension (Berrueta et al., 2007).

In general, the first PC holds the greatest variability between the samples, followed by the second PC, and so on, until all the variance is explained. Consequently, in order to choose a suitable number of PCs, it is ideal to analyze the percentage of variance of each PC and the cumulative percentage. Then, it is possible to better investigate the relationship between the groups of samples, where this relationship may be based, and, also, to identify outliers (Berrueta et al., 2007; Manley, 2014).

The results of a PCA are basically represented in two plots: scores and loadings. The scores plot gives the information regarding samples, their behavior, clustering, and anomalous samples. On the other hand, the loadings plot provides the information about the variables, and their contribution in separating the samples (Bona et al., 2018).

3.20.7.2 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) appears to be the most used supervised qualitative discrimination method. As PCA, this is a pattern recognition method that seeks dimensionality reduction by defining a smaller dimension of the hyperplane and projecting the original higher dimension data in this new plan. However, unlike PCA, this projection is performed by calculating the variance between classes and within a class. Thus, a linear projection is constructed by maximizing the variance between-class and minimizing the variance within-class (Abdolmaleki et al., 2015; Sendin et al., 2018). The new variables generated by LDA, called latent variables (LV) are, then, a linear combination of the original variables (Berrueta et al., 2007).

The reliability of an LDA model is based on parameters of sensitivity and specificity. Sensitivity represents the model capacity in correctly classify the samples to the interest class, whereas the specificity expresses the model capacity in correctly classify the samples when they do not belong to the interest class (Su and Sun, 2016). Still, it is worth mentioning that, as a discriminant analysis, the samples will always be classified into one of the given classes, what may be a disadvantage if the sample does not belong to any class. In addition, a constraint of LDA models is the number of variables that should not exceed the number of samples, being necessary the variable selection prior development of the models (Berrueta et al., 2007). Commonly, LDA is processed after a PCA analysis, using the scores of a few significant principal components as the reduced set of variables, as studied by (De Girolamo et al., 2019).

NIR spectroscopy combined with LDA have been used in the discrimination of pure and adulterated meat samples (Kuswandi et al., 2015; Nolasco-Perez et al., 2019; Rady and Adedeji, 2018a); and honey (Aliaño-González et al., 2019; Ferreira-González et al., 2018).

3.20.7.3 Soft Independent Modeling by Class Analogy (SIMCA)

Soft independent modeling by class analogy (SIMCA) is a supervised qualitative method in multivariate analysis. Unlike LDA, SIMCA is not only a discrimination, but a classification method, that is, SIMCA is able to classify samples as belonging to one class, more than one or neither class. SIMCA model is built based on principal component analysis, with one PCA performed for each class. Each PCA is individually developed and evaluated, with the appropriate number of principal components chosen according to the minimum cross-validation error. Moreover, Q-statistic and Hotelling T^2 are evaluated to set the boundary of each class (Badaró et al., 2019; Rodríguez et al., 2019).

Furthermore, SIMCA calculates the class distance, discriminatory power and modeling power. The class distance can be determined by the geometric distance from the principal component models. However, another concept is commonly used, which consider the boundary as a region of space with a 95% confidence level that a sample belongs to a class. The discriminatory power is determined as the ability of a variable to discriminate two classes, whereas the modeling power describes how the variable contribute to the model in classifying samples (Berrueta et al., 2007; Brereton, 2003).

The results of SIMCA models can be interpreted by observing the Coomans plot. This graph evaluates the distance between two classes and it is divided in four regions (with a confidence level of 95%): one representing samples from class 1, one for class 2, one for overlapping samples and one for those samples that do not belong to any class. In addition, the performance of SIMCA models can be evaluated as mentioned for LDA models, based on parameters of sensitivity and specificity (Badaró et al., 2019; Berrueta et al., 2007).

SIMCA can define the class of an authentic food in an untargeted way, exhaustively considering the variability of the authentic food during the modeling stage. Therefore, this approach can lead to a more robust classification, as any change from several origin of the food composition, detectable by the used analytical technique, can be potentially identified (Brereton, 2003).

SIMCA have been applied in NIR data to classification of pure and adulterated milk powder (Capuano et al., 2015); hazelnut (Márquez et al., 2016); and edible oil (Basri et al., 2017; Vanstone et al., 2018).

3.20.7.4 Partial Least Squares (PLS) and Partial Least Squares Discriminant Analysis (PLS-DA)

Partial Least Squares (PLS) is the most known and used quantitative method in chemometrics. PLS is a supervised method that establish a relationship between the dependent (Y) and independent (X) variables. The matrix X can be spectral information, whereas the matrix Y is a parameter, as protein content or percentage of adulteration. During the development of a PLS model, those independent variables that most contributed to describe the dependent variables are considered. Then, the software algorithm gives more importance or weight to the variables that are more correlated to the output and less importance to those that are irrelevant or carry noise. So, PLS is a reduction method that defines the latent variables (LV) (or factors) that well predict the parameter of interest (Berrueta et al., 2007; Porep et al., 2015).

These LV are a linear combination of the original variables that maximize the covariance between the matrices of dependent and independent variables. The first few LV carry most of the relevant information necessary for the prediction, and an appropriate number of factors must be selected in order to obtain a good performance of the model. If it is selected fewer samples than needed, the model will not cover all the relevant information, while more samples than needed will generate a noisy model. In both cases, the models developed will result in a poor prediction of the parameter of interest. Therefore, a wise choice of the number of factor together with similar prediction error is fundamental. The cross-validation in calibration models plays an important role in selecting the suitable number of latent variables (Porep et al., 2015).

The performance of PLS models is evaluated based on the coefficients of determination (R^2), and root mean square errors (RMSE) of cross-validation (R^2_{CV} , RMSECV) and prediction (R^2_p , RMSEP). The R^2 , represents the fraction of variance of the property or concentration explained by the spectral data variability and ranges from 0 to 1 (with 1 indicating 100% explanation). Then, a value close to one is desired to have a good performance of the model. R^2 is heavily dependent of the range and distribution of the values of the property or concentrations, and must be reported with care to avoid misinterpretation of the model quality. On the other hand, RMSE estimates the difference between the reference and predicted values. In the development of calibration models, samples are selected according to diverse criteria and left out of the modeling stage. Equations are developed in order to predict those samples. This is called cross-validation, and this process is conducted until all the samples in the dataset is validated. The RMSECV is, then, calculated in relation to reference and predict values from cross-validation. In contrast, the RMSEP estimates the difference between the experimental and predict values of an independent dataset (external validation) (Porep et al., 2015).

Author's personal copy

238 Food Quality and NIR Spectroscopy in the Omics Era

NIR coupled to PLS have been applied to quantify the level of adulteration in meat (Kuswandi et al., 2015; Nolasco-Perez et al., 2019); milk (Capuano et al., 2015; Mabood et al., 2017); coffee (Correia et al., 2018; Winkler-Moser et al., 2015); starch (Ma et al., 2017); oil (Basri et al., 2017; Farres et al., 2019); butter (Mabood et al., 2018); spices (Kar et al., 2018); honey (Aliaño-González et al., 2019; Ferreiro-González et al., 2018; Kumaravelu and Gopal, 2015); flour (Faith et al., 2019); and cocoa powder (Quelal-Vásquez et al., 2018).

Another approach is a PLS-based method called Partial Least Squares Discriminant Analysis (PLS-DA). The principals of PLS-DA do not differ from PLS, where a model is developed to relate a matrix X of independent variables with a matrix Y of dependent variables. However, in a PLS discriminant analysis, the matrix Y is a class, and the optimal number of factors (also determined by cross-validation) is selected in order to discriminate samples in the dataset. In contrast to the discriminant methods mentioned before, this method discriminates only two classes at a time. Therefore, a new Y matrix is created typically assigning the value 1 for one class and 0 for all the others. An interesting point is that, in PLS-DA, the information regarding matrices X and Y are considered. In other words, the scores from matrix Y are used to calculate the loadings of matrix X and vice versa. Thus, a model is built for each class (Berrueta et al., 2007).

In this context, if the value predict for the unknown sample is close to 1, it is more likely the sample belongs to the studied class, and 0 otherwise. The coefficients of determination (R^2) and root mean square errors (RMSE) are considered to evaluate the model performance, though, as discriminant method, the parameters of sensitivity and specificity defines the effectiveness of the model (Berrueta et al., 2007).

PLS-DA have been applied in NIR data for discrimination between pure and adulterated samples of milk (Mabood et al., 2017); meat (Rady and Adedeji, 2018b); butter (Mabood et al., 2018); oil (Jiménez-Carvelo et al., 2019b); cocoa powder (Quelal-Vásquez et al., 2018); starch (Ma et al., 2017). This method also was applied in the discrimination of sea bass samples according to their production method, stocking density and geographical origin (Chidini et al., 2019); characterization and authentication of the wine vinegar (Rios-Reina et al., 2019; Rios-Reina et al., 2018); prediction of aging time of beef steaks (Moran et al., 2018); authentication of organic milk (Liu et al., 2018); and discrimination between organic and conventional rice (Xiao et al., 2019).

3.20.8 Non-linear Methods

3.20.8.1 Neural Networks (NN)

Artificial neural network (ANN) or simply "neural networks" is a nonlinear learning method of Machine learning (ML). ANNs are parallel structures composed by processing units called neurons (or nodes), which have the ability to perform the data processing. A neural network contains an input layer, composed by neurons structured in parallel and representing the independent variables (one neuron for each variable), hidden layers (one or more) that create the nonlinear relationship between the layers, and an output layer, composed by the dependent variables (one neuron for each parameter). Basically, the input variables provide the information to the hidden layers, which use it for the learning process. Then, the ANN algorithm finds the weights of each neuron that better predict the output feature (Berrueta et al., 2007; Haykin, 2010). Fig. 2 schematizes how an ANN is disposed.

Usually, the most common ANNs are Kohonen, counter propagation, radial basis function and probabilistic neural networks (PNN); however, the most used is back-propagation ANN. In this approach, the information goes through the network via synapses; then, the ANN algorithm predicts the results and compare them to the expect values. After that, the error of prediction is calculated and propagated backwards from the output to the input layers, passing through all the hidden layers. This error is used by the algorithm to correct the weight of each neuron in each layer according to the previously one. This learning process is repeated until all the weights are able to predict the output with high precision (Berrueta et al., 2007).

The learning process of a network initiates with random values of weights, and the parameters used must be well selected. The number of hidden layers and training cycles (epochs) are tested until the error of prediction is minimum. This test is based on trial and error, and in the prediction accuracy of the model. A large number of nodes in the hidden layers, for example, can over fit the model, which decreases the calibration error, but does not improve the validation. On the other hand, if fewer hidden nodes than necessary are chosen, the algorithm will not be able to accurately predict the outputs (Berrueta et al., 2007; Haykin, 2010).

Usually, one hidden layer is enough for learning processing; in despite of cases where there are discontinuities. In this case, two hidden layers may be indicated. It is still worth to mention that it is not necessary to know deeply the relationship between input and output variables, since they will learn this relation through the training process. However, the data must be always balanced to avoid bias. Furthermore, ANN can be used in prediction or classification tasks. Then, the model performance is based on root mean square error and the determination coefficient of prediction for regression analysis, and the model ability in classify correct or incorrectly the samples for classification analysis (Berrueta et al., 2007). Feedforward Artificial Neural Networks coupled to NIR data have been used to classify pure and adulterated samples of meat (Rady and Adedeji, 2018b).

3.20.8.2 Deep Learning (DL)

Deep learning is another tool of machine learning method which is based on neural networks. In comparison to ANN, deep learning has more hidden layers, deals with a lot more data, and has stronger learning ability. Therefore, based on these features, DL is able to obtain very great results (Zhang et al., 2019).

Author's personal copy

Food Quality and NIR Spectroscopy in the Omics Era 239

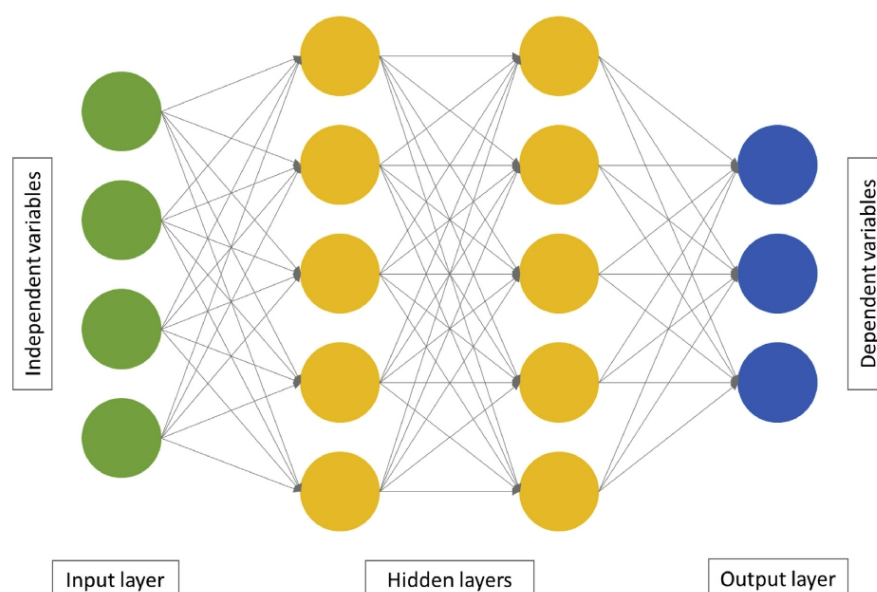


Figure 2 – Scheme of a neural network with inputs, output and hidden layers.

Deep learning uses the hidden layers with integrated data and passes the information from one layer to another until the suitable prediction. In DL, the prediction is done without many information from the object to be studied, but through training. Based on this, the machine finds more improved patterns and is able to learn complicated functions (Navarro, 2019).

The omics science involves a huge amount of data, including genome, transcript and proteome data. Moreover, to solve some issues, it is necessary to combine different source of data, what makes much more difficult to extracted relevant information and build reliable models. Therefore, deep learning is a good alternative to handle these complex data and obtain more accurate results (Zhang et al., 2019).

3.20.9 Support Vector Machine (SVM)

Support Vector Machine is another feedforward network used in linear or nonlinear cases. This supervised method is based on a learning algorithm that separates a dataset by a hyperplane and its application includes classification or regression (Haykin, 2010). This hyperplane is constructed by a kernel method which is an algorithm that seeks to minimize the classification error and maximize the separation margin between samples. The margin separation represents the maximum this limit can extend before meeting the closest samples in the dataset. These samples are known as support vectors (Mahadevan et al., 2008).

The kernel algorithms will transform the data according to their behavior. That means, if the input data and the expected response have a linear relationship and a hyperplane can be designed in the original space, the SVM model uses a linear kernel. Otherwise, if the data do not have a linear relationship, they are projected in a higher dimension space by the inner product kernel between these support vectors and a vector of the original data. Thus, a linear relation is created in this new higher space, also known as a kernel trick. Then, SVM methods can deal with higher complex and nonlinear data if compared to other multivariate methods such as PCA, PLSR or PLS-DA (Jiménez-Carvelo et al., 2019a; Mahadevan et al., 2008).

3.20.10 Random Forests (RFs)

Beyond the machine learning methods aforementioned, Random Forests have also been widely used as supervised method for data analysis. However, there is not many works regarding RFs in the food field as for SVM (Xu and Sun, 2018). RF is one of the decision tree (DT) methods used for pattern recognition. A DT consists of an initial node, internal nodes or branches, which represent the

Author's personal copy

240 Food Quality and NIR Spectroscopy in the Omics Era

characteristics used to classify the samples into their respective classes, and end nodes or leaves, which are the final labels (or classes) (Jiménez-Carvelo et al., 2019a; Kucheryavskiy, 2018).

This method separates the samples into random subsets based on the variables, and predicts them according to the suitable outcome. The highest rated result creates additional nodes that are branched into other possibilities until the final determination, thus generating a tree-shape structure, as the name suggest (Ai et al., 2014; Jiménez-Carvelo et al., 2019a).

In DT, the subsets are selected by using boosting or bagging, which increases the model prediction power. Their main difference is the iteration: while boosting weights the outcome in each node and learns successive models according to the misclassified samples, bagging creates different models for each sub-dataset. The RF is one of the most known bagging methods (Jiménez-Carvelo et al., 2019a). Therefore, since the calibration models of each subset are random and do not contain all the samples, those not included in the model are used as an internal validation set, dismissing the need of a cross-validation (Kucheryavskiy, 2018). Consequently, the RF method generates classification parameters for each subset model, as accuracies and out-of-bag (OOB) errors. Thus, the final predicted class is determined by averaging all the parameters (Xu and Sun, 2018).

3.20.11 Data Fusion

In this growing search for new strategies that will improve the food quality and authentication, emerged the data fusion technique. Fusion of data together with chemometrics have shown a great power in predict more reliable results. However, Borràs et al. (2015) point out the fact that this strategy have been used for more time that it seems, since single chemical parameters evaluated by different analysis have been combined in order to improve food authentication. Nowadays, though, the challenge is to fuse big amount of complementary data rather than individual data.

In this context, data fusion has been used in the combination of different techniques, such as NIR, MIR and Raman spectroscopy, and liquid or mass chromatography. This strategy can be applied in three levels: low-, mid-, and high-level (Cuevas et al., 2017).

In low-level data fusion, the original data from both techniques are pre-processed and, then, combined in one single matrix for further multivariate analysis and extraction of relevant information. Even though the great advantage of this strategy in straightforward results, it also has two main disadvantages. First, a scaling step is necessary to balance the variance in each dataset to make them compatibles; and second, as usually these datasets carry a great amount of irrelevant information, there is a big risk in increasing the number of insignificant variables in the model. In mid-level data fusion, optimum variables are first selected from the original and unprocessed data to, then, be subjected to multivariate analysis. This strategy is more used with the aim at prediction rather than interpretation. On the other hand, high-level data fusion evaluates each data individually and, then, the models obtained by each technique are merged (Bevilacqua et al., 2017; Chen et al., 2017; Oliveira et al., 2019). Therefore, this approach combines the data of different techniques, and in different levels, in order to generate more accurate classifications and predictions than an individual technique (Chen et al., 2017).

The biological system is characterized by all the levels analyzed in omics science, including proteins, metabolites and DNA. Therefore, more than one platform can be used in order to have an idea of the whole system. In this regard, the use of data fusion from different techniques in the omics emerges as an alternative for improving the identification of biomarkers or the knowledge about the metabolic pathways. In addition, the fusion of data in omics science integrate genes, proteins and metabolites, giving an overview of a system based on more reliable information. Even though the use of high-level data fusion has constantly increased in omics science, mid-level is more used, especially in food authentication (Bevilacqua et al., 2017).

High- and mid-level data fusion were used to verify the combination of FT-Raman and NIR in the detection of hazelnut adulteration (Márquez et al., 2016). Mid-level data fusion, and Common Component and Specific Weights analysis multiblock method were applied in four techniques, including Fourier-transform mid-infrared spectroscopy (MIR), near infrared spectroscopy (NIR), multi-dimensional fluorescence spectroscopy (EEM) and proton nuclear magnetic resonance (¹H-NMR) for characterization and classification of wine vinegars with protected designations of origin (Ríos-Reina et al., 2019).

References

- Abbas, O., Zadravec, M., Baeten, V., Mikuš, T., Lesić, T., Vulić, A., et al., 2018. Analytical methods used for the authentication of food of animal origin. *Food Chem.* 246, 6–17. <https://doi.org/10.1016/j.foodchem.2017.11.007>.
- Abdolmaleki, A., Ghasemi, J., Shiri, F., Pirhadi, S., 2015. Application of multivariate linear and nonlinear calibration and classification methods in drug design. *Comb. Chem. High Throughput Screen.* 18 (8), 795–808. <https://doi.org/10.2174/1386207318666150803142158>.
- Abernethy, G., Higgs, K., 2013. Lactose semicarbazone as a marker for semicarbazide adulteration in milk. *J. Chromatogr. A* 1295, 152–155. <https://doi.org/10.1016/j.chroma.2013.03.075>.
- Ai, F.F., Bin, J., Zhang, Z.M., Huang, J.H., Wang, J.B., Liang, Y.Z., et al., 2014. Application of random forests to select premium quality vegetable oils by their fatty acid composition. *Food Chem.* 143, 472–478. <https://doi.org/10.1016/j.foodchem.2013.08.013>.
- Allaño-González, M.J., Ferrero-González, M., Espada-Bellido, E., Palma, M., Barbero, G.F., 2019. A screening method based on Visible-NIR spectroscopy for the identification and quantification of different adulterants in high-quality honey. *Talanta* 203, 235–241. <https://doi.org/10.1016/j.talanta.2019.05.067>.
- Azadmard-Damirchi, S., 2010. Review of the use of phytosterols as a detection tool for adulteration of olive oil with hazelnut oil. *Food Addit. Contam. Part A Chem. Anal. Control. Expo. Risk Assess.* 27 (1), 1–10. <https://doi.org/10.1080/02652030903225773>.
- Badaró, A.T., Morimitsu, F.L., Ferreira, A.R., Clerici, M.T.P.S., Fernandes Barbin, D., 2019. Identification of fiber added to semolina by near infrared (NIR) spectral techniques. *Food Chem.* 289, 195–203. <https://doi.org/10.1016/j.foodchem.2019.03.057>.

Author's personal copy

Food Quality and NIR Spectroscopy in the Omics Era 241

- Barreto, A., Cruz-Tirado, J.P., Siche, R., Quevedo, R., 2018. Determination of starch content in adulterated fresh cheese using hyperspectral imaging. *Food Biosci.* 21, 14–19. <https://doi.org/10.1016/j.food.2017.10.009>.
- Basri, K.N., Hussain, M.N., Bakar, J., Sharif, Z., Khir, M.F.A., Zoolfakar, A.S., 2017. Classification and quantification of palm oil adulteration via portable NIR spectroscopy. *Spectrochim. Acta Mol. Biomol. Spectrosc.* 173, 335–342. <https://doi.org/10.1016/j.saa.2016.09.026>.
- Bernueta, L.A., Alonso-Salces, R.M., Héberger, K., 2007. Supervised pattern recognition in food analysis. *J. Chromatogr. A* 1158 (1–2), 196–214. <https://doi.org/10.1016/j.chroma.2007.05.024>.
- Bevilacqua, M., Bro, R., Marini, F., Rinnan, Å., Rasmussen, M.A., Skov, T., 2017. Recent chemometrics advances for foodomics. *Trac. Trends Anal. Chem.* 96, 42–51. <https://doi.org/10.1016/j.trac.2017.08.011>.
- Böhme, K., Calo-Mata, P., Barros-Velázquez, J., Ortea, I., 2019. Recent applications of omics-based technologies to main topics in food authentication. *Trac Trends Anal. Chem.* 110, 221–232. <https://doi.org/10.1016/j.trac.2018.11.005>.
- Bona, E., Marçó, P.H., Valderrama, P., 2018. Chemometrics applied to food control. In: *Food Control and Biosecurity*, vol. 16, pp. 105–133. <https://doi.org/10.1016/B978-0-12-811445-2.00004-0>.
- Borrás, E., Ferrá, J., Boqué, R., Mestres, M., Acedia, L., Busto, O., 2015. Data fusion methodologies for food and beverage authentication and quality assessment - a review. *Anal. Chim. Acta* 891, 1–14. <https://doi.org/10.1016/j.aca.2015.04.042>.
- Bouzemrak, Y., Marvin, H.J.P., 2016. Prediction of food fraud type using data from Rapid Alert System for Food and Feed (RASFF) and Bayesian network modelling. *Food Control* 61, 180–187. <https://doi.org/10.1016/j.foodcont.2015.09.026>.
- Brereton, R.G., 2003. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, vol. 8. John Wiley & Sons, Ltd, England.
- Capuano, E., Boerigter-Enling, R., Koot, A., van Ruth, S.M., 2015. Targeted and untargeted detection of skim milk powder adulteration by near-infrared spectroscopy. *Food Anal. Methods* 8 (8), 2125–2134. <https://doi.org/10.1007/s12161-015-0100-3>.
- Chen, J., Ye, F., Zhao, G., 2017. Rapid determination of farinograph parameters of wheat flour using data fusion and a forward interval variable selection algorithm. *Anal. Methods* 9 (45), 6341–6348. <https://doi.org/10.1039/c7ay02065a>.
- Correia, R.M., Tosato, F., Domingos, E., Rodrigues, R.R.T., Aquino, L.F.M., Figueiras, P.R., et al., 2018. Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta* 176, 59–68. <https://doi.org/10.1016/j.talanta.2017.08.009>.
- Cuevas, F.J., Pereira-Caro, G., Moreno-Rojas, J.M., Muñoz-Redondo, J.M., Ruiz-Moreno, M.J., 2017. Assessment of premium organic orange juices authenticity using HPLC-HR-MS and HS-SPME-GC-MS combining data fusion and chemometrics. *Food Control* 82, 203–211. <https://doi.org/10.1016/j.foodcont.2017.06.031>.
- Curzon, A.Y., Chandrasekhar, K., Nashif, Y.K., Abbo, S., Bonfil, D.J., Reifen, R., et al., 2019. Distinguishing between bread wheat and spelt grains using molecular markers and spectroscopy. *J. Agric. Food Chem.* 67 (13), 3837–3841. <https://doi.org/10.1021/acs.jafc.9b00131>.
- De Girolamo, A., von Holst, C., Cortese, M., Cervellieri, S., Pascale, M., Longobardi, C.L., Porricelli, A.C.R., Lippolis, V., 2019. Rapid screening of ochratoxin A in wheat by infrared spectroscopy. *Food Chem.* 282, 95–100. <https://doi.org/10.1016/j.foodchem.2019.01.008>.
- De Jonge, J., Frewer, L., Van Trijp, H., Jan Renes, R., de Wit, W., Timmers, J., 2004. Monitoring consumer confidence in food safety: an exploratory study. *Br. Food J.* 106, 837–849. <https://doi.org/10.1108/00070700410561423>.
- Deidda, R., Sacre, P.Y., Clavud, M., Coïc, L., Avohou, H., Hubert, P., Ziemons, E., 2019. Vibrational spectroscopy in analysis of pharmaceuticals: critical review of innovative portable and handheld NIR and Raman spectrophotometers. *Trac Trends Anal. Chem.* 114, 251–259. <https://doi.org/10.1016/j.trac.2019.02.035>.
- Dvash, L., Afk, O., Shafir, S., Schaffer, A., Yeselson, Y., Dag, A., Landau, S., 2002. Determination by near-infrared spectroscopy of perseitol used as a marker for the botanical origin of avocado (*Persea americana* Mill.) honey. *J. Agric. Food Chem.* 50 (19), 5283–5287. <https://doi.org/10.1021/jf020329z>.
- Eisenstecken, D., Stürz, B., Robatscher, P., Lozano, L., Zanella, A., Oberhuber, M., 2019. The potential of near infrared spectroscopy (NIRS) to trace apple origin: study on different cultivars and orchard elevations. *Postharvest Biol. Technol.* 147 (May 2018), 123–131. <https://doi.org/10.1016/j.postharbio.2018.08.019>.
- Esslinger, S., Riedl, J., Fahl-Hassek, C., 2014. Potential and limitations of non-targeted fingerprinting for authentication of food in official control. *Food Res. Int.* 60, 189–204. <https://doi.org/10.1016/j.foodres.2013.10.015>.
- Faith, P., Lembe, N., Magwaza, S., Zeray, S., 2019. Rapid visible – near infrared (Vis – NIR) spectroscopic detection and quantification of unripe banana flour adulteration with wheat flour. *J. Food Sci. Technol.* 56 (4).
- Farres, S., Srata, L., Fethi, F., Kadaoui, A., 2019. Argan oil authentication using visible/near infrared spectroscopy combined to chemometrics tools. *Vib. Spectrosc.* 102, 79–84. <https://doi.org/10.1016/j.vibspec.2019.04.003>.
- Fernández Piñera, J.A., Vermeulen, P., Amand, O., Tossens, A., Dardenne, P., Baeten, V., 2012. NIR hyperspectral imaging spectroscopy and chemometrics for the detection of undesirable substances in food and feed. *Chemometr. Intell. Lab. Syst.* 117, 233–239. <https://doi.org/10.1016/j.chemolab.2012.02.004>.
- Ferreiro-González, M., Espada-Bellido, E., Guillén-Cuelo, L., Palma, M., Barroso, C.G., Barbero, G.F., 2018. Rapid quantification of honey adulteration by visible-near infrared spectroscopy combined with chemometrics. *Talanta* 188, 288–292. <https://doi.org/10.1016/j.talanta.2018.05.095>.
- Ferrì, E., Galimberti, A., Casiraghi, M., Airolidi, C., Ciaramelli, C., Palmioli, A., et al., 2015. Towards a universal approach based on omics technologies for the quality control of food. *Biomed Res. Int.* 2015, 365794. <https://doi.org/10.1155/2015/365794>.
- Gallo, M., Ferranti, P., 2016. The evolution of analytical chemistry methods in foodomics. *J. Chromatogr. A* 1428, 3–15. <https://doi.org/10.1016/j.chroma.2015.09.007>.
- Galvin-King, P., Haughey, S.A., Elliott, C.T., 2018. Herb and spice fraud: the drivers, challenges and detection. *Food Contr.* 88, 85–97. <https://doi.org/10.1016/j.foodcont.2017.12.031>.
- Geng, F., Wang, J., Liu, D., Jin, Y., Ma, M., 2017. Identification of N-glycosites in chicken egg white proteins using an omics strategy. *J. Agric. Food Chem.* 65 (26), 5357–5364. <https://doi.org/10.1021/acs.jafc.7b01706>.
- Ghidini, S., Varà, M.O., Dall'Asta, C., Badiani, A., Ianieri, A., Zanardi, E., 2019. Rapid authentication of European sea bass (*Dicentrarchus labrax* L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics. *Food Chem.* 280, 321–327. <https://doi.org/10.1016/j.foodchem.2018.12.075>.
- Han, Y., Wang, X., Liu, Y., Han, L., Yang, Z., Liu, X., 2019. A novel FTIR discrimination based on genomic DNA for species-specific analysis of meat and bone meal. *Food Chem.* 294, 526–532. <https://doi.org/10.1016/j.foodchem.2019.05.088>.
- Haykin, S., 2010. *Neural Networks and Micromechanics*. Neural Networks and Micromechanics, third ed. Pearson. <https://doi.org/10.1007/978-3-642-02535-8>.
- Jiménez-Carvelo, A.M., González-Casado, A., Bagur-González, M.G., Cuadros-Rodríguez, L., 2019a. Alternative data mining/machine learning methods for the analytical evaluation of food quality and authenticity – a review. *Food Res. Int.* 122, 25–39. <https://doi.org/10.1016/j.foodres.2019.03.063>.
- Jiménez-Carvelo, A.M., Lozano, V.A., Olivieri, A.C., 2019b. Comparative chemometric analysis of fluorescence and near infrared spectroscopies for authenticity confirmation and geographical origin of Argentinean extra virgin olive oils. *Food Control* 96, 22–28. <https://doi.org/10.1016/j.foodcont.2018.08.024>.
- Johnson, R., 2014. Food Fraud and “Economically Motivated Adulteration” of Food and Food Ingredients. Retrieved from: www.crs.gov/R43358.
- Kar, S., Tudu, B., Bag, A.K., Bandyopadhyay, R., 2018. Application of near-infrared spectroscopy for the detection of metanil yellow in turmeric powder. *Food Analytical Methods* 11 (5), 1291–1302. <https://doi.org/10.1007/s12161-017-1106-9>.
- Kemsley, E.K., Defezze, M., Marini, F., 2019. Multivariate statistics: considerations and confidences in food authenticity problems. *Food Control* 105, 102–112. <https://doi.org/10.1016/j.foodcont.2019.05.021>.
- Kim, J.M., Kim, H.J., Park, J.M., 2015. Determination of milk fat adulteration with vegetable oils and animal fats by gas chromatographic analysis. *J. Food Sci.* 80 (9), 1945–1951. <https://doi.org/10.1111/1750-3841.12979>.
- Kucheravskiy, S., 2018. Analysis of NIR spectroscopic data using decision trees and their ensembles. *J. Anal. Test.* 2 (3), 274–289. <https://doi.org/10.1007/s41664-018-0078-0>.

Author's personal copy

242 Food Quality and NIR Spectroscopy in the Omics Era

- Kumaravelu, C., Gopal, A., 2015. Detection and quantification of adulteration in honey through near infrared spectroscopy. *Int. J. Food Prop.* 18 (9), 1930–1935. <https://doi.org/10.1080/10942912.2014.919320>.
- Kuswandi, B., Cendekiawan, K.A., Kristiningrum, N., Ahmad, M., 2015. Pork adulteration in commercial meatballs determined by chemometric analysis of NIR Spectra. *J. Food Meas. Charact.* 9 (3), 313–323. <https://doi.org/10.1007/s11694-015-9238-3>.
- Leme, L.M., Nakamura, F., Coelho Tanamati, A.A., Valderrama, P., Março, P.H., 2019. Fast non-invasive screening to detect fraud in oil capsules. *LWT* 109, 179–185. <https://doi.org/10.1016/j.lwt.2019.03.088>.
- Liebler, D.C., Zimmerman, L.J., 2013. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* 52 (22), 3797–3806. <https://doi.org/10.1021/bi400110b>.
- Liu, N., Parra, H.A., Pustjens, A., Hettling, K., Mongondry, P., van Ruth, S.M., 2018. Evaluation of portable near-infrared spectroscopy for organic milk authentication. *Talanta* 184, 128–135. <https://doi.org/10.1016/j.talanta.2018.02.097>.
- Lohumi, S., Lee, S., Lee, H., Cho, B.-K., 2015. A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends Food Sci. Technol.* 46 <https://doi.org/10.1016/j.tifs.2015.08.003>.
- Ma, H.L., Wang, J.W., Chen, Y.J., Cheng, J.L., Lai, Z.T., 2017. Rapid authentication of starch adulterations in ultrafine granular powder of Shanyao by near-infrared spectroscopy coupled with chemometric methods. *Food Chem.* 215, 108–115. <https://doi.org/10.1016/j.foodchem.2016.07.156>.
- Mabood, F., Abbas, G., Jabeen, F., Naureen, Z., Al-Harasi, A., Hamaed, A.M., et al., 2018. Robust near NIRS coupled with multivariate methods for the detection and quantification of talow adulteration in clarified butter samples. *Food Addit. Contam. Part A Chem. Anal. Control. Expo. Risk Assess.* 35 (3), 404–411. <https://doi.org/10.1080/19440049.2017.1418090>.
- Mabood, F., Jabeen, F., Ahmed, M., Hussain, J., Al Mashaykhi, S.A.A., Al Rubaiey, Z.M.A., et al., 2017. Development of new NIR-spectroscopy method combined with multivariate analysis for detection of adulteration in camel milk with goat milk. *Food Chem.* 221, 746–750. <https://doi.org/10.1016/j.foodchem.2016.11.109>.
- Mahadevan, S., Shah, S.L., Marrie, T.J., Slupsky, C.M., 2008. Analysis of metabolomic data using support vector machines. *Anal. Chem.* 80 (19), 7562–7570. <https://doi.org/10.1021/ac800954c>.
- Manley, M., 2014. Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chem. Soc. Rev.* 43, 8200–8214. <https://doi.org/10.1039/c4cs00062e>.
- Márquez, C., López, M.I., Ruisánchez, I., Callao, M.P., 2016. FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud. *Talanta* 161, 80–86. <https://doi.org/10.1016/j.talanta.2016.08.003>.
- McGrath, T.F., Haughey, S.A., Patterson, J., Fahl-Hassek, C., Donarski, J., Alewijn, M., et al., 2018, June. What are the scientific challenges in moving from targeted to non-targeted methods for food fraud testing and how can they be addressed? – spectroscopy case study. *Trends Food Sci. Technol.* <https://doi.org/10.1016/j.tifs.2018.04.001>.
- Moran, L., Andres, S., Allen, P., Moloney, A.P., 2018. Visible and near infrared spectroscopy as an authentication tool: preliminary investigation of the prediction of the ageing time of beef steaks. *Meat Sci.* 142, 52–58. <https://doi.org/10.1016/j.meatsci.2018.04.007>.
- Navarro, C., 2019. How to train your genome. *Cell.* <https://doi.org/10.1016/j.cell.2019.03.003>.
- Nolasco-Perez, I.M., Rocco, L.A.C.M., Cruz-Tirado, J.P., Pollonio, M.A.R., Barbon, S., Barbon, A.P.A.C., Barbin, D.F., 2019. Comparison of rapid techniques for classification of ground meat. *Biosyst. Eng.* 183, 151–159. <https://doi.org/10.1016/j.biosystemseng.2019.04.013>.
- Oliveira, M.M., Cruz-Tirado, J.P., Barbin, D.F., 2019. Nontargeted analytical methods as a powerful tool for the authentication of spices and herbs: a review. *Compr. Rev. Food Sci. Food Saf.* 18 (3), 670–689. <https://doi.org/10.1111/1541-4337.12436>.
- Ortea, I., O'Connor, G., Maquet, A., 2016. Review on proteomics for food authentication. *J. Proteomics* 147, 212–225. <https://doi.org/10.1016/j.jprot.2016.06.033>.
- Pasquini, C., 2003. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *J. Braz. Chem. Soc.* 14, 198–219.
- Pasquini, C., 2018. Near infrared spectroscopy: a mature analytical technique with new perspectives - a review. *Anal. Chim. Acta* 1026, 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>.
- Perep, J.U., Kammerer, D.R., Carle, R., 2015. On-line application of near infrared (NIR) spectroscopy in food production. *Trends Food Sci. Technol.* 46 (2), 211–230. <https://doi.org/10.1016/j.tifs.2015.10.002>.
- Pustjens, A.M., Weesepoel, Y., van Ruth, S.M., 2015. Food fraud and authenticity: emerging issues and future trends. In: *Innovation and Future Trends in Food Manufacturing and Supply Chain Technologies*, pp. 3–20. <https://doi.org/10.1016/B978-1-78242-447-5.00001-0>.
- Quelal-Vásconez, M.A., Pérez-Estève, E., Arnaú-Bonachera, A., Barat, J.M., Talens, P., 2018. Rapid fraud detection of cocoa powder with carob flour using near infrared spectroscopy. *Food Contr.* 92, 183–189. <https://doi.org/10.1016/j.foodcont.2018.05.001>.
- Rachmawati, Rohaeti, E., Rafi, M., 2017. Combination of near infrared spectroscopy and chemometrics for authentication of taro flour from wheat and sago flour. In: *Journal of Physics: Conference Series*, vol. 835. <https://doi.org/10.1088/1742-6596/835/1/012011>.
- Rady, A., Adediji, A., 2018a. Assessing different processed meats for adulterants using visible-near-infrared spectroscopy. *Meat Sci.* 136, 59–67. <https://doi.org/10.1016/j.meatsci.2017.10.014>.
- Rady, A., Adediji, A., 2018b. Assessing different processed meats for adulterants using visible-near-infrared spectroscopy. *Meat Sci.* 136, 59–67. <https://doi.org/10.1016/j.meatsci.2017.10.014>.
- Ríos-Reina, R., Callejón, R.M., Savorani, F., Amigo, J.M., Cocchi, M., 2019. Data fusion approaches in spectroscopic characterization and classification of PDO wine vinegars. *Talanta* 198, 560–572. <https://doi.org/10.1016/j.talanta.2019.01.100>.
- Ríos-Reina, R., García-González, D.L., Callejón, R.M., Amigo, J.M., 2018. NIR spectroscopy and chemometrics for the typification of Spanish wine vinegars with a protected designation of origin. *Food Contr.* 89, 108–116. <https://doi.org/10.1016/j.foodcont.2018.01.031>.
- Rodríguez, S.D., Rolandelli, G., Buera, M.P., 2019. Detection of quinoa flour adulteration by means of FT-MIR spectroscopy combined with chemometric methods. *Food Chem.* 274 (March 2018), 392–401. <https://doi.org/10.1016/j.foodchem.2018.08.140>.
- Rospal, R., 2010. Nonlinear partial least squares: an overview. In: *Cheminformatics and Advanced Machine Learning Perspectives: Complex Computational Methods and Collaborative Techniques*, pp. 169–189. <https://doi.org/10.4018/978-1-61520-911-8.ch009>.
- Sandin, K., Williams, P.J., Manley, M., 2018. Near infrared hyperspectral imaging in quality and safety evaluation of cereals. *Crit. Rev. Food Sci. Nutr.* 58 (4), 575–590. <https://doi.org/10.1080/10408398.2016.1205548>.
- Spink, J., Moyer, D.C., 2011. Defining the public health threat of food fraud. *J. Food Sci.* 76 (9), R157–R163. <https://doi.org/10.1111/j.1750-3841.2011.02417.x>.
- Su, W.-H., Sun, D.-W., 2016. Facilitated wavelength selection and model development for rapid determination of the purity of organic spelt (*Triticum spelta* L.) flour using spectral imaging. *Talanta* 155, 347–357. <https://doi.org/10.1016/j.talanta.2016.04.041>.
- Teye, E., Amuah, C.L.Y., McGrath, T., Elliott, C., 2019. Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. *Spectrochim. Acta Mol. Biomol. Spectrosc.* 217, 147–154. <https://doi.org/10.1016/j.saa.2019.03.085>.
- Van Emon, J.M., 2016. The omics revolution in agricultural research. *J. Agric. Food Chem.* <https://doi.org/10.1021/acs.jafc.5b04515>.
- van Ruth, S.M., Huisman, W., Luning, P.A., 2017. Food fraud vulnerability and its key factors. *Trends Food Sci. Technol.* 67, 70–75. <https://doi.org/10.1016/j.tifs.2017.06.017>.
- Vanstone, N., Moore, A., Martos, P., Neethirajan, S., 2018. Detection of the adulteration of extra virgin olive oil by near-infrared spectroscopy and chemometric techniques. *Food Qual. Saf.* 2 (4), 189–198. <https://doi.org/10.1093/qsaf/fyy018>.
- Velásquez, L., Cruz-Tirado, J.P., Siche, R., Quevedo, R., 2017. An application based on the decision tree to classify the marbling of beef by hyperspectral imaging. *Meat Sci.* 133, 43–50. <https://doi.org/10.1016/j.meatsci.2017.06.002>.
- Winkler-Moser, J.K., Singh, M., Pennick, K.A., Bakota, E.L., Jham, G., Liu, S.X., Vaughn, S.F., 2015. Detection of corn adulteration in Brazilian coffee (coffee arabica) by tocopherol profiling and near-infrared (NIR) spectroscopy. *J. Agric. Food Chem.* 63 (49), 10662–10668. <https://doi.org/10.1021/acs.jafc.5b04777>.

Author's personal copy

Food Quality and NIR Spectroscopy in the Omics Era 243

- Xiao, R., Li, L., Ma, Y., 2019. A label-free proteomic approach differentiates between conventional and organic rice. *J. Food Compos. Anal.* 80, 51–61. <https://doi.org/10.1016/j.jfca.2019.04.004>.
- Xiao, R., Liu, L., Zhang, D., Ma, Y., Ngadi, M.O., 2019. Discrimination of organic and conventional rice by chemometric analysis of NIR spectra: a pilot study. *J. Food Meas. Charact.* 13 (1), 238–249. <https://doi.org/10.1007/s11694-018-9937-7>.
- Xiao, R., Ma, Y., Zhang, D., Qian, L., 2018. Discrimination of conventional and organic rice using untargeted LC-MS-based metabolomics. *J. Cereal. Sci.* 82, 73–81. <https://doi.org/10.1016/j.jcs.2018.05.012>.
- Xu, J.-L., Sun, D.-W., 2018. Imaging spectroscopic technique: hyperspectral imaging. In: *Modern Techniques for Food Authentication*, pp. 253–286. <https://doi.org/10.1016/b978-0-12-814264-6.00008-6>.
- Xue, X., Wang, Q., Li, Y., Wu, L., Chen, L., Zhao, J., Liu, F., 2013. 2-acetylfuran-3-glucopyranoside as a novel marker for the detection of honey adulterated with rice syrup. *J. Agric. Food Chem.* 61 (31), 7488–7493. <https://doi.org/10.1021/jf401912u>.
- Zhang, Z., Zhao, Y., Liao, X., Shi, W., Li, K., Zou, Q., Peng, S., 2019. Deep learning in omics: a survey and guideline. *Brief. Funct. Genomics* 18 (1), 41–57. <https://doi.org/10.1093/bfpg/ely030>.

09/07/2021

RightsLink Printable License

ELSEVIER ORDER DETAILS

Jul 09, 2021

Order Number	501666460
Order date	Jul 09, 2021
Licensed Content Publisher	Elsevier
Licensed Content Publication	Elsevier Books
Licensed Content Title	Comprehensive Foodomics
Licensed Content Author	Amanda Teixeira Badaró,Celio Pasquini,Douglas Fernandes Barbin
Licensed Content Date	2021
Licensed Content Pages	13
Start Page	231
End Page	243
Type of Use	reuse in a thesis/dissertation
I am an academic or government institution with a full-text subscription to this journal and the audience of the material consists of students and/or employees of this institute?	No
Portion	full chapter
Circulation	10

09/07/2021

RightsLink Printable License

Format	both print and electronic
Are you the author of this Elsevier chapter?	Yes
How many pages did you author in this Elsevier book?	13
Will you be translating?	No
Title	NIR spectral techniques and chemometrics applied to food processing
Institution name	University of Campinas and Universitat Politécnica de València
Expected presentation date	Oct 2021
Requestor Location	Miss. Amanda Badaró Rua Pedro Geraldo Costa, 291
Publisher Tax ID	Tambaú, Sao Paulo 13710-000 Brazil Attn: Miss. Amanda Badaró
Customer VAT ID	GB 494 6272 12
Total	BR09939021623
	Not Available

APPENDIX II*Automated method for determination of cheese
meltability by computer vision*



Automated Method for Determination of Cheese Meltability by Computer Vision

Amanda Teixeira Badaró¹ · Gustavo Vechin de Matos² · Caroline Bilhar Karaziack³ · Walkiria Hanada Viotto³ · Douglas Fernandes Barbin¹

Received: 21 April 2021 / Accepted: 9 July 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Meltability is the property of a cheese to flow and spread, as well as the loss of the integrity of the cheese structure by heating, and is of great importance for cheeses used as ingredients in other products. Analytical methods for determination of cheese meltability are laborious and time-consuming, requiring heating of cheese samples followed by assessment of the dimensional changes, such as diameter and height. In this study, computer vision is proposed for rapid and accurate determination of meltability for five different types of cheese and compared to the reference Schreiber method. Digital images were acquired from the samples before and after melting, and the variation of samples was measured after segmentation and identification of the region of interest (ROI). The meltability measured by the Schreiber method varied from 0 to 45.37%, while the results determined by the computer vision method varied from 0.37 to 99.64%. The computer vision method presented high correlation with the Schreiber method, providing better results for samples that had irregular shape after melting, when compared to the traditional method. Additionally, the correlation between the manual and the automatic methods was calculated and the results showed a perfect correlation between all the computer vision methods (> 0.99). The results indicate the potential application of computer vision as a standard analytical method for determination of cheese meltability.

Keywords Schreiber method · Image analyses · Food; Process analytical technology · Mozzarella · Meltability

Introduction

Food quality, safety, and physicochemical properties are primary concern of consumers, industry, and federal regulations directly connected to the food field. Over the years, several analytical methods have been developed and validated for the determination of food composition, such as moisture, protein, fat, and carbohydrates. Most of these techniques, such as the Kjeldahl procedure for protein or Soxhlet extraction for fat determinations, are still used these days due to the easy availability of chemicals or equipment in the laboratories. However, there has been major concern regarding the fast-paced processing industry, which demands fast analytical methods with reduced need for chemicals and “green technologies” with reduced amount of residues. In addition, several food analytical techniques are laborious and time-consuming; therefore, fast and accurate techniques could help to expedite food analysis, finding applications in laboratories and in the industrial environment (Gallo and Ferranti 2016).

✉ Douglas Fernandes Barbin
dfbarbin@unicamp.br
Amanda Teixeira Badaró
tbadaro.amanda@gmail.com
Gustavo Vechin de Matos
gustavo.vdematos@gmail.com
Caroline Bilhar Karaziack
carol.karaziack@gmail.com
Walkiria Hanada Viotto
walkiyoto@unicamp.br

¹ Department of Food Engineering, School of Food Engineering, University of Campinas, Campinas, Brazil
² School of Mechanical Engineering, University of Campinas, Campinas, Brazil
³ Department of Food Technology, School of Food Engineering, University of Campinas, Campinas, Brazil

Functional properties are of great importance for cheeses used as ingredients in other products. Due to the wide range of characteristics that cheese presents, they are used for different purposes in food, according to the desired final product, such as used *in natura* or processed, in several foods like pizzas and sandwiches (Guinee and Kilcawley 2004).

The ability of a cheese to flow and spread, as well as the loss of the integrity of cheese structure by heating, is defined as cheese meltability (Lucey et al. 2003). Meltability is one of the main functional properties of cheeses, specifically when used as fillings and toppings for some processed food (Lucey 2008). In this case, the cheese must have an attractive elastic appearance after heat treatment (Altan et al. 2005). Cheese with poor meltability is tough, while excessive melting may result in undesirable texture and appearance, thus being avoided by the food industry.

The attempts to measure meltability in cheese include empirical and objective methods. The current empirical methods for determination of cheese meltability require accurate and reliable measurements methods that consist of heating cheese samples and measuring the dimensional changes, such as diameter and height (Altan et al. 2005). There are some chemical and textural differences between the cheese core and the rind, which can influence the melting and negatively affect the sensorial properties of the product. The precision of the test should also be able to differentiate between the types of cheese depending on its application, for example, Mozzarella is expected to melt when used in pizza, whereas Coalho cheese is grilled, so it is not expected to melt during heating.

One of the methods used to measure cheese meltability is Schreiber's test, which is based on the increase in sample diameter after melting (Kosikowski 1982). Although less common than Schreiber's test, the Arnott test is another method used for determination of cheese meltability. In this test, meltability is determined by difference of height between samples, in percentage, before and after melting (Wang and Sun 2002a). The precision of the Schreiber method depends on the shape of the sample and how the measurements are taken, since they must represent the whole sample.

Other objective but more complex methods are related to the rheological properties of cheese. These methods include dynamic stress rheometry (DSR), melt profile analysis (UW Meltmeter), and rapid visco analyzer (RVA). Although widely adopted, these methods have some inconsistencies, as cheese samples melt in irregular ways and several directions, thus making it difficult to achieve precise determination of meltability (Wang et al. 1998). Therefore, the Schreiber method has good repeatability and it is simpler than other more accurate methods, making it the most used in the food industry and the reference for the development

of other methods for cheese meltability (Altan et al. 2005; Calvini et al. 2020).

Recently, a significant growth has been observed both in the plurality of process analytical technologies (PAT) and in the diversity of applications for food samples (Tsakanikas et al. 2015). Computer vision is one of these techniques, where image acquisition and processing techniques allow transferring analytical methods from laboratories to processing lines. Image analysis relies on the extraction of various image features such as color, texture, shape, and size, among others, that can be related to sample attributes, finding several applications in the food industry. Some image features are used for predictors of physical attributes, grading, and classification of several food and agricultural samples (Shetty et al. 2019), such as meat (Nolasco-Perez et al. 2019), chicken (Barbin et al. 2016; Geronimo et al. 2019), cereals and grains (Lopes et al. 2019; Li et al. 2009), pasta and bread (Mastelini et al. 2018; Yin et al. 2021), fruits (Zhang et al. 2014; Santos Pereira et al. 2018; Udomkun et al. 2019), and cocoa (Oliveira et al. 2021). As a result, these resources can also be used for automated processes aimed at quality inspection through application of image processing algorithms in industrial computer vision systems (CVS) (Valous et al. 2009).

Analytical techniques based on image analysis (Wang and Sun 2002a, b, c; Minz and Saini, 2019) and spectroscopic analysis (Amamcharla and Metzger 2015; Fagan et al. 2007) have been proposed for determination of cheese properties. However, these techniques were tested for few types of cheese (Cheddar and Mozzarella), thus lacking further investigation for other types of cheeses from different regions. In addition, the correlation between results obtained by imaging techniques and analytical methods presented high levels of variation, as image analyses are highly dependent of steps adopted during image processing. Differences in color and texture influence image processing tasks, such as segmentation and feature extraction. Hence, these techniques have not been adopted as analytical methods. It is majorly relevant to develop methods that can be applied for different types of samples, to reduce the effect of sample variation in image analysis, allowing for its adoption as a reference method.

In the current research, an image analyses approach is proposed for determination of cheese meltability for five different types of cheeses, with a wide range of technological attributes (color, texture, meltability). Image features were extracted from cheese samples and compared to the traditional method (Schreiber's test) in order to provide an objective and accurate method for determination of meltability. The objective was to propose a robust method, validated on five different types of cheese, that can later be expanded to other samples and thus become an analytical method of reference.

Material and Methods

Sample Preparation

Five types of cheese from seventeen different brands were acquired in the local market: Coalho ($n=3$), Meia Cura ($n=5$), Mozzarella ($n=3$), Prato ($n=3$) and Requeijão do Norte ($n=3$). Samples were kept in a freezer at $-18\text{ }^{\circ}\text{C}$ and transferred to a refrigerator at $5\text{ }^{\circ}\text{C}$ for 60 min before used for determination of meltability. Different cheese types were chosen to cover a wide range of meltability, among the most consumed in Brazil.

Mozzarella cheese is the most consumed cheese in Brazil and largely consumed worldwide, mainly due to its use as an ingredient in pizzas and frozen foods. Melting is the main expected characteristic of Mozzarella. Prato cheese is a cheese with similar characteristics to Gouda and Danbo cheeses, while Minas Meia Cura, Coalho, and Requeijão do Norte are handmade products with regional characteristics. Minas Meia Cura is a semi-cooked cheese, sold with at least 10 days of storage, mainly for direct consumption. Coalho, on the other hand, is a semi-hard and elastic cheese, originating in the northeastern region of Brazil, but with a considerable increase in industrially produced Coalho cheese, countrywide. Coalho is grilled for consumption; thus, meltability is an undesired characteristic in this type of cheese. Requeijão do Norte is a processed cheese that has a large free oil formation as a characteristic. Regarding the technical properties, Prato should present low meltability, while Minas Meia Cura and Requeijão do Norte should not melt when exposed to heating.

Determination of Meltability

Before analysis, samples were removed from the refrigerator and a probe was used to slice similar cylindrical subsamples from each cheese. Six subsamples with 36 mm of diameter and 5 mm of height were obtained from each cheese, totalizing 102 samples. These cylindrical samples were used for modified Schreiber's test, as they were placed in Petri dishes, and left in room temperature for 30 min. Subsequently, the Petri dishes containing the samples were placed inside an oven at $100\text{ }^{\circ}\text{C}$ for 7 min. Samples were removed from the oven and cooled at room temperature for 30 min. The diameter of the samples after melting was determined as the average of four different measurements across the sample, with 45° between each measurement, using a ruler. The increase in diameter was determined as follows:

$$M_d = \frac{(D_f - D_i)}{D_i} \times 100\% \tag{1}$$

where M_d (%) is the percentage of increase in diameter, and D_f and D_i are, respectively, the final and the initial diameters, in millimeters. This measurement was adopted as the reference value for meltability of the respective sample.

Computer Vision System and Image Acquisition

The computer vision system was previously described (Santos Pereira et al. 2018), and includes a CCD camera (Sony DSC-H55, Japan) installed inside a dark chamber, with an illumination unit comprised of two LED lamps (Natural daylight, 100 W). As standard setting conditions, the angle between the camera lens and the lighting source was approximately 45° , since diffuse reflection occurs at this angle (Valous et al. 2009). The surface used to acquire the images included two rulers positioned perpendicularly, used as a source of information for the calibration of colors in the image. Moreover, they allowed the conversion of the number of pixels to mm. The system was started 30 min before tests. Images were acquired using the following camera settings: manual exposure with shutter speed of 1/60 s (zoom and flash functions off) and ISO number of 200. One image was taken before and another after melting for each sample, and saved in TIFF format for further analyses.

Image Analyses

RGB images are comprised of three two-dimensional gray-scale matrices, where pixel values vary from 0 to 255, as 0 represents pure black, 255 represents pure white, and any intermediary values are shades of gray (Gunasekaran 1996). The three channels from an RGB image represent the intensity of different colors: R (red), G (green), and B (blue). In the proposed method, two approaches were compared, one using manual identification of the region of interest (ROI), and the other using an automatic method of ROI selection.

Manual ROI Selection

During image processing, images were segmented for isolation of the cheese sample from the background. First, it was performed a manual selection of the ROI as a reference to test the reliability of the automatic methods. In this case, ROI was manually selected from three different images: (1) the original RGB image; (2) the original RGB image was split into its 3 color channels, channels R and G were added and channel B was subtracted, and the resulting image used to select the ROI; (3) the original RGB image was converted to $L^*a^*b^*$ and the b^* channel was used to manually select the ROI. Then, the number of pixels from the selected ROI using these three methods was averaged. This procedure was used for images before and after melting, to provide an average result and reduce misinterpretation due to difficulties in

identifying the edges of the ROI. Similarly to the modified Schreiber method, the increase in area was determined as follows:

$$M_a = \frac{(A_f - A_i)}{A_i} \times 100\% \quad (2)$$

where M_a (%) is the percentage of area increase and A_f and A_i are, respectively, the final (after melting) and initial (before melting) areas, measured in number of pixels (Wang and Sun 2002a).

Automatic Selection of ROI

Four different approaches were investigated for automatic selection of ROI, for comparison purposes and process automation:

- (a) **R + G-B + Otsu + median filter:** Each RGB image (Fig. 1A) was initially separated in its 3 color channels. Since the color yellow is formed by high values in R and G channels and low values in B channel, and the white color is formed by high values in all channels, the channels R and G were summed up (Fig. 1B) to enhance contrast between the cheese sample (yellow) and the background (black). After, channel B was subtracted from this image to eliminate brightness from the oil in the cheese borders (Fig. 1C), which could affect the measurements. Otsu's method for thresholding was applied in the resulting image (Fig. 1D) and a median filter was used to eliminate background noise and to reduce the error from the bright spots in the borders

(Fig. 1E). After, other large objects remaining in the image, like the Petri dish, were removed (Fig. 1F).

- (b) **R + G-B + median filter + Otsu + median filter:** Similar to proposed approach (a); however, in this case, the median filter was also applied before the segmentation by Otsu's method.
- (c) **b* + Otsu + median filter:** Each RGB image (Fig. 2A) was initially converted in $L^*a^*b^*$ space color (Fig. 2B). After, Otsu's method for thresholding was applied on channel b^* (Fig. 2C) and a median filter was used to eliminate background noise and to reduce the error from the bright spots in the borders (Fig. 2D). Then, other large objects remaining in the image were removed (Fig. 2E).
- (d) **b* + median filter + Otsu + median filter:** Similar to proposed approach (c); however, in this case, the median filter was also applied before the segmentation by Otsu's method.

The remaining area of the cheese samples was measured, and similarly to the manual method, the increase in area was determined by Eq. 2. All results were reported as the average of three measurements. All image processing steps were performed in a script developed by the authors using ImageJ (open access) and Matlab R2015a software (Mathworks, Natick, MA, USA). Statistical analyses were carried out using Microsoft Excel (Windows Office, U.S.A.).

Fig. 1 Image processing steps: 1A—original RGB image; 1B—sum of channels R and G; 1C—subtraction of channel B; 1D—image after segmentation by Otsu's method; 1E—image filled for gaps; 1F—segmented cheese sample

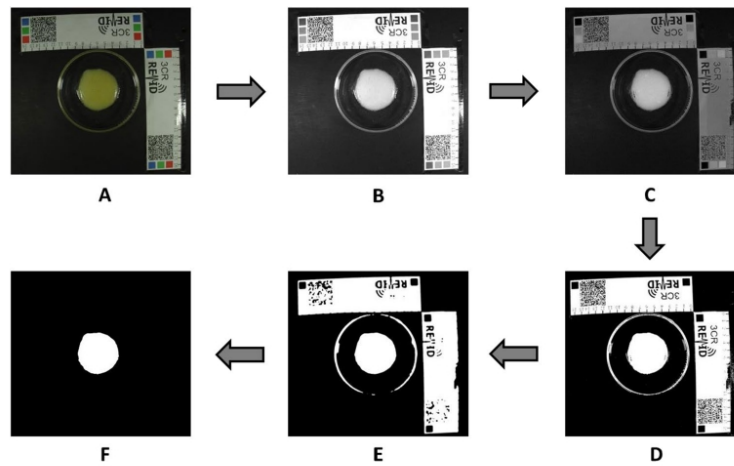
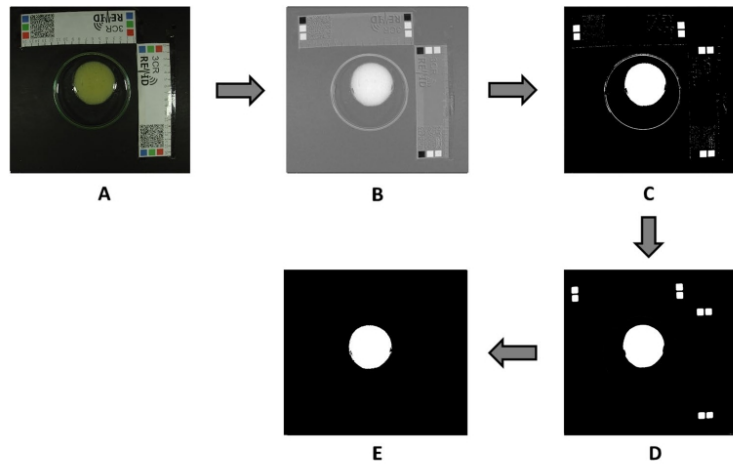


Fig. 2 Image processing steps: 2A—original RGB image; 2B—image in the b* channel; 2C—image after segmentation by Otsu’s method; 2D—image after median filter; 2E—segmented cheese sample



Results and Discussion

Sample Diameter

Table 1 shows the diameter measurements taken from the 102 cheese samples (17 samples × 6 subsamples) and the respective standard deviation (SD) of the four diameter measurements. In general, samples showed a small variation from the average diameter, what can be explained by the low SD. However, some samples showed a high SD, as Meia Cura D (subsamples 1 and 3), Mozzarella B (subsamples 4), Prato A (subsamples 1), and Requeijão B (subsamples 5). This high variation may be due to the fact that these samples spread irregularly after melting, so the positions where the diameters were measured were very different.

Chemical and textural differences occur among different types of cheese or even between the same types, when they are from different regions of the sample (cheese core or rind) from where the subsamples are taken, and this can influence melting. Thus, the four diameter measurements were averaged and the results are summarized in Table 1. The high SD values obtained by Coalho B and Meia Cura B suggest that there was a high difference between the subsamples of this type of cheese, what may be explained by the region of where samples were taken. Moreover, this table showed that even samples from the same type of cheese had significant difference between the diameters after melting, as Meia Cura A. This sample was not significantly different from Meia Cura C or E, but it was significantly different from Meia Cura B and D. Samples of Meia Cura C differ significantly from all the other samples of Meia Cura, except Meia Cura A. Requeijão samples presented similar results for melting, demonstrating that even different samples had

related properties. This difference in diameter after melting may be due to the chemical composition and differences in processing steps of each cheese, which might be different, even for samples of the same type.

Comparison Between the Schreiber Method and Computer Vision

The meltability results of 17 different cheeses determined by both the Schreiber and computer vision methods (manual selection of the ROI) are presented in Table 2. The meltability of cheeses determined by the Schreiber method varied from 0 to 45.37% among the different types and brands of cheese, while the results found by the computer vision method varied from 0.37 to 99.64%. As expected, Mozzarella samples presented higher values of meltability. Figure 3 shows the appearance of a Mozzarella sample before and after melting. On the other hand, although Coalho was expected to have very low meltability, this was not observed for sample Coalho C. Prato may present some degree of meltability, as observed in sample A, since its use in processed food has increased. The observed variation between samples of the same type of cheese may occur due to some lack of standardization in cheese production from different brands, which could lead to differences in composition that might affect cheese properties (Kapoor and Metzger 2008).

The result of meltability measured by the Schreiber method for Coalho A, Meia Cura A, and Requeijão A was zero, indicating that the increase was so subtle that the reference method was not capable to detect any difference between the samples before and after heating. However, it was possible to identify a slight increase in sample area, by computer vision.

Food Analytical Methods

Table 1 Sample diameter measurements determined by the modified Schreiber method

Cheese	Subsample	Diameter after melting (mm)				Average \pm SD (mm)
Coalho A	1	36	36	36	36	36.00 \pm 0.0
	2	36	36	36	36	36.00 \pm 0.0
	3	36	36	36	36	36.00 \pm 0.0
	4	36	36	36	36	36.00 \pm 0.0
	5	36	36	36	36	36.00 \pm 0.0
	6	36	36	36	36	36.00 \pm 0.0
	Average*					36.00 ^b \pm 0.00
Coalho B	1	44	43	45	44	44.00 \pm 0.82
	2	38	39	38	37	38.00 \pm 0.82
	3	45	45	44	43	44.25 \pm 0.96
	4	38	38	39	38	38.25 \pm 0.50
	5	41	42	42	41	41.50 \pm 0.58
	6	36	36	36	36	36.00 \pm 0.00
	Average*					40.33 ^{bc} \pm 3.43
Coalho C	1	48	51	49	49	49.25 \pm 1.26
	2	49	50	47	47	48.25 \pm 1.50
	3	45	45	45	46	45.25 \pm 0.50
	4	50	52	49	49	50.00 \pm 1.41
	5	49	51	50	51	50.25 \pm 0.96
	6	47	50	48	46	47.75 \pm 1.71
	Average*					48.46 ^{bc} \pm 1.85
Meia Cura A	1	36	36	36	36	36.00 \pm 0.00
	2	36	36	36	36	36.00 \pm 0.00
	3	36	36	36	36	36.00 \pm 0.00
	4	36	36	36	36	36.00 \pm 0.00
	5	36	36	36	36	36.00 \pm 0.00
	6	36	36	36	36	36.00 \pm 0.00
	Average*					36.00 ^b \pm 0.00
Meia Cura B	1	40	39	39	40	39.50 \pm 0.58
	2	51	53	51	52	51.75 \pm 0.96
	3	35	35	35	35	35.00 \pm 0.00
	4	50	46	46	47	47.25 \pm 1.89
	5	49	49	48	47	48.25 \pm 0.96
	6	48	48	48	48	48.00 \pm 0.00
	Average*					44.96 ^{cd} \pm 6.33
Meia Cura C	1	34	34	34	34	34.00 \pm 0.00
	2	34	34	34	34	34.00 \pm 0.00
	3	33	33	33	33	33.00 \pm 0.00
	4	35	35	35	35	35.00 \pm 0.00
	5	34	34	34	34	34.00 \pm 0.00
	6	33	34	33	34	33.50 \pm 0.58
	Average*					33.92 ^a \pm 0.66
Meia Cura D	1	41	43	40	38	40.50 \pm 2.08
	2	49	49	48	48	48.50 \pm 0.58
	3	43	52	52	43	47.50 \pm 5.20
	4	50	46	47	48	47.75 \pm 1.71
	5	48	45	45	47	46.25 \pm 1.50
	6	48	46	45	46	46.25 \pm 1.26
	Average*					46.13 ^d \pm 2.89

Food Analytical Methods

Table 1 (continued)

Cheese	Subsample	Diameter after melting (mm)				Average \pm SD (mm)
Meia Cura E	1	42	42	40	41	41.25 \pm 0.96
	2	38	38	38	38	38.00 \pm 0.00
	3	41	39	40	41	40.25 \pm 0.96
	4	37	37	37	37	37.00 \pm 0.00
	5	38	39	39	38	38.50 \pm 0.58
	6	40	40	40	40	40.00 \pm 0.00
	Average*					39.17 ^b \pm 1.59
Mozzarella A	1	43	43	43	42	42.75 \pm 0.50
	2	46	44	45	45	45.00 \pm 0.82
	3	48	50	51	48	49.25 \pm 1.50
	4	47	47	46	45	46.25 \pm 0.96
	5	45	44	46	46	45.25 \pm 0.96
	6	50	49	49	48	49.00 \pm 0.82
	Average*					46.25 ^d \pm 2.50
Mozzarella B	1	51	53	50	52	51.50 \pm 1.29
	2	51	50	49	50	50.00 \pm 0.82
	3	55	53	54	56	54.50 \pm 1.29
	4	56	53	54	49	53.00 \pm 2.94
	5	53	55	54	53	53.75 \pm 0.96
	6	52	50	50	53	51.25 \pm 1.50
	Average*					52.33 ^c \pm 1.70
Mozzarella C	1	38	38	38	38	38.00 \pm 0.00
	2	39	39	40	39	39.25 \pm 0.50
	3	38	38	38	38	38.00 \pm 0.00
	4	34	34	34	34	34.00 \pm 0.00
	5	41	42	41	40	41.00 \pm 0.82
	6	37	37	36	38	37.00 \pm 0.82
	Average*					37.88 ^{ab} \pm 2.34
Prato A	1	48	48	42	44	45.50 \pm 3.00
	2	47	46	48	48	47.25 \pm 0.96
	3	52	51	52	52	51.75 \pm 0.50
	4	48	50	51	48	49.25 \pm 1.50
	5	44	45	45	46	45.00 \pm 0.82
	6	48	47	48	47	47.50 \pm 0.58
	Average*					47.71 ^{de} \pm 2.50
Prato B	1	40	41	40	38	39.75 \pm 1.26
	2	43	41	39	39	40.50 \pm 1.91
	3	40	40	39	42	40.25 \pm 1.26
	4	40	41	42	40	40.75 \pm 0.96
	5	44	42	41	41	42.00 \pm 1.41
	6	40	40	40	40	40.00 \pm 0.00
	Average*					40.54 ^{bc} \pm 0.80
Prato C	1	40	41	41	41	40.75 \pm 0.50
	2	38	38	38	37	37.75 \pm 0.50
	3	42	42	41	41	41.50 \pm 0.58
	4	39	39	40	40	39.50 \pm 0.58
	5	40	40	40	40	40.00 \pm 0.00
	6	44	42	41	41	42.00 \pm 1.41
	Average*					40.25 ^{bc} \pm 1.53

Table 1 (continued)

Cheese	Subsample	Diameter after melting (mm)				Average \pm SD (mm)
Requeijão A	1	36	36	36	36	36.00 \pm 0.00
	2	36	36	36	36	36.00 \pm 0.00
	3	36	36	36	36	36.00 \pm 0.00
	4	36	36	36	36	36.00 \pm 0.00
	5	36	36	36	36	36.00 \pm 0.00
	6	36	36	36	36	36.00 \pm 0.00
	Average*					36.00 ^{ab} \pm 0.00
Requeijão B	1	36	37	36	36	36.25 \pm 0.50
	2	38	37	36	36	36.75 \pm 0.96
	3	36	39	37	35	36.75 \pm 1.71
	4	37	36	35	35	35.75 \pm 0.96
	5	40	37	35	36	37.00 \pm 2.16
	6	36	36	36	37	36.25 \pm 0.50
	Average*					36.46 ^{ab} \pm 0.46
Requeijão C	1	36	36	36	36	36.00 \pm 0.00
	2	36	36	36	36	36.00 \pm 0.00
	3	37	37	36	37	36.75 \pm 0.50
	4	36	36	36	36	36.00 \pm 0.00
	5	38	36	36	37	36.75 \pm 0.96
	6	36	37	38	38	37.25 \pm 0.96
	Average*					36.46 ^{ab} \pm 0.53

Results for each subsample are reported as averages of four measurements \pm standard deviation

*Results for each sample are reported as averages of six subsamples (measurements \pm standard deviation). Different letters in the same column are significantly different as determined by Tukey's test ($p < 0.05$)

SD standard deviation

The most important value to be observed is C2 (Fig. 4), the correlation between final diameter measured by the Schreiber method and final increase in area measured by computer vision. Due to the irregular variation of the sample after melting (Fig. 5), measuring the diameter sometimes is not accurate, as samples may not spread regularly, but in different directions. Although four diameter measurements were taken and averaged, this was not enough to detect small differences in the sample diameter after melting. However, these differences were identified by the computer vision method. Since the proposed method takes into account the whole area of the sample, it is feasible to claim that it provides more accurate results. The results also demonstrate that the increase in the diameter did not cause the same percentage increase in the area. Additionally, values for standard deviation obtained by computer vision were very high, indicating a wider range among samples that indicates the sensitiveness of the proposed approach, compared to the Schreiber method.

The correlations between the percentage of increase in diameter and area (C1) and between the final values of diameter and area (C2) are shown in Fig. 4. Overall, most of the cheese samples showed high correlation, with high values of

C1 and C2. However, some of them, such as Coalho C and Meia Cura C, showed a negative correlation (C1). This result may be explained by the fact that C1 is a correlation between the difference of final and initial diameters, and final and initial areas. Therefore, considering that the diameter is the value measured by the Schreiber method and the area by the computer vision method, it was observed that some of the samples that had the highest difference of diameter had the lowest difference of area, resulting in a negative correlation.

However, in the case of Meia Cura C samples, apart from the C1 value, cheese samples showed negative results for both diameter and area increases, indicating that samples shrunk after heat treatment. Moreover, samples of Requeijão B showed a negative result for area increase, suggesting that the computer vision method detected sample shrinkage rather than expansion. Shrinkage and reduction of area has been previously related to evaporation of moisture and oiling-off during cheese heating (Wang and Sun 2002b).

Furthermore, the low and the negative correlations can also be explained by the fact that the script was optimized for all samples, which varied between different shades of yellow. The negative correlations were observed for samples that presented very subtle variations in the diameter after

Food Analytical Methods

Table 2 Comparison of meltability results obtained by the modified Schreiber method and computer vision with manual segmentation, for different cheese types

Sample	Average increase in diameter \pm SD (%)*	Average increase in area \pm SD (%)*
Coalho A	0.00 \pm 0.00	0.41 \pm 0.82
Coalho B	12.04 \pm 8.69	26.08 \pm 20.63
Coalho C	34.61 \pm 4.68	70.53 \pm 7.94
Meia Cura A	0.00 \pm 0.00	0.37 \pm 0.42
Meia Cura B	24.88 \pm 16.05	54.48 \pm 40.40
Meia Cura C	-5.79 \pm 0.60	-3.15 \pm 0.42
Meia Cura D	28.13 \pm 2.45	51.54 \pm 10.91
Meia Cura E	8.80 \pm 4.04	10.38 \pm 6.50
Mozzarella A	28.47 \pm 6.35	56.52 \pm 16.42
Mozzarella B	45.37 \pm 4.31	99.64 \pm 14.42
Mozzarella C	5.21 \pm 2.22	18.95 \pm 4.18
Prato A	32.52 \pm 4.63	71.76 \pm 1.76
Prato B	12.62 \pm 3.89	18.22 \pm 8.75
Prato C	11.81 \pm 3.89	24.41 \pm 8.75
Requeijão A	0.00 \pm 0.00	0.43 \pm 0.36
Requeijão B	1.27 \pm 1.16	-0.31 \pm 1.06
Requeijão C	1.27 \pm 1.35	0.38 \pm 1.02

*Results for each sample are reported as averages of six replicates
 Increase in diameter: result obtained by the modified Schreiber method
 Increase in area: result obtained by computer vision
 SD standard deviation

melting. The subtle variation was also detected by computer vision, but the replicates were not necessarily similar for both methods, thus resulting in negative correlation.

Comparison Between Manual and Automatic Selection of ROI

Manual selection can be quite tedious and laborious depending on the number of images to be analyzed. Therefore, automatic techniques of segmentation were tested to achieve an accurate and faster method for determination of cheese meltability that could be used in laboratories and food processing lines. Figure 6 reports the results of the correlation between percentage of increase in diameter and percentage of increase in area (C1'), and between the final diameter and area increase (C2'), determined by the Schreiber method and image analysis using automatic segmentation techniques previously described in **Material and Methods**.

Again, it is important to analyze the values of C2', the correlation between the final diameter measured by the modified Schreiber method and the final area measured by the computer vision with automatic ROI segmentation. Overall, the results presented high values of correlation with the Schreiber method. Since this method was not able to identify any difference in sample diameter after heating, samples of Coalho A, Meia Cura A, and Requeijão do Norte A showed a coefficient of correlation of zero, similar to the correlation obtained using the manual ROI selection.

Additionally, the correlation between the manual and the automatic methods was calculated and the results showed a perfect correlation between all the computer vision methods (> 0.99), showing the feasibility of automation of image analysis for cheese meltability.

Compared to previous works (Wang and Sun 2002a, b, c), that used features such as diameter, perimeter, area, and circularity, our method presented a simpler and more straightforward approach, reducing processing time and providing highly accurate results.

It was observed that other factors may influence the results: some samples presented bubbles while melting making it difficult for identification of ROI; Requeijão do Norte

Fig. 3 Mozzarella sample before (A) and after melting (B) (for interpretation of the references to color in the figure, the reader is referred to the web version of this article)

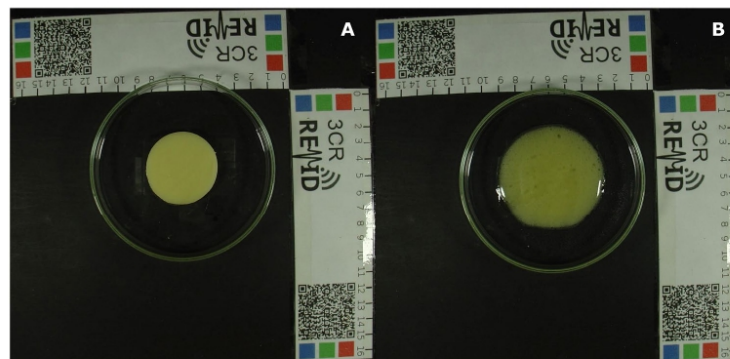


Fig. 4 Correlation between the Schreiber method and manual computer vision method (C1: correlation between percentage of increase in diameter and percentage of increase in area and C2: correlation between the final diameter and area increase using manual method of segmentation) (for interpretation of the references to color in the figure, the reader is referred to the web version of this article)

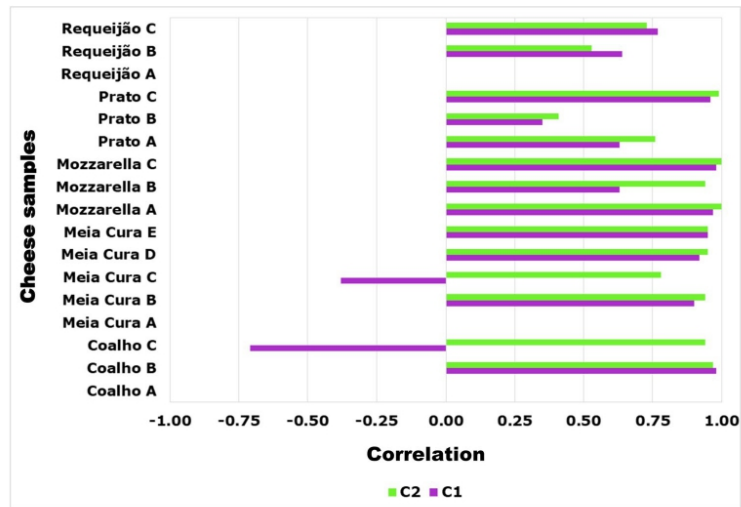
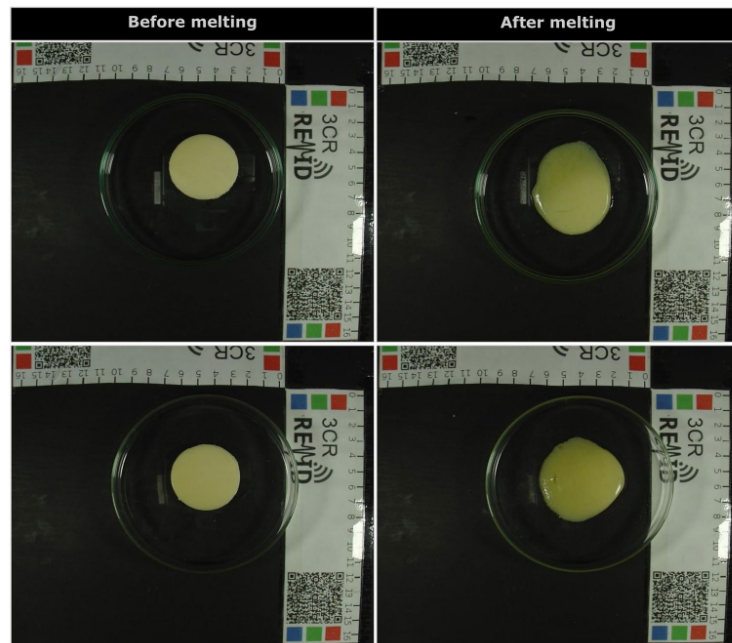


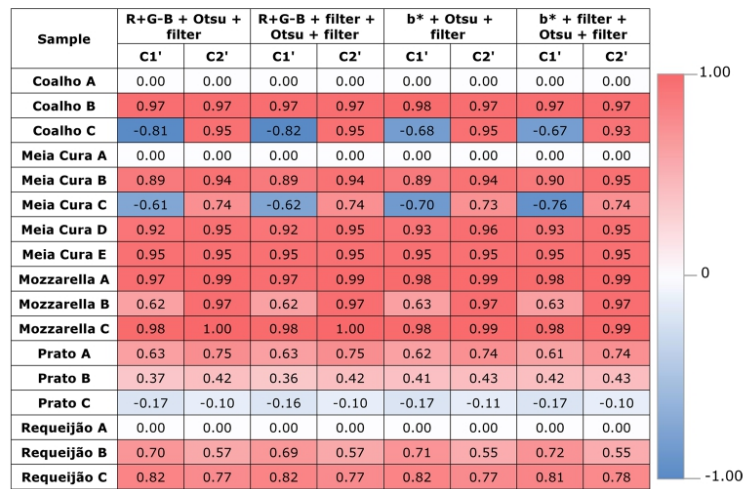
Fig. 5 Cheese samples showing irregular meltability after heating (for interpretation of the references to color in the figure, the reader is referred to the web version of this article)



released oil under heating; these phenomena could influence model precision. Nevertheless, the results obtained show a

high correlation between the traditional analytical method (Schreiber) and the automated method this article proposes.

Fig. 6 Correlation between manual and automatic computer vision methods (C1': correlation between percentage of increase in diameter and percentage increase of area and C2': correlation between the final diameter measured by the Schreiber method and area increase calculated by automatic method of segmentation) (for interpretation of the references to color in the figure, the reader is referred to the web version of this article)



Conclusions

Results for determination of cheese meltability demonstrated good correlation between the computer vision method proposed in this study and the traditional method (modified Schreiber method). The Schreiber method is susceptible to precision of human visual imprecision and errors caused by human fatigue, which could interfere in the results. These errors are prevented by using a fully automated computer vision method. Moreover, the proposed method showed its ability to measure the whole area of cheese samples rather than just the diameter, which is a great advantage for cheese samples with irregular meltability. Thus, the computer vision could be more accurate than the reference method adopted in this study. Additionally, the system could be applied for simultaneous imaging of several samples, and after image acquisition, the script can rapidly analyze the images and provide accurate results for cheese meltability. Considering all the advantages, the proposed automated method is a reliable, precise, and fast tool to determine cheese meltability, and it can be improved for further types of samples or adapted to other applications.

Author Contributions Amanda Teixeira Badaró: formal analysis, writing—original draft, writing—review and editing. Gustavo Vechin de Matos: methodology, software, validation, investigation, data curation, writing—original draft. Caroline Bilhar Karaziack: investigation, methodology. Walkiria Hanada Viotto: conceptualization, methodology, resources, writing—review and editing. Douglas Fernandes Barbin: conceptualization, resources, writing—original draft, writing—review and editing, supervision, project administration, funding acquisition.

Funding This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES) [Finance Code 001] and São Paulo Research Foundation (FAPESP) [grant numbers 2015/24351–2, 2017/17909–2, 2017/17628–3, 2019/06842–0].

Data Availability All data and other material will be made available upon request.

Declarations

Informed Consent Informed consent not applicable.

Conflict of Interest Amanda Teixeira Badaró declares that she has no conflict of interest. Gustavo Vechin de Matos declares that he has no conflict of interest. Caroline Bilhar Karaziack declares that she has no conflict of interest. Walkiria Hanada Viotto declares that she has no conflict of interest. Douglas Fernandes Barbin declares that he has no conflict of interest.

References

Altan A, Turhan M, Gunasekaran S (2005) Short communication: comparison of covered and uncovered Schreiber test for cheese meltability evaluation. *J Dairy Sci* 88(3):857–861. [https://doi.org/10.3168/jds.S0022-0302\(05\)72751-X](https://doi.org/10.3168/jds.S0022-0302(05)72751-X)

Amamcharla JK, Metzger LE (2015) Prediction of process cheese instrumental texture and melting characteristics using dielectric spectroscopy and chemometrics. *J Dairy Sci* 98(9):6004–6013. <https://doi.org/10.3168/jds.2015-9739>

Barbin DF, Mastelini SM, Barbon S, Campos GF, Barbon APA, Shimokomaki M (2016) Digital image analyses as an alternative tool for chicken quality assessment. *Biosyst Eng* 144:85–93. <https://doi.org/10.1016/j.biosystemseng.2016.01.015>

Calvini R, Michelini S, Pizzamiglio V, Foca G, Ulrici A (2020) Exploring the potential of NIR hyperspectral imaging for automated quantification of rind amount in grated Parmigiano Reggiano

- cheese. *Food Control* 112:107111. <https://doi.org/10.1016/j.foodcont.2020.107111>
- Fagan CC, Everard C, O'Donnell CP, Downey G, Sheehan EM, Delahunty CM, O'Callaghan DJ, Howard V (2007) Prediction of processed cheese instrumental texture and meltability by mid-infrared spectroscopy coupled with chemometric tools. *J Food Eng* 80(4):1068–1077. <https://doi.org/10.1016/j.jfoodeng.2006.04.068>
- Gallo M, Ferranti P (2016) The evolution of analytical chemistry methods in foodomics. *J Chromatogr A* 1428:3–15. <https://doi.org/10.1016/j.chroma.2015.09.007>
- Geronimo BC, Mastelini SM, Carvalho RH, Barbon Júnior S, Barbin DF, Shimokomaki M, Ida EI (2019) Computer vision system and near-infrared spectroscopy for identification and classification of chicken with wooden breast, and physicochemical and technological characterization. *Infrared Phys Technol* 96:303–310. <https://doi.org/10.1016/j.infrared.2018.11.036>
- Guinee TP, Kilcawley KN (2004) Cheese as an ingredient. In: Fox PF, McSweeney PLH, Cogan TM, Guinee TP (eds) *Cheese: chemistry, physics and microbiology*, 3rd edn. Academic Press, pp 395–428
- Gunasekaran S (1996) Computer vision technology for food quality assurance. *Trends Food Sci Technol* 7:245–256. [https://doi.org/10.1016/0924-2244\(96\)10028-5](https://doi.org/10.1016/0924-2244(96)10028-5)
- Kapoor R, Metzger LE (2008) Process cheese: scientific and technological aspects—a review. *Comp Reviews Food Sci Food Safety* 7(2):194–214. <https://doi.org/10.1111/j.1541-4337.2008.00040.x>
- Kosikowski F (1982) *Cheese and fermented milk foods*, 2nd edn. F.V. Kosikowski and Associates, New York
- Li H, Liang Y, Xu Q (2009) Support vector machines and its applications in chemistry. *Chemom Intell Lab Syst* 95:188–198. <https://doi.org/10.1016/j.chemolab.2008.10.007>
- Lopes JF, Ludwig L, Barbin DF, Grossmann MV, Barbon S (2019) Computer vision classification of barley flour based on spatial pyramid partition ensemble. *Sensors (switzerland)* 19(13):2953. <https://doi.org/10.3390/s19132953>
- Lucey JA (2008) Some perspectives on the use of cheese as a food ingredient. *Dairy Sci Technol* 88:573–594. <https://doi.org/10.1051/dst:2008010>
- Lucey JA, Johnson ME, Horne DS (2003) Invited review: perspectives on the basis of the rheology and texture properties of cheese. *J Dairy Sci* 86(9):2725–2743. [https://doi.org/10.3168/jds.S0022-0302\(03\)73869-7](https://doi.org/10.3168/jds.S0022-0302(03)73869-7)
- Mastelini SM, Sasso MGA, Campos GFC, Schmiele M, Clerici MTPS, Barbin DF, Barbon S (2018) Computer vision system for characterization of pasta (noodle) composition. *J Electron Imag* 27(5):053021. <https://doi.org/10.1117/1.JEI.27.5.053021>
- Minz PS, Saini CS (2019) Evaluation of RGB cube calibration framework and effect of calibration charts on color measurement of mozzarella cheese. *J Food Meas Charact* 13:1537–1546. <https://doi.org/10.1007/s11694-019-00069-9>
- Nolasco-Perez IM, Rocco LACM, Cruz-Tirado JP, Pollonio MAR, Barbon S, Barbon APAC, Barbin DF (2019) Comparison of rapid techniques for classification of ground meat. *Biosyst Eng* 183:151–159. <https://doi.org/10.1016/j.biosystemseng.2019.04.013>
- Oliveira MM, Cerqueira BV, Barbon S, Barbin DF (2021) Classification of fermented cocoa beans (cut test) using computer vision. *J Food Compos Anal* 97:103771. <https://doi.org/10.1016/j.jfca.2020.103771>
- Santos Pereira LF, Barbon S, Valous NA, Barbin DF (2018) Predicting the ripening of papaya fruit with digital imaging and random forests. *Comput Electron Agric* 145. <https://doi.org/10.1016/j.compag.2017.12.029>
- Shetty DK, Acharya, UD, Malarout N, Gopakumar R, Prajwal PJ (2019) A review of application of computer-vision for quality grading of food products. In 2019 Int. Conf. Autom. Comput. Technol. Manag. ICACTM 2019, pp 297–303. <https://doi.org/10.1109/ICACTM.2019.8776763>
- Tsakanikas P, Pavlidis D, Nychas GJ (2015) High throughput multi-spectral image processing with applications in food science. *PLoS One* 10(10):e0140122. <https://doi.org/10.1371/journal.pone.0140122>
- Udomkun P, Innawong B, Jeepetch K (2019) Computer vision system (CVS) for color and surface oil measurements of durian chips during post-frying. *J Food Meas Charact* 13(3):2075–2081. <https://doi.org/10.1007/s11694-019-00128-1>
- Valous NA, Mendoza F, Sun DW, Allen P (2009) Colour calibration of a laboratory computer vision system for quality evaluation of pre-sliced hams. *Meat Sci* 81(1):132–141. <https://doi.org/10.1016/j.meatsci.2008.07.009>
- Wang HH, Sun DW (2002a) Correlation between cheese meltability determined with a computer vision method and with Arnott and Schreiber tests. *J Food Sci* 67(2):745–749. <https://doi.org/10.1111/j.1365-2621.2002.tb10670.x>
- Wang HH, Sun DW (2002b) Melting characteristics of cheese: analysis of effects of cooking conditions using computer vision techniques. *J Food Eng* 51(4):305–310. [https://doi.org/10.1016/S0260-8774\(01\)00072-3](https://doi.org/10.1016/S0260-8774(01)00072-3)
- Wang HH, Sun DW (2002c) Melting characteristics of cheese: analysis of effect of cheese dimensions using computer vision techniques. *J Food Eng* 52(3):279–284. [https://doi.org/10.1016/S0260-8774\(01\)00116-9](https://doi.org/10.1016/S0260-8774(01)00116-9)
- Wang YC, Muthukumarappan K, Ak MM, Gunasekaran S (1998) A device for evaluating melt/flow characteristics of cheeses. *J Texture Stud* 29(1):43–55. <https://doi.org/10.1111/j.1745-4603.1998.tb00152.x>
- Yin J, Hameed S, Xie L, Ying Y (2021) Non-destructive detection of foreign contaminants in toast bread with near infrared spectroscopy and computer vision techniques. *J Food Meas Charact* 15:189–198. <https://doi.org/10.1007/s11694-020-00627-6>
- Zhang Y, Wang S, Ji G (2014) Phillips P (2014) Fruit classification using computer vision and feedforward neural network. *J Food Eng* 143:167–177. <https://doi.org/10.1016/j.jfoodeng.2014.07.001>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

27/09/2021 11:13

RightsLink Printable License

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Sep 27, 2021

This Agreement between Miss. Amanda Badaró ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number 5157100039305

License date Sep 27, 2021

Licensed Content Publisher Springer Nature

Licensed Content Publication Food Analytical Methods

Licensed Content Title Automated Method for Determination of Cheese Meltability
by Computer Vision

Licensed Content Author Amanda Teixeira Badaró et al

Licensed Content Date Aug 3, 2021

Type of Use Thesis/Dissertation

Requestor type academic/university or research institute

Format print and electronic

Portion full article/chapter

Will you be translating? no

Circulation/distribution 100 - 199

Author of this Springer Nature content yes

27/09/2021 11:13

RightsLink Printable License

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Sep 27, 2021

This Agreement between Miss. Amanda Badaró ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5157100039305
License date	Sep 27, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Food Analytical Methods
Licensed Content Title	Automated Method for Determination of Cheese Meltability by Computer Vision
Licensed Content Author	Amanda Teixeira Badaró et al
Licensed Content Date	Aug 3, 2021
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	100 - 199
Author of this Springer Nature content	yes

27/09/2021 11:13

RightsLink Printable License

Title	NIR spectral techniques and chemometrics applied to food processing
Institution name	University of Campinas and Universitat Politècnica de València
Expected presentation date	Oct 2021
Order reference number	2
Requestor Location	Miss. Amanda Badaró Rua Pedro Geraldo Costa, 291 Tambaú, Sao Paulo 13710-000 Brazil Attn: Miss. Amanda Badaró
Customer VAT ID	BR09939021623
Total	0.00 USD

Terms and Conditions

**Springer Nature Customer Service Centre GmbH
Terms and Conditions**

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

- 1. 1.** The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.
- 1. 2.** The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).
- 1. 3.** If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

27/09/2021 11:13

RightsLink Printable License

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the [STM Permissions Guidelines](#), as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

27/09/2021 11:13

RightsLink Printable License

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. **BOOKS ONLY:** Where 'reuse in a dissertation/thesis' has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the [STM Permissions Guidelines](#), which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

27/09/2021 11:13

RightsLink Printable License

For Journal Content:

Reprinted by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)]

For Advance Online Publication papers:

Reprinted by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)]

For Adaptations/Translations:

Adapted/Translated by permission from [the Licensor]: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)]

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [the Licensor]: on behalf of Cancer Research UK: : [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication)]

For Advance Online Publication papers:

Reprinted by permission from The [the Licensor]: on behalf of Cancer Research UK: [Journal Publisher (e.g. Nature/Springer/Palgrave)] [JOURNAL NAME] [REFERENCE CITATION (Article name, Author(s) Name), [COPYRIGHT] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])]

For Book content:

Reprinted/adapted by permission from [the Licensor]: [Book Publisher (e.g. Palgrave Macmillan, Springer etc)] [Book Title] by [Book author(s)] [COPYRIGHT] (year of publication)

Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

APPENDIX III

*Identification of turkey meat and processed products
using near infrared spectroscopy*



Contents lists available at ScienceDirect

Food Control

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

Identification of turkey meat and processed products using near infrared spectroscopy



Douglas Fernandes Barbin^{a,b,*,1}, Amanda Teixeira Badaro^{b,1}, Danielle C.B. Honorato^c, Elza Youko Ida^c, Massami Shimokomaki^{a,2}

^a Department of Food Science, Federal Technological University of Parana (UTFPR), Rodovia Celso Garcia Cid, PR 445 Km 380, Campus Universitário, Londrina, PR, 86055-900, Brazil

^b Department of Food Engineering, University of Campinas (UNICAMP), Rodovia Celso Garcia Cid, PR 445 Km 380, Campus Universitário, Londrina, PR, 86055-900, Brazil

^c Department of Food Science, State University of Londrina (UEL), Rodovia Celso Garcia Cid, PR 445 Km 380, Campus Universitário, Londrina, PR, 86055-900, Brazil

ARTICLE INFO

Keywords:

Authentication
Ham
Partial least squares
Discriminant analysis
Principal component analysis

ABSTRACT

Meat processors and consumers are greatly concerned about nutritional value, safety and quality of food products. Besides the reproducibility, which is a significant quality parameter of processed food product, the adulteration of meat products is a crucial concern for manufactures and consumers. Therefore, fast and objective techniques are demanded to ensure the quality of raw or processed meat. In the current study, near-infrared (NIR) spectroscopy was verified as a prospective technique to discriminate turkey cuts and processed turkey meat products. Spectral information in the wavelengths between 400 and 2500 nm of raw material and ready-to-eat turkey products were acquired and studied for their potential application for quality control and authentication. Principal component analysis (PCA) was explore the spectra information and samples were classified using linear discriminant analysis (LDA). PCA carried out on NIR dataset revealed the effect of chemical composition and quality features on the spectra. This investigation suggested that NIR spectroscopy is a convenient tool for quality evaluation of turkey meat.

1. Introduction

The designation 'ham' entails the thigh and rump of pork; however, the name can be used to processed poultry products if the meat have as a source the breast or leg of the animal, although this is usually qualified with the species of the animal as with 'turkey ham' (Feiner, 2006). The production and commercialization of a wide variety of processed meat products has increased as required by the industry and consumers to be used as catering, delicatessen, and ingredient (Iqbal, Valous, Mendoza, Sun, & Allen, 2010).

Consumers are greatly concerned about nutritional value, safety and quality in terms of health and sensory properties of food products. On the other hand, meat processors are concerned about products technological quality and safety (Zamora-Rojas, Garrido-Vado, Pedro-Sanz, Guerrero-Ginel, & Pérez-Marín, 2011). Low variability in product attributes is an essential feature of processed food quality. There are

currently few studies concerning reproducibility of consecutive batches in production lines. The assessment of the variability in product characteristics demands a considerable number of data, which often makes it difficult to analyse the results (Probola & Zander, 2007).

In addition, meat adulteration is a crucial concern for both the meat handlers and consumers. Fraudulent substitution in ground meat products with low-price meat or non-meat substances is a growing issue that involves not only economy, quality and safety, but also religion (Deniz et al., 2018). Ground meat production homogenizes the muscles characteristics and replacement, thus making adulterant identification an extremely difficult task (Kamruzzaman, Makino, Oshita, & Liu, 2015).

Therefore, it is necessary to rigorously control the procedures to ensure the quality of raw or processed meat (Prieto, Roehle, Lavín, Batten, & Andrés, 2009). The meat industry and consumers demand fast screening techniques in order to determine meat quality and verify

* Corresponding author. Department of Food Science, Federal Technological University of Parana (UTFPR), Rodovia Celso Garcia Cid, PR 445 Km 380, Campus Universitário, Londrina, PR, 86055-900, Brazil.

E-mail address: dobarbin@unicamp.br (D.F. Barbin).

¹ These authors contributed equally.

² In memoriam

<https://doi.org/10.1016/j.foodcont.2019.106816>

Received 17 May 2019; Received in revised form 8 August 2019; Accepted 9 August 2019

Available online 09 August 2019

0956-7135/ © 2019 Elsevier Ltd. All rights reserved.

product label (Zamora-Rojas et al., 2011). Consequently, the determination of quality parameters and inspection becomes continuously important for both manufacturers and consumers (Iqbal et al., 2010).

In recent years, many works have been reported investigating techniques able to identify meat adulteration and verify labelling statements. Methods based on ELISA, nucleic acid amplification and mass spectrometry have been used; however, they are time-consuming and demand technical knowledge (Masiri et al., 2017). Thus, an easy, fast and accurate method implemented for the regulation of food label and product quality control is a good alternative (Meza-Marquez, Gallardo-Velazquez, & Osorio-Revilla, 2010). Process analytical technology (PAT) is an effective tool to delineate, inspect and supervise manufacturing processes by measuring critical process parameters (CPP) and monitoring them, since this parameters can influence the quality attributes of food products.

Spectroscopic measurements span throughout the visible, near-, mid- and far-infrared region. The NIR region (750–2500 nm) involves the radiation associated to overtones and combinations of fundamental vibrational modes of carbon, nitrogen, oxygen and sulphur, each bonded to a hydrogen molecule. On the other hand, visible wavelength ranges are useful in muscle food applications due to pigments in the raw material (Kumar & Chandrakant Karne, 2017).

Spectral data comprise a large amount of co-linear and redundant information. Hence, multivariate statistical approach is essential to reduce the data dimension without missing functional information. Multivariate statistical analysis methods have been often used in such situations, allowing for variables reduction and clearer interpretation of results (Lohumi, Lee, Lee, & Cho, 2015). Principal component analysis (PCA) is usually the first move for exploring spectral data. This technique is able to linearly reduce the large amount of variables in the original data into principal components, which contain the majority of variability of source data (Kumar & Chandrakant Karne, 2017).

NIR spectroscopy has been used as in-line and on-line monitoring tool to reduce production-cycling time, avoid batches rejection and allow real time monitoring of parameters (Roggo et al., 2007). Spectroscopy techniques are fast, often non-destructive, and once the models are calibrated, they are easy to use, representing an interesting option for food inspection (Alamprese, Casale, Sinelli, Lanteri, & Casiraghi, 2013). Rady and Adedeji (2018) built classification and prediction models based on Vis-NIR data for detecting plant and animal-origin adulterants on minced beef and pork. Promising results were found by De Marchi et al. (2017) in the prediction of sodium content in commercial processed meat products using near infrared spectroscopy. However, this technique has not been used with the specific propose of identifying turkey cuts and turkey processed products.

Therefore, the aim of the current work was to explore the potential use of NIR spectroscopy as a rapid and chemical free technique to discriminate different turkey cuts and turkey meat products. Specific objectives were (1) to characterize the different products according to chemical composition during processing steps, (2) acquiring NIR spectral information from the samples and investigating the influence of samples chemical composition on spectra, (3) to apply chemometric methods to discriminate between different samples, (4) identifying most significant wavelengths to discriminate samples with reduced spectral information.

2. Materials and methods

2.1. Sample preparation

The experiment was carried out in two parts. In the first part, 100 fresh meat samples of turkey cuts (20 wings, 20 legs, 20 drumsticks, 20 breasts and 20 skin) were analysed. In the second part of the experiment, four different processed turkey products (blanquet, cooked ham, turkey breast and smoked breast) were examined (Table 1). Samples

were analysed in three processing stages: (1) sample mixture without addition of brine, (2) sample mixture after addition of brine, and (3) final products after processing.

Several different batches were acquired for each sample from meat processing plant in Parana, Brazil. Each batch, containing three units ($n_{total} = 158$ samples), was randomly taken and transported under refrigeration to the Laboratory of Food Science at the State University of Londrina, Londrina-PR, Brazil, for analysis. A multiprocessor was used for grinding and blending of the samples, and 50 g of each sample were used for analysis. The experiment was performed in duplicate.

2.2. Analytical measurements

Samples were thawed at 10 °C for 24 h. Colour was determined using a Minolta colorimeter (CR 400, Konica-Minolta Sensing Inc., Osaka, Japan) with a D65 illuminant and a 10° observer; 4 measurements of each sample were taken and averaged. The calibration was performed with a standard ceramic tile. Colour was expressed in terms of values for lightness (L^*), redness (a^*), yellowness (b^*), Chroma (C^*), and hue angle (h^*) according to the Commission Internationale de l'Eclairage (CIE) colour system (CIE, 1976; Feiner, 2006; Honikel, 1998).

2.3. Chemical analysis

Moisture, fat, ash and protein content were determined using the whole parts and ground samples and used as reference values for further NIR data. Samples were evaluated by traditional analytical methods immediately after NIRS analysis. The moisture content was determined according to ISO-1442 (1997) by oven drying the samples until constant weight at 100 °C. Fat content was measured following the Soxhlet procedure (ISO-1443, 1973), ash content was determined after incineration at 550 °C following ISO-936 (1998), total protein was determined by the Kjeldahl method (ISO-937, 1978) and carbohydrates were calculated by difference.

2.4. Near-infrared spectroscopy

The sample spectral information was acquired using an XDS Near-Infrared model XM 1100 series – Rapid Content Analyser (Foss NIRSystems, Denmark), which covers the wavelength range from 400 to 2498 at 2-nm intervals. Measurements were collected in reflectance mode and converted as absorbance ($\log 1/R$). After each measurement, the sample cell was washed with ethanol (70% v/v), rinsed with distilled water and dried using soft paper tissue. Turkey cut spectra were collected directly from the sample.

2.5. Statistical analysis

2.5.1. ANOVA

The results were assessed by analysis of variance (ANOVA) and the Tukey test with a significance level of 0.05 using Statistica software 7.0 (StatSoft Inc., USA).

2.5.2. PCA and classification models and selection of optimal wavelengths

Spectral data were, first, evaluated by a PCA in order to explore the variance among samples. The method was used in two separate data sets: Set 1 - turkey cuts; and Set 2 - processed samples (without and with addition of brine) and final products.

The near infrared spectrum comprises many overtones and combination bands. Therefore, designating some important wavelengths can reduce in a significant way the amount of data regarding the response of interest. The loadings obtained from PCA using full spectra can be useful for selecting the optimum wavelengths, yielding fairly similar results compared with model using all the wavelengths.

In order to classify turkey cuts and processed meat turkey samples

Table 1
Basic composition of processed turkey meat products.

Product	Ingredients
Blanquet	White breast meat, turkey fat, water, soy protein, malt dextrin, cassava starch, salt, natural spices, sugar, sodium polyphosphate (INS 452i), carrageenan (INS 407), sodium nitrite (INS 250) and sodium erythorbate (INS 316).
Cooked ham	Dark meat from whole drumstick, salt, malt dextrin, soy protein, natural spices, sugar, carmine natural colour (INS 120), sodium nitrite (INS 250), carrageenan (INS 407), sodium polyphosphate (INS 452i), monosodium glutamate (INS 621) and sodium erythorbate (INS 316).
Turkey breast	White meat from whole breast, sugar, salt, malt dextrin, soy protein, natural spices, carrageenan (INS 407), sodium nitrite (INS 250), sodium polyphosphate (INS 452i) and sodium erythorbate (INS 316).
Smoked breast	White breast meat, dark meat from drumstick, turkey fat, water, salt, sugar, soy protein, malt dextrin, cassava starch, natural spices, sodium polyphosphate (INS 452i), carrageenan (INS 407), sodium nitrite (INS 250) and sodium erythorbate (INS 316), colorant.

according to product characteristics, linear discriminant analysis (LDA) was performed (Massart et al., 1997). LDA is considered a probabilistic classification method, since it optimizes the separation among categories. NIR spectroscopy encompasses a large number of information from the samples. Since the amount of variables must be lower than samples, LDA was determined using few selected wavelengths. LDA was performed with a cross-validation (leave one out) and external validation, where samples were split in two groups: calibration (70% of samples) and validation (30% of samples). The models performance was tested with raw and pre-treated data, using Multiplicative Scattering Correction (MSC), Standard Normal Variate (SNV), 1st and 2nd Derivative pre-treatments. All multivariate analysis were performed with the Unscrambler software (version 9.7, CAMO, Trondheim, Norway).

3. Results and discussion

3.1. Chemical composition of processed Turkey products

Table 2 summarizes chemical composition of tested turkey cuts, while Table 3 presents the variations in colour (L^* , a^* , b^* , Chroma and Hue angle and chemical composition of tested turkey products. The quality characteristics results suggested that there is a wide variance range in these attributes. Breast and wing samples had the greatest protein content (24.80 and 23.98%, respectively), differentiating significantly from other samples. Skin had the smallest protein content (9.94%). Even though the moisture content were very similar between the turkey cuts, samples were significantly different between each other. The fat content was significantly higher in skin samples (15.28%), while leg and drumstick, and wing and breast samples had no significant difference between them. Wing and drumstick had no difference among ash content, while skin samples that had the smallest content, varied significantly from other samples.

Cooked ham samples presented the smallest L^* values and differentiated significantly from all the samples, both in samples without brine, with brine and final products. Since a^* represents the samples redness, this parameter can be affected by the myoglobin content, which is different in samples with only breast meat or only drumstick meat (Wideman, O'Bryan, & Crandall, 2016), showing significant difference between samples. Cooked ham is a product composed by dark meat from drumstick (Table 1) and showed the highest a^* values, regardless

of whether the samples were without brine, with brine or final products. Moisture is the largest component in the samples and, even though had not varied much between samples and processing stages (no brine, with brine and final products), the differences were significant. Blanquet and cooked ham had the highest moisture content in samples without brine, which differentiated significantly between the other samples. In samples with brine and final products, this parameter was higher in cooked ham, which was different from the other samples. Protein content had no significant difference in samples without brine, but after the stage with addition of brine, the protein content decreased and cooked ham differed from other samples. Turkey breast had the highest protein content in final products, differing from other samples. The carbohydrate content of turkey cuts and processed turkey products was calculated as a difference of 100 minus the sum of the other proximate components means (eq. (1)); hence, those who the other components exceed 100%, the carbohydrate content was not calculated.

$$c = 100 - (\bar{p} + \bar{m} + \bar{f} + \bar{a}) \quad (1)$$

Where, c is the carbohydrate content, \bar{p} is the mean protein content, \bar{m} is the mean moisture content, \bar{f} is the mean fat content and \bar{a} is the mean ash content.

3.2. Spectral characteristics of samples

The average NIR spectra acquired from each cut and product in the range of 400–2500 nm are shown in Fig. 1. The NIR range comprises spectral features regarding absorption bands of many chemical compounds. Overtones and combinations of fundamental vibrations of C–H, N–H, O–H and S–H functional groups are most prominent absorption bands in the NIR region (Kumar & Chandrakant Karne, 2017).

Local absorption peaks at 446, 560, 980, 1210, 1450, 1950 nm, and a slope starting at 2240 nm can be noticed in the turkey meat spectra. Even though the spectral profiles of the samples presented similar pattern, there is difference in absolute reflectance values. Leg and drumstick samples (Fig. 1a) had the highest absorbance values in the visible range. These cuts have high myoglobin content, which are responsible for the darker colour (Wideman et al., 2016). Therefore, cooked ham (Fig. 1b), which contains drumstick meat as main ingredient, showed the same behaviour, indicating its darker colour in comparison with the other products.

Table 2
Experimental results for chemical composition of the tested turkey cuts.

Turkey cuts	Protein (%)	Moisture (%)	Fat (%)	Ash (%)	Carbohydrate (%)
Wing	23.98 ^a ± 0.24	72.77 ^a ± 0.12	2.44 ^a ± 0.30	1.06 ^a ± 0.01	–
Leg	19.42 ^b ± 0.41	74.87 ^b ± 0.29	3.77 ^b ± 0.21	1.13 ^b ± 0.01	0.81
Drumstick	20.84 ^c ± 0.14	74.36 ^c ± 0.12	3.66 ^b ± 0.28	1.10 ^{ab} ± 0.02	0.04
Breast	24.80 ^d ± 0.11	73.23 ^d ± 0.21	1.53 ^a ± 0.42	1.13 ^b ± 0.02	–
Skin	9.94 ^d ± 0.63	74.40 ^{bc} ± 0.11	15.28 ^c ± 0.76	0.43 ^c ± 0.02	–

Results are reported as averages of three or more determinations ± standard deviation. Different letters in the same column are significantly different as determined by Tukey test ($p < 0.05$).

D.F. Barbin, et al.

Food Control 107 (2020) 106816

Table 3
Experimental results for physical attributes and chemical composition of the tested turkey samples.

Product/ Attribute	L*	a*	b*	C*	h*	Protein (%)	Moisture (%)	Fat (%)	Ash (%)	Carbohydrates (%)
NO BRINE										
Blanquet	63.4 ^a ± 1.8	8.8 ^{ac} ± 1.7	14.8 ^a ± 1.4	17.3 ^{ab} ± 1.8	0.54 ^a ± 0.08	21.7 ^a ± 0.9	74.3 ^a ± 0.5	2.20 ^a ± 0.55	1.08 ^a ± 0.06	0.72
Cooked ham	41.0 ^b ± 4.1	15.7 ^b ± 0.9	10.4 ^b ± 3.9	19.0 ^b ± 2.9	1.00 ^b ± 0.13	20.5 ^a ± 1.4	74.1 ^a ± 0.7	2.23 ^a ± 0.44	1.46 ^b ± 0.01	1.71
Turkey breast	58.6 ^a ± 3.6	6.7 ^a ± 1.2	8.6 ^b ± 1.6	11.0 ^c ± 1.6	0.66 ^{bc} ± 0.11	20.7 ^a ± 3.1	72.3 ^b ± 0.5	3.96 ^b ± 0.39	1.03 ^a ± 0.03	2.01
Smoked breast	62.4 ^a ± 3.8	10.9 ^c ± 1.6	11.4 ^b ± 0.9	15.8 ^a ± 1.3	0.76 ^c ± 0.08	20.2 ^a ± 0.6	70.8 ^b ± 0.9	2.51 ^a ± 0.16	1.29 ^c ± 0.01	5.20
WITH BRINE										
Blanquet	61.9 ^a ± 3.6	10.9 ^a ± 1.8	13.9 ^a ± 2.4	17.7 ^{ab} ± 3.0	0.66 ^a ± 0.04	17.8 ^a ± 0.3	70.9 ^a ± 0.9	4.44 ^a ± 0.34	3.30 ^a ± 0.2	3.56
Cooked ham	48.8 ^b ± 5.3	14.6 ^b ± 2.9	10.7 ^{ab} ± 2.8	18.3 ^a ± 3.0	0.93 ^b ± 0.17	14.6 ^b ± 0.3	75.2 ^b ± 0.3	3.36 ^{ab} ± 0.50	3.91 ^b ± 0.01	2.93
Turkey breast	64.6 ^a ± 2.6	6.1 ^c ± 1.4	10.0 ^b ± 1.1	11.7 ^c ± 1.4	0.54 ^{bc} ± 0.09	16.5 ^a ± 0.3	77.2 ^c ± 0.3	2.21 ^c ± 0.23	3.4 ^a ± 0.3	0.69
Smoked breast	63.7 ^a ± 4.2	6.6 ^c ± 1.9	12.5 ^{ab} ± 2.0	14.2 ^{bc} ± 2.5	0.48 ^a ± 0.08	17.7 ^a ± 1.2	71.6 ^a ± 0.7	2.43 ^{bc} ± 0.26	2.13 ^c ± 0.09	6.14
FINAL PRODUCTS										
Blanquet	70.9 ^a ± 2.0	8.5 ^a ± 1.3	8.9 ^a ± 0.9	12.4 ^a ± 0.6	0.76 ^{ab} ± 0.11	18.2 ^{ab} ± 0.7	70.1 ^a ± 0.2	4.03 ^a ± 0.48	4.71 ^a ± 0.02	2.96
Cooked ham	57.8 ^b ± 2.6	15.6 ^b ± 1.5	8.6 ^a ± 0.9	17.8 ^b ± 1.6	1.06 ^c ± 0.04	17.1 ^a ± 0.3	74.7 ^b ± 0.8	4.71 ^a ± 0.16	3.32 ^b ± 0.02	0.17
Turkey breast	69.2 ^a ± 1.7	8.6 ^a ± 0.8	8.2 ^a ± 0.4	11.9 ^a ± 0.5	0.80 ^a ± 0.06	22.51 ^c ± 0.08	72.5 ^c ± 0.1	4.29 ^a ± 0.77	2.9 ^c ± 0.2	-
Smoked breast	70.5 ^a ± 1.4	7.0 ^a ± 0.4	8.4 ^a ± 0.5	10.9 ^a ± 0.6	0.69 ^b ± 0.01	19.0 ^b ± 0.9	73.0 ^c ± 0.1	5.00 ^a ± 0.32	3.02 ^{bc} ± 0.01	-

Results are reported as averages of three or more determinations ± standard deviation. Different letters in the same column of each section (no brine, with brine and final products) are significantly different as determined by Tukey test (p < 0.05).

A peak at 1730 nm and 2310 nm can be observed in the spectra of skin samples. These regions correspond to stretching first overtone of CH₂ associated to fat and fatty acids and CH combinations related with fat, respectively (Cozzolino & Murray, 2004).

The most intensive absorption peaks identified in the NIR range at 980 nm, 1450 nm and 1950 nm are due to O-H stretching third, second and first overtones, respectively, mainly related to water (Osborn,

Fearn, & Hindle, 1993). After addition of brine (Fig. 1c), cooked ham and breast samples have shown the highest absorption in this range, while smoked breast had lowest values, corroborating the experimental results obtained for moisture content. Absorption bands around 1210 nm are related to stretching and deformation of C-H bonds, while the region around 2240 nm is influenced by N-H stretching and deformation associated to amino acids (Osborn et al., 1993).

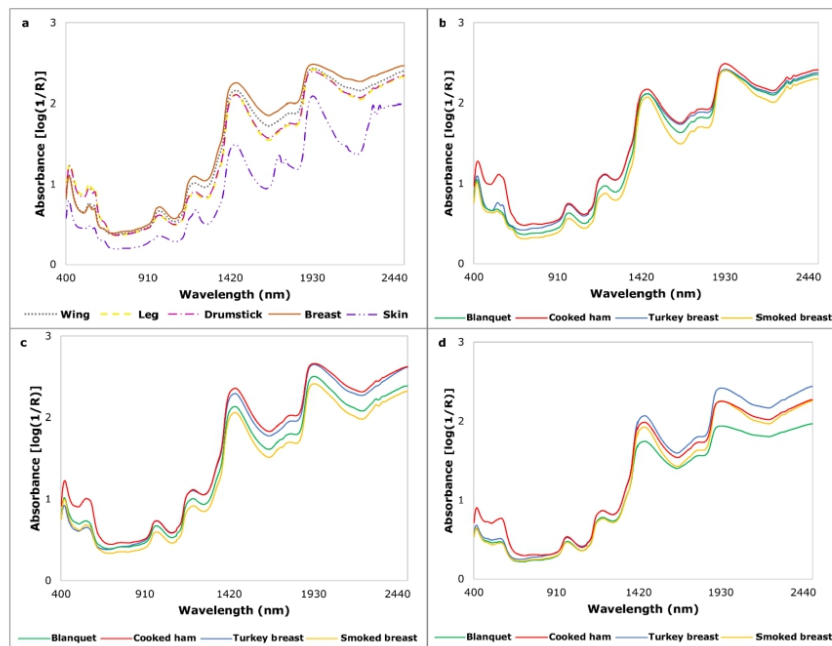


Fig. 1. Spectral information for samples a) turkey cuts, b) raw material without brine, c) raw material with brine, d) final products.

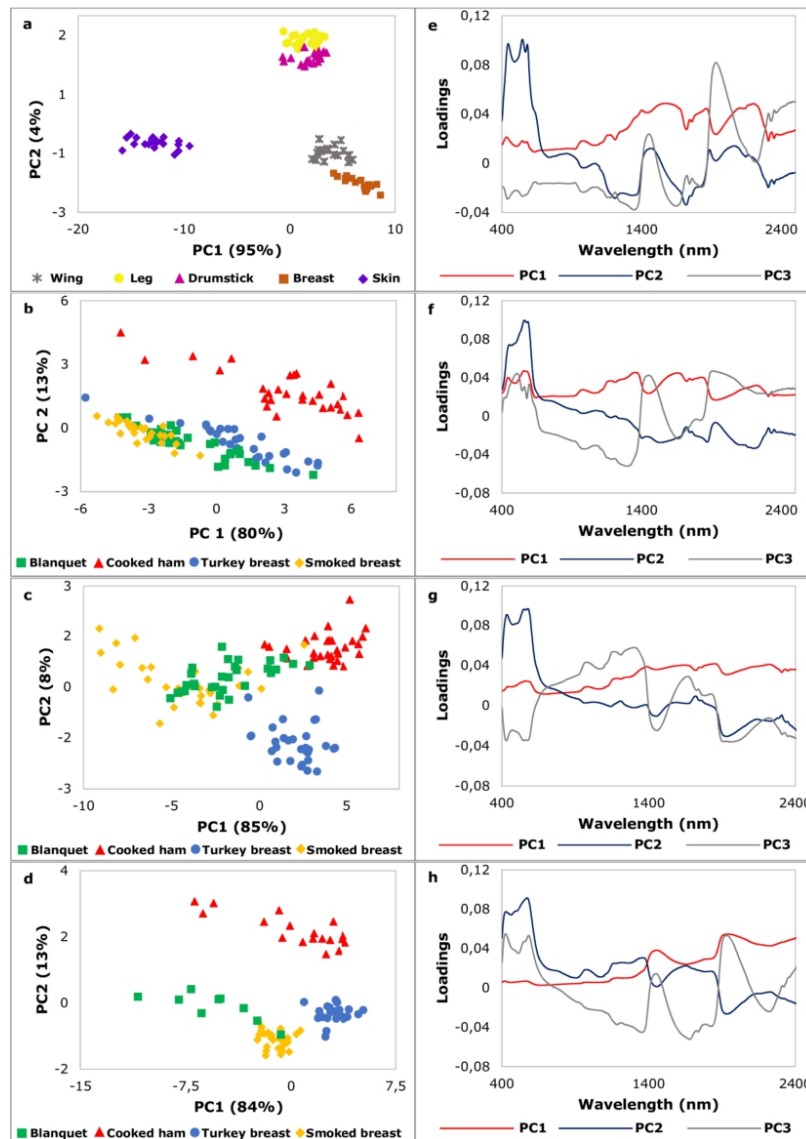


Fig. 2. Score plot of the first two principal components and loadings plot (PC1, PC2 and PC3) for spectral data of turkey samples: a) turkey cuts, b) raw samples without brine, c) raw samples with brine, d) final products.

3.2.1. Spectral multivariate analyses

Qualitative discrimination among samples was evaluated by a PCA, which was performed with a cross-validation (leave one out) on the spectral data. The original spectral data was rebuilt to a reduced the amount of variables, so that, samples were grouped according to their PCA scores instead of reflectance intensities. For all groups of samples, the first two principal components explained over 90% of the total variance among meat samples. The PCA is normally interpreted

visualizing the PC scores. Scores of the first two principal components of Set 1 and 2 were plotted and are presented in Fig. 2.

Fig. 2a represents the score plots of turkey cuts, with the first two PCs explaining 99% of the total variance. Five different clusters can be clearly observed, where skin samples are in the negative region of PC1 while the other samples are in the positive region. Still, the two groups composed by leg and drumstick samples are very distant from wing and breast samples and can be observed in the upper region of PC2. These

D.F. Barbin, et al.

Food Control 107 (2020) 106816

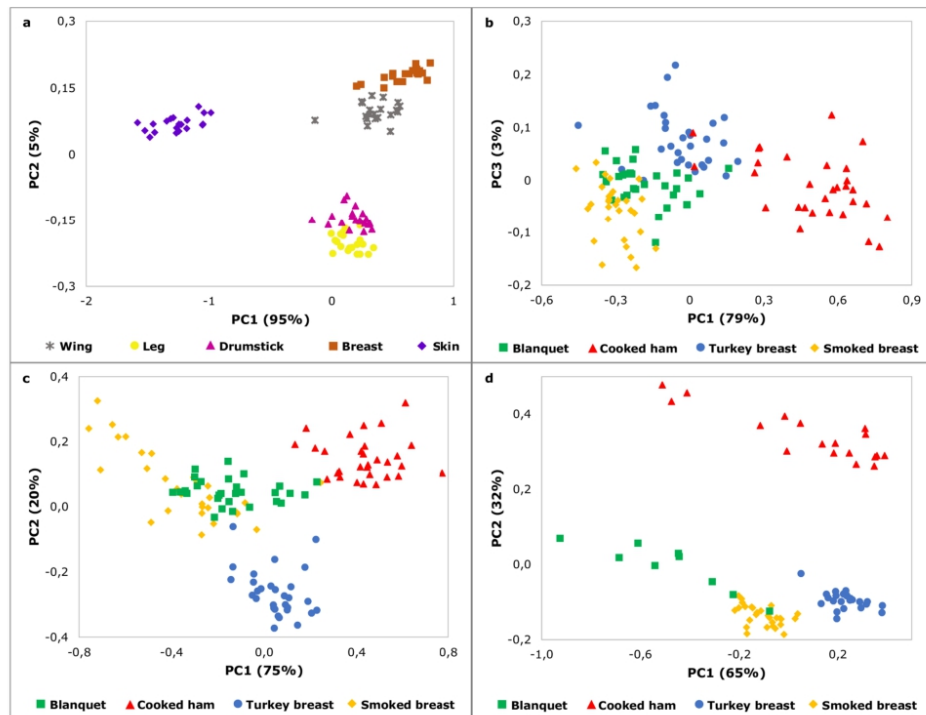


Fig. 3. Score plot of the first three principal components for spectral data of turkey products samples: a) turkey cuts, b) without brine, c) with brine, d) final products.

principal components have shown to be in accordance to the fat content of turkey cuts.

Regarding the processed products, the first two principal components shown in Fig. 2b suggested that the samples without brine addition were spectrally different, being able to discriminate them into separate classes. Samples were grouped into four distinct groups, indicating that the sample characteristic in the same cluster are very similar and those in different clusters are distinct. The cluster representing cooked ham is isolated in the positive region of PC2, meanwhile other samples are very close in the negative region of PC2. The blanquet is observed in the region between breast and smoked breast clusters.

Fig. 2c represents the score plot after addition of brine and shows breast samples in the negative region of PC2, while the other clusters are in the positive region of PC2. Final products (Fig. 2d) had very different profiles and can be clearly distinguished in the score plot of PC1 vs PC2.

These principal components seem to be related mainly to the colour characteristics values (L^* and a^*) of the samples without addition of brine and final products, while moisture content is mostly influencing the spectral information after addition of brine. Thus, samples showing similar characteristics regarding spectral information tend to be close predicted in the principal component space.

3.2.2. Identification of optimal wavelengths

Models based on a few number of wavelengths, which have most of the information in the data set, may be equal or better than models based on full wavelengths (Wold, Jakobsen, & Krane, 1996). Selecting a few wavelengths can be sufficient to perform most of the classification,

being suitable for online application as an alternative to remove redundant information, improve data processing and posterior analysis (Ariana, Lu, & Guyer, 2006; Nakariyakul & Casasent, 2008; Vila et al., 2005).

In this study, optimal wavelengths were identified and selected using the PCA loadings (Fig. 2e, f, g and h). The PCA loadings can be interpreted as the regression coefficients of each wavelength at each principal component, indicating which wavelengths the most influence the discrimination (Lawrence, Windham, Park, Smith, & Poole, 2004). The loadings of PC1, PC2 and PC3 were chosen for wavelength selection since they explained most of the variance in the data. The peaks and valleys at these loadings were selected, which corresponded to some of the main wavelengths observed in the spectra (446, 516, 560, 980, 1380, 1450, 1950 and 2220 nm), to test the ability of the selected wavelengths in discriminate the samples.

A new PCA was built with only the optimum wavelengths and Fig. 3 exhibits for the first three PCs the score plots of turkey cuts and processed products. After variable selection, the clusters were further separated from each other, even though they presented the same pattern. In Fig. 3a, PC1 and PC2 explained 100% of the difference among turkey cut samples, according to fat content determined by chemical analysis. After variable selection, skin samples remained in the negative side of PC1 and the other samples are in the positive side; however, leg and drumstick samples grouped together in the lower part of PC2, meanwhile wing and breast samples were in the upper part. Nevertheless, there are a clear separation among the clusters.

In Fig. 3b, PC1 and PC3 explained the difference among samples without addition of brine related to colour parameters, with cooked ham in the positive region of PC1. The cluster represented by breast

Table 4
LDA parameters for classification of turkey cuts using selected wavelengths.

TURKEY CUTS	PRE-TREATMENT	CALIBRATION		VALIDATION	
		Sensitivity	Selectivity	Sensitivity	Selectivity
WING	None	1.00	1.00	0.86	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	1.00	0.86	1.00
	2nd Derivative	1.00	1.00	1.00	1.00
LEG	None	0.93	1.00	0.80	0.80
	MSC	0.93	1.00	0.80	1.00
	SNV	1.00	1.00	0.80	1.00
	1st Derivative	1.00	0.94	0.60	1.00
	2nd Derivative	1.00	0.88	0.80	0.80
DRUMSTICK	None	1.00	0.98	0.83	0.83
	MSC	1.00	0.93	1.00	0.86
	SNV	1.00	1.00	1.00	0.86
	1st Derivative	0.92	1.00	1.00	0.75
	2nd Derivative	0.85	1.00	0.83	0.83
BREAST	None	1.00	1.00	1.00	0.83
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	1.00	1.00	0.83
	2nd Derivative	1.00	1.00	1.00	1.00
SKIN	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	1.00	1.00	1.00
	2nd Derivative	1.00	1.00	1.00	1.00

after addition of brine (Fig. 3c) is in the negative region of PC2, with other samples scattered in the positive part of PC2, corroborating with moisture content values. Different clusters can be clearly noticed in the recalculated PCA of final products (Fig. 3d), where smoked ham samples are isolated in the positive part of PC2, while the remaining samples are in the negative part.

In this way, as the optimum wavelengths were able to discriminate the samples, data processing time can be significantly reduced by implementing a powerful PAT in the meat processing industry using only these selected wavelengths. However, these plots only refer to qualitative variations of samples, without mentioning their quantitative attributes.

3.2.3. Classification of samples using selected wavelengths

The selected data were also used to classify turkey meat by linear discriminant analysis (LDA). This technique provides automatic object classification, by improving the variance ratio between classes and reducing the variance ratio within classes, and, consequently, finding an optimum limit among classes (Mignani, Giaccheri, Cimato, Attilio, & Smith, 2005). The selected wavelengths were used instead of the full spectral data to perform the LDA.

Table 4 shows the classification parameters of turkey cuts using selected wavelengths. The models based on reduced spectra showed high values of sensitivity and selectivity (over 75%), indicating robustness of the discrimination models. Raw data were able to correctly classify samples with over 80% certainty. Data pre-treated with MSC and SNV also showed good results classifying the turkey cuts samples. After 1st and 2nd derivative pre-treatments, the data did not show any improvement in the model performance. Table 5 shows the classification parameters of processed turkey products using the selected wavelengths. Overall, raw data were able to correctly classify turkey products (sensitivity and selectivity over 71%); however, the MSC and SNV pre-treatments clearly improved the models performance, with only few misclassifications. The high classification sensitivity and selectivity

values point out that the optimum wavelengths had the potential to discriminate the samples. The results suggest that NIR spectroscopy was able to distinguish these samples without any physicochemical information integrated with spectral data, representing a potential PAT for food processing industry.

4. Conclusion

The traditional techniques used for determination of turkey cuts and processed meat products have the disadvantage of being time-consuming, arduous and demanding technical knowledge, thus not suitable for evaluation of many samples in the production line. It would be a great advantage to determine at the same time distinct attributes of a batch, by using only a single device that is non-destructive.

NIR spectroscopy was explored as a non-destructive technique for classification of several turkey cuts and processed turkey meat products during different stages along the processing line. A few established wavelengths indicated presence of water and other chemical components in the samples. Identifying and selecting the most important wavelengths suggest the use of low-price multispectral NIR equipment for the desired application.

Experimental results indicate that NIR spectroscopy can be an asset over traditional method used for processed turkey cuts and processed meat products. The results highlighted the ability of visible and near infrared spectral information to identify samples without any background of physicochemical analysis. The technique has potential to be applied in the poultry industry as a key component in manufacturing, offering several advantages, which includes elimination of subjective analysis and the construction real time data of meat products for documentation, traceability, and labelling.

Declaration of interest

The authors declare no competing financial interest.

Table 5
LDA parameters for classification of processed turkey products using selected wavelengths.

TURKEY CUTS	PRE-TREATMENT	CALIBRATION		VALIDATION	
		SENSITIVITY	SELECTIVITY	SENSITIVITY	SELECTIVITY
NO BRINE					
BLANQUET	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	0.96	1.00	1.00
	2nd Derivative	0.96	1.00	0.80	1.00
COOKED HAM	None	0.88	0.88	0.71	0.83
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	0.84	1.00	0.82
	2nd Derivative	1.00	0.94	0.93	0.81
TURKEY BREAST	None	0.88	0.88	0.86	0.75
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	0.81	1.00	0.79	1.00
	2nd Derivative	0.94	0.94	0.79	0.92
SMOKED BREAST	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	0.96	1.00	1.00	1.00
	2nd Derivative	0.96	0.96	1.00	0.75
WITH BRINE					
BLANQUET	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	0.95	1.00	0.78	1.00
	2nd Derivative	0.90	0.86	0.67	0.60
COOKED HAM	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	1.00	1.00	1.00
	2nd Derivative	0.92	0.92	0.80	0.57
TURKEY BREAST	None	1.00	0.95	1.00	0.73
	MSC	1.00	0.95	1.00	0.73
	SNV	1.00	0.95	1.00	0.73
	1st Derivative	0.95	0.95	1.00	0.80
	2nd Derivative	1.00	1.00	1.00	0.80
SMOKED BREAST	None	0.94	1.00	0.77	1.00
	MSC	0.94	1.00	0.77	1.00
	SNV	0.94	1.00	0.77	1.00
	1st Derivative	0.94	0.89	1.00	1.00
	2nd Derivative	0.82	0.88	0.54	0.88
FINAL PRODUCT					
BLANQUET	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	1.00	1.00	1.00
	2nd Derivative	1.00	1.00	1.00	1.00
COOKED HAM	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	1.00	1.00	1.00
	2nd Derivative	1.00	1.00	1.00	1.00
TURKEY BREAST	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	1.00	1.00	1.00
	2nd Derivative	1.00	1.00	1.00	1.00
SMOKED BREAST	None	1.00	1.00	1.00	1.00
	MSC	1.00	1.00	1.00	1.00
	SNV	1.00	1.00	1.00	1.00
	1st Derivative	1.00	1.00	1.00	1.00
	2nd Derivative	1.00	1.00	1.00	1.00

Acknowledgements

The authors gratefully acknowledge the financial support from the National Council for Scientific and Technological Development (CNPq) strategic research initiative under the Brazilian Ministry of Science and Technology, project number 404852/2016-5, and São Paulo Research

Foundation (FAPESP) project 2015/24351-2 and 2017/17628-3\.. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

D.F. Barbin, et al.

Food Control 107 (2020) 106816

References

Alamprese, C., Casale, M., Sinelli, N., Lanteri, S., & Casiraghi, E. (2013). Detection of minced beef adulteration with Turkey meat by UV-vis, NIR and MIR spectroscopy. *Lebensmittel-Wissenschaft und -Technologie- Food Science and Technology*, 53, 225–232. <https://doi.org/10.1016/j.lwt.2013.01.027>.

Ariana, D. P., Lu, R., & Guyer, D. E. (2006). Near-infrared hyperspectral reflectance imaging for detection of bruises on pickling cucumbers. *Computers and Electronics in Agriculture*, 53, 60–70. <https://doi.org/10.1016/j.compag.2006.04.001>.

CIE (1976). *International commission on illumination, colorimetry: Official recommendations of the international commission on illumination*. Paris, France: Bureau Central de la CIE.

Cozzolino, D., & Murray, I. (2004). Identification of animal meat muscles by visible and near infrared reflectance spectroscopy. *Lebensmittel-Wissenschaft und -Technologie- Food Science and Technology*, 37(4), 447–452. <https://doi.org/10.1016/j.lwt.2003.10.013>.

De Marchi, M., Manuelian, C. L., Ton, S., Manfrin, D., Meneghesso, M., Cassandro, M., et al. (2017). Prediction of sodium content in commercial processed meat products using near infrared spectroscopy. *Meat Science*, 125, 61–65. <https://doi.org/10.1016/j.meatsci.2016.11.014>.

Deniz, E., Günes Altuntaş, E., Ayhan, B., İğci, N., Özel Demiralp, D., & Candoğan, K. (2018). Differentiation of beef mixtures adulterated with chicken or Turkey meat using FTIR spectroscopy. *Journal of Food Processing and Preservation*, 42(10), 1–12. <https://doi.org/10.1111/jfpp.13767>.

Feiner, G. (2006). Colour in fresh meat and in cured meat products. *Meat products handbook: Practical science and technology* (pp. 142–156). Cambridge, England: Woodhead Publishing Ltd.

Honikel, K. O. (1998). Reference methods for the assessment of physical characteristics of meat. *Meat Science*, 49, 447–457. [https://doi.org/10.1016/s0309-1740\(98\)00034-5](https://doi.org/10.1016/s0309-1740(98)00034-5).

Iqbal, A., Valous, N. A., Mendoza, F., Sun, D.-W., & Allen, P. (2010). Classification of pre-sliced pork and Turkey ham qualities based on image colour and textural features and their relationships with consumer responses. *Meat Science*, 84, 455–465. <https://doi.org/10.1016/j.meatsci.2009.09.016>.

ISO 1442, I. O. for S (1997). *Determination of moisture content - international standards meat and meat products*. Geneva: International Organization for Standardization 1973.

ISO 936, international O. for S (1998). *Determination of total ash - international standards meat and meat products*. Geneva, Switzerland: International Organization for Standardization.

ISO-1443, I. O. for S (1973). *Determination of total fat content - international standards meat and meat products*. Genève, Switzerland: International Organization for Standardization.

ISO-937, I. O. for S (1978). *Determination of nitrogen content - international standards meat and meat products*. Genève, Switzerland: International Organization for Standardization.

Kamruzzaman, M., Makino, Y., Oshita, S., & Liu, S. (2015). Assessment of visible near-infrared hyperspectral imaging as a tool for detection of horsemeat adulteration in minced beef. *Food and Bioprocess Technology*, 8(5), 1054–1062. <https://doi.org/10.1007/s11947-015-1470-7>.

Kumar, Y., & Chandrakant Karne, S. (2017). Spectral analysis: A rapid tool for species detection in meat products. *Trends in Food Science & Technology*, 62, 59–67. <https://doi.org/10.1016/j.tifs.2017.02.008>.

Lawrence, K. C., Windham, W. R., Park, B., Smith, D. P., & Poole, G. H. (2004). Comparison between visible/NIR spectroscopy and hyperspectral imaging for detecting surface contaminants on poultry carcasses. *Proceedings of SPIE*, 5271, 35–42. <https://doi.org/10.1117/12.516153>.

Lohumi, S., Lee, S., Lee, H., & Cho, B.-K. (2015). A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends in Food Science & Technology Journal*, 46 <https://doi.org/10.1016/j.tifs.2015.08.003>.

Masiri, J., Benoit, L., Thienes, C., Kainrath, C., Barrios-Lopez, B., Agapov, A., et al. (2017). A rapid, semi-quantitative test for detection of raw and cooked horse meat residues. *Food Control*, 76, 102–107. <https://doi.org/10.1016/j.foodcont.2017.01.015>.

Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., De Jong, S., Lewi, P. L., & Smeyers-Verbeke, J. (1997). Supervised pattern recognition. In B. G. M. Vandeginste, & S. C. Rutan (Vol. Eds.), *Handbook of chemometrics and qualimetrics: Part A*: 20A, (pp. 280–282). Amsterdam: Elsevier.

Meza-Marquez, O. G., Gallardo-Velazquez, T., & Osorio-Revilla, G. (2010). Application of mid-infrared spectroscopy with multivariate analysis and soft independent modeling of class analogies (SIMCA) for the detection of adulterants in minced beef. *Meat Science*, 86, 511–519. <https://doi.org/10.1016/j.meatsci.2010.05.044>.

Mignani, A. G., Giaccheri, L., Cimato, A., Attilio, C., & Smith, P. R. (2005). Spectral nephelometry for the geographic classification of Italian extra virgin olive oils. *Sensors and Actuators B*, 111–112, 363–369. <https://doi.org/10.1016/j.snb.2005.03.023>.

Nakariyakul, S., & Casasent, D. (2008). Hyperspectral waveband selection for contaminant detection on poultry carcasses. *Optical Engineering*, 47(8), 087202–087209 <https://doi.org/10.1117/1.2968693>.

Osborn, B. G., Fearn, T., & Hindle, P. H. (1993). Theory of near infrared spectroscopy. *Practical NIR spectroscopy with applications in food and beverage analysis* (pp. 13–35). (2nd ed.). London: Longman Singapore Publishers (Pte) Ltd.

Prieto, N., Roche, R., Lavín, P., Batten, G., & Andrés, S. (2009). Application of near infrared reflectance spectroscopy to predict meat and meat products quality: A review. *Meat Science*, 83, 175–186. <https://doi.org/10.1016/j.meatsci.2009.04.016>.

Probola, G., & Zander, L. (2007). Application of PCA method for characterisation of textural properties of selected ready-to-eat meat products. *Journal of Food Engineering*, 8, 93–98. <https://doi.org/10.1016/j.jfoodeng.2007.02.052>.

Rady, A., & Adedeji, A. (2018). Assessing different processed meats for adulterants using visible-near-infrared spectroscopy. *Meat Science*, 136, 59–67. <https://doi.org/10.1016/j.meatsci.2017.10.014>.

Roggo, Y., Chalou, P., Maurer, L., Lemamartinez, C., Edmond, A., & Jent, N. (2007). A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, 44(3), 683–700. <https://doi.org/10.1016/j.jpba.2007.03.023>.

Vila, J., Calpe, J., F. Pla, L., Gomez, J., Connell, J., et al. (2005). SmartSpectra: Applying multispectral imaging to industrial environments. *Real-Time Imaging*, 85–98 <https://doi.org/10.1016/j.rti.2005.04.007>.

Wideman, N., O'Bryan, C. A., & Crandall, P. G. (2016). Factors affecting poultry meat colour and consumer preferences - a review. *World's Poultry Science Journal*, 72(02), 353–366. <https://doi.org/10.1017/s0043933916000015>.

Wold, J. P., Jakobsen, T., & Krane, L. (1996). Atlantic salmon average fat content estimated by near-infrared transmittance spectroscopy. *Journal of Food Science*, 61(1), 74–77. <https://doi.org/10.1111/j.1365-2621.1996.tb14728.x>.

Zamora-Rojas, E., Garrido-Vado, A., Pedro-Sanz, E., Guerrero-Ginel, J. E., & Pérez-Marín, D. (2011). Monitoring NIRS calibrations for use in routine meat analysis as part of Iberian pig-breeding programs. *Food Chemistry*. 1889–1897 <https://doi.org/10.1016/j.foodchem.2011.05.139>.



RightsLink®



Identification of turkey meat and processed products using near infrared spectroscopy

Author: Douglas Fernandes Barbin, Amanda Teixeira Badaró, Danielle C.B. Honorato, Elza Youko Ida, Massami Shimokomaki

Publication: Food Control

Publisher: Elsevier

Date: January 2020

© 2019 Elsevier Ltd. All rights reserved.



Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

APPENDIX IV***Supplementary material of Chapter 2***



RightsLink®



Home



Help



Email Support



Sign in



Create Account

Identification of fiber added to semolina by near infrared (NIR) spectral techniques



Author: Amanda Teixeira Badaró, Fernanda Lie Morimitsu, Amanda Rios Ferreira, Maria Teresa Pedrosa Silva Clerici, Douglas Fernandes Barbin

Publication: Food Chemistry

Publisher: Elsevier

Date: 15 August 2019

© 2019 Elsevier Ltd. All rights reserved.

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

APPENDIX V***Supplementary material of Chapter 3***



RightstLink®



Home



Help



Email Support



Sign in



Create Account

Near infrared hyperspectral imaging and spectral unmixing methods for evaluation of fiber distribution in enriched pasta

Author:

Amanda Teixeira Badaró, José Manuel Amigo, José Blasco, Nuria Aleixos, Amanda Rios Ferreira, Maria Teresa Pedrosa Silva Clerici, Douglas Fernandes Barbin

Publication: Food Chemistry

Publisher: Elsevier

Date: Available online 31 October 2020

© 2020 Elsevier Ltd. All rights reserved.



Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW

APPENDIX VI***Supplementary material of Chapter 5***



RightsLink®



Home



Help



Email Support



Sign in



Create Account

Determination of pectin content in orange peels by near infrared hyperspectral imaging



Author:
Amanda Teixeira Badaró, Juan Francisco García-Martin, María del Carmen López-Barrera, Douglas Fernandes Barbin, Paloma Alvarez-Mateos

Publication: Food Chemistry

Publisher: Elsevier

Date: Sep 1, 2020

Copyright © 2020, Elsevier

Please note that, as the author of this Elsevier article, you retain the right to include it in a thesis or dissertation, provided it is not published commercially. Permission is not required, but please ensure that you reference the journal as the original source. For more information on this and on your other retained rights, please visit: <https://www.elsevier.com/about/our-business/policies/copyright#Author-rights>

BACK

CLOSE WINDOW