Detection of minor compounds in food powder using near infrared hyperspectral imaging

Antoine Laborde

université PARIS-SACLAY

Detection of minor compounds in food powder using near infrared hyperspectral imaging

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 581 : agriculture, alimentation, biologie, environnement et santé (ABIES)

Spécialité de doctorat: Chimie Analytique
Unité de recherche : Université Paris-Saclay, AgroParisTech, INRAE, UMR PNCA, 75005, Paris, France.
Référent : AgroParisTech

Thèse présentée et soutenue à Paris-Saclay, le 03/12/2020, par

Antoine LABORDE

Composition du Jury

Evelyne VIGNEAU Professeure, ONIRIS Nantes	Présidente
Nathalie DUPUY Professeure, Université Aix Marseille	Rapporteur et Examinatrice
Cyril RUCKEBUSCH Professeur, Université Lille 1	Rapporteur et Examinateur
Sílvia MAS GARCÍA Chargée de recherche, INRAE – Centre Occitanie-Montpellier	Examinatrice
Dalila AZZOUT-MARNICHE Professeure, AgroParisTech	Examinatrice
Christophe CORDELLA Ingénieur de recherche (HDR), INRAE – Centre IdF-Versailles-Grignon	Directeur de thèse
Benoît JAILLAIS Ingénieur de recherche, INRAE – Centre Pays de la Loire	Co-Encadrant et Examinateur
Anna DE JUAN Associate Professor, University of Barcelona	Invitée
Jean-Michel ROGER Directeur de Recherche, INRAE – Centre Occitanie-Montpellier	Invité
Douglas N. RUTLEDGE Professeur émérite, AgroParisTech	Invité

Thèse de doctorat

NNT : 2020UPASB017

Remerciements

Je tiens à remercier l'ensemble de l'équipe m'ayant encadré pendant ces trois années de thèse aussi bien pour leur confiance accordée, leur soutien et leurs encouragements. En particulier mon directeur de thèse Christophe Cordella, mon encadrant Benoît Jaillais qui m'a accueilli à Nantes pour des séjours enrichissants et agréables, ainsi que mes co-encadrants Delphine Jouan-Rimbaud Bouveresse et Luc Eveleigh qui ont su être disponibles lorsque j'en avais besoin. Je remercie aussi particulièrement Dominique Bertrand donc j'ai beaucoup apprécié l'aide et les conseils.

Je souhaite aussi grandement remercier Jean-Michel Roger pour son accueil et ses précieux conseils dans cette aventure. Je suis aussi très reconnaissant de l'aide et de l'accueil de l'équipe de Montpellier : Ryad Bendoula, Maxime Metz et Sílvia Mas García.

Je n'oublie pas ma visite très agréable au CRA-W de Gembloux où j'ai été très bien accueilli par Juan Antonio Fernández Pierna et Vincent Baeten.

Je remercie également Francesc Puíg-Castellví pour le travail réalisé ensemble, les discussions et l'aide apportée ainsi qu'Olivier Chapleur et Laetitia Cardona pour leurs conseils.

Je souhaite aussi remercier les personnes et entités ayant rendu ce projet possible malgré le déroulement de certains évènements. Je souhaite remercier les collègues qui m'ont soutenu : Katia et Elina ; Douglas Rutledge avec qui cette thèse a été initiée ; et Alexandre Péry ainsi que l'école doctorale qui m'ont soutenu dans la gestion de ce projet.

Enfin je voudrais remercier les membres du jury qui ont accepté d'évaluer mon travail. Je remercie mes deux rapporteurs Nathalie Dupuy et Cyril Ruckebusch, mes examinateurs Evelyne Vigneau, Sílvia Mas García, Dalila Azzout-Marniche ainsi que mes invités Douglas Rutledge et Anna De Juan.

Acronyms

AMSD	adaptive matched subspace detector
CSEL	concentration selectivity
EM	expectation maximization
GLR	generalized likelihood ratio
GMM	Gaussian mixture model
HSI	hyperspectral imaging
ICA	independent component analysis
LMM	linear mixing model
MCR-ALS	multivariate curve resolution alternating least-squares
MIR	mid-infrared
MLE	maximum likelihood estimator
MLR	multivariate linear regression
MSD	matched subspace detector
MVSA	minimum volume simplex analysis
NIR	near-infrared
NIRS	near-infrared spectroscopy
NIR HSI	near-infrared hyperspectral imaging
NMF	non negative matrix factorization
OLS	ordinary least-squares
PC	principal component
PCA	principal component analysis
PLA	polylactic acid
PLS	partial least-squares
PLSDA	partial least-squares discriminant analysis
RTE	radiative transfer equation
RMSE	root mean square error
RMSECV	root mean square error of cross-validation
SNR	signal to noise ratio
SNV	standard normal variate
SRS	spatially resolved spectroscopy
SVD	singular value decomposition
VNIR	visible near-infrared

Notations

Physics

I	intensity of an electromagnetic wave
λ	wavelength of an electromagnetic wave
R	reflectance
K	absorbance coefficient for the Kubelka-Munk theory
S	scattering coefficient for the Kubelka-Munk theory

Chemometrics

General notations

n	number of spectral measurements
m	number of wavelengths in spectral data
k	number of spectral constituents
$\boldsymbol{x}$	a spectrum vector
$\mathbf{X}$	a matrix containing one spectrum for each row
$\mathbf{P}$	loading matrix of Principal Component Analysis
$\mathbf{T}$	score matrix of Principal Component Analysis
$\boldsymbol{w}$	model residuals
$\mathbf{I}$	identity matrix
$\mathcal{N}$	Normal distribution
P	probability
$\boldsymbol{\mu}$	mean vector
$\boldsymbol{\Sigma}$	covariance matrix
$\boldsymbol{\omega}$	Gaussian mixture weights
M	number of Gaussians in the Gaussian Mixture Model

Matched Subspace Detector

$\boldsymbol{s}$	spectral component
$\boldsymbol{a}$	spectral contribution
L	number of components in the model of the regular sample
J	number of components in the model of the adulterant sample
$\mathbf{M}$	matrix containing the spectral components
c	simulated concentration of standard sample
$\tilde{\boldsymbol{x}}$	simulated spectrum
$\tilde{\mathbf{T}}$	simulated score matrix

Multivariate Curve Resolution Alternating Least-Squares

$\mathbf{C}$	concentration profile matrix, vector
$\mathbf{S}$	spectral profile matrix, vector
$\mathbf{E}$	model error matrix

Table of contents

Introduction

1. Context and objectives

This thesis was initiated by a collaboration between AgroParisTech and ONIRIS to develop monitoring solutions in the food industry. This thesis focuses on the case of food powder like flours using near-infrared (NIR) hyperspectral imaging (HIS) solutions. This technology that enables to characterize and localize food or chemical compounds in a sample can be used to detect adulteration. Detection of foreign components is a major application for enhancing food safety in the industry. As the chemical interactions between nutrients during food processing may affect their spectral signature, the detection should be done as soon as possible in the process. It means hyperspectral imaging should be applied on raw material such as powders which are intensively used in the food industry. Ce projet de thèse est issu d'une collaboration entre l'AgroParisTech et l'ONIRIS dans le but de développer des solutions de contrôle de procédés de fabrication dans l'industrie agroalimentaire.

However, HSI pixels may be larger than a food particle. In that case, one pixel measures the signal of several particles with different spectral signatures which results in a mixed signal. In addition, food samples are composed of the main nutrients that have similar spectral signatures in the NIR region. Therefore, the detection of foreign particles requires the use of specific chemometrics algorithms. One objective of this work is to propose performant algorithms for the signal analysis of mixed pixels.

NIR radiations can penetrate solid samples at a given distance before being absorbed. Consequently, the NIR spectral signatures of foreign particles can only be detected on a finite depth of raw material. The second objective of this work is to propose a method to assess the detection depth of a measurement system in a food powder.

2. Structure of the manuscript

The first chapter studies a new method to determine the detection depth of a material in a food powder. This method relies on the reflectance profile analysis derived from the Kubelka-Munk theory. An original sample holder is designed and produced using polylactic acid (PLA) which has a typical spectral signature. The design of the sample holder makes the thickness of powder material vary and is measurable using a NIR HSI system. A multivariate method based on the PLS regression is developed to determine the detection depth of the PLA in the powder material. The results are compared with the reflectance profile analysis. This chapter has the same structure as the published article. An additional part is provided to discuss the concept of detection depth.

The second chapter of this manuscript deals with the detection of peanut flour in wheat flour using a NIR HSI system and the MSD. Mixed samples with various

concentration of peanut flours are prepared (from 10 % to 0.02 %) and measured by HSI. The MSD is designed using the measurements of the pure samples. A spectral simulation method is proposed to provide a validation dataset for three MSD designs. The detectors are then applied to the real mixed samples to detect subpixel peanut adulteration in pixels. This chapter has the same structure as the published article.

The third chapter of this manuscript is dedicated to the detection of peanut flour in chocolate powder using the MCR-ALS. The chocolate powder is an industrial mix of cocoa and sugar which is adulterated with peanut flour in various concentration (from 10 % to 0.02 %). The samples are measured using a NIR HSI system. The spectral data of mixed samples are unmixed using MCR-ALS with an augmented matrix strategy and a selectivity constraint. The obtained concentration profiles are then processed by an outlier detection algorithm for detecting pixels adulterated with peanuts. This chapter has the same structure as the published article. An additional part provides a discussion of the unmixing and detection strategies.

A. Main contributions

First chapter

- The design of an original sample holder for the detection depth assessment of powder products.
- The development of a multivariate method for measuring the detection depth on the sample holder.

Second chapter

- The tuning of a MSD algorithm for the detection of peanut in wheat flour.
- A method for spectral data simulation used for the MSD design validation.

Third chapter

- The tuning of MCR-ALS with a selectivity constraint for detecting peanut in chocolate powder.
- The combination of an outlier detection algorithm with MCR-ALS.

B. List of the published works

- Laborde, B. Jaillais, R. Bendoula, J. Roger, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, D. Bertrand, A. Boulanger, C.B.Y. Cordella, A partial least squares-based approach to assess the light penetration depth in wheat flour by near infrared

hyperspectral imaging, J. Near Infrared Spectrosc. (2019). https://doi.org/10.1177/0967033519891594.

- A. Laborde, B. Jaillais, J.M. Roger, M. Metz, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, C.B.Y. Cordella, Subpixel detection of peanut in wheat flour using a matched subspace detector algorithm and near-infrared hyperspectral imaging, Talanta. 216 (2020). https://doi.org/10.1016/j.talanta.2020.120993.

- A. Laborde, F. Puig-Castellví, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, C.B.Y. Cordella, B. Jaillais, Detection of chocolate powder adulteration with peanut using near-infrared hyperspectral imaging and Multivariate Curve Resolution, Food Control. 119 (2021). https://doi.org/10.1016/j.foodcont.2020.107454.

C. List of communications

- A.Laborde, B. Jaillais, A. Boulanger, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, C.B.Y. Cordella, 7th June 2018, Detection of adulteration in powders in agro-industry, Conference on Hyperspectral Imaging in Industry (Graz, Austria).

- A. Laborde, R. Bendoula, D. Héran, A. Boulanger, J.M. Roger, B. Jaillais, C.B.Y. Cordella, 26th September 2018, Improving the scan depth of near-infrared hyperspectral imaging using spatially resolved spectroscopy, 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (Amsterdam, Netherlands).

- A.Laborde, B. Jaillais, D. Jouan-Rimbaud Bouveresse, A. Boulanger, C.B.Y. Cordella, 30th January 2019, Subpixel detection of peanut in wheat flour using near infrared hyperspectral imaging, Chimiométrie 2019 (Montpellier, France).

- A. Laborde, A. Boulanger, E. Siurdyban, A. Biloé, B. Jaillais, D. Jouan-Rimbaud Bouveresse, C.B.Y. Cordella, 14th March 2019 , Subpixel detection of peanut in wheat flour using near infrared hyperspectral imaging, 4th Conference on Optical Characterization of Materials (Karlsruhe, Germany). Best paper award.

3. The interest of powder in the food industry

Product stability is a major issue in food industry. Low water content products are usually stable, but for convenience, are often supplied as powder [4]. This product form induces multiple challenges for the food industry like maintaining the functionality of ingredients or preventing segregation of food ingredient mixes and particle stickiness [4]. Since food powders are finally consumed by humans, there is a high importance for preventing contamination with undesirable bio-life forms and chemical components. Contamination is difficult to prevent because of the dust formation which leads to particle settling and sticking onto equipment. The contamination may also happen in different stages of the product making and/or between several recipes, and could become a sanitary problem when considering food allergens.

The food allergy is defined as "an adverse health effect arising from a specific immune response that occurs reproducibly on exposure to a given food. [5]". It is a worrisome public health problem as it is responsible for 200 deaths per year in the United States [6]. Adult food allergy represents a population of 3.7 % in United States and 3.2 % in France [7]. In addition, the prevalence may be on the rise worldwide [8]. Some common foods with an allergen power are frequently used in the industry like wheat, egg, peanut, or milk and may have hazardous effects on the consumer [6]. For these reasons, the contamination of products by food allergens is highly probable and the industry tries to avoid it.

Pulverulent agri-food products are subject to numerous analyses, for example to control mixing and homogeneity [4]. In such cases, the presence of minor compounds often poses a problem. Firstly because they can be difficult to detect. This is the case of food contamination or food ingredients consciously added in low quantity in the mixture. Secondly, because the repartition of such ingredients has to be homogeneous. For this purpose, precise information about the chemical nature of particles and their position in the sample must be known to quantify the homogeneity. During the last decades, the near-infrared spectroscopy (NIRS) has been widely used to assess these features. Near infrared hyperspectral imaging (NIR HSI) has been recently applied and considered as a promising tool to analyze food samples. This technology enables to measure the spectrum of tiny localized areas on the sample and characterize it with more precision.

4. Near-infrared spectroscopy

The NIR spectroscopy (NIRS) studies the interaction of light radiations with the matter in the range between 780 nm and 2 500 nm [9]. When absorbed, these radiations make the molecules vibrate and produce a NIR spectrum. The type of vibration determines the wavelength at which a spectrum absorbs energy. The amplitude of absorption is described by the Beer Lambert's law and depends on the absorptivity of the molecule, its concentration in the sample and the radiation's pathlength. Figure 1 represents the

spectrum (A) as a consequence of the periodic stretching (C) of the methyl group. An incident light beam of intensity I_0 is directed towards two samples of varying concentration in methyl group (B). The intensity I_2 that goes out from the most concentrated sample is weaker than the other I_1 because of the molecule absorption.

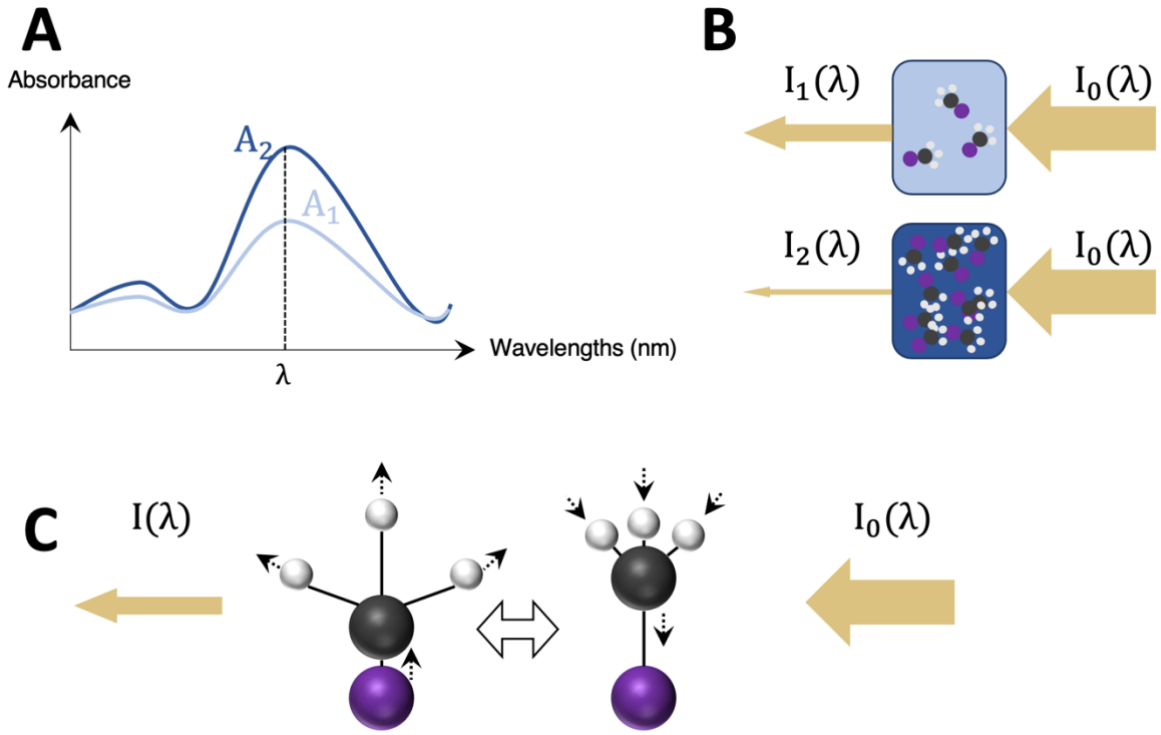

Figure 1: (A) Two absorbance spectra with an two absorption levels; (B) the sample concentration affecting the absorbance spectrum (C) and the methyl symmetrical stretching of ν_s CH_3.

Chemical bonds such as C = O, C – H, and N – H are the most active from the infrared point of view. They are present in organic compounds and, consequently, in food. Many advantages explain why this technology has been widely used in several fields like pharmaceutical field, petrochemistry, medicine, earth observation or environment. In particular, the use of NIR spectroscopy requires less sample preparation than other spectroscopic techniques.

5. Near-infrared hyperspectral imaging

NIR HSI consists of combining NIR spectroscopy and conventional imaging. This technology can be seen as an extension of RGB imaging by providing hundreds of spectral channels [1]. For each measured pixel, a complete NIR spectrum is acquired.

Especially, NIR HSI provides a spatialized and resolved NIR measurement of entire samples. The pixels of an HSI system have a field of view of 0.2 × 0.2 mm depending on the type of objective [2]. It considerably increases the chance to isolate a single chemical in one NIR measurement compared to conventional spectroscopy.

NIR HSI for food quality and safety inspection was proposed at the end of the 90s thanks to the advances in computer technologies [3]. Multiple measurement methods were used such as reflectance, transmittance or fluorescence and various wavelength ranges: Visible Near-Infrared (Vis-NIR), 400 – 1000 nm; NIR, 900 – 1700 nm; and Short Wave Infrared (SWIR), 1000 nm – 2500 nm [1]. The applications of NIR HSI in food safety are presented in three categories inspired by the classification provided by Feng and Sun [4].

D. Physical contaminations

Physical contaminations in food correspond to the presence of foreign materials in a product. They should be avoided to prevent hazardous effects for the consumers.

Hyperspectral imaging was intensively used to detect foreign materials in food matrices in the 2010s. In 2011, Bhuvaneswari *et al.* demonstrated the use of NIR HSI for detecting insect fragments in semolina [9]. Melamine and cyanuric acid contamination in soybean were investigated using NIR HSI by Fernández Pierna *et al.* in 2014 following the scandal of contaminated milk in China. Many other studies tackled the detection of melamine in milk powder using various chemometrics methods [10-13]. In 2015, Mishra *et al.* showed the detection of peanut particles in wheat flour using Principal Component Analysis (PCA) and Independent Component Analysis (ICA) [14-15] Later in 2018, Zhao *et al.* studied the detection of peanut and walnut powders in wheat flour using a Partial Least Squares (PLS) regression in NIR hyperspectral images [16]. These studies show the interest in detecting food allergen like nuts in raw materials and illustrated the capacity of NIR HSI for its detection. More recently, detecting foreign materials as bone fragments in chicken and microplastics in fish was demonstrated using NIR HSI [17-18].

E. Defects

Besides detecting foreign materials, NIR HSI was used to characterize food samples to detect their defects characterized by chemical changes.

Many applications of defects detection on fruits were demonstrated in the 2010s using Vis-NIR HSI. As an example, Gowen *et al.* showed the identification of mushrooms subjected to freeze damage using PCA and Linear Discriminant Analysis [19]. Apples were subject to numerous analyses using Vis-NIR HSI for detecting bruises, black pox or bitter pits [4]. More recently, Li *et al.* used Vis-NIR HSI for detecting skin defects on peaches showing this measurement method is still investigated for studying fruit defects [20]. Singh *et al.* used NIR HSI for studying insect-damaged wheat kernels [21].

This kind of applications consists of detection chemical modifications in a localized area of a sample. As such, it can be compared to the detection of contaminants in powder samples.

F. Microbiological contaminations

Hyperspectral imaging was used for bacterial determination in various studies tackling the freshness of fish and meat. Grau et al. studied the meat freshness on chicken breasts using SWIR HSI [22]. Fish freshness was also investigated and showed various model performances according to the measured portion of the sample [4].

Contamination of crops by fungi is another type of microbiological contamination that was studied using hyperspectral imaging. Jin *et al.* showed the potential of HSI for detecting toxigenic strains of fungi [23]. Later, Del Fiore *et al.* and Williams *et al.* both showed the potential of using HSI for detecting fungal contamination in maize kernels [24-25].

The detection of fecal contamination on food surfaces was one of the first applications in food safety at the beginning of the 2000s. Park and Lawrence developed a Vis-NIR HSI system for detecting feces and ingesta on poultry carcasses using a band ratio method [5]. This work was followed by other studies showing the possibility of implementing such a solution for the food industry [6]. Fecal contamination was also investigated using hyperspectral fluorescence imaging on leafy greens and apples [7] [8].

6. The penetration and the detection depth of NIR radiations

A. The penetration depth

The main theories of light scattering and absorption were established in the past century. The Radiative Transfer Equation (RTE) is the most general equation that describes the variation in intensity of an incident radiation when passing through an absorbing and scattering element [26]. Chandrasekhar proposed a solution in the case of two plane-parallel layers [27]. In practice, such calculations are not easily transposable to practical cases. Instead, other more straightforward solutions were proposed and known as the methods with one (Beer-Lambert) or two (Kubelka-Munk) fluxes.

The Kubelka-Munk theory is based on a model developed by Schuster in 1905 [28]. The goal was to simplify the RTE by proposing a two-fluxes model considered as isotropic (which means we neglect the fluxes' angular distribution as it occurs in diffuse lightning). In 1931, Kubelka and Munk proposed a similar approach to study the optical properties of paint layers [29]. Their theory defines the diffuse reflectance of material due to the interaction between inward and backward intensity fluxes. On the one hand,

because of the scattering effect, the diffuse reflectance signal comes from multiple infinitesimal layer of particles. On the other hand, the light intensity decreases with the layer's depth because of the absorbance effect. Consequently, the diffuse reflectance signal increases up to a given depth of material for which it becomes stable. Kubelka and Munk called this reflectance value R_∞ where the infinity symbol stands for the fact the material is thick enough. They propose the following formulation that involves the scattering and the absorption coefficients of the material S and K:

$$\frac{(1-\mathrm{R}_\infty)^2}{2\mathrm{R}_\infty} = \frac{\mathrm{K}}{\mathrm{S}}$$

Equation 1: The Kubelka-Munk's formulation for the reflectance with an infinite optical depth.

The main consequence of this concept is that the spectral information of a diffuse reflectance measurement is obtained on a limited thickness of the material. The concepts of penetration depth and effective sample size were introduced to take this phenomenon into account.

Berntsson *et al.* studied the effective sample size in diffuse reflectance and transmittance NIR spectroscopy by comparing two methods [30]. The first method uses calculations from the RTE and the 3-fluxes approximation detailed by Kuhn *et al.* [31]. The reflectance values R for increasing value of optical depth are calculated and compared to the theoretical R_∞. The optical depth such that R reaches a certain percentage of R_∞ is defined as the penetration depth of the sample. The second method is empirical and consists of measuring the sample several times by increasing its thickness (Figure 2). It results in a reflectance profile obtained for on a range of sample thicknesses. According to the RTE, the reflected signal increases until a certain limit noted R_∞. This reflectance profile is fitted using a negative exponential function as illustrated in Figure 2. This fit can be used to calculate the depth corresponding to a percentage of R_∞.

In both methods, the authors refer to the effective mass sample. This notion is equivalent to the effective volume sample or the effective depth sample. They all refer to the actual quantity of sample that is responsible for the reflectance signal.

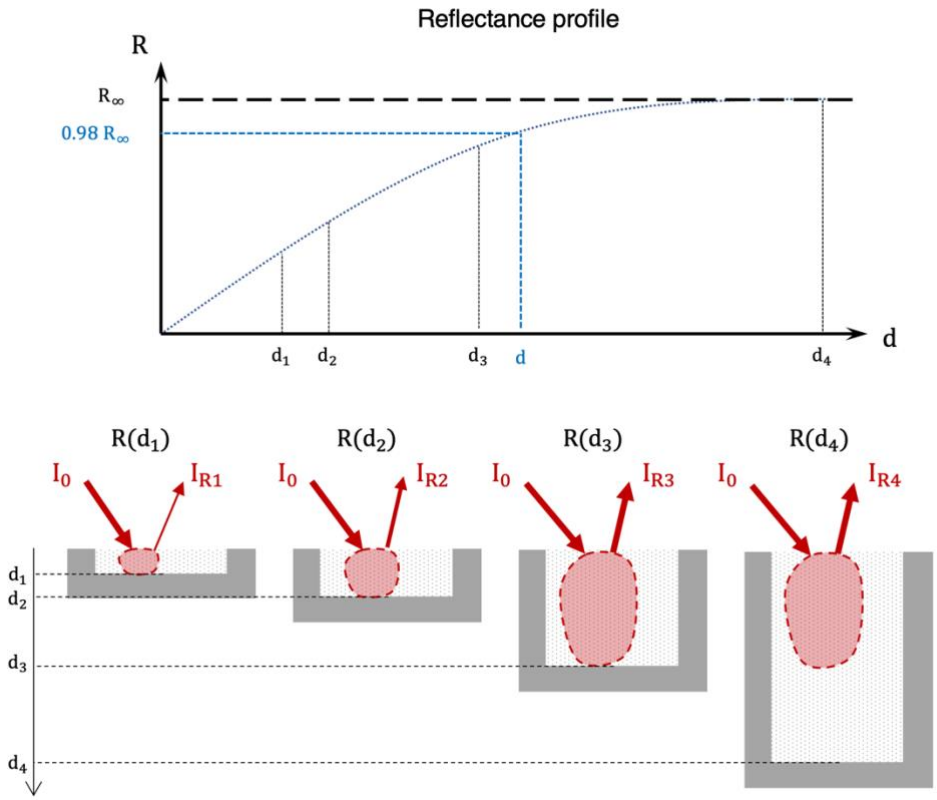

Figure 2: The empirical method for the determination of the effective sample depth.

Similar methods were used to determine the light penetration of radiations in different samples. Stolik *et al.* provided the light penetration depth of four Vis-NIR wavelengths in thirty types of "ex vivo" human tissues [32]. Ciani *et al.* studied the penetration of visible light in 19 different soils [33]. Lammertyn *et al.* focused on light penetration in apple slices [34].

B. The detection depth

The same context as Berntsson *et al.* can be considered (Figure 2) to define the concept of detection depth. A powder sample is lying on a flat and solid surface represented by the sample holder (in grey). In the study of Berntsson *et al.*, the flat surface is a black polyamide plate that absorbs all the radiations indifferently [30]. Because of this, there is no specific absorption pattern that can be identified in the reflectance signal.

Let us consider a new situation where a material with a specific absorption pattern replaces the polyamide black plate. The detection depth corresponds to the maximal thickness of powder that enables the absorption signal of the bottom material to be identified. Its detection can be achieved when its spectrum exhibits a strong absorbance at a given wavelength. If the powder sample does not have an absorbance pattern for the surrounding wavelengths, an absorption peak in the diffuse reflectance spectrum can be attributed to the bottom material. This is illustrated in Figure 3 which shows the influence of the bottom material identified for a specific wavelength λ_m.

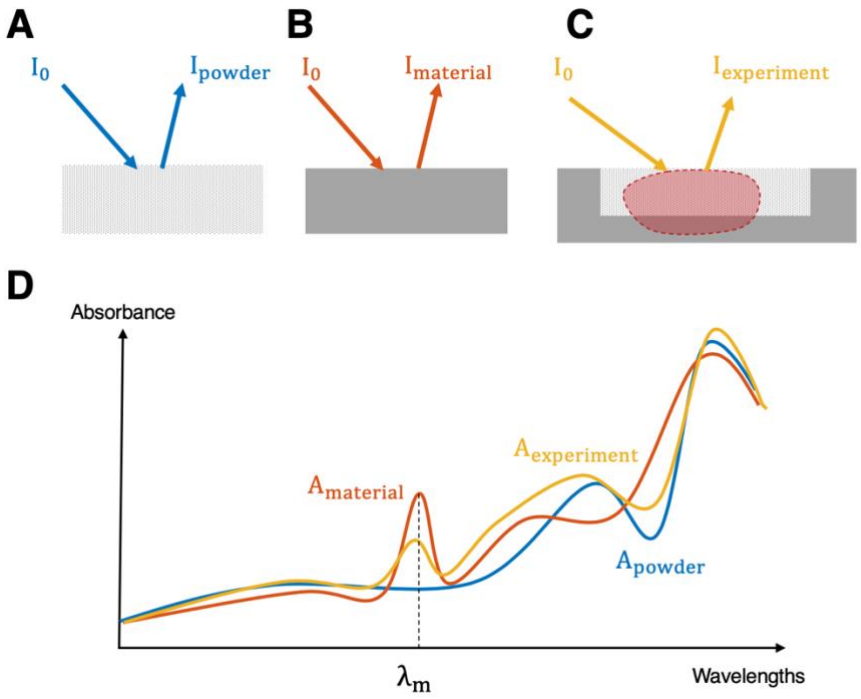

Figure 3: The comparison of three absorbance spectra showing the influence of the bottom material for in-depth reflectance measurement.

The theories regarding the diffuse reflectance are useful in describing simple situations where the scattering material is homogeneous. However, they are not efficient in modeling complex cases like the medium discontinuity between the powder and the bottom material because no simplification assumptions can be made. Instead, the literature shows that empirical approaches appear to be more realistic to achieve.

This approach was investigated by Huang et al. with regard to melamine detection in milk powder [35]. The authors prepared a few samples with a first layer of increasing milk powder thicknesses (from 1 mm to 5 mm) above a melamine powder layer. They applied a pre-trained classification model (Partial Least Square Discriminant Analysis) on the spectra measured by reflectance with a NIR HSI system. This model aimed to assess whether each pixel spectrum was pure milk or milk and melamine. The authors described that the performance of the model decreases in detecting milk-melamine pixels when the thickness of the milk layer increases. Thanks to this analysis, the authors were able to justify that a layer of 2 mm of milk powder was recommended to ensure the melamine beneath could be detected. This work provides a clear illustration of the detection depth and shows that the detection using NIR HSI should be made with a thin layer of material; otherwise, the melamine may be missed. However, there is still no general method to estimate the material depth at which the detection application is compromised. Moreover, this detection depth was not compared or discussed with the concept of penetration depth.

Despite the previously cited works on the field of penetration depth in NIR spectroscopy, there is still a lack of empirical study in the context of contamination

detection. To our knowledge, there is a need for a method measuring the detection depth of a given couple of materials in the context of detection using NIR HSI.

7. The detection of subpixel food particles

Raw materials in the food industry are often provided in powders because they guarantee a better stability [36]. Except for fish and meat, contamination detection in food using NIR HSI was often applied on powder. The particle size of food products may be smaller than 200 µm. For example, the Codex Standard defines that the particle size of wheat flour should be such that more than 98% of it should pass through a 212 µm sieve [37]. As a result, when measuring a mixture of different flours, the spectrum of a pixel $\boldsymbol{x}$ may not be representative of a pure chemical compound. The situation is comparable to that of the measurement of a heterogeneous sample using conventional spectroscopy. At this stage, the acquisition of a hyperspectral image needs to be coupled to the signal unmixing process. It means that the spectrum is modeled as containing two or more different pure chemical signatures in given proportions. These spectral signatures and proportions are generally unknown, and some chemometric tools must be used to solve this problem.

The unmixing problem is an essential subject in remote sensing. When the goal is to detect a target in a hyperspectral image, it is known as the subpixel detection problem. Various algorithms were designed to take this problem into account and tested on multiple hyperspectral datasets [38-39]. Most of these algorithms were developed using annotated datasets. They consist of a hyperspectral image containing various types of materials like trees, asphalt, corn, wheat, and so forth. An annotated image is provided to give the ground truth of each pixel of the image. For example, the dataset *University of Pavia*[1] is an image of a scene acquired by a hyperspectral sensor during a flight over Pavia, northern Italy. The ground-truth image gives a class to every pixel of the hyperspectral data. There also exist hyperspectral datasets for hyperspectral unmixing [40]. It means that unmixing algorithms can be evaluated on real data, which reveal their actual performance. This is the standard procedure to produce detection algorithms.

For food powder products, it is not possible to obtain similar ground-truth information. It would require knowing the spatial distribution of millions of particles of 100 µm of diameter and their chemical nature. In addition, annotated hyperspectral images for remote sensing analysis focus on the surface signal and do not consider multiple layers, which is essential in the case of food powders. Therefore, the result of the application of an unmixing algorithm is difficult to compare with the real situation.

The literature shows several applications of contamination detection in powders

[1] The dataset can be found on the website: http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

using hyperspectral imaging. According to the situation, the authors chose different modeling strategies.

A. Classification algorithms

One strategy consists of using classification algorithms to label each pixel of the hyperspectral image. This technique requires to have a clear spectral definition of both the adulterant and the material. Vermeulen *et al.* [41] studied the adulteration of cereals by ergot bodies on a conveyer belt using a NIR HSI. The authors used Support vector Machine (SVM) and Partial Least Square Discriminant Analysis (PLSDA) to discriminate the spectra. As ergot bodies and cereals are larger than one pixel of the camera, each pixel likely contains either ergot or cereals but not both at the same time. Hence, there is no spectral mixing to consider in the pixel. In that case, Vermeulen *et al.* showed that the classification method is efficient for the detection.

B. Spectral similarities

Another method consists of using spectral similarity analysis to compare spectra from the hyperspectral image with a reference. Fu et *al.* used this method to detect melamine adulteration in milk powders down to a concentration of 0.02 % [10]. In this case, the particles of melamine and milk powders are smaller than the pixel meaning the spectral signal may be mixed in the pixels. For this reason, the authors used a threshold on the spectral similarity scores to identify pixels containing melamine. Huang et *al.* studied the same melamine and milk powder case [12]. They used the band ratio method that consists of analyzing the reflectance ratio of two wavelengths. As previously, the authors proposed a detection algorithm based on a threshold of the band ratio.

C. Quantification methods

A third method consists of the calibration of an algorithm that quantifies the adulterant proportion in each pixel. Lim *et al.* used a PLS regression to detect and quantify the melamine in milk powders [13]. As shown in the previous works, the melamine could be detected at 0.02 % global adulteration. The PLS regression coefficients show the same important wavelengths as the one selected by the band ratio method in [12]. Zhao *et al.* studied the adulteration of wheat flour by walnut and peanut flour using PLS model calibration [16]. The authors concluded that the model could detect adulteration over 1 % of global concentration. They noticed that the localization of peanut and walnut particles was impossible with this methodology because of two reasons: the particle size which is smaller than the pixel, and the similar trends of spectral curves among pure samples.

D. Unmixing methods

The Linear Mixing Model

When subpixel particles are measured in one pixel, their spectral signatures all contribute to the observed mixed spectrum. The most widely used model is the linear mixing model (LMM) [42]. It assumes the resulting spectrum from a pixel is generated by the linear combination of the spectra of the constituents, here the powder particles.

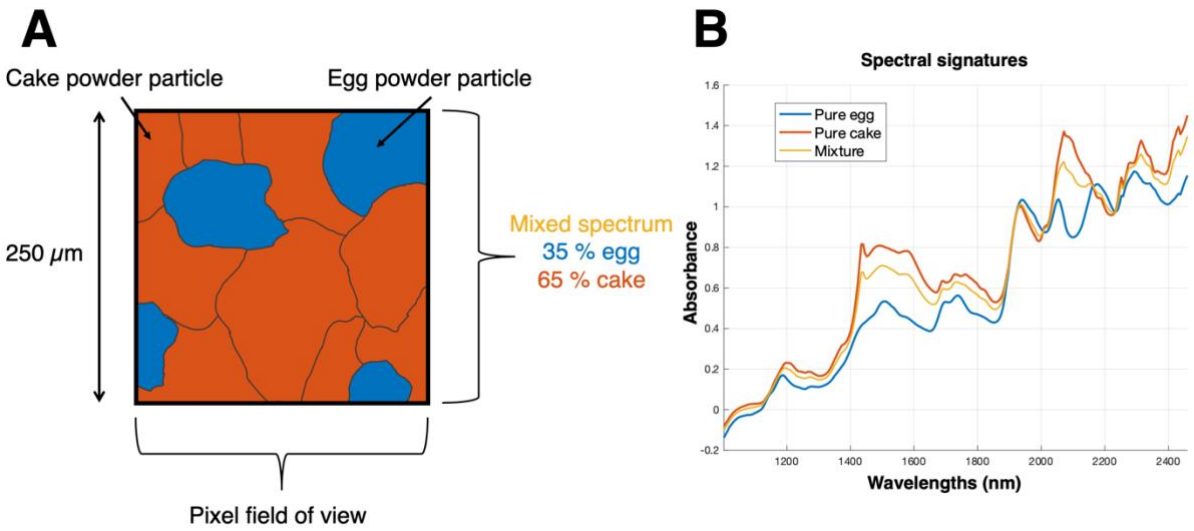

Figure 4: (A) A pixel is represented with several particles in its field of view. (B) The spectra of the pure ingredients and the mixture are represented.

The mathematical model of the LMM is given by [42]:

$$\boldsymbol{x} = \sum_{i=1}^{k} a_i \boldsymbol{s}_i + \boldsymbol{w} = \boldsymbol{S}\boldsymbol{a} + \boldsymbol{w}$$

Equation 2: The Linear Mixing Model.

$\boldsymbol{x}$ is the pixel spectrum; k is the number of constituents in the pixel; the $(\boldsymbol{s_i})_{1 \le i \le k}$ are the k components representing the spectra of pure constituents of the pixel; $(\boldsymbol{a_i})_{1 \le i \le k}$ are the contribution coefficients associated to the spectra; $\boldsymbol{w}$ is an additive noise vector.

In the case of adulteration detection, spectral unmixing consist of finding the appropriate spectral profiles. The contribution coefficients can be used to detect adulteration for each pixel. Several algorithms can decompose a matrix of spectral measurements into two matrices of pure spectral components and coefficients. For example, PCA assumes the spectral components are orthogonal to each other. On the other hand, ICA assumes they are independent of each other. Mishra *et al.* studied the detection of peanut particles in wheat flour using both methods [14] [15]. They obtained two different sets of components and performed peanut detection using a

threshold on the scores. Even if the peanut particles were larger than the pixel, the unmixing method helped discriminate the neighboring pixels containing both peanut and wheat contribution. The authors noticed that the presence of fatty acids in peanut help to discriminate the particle in wheat flour.

The Multivariate Curve Resolution model

For melamine detection in milk powders, the particle sizes are smaller than the standard pixel sizes in NIR HSI (0.2 × 0.2 mm) [2]. Huang *et al.* uses an unmixing approach to determine the concentration map of melamine and milk in mixed samples [11]. They compare the PCA, the Classical Least Squares (CLS) and the Multivariate Curve Resolution Alternative Least-Squares (MCR-ALS) approaches. The authors showed that the MCR-ALS approach provides the best quantitative results. This method is based on the following bilinear model:

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{\mathbf{T}} + \mathbf{E}$$

Equation 3 : The bilinear model for MCR.

This model considers a matrix $\mathbf{X}$ of size $n \times m$ that contains n spectra $\boldsymbol{x}$ stacked by rows. It decomposes $\mathbf{X}$ into a combination of k spectral components. Their spectral profiles are described in $\mathbf{S}$ of size $k \times m$. The combination of the spectral components for each pixel is described in the concentration matrix $\mathbf{C}$ of size $n \times k$. $\mathbf{E}$ describes the model residuals for each pixel in row.

The Alternating Least-Squares algorithm

Many algorithms were used in the literature to solve the MCR model [43]. The iterative methods are the most widely used solutions because they enable to introduce mathematical constraints during the optimization process [44]. The association of MCR with the Alternative Least-Squares (ALS) algorithm developed in 1995 is the most popular approach [45]. The cost function of the MCR model is given by:

$$\|\mathbf{X} - \mathbf{C}\mathbf{S}^{\mathbf{T}}\|_2^2 = \sum_{i,j} (x_{i,j} - \boldsymbol{c}_i \boldsymbol{s}_j)^2$$

Equation 4: Cost function of the MCR model.

where $\boldsymbol{c}_i$ and $\boldsymbol{s}_j$ are the parameters to optimize. This is not a convex cost function because of the interaction between $\boldsymbol{c}_i$ and $\boldsymbol{s}_j$. The ALS algorithm consists of alternatively fixing one parameter to have a simpler cost function. The solution can be calculated using the OLS similarly to the linear regression case. Hence the solution is given by:

$$\hat{\mathbf{C}} = (\mathbf{S}^{\mathbf{T}}\mathbf{S})^{-1}\mathbf{S}^{\mathbf{T}}\mathbf{X}$$
$$\hat{\mathbf{S}} = (\mathbf{C}^{\mathbf{T}}\mathbf{C})^{-1}\mathbf{C}^{\mathbf{T}}\mathbf{X}$$

Equation 5: Estimation of the concentration and spectral profiles using the ALS algorithm.

As the ALS is an iterative process, the algorithm starts with an initial guess either for $\mathbf{C}$ or the $\mathbf{S}$ matrix. Then, for each step of the process, $\hat{\mathbf{C}}$ and $\hat{\mathbf{S}}$ are calculated using the OLS (Equation 5) and enforcing the additional constraints. The hat notation indicates that the matrices are estimated. The algorithm continuously performs this two-steps estimation until a given stop criterion is reached. The criterion should evaluate how well the data are reconstructed using the $\mathbf{C}$ and $\mathbf{S}$ matrices. The lack of fit (LOF) is such a criterion:

$$\mathrm{LOF} = \sqrt{\frac{\sum_{i,j} e_{i,j}^2}{\sum_{i,j} x_{i,j}^2}}$$

Equation 6: The lack of fit for the MCR model.

In practice, the required LOF may never be reached by the algorithm. For instance, it may be the case when the constraints are too important or when the noise level is high. For this reason, a maximum number of iterations is also taken into account as a stop criterion as shown in Figure 5.

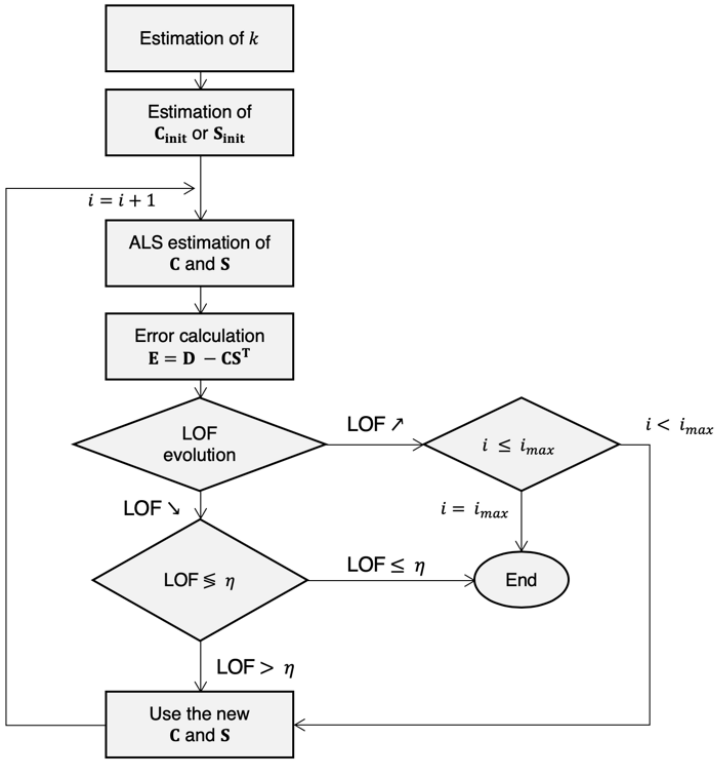

Figure 5: The schema of the MCR-ALS algorithm. The variable i denotes the number of iterations of the algorithm and i_{max} is the maximum number of iterations set by the user.

The ambiguity of the Multivariate Curve Resolution model

The MCR model (Equation 3) is ambiguous. It means that many solutions ($\mathbf{C}$ and $\mathbf{S}$ matrices) fit the model equally well. Although it is mathematically valid to consider that various solutions provide a model of the $\mathbf{X}$ matrix with the same precision, it is a problem for interpretation. For practical applications, the pure components should be unique and represent the pure constituents of the $\mathbf{X}$ matrix. The MCR-ALS literature identifies three kinds of ambiguities [43]:

- The **permutation ambiguity**: the order of the MCR components is not guaranteed in the $\mathbf{C}$ and $\mathbf{S}$ matrices. Let us assume a MCR model with $k = 3$; $\mathbf{C}$ of size $n \times 3$, $\mathbf{S^T}$ of size $3 \times m$. This model is mathematically identical if $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$, as well as $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$, are switched.

- The **intensity ambiguity**: the spectral intensity in the reconstruction of $\mathbf{X}$ is indifferently attributed to the concentration profile $\mathbf{C}$ or the spectral profile $\mathbf{S^T}$. In other words, the two following models are mathematically equivalent:

$$\mathbf{X} = \sum_{i=1}^{k} \boldsymbol{c}_i \boldsymbol{s}_i^T \text{ and } \mathbf{X} = \sum_{i=1}^{k} (\boldsymbol{c}_i a_i)(\boldsymbol{s}_i^T \frac{1}{a_i}) \text{ where } a_i \text{ are non-null scalar values.}$$

Because of this ambiguity, the spectral profiles obtained using MCR may have different scales which is a problem for interpretation.

- The **rotational ambiguity**: spectral profiles with different shapes can reconstruct the $\mathbf{X}$ matrix with the same precision. This ambiguity is probably the most problematic since it leads to a change in the shape of the spectral profiles. Hence, the pure spectral signature may not be recognized during result interpretation (Figure 6). Mathematically, the rotational ambiguity is expressed as follows: any rotation matrix $\mathbf{R}$ can be introduced in the model without changing the reconstruction of $\mathbf{X}$.

$\mathbf{X} = \mathbf{C}\mathbf{S}^{\mathbf{T}}$ and $\mathbf{X} = (\mathbf{C}\mathbf{R})(\mathbf{R}^{-1}\mathbf{S}^{\mathbf{T}})$ where $\mathbf{R}$ is a rotation matrix which satisfies: $\mathbf{R}\mathbf{R}^{\mathbf{T}} = 1$.

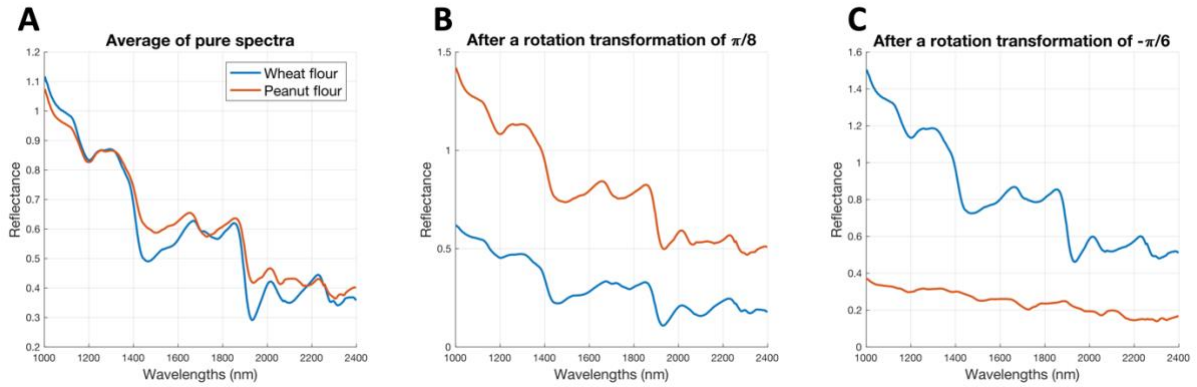

Figure 6: Effect of the rotational ambiguity on the spectral profiles. (A) The average of pure peanut and wheat spectra; (B) the same spectra after a rotation transformation of π/8 (C) and -π/6.

The ambiguities of the MCR model show that minimizing the reconstruction error of $\mathbf{X}$ is not sufficient to find the purest spectral signatures. Instead, additional criteria should be applied to reduce the set of possible solutions, i.e. the ambiguity of the model.

The constraints of the Multivariate Curve Resolution model

The non-negativity constraint imposes the concentration and/or spectral profiles only contain positive values. This constraint is relevant with NIR spectra because intensity, reflectance or, absorbance values should always be positive. The closure constraint imposes the sum of all contributions is equal to a constant (often 1), which implies interdependencies between the species contribution. This constraint can be seen as a type of normalization which affects the intensity ambiguity [46]. The knowledge of pure spectra or concentration profiles can be introduced as a constraint in the ALS procedure. It should be used when a profile is known.

Although many constraint methods were implemented in the MCR-ALS, it is still tricky to reduce the rotational ambiguity because it often requires a prior knowledge of the spectral profiles, which is not always possible. The consequence of a rotational ambiguity could be that the spectral signals are not well unmixed. It may lead to misleading conclusion, in particular for detection purposes.

Many constraints have been developed and could be useful to reduce the rotational ambiguity. The correspondence of species can be applied in the case of a multiset analysis. In this situation, a column-wise augmented matrix is used to indicate which experiment contains or does not contain a specific component.

The matrix augmentation is not a constraint as such, but it enables to reduce the rotational ambiguity of the MCR solutions [47]. It enables to introduce several matrices that share one dimension. The augmented matrix strategy is a possible solution by stacking the unfolded matrices on top of each other:

$$\begin{pmatrix} \boldsymbol{X_1} \\ \boldsymbol{X_2} \\ \boldsymbol{X_3} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C_1} \\ \boldsymbol{C_2} \\ \boldsymbol{C_3} \end{pmatrix} \boldsymbol{S^T} + \begin{pmatrix} \boldsymbol{E_1} \\ \boldsymbol{E_2} \\ \boldsymbol{E_3} \end{pmatrix} = \boldsymbol{C_{aug} S^T} + \boldsymbol{E_{aug}}$$

Equation 7 : The column-wise augmented matrix MCR model.

The selectivity constraint consists of providing concentration information to the ALS [48]. It is done by forcing the concentration matrix $\mathbf{C}$ to fulfill the constraints provided by the prior knowledge. For example, the concentration coefficient of one spectral component can be set to 0 for some spectra. The correlation constraint imposes than the concentration profiles in $\mathbf{C}$ have a sufficiently high correlation with reference data. It is a smoother way to impose that the concentration coefficients follow the prior knowledge.

Cordeiro Dantas *et al.* used MCR-ALS with a correlation concentration to detect adulterants in petroleum diesel using Raman spectroscopy [49]. They showed the constraint and the data augmentation strategy improve the quantification and the detection. Boiret *et al.* used a method to set local rank constraints for image resolution analysis with MCR-ALS [50]. The application of their method was the detection of a low dose compound in pharmaceutical compound using Raman microscopy. Although these two works show the potential of MCR-ALS for detection in spectroscopy, there is still no study aiming to detect food adulterant using NIR HSI. One reason can be that it is difficult to unmix signal composed of similar spectral profiles as it for food products in NIR spectroscopy.

The current methods for pixel unmixing in NIR HSI could be limited because of the subpixel problem and the ambiguity of spectral signatures in food products. Mishra *et al.* showed that the detection of peanut is feasible thanks to its fatty acids signature which is not present in wheat flour [14]. Quantitative models may be also limited in the case where subpixel detection is required as shown by Zhao et al. [16]. The MCR-ALS approach has the flexibility to provide constraints and solve the ambiguity problem. To our knowledge, the application of MCR-ALS for detection in food products using NIR

HSI has been quite limited. The application of constraints to reduce the model ambiguity seems to be the most promising technique.

E. Subspace detector

The last approach for the detection consists of modeling the spectral variability of samples to design a detector. The geometrical approach proposes a way to model the spectral variability. One spectrum is considered as a vector in a m-dimensional space, m being the number of wavelengths in the spectrum. The principle of the geometrical approach is to restrict the spectrum's variations in a lower dimensional space [42]. A spectrum $\boldsymbol{x}$ is described by:

$$\boldsymbol{x} = \sum_{i=1}^{L} a_i \boldsymbol{s}_i = \boldsymbol{S}\boldsymbol{a}$$

Equation 8: The model for spectral variability.

In Equation 8, $L < m$, and the vectors $\boldsymbol{s_i}$ define the variability subspace. These vectors are spectral signatures that can have multiple origins. It may be the spectral signature of a pure sample, or it may be obtained from a statistical method like PCA. The spectral signatures are the axis of the subspace, whereas the coefficients $\boldsymbol{a}$ are the coordinates of the spectrum.

The variability subspace described in Equation 8 can be obtained using PCA. The loading vectors are the weights of the PCs in the original space. Hence, they describe how the original variables are affected by the different sources of variability, i.e. the PCs. The scores describe the extend of the variability of the individuals for each component.

Figure 7 shows how spectral data are represented in the PC space. Figure 7A shows the feature space where one spectrum is represented by a black cross (or a vector). Its coordinates correspond to the contribution of each PC. Figure 7B shows the spectral contribution of the loadings and the reconstruction of the vector $\boldsymbol{x}$ by combining the first two components. Figure 7C shows the resulting spectrum for the spectrum in (A) after adding the average spectrum represented in dash line[2].

[2] PCA is commonly applied on the data matrix after mean centering.

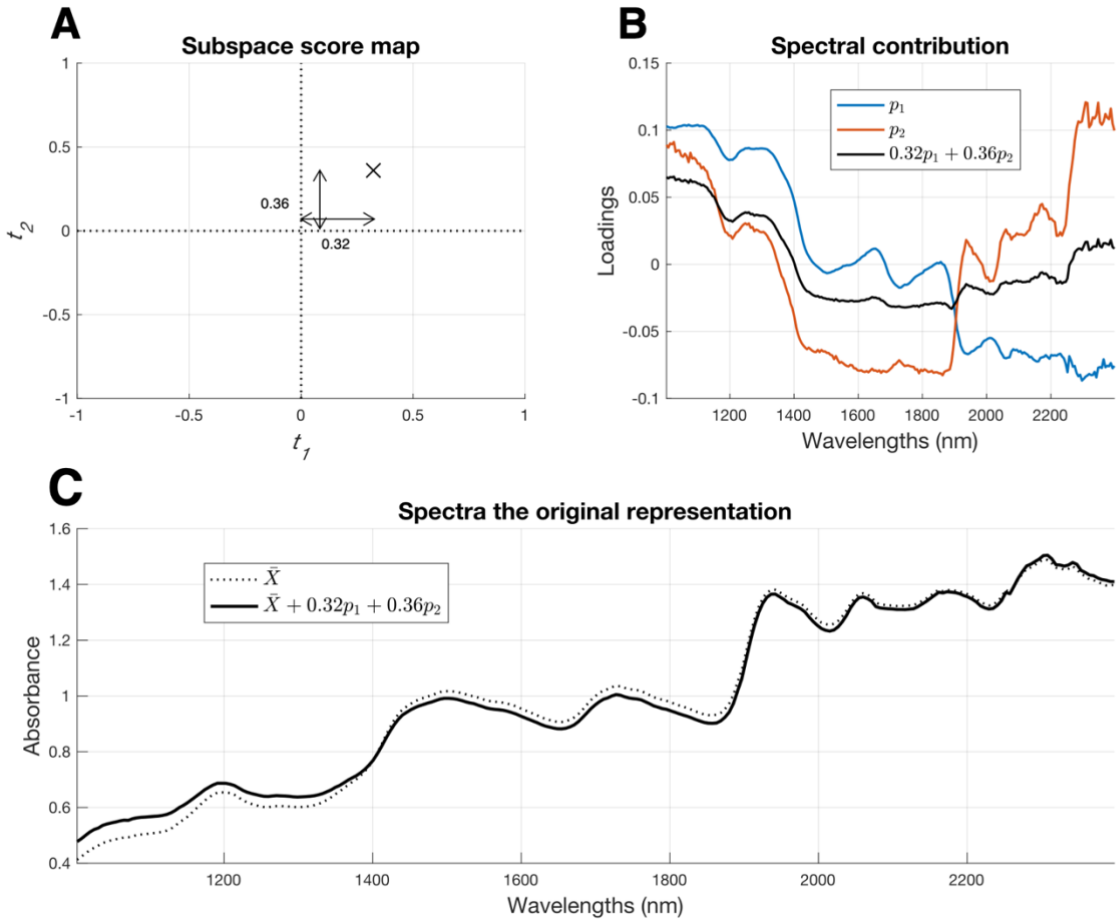

Figure 7: Representation of the variability using PCA. (A) the subspace PCA score map; (B) the two first loadings their combination; (C) the spectral rebuilding in the original space after adding the average spectrum.

The literature of NIR HSI for earth observation described multiple algorithms based on this modeling [51]. The Matched Subspace Detector (MSD) assumes that each pixel of a hyperspectral image falls into one of the two situations:

- Only particles from the standard sample are measured;
- Particles from both the standard and the adulterant samples are measured.

The situation where only particles from the adulterant sample are measured is excluded because this case can be treated by the second situation without difficulty.

In the first situation, the pixel spectrum can be described by Equation 8 considering only the spectral components that describe the food sample. In the second situation, the spectral components describing the adulterant spectral signature are added to the model as a linear contribution as stated by the LMM. Both hypotheses leads to the design of two matrices that are used to build the MSD. Each matrix describes a variability subspace were the sample to detect is described.

This detector was successfully used in the earth observation applications for target detection. In this case, the pixels' field of view is larger than targets of interest and the MSD was successfully applied to detect them. Du et al. used MSD to detect targets on

hyperspectral images [52]. Manolakis et al. provided an extensive description of subpixel target detector showing the MSD is particularly suitable for subpixel targets [51].

To our knowledge, such methods are not used for NIR HSI applications on food product. However, such a detector could be of great interest because it explicitly take into account the variability of the samples to detect. There is a need to develop this type of detector on food detection application.

I. The detection depth of a near-infrared hyperspectral imaging system

This part has been adapted from the publication:

A. Laborde, B. Jaillais, R. Bendoula, J.M. Roger, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, D. Bertrand, A. Boulanger, C.B.Y. Cordella, A partial least squares-based approach to assess the light penetration depth in wheat flour by near infrared hyperspectral imaging, J. Near Infrared Spectrosc. (2019). https://doi.org/10.1177/0967033519891594.

1. Introduction

There is a need for the study of powdered samples as they are often used in food industry processes and adulteration issues may occur in powdered raw materials. Even though HSI is able to resolve spectroscopic measurement on the surface of the sample, the volume that is screened is restricted. In fact, the penetration of light radiation is known to be limited because of scattering and absorption phenomena so that only a part of the product can be analyzed. This limitation is critical for quality-control applications and particularly for detection problems [54]. Indeed, when a sample is screened for adulteration checking, the whole sample should be analyzed to make sure the product is not contaminated. In this context, it is very important to know the actual volume of the screened sample. This knowledge enables technicians to know the best measurement conditions to ensure the detection. This is also an issue for powder homogeneity assessment using near-infrared spectroscopy (NIRS) [55] as the scale of scrutiny may be limited.

Powdered samples are mainly measured in reflectance mode. For diffuse reflectance, the spectroscopic sensor measures photons that are back-scattered in the sample or reflected at its surface. As the path length increases in the sample, the chance to be absorbed increases. As a consequence, there are much fewer photons that come back from the deepest layers of the sample than from the surface. For a certain depth, the amount of signal received by the sensor is similar to the noise measurement and it is not possible to retrieve any spectral information from this depth. Additionally, the amount of spectral information needed for detection may vary according to the chemical species. For example, melamine and milk powder [56] have two very distinct spectral signatures but it is less true for wheat flour and peanut particles [14]. As a consequence, the required signal-to-noise ratio for melamine detection in milk is likely to be smaller than those for peanut detection in wheat flour. Thus, the perceived detection depth may be different for the two cases whereas neither the sensor nor the physical phenomenon have changed.

The problem of light penetration depth for detection involves three considerations: first, the physical light penetration phenomenon in powders; second, the sensor dynamic range and finally the spectral signatures that have to be unmixed.

As hyperspectral imaging is more frequently used for detection problems, there are two important needs with respect to this light penetration issue. First, the need for an empirical method to be able to determine the maximal depth for detection regarding a given application. Then, a better understanding of the phenomenon that encompasses the physical phenomenon of light penetration, the sensor dynamic and the unmixing application case.

The Kubelka-Munk theory [29][57-58] provides some understanding about the penetration depth of near infrared radiations. Using a model with two fluxes of photons, it shows the diffuse reflectance for an infinitely deep sample (R_∞) depends on the ratio between the scattering and absorption constants [54]. Although this theory assumes the sample is isotropic, the derived formula for diffuse reflectance is a central point for the study of penetration depth in powdered samples. According to this theory, the diffuse reflectance signal comes from different layers of the sample. When the thickness of the sample increases, the measured reflectance R increases to a given limit R_∞. Deeper layers of the sample do not contribute to the reflectance. This concept is relevant for the determination of penetration depth. In past decades, many authors have studied the penetration depth subject according to different points of view.

Olinger et al. proposed an approach to determine the number of interrogated particles by comparing the baseline-corrected value of the pseudo-absorbance $\log_{10}(1/R)$ to the absorbance per particle [59]. The authors deduced that for an absorbing matrix like carbazole, the penetration depth is less than 1 mm. Berntsson et al. have provided further understanding about the effect of sample thickness on diffuse reflectance measurements [30]. According to them, penetration depth in a sample is related to the depths from which the diffuse reflectance signal originates. Following this concept, they introduced the effective sample size which defines the sample mass which is sufficient to reach 98% of the diffuse reflectance of a corresponding optically thick sample (R_∞). These results show that, for a powdered sample, the diffuse reflectance signal comes from different depths down to a certain level. This level is defined as the penetration depth or the effective sample size. Berntsson et al. proposed two methods for determining this depth and provided results for radiations between 400 nm and 2500 nm. For microcrystalline cellulose powder, the penetration depth shows a global decreasing behavior between 1000 nm and 2500 nm with penetration depth varying between 2.0 mm and 0.33 mm.

Stolik et al. measured human tissues in transmittance mode in order to determine their penetration depth [32]. The authors used the one-dimensional diffusion model where the penetration depth plays the role of the distance constant in the exponential decreasing law of intensity. By measuring the transmitted flux through the tissue for different thicknesses, the authors determine the penetration depth at different wavelengths. According to this definition, the penetration depth is the thickness of material that attenuates 63% of the incoming flux. Reported results show

penetration depth values vary between 0.1 and 3.0 mm for different kinds of human tissues at different visible light wavelengths.

Lammertym et al. used reflectance diffuse measurements on apple slices in order to determine the penetration depth of near infrared radiations [34]. By successively slimming the apple slice, the authors obtained the reflectance measurement for different thicknesses and fitted a decreasing exponential curve for each wavelength of the range. Results were similar to Berntsson et al. and the authors found a penetration depth between 2 mm and 4 mm in apples.

More recently, Padalkar and Pleshko have worked on the light penetration depth into cartilage to ensure the signal is not corrupted by underlying subchondral bone [60]. An empirical method was employed using a disk of polystyrene placed behind cartilages of different thicknesses. As this thickness increased, authors showed the signal of polystyrene decreased until it became invisible at visual inspection of spectra. From their protocol, the penetration depth is defined as the sample thickness for which the signal of the polystyrene target does not contribute to the diffuse reflectance measurement.

Huang et al. used a similar protocol by placing melamine under different thicknesses of milk powder [35]. The authors showed that a PLSDA failed to detected melamine contribution for a thickness of milk powder larger than 2 mm.

The literature shows that light penetration depth in spectroscopy can be studied through different underlying definition of the phenomenon. Some authors rely on theoretical models such as Kubelka-Munk or the diffusion model, whereas others use an empirical method that is specific to the application such as melamine and milk powder [35] or light penetration into cartilage [60]. As the literature shows, and to our best knowledge, no study offers a multivariate chemometric approach, sensor considerations and theoretical interpretation of the phenomenon.

This work studied the penetration depth of near infrared radiation into wheat flour using a sample holder of PLA. Hyperspectral imaging was used to acquire a great number of spectra with spatial information. PLS regression was used in order to quantify the amount of spectral signature coming from PLA along the sample holder. Finally, an interpretation of the phenomenon is proposed using the Kubelka-Munk theory and sensor considerations.

2. Material and methods

A. Samples

A sample holder has been designed and manufactured for this experiment using a 3-dimensional printer (Figure 8). The central cavity was designed on a gradient to contain powdered samples of varying thickness. The powder is skimmed on the top of the sample holder so that the thickness is graduated from 3.5 mm to 0.5 mm. The sample holder is made of PLA which has a specific absorbance peak in the near infrared spectral

range (1168 nm). White wheat flour (Francine, batch number 138, France) was used for the investigation of light penetration depth. Two replicates from the same pack were used for the measurement. Since packing density may have an effect on light behavior in the sample, the powder was not forced into the sample holder. Instead, wheat flour was sprinkled over the sample holder and skimmed in order to ensure the repeatability of the sample packing.

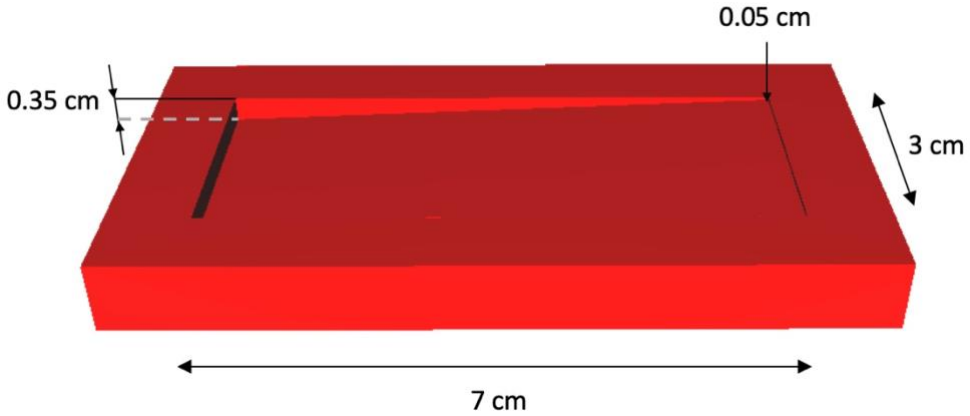

Figure 8: Schema of the sample holder.

B. Hyperspectral imaging system

A line-scan pushbroom HySpex SWIR-320m-e camera (Norsk Elektro, Skedsmokorse, Norway) was used to acquire hyperspectral images. The spectral range was 1000 – 2500 nm and 256 spectral bands were acquired, leading to a spectral resolution of 6 nm. The camera acquired 320 pixels per line. Two halogen lamps were used to illuminate the sample. A standard white diffuse reflectance standard (Spectralon®, SRS-99-010, Labsphere) was used to acquire the white reference image. The black reference image was acquired by closing the shutter of the camera.

C. Data processing

Images were cropped to focus on the central cavity of the sample holder leading to 100 × 246-pixels images. The white reference image was averaged to obtain one spectrum for every pixel of sensor's line (I_0). The black measurement (I_B) and the white reference were used to calculate the reflectance signal from the raw measurement (I) using:

$$R = \frac{I - I_B}{I_0 - I_B}$$

Equation 9: The reflectance calculation.

As they exhibit a low signal-to-noise ratio, the first wavelengths (smaller than 1100 nm) of the spectra were removed. A Savitzky-Golay filter was applied (2nd order polynomial, 7-points window and no derivative) to smooth the spectra. Finally, a log transformation ($-\log_{10}$) was applied to obtain absorbance spectra only for PLS application.

D. Thickness target values

The sample holder designed for the study is made such that the thickness of wheat flour varies. In the following, wheat flour thickness is referred as the $\boldsymbol{y}$ target value. This thickness depends on the sample holder geometry. As a consequence, the $\boldsymbol{y}$ target vector is constructed using the geometry of the central pit of the sample holder. Since it is designed as a slope between 0.05 cm and 0.35 cm, a linear interpolation vector was created and assigned to each of the 100-pixel lines across the sample holder. This procedure leads to a 2-dimensional mask that can be applied on the hyperspectral image (Figure 9). For spectral analysis, hyperspectral cubes are unfolded to obtain matrices of 24 600 lines and 256 columns. The 2-dimensional mask for $\boldsymbol{y}$ values is unfolded in the same way so that each spectrum of the matrix is associated with the appropriate $\boldsymbol{y}$ target.

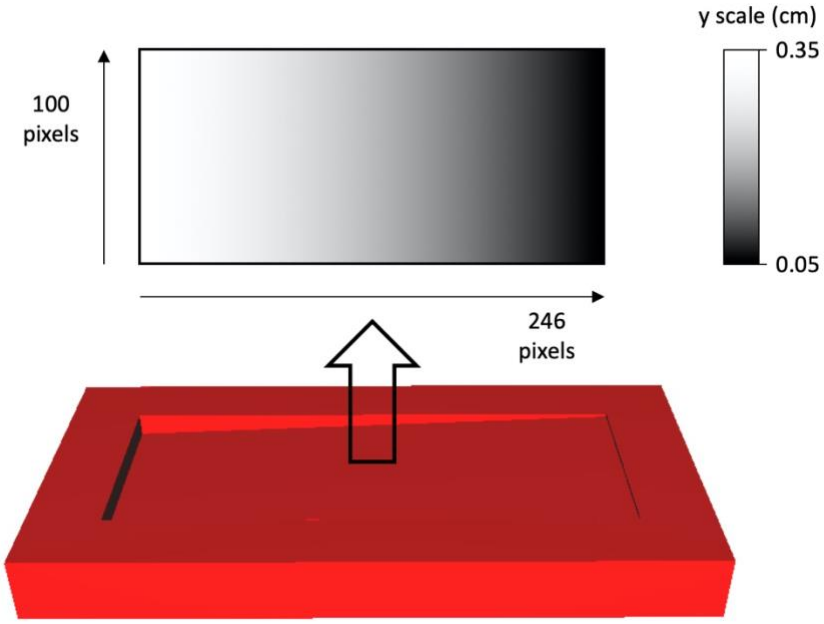

Figure 9: The construction of the two-dimensional mask for thickness target values.

E. Reflectance profile extraction

The reflectance profiles across the sample holder were extracted for each wavelength following the procedure described in Figure 10. All the pixels on the same vertical line

were averaged in order to obtain one spectrum for each $\boldsymbol{y}$ thickness value (step 1 to 2). As a result, a 2-dimensional matrix was obtained as well as the corresponding vector of $\boldsymbol{y}$ target values. After selecting a wavelength band, all the corresponding reflectance values were extracted and plotted against the $\boldsymbol{y}$ thickness values giving the reflectance profile (steps 3 to 4).

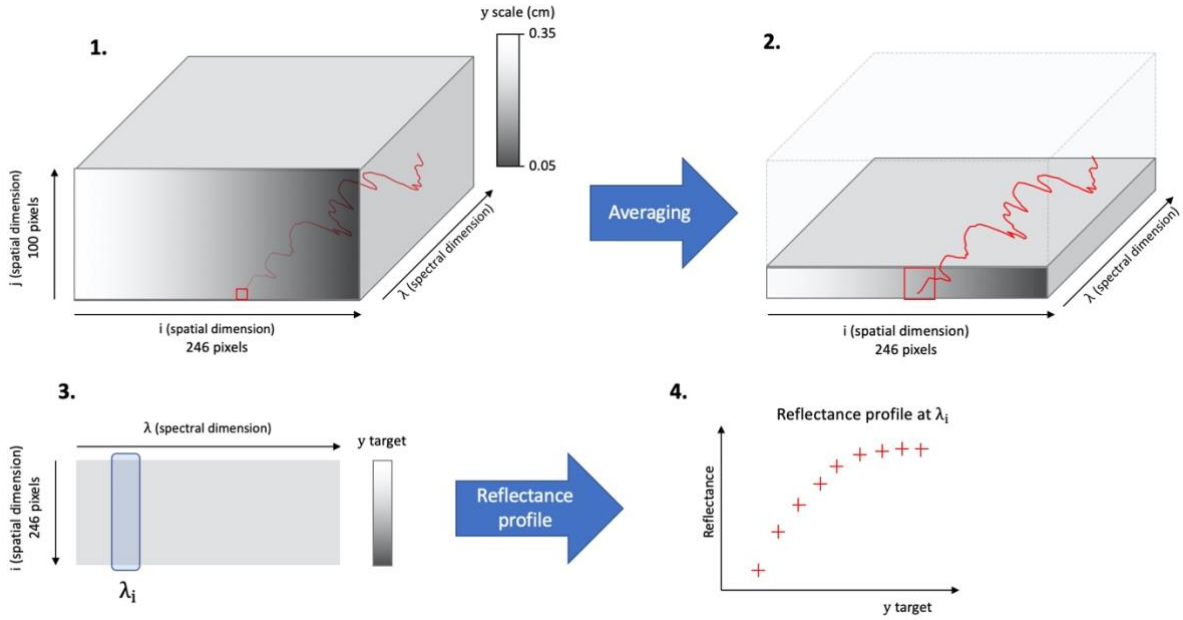

Figure 10: The procedure for the reflectance profile analysis.

F. Partial Least-Squares Regression

The PLS regression is an algorithm used for predicting a target value $\boldsymbol{y}$ using predictors $\mathbf{X}$ with a linear relationship: $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$. PLS is a good alternative to classical Multiple Linear Regression (MLR) or Principal Component Regression (PCR) when predictors are NIR spectral data. For this kind of data, there are a great number of variables (several hundreds) that are mostly correlated to each other. As a consequence, the construction of orthogonal latent variables is required for applying multiple linear regression. PCA is one method used for constructing such variables that are orthogonal and ranked according to the amount of variance they represent in $\mathbf{X}$. PCR is achieved by performing MLR on these new variables. However, PCR does not take into account the relationship with target values $\boldsymbol{y}$ in the construction of the orthogonal latent variables. PLS solves this problem by constructing latent variables based on the covariance between $\mathbf{X}$ and $\boldsymbol{y}$ [20 - 21]. PLS has been widely used in chemometrics as it is particularly suitable for near infrared spectral data [61]. In this study, PLS is used in order to quantify the amount of PLA signal in the diffuse reflectance measurements. It is assumed that the signal of PLA is linked to the wheat flour thickness in the sample holder. As a consequence, the $\mathbf{y}$ thickness vector is used as target for the PLS calibration. The training was performed using cross-validation on the first sample replicate. 70% of the spectra from the cube were used for calibration and 30% for validation. This procedure

was repeated 10 times to select the number of latent variables associated to the averaged minimum root mean square error of cross-validation (RMSECV).

3. Results and discussions

A. Reflectance evolution for each wavelength

Figure 11 shows the reflectance spectral signatures of PLA and wheat flour. The spectrum of PLA exhibits high and resolute absorption peaks all along the near infrared range. The absorption peak at 1168 nm represents a high difference in reflectance between PLA and wheat flour. Figure 12 shows the reflectance profile at 1168 nm corresponding to this absorption peak. The experimental points exhibit a curve showing two behaviors. The first part of the curve corresponds to low thickness values and shows an increasing reflectance profile. The second part shows a stabilization of the reflectance level for high thickness values.

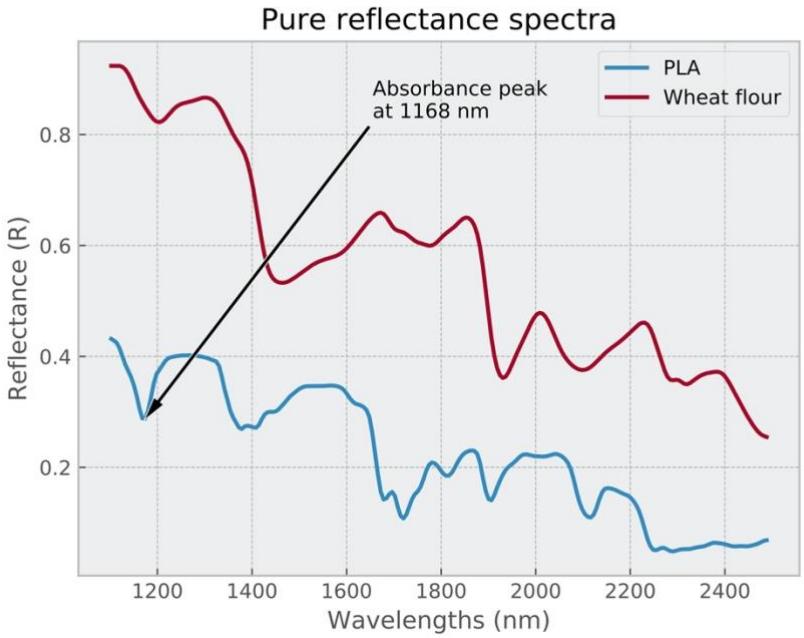

Figure 11: The pure reflectance spectra of wheat flour and PLA.

When the wheat flour thickness is low, the PLA plays an important role in the resulting diffuse reflectance signal. As it absorbs radiation around 1168 nm, the reflectance profile at this wavelength starts with low reflectance values. When the thickness increases, the role of wheat flour becomes more important than PLA in the resulting reflectance spectrum. Since wheat flour absorbs much less than PLA at 1168 nm, the reflectance level increases. This behavior can be interpreted using the theory of Kubelka-Munk presented in the next section.

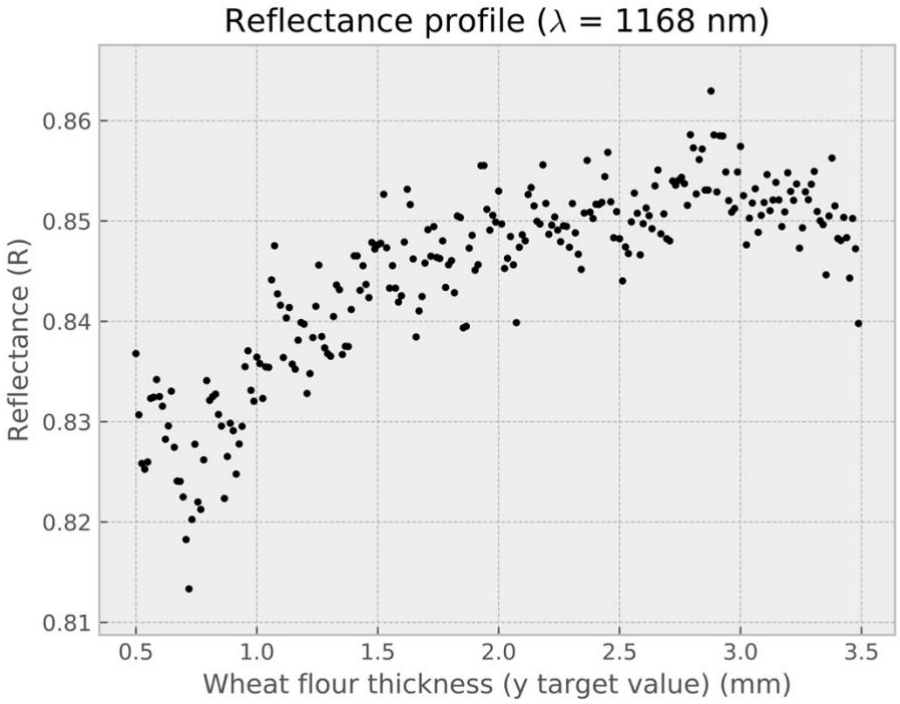

Figure 12: The reflectance profile at 1168 nm through increasing depths of wheat flour in the PLA sample holder.

B. Physical interpretation

The Kubelka-Munk model

The Kubelka-Munk model developed on a monochromatic case will be applied to the following one. It is assumed the results are applicable to every wavelength of the detector range between 1100 and 1350 nm. Let us consider a layer of wheat flour of thickness $\boldsymbol{y}$ as shown in Figure 13. This case corresponds to a slice of the sample holder for a fixed thickness value. The surface of the sample holder is assumed to be wide enough so that the influence of borders can be neglected for the application of the Kubelka-Munk theory. The wheat flour is lying on a layer of PLA with a reflectance $\mathrm{R_g}$. An infinitesimal layer of thickness dz is considered in the wheat flour at the height z. This layer is crossed by two fluxes: the descending flux $\mathrm{i(z)}$ and the ascending flux $\mathrm{j(z)}$. The wheat flour is assumed to be isotropic so that global absorption and scattering coefficients can be defined by K and S respectively. Taking into account the changes for both fluxes when crossing the layer of wheat flour leads to the following equations:

$$-\mathrm{di(z)} = -\mathrm{Ki(z)dz} - \mathrm{Si(z)dz} + \mathrm{Sj(z)}$$

$$\mathrm{dj(z)} = -\mathrm{Ki(z)dz} - \mathrm{Si(z)dz} + \mathrm{Si(z)}$$

These two equations result in a differential equation system:

$$\begin{cases} \frac{di(z)}{dz} = (K+S)i(z) - Sj(z) \\ \frac{dj(z)}{dz} = -(K+S)i(z) + Si(z) \end{cases}$$

Kubelka and Munk proposed a solution to this system [23] which only involves the coefficients K and S as well as the reflectance of the sample holder R_g:

$$R = \frac{1-R_g(a-b\coth(bSy))}{a-R_g+b\coth(bSy)} \text{ where } a = \frac{K+S}{K} \text{ and } b = \sqrt{a^2-1}$$

Equation 10: Solution of the Kubelka-Munk model.

This solution (Equation 10) is suitable to describe the situation in the sample holder. However, only the wavelength range 1100 – 1350 nm is considered because it exhibits significant signal from PLA at lowest thickness values. In that context, R_g, K and S remain constant at each wavelength whereas y increases according to one dimension. As a consequence, each reflectance profile can be modeled by this relationship by considering y as a variable.

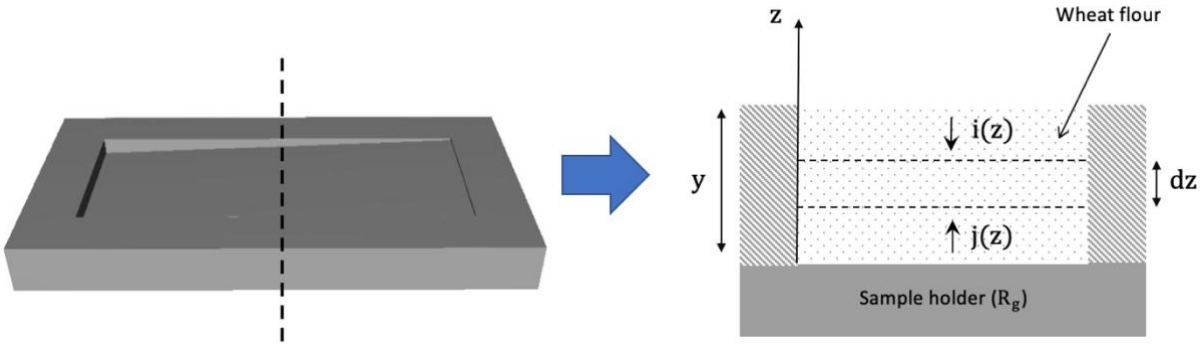

Figure 13: The Kubelka-Munk formalism applied on a slice of the sample holder.

The derivation of the Kubelka-Munk model shows the reflectance measurement (R) approaches a limit R_∞ when the thickness value approaches infinity. In this situation, the sample is so thick that the presence of the reflective background has no effect on the measurement. This theoretical value can be obtained using the boundary condition $i(0) = 0$ (with $j(0) \neq 0$). The final expression of R_∞ only depends on K and S [62]:

$$R_\infty = 1 + \frac{K}{S} - \sqrt{\frac{K^2}{S^2} + 2\frac{K}{S}}$$

Equation 11: Reflectance for an infinitely thick sample according to Kubelka-Munk.

Theoretically, the value of R_∞ is never reached. In other words, the sample holder has an influence on the reflectance measurement for every value of thickness $\boldsymbol{y}$.

Sensor considerations

In practice, the reflectance measurement is always corrupted by noise. This measurement noise is mainly caused by electronic components of the hyperspectral camera (thermal noise, shot noise) [63] and leads to variations in the reflectance measurement for the same configuration. As a consequence, the reflectance level is measured with a certain level of uncertainty. This explains why the measurement points on the reflectance profile (Figure 12) describe a curve with a certain width.

This kind of noise can be counteracted by increasing the exposure time of the measurement as long as the sensor is working in its linear phase. With this method, the signal-to-noise ratio (SNR) is improved so that the uncertainty of reflectance measurement is decreased. However, most of the time, increasing the exposure time is not possible because of pixel saturation. The sensor of a hyperspectral camera has a linear response on a finite range of photon flux. If the photon flux is more powerful, the pixel sensor is saturated, and the information is lost. If the photon flux is not powerful enough, the signal is drawn into measurement noise. The ratio between this largest and smallest flux corresponds to the dynamic range of the sensor. In practice, the exposure time of a camera is tuned so that the exposure time is high enough to have a high SNR. The limit is set to avoid pixel saturation. The saturation is mainly due to specular reflectance on the surface of the sample. Indeed, this source of the signal is, by nature, more powerful than the signal of interest that is partially absorbed.

Even if the value of R_∞ cannot be reached theoretically, it can be measured in practice because of measurement noise. As a result, for a given thickness value y, the reflectance value R is so close to the theoretical limit R_∞ that it can be reached because of uncertainty. For this thickness value, the corresponding reflectance measurements do not carry any distinguishable information about the reflectance signal of the sample holder. In this context, the penetration depth of the signal is reached. Since this notion is depending on sensor considerations as well as the nature of the sample holder, the notion of detection depth should be more suitable. However, in the following, the penetration depth notion is used to describe results obtained using the Kubelka-Munk theory.

C. Determination of the penetration depth

The reflectance profiles obtained for each wavelength (Figure 12) correspond to the context of the Kubelka-Munk theory. The experimental points are used to fit the model defined by Equation 10 as a function of the sample thickness y. The three parameters

K, S and R_g are estimated using non-linear least squares fitting. The penetration depth is estimated by applying the following procedure for each wavelength (Figure 7):

1. The root-mean-square error (RMSE) of the modeling is calculated between the experimental points and the fitted function (Equation 10). This value is chosen to represent the spread of the experimental points around the fitted curve $R(y)$.

2. The value of R_∞ is calculated using Equation 11.

3. The threshold reflectance value R_T is calculated as the difference between the reflectance limit R_∞ and the RMSE: $R_T = R_\infty - \mathrm{RMSE}$.

4. The penetration depth y_p is determined such that $R(y_p) = R_T$.

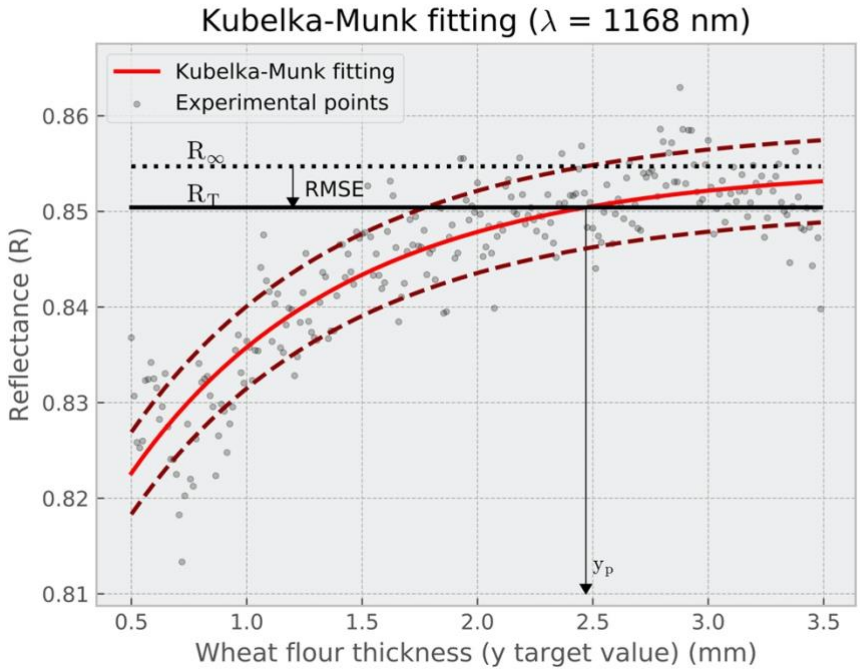

Figure 14: Procedure for the calculation of the penetration depth at 1168 nm.

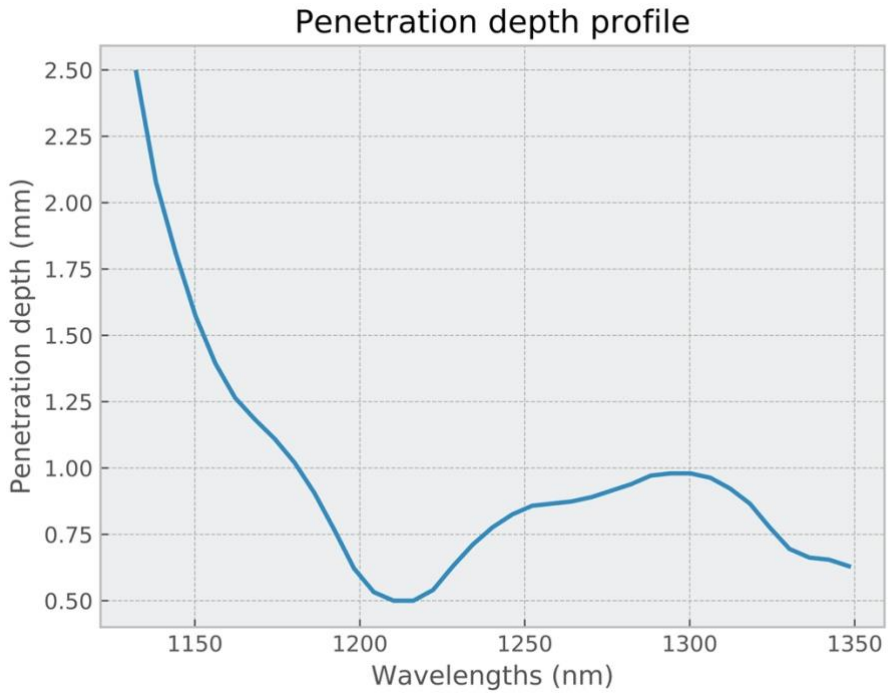

Figure 15: Penetration depth profile obtained from the reflectance profile measurements.

The penetration depth profile obtained is presented on the Figure 15. The profile is estimated on the range 1100 – 1350 nm because the Kubelka-Munk model fitting cannot be applied on higher wavelengths. In these conditions, the reflectance profile is flat because the PLA spectral signature is not visible for any wheat flour thicknesses. Otherwise, the resulting curve shows the penetration depth is highly dependent on the wavelength of NIR radiations. It is higher for smaller wavelengths (from 1100 nm to 1150 nm). The profile can be explained by the absorption coefficients of the pure materials (Figure 11). Indeed, it is similar to the reflectance spectral signature of wheat flour. Consequently, the penetration depth is smaller for wavelengths at which wheat flour absorbs (1210 nm). The lower values for penetration depth between 1100 nm and 1170 nm can be explained by the strong absorption of PLA between 1123 nm and 1211 nm.

D. Partial Least-Squares regression results

In this study, the use of PLS regression can be compared to unmixing problems. In this situation, the aim consists in finding the k pure spectral endmembers ($\boldsymbol{s_i}$) associated with their proportions (a_i) in the linear mixing model (Equation 2) that decomposes the signal of the measured spectrum ($\boldsymbol{x}$).

In the situation of the sample holder filled with wheat flour, the spectrum of each pixel can be modeled by a mixture between the spectral contributions of wheat flour and PLA with an additional residual vector ($\boldsymbol{w}$). However, solving the linear mixing model for each pixel requires some knowledge about the endmembers ($\boldsymbol{s_i}$). In our

study, the pure spectra of wheat flour and PLA may not be relevant. Indeed, linear combination of spectral signatures is an appropriate model when the spectral mixture occurs in the sensor. In our case, the mixture relies on nonlinear mixture effects [64] [65] [66]. As a consequence, making assumptions on the form of spectral endmembers in the case of linear mixture model may be not appropriate. As we can consider this problem is a two spectral signature mixing, quantifying the proportion of one element is sufficient. Consequently, PLS can be seen as a method to solve the mixture model within the sample holder. Thus, the resulting prediction $\boldsymbol{y}$ represents the contribution of PLA in the spectra as well as the wheat flour thickness. Instead of using assumption on the spectral endmembers, PLS must be trained with a training set of predictors $\mathbf{X}$ and target values $\boldsymbol{y}$.

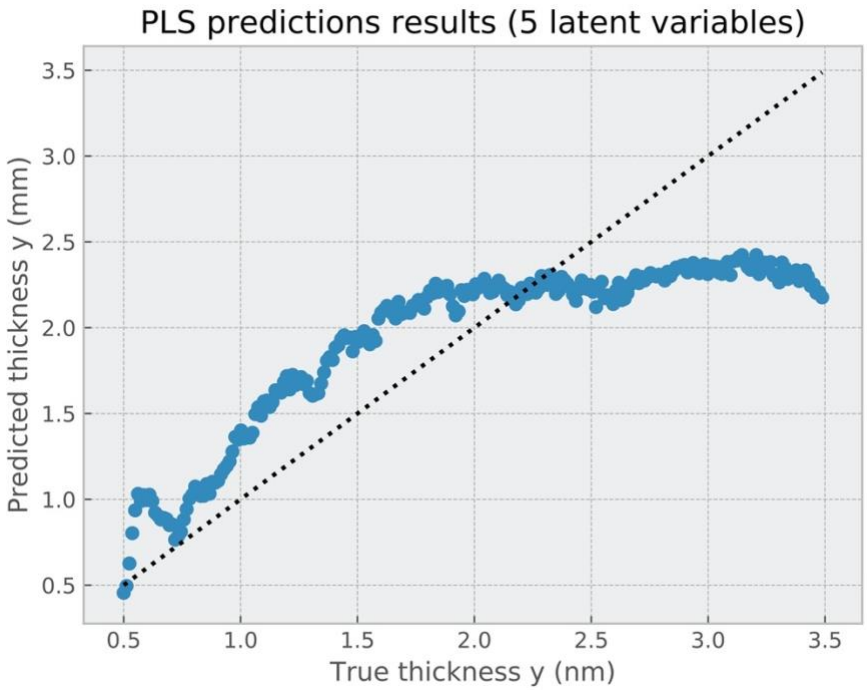

Figure 16: PLS prediction results for wheat flour thickness across the sample holder.

PLS model was trained on the spectra issued from the training images. Five latent variables were chosen as more variables did not improve the cross-validation error. The obtained model was then applied on the image of a different wheat flour sample. Each pixel gave a prediction leading to 24 600 prediction points. All predictions coming from pixels of the same line (same wheat flour thickness $\boldsymbol{y}$) were averaged to improve visualization. As a result, 246 prediction points were obtained and plotted on a graph (Figure 16) against the real wheat flour thickness values. These results show two types of behaviors. The first part of the prediction curve shows a monotonic increasing behavior for low thickness values (between 0.5 mm and 1.5 mm). Thus, PLS regression exhibits a correlation between the measured spectra and the corresponding wheat flour thickness. The measured reflectance data follow a mixing model between wheat flour and PLA spectra. The results show that PLS is able to fit this mixing model

for low thickness values. For thickness values higher than a given value, PLS prediction results do almost not evolve. The PLS model does not exhibit any correlation between the measured spectra and the wheat flour thickness. The mixing model between wheat flour and PLA spectra is not fitted by the PLS model. As a consequence, there is a change in the mixing behavior between PLA and wheat flour. As the predictions remains constant, the PLS model interprets high thickness wheat flour spectral data as pure wheat flour spectra. In this situation, the signal coming from the PLA is so weak that its influence becomes comparable to the measurement noise.

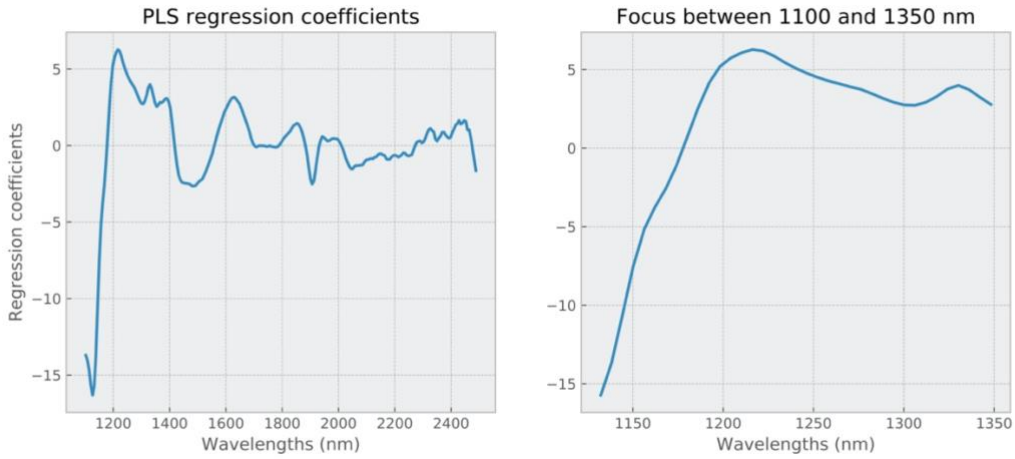

Figure 17: Regression coefficient of the PLS model used for the prediction of the wheat flour thickness.

Figure 17 shows the PLS regression coefficients for each wavelength. The high weights attributed to the low wavelengths show the importance of these variables for quantifying the PLA spectral signature across the sample holder. The right part of Figure 17 focuses on the range 1100 – 1350 nm to highlight the similarities between the regression coefficients and the penetration depth results obtained (Figure 15).

As these similarities show, the PLS unmixing method and the reflectance profile fitting method are relevant to each other. The PLS method provides a multivariate analysis of the problem so that all wavelengths contributions are taken into account in the prediction result profile (Figure 16). Thus, by extrapolation, the Kubelka-Munk theory and the sensor considerations explain the behavior observed on the PLS prediction results (Figure 16).

Indeed, for a given wheat flour thickness y, the signal of the PLA cannot be extracted from the diffuse reflectance spectrum. This may be an issue for detection problems when the target is buried under a layer of sample. For this reason, it is important to define the higher thickness y_d for which a detection is possible. This limit can be defined by using the PLS prediction results. In the context of the sample holder, when the perceived concentration of PLA in the spectral measurement remains constant, the limit of detection is reached. For determining this limit, the PLS prediction results were used to fit two linear regression models. The intersection of the two regression lines is considered to be the maximum acceptable thickness for which the PLA concentration evolution can be interpreted by a multivariate unmixing method.

The detection depth obtained by this method is $y_d = 1.80$ mm (Figure 18). As a consequence, the maximum wheat flour thickness to use in order to detect the background in PLA is 1.80 mm. This result is specific to the case of PLA under a thickness of wheat flour. However, the method of determination may be used for any kind of sample and background or target.

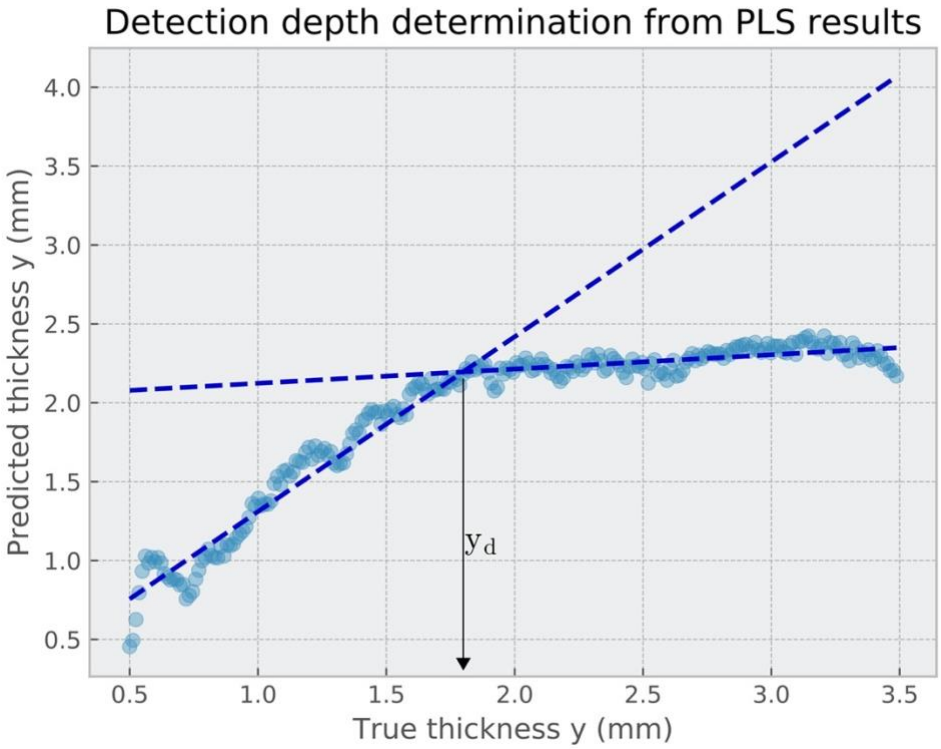

Figure 18: Determination method for the detection depth from the PLS prediction results.

Additional experiments (results not shown) were performed using different particle sizes for wheat flour, other samples such as chocolate powder or almond powder, and different designs for the sample holder. The hyperspectral imaging measurement followed by PLS analysis showed to be consistent for each application. Globally, the same behavior for the penetration depth were observed with some variations according to the samples. The particle size and the density of the powder seem to be important parameters that influence penetration depth. For future work, the influence of these parameters may be investigated.

4. Additional discussions

A. The detection depth versus the penetration depth

This study introduces the notion of detection depth which is different from the penetration depth. The penetration depth is defined as the depth (into a given material) at which the intensity of an incident beam decreases by 99 % [35]. However, this

definition is difficult to use for a range of wavelengths because of the absorption variations. Although it is a convenient notion to describe the signal theoretically, it may not be suitable for the study of complete spectra. Indeed, the analysis of spectral data is commonly made using multivariate methods that take all the wavelengths into account. Hence, the notion of penetration depth is not adapted for the characterization of an entire spectrum in the context of detection.

Moreover, the detection of PLA under a layer of wheat flour is not possible for each wavelength. Examples are the wavelengths where PLA does not exhibit an intense absorption peak (Figure 15). In this situation, the obtained reflectance profile does not show any evolution depending on the powder thickness. As a result, the curve fitting technique cannot be adequately applied and the corresponding penetration depth value is ill-defined. This is the reason why the penetration depth cannot be evaluated for high wavelengths in the study of wheat flour and PLA. For wavelengths higher than 1350 nm, the absorption of wheat flour is too strong so that the detection of PLA is not possible. Berntsson et al. had the same limitation at 1400 nm using the empirical method [30]. As the NIR spectra often have an increasing absorption baseline, this would suggest the higher the wavelength, the more difficult is to detect its signal in depth. Berntsson et al. have shown, using the theoretical method, that the penetration depth for an absorbent material is below 0.5 mm for wavelengths higher than 1500 nm.

As a consequence, the notion of detection depth may be more suitable regarding detection problems in NIR spectroscopy. This consists of characterizing the entire spectrum instead of each wavelength.

B. The effective detection depth

Our study shows that the detection depth of PLA in the wheat flour is about 1.80 mm on the NIR range [53]. This result can be compared to those obtained by Huang et al. [35] who worked on the detection of melamine through milk powder. Indeed, the measurement contexts are similar. Moreover, the samples play similar roles: the wheat flour and the milk are both powders with smooth spectra (i.e. no sharp absorption peak). On the other hand, the melamine and the PLA both exhibit a sharp absorption peak in their spectra. In the study of Huang et al., only five thickness values could be tested and the authors concluded that the detection depth is below 2 mm. The two works are relevant to show that even if there is a strong absorption of a material, it can be strongly attenuated by a small thickness of powder in the order of some millimeters (Figure 19). As a reference, 10 to 20 wheat flour particles can strongly attenuate the NIR radiations.

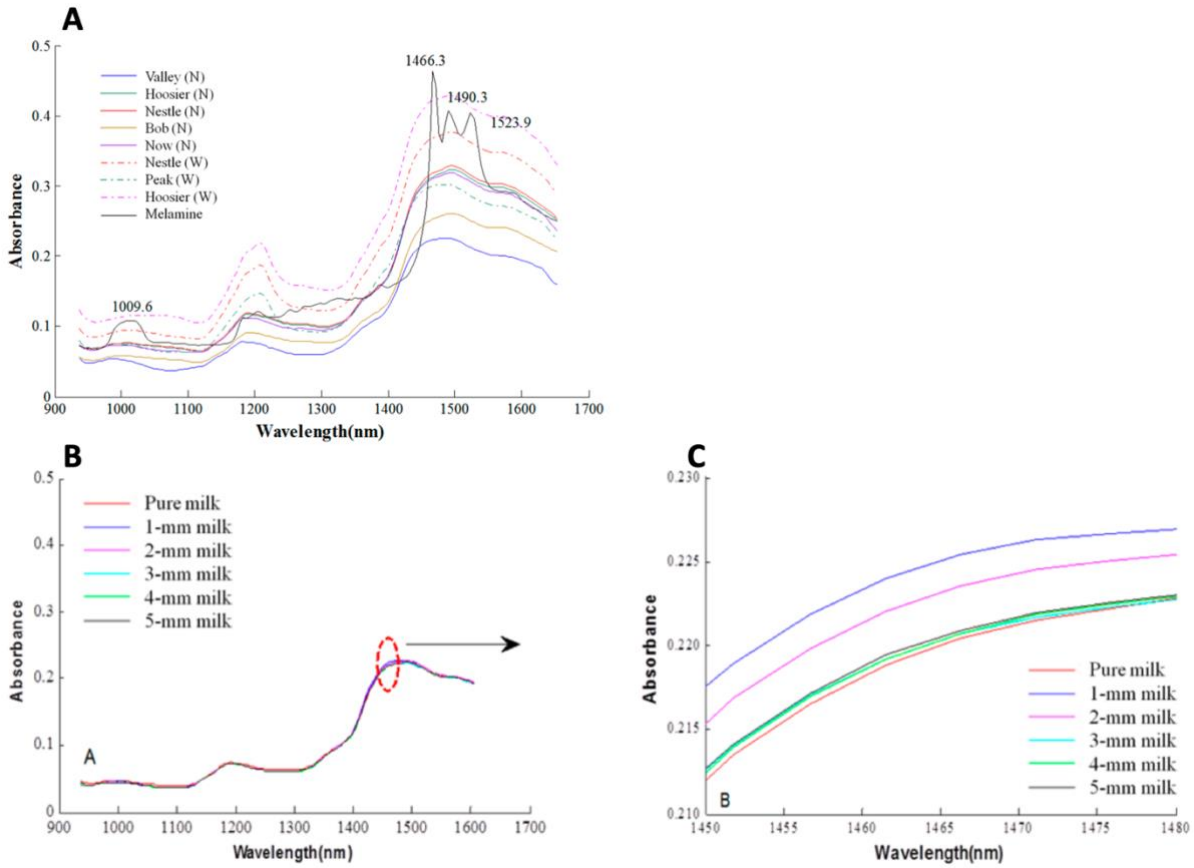

Figure 19: (A) the absorbance profiles of milk powders and melamine samples separately and (B-C) measured together with low amount of melamine [35].

C. The consequences of the detection depth

The previous results show that the detection depth is very small compared to the typical sample size: only 10 to 20 particles of wheat flour. As a result, the signal measured using NIR spectroscopy is mostly representative of the sample surface. It is essential to check that the corresponding analysis depth is relevant to the sample property to be evaluated. For example, Lammertyn et al. study the penetration depth in apple slices to ensure their internal properties can be inferred from standard NIR measurements [34]. In a perfect world, such a study should be performed systematically to be sure to know what part of the sample is screened.

The detection depth may have more hazardous implications when the application is about detecting adulterations. Because of the NIR limitation, the detection can only be performed for a certain depth which may be even more reduced than the results shown before. Indeed, the chemical to be detected may have no significant absorbance peak to help for the detection. As a result, the conclusion of a NIR inspection should be restricted to the sample thickness defined by the detection depth. Such a limitation may be a problem for the detection of some default in food samples. However, this limitation could be overcome when dealing with powders. One solution could be to spread out the powder sample on a large surface to reduce the

thickness. Another solution could be to analyze the same powder sample several times with a HSI. The powder is shuffled each time before measuring. In this way, several random layers of powder are presented to the sensor. If there is a particle to be detected, both techniques give the sensor a better chance of discovering and measuring the minor compound.

D. The parameters influencing the detection depth

The detection depth depends on several parameters. They can be separated into two categories: the parameters linked to the physical properties of the sample and those linked to the properties of the measurement system. The first category has been discussed in the previous sections through the scattering and absorption coefficients of the medium. These coefficients are dependent on the wavelength and on various characteristics of the sample. The scattering coefficient is mainly influenced by the particle size (powders) while the absorption coefficient is influenced by the chemical nature of the molecules in the sample.

The properties of the measurement system are also highly responsible for the detection depth. In particular, the dynamic of the sensor has a great importance in the ability to detect NIR signal in reflectance. Let us consider a NIR incident beam that irradiates a powder sample with a bottom material (Figure 20). The probability that the photon is scattered or absorbed increases with the path length. Hence, there are a few chances a light path reaches the bottom material so that its absorption signature can be taken into account by the measurement. Let us present this differently by focusing on a specific wavelength at which the bottom material has a sharp absorption peak (for example 1168 nm for PLA). Even if the powder absorption for this wavelength is not high, the deeper the light beam goes into the powder the higher is the probability it is absorbed. As a result, the absorbance peak of the bottom material does not affect the measurement.

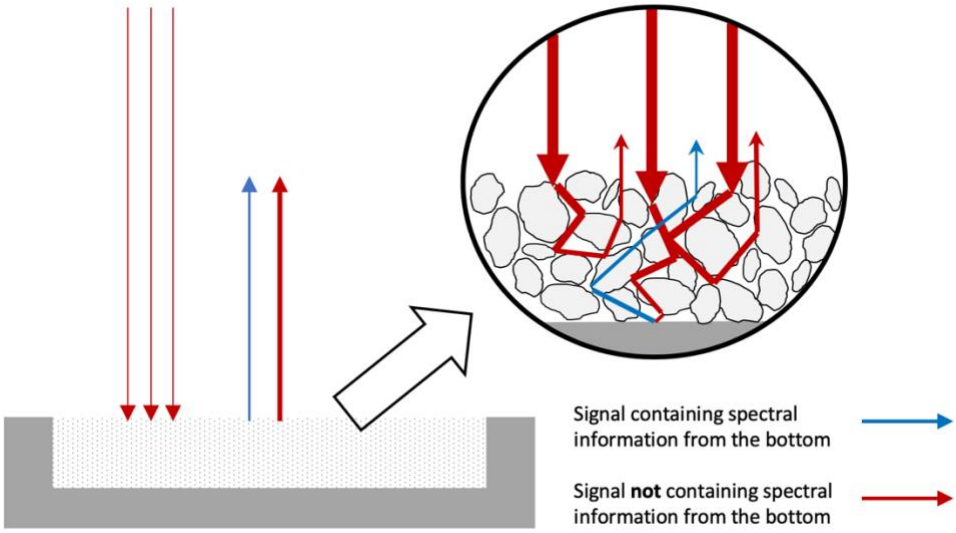

Figure 20: The light beam behavior when measuring in-depth material through a scattering medium.

However, the reflectance measurement is the result of a massive number of path lengths. Among them, there probably exist some light beam that successfully reach the bottom material. As these path lengths are in minority, their contribution is very weak and invisible in the measured signal. It is because this intensity is comparable to the measurement noise. Hence, the signal of interest cannot be differentiated from the variance of the measurement. The standard way to reduce the importance of the noise is to increase the integration time of the sensor. By this process, the signal to noise ratio increases. However, increasing the integration time of the sensor also increases the amount of counted photons in the sensor. This count is limited by the technology and causes the sensor saturation. As a result, the integration time cannot be increased indefinitely and the saturation depends on the amount of received signal. Unfortunately, the reflectance measurement creates a signal of a strong intensity signal from the surface of the sample. In particular, the specular reflectance has a strong absorbance that is responsible for the sensor saturation.

The dynamic of a sensor is the ratio between the smallest and the largest signal quantity it can measure. For a spectrometer, it is possible to tune the integration time to decide how long the sensor counts the incoming photons. If this time is too short, photons from interesting signal maybe confounded with the measurement noise. If this time if too long, the sensor maybe saturated and no quantitative information can be recovered. For this reason, if a sensor has a high dynamic range, it is possible to increase the integration time – meaning measuring more signal – without saturating. Hence it can measure signals coming from deeper layers in the sample.

5. Conclusion and perspectives

This work has proposed a method using hyperspectral imaging, a PLA sample holder and the PLS regression method to study the light penetration depth in a wheat flour sample. Reflectance profiles were extracted and interpreted using the Kubelka-Munk theory. Using this model derivation as well as sensor considerations, a criterion for light penetration depth was given and calculated for the range 1100 – 1350 nm. Results have shown that it is highly dependent on the wavelength value. The PLS method has been shown to be an efficient solution for fitting spectral data on the linear mixing model up to a given thickness. The use of this multivariate technique has provided a criterion for defining the detection depth, the maximum thickness of wheat flour for which PLA can be quantified from the signal. Two linear models were fitted on the PLS prediction results in order to calculate a detection depth of 1.80 mm. This value provides an estimation of the maximum depth for which a spectral target can be detected into wheat flour. It also corresponds to the minimum thickness of wheat flour to ensure the signal of the background does not have influence in the diffuse reflectance measurement. Unlike the concept of penetration depth, the detection depth is related to the application and gives a more suitable value for in-depth detection purposes. The procedure used in this work could be reproduced using another material as a target. By applying a thin layer of target particles at the bottom of the sample holder, the same PLS regression procedure could be applied to obtain the detection depth results for a given application.

At this moment, the detection of PLA in wheat flour does not have a direct application in the industry, however the detection of contaminants is carried out. A solution could be to spread out the target chemical on the bottom of the sample holder. By doing this, the target material becomes the chemical of interest. Huang et al. proposed a similar approach for studying the detection of melamine in milk powder [35].

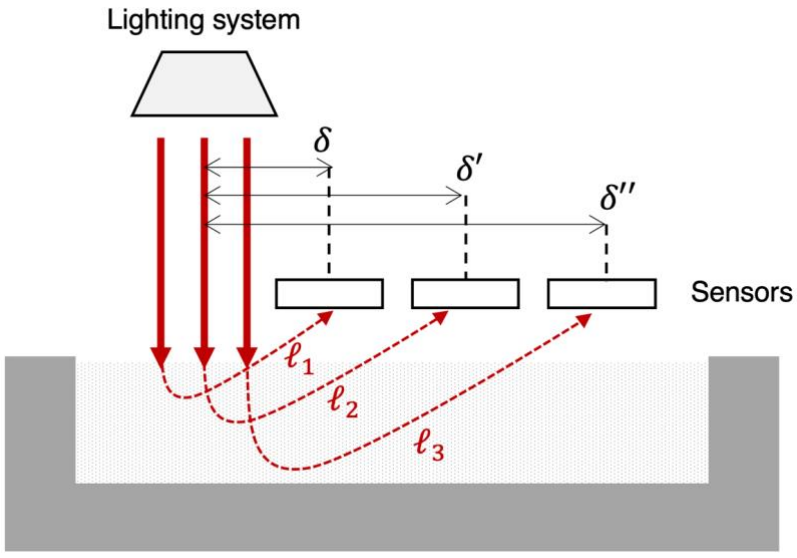

Figure 21: Principle of the SRS measurement.

Finally, the way the reflectance measurement is performed could be changed to increase the detection depth. As presented before, the detection depth is limited by the fact that the surface contribution dominates the intensity of the reflectance measurement. Consequently, there is an interest in being able to measure the reflectance signal from the surface and from the sample depth independently. The Spatially Resolved Spectroscopy (SRS) is a measurement method that is used to study the scattering properties of a sample. When using this technique, one assumes that if a light path goes deep in the sample, there is more chance it goes out from the sample at a longer distance from the light source. Hence, by illuminating a restricted area of the sample the reflected light with a higher pathlength can be isolated and measured by a sensor. Figure 21 shows this principle: three different light paths (ℓ_1, ℓ_2 and ℓ_3) are shown. As the light goes deeper in the powder, it is scattered back at a longer distance from its entry point. Multiple sensors can be placed at different distances (δ, δ', δ'') from the lighting to get the information from a specific light path. This sensor must target a specific area of the sample, for example a given distance from the lighting system. Figure 21 shows that since the surface reflectance does not irradiate the sensor, the detection of the bottom material could be done for higher depths.

II. The detection of peanut flour using the Matched Subspace Detector

This part has been adapted from the publication:

A. Laborde, B. Jaillais, J.M. Roger, M. Metz, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, C. Cordella, Subpixel detection of peanut in wheat flour using a matched subspace detector algorithm and near-infrared hyperspectral imaging, Talanta. 216 (2020) 120993. https://doi.org/10.1016/j.talanta.2020.120993.

1. Introduction

Peanut is a primary product used in the food industry for its fats and proteins [68]. However, it is also a major food allergen. As such, it is hazardous for allergic people: the ingestion of some mg of protein [69] may have a dramatic effect up to death. For this reason, the food industry enforces good manufacturing practices to reduce the risk of contamination and to inform the probable presence of food allergen in their products [70].

Hyperspectral Imaging is a technique combining spectroscopy and imaging to obtain both spatial and spectral information from a sample. When associated with the NIR spectral range, HSI is a powerful technique to provide a fast, non-destructive, and cost-effective control method. It is an emerging technology for food inspection since it provides non-destructive analysis of heterogeneous samples [4], and hyperspectral images allow the visualization of chemical maps of the samples. As an example, Elmasry et al. use this technology to provide water, fat and protein distributions on beef samples [71]. The estimation of food nutrients in samples leads to their quality assessment [1].

Hyperspectral imaging technique enables to obtain a spectral measurement for every pixel of an image. As a consequence, each spectrum is representative of a small surface of the sample. The spectral measurement is then more sensitive to the minor components of this sample since they have more influence on the field of view of one pixel than on the entire sample. It offers the opportunity to detect adulterants in food by characterizing each pixel of the image [72][73][74]. In practice, the detection consists of classifying each pixel of the hyperspectral image as an adulterant or a sample spectral signature. As NIR spectroscopy has been a powerful technique for characterizing organic matter [75], [76], HSI appears to be a promising tool for adulteration detection in the food industry. The literature shows plenty of such applications: Vermeulen et al. studied the detection of ergot bodies in cereal flour [77], Fernández Pierna et al. investigated the detection of melamine in milk powder [78] [56], Verdú et al. proposed to study the adulteration of wheat products [79]. Mishra et al. studied the detection of crushed peanut in wheat flour [15] [14] using two different

detection methods. However, performing this detection may be more difficult in another context. Defatted peanut flour is a product obtained from milled peanut. In this condition, peanut flour particles are smaller than the crushed peanuts: 500 µm to 1000 µm [15] for crushed peanut against smaller than 200 µm for peanut flour (similar to wheat flour [37]). Additionally, as the fatty acids of peanuts are removed, their spectral contribution cannot be used for the discrimination of peanut particles against wheat particles. For these reasons, the use of HSI for detection may be a challenge. The purpose is to identify different materials based on their spectral signature to label each pixel as an adulterant or a sample pixel. However, there is not a unique spectral signature for each material [51] because the reflectance value at each wavelength of a given material is not deterministic but is a random variable. Its variability is linked to the lighting conditions, the material surface, the sample heterogeneity, and many other factors. On the other hand, two different materials may have very similar spectra. In particular food industry products, the shapes of NIR spectra are often similar since they are the result of the mix of the main nutrients. It is the case for wheat and peanut as they are similar products once transformed into flour. Thus, the ambiguity of spectral information and the spectral variability issue are two main challenges for detection purposes in the food industry.

Another difficulty arises when dealing with powder samples because the particle size may be smaller than the pixel size. For a hyperspectral camera, a pixel integrates the radiance signal from all the material in its field of view. If it contains particles with different spectral signatures, the pixel is considered as mixed and the resulting spectrum does not correspond to any pure chemical defined as target or background. This problem is known as the subpixel detection [51].

In the literature, some detection algorithms have been developed to take into consideration both variability and subpixel issues using spectral modeling [51]. On the one hand, the mixed pixel issue is tackled using the LMM [64], which assumes the radiance measured by a pixel is the sum of the chemical radiances weighted by their surface contribution in the pixel field of view. On the other hand, the detection problem can be addressed using a MSD [51] [80], which is derived from a hypothesis test.

The design of the MSD algorithm requires to fine-tune its parameters as well as to evaluate its detection performance. It implies to have annotated hyperspectral images in which adulterated and regular pixels are identified. However, such data cannot be obtained easily in the context of food inspection. When two powders of similar aspects are mixed, it is not possible to identify adulterated pixel by human eye inspection. On the other hand, it is not possible to control the spatial constitution of individual pixels since they represent a very small field of view (0.2 mm × 0.2 mm) [2].

Instead, spectral simulations can be used to generate new spectral data from a known statistical distribution. With this technique, synthetic spectral data with known adulteration concentrations can be used to validate the design of the MSD detection algorithm.

The use of a near-infrared hyperspectral imaging for the detection of adulterant in food has been studied in plenty of applications [4]. However, the issues regarding the variability of the samples, the spectral ambiguity between species, the mixed pixels and the lack of reference data for the detector design make the detection difficult. To our knowledge, no study has been proposed to tackle detection for such samples with particle size smaller than the pixel. The purpose of this study is to evaluate how the MSD approach using the LMM and the modeling of spectral variability can provide performant detection for such a detection problem. As no reference values for the detector design are available, a spectral simulation method is proposed. After studying the performances of the detector on simulated data, we propose to study the detection using real hyperspectral measurements of flour mixtures at graduated concentrations of peanut.

2. Material and methods

A. Samples

In this study, the aim is to detect adulteration of peanut flour in wheat flour. White wheat flour (Grands Moulins de Paris, Francine batch number **ER510 – FTU104**, France) was used as the regular sample. Samples were taken from two different packs. Defatted peanut flour (KoRo Handels GmbH, batch number **C170151**, Germany) was used for the target sample. Flour samples were mixed together to obtain 8 different mass concentrations of peanut flour: 20 %, 10 %, 5 %, 2 %, 1 %, 0.5 %, 0.2 % and 0.02 % for a total mass of 13.75 g. Pure peanut flour and pure wheat flour were prepared as well. Mass measurements were performed using a precision balance (**Sartorius Entris**, 0.01 mg precision).

Mixed samples were put in a container to be shaken and mixed with a spatula. Samples were put in a rectangular sample holder (30 mm width × 70 mm length) made of a 7 mm depth cavity. The top of the sample holder was skimmed to remove excess powder without affecting the packing density. In the following, the wheat flour is referred to as the regular sample with letter S, and the peanut flour is referred as the adulterant with letter A.

B. Hyperspectral imaging system

A line-scan pushbroom Specim SWIR camera (SPECIM, Oulu, Finland) was used for the image acquisition. The hyperspectral camera acquired 288 spectral bands from 900 nm to 2500 nm with a 5.6 nm step. The camera acquired 392 pixels per line and the pixel size was 250 µm × 250 µm. Six halogen lamps were used for the measurement and heated up for 30 min before the acquisition. The integration time was 0.973 ms. A white reference measurement was performed before each acquisition using a white diffuse

reflectance standard (Spectralon®, SRS-99-010, Labsphere). Additionally, the dark reference image was acquired after closing the shutter of the camera. Each sample was measured independently, leading to 30 data cubes.

C. Data processing

Each image was cropped to focus on the sample in the central cavity of the sample holder leading to data cubes of size 200 × 320 × 188. The white reference image was averaged along the perpendicular direction of the sensor array to obtain one spectrum for every pixel of the sensor line. The reflectance image is calculated using Equation 9. First (under 1200 nm) and last (over 2200 nm) wavelengths were removed as spectra were too noisy. Spectra were processed using a Savitsky-Golay filter to reduce the noise for the remaining wavelengths (2nd order polynomial, 7-points window, and no derivative). A Standard Normal Variate (SNV) transformation was applied to compensate for scattering effects.

D. Spectral simulation using Principal Component Analysis

A spectrum can be considered as a vector of m absorbance values, one for each wavelength of the spectral range. The variability of spectral data is the variance of the reflectance for each wavelength. NIR spectral data exhibits a high correlation between the variables. As a consequence, defining the spectral variability independently for each variable is not efficient. Instead, new variables can be calculated using PCA.

PCA is a method for dimensionality reduction that decomposes the data matrix $\mathbf{X} \in \mathbb{R}^{n \times m}$ according to orthogonal sources of the highest possible variance. The data matrix can be decomposed as follow [81]:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}}$$

Equation 12: Matrix decomposition by PCA.

where $\mathbf{T} \in \mathbb{R}^{n \times m}$ is the score matrix, $\mathbf{P} \in \mathbb{R}^{m \times m}$ is the loading matrix and the upper script symbol T refers to the transposed matrix. The dimensions of the score and the loading matrices hold if $m \leq n$.

Under this new representation, each spectrum is represented by a set of scores (from the $\mathbf{T}$ matrix) in the Principal Component space and their corresponding loading vectors (from the $\mathbf{P}$ matrix) that describe the spectral variability of the $\mathbf{X}$ image. In this context, each loading or component vector from the $\mathbf{P}$ matrix describes the subspace of the hyperspectral image's variability. On the other hand, the scores describe the coordinates of the pixels on the corresponding subspace.

For each component, the distribution of these scores can be considered as Gaussian with mean μ - which is, for PCA applied on centered data, null - and variance

σ^2. As a consequence, a new spectrum can be simulated by randomly generating its scores coordinates on the principal components.

The spectral simulation procedure consists of combining the PCA and the LMM (Equation 2). In this context, the vector $\boldsymbol{s_i}$ can represent a pure spectral signature, like the average spectrum of a product. Or it may represent a spectral signature describing one variability subspace, like a component from the PCA. The process of data simulation used in this study is described by the following procedure:

1. A PCA was performed on the centered data matrix of the regular sample $\mathbf{X_S}$ and of the adulterant sample $\mathbf{X_A}$ distinctly. The average spectra of both matrices $\bar{\mathbf{X}}_S$ and $\bar{\mathbf{X}}_\mathbf{A}$ were calculated and considered as the pure spectral signatures of the materials.

2. For every principal component index $i \in [1, \mathrm{m}]$, the distribution of scores $\boldsymbol{T^i}$ was assumed to be Gaussian with mean μ^i (which is equal to 0 in case of a centered PCA) and standard deviation σ^i. These parameters were estimated from the PCA performed in the first step of the procedure.

3. For a given peanut proportion c varying between 0 and 1, the average spectrum was simulated using the LMM:

$$\tilde{\bar{X}}_0 = c\bar{X}_A + (1-c)\bar{X}_S$$

Equation 13: Simulation of the average spectrum with the LMM.

where the tilde symbol designates the simulated matrices.

4. The simulated scores for each PC were randomly generated from Gaussian distributions with the parameters estimated in step 2. Then, the simulated variability was calculated by multiplying the random scores $\tilde{\mathbf{T}}_\mathbf{A}$ and $\tilde{\mathbf{T}}_\mathbf{S}$ by the principal components of the corresponding PCA $\mathbf{P_A}$ and $\mathbf{P_S}$. The total variability attributed to $\tilde{\mathbf{X}}$ is a balance between peanut and wheat controlled by the proportion c. This simulated variability was finally added to the averaged spectrum calculated in step 3:

$$\tilde{\mathbf{X}} = \tilde{\bar{\mathbf{X}}}_\mathbf{0} + c\tilde{\mathbf{T}}_\mathbf{A}\mathbf{P}_\mathbf{A}^\mathbf{T} + (1-c)\tilde{\mathbf{T}}_\mathbf{S}\mathbf{P}_\mathbf{S}^\mathbf{T}$$

Equation 14: Simulation of the spectral data with PCA and the LMM.

The simulation procedure was applied to obtain 100 spectra for each peanut concentration 5%, 10%, 15% and 20%.

E. Detection using the Matched Subspace Detector

Two competing hypotheses were tested to address the subpixel detection problem. In the null hypothesis, the pixel was assumed to contain only the regular sample (wheat flour). The LMM decomposed the pixel spectrum according to the variability subspace associated with the regular sample $\boldsymbol{s_i^S}$. In the alternative hypothesis, the pixel was assumed to contain adulterant particles as well as the regular particles. Thus, $\boldsymbol{x}$ was modeled using the LMM with the variability subspaces associated with the regular sample $\boldsymbol{s_i^S}$ and the adulterant $\boldsymbol{s_i^A}$. The detection algorithm was based on the following statistical test:

$$\mathrm{H}_0: x = \sum_{i=1}^{L} a_i s_i^S$$

$$\mathrm{H}_1: x = \sum_{i=1}^{L} a_i^S s_i^S + \sum_{i=1}^{J} a_i^A s_i^A$$

Equation 15: The two hypothesis for the subpixel detection.

where L and J defined the dimension of the variability subspace for the regular sample and the adulterant sample, respectively. Two matrices were defined corresponding to these hypotheses: $\mathbf{M_S}$ contains the vectors $\boldsymbol{s_i^S}$ in columns, and $\mathbf{M_A}$ contains the vectors $\boldsymbol{s_i^S}$ and $\boldsymbol{s_i^A}$ in columns: $\mathbf{M_S} = \left(\boldsymbol{s_1^S}, \boldsymbol{s_2^S}, \dots, \boldsymbol{s_L^S}\right)$ and $\mathbf{M_A} = \left(\boldsymbol{s_1^S}, \boldsymbol{s_2^S}, \dots, \boldsymbol{s_L^S}, \ \boldsymbol{s_1^A}, \boldsymbol{s_2^A}, \dots, \boldsymbol{s_J^A}\right)$.

The generalized likelihood ratio approach gives the detection statistic for the MSD algorithm [51]:

$$\mathrm{T}_{\mathrm{MSD}}(x) = \frac{\boldsymbol{x}^T(\mathbf{Q_S^{\perp}} - \mathbf{Q_A^{\perp}})\boldsymbol{x}}{\boldsymbol{x}^T\mathbf{Q_A^{\perp}}\boldsymbol{x}}$$

Equation 16: The statistic of the MSD.

where $\mathbf{Q_S^{\perp}}$ and $\mathbf{Q_A^{\perp}}$ are the projection matrices on the orthogonal subspace of $\mathbf{M_S}$ and $\mathbf{M_A}$ respectively. These projectors were obtained using the following formula:

$$\mathbf{Q_X^{\perp}} = \mathbf{I} - \mathbf{X}(\mathbf{X^T X})^{-1}\mathbf{X^T}$$

Equation 17: The definition of a projection matrix on an orthogonal subspace.

In the first step, the detection statistic was calculated for each pixel spectrum of a sample using Equation 16. In a second step, a threshold must be chosen to classify each pixel between two classes: regular pixel or adulterant pixel. This threshold was chosen by applying the following procedure. First, the detection statistic was applied to all the pixels of the regular samples. When the value of $\mathrm{T}_{\mathrm{MSD}}$ increases, the probability that the corresponding spectrum contains adulterant particles increases as well. Hence, the maximum value of $\mathrm{T}_{\mathrm{MSD}}$ calculated on the regular sample was chosen as the threshold. This approach consisted of maximizing the detection rate by keeping

the false alarm rate under a limit. For new pixels, the detection procedure is defined by:

$$\mathrm{T}_{\mathrm{MSD}}(x) \gtrless \eta_{NP}$$
$$\text{with } \eta_{NP} = max\left(\mathrm{T}_{\mathrm{MSD}}(\mathbf{X_S})\right)$$

Equation 18: The test decision inequality.

where the two other pure wheat replicates were used in order to assess the robustness of the thresholding method.

As Equation 16 shows, the design of the MSD algorithm depends on the design of $\mathbf{M_S}$ and $\mathbf{M_A}$ which directly depend on the parameters L and J. As a consequence, the design of the MSD algorithm consists of finding optimal values for L and J. In the following, the performance of the detector is qualified using its sensitivity which refers to the minimum local concentration (at the pixel scale) for which the detection rate is over 99 %.

F. Software

The data processing was performed using Python 3.7. For data simulation, the PCA was performed using the Scikit-Learn 0.18.1 implementation consisting of a Singular Value Decomposition (SVD). The PCA on non-centered data was performed using the eigenvalue decomposition of $\mathbf{X^T X}$ on Numpy 1.16.4.

3. Results and discussions

A. Evaluation of data simulation for the detector design

Figure 22 shows the factorial PC1-PC2 plan of PCA performed on the pure wheat and the pure peanut samples. Their corresponding scores are plotted with empty squares. On the other hand, the simulated data between 5 % and 20 % of peanut adulteration are plotted with filled markers. These scores are obtained by projecting the simulated spectra onto the first two loading vectors of the PCA. The figure shows that the lowest peanut concentrations are closer to the pure wheat sample on the left of the principal component. Conversely, the highest peanut concentrations are on the right side of the plot. It shows that the simulated data are ordered according to the first principal component, which describes the main variability between peanut and wheat. Hence the data simulation procedure is relevant with the expected concentration level of peanut.

The figure contains a focus which enables to evaluate the distribution of the scores for the simulated data and the pure wheat flour data. This observation shows

that the simulated data have similar variability as real measurements and can be used to assess the sensitivity of the detection algorithm.

Additionally, Figure 22 shows that the pure peanut flour measurements exhibit a higher variance on the second principal component that the wheat flour and the simulated data. It is explained by the fact that the surface of the peanut sample exhibits more heterogeneity as well as the peanut particles are more diverse than wheat. On the other hand, the variability of the simulated data was generated using a combination of information coming from the pure wheat and peanut measurements. This combination is controlled by the concentration parameter c, as shown by Equation 14. As a result, for small concentrations, the simulated data have a variability which is closer to that of the wheat flour sample.

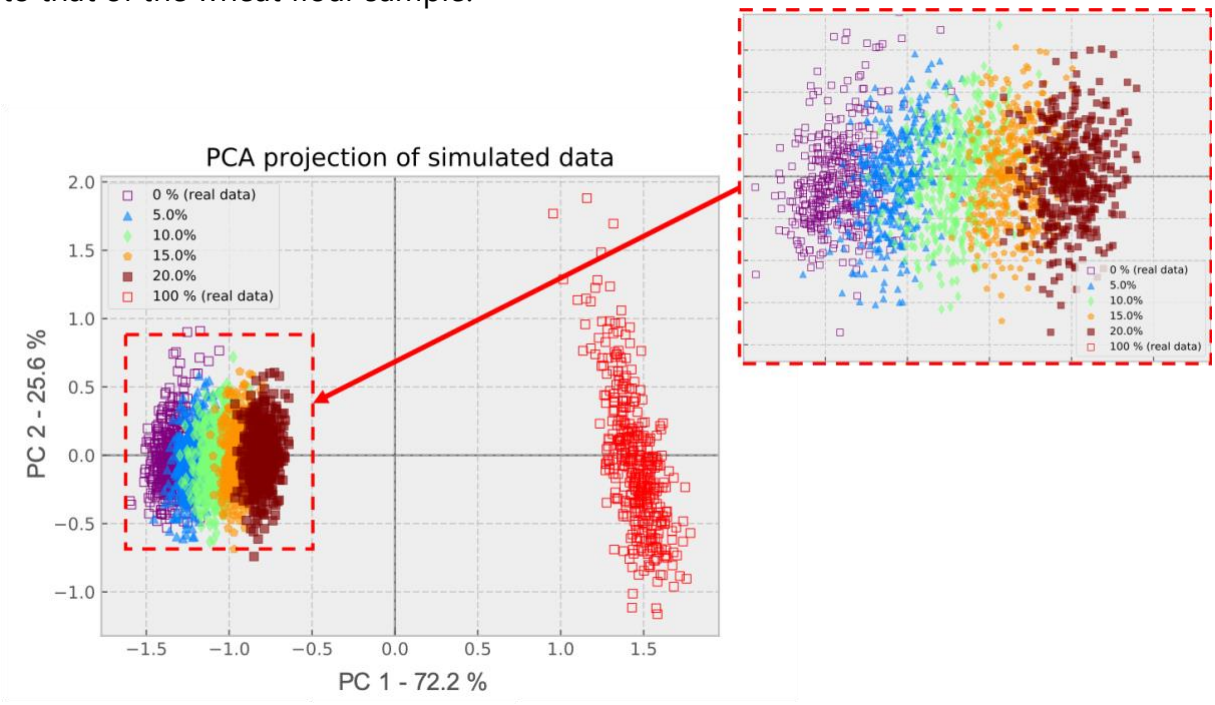

Figure 22: Simulated data are projected on the score plot of the PCA performed on real measurements of pure samples.

Table 1 shows the details of the design (the values for L and J) for three MSD algorithms. These designs were selected because they show the most interesting results among all those which were tested. The next section shows the results for other designs and focuses on the choice of these parameters. In the current section, we focus on the relevancy of the use of simulated data to evaluate the MSD algorithms performances.

Table 1: Design parameters L and J for the three MSD designs of interest.

Design parameters	L	J
MSD 1	1	1
MSD 2	2	1
MSD 3	2	2

The design of the MSD algorithms requires to evaluate its sensitivity to optimize the choice of the parameter values (L and J). Figure 23 shows the detection rate of the three MSD designs described in Table 1. The detection rate indicates the fraction of detected targets for a given peanut concentration of the simulated data. For zero peanut concentration, spectra from the real wheat flour images were used. The graph shows that no detectors have any false alarm on real wheat measurements. It means that the thresholding method is robust for all three MSD designs. MSD 2 and 3 reach a detection rate of 100% for a simulated peanut concentration of 20% and they both have similar detection rates for smaller peanut concentrations. MSD 1 exhibits a lower detection rate for every concentration and does not reach a 100% detection rate for 20% of peanut adulteration. According to the simulated data, MSD 2 and 3 have a similar sensitivity, which is higher that of MSD 1.

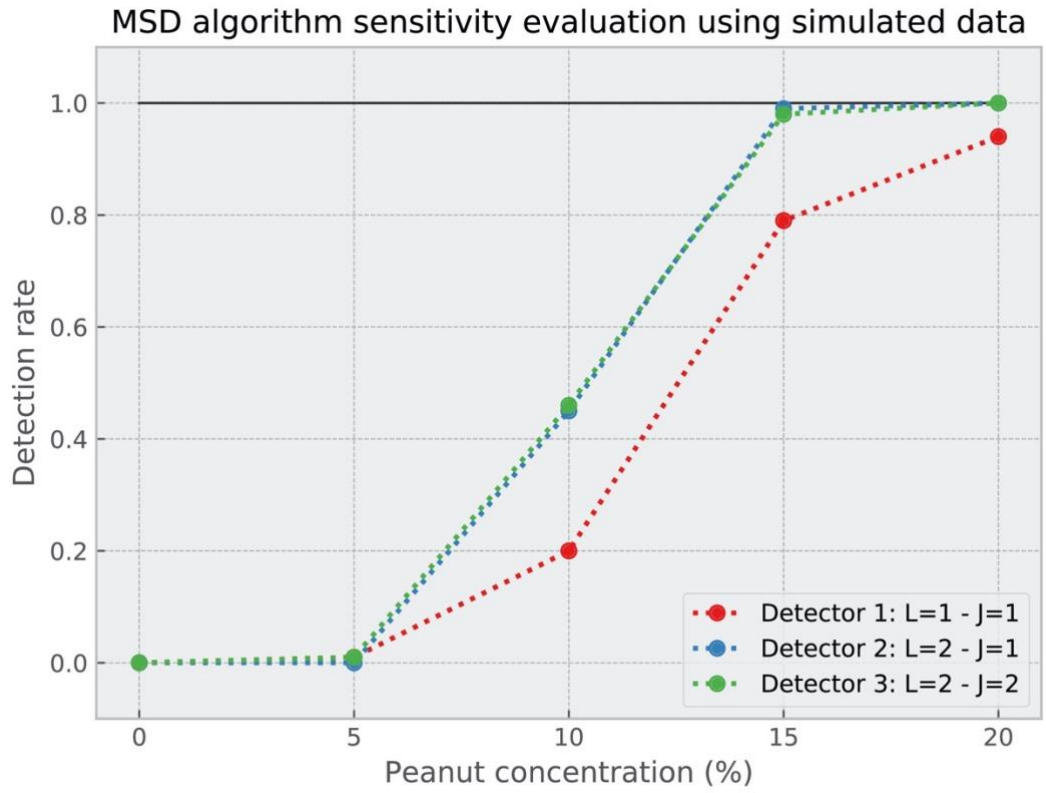

Figure 23: The detection rate according to the peanut concentration in simulated data for three MSD designs.

Figure 24 shows the comparison of the detection maps for the three MSD presented in Figure 23 and Table 1. For this purpose, a focus is made on the hyperspectral image measured on a sample containing 2% of peanut flour. The map represents an area of 104 × 62 pixels (2.6 × 1.5 cm) and each color corresponds to the output of the comparison of 2 MSD algorithms.

The top map shows several groups of blue pixels (top left-hand corner), meaning that 30 pixels are only detected by MSD 2 and not by MSD 1. Since real measurements do not contain any reference value at the pixel scale, the real position of the targets is

unknown. However, the fact that neighbor pixels are simultaneously detected strengthens the probability that there is effectively peanut in these pixels. In other words, the detection of a neighborhood of pixels is more credible than the detection of an isolated pixel. On the other hand, MSD 1 only detects one pixel exclusively. The comparison of the detection maps shows that MSD 1 is less sensitive than MSD 2.

The map below shows less colored pixels, which means that MSD 2 and 3 have similar performances. MSD 3 detects 13 more pixels than MSD 2 which are located close to clusters of pixels that are detected by both MSD designs. On the other hand, MSD 2 only detects one more pixel than MSD, which can be considered suspicious since it is isolated. These observations show that MSD 3 has better performances that MSD 2.

These conclusions drawn using real measurements are relevant to the results obtained with the simulation method (Figure 23). They show that the simulated data can be relevantly used for the analysis of the MSD designs.

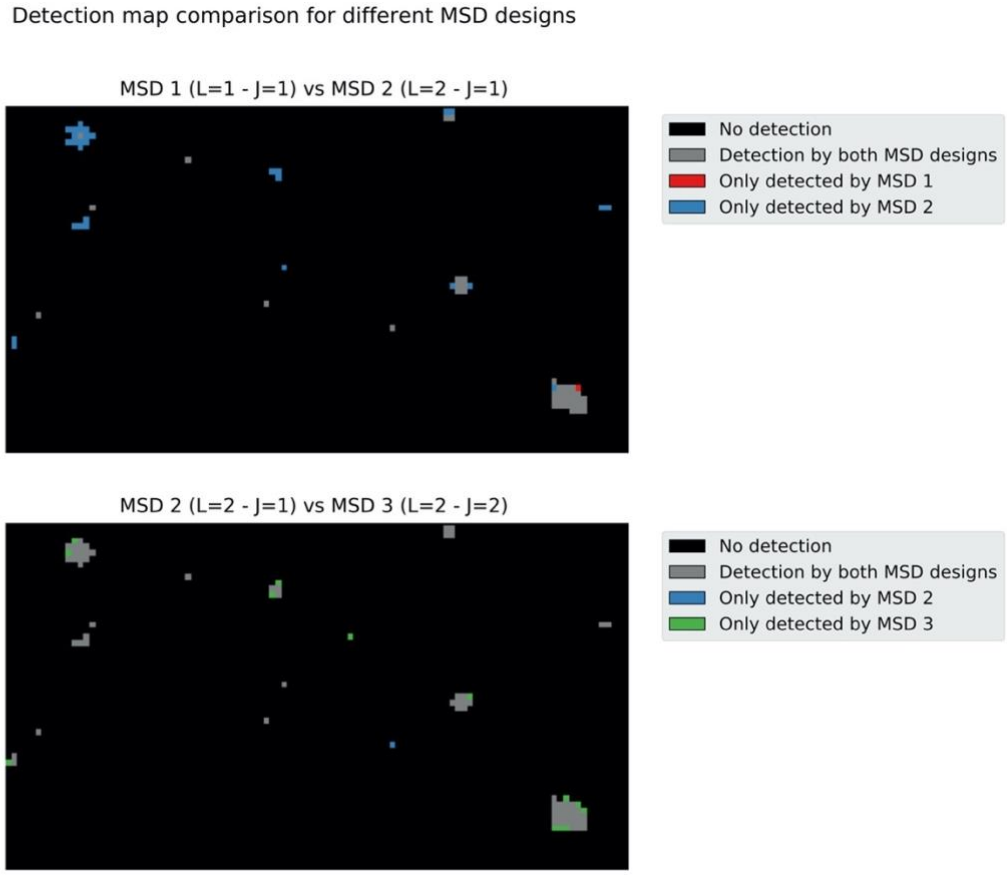

Figure 24: Focus on the detection map comparison for the sample with 2% of peanut (replicate A).

B. Evaluation of the Matched Subspace Detector Algorithm

Design of the Matched Subspace Detector

Figure 25 shows the sensitivity of several MSD designs calculated with the simulated data. The first graph on the left shows the effect of varying L with $J = 2$. It means the dimensions of the subspace which represents the regular sample is varying whereas a subspace of dimension two is chosen to represent to adulterant sample. $L = 1$ gives low performances since no spectra are detected even for a peanut concentration of 20%. The best performances are obtained for $L = 2$. Then, increasing L leads to lower detection rates. The graph on the right shows the evolution of the sensitivity when fixing $L = 2$. In this condition, the performances of the MSD algorithms are identical for $J = 1$ and $J = 2$. Then, choosing a higher value for J decreases the detection rate. Figure 23 also shows that the design with $J = 1$ and $L = 1$ provides a lower sensitivity than MSD 3. Consequently, the results show that there is an optimal design for the MSD regarding the performances on the simulated data: $L = 2, J = 2$.

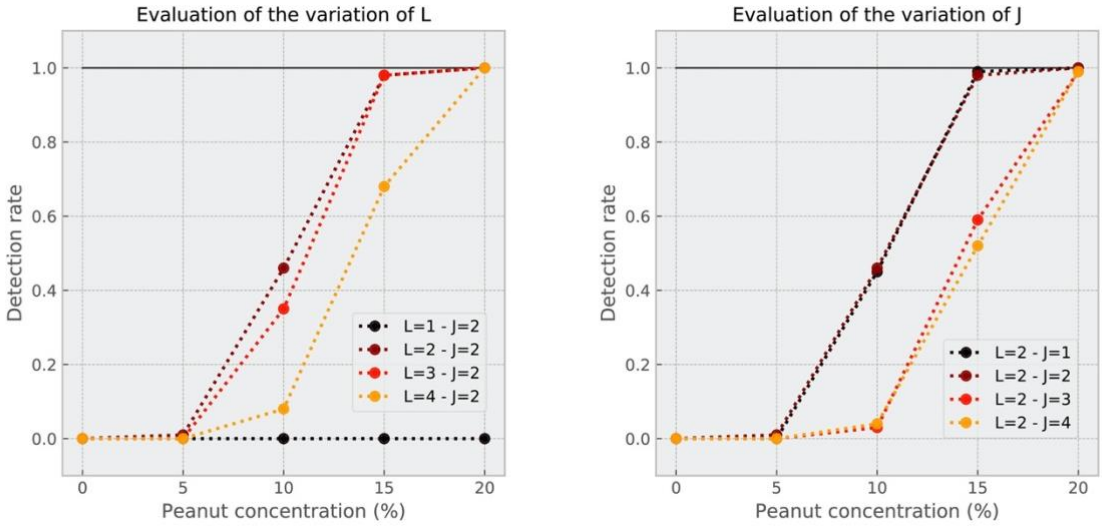

Figure 25: MSD algorithms sensitivity evaluation for varying L and J. The detection rate was calculated on simulated data for concentration from 5% to 20% and on real wheat measurements data for 0%.

According to the test statistic of the MSD algorithm (Equation 16), the pixel spectrum $\boldsymbol{x}$ is modeled using two different models: the model using only the regular sample subspace $\boldsymbol{s^S}$, and the model using the adulterant sample subspace $\boldsymbol{s^A}$ besides. Hence, the spectrum $\boldsymbol{x}$ can be described by the two models by projection on the subspaces $\mathbf{M_S}$ and $\mathbf{M_A}$. The corresponding orthogonal subspaces describe the residuals of $\boldsymbol{x}$ under each model. The projection matrices $\mathbf{Q_S^{\perp}}$ and $\mathbf{Q_A^{\perp}}$ are used to project a spectrum $\boldsymbol{x}$ on these subspaces. Precisely, the quantity $\mathbf{x^T Q_S^{\perp} x}$ represents the norm of the residuals of $\boldsymbol{x}$ under this hypothesis H_0. This interpretation shows that the statistic

of the MSD (Equation 16) is the normalized comparison of the residuals under each hypothesis. The geometrical interpretation of T_{MSD} in three dimensions is proposed in the supplementary materials.

According to the previous interpretation, the parameters L and J play a significant role in the design of the MSD algorithm. The dimensions of the subspaces highly influence the residuals of the projection when projecting $\boldsymbol{x}$ on $\mathbf{M_S}$ or $\mathbf{M_A}$. We should highlight the fact that J describes the number of additional dimensions chosen in $\boldsymbol{s^A}$ to construct $\mathbf{M_A}$. It means that $\mathbf{M_A}$describes a model which has a higher number of dimension than $\mathbf{M_S}$. As a result, when $L < J$, the models for the two competitive hypotheses become highly unbalanced in terms of dimension. For instance, if $L = 2$, and $J = 3$, $\boldsymbol{x}$ is modeled using a 2-dimensional subspace under H_0 compared to a 5-dimensional subspace under H_1. This unbalanced situation is expected to be solved by the normalization in the MSD statistic T_{MSD} (Equation 16) and the use of the threshold based on the regular sample. However, when the dimension of $\boldsymbol{s^A}$ increases, the residuals under H_1 can be arbitrarily reduced. It is the case because the peanut flour and the wheat flour samples exhibit spectral similarities. As a consequence, the vectors from $\boldsymbol{s^A}$ have similarities with vectors from $\boldsymbol{s^S}$. Ultimately, the scalar product between $\boldsymbol{x}$ and the vectors of $\boldsymbol{s^A}$ is not null even if $\boldsymbol{x}$ only contains regular particles. This explains why the value of J should be kept under the value of L to prevent for a degenerated design of the MSD algorithm. Figure 25 shows this type of design leads to low detection performances ($L = 1$, $J = 1$ on the left graph; $L = 2$, $J = 3$ and $L = 2$, $J = 4$ on the right graph).

When L and J are high, each model takes a large variability into account. As explained before, this is not a good strategy because of the similarities between peanut and wheat. Additionally, it leads to include variability subspaces that are less significant in the model which leads to overfitting. This explains why the MSD designs with $L > 2$ are not optimal (Figure 25). Finally, when L and J are too small ($L = 1$ and $J = 1$, Figure 23), the design only takes the average shape of the materials as a reference into account. However, considering the variability of the sample is detrimental for this application.

Analysis of the number of detected pixels

Figure 26 shows the detection rate of the three selected MSD (Table 1) calculated on the real samples. The $\boldsymbol{x}$ axis represents the global peanut concentrations that were introduced in the samples. On the other hand, the y axis shows the detection rate: the total number of detected pixels using the MSD algorithms divided by the total number of pixels in the image. The scatter plot shows that the detection rate increases when the sample concentration increases. Hence, the application of the MSD algorithm on the real measurements is relevant to the results obtained on simulated data. The results also show a high variance in the detection rates for a given MSD applied to the three replicate samples with the same concentration. The experimental conditions can

explain it. Hyperspectral measurements are representative of the material through a depth of some millimeters [53]. As a result, the assessment of the detection rate is made on a thin layer of flour. On the other hand, the peanut concentration on the x-axis is a global characteristic of the sample volume. Consequently, both metrics do not describe the same aspect of the sample. As it is complicated to ensure the sample is homogeneous in peanut concentration through all volume, the apparent concentration on the surface of the sample may not be representative of the global concentration. Consequently, sample replicates may exhibit different peanut concentrations in the surface layer despite the fact that they have the same global concentration, which explains the variance observed in Figure 26.

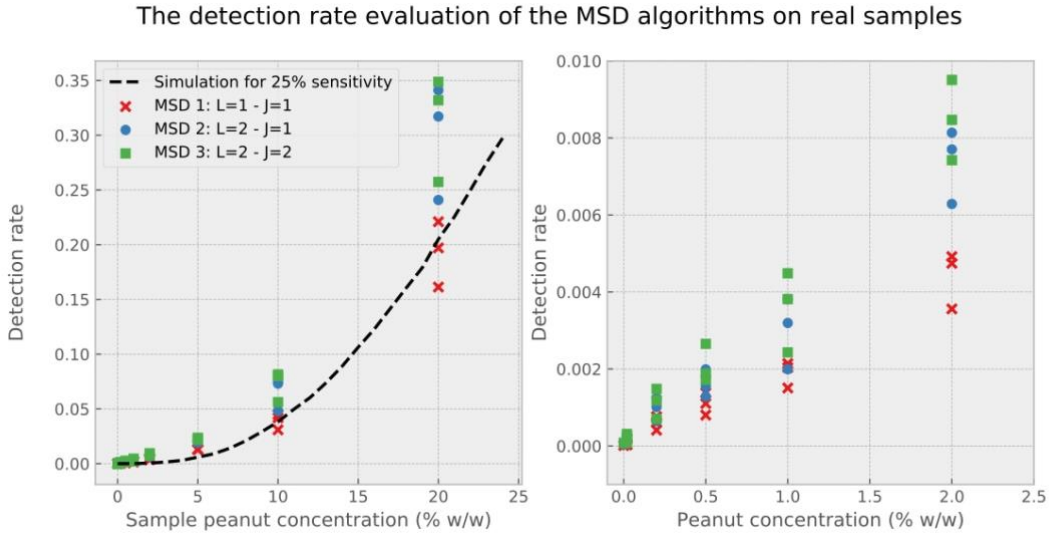

Figure 26: The detection rates of the MSD algorithms evaluated on the real samples (from 0.02% to 20% of peanut concentration) against the global peanut concentration.

The results show a non-linear behavior in the evolution of the detection rate according to the peanut concentration. It is visible on the left-graph of Figure 26. There are three phenomena to take into account to explain this behavior.

1. Let us assume a sample is perfectly homogeneous at the pixel-scale with a peanut concentration of 20%. One MSD algorithm with a detection sensitivity of 10% is applied to the hyperspectral data cube. In this configuration, each pixel has a contribution of 20% of peanut and is detected by the algorithm. As a result, all the pixels are detected on the image, and the detection rate is 1. Let us consider another sample with a peanut concentration of 5%. Following the same reasoning, no pixel is detected and the detection rate is 0. This example shows that since a pixel can only have two states (detected or not detected), the resulting global detection rate has a non-linear relationship with the real concentration of the sample. It also shows that comparing these two values may not be relevant if the sample is entirely homogeneous. As a consequence, the comparison only holds if we assume the scale of scrutiny is higher than the pixel:

the sampling size for which the homogeneity is guaranteed is bigger than the pixel size.

2. More realistically, the fact that a sample has a global concentration of 20% does not mean that each pixel surface has the same concentration. We have to assume the sample is heterogeneous at the pixel-scale. Also, if we assume there is no spatial relationship between neighbor pixels, an image of 100 000 pixels can be considered as 100 000 independent experiments. Each one can be seen as a series of Bernoulli processes with as many trials as the number of particles in the pixel. The probability of selecting a peanut particle corresponds to the global concentration of peanut. A binomial distribution thus gives the pixel-wise concentration. Such a simulation provides a detection rate curve shown on the left graph of Figure 26. The result obtained using a sensitivity of 25%, which is relevant to the experimental observations. It shows that the sensitivity of the MSD algorithm is close to 25% which was shown using the simulation data as well (Figure 23).

In practice, pixels are not independent for two main reasons. Firstly, because flour samples contain several particle sizes and some may be higher than 150 µm. With such a size, some particles may overlap multiple pixels and make them detectable. Secondly, because the particles of flours tend to agglomerate with each other, this is known as the stickiness [36]. Despite the fact the median particle size is approximately 50 µm in wheat and peanut flours, some particle clusters may have a size of several millimeters which is higher than the pixel size (250 µm × 250 µm). The agglomeration occurs particularly often for peanut flour because of the remaining fatty acids.

These arguments show that all the MSD designs provide relevant results regarding the real samples with different concentrations. However, even if the relationship between the number of detections and the global peanut concentration of samples is useful to validate the results, it is complex to interpret. Hence, it should not be the only metric to validate the results of a detection problem.

Analysis of the detection positions

The previous results show that MSD 3 provides the most sensitive results. Figure 27 shows the detection maps for three different concentrations and their replicates. These maps show that the number of detections is repeatable among the replicates as Figure 26 showed. They also show the detection locations are credible: for a high concentration, most of them are made on neighbor pixels so that peanut agglomeration can be seen. Furthermore, the location of these agglomerations is randomly distributed across the sample. These results show the MSD algorithm can be used to detect challenging targets as peanut flour in wheat flour and give their position.

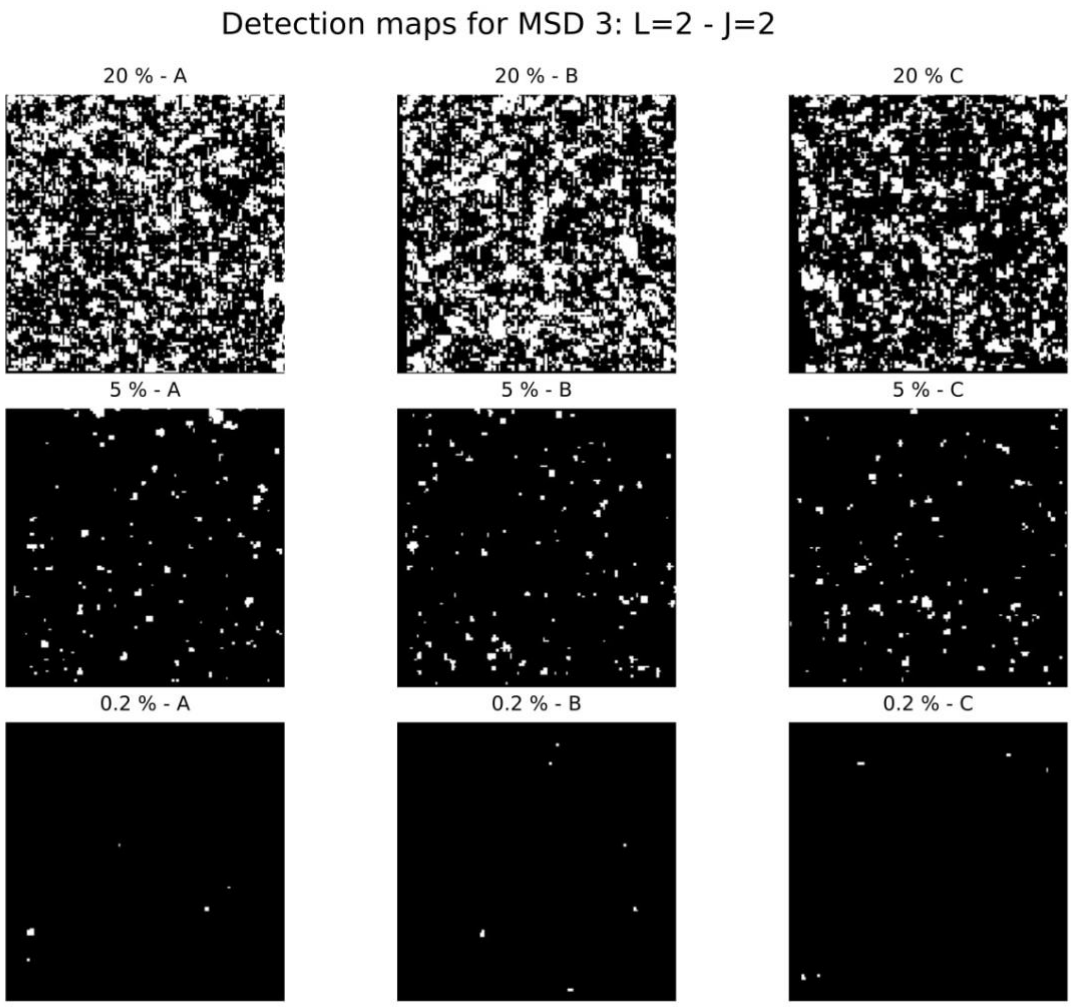

Figure 27: The detection maps obtained by applying MSD 3 on the real samples of concentrations: 20 %, 5 %, and 0.2 %.

4. Conclusions

The purpose of this study was to tackle a challenging detection problem dealing with similar materials with high spectral variability and a particle size involving subpixel detection. The development of the Matched Subspace Detector algorithm was proposed to overcome these difficulties. The spectral variability was tackled using subspace modeling by PCA whereas the Linear Mixing Model was used to consider the subpixel detection context. Moreover, data simulation of several peanut concentrations was proposed to provide an estimation of the sensitivity for the MSD. This technique was used to help in choosing the most appropriate design.

The data simulation method provided realistic data regarding the measurement variability. The MSD designs giving a high detection rate on the simulation were conserved and used on the real measurements. Despite the lack of local reference

values, the number and the positions of the detections show that MSD and the data simulation were relevant to overcome the detection issue.

Additional work could be provided for further improvements to this kind of detection situation. Firstly, the data simulation process could be improved by selecting a subset of loadings to simulate the data. It may provide more reliability in the simulation. Indeed, the expected sensitivity obtained on simulated data (20 %) does not seems to be reached in practice. Then, even if no spatial a priori hypothesis can be made regarding the particle size, the detection results on real data show that most of the detections are made on neighbor pixels. This is because of particle agglomeration which is a phenomenon that applies to the smallest flour particles. Such an effect could be taken into account to improve the detection by adding some spatial dependence in the MSD algorithm. Finally, the statistical simulation for the number of detected pixels according to the peanut concentration may be improved. For example, the hypothesis that each pixel is independent of the other could be changed to get a more accurate approach.

III. The detection of peanut flour in chocolate powder using Multivariate Curve Resolution

This part has been adapted from the publication:

A. Laborde, F. Puig-Castellví, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, C.B.Y. Cordella, B. Jaillais, Detection of chocolate powder adulteration with peanut using near-infrared hyperspectral imaging and Multivariate Curve Resolution, Food Control. 119 (2021) https://doi.org/10.1016/j.foodcont.2020.107454.

1. Introduction

In the industry, there is a high interest in detecting contaminations in powders [36]. Food industry concerns about cross-contamination of ingredients are increasing due to the increasing prevalence of food allergies. On the one hand, for some people, allergic reactions can be triggered with just a few milligrams of pure allergen [83]. On the other hand, a significant 4 % of the total world population suffers from a form of food allergy [84].

Due to the existence of major food allergens such as peanuts [85], food handling in the industry can be especially challenging [86]. In addition, processing methods that focus on powder foods (i.e., producing a dry mix) can increase the risk of food cross-contamination due to the difficulty in cleaning the equipment between two lines of food products in progress [85] and because it is difficult to assess whether all contaminants have been removed after the cleaning phase [84]. However, several strategies exist to detect product contamination. Most detection strategies consist of direct methods that search for specific target molecules (i.e., proteins, DNA) from the adulterant/contaminant, i.e. allergen, using molecular biology and immunological methods. For example, the immunological method ELISA detects proteins with a very low sensitivity (from 1 to 2.5 ppm) [85]. However, ELISA is a destructive method employed on small sample volumes, which might not be representative enough of the whole sample. Additionally, immunological or molecular biology methods are not optimal for automatic screening because they are expensive and time-consuming.

Some examples of product contamination detection using spectroscopy methods can be found in the literature. For instance, the detection of melamine in milk powder was investigated using line scan NIR hyperspectral imaging [10-12][56]. The detection of crushed peanut in wheat flour was performed by applying Independent Component Analysis (ICA) [15] and Principal Component Analysis (PCA) [14] to the hyperspectral data. In the pharmaceutical domain, the amount of low dose of magnesium stearate was analyzed using Multivariate Curve Resolution (MCR) and Raman hyperspectral imaging [88].

Detecting a spectral signature in spectroscopic signals implies two main steps. The first step consists of describing the features of the problem, which are the expected standard spectral profiles as opposed to the adulterant spectral profiles. It implies the description of their average behavior and their variability. In an adulteration detection context, "standard" refers to the expected sample in opposition with the "adulterant" which is to be detected. The second step consists of detecting the adulterant spectral observations in the feature space. Despite the fact that the two steps are equally important, the first one appears to be the main challenge when dealing with detection in NIR.

The previous chapter showed that hyperspectral imaging is subject to the subpixel detection problem [64]. Various chemometric methods (i.e., ICA ([15]) ,MCR ([88]), Non-Negative Least Squares (NNLS) [89]) have been used to address it. All these methods showed high-performance results to unmix hyperspectral images, although the data analysis may become challenging when the particles to detect (from the contaminant) have a spectral signature very close to the background particles (belonging to food sample).

When dealing with food contamination, there is no guarantee the spectral features of the contaminant differ from those of the background. For example, this is the case for melamine contamination in milk powder [56]. In the same line, the peanut NIR spectral signature is very similar to cocoa especially when the products are transformed into flour or chocolate powder. These two food compounds are involved in potential allergen contamination cases in the food industry. In agreement with this, a study of the French market showed that 67% of snacking products labels advised of the possible unwanted presence of peanuts in their product [90]. Thus, assuring the absence of peanut traces in chocolate products is a matter of utmost importance for the food industry.

In this study, we propose to tackle the problem of detecting peanut flour particles in chocolate powder using both the NIR hyperspectral imaging technique and chemometrics methods. First, a PCA was performed as a reference technique and then a detection algorithm based on the MCR-ALS chemometric method was applied. This work provides novel insights into the spectral features of the chocolate-peanut system and presents a methodology to address subpixel detection using hyperspectral imaging with the potential to be implemented in food processing industries. To our knowledge, the study of the detection of food allergens by combining NIR hyperspectral imaging and chemometrics is not common in the literature.

2. Material and methods

A. Sample preparation

A chocolate powder mix to prepare milk beverages was purchased in a French supermarket. The main ingredients of the mixture are sucrose, cocoa, dextrose, and

soya lecithin. Defatted peanut flour was bought on the German market. The powders were mixed in different mass proportions of peanut flour: 10%, 1%, and 0.1%. For each concentration, three replicate samples of 13 g were prepared. In addition, pure chocolate powder and pure peanut flour samples were also prepared in triplicate, leading to a total of 15 samples. For the spectral measurements, the powders were put in a plastic sample holder made of polylactic acid. The powder was skimmed on the top to achieve a thickness of 7 mm. This thickness ensured that the bottom of the sample holder did not have an influence on the near-infrared reflectance signal measured by the hyperspectral system [53].

B. Hyperspectral imaging system

A line-scan pushbroom Specim SWIR camera (Specim, Spectral Imaging Ltd, Oulu, Finland) was used to acquire the hyperspectral images. The system acquires 288 spectral bands from 1000 to 2500 nm with a spectral sampling of 5.6 nm. The spectral range 1000 – 2500 nm was chosen as it proved its superior capability to analyze the chemical content of peanut compared to the range 400 – 1000 nm [91]. Additionally, it was successfully used for the measurement of sugar in a similar chocolate powder [92]. The camera was moving along the y-axis and acquired 320 pixels per line to form the hyperspectral cube. Six halogen lamps were used to illuminate the sample. A white diffuse reflectance standard in Teflon was used to acquire the white reference image before each measurement. The dark reference image was acquired by closing the shutter of the camera.

C. Data Processing

Hyperspectral cubes were cropped in the spatial dimension to obtain the same image size for each sample. Each image was composed of 61 × 61 pixels, which corresponded to a field of view of 1.5 cm × 1.5 cm. The reflectance cube was calculated using Equation 9. Wavelengths from 1000 nm to 1100 nm and from 2400 nm to 2500 nm were removed because they were mainly representative of electronic noise. The remaining absorbance were smoothed using a Savitsky-Golay filter (second order polynomial, 7-point window, and with no derivative).

D. Hyperspectral cube unfolding

Hyperspectral images can be regarded as tridimensional hyperspectral cubes, where the $\boldsymbol{x}$ and $\boldsymbol{y}$ planes correspond to the spatial dimensions and the $\boldsymbol{z}$ plane contains the hyperspectral data for every pixel (left side in Figure 28). This cube can be unfolded into a two-dimensional matrix, with as many rows as pixels and as many columns as

measured wavelengths (right side in Figure 28). Every row in this matrix contains the spectrum relative to one pixel. This data unfolding strategy is required to investigate tridimensional data with bilinear methods such as MCR-ALS [93].

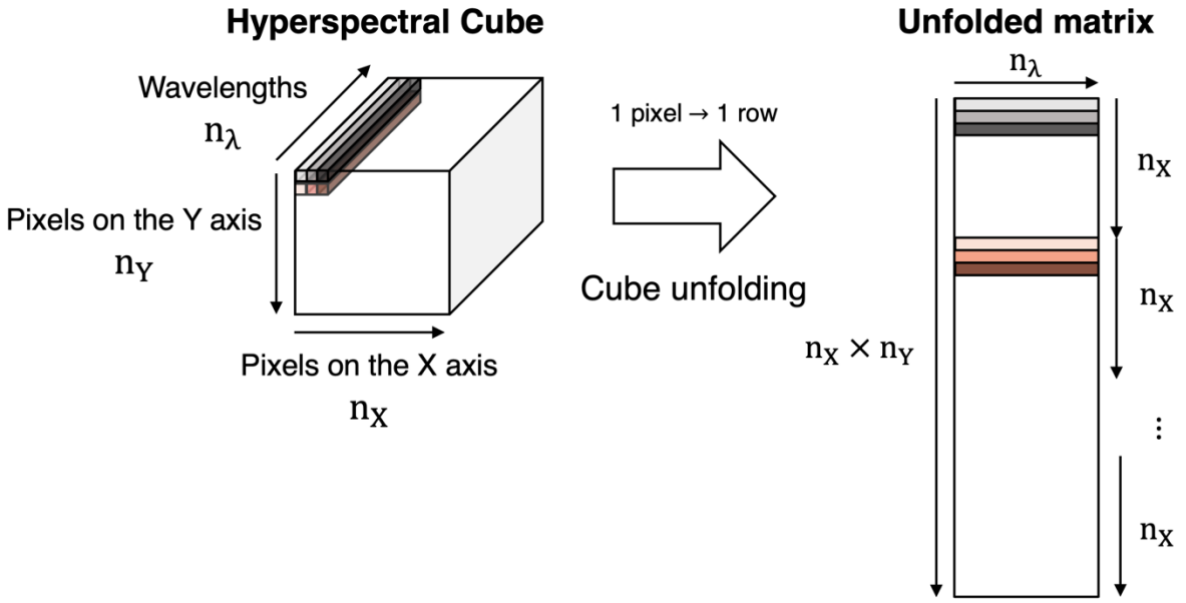

Figure 28: Hyperspectral cube unfolding.

E. Multivariate Curve Resolution – Alternating Least Squares

MCR-ALS is a chemometric method used to solve the unmixing problem [94]. According to MCR-ALS [43] [95], a matrix $\mathbf{X}$ containing mixed signal measurements can be decomposed as the product of the pure spectral profiles, $\mathbf{S}$, associated with their pure concentration contributions, $\mathbf{C}$, using Equation 3.

In this decomposition, the number of pure profiles to be resolved is determined by the number of components, which needs to be estimated before running the analysis. The number of components can be estimated using the SVD algorithm [96]. MCR-ALS algorithm is an iterative method that optimizes a set of initial estimates for the concentration ($\mathbf{C}$) or the spectrum profiles ($\mathbf{S}$) under constraints while minimizing the residual part $\mathbf{E}$.

In this study, the initial estimates used were the most dissimilar spectra found in the hyperspectral images of the pure samples of peanut flour and chocolate powder. One spectrum was selected for peanut flour, while two spectra were chosen for chocolate powder due to its higher complexity. Hence, during the MCR-ALS analysis of the mixture samples, three components were used.

Concentration and spectral profiles obtained by the application of the bilinear model (Equation 3) may not be the correct ones due to the existence of rotational ambiguities [97]. The bilinear model allows that several sets of concentration profiles and spectra with different shapes can reproduce the data $\mathbf{X}$ with the same precision [97]. In other words, the optimization problem is under-constrained which leads to a

great deal of possible solutions. To drive the iterative analysis towards the purest solution, some known information from the system of study can be used as constraints. Examples of constraints include non-negativity and selectivity constraints. The non-negativity constraint can be applied when the resolved spectral data must be positive. The selectivity constraint is used to fix some values in the **S** or **C** matrices. Finally, the spectral profiles can also be normalized to reduce the intensity ambiguity of the solutions [98]. This ambiguity is caused by the fact that the intensity of spectra can be reproduced by multiple dyads of profiles and concentration with different arbitrary scales [43].

Augmented matrix

By analyzing more than one sample at once, more information relative to the pure components is introduced in the analysis and the possible ambiguities can be substantially reduced. In this work, MCR-ALS was applied to the column-wise augmented data matrix resulting from stacking the 15 unfolded hyperspectral image samples, **X.** This matrix has 55,815 rows and 248 columns (Figure 29).

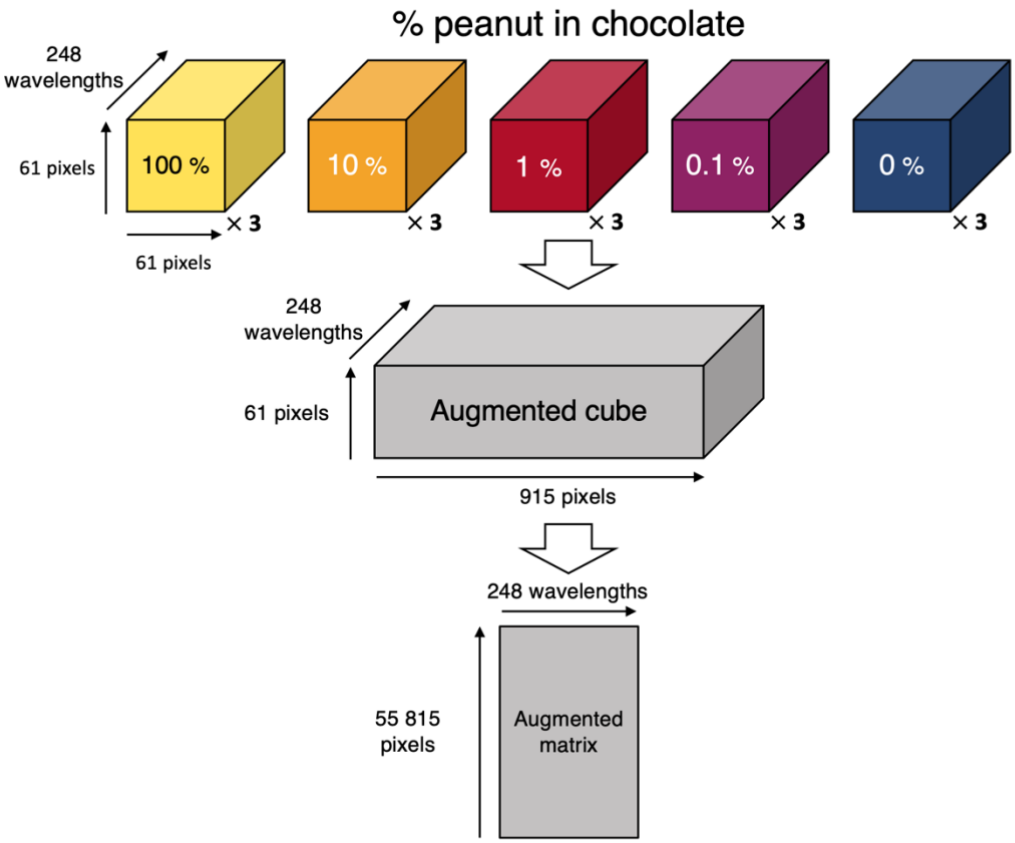

Figure 29: The hyperspectral cubes are horizontally stacked as an augmented cube.

Selectivity constraint in the concentration matrix

As stated before, an additional constraint can be applied to introduce known information in the MCR-ALS analysis and reduce the ambiguity of the final solution. In the case of the MCR-ALS multiset analysis, the correspondence among constituents and MCR-ALS components allows the introduction of information about the presence or the absence of these constituents in the pixels. In the situation depicted in section 2.A, the composition of the pixels of the pure samples is known, while the composition of every pixel in the mixed powder image is not and therefore it could not be inferred. The correspondence among species, i.e., the selectivity constraint in the $\mathbf{C}$ matrix (abbreviated as "CSEL") was implemented to constrain only the pure sample images.

With this constraint, components representative of chocolate powder (components 1 and 2) were imposed to not show contribution in pure peanut pixels. Analogously, the component representative of peanut flour (component 3) was imposed to not show contribution in chocolate powder pixels.

In this study, we tested and compared both MCR-ALS resolution methods with and without a selectivity constraint. For simplicity, in this paper, these methods will be referred to as MCR-ALS and MCR-ALS-CSEL, respectively.

F. Detection algorithm

After decomposing the signal of the pixels by MCR-ALS, the resulting $\mathbf{C}$ matrix was investigated with a detection algorithm to determine the presence or absence of peanut. In this detection algorithm, the $\mathbf{C}$ matrix was used to build a Gaussian Mixture Model (GMM) [99] on the pure chocolate powder pixels. The GMM consisted of modeling the distribution of points in the 3-dimensional sub-space defined by the 3 MCR-ALS components using several 3D Gaussian distributions. In this study, 2 Gaussians were fitted using the Expectation-Maximization (EM) algorithm [99] to model the distribution of chocolate powder pixels. The Mahalanobis distances between each pixel (represented each with the 3 concentration profiles from the $\mathbf{C}$ matrix) and both Gaussians were then computed to obtain the score for detection. The threshold for detection was chosen as the highest Mahalanobis distance measured between the GMM and a pixel from the pure chocolate powder samples.

G. Software

Data processing were performed in Matlab R2016a (MathWorks Inc.) using the SAISIR toolbox [100]. Processed samples were analyzed with the MCR-ALS method using the MCR-ALS GUI 2.0 under Matlab environment [101].

3. Results and discussions

A. Principal Component Analysis

A PCA was first applied to the $\mathbf{X}$ augmented matrix containing the hyperspectral data from the pixels of all the sample images The PCA score plot on the first two PCs is given in Figure 30. In this figure, the scores from the pure peanut flour pixels are clearly separated from those from the pure chocolate powder pixels. Regarding the scores from the mixture samples, they were found clustered around the distribution of pure chocolate pixels scores.

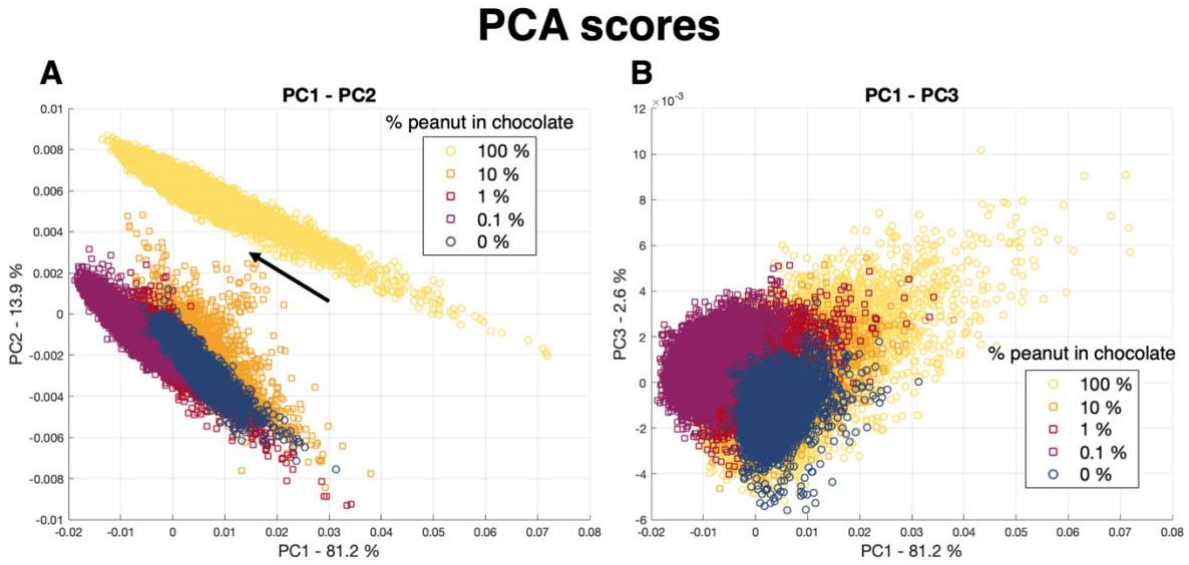

Figure 30: PCA score plots for chocolate powder with peanut.

PC1 (81.2 % of the total explained variance of the dataset) mainly describes the variability among pixels within the same sample type, while PC2 (13.9 % of the explained variance) allows the discrimination of the pure chocolate pixels cluster from the pure peanut pixel cluster. PC3 (2.6 % of the explained variance) does not show any separation of the pixels between the pure peanut and the pure chocolate. The same observations were made on the following principal components (results not shown).

The results from this PCA illustrate the so-called "subpixel detection problem" [64], since the scores distribution shows that no pixel from the mixture samples is of "pure peanut" as they do not fall into the area of the pure peanut cluster. However, a few scores from the mixture pixels were found between the pure chocolate and the pure peanut variability distributions (see the black arrow in Figure 30A). These scores represent the pixels from the mixture samples containing, in addition to the spectral signatures from chocolate, the most important spectral contribution from peanut. Nevertheless, these intermediate scores were not observed for pixels from samples containing 1% of peanut, proving that PCA cannot detect peanut adulteration at this

concentration level. Therefore, PCA is not an optimal tool to detect peanut pixels in chocolate matrices. We can argue that PCA limitation is derived from the fact that spectral differences linked to the adulteration do not dominate the dataset, since these spectral differences are very small and only occur in a small fraction of the total number of pixels. PCA also imposes the components to be orthogonal to each other. However, this constraint is not representative of the real chemical signatures. Hence, the principal components calculated by PCA does not reflect the true chemical signatures of the mixture problem [43].

B. MCR-ALS

Since PCA could not extract the spectral components in the pixels relative to peanut flour and chocolate powder, MCR-ALS method was applied instead. Figure 31 shows the $\mathbf{C}$ concentration profiles for the three MCR-ALS resolved components, which are representative of the pixels. Thus, MCR-ALS results shown in Figure 31can be directly compared to the PCA scores shown in Figure 30. For instance, in Figure 31, the pure pixels from peanut and chocolate can be discriminated using the third component ($\boldsymbol{c_3}$) only. On the other hand, two principal components were needed to discriminate the pure samples on the PCA score plot (Figure 30), demonstrating that MCR-ALS components are much easily interpretable than the PCA components. The better outcome found for MCR-ALS is mainly derived from the use of the non-negativity constraint that reduced the ambiguity of the data decomposition. Since negative concentrations were not allowed in the resolution of the MCR-ALS model, the spectral components could not compensate for each other. As a result, the spectral shapes of the MCR-ALS component were closer to the pure chemical compounds (Figure 32).

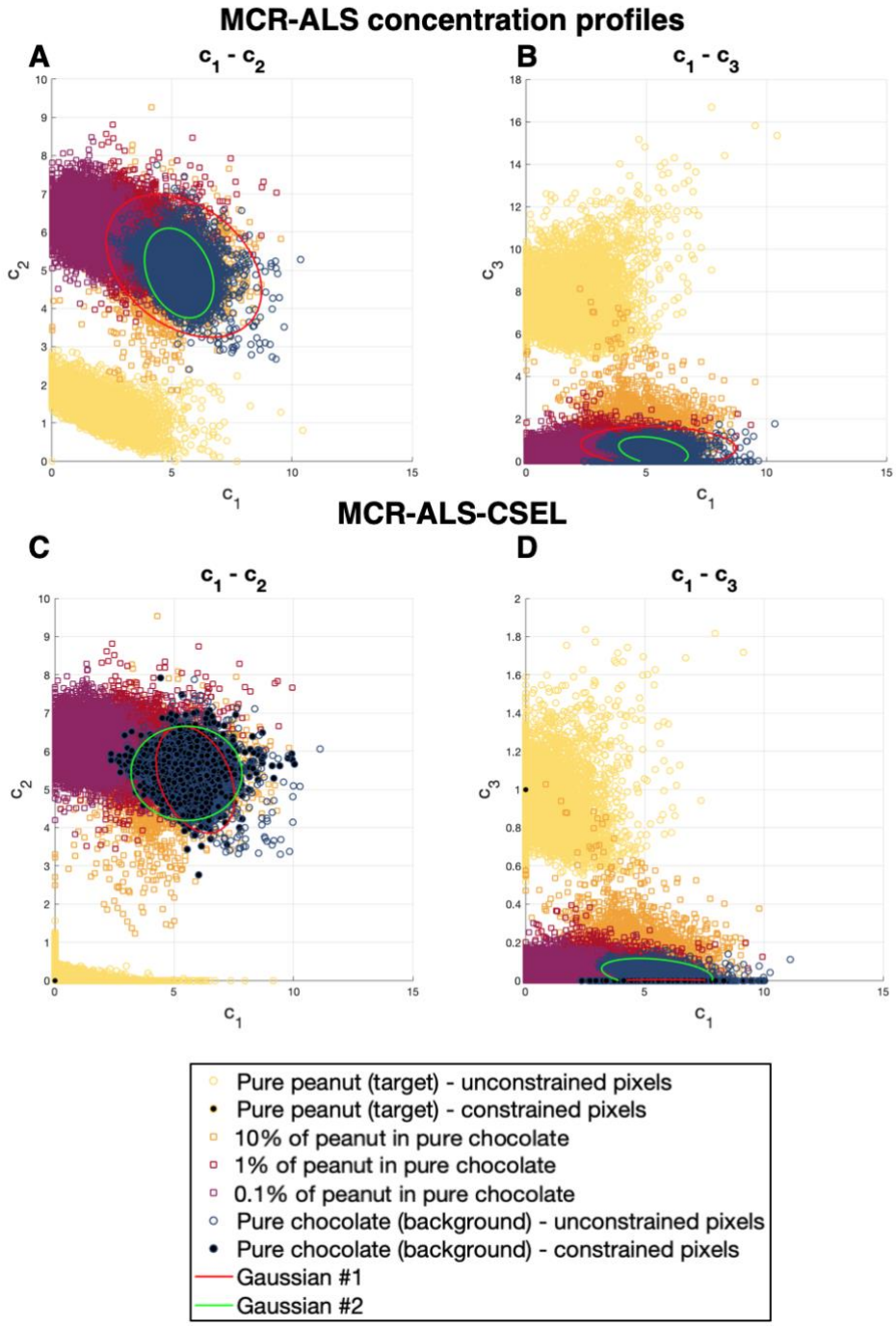

Figure 31: Scatter plots of the resolved MCR-ALS $\mathbf{C}$ concentration profiles with the ellipses of the GMM.

Consequently, some pixels from the samples containing 10 % peanut were found within the pure peanut variability whereas this was not observed in the PCA, which illustrates that MCR-ALS provided a more suitable model for the detection problem of peanut in chocolate powder than the PCA.

Figure 32 shows the resolved MCR-ALS spectral profiles in blue compared to their initial estimates in black.

The first spectral profile ($\boldsymbol{s_1}$, Figure 32A) is descriptive of cocoa as it includes some spectral features characteristics from this constituent as reported elsewhere: the low intensity and spread absorption peak at 1208 nm is due to the combination of the

second overtone of CH, CH_2, and CH_3; a higher intensity level absorption can be observed at 1491 nm corresponding to the N-H group characteristic of the proteins in cocoa (Krähmer et al., 2015); the absorption at 1935 nm is attributed to the second overtone of amide C=O (Workman Jr. & Weyer, 2012) which is also descriptive of proteins in cocoa.

The second spectral profile ($\boldsymbol{s_2}$, Figure 32 B) is descriptive of the sucrose content present in the chocolate powder. The intense absorption peaks at 1435 nm and 2072 nm in $\boldsymbol{s_2}$ are associated with the C-H stretching and the combination of O-H stretching in sucrose, respectively [92]. Additionally, a double peak at 2280 nm was attributed to the C-H stretching and CH_2 deformation of polysaccharides [102].

The third component is associated with the peanut spectral signature. The peaks at 1474 nm and 1735 nm are related to the N-H second overtone from proteins and to C-H group of amine polysaccharides respectively [102]. The two main absorption peaks at 1200 nm and 1942 nm are representative of the water absorption contained in the flour. Since the peanuts were defatted before being milled into flour, the fatty acids of peanut were not present in the final product. As a consequence, the near-infrared spectrum of peanut was mainly characterized by its proteins, causing the detection problem to become more difficult as there were fewer spectral signatures characteristic of peanut to be detected in the mixture samples.

Differences between the initial estimates and the final resolved MCR-ALS spectral profiles varied across components. On the one hand, $\boldsymbol{s_2}$ and $\boldsymbol{s_3}$ did not change significantly after the MCR-ALS iterative process. This occurred because the pixels chosen as initial estimates were mainly constitutive of the compounds they describe (sucrose and peanut flour, respectively). However, $\boldsymbol{s_1}$ showed a more prominent alteration of the spectral profile after the iterative process is performed, allegedly because all pixels from cocoa samples contain sucrose and therefore an initial estimate from cocoa without sucrose could not be used. Nevertheless, even without the proper initial estimate for $\boldsymbol{s_1}$, MCR-ALS reached a satisfactory resolution after the iterative process. This result highlights that the MCR-ALS method is robust enough to extract the pure spectral components of chocolate powder.

The coefficient of determination between $\boldsymbol{s_1}$ and $\boldsymbol{s_3}$ is high ($r^2 = 0.96$), indicating that these spectral profiles of cocoa (without sucrose) and peanut flour are very similar in spectral shape. As a consequence, the obtained MCR-ALS solutions may not be the purest ones due to ambiguities between the two spectra. In fact, this ambiguity could explain why some coefficients $\boldsymbol{c_1}$ for the pure peanut pixels were not null (Figure 31). For this reason, the $\mathbf{C}$ concentration profiles must be read with caution and therefore they should not be used directly for detection purposes.

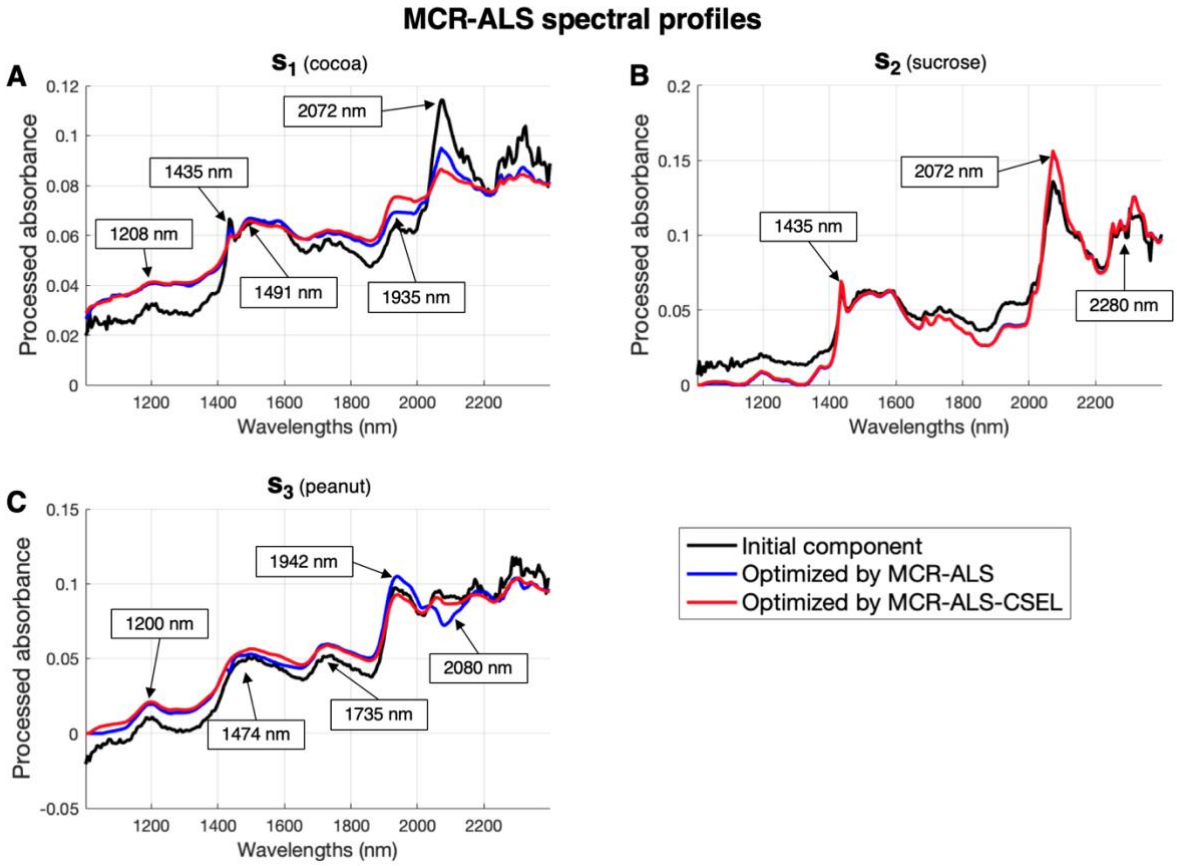

Figure 32: MCR-ALS optimization of the spectral profiles.

C. MCR-ALS-CSEL

After noticing some spectral ambiguities in the MCR-ALS model performed in the previous section, the MCR-ALS analysis was repeated using CSEL as an additional constraint (see Methods). With CSEL, the value of the $\mathbf{C}$ concentration profiles can be imposed for certain known pixels, resulting in a reduction of the ambiguity of the system of study. In this work, we imposed that pure peanut pixels only contain spectral signatures from $\boldsymbol{s_3}$, while pure cocoa pixels only contain spectral signatures from $\boldsymbol{s_1}$ and $\boldsymbol{s_2}$.

Table 2 shows the fitting performances of MCR-ALS and MCR-ALS-CSEL. Both methods exhibit high performances with $R^2 > 0.99$. The addition of constraint to the MCR-ALS algorithm leads to consider a tradeoff between the fitting of the data matrix $\mathbf{X}$ and the satisfaction of the constraints. Table 2 shows that despite the addition of constraints in the MCR-ALS-CSEL method, the fitting performances are still similar to the MCR-ALS. This indicates that the optimal solution was reached regardless the MCR-ALS method was more constrained. The effect of CSEL can be observed in Figure 32**.** In this figure, the spectral profiles resolved with MCR-ALS-CSEL are shown in red, while those resolved with MCR-ALS are shown in blue. Interestingly, some spectral differences can be observed for $\boldsymbol{s_1}$ and $\boldsymbol{s_3}$ spectral profiles. On the other hand, s_2

spectral profile remains exactly the same for both methods indicating that the pure profile for this component can be achieved without using the given constraint.

Table 2: The performance in lack of fit and coefficient of determination (R^2) for MCR-ALS and MCR-ALS-CSEL.

	MCR-ALS	MCR-ALS-CSEL
Lack of fit (%)	3.16 %	3.27 %
R^2	0.9990	0.9989

When looking at the differences among the resolved spectral profiles, it can be observed that $\boldsymbol{s_1}$ and $\boldsymbol{s_3}$ profiles both differ on similar wavelengths around 2080 nm and 1940 nm. These two spectral patterns are associated with the second overtone of amide group and to the combination O-H stretching, and they can be attributed to both cocoa or peanut constituents. Regarding the unconstrained MCR-ALS method, it resulted in a higher absorption peak at 2072 nm and a smaller absorption peak at 1935 nm in $\boldsymbol{s_1}$, and a smaller absorption at 2072 nm and a higher absorption at 1935 nm in $\boldsymbol{s_3}$.

The CSEL constraint had visible effects on the distribution of pixels in the MCR space (Figure 31). Figure 31A and Figure 31C show the evolution of the distributions for the $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ contributions with and without the application of the CSEL constraint. The main effect observed was the shrinkage of the peanut pixels' distribution, which means the variability of the $\mathbf{C}$ concentration profiles from peanut pixels was reduced after the application of the constraint. The small variability of pure peanut pixels was expected because they do not contain any sucrose nor cocoa and thus the resolved $\boldsymbol{c_1}$ and $\boldsymbol{c_3}$ concentrations should be very low.

Figure 31C and Figure 31D also showed that, in overall, the distribution of the mixture pixels was not changed by the application of the constraint. Consequently, some scores from mixture pixels were closer to the cluster of peanut distribution for the MCR-ALS-CSEL method. On the other hand, Figure 31C shows that the cluster of constrained pixels from the chocolate powder pixels is very similar to the cluster of unconstrained pixels showing that the MCR-ALS-CSEL is robust.

These results showed the MCR-ALS-CSEL method was more suitable for solving the unmixing problem than the unconstrained one since it provided concentration profiles more suitable for detection purposes.

D. The detection results

In the previous sections, two MCR- ALS methods were used to decompose the pure spectral signatures found in every pixel into different components representative of the pure sample constituents. In these analyses, two of the resolved components ($\boldsymbol{s_1}$ and $\boldsymbol{s_2}$) were representative of chocolate powder while the other ($\boldsymbol{s_3}$) was of peanut. However, we observed that, for the two MCR-ALS analyses, some pixels from pure

(chocolate powder and peanut) samples were decomposed as a mixture of non-zero contributions when at least one of the contributions should have been of 0. The fact that some pure peanut pixels presented $\boldsymbol{c_1} \neq 0$ and $\boldsymbol{c_2} \neq 0$ and some chocolate powder pixels presented $\boldsymbol{c_3} \neq 0$ denote that there was still some ambiguity in the MCR-ALS resolution. This phenomenon also occurred for the non-constrained pure pixels when the MCR-ALS-CSEL method was used, although the abovementioned coefficient contributions were closer to 0 than when the resolution was performed with MCR-ALS. This result suggests that the value given of 0 in $\boldsymbol{c_3}$ (the component representing peanut) must not be used directly as the threshold level to determine the presence or absence in our samples.

Alternatively, using the lowest value found in c_3 of pure peanut pixels as the threshold level would cause that the detection algorithm is very restrictive (Figure 31B-D shows that only a few pixels from the 10 % adulterated sample would be selected). Instead, a more sophisticated approach is needed to determine peanut adulteration at the pixel level from these MCR-ALS results. For instance, to reduce to a larger extent the ambiguity of the system, it would have been desirable to constrain some pixels from the mixture samples with the CSEL constraint. However, this is unfeasible as the spatial location of the adulterated pixels is not known after performing the mixing.

To overcome this limitation, we implemented a strategy to determine whether a mixture pixel is within the pure chocolate powder variability or not. In the latter situation, then the pixel could be considered adulterated. Therefore, to use this method, the only information needed is the product (chocolate powder) variability.

The distribution of pure chocolate pixels in the MCR space was modeled with two Gaussian models using the Expectation-Maximization algorithm. Two Gaussian models were needed to account for the specific variabilities of the two major ingredients found in chocolate (sucrose and cocoa). The resulting Gaussian models are represented by the projection of their 99% confidence ellipses in the MCR space in Figure 31. Important observations can be made from the analysis of the shape of these ellipses. First, the variability of the chocolate powder was effectively reduced on $\boldsymbol{c_3}$ when using the CSEL method since the ellipses' areas from the $\mathbf{C}$ concentration profiles obtained in this method were smaller than when the non-constrained method was used. Second, there were a significant amount of mixture pixels with lower $\boldsymbol{c_1}$ concentration profile than in the pure chocolate powder pixels. Similarly, some other mixture pixels have higher $\boldsymbol{c_3}$ concentration profile than in the pure chocolate powder pixels. A low $\boldsymbol{c_1}$ coefficient cannot be regarded as a robust proof of peanut adulteration, as it only indicates that a pixel does not contain a standard amount of cocoa. Conversely, a high $\boldsymbol{c_3}$ coefficient is undoubtedly indicative of peanut since the associated spectral profile is from this ingredient (Figure 32). However, since the interpretation of $\boldsymbol{c_1}$ and $\boldsymbol{c_2}$ concentration profiles may be required for cases when peanut adulteration is very low, we implemented a detection algorithm that used the information relative to these three concentration profiles.

Specifically, to account for $\mathbf{C}$ concentration profiles variability, the designed algorithm establishes the peanut adulteration in every pixel on the basis to their Mahalanobis distance with the GMM.

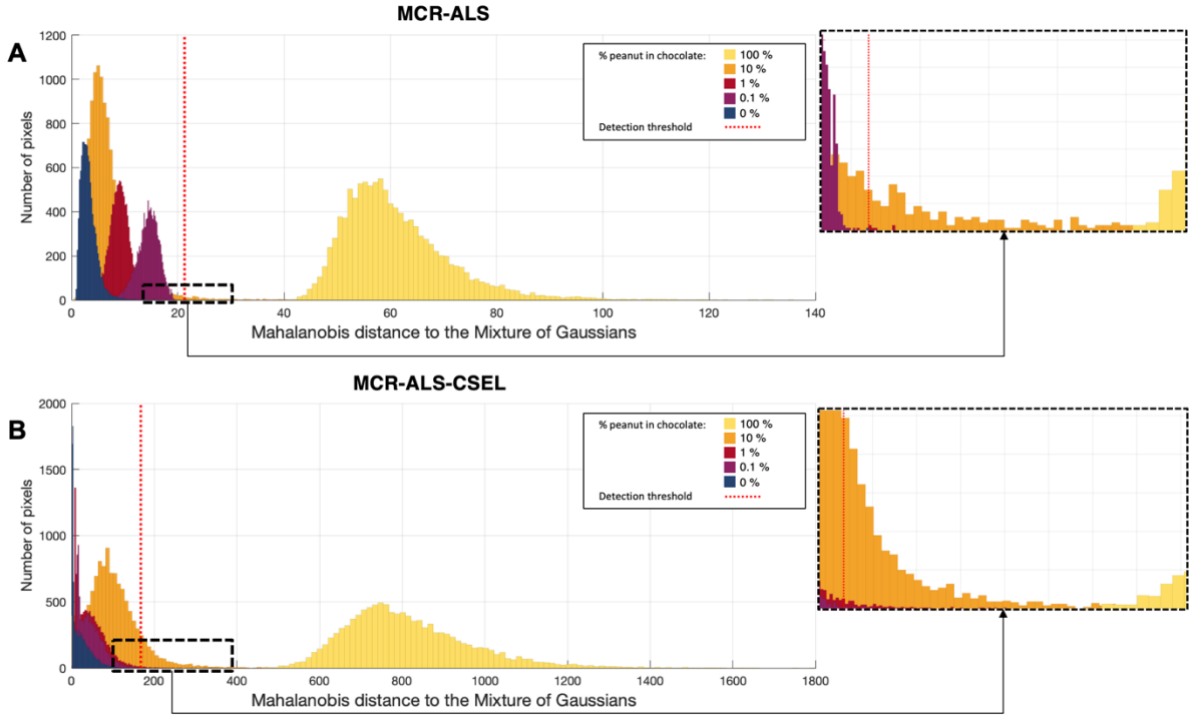

Figure 33: Histograms of the Mahalanobis distance of every pixel to the GMM for the detection of peanut particles.

Figure 33 shows the histograms of the Mahalanobis distances calculated between every pixel and the Mixture of Gaussians from chocolate distribution obtained under both MCR-ALS methods. From this histogram representation, a threshold value that discriminates pure chocolate powder pixels (blue distribution in Figure 33) from the rest can be estimated. This threshold value corresponds to the highest Mahalanobis distance found for the pure chocolate pixels. Thus, any pixel surpassing this threshold was considered to be adulterated.

Figure 33 revealed important differences between the distributions obtained from the two MCR-ALS analyses. For instance, the distribution of the Mahalanobis distances for chocolate pixels was narrower when using the CSEL constraint. As a consequence, the detection threshold could be set smaller leading to more positive detection pixels in the mixture samples. This additional detection power derived from the use of the CSEL constraint can be appreciated by comparing the two highlighted regions in Figure 33. In this region, it can be observed that a higher number of mixture pixels were considered to be adulterated when the constraint was used.

To assess the reliability of the detection algorithm, the numbers of positive peanut adulteration pixel detections for each sample image were calculated (Table 3). From these numbers, it is observed that all the pure pixels were correctly assigned to chocolate powder or peanut flour. Therefore, the sensitivity and the specificity (calculated from the pure samples) are 1 for the two methods. This shows that the two

methods are reliable to confirm the purity of the samples. Regarding the mixture samples, it is not possible to confirm the validity of the assignment since the actual position of the adulterated pixels in the mixture cannot be known. However, by comparing the proportion of detected adulterated pixels found with the corresponding peanut concentrations, we still can evaluate the quality of the detection algorithm.

Table 3: Number of detected pixels in each sample image when using the MCR-ALS and the MCR-ALS-CSEL methods.

Concentration of peanut (%) - replicate	**MCR-ALS-CSEL**		**MCR-ALS**	
	Number of detections [a]	Percentage of detection (%) [b]	Number of detections [a]	Percentage of detection (%) [b]
100 % - A	3,721	100	3,721	100
100 % - B	3,721	100	3,721	100
100 % - C	3,721	100	3,721	100
10 % - A	382	10.30	45	1.21
10 % - B	325	8.73	39	1.05
10 % - C	633	17.00	82	2.20
1 % - A	29	0.78	2	0.05
1 % - B	24	0.64	2	0.05
1 % - C	44	1.12	1	0.03
0.1 % - A	9	0.24	0	0
0.1 % - B	11	0.30	0	0
0.1 % - C	1	0.03	0	0
0 % - A	0	0	0	0
0 % - B	0	0	0	0
0 % - B	0	0	0	0

[a] Number of detected pixels in the image containing a total of 3,721 pixels.

[b] Proportion of detected pixels with respect to the total number of pixels in the image (3,721 pixels).

Table 3 shows that the number of detections was in line with the global adulteration level. For the mixture samples containing 10% peanut, the MCR-ALS-CSEL method detected between 8.73% and 17.00% of adulterated pixels. On the other hand, for the mixture samples containing 1% peanut, between 0.64% and 1.12% of the pixels were found to be adulterated. The unconstrained MCR-ALS method also found adulterated pixels at these two concentration levels. Nevertheless, the number of positive detections for the unconstrained MCR-ALS method was inferior. The coefficient of correlation was used by Vermeulen et al. to measure the relevancy of the detection rate between adulterated samples [77]. After considering only the adulterated samples, the coefficients of correlation were 0.92 and 0.93 for the detection methods using the MCR-ALS and the MCR-ALS-CSEL respectively.

Regarding the mixture samples with 0.1% adulteration, a few adulterated pixels were found when data were analyzed with the detection algorithm based on the MCR-ALS-CSEL method. However, the same samples were considered to be pure chocolate powder samples when analyzed with the detection algorithm based on the unconstrained MCR-ALS method. Hence, the algorithm based on MCR-ALS-CSEL has a lower limit of detection than the algorithm based on the unconstrained MCR-ALS.

Figure 34 shows the spatial positions of the adulterated pixels on the mixture sample image according to both methods. In this figure, it is observed that the adulterated pixels are aggregated in clusters, and these clusters are larger proportionally with the adulteration level. The presence of these clusters, rather than a random distribution of the adulterated pixels, suggests that peanut flour cannot be homogenized easily in the chocolate powder.

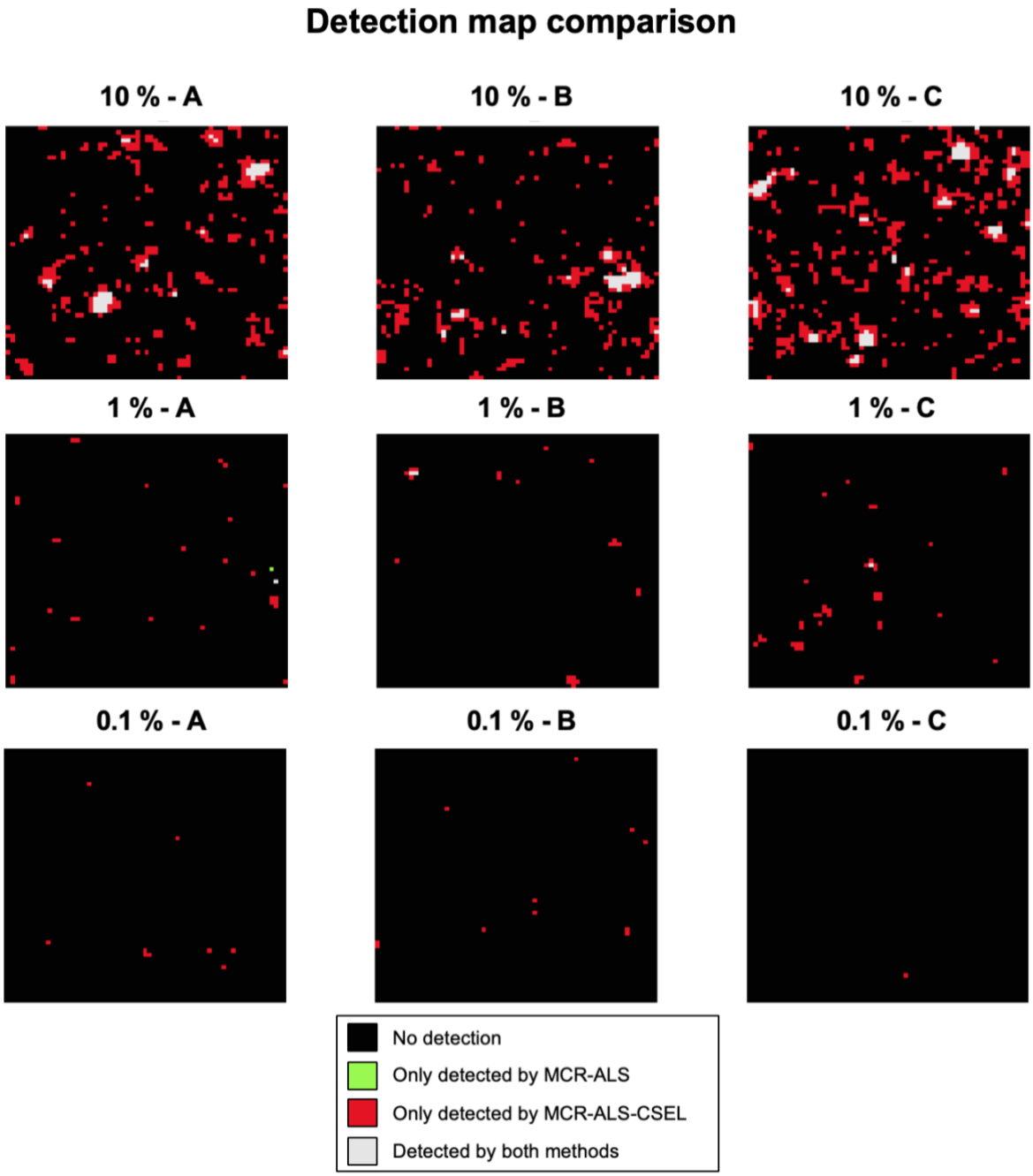

Figure 34: Comparison map for each adulterated sample and the two detection methods.

Figure 35 shows the repartition of the particle size of peanut flour and chocolate powder. For peanut, two main modes were observed at 250 µm and 600 µm and the maximal particle size reached 1 mm. However, this particle size distribution did not correspond to the expected particle size of the product. For example, the FAO standard for wheat flour limits the particle size to be less than 212 µm for at least 98% of the particles [37]. The apparition of bigger particle clusters could be explained by the agglomeration phenomena of small particles, which might have occurred during the mixing process. Therefore, the observed clusters of detected pixels in the spectral data correspond to agglomerated peanut particles.

To conclude, the detection maps in Figure 34 compare both detection methods. These figures, apart from revealing that the MCR-ALS-CSEL method detected much more adulterated pixels than without the constraint, also show that the coincident adulterated pixels were found in the center of the clusters. This particular pixel distribution can be explained because those pixels contain a higher quantity of peanut, and therefore are more likely to be recognized by the detection methods as adulterated pixels than those from the outer regions of the clusters.

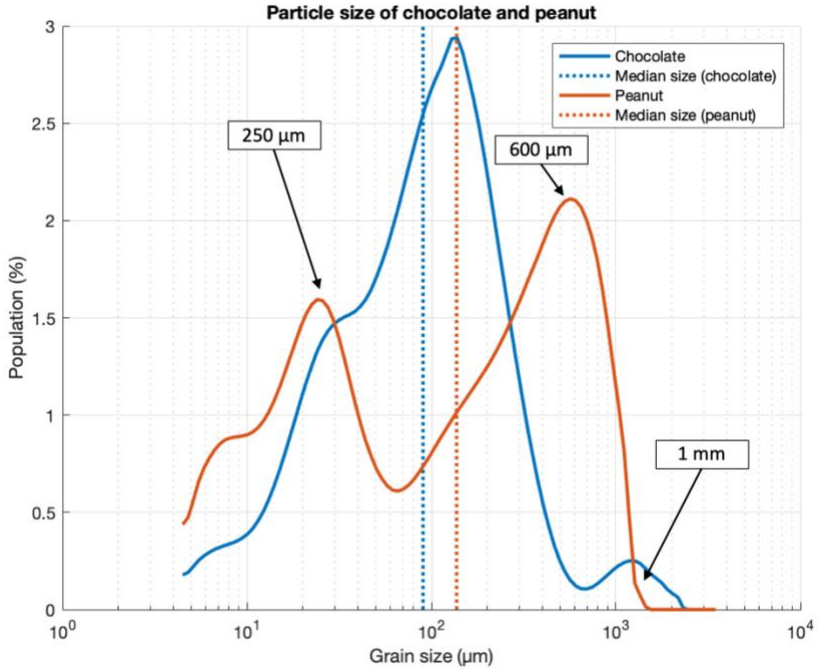

Figure 35: The particle size distribution of chocolate powder and peanut flour.

These results showed that the proposed methods, and in particular the one using MCR-ALS in combination with the CSEL constraint, are very reliable to detect food adulterations even at low concentration levels. The analysis of the adulterated pixels highlighted that the method presented a promising detection power, as the number of adulterated pixels was coincident with the experimental conditions in the terms of signal-response. Moreover, this detection method also revealed that peanut adulteration in chocolate powder is not homogeneous and forms clusters of particles. Despite the spectral similarities of chocolate powder and peanut flour, the presented approach could lead to satisfactory detections of the adulterated pixels. Therefore, this methodology has potential to be used for solving other complex food adulteration systems. Additionally, the detection results showed that low adulteration levels can be detected by screening a small number of pixels. In our case, we were able to detect peanut adulteration in a chocolate powder matrix mixed with 0.1% peanut by only screening 61 × 61 pixels.

4. Additional discussions

In this part, the detection strategies depicted in Chapter II and Chapter III are discussed.

A. The pixel unmixing strategy

The detection method

The literature shows at least five main chemometrics methods to provide a bilinear model of spectral data [103]. PCA and MCR-ALS are two popular techniques that use different assumptions to solve the bilinear model. PCA assumes that the spectral components are orthogonal to each other. On the other hand, MCR-ALS applies various constraints on the system such as non-negativity on the spectral profiles. Besides, the Independent Component Analysis (ICA) is another method that does not impose the orthogonality of spectral components like PCA does. Instead, it looks for the statistical independence of the spectral sources. According to the Central Limit Theorem, linear combinations of independent signal sources tend to be more Gaussian than the sources. Hence, ICA aims to find the least Gaussian source signals in a data matrix [104]. Non-Negative Matrix Factorization (NMF) is another method that provides strictly positive components. It is an appropriate technique to use with NIR signals which have positive values [105]. Minimum Volume Simplex Analysis (MVSA) is another method that is mainly used in remote sensing and satellite imaging [103]. This technique uses a linear model (similar to the LMM) with the constraints that the abundance coefficients are all positive, and their sum is 1. As for the other methods, MVSA is looking for the spectral components that describe the system. In this case, the criterion is the minimization of the simplex volume that encompasses the spectral observations. Figure 36 shows the geometrical principle of this method: the spectral endmembers (m_1, m_2, m_3) are the vertices of the simplex that encompasses the spectral observations.

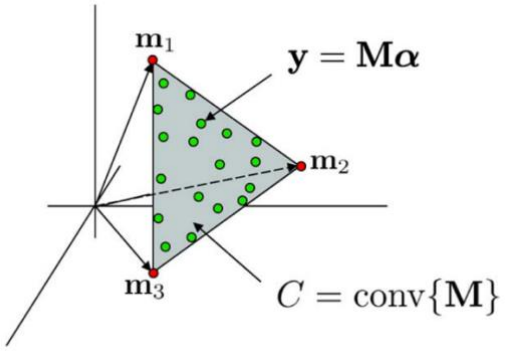

Figure 36: Geometrical principle of the MVSA approach [106].

The spectral unmixing of hyperspectral data is a complex problem, and each of the methods above can be efficient for this. When it comes to detection, it is important to evaluate the relevancy of these techniques with regard to the unmixing situation and the detection algorithm to be used.

In the case of MSD, it is important to obtain linearly independent components to build a projection matrix. In this purpose, PCA is more suitable than MCR-ALS because it assumes the spectral components are non colinear. On the other hand, the detection algorithm used with MCR-ALS (GMM with Malahanobis method) is more flexible. Since this method directly works on the concentration profiles, no assumptions are made on the spectral profiles. Moreover, the GMM enables to fit complex distribution in the feature space.

Choosing the appropriate threshold for detection is also an issue in the tuning phase. For the MSD, the metric for target detection is directly linked to a hypothesis test which is clearly defined. The threshold determination for outlier detection is fuzzier in the case of GMM: a high Mahalanobis distance may be due to an adulterated pixel or a standard pixel that follow the main variability of the chemical species and which is an extreme observation. As a consequence, this solution requires more interpretation and confidence in the design of the feature space, i.e., the determination of the spectral profiles.

The literature shows that ICA [107-108] or vertex-based [109] techniques are widely used for hyperspectral unmixing and detection. However, it is not so much used in food industry applications. Such methods combined with appropriate detection algorithms could be a way forward for hyperspectral imaging in the food industry.

Pixel unmixing and mathematical issues

In our work, the pixel unmixing phase is equivalent to the feature space design, which is finding the most interesting spectral components for detection. We assumed these spectral profiles should represent the signatures of the pure chemical in the system. This choice is justified for improving results interpretability. For instance, in the case of the MCR-ALS study (Figure 31) the fact that $\boldsymbol{c_3}$ is associated to the pure peanut spectral signature helps in the interpretation and detection process. However, the mathematical problem of finding the purest spectral component in a system is complex. Fundamentally, this may be considered as an ill-posed problem for two main reasons.

Solving the bilinear model implies that the number of spectral components, often referred to as the chemical rank [110], has been chosen beforehand. This choice is detrimental for techniques like MCR-ALS or ICA because the solutions will be different according to the chosen rank number, i.e. they are not nested method like PCA [111]. The link between the chemical rank and its mathematical meaning is not completely clear [110]. Consequently, the right number of chemical species in the system may not always be an appropriate guide.

In practice, it is complex to estimate all the independent chemical species that influence the NIR signal before analyzing it. As a result, the chemical rank is often deduced from a first analysis of the signal, and is not really associated to the number of pure existing compounds in the sample. Instead, it should be described as the number of different pure signal signatures that can be interpreted based on the pure sample analysis. The case of the chocolate powder is a good example. In the context of adulteration with peanut, this sample was considered to be a mixture of two main ingredients: cocoa and sucrose [82]. However, we know from the sample description that it contains many more ingredients like soybean lecithin (Table 5). Hence, it should be relevant to consider all the elements of this list. However, it turns out that only the sucrose and the cocoa can be distinguished using hyperspectral imaging. Hence only these independent spectral signatures were considered for the choice of the MCR-ALS rank. Other settings of the camera could lead to distinguish other ingredients by measuring smaller pixels (i.e. a smaller field of view). In that case, it could be relevant to consider additional components in the unmixing problem.

Table 4: List of ingredients of the chocolate powder sample used in [82].

Ingredient	Proportion
Saccharose	-
Powder cocoa	21,3 %
Dextrose	-
Soya lecithin	-
Salt	-
Cinnamon	-
Aroma	-

The ambiguity in spectral profiles is a major difficulty in pixel unmixing and detection for two main reasons.

First, because the true spectral signatures may not be known completely. Although the main pure ingredients could be identified for chocolate powder and peanut flour, their associated NIR signatures were still ambiguous. As the NIR spectra may change according to multiple factors, it may not be obvious to compare chocolate spectral signature from one study to another. As another example, studies about the spectral signature of crushed peanut can be found in the scientific literature. However, it was not possible to find any study about peanut flour. There is a need to have more interpretative studies of food powder NIR spectra in the literature. It could help to understand the type of NIR pattern that is expected in food powders despite their complexity.

Besides, finding the spectral components of a system may be tricky because of mathematical reasons. The unmixing process implies that the spectral signatures exhibit different shapes. For instance, techniques like PCA or ICA are looking for spectral profiles that are orthogonal and independent, respectively. This assumption

becomes an issue when two spectral components actually share some NIR patterns. The case of peanut in chocolate powder is a good example of such a situation. The NIR signatures of peanut and cocoa can be considered to be quite similar (see $\boldsymbol{s_1}$ and $\boldsymbol{s_3}$ on Figure 32). Hence, when considering a pixel containing mainly sucrose, it is not clear at all if the remaining substance is cocoa, peanut, or even a mixture of both. In this situation, the mathematical problem is ill-defined once again.

This issue of ambiguity has been addressed, mainly in the case of MCR-ALS, with the MCR-BANDS technique [112]. It gives the magnitude of the ambiguity and the MCR-BANDS which represents the boundaries of possible solutions for each component. This method could be an efficient tool to understand how the additional constraints reduce the ambiguity of the solutions. Indeed, the number and the types of the constraints that should be applied for a given problem is not straightforward to determine. If the problem is under-constrained, the MCR-BANDS may show that the system resolution is ambiguous. Hence, not so much confidence can be attributed to the spectral and concentration profiles obtained by the optimization process.

On the other hand, too many constraints may be applied to the system. The data matrix cannot be reconstructed by satisfying all the constraints and a sufficiently low lack of fit. This kind of situation may be quite common regarding complex unmixing situations. One perspective of this work could be to investigate in what extent the unmixing situation is over-constrained and the most suitable constraints to apply.

Finally, hyperspectral imaging offers spatial information about a sample. In our work, this information was never used as a reference to help the unmixing process. MCR-ALS was initially developed to handle a 2-dimensional matrix that does not hold any spatial information. However, it is possible to use an alternative approach to the MCR-ALS procedure called HSI-MCR-ALS [113]. This method consists of refolding the image structure of the concentration profiles for each component during the ALS. Once the spatial information is retrieved, image processing methods can be applied to the concentration profiles. Recent works have shown the efficiency of spatial constraints in MCR-ALS [114][115]. One perspective of our work could be to implement such constraints to improve the unmixing process and the detection. The challenge is that the adulterant particles do not follow a clear spatial pattern. The results of the MSD and MCR-ALS studies show the detected pixels produce clusters of different sizes. Hence, it could be challenging to assume a spectral continuity between neighbor pixels.

Dynamic context

The detection approach of this work is used in a static context. It means that the adulterant and standard samples are considered to be known. However, the products may change according to the seasonality (wheat flour for example) or the product formulation (chocolate powder). In a real industrial environment, these changes may affect the hyperspectral measurements which may cause a drift in the detection algorithms.

Some solutions are investigated in the literature. One main idea could be to learn the statistical behavior of the standard sample directly on the analyzed image. It implies there are potentially adulterated pixels in the image. Hence, this method only works with the assumption that the adulterant is rare in the sample. This method is used to implement the Adaptative Matched Subspace Detector (AMSD) which is the alternative version of the MSD used in the present work. More developed methods can be used to select random pixels, remove potential outliers and perform the standard sample design [52]. Another approach consists of the online estimation of the spectral and abundance profiles. Each time a new line of the hyperspectral image is measured, the model is updated using the information already collected together with the new line. This approach was used with NMF to implement online hyperspectral processing on wood samples [105].

B. The detection sensitivity

The importance of the pixel detection sensitivity

The notion of detection sensitivity is tackled in this thesis regarding the peanut detection in wheat flour. In this work, we stated that detecting a particle in a subpixel context requires the investigation of the detector sensitivity.

Since the particles to detect were assumed to be potentially smaller than the pixel field of view, no assumption could be done on their spatial pattern. Hence, the validation and the explanation of the detections could not be done using the detection map only. Each pixel detection should be treated independently in the first place, regardless of the image structure. This means that only spectral information could be used to validate the detected pixels. In this context, it is important to know the minimal amount of adulterant particles the detection system (hyperspectral camera and algorithm) can detect in standard pixels. That is why the sensitivity at the pixel scale is important to determine.

The detection results obtained for each pixel are usually extrapolated to all samples. For example, detecting wheat flour with a peanut adulteration of 1 % only depends on the detectability of each pixel. For this reason, the investigation of the detector sensitivity should be, once again, considered as a significant interest in the case of subpixel detection.

The lack of reference method

In practice, studying the detector sensitivity can be conducted as follows. Some pixels with known concentrations of target chemicals are selected. The detection algorithm is applied to these pixels. Then, the minimal concentration that can be detected with a high probability (99 % of the tested pixels with this concentration) can be used as the

reference sensitivity. Although this process appears trivial, there is a significant challenge: knowing the concentration of the target chemical for real pixels.

In remote sensing, such information can be obtained for earth observation images. The hyperspectral dataset HYDICE is partially manually annotated so that each pixel contains the percentage of four targets of interest (asphalt, roof, grass, and tree) [116]. Such a result could be obtained assuming the linear mixing model holds. However, this is a reasonable assumption only for macroscopic spectral mixture [106]. In case of lower scale interactions, like for food powders, the multiple scattering interactions makes the resulting signal much more challenging to calculate. Besides, being able to characterize the particles in a 3D volume raises many complex problems. It is not clear what particles and at which scale the particles should be considered. As for the unmixing problem, there is no clear guidance about how to choose the spectral components. Moreover, the deduction of the signal for one pixel would require to know the chemical nature and the position of a large number of particles in a small 3D volume. Thus, obtaining the reference values of a hyperspectral image of food products using the same approach than earth observation image was not considered in our work.

One solution could be to change the scale of the problem to a larger one. Instead of considering the real pixels obtained from the hyperspectral images, one may use macro-pixels (Figure 37). These are obtained by combining real pixels from the standard and the adulterant samples because their compositions are assumed to be pure and known. Once the macro-pixel is obtained, one may assume its signal can be obtained by averaging the original pixels' signals. This equivalent to considering the LMM is valid in this context. Therefore, pixels with known and varying target proportions are obtained by simulating a larger pixel field of view. Such a technique may look promising for testing detection algorithm. However, one major problem of this model is that the signal of the macro pixel is, by nature, less noisy than the signals coming from the original pixel. Mathematically, it happens because the noise is random and the averaging process decreases it. In practice, even if a pixel is larger and should provide less noise, the camera's electronic noise is still added to the measurement. Hence, such a method leads to smoother spectra than those obtained in the real pixels. This is a problem for the estimation of the detector sensitivity. Spectra with less noise exhibit less variability and are closer to average behavior. Hence, the detection of such pixels is likely to be much easier to perform. A simulation method that can take the pixel variability into account is a must-have.

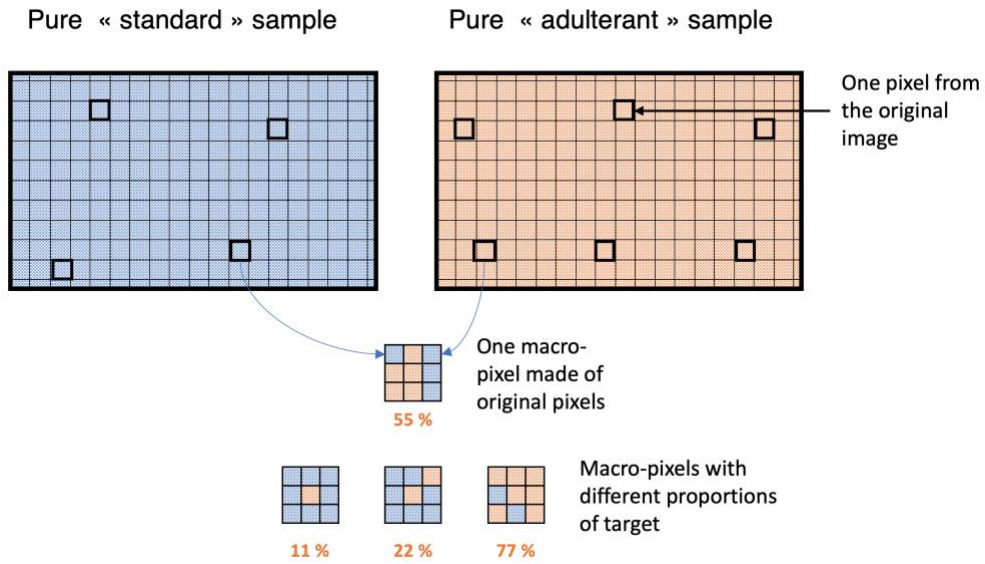

Figure 37: A method to generate macro-pixels with known concentrations from pixel of pure samples.

The simulation of hyperspectral data

The simulation method proposed in [67] shows its relevancy in the study of subpixel detection. One main significant result is that the simulated dataset enables to draw consistent conclusions compared to the observations on real data. We previously showed that a significant proportion of simulated spectra (> 95%) were detected by the algorithm as soon as the concentration of peanut is above 20%. Hence, it could be stated the detection system has a pixel-scale sensitivity of 20%.

To compare this result with some figures, we can model the pixel field of view as a square of size 250 µm. The NIR signal comes from particles under a layer of 1 mm [53]. Hence, the particles that contribute to the signal are in the volume described in Figure 38. This scheme helps to put the result in perspective and make several observations:

- Knowing that the signal of the underlying particles is weak, there is a big chance that if a median peanut particle is on the surface material, it will be detected. In this case, the particle represents approximately 20 % of the pixel surface.
- Lower concentrations of peanut in the pixel can only be obtained by considering underlying layers. In this case, the signal from peanut may be highly attenuated by the surface layer.
- The signal contribution of a pixel seems to be much more explained by the signal from multiple layers than from neighbor particles at the surface.

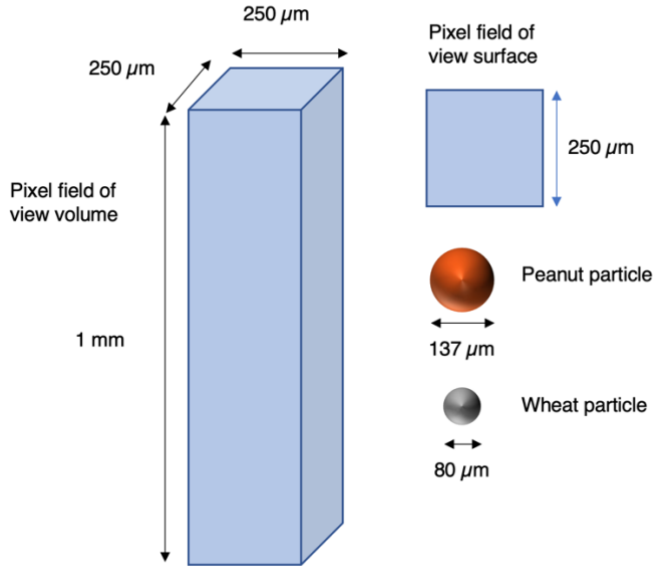

Figure 38: The pixel scale geometry compared to the peanut and wheat particles.

The last observation tends to show that the multiple scattering in powders may have a non-negligible effect on the pixel signal. According to the detection depth study results, at least ten layers of particles would be involved in the signal of a pixel. This is without considering the influence of neighbor pixels. This means that nonlinear models for spectral simulation should be also considered, as already suggested in the literature [117]. Nonlinear spectral simulation is still a challenge, and it should be a way for perspective.

Besides, despite it not being directly visible on spectral data, the real pixels from the MSD study exhibit a variability that derives from the Gaussian distribution. Figure 39 shows that the scores of real pixels on PC1 does not entirely overlap the Gaussian distribution of the simulated pixels obtained in the study. One may observe that the mode of the distribution of the real pixels is slightly shifted from the zero mean because of the extreme scores on the right-hand side of the axis. Hence, using more advanced simulation methods for fitting the real distribution of scores may improve the results for a better estimation of the detector sensitivity. For example, the distribution of each score could be estimated using a kernel distribution.

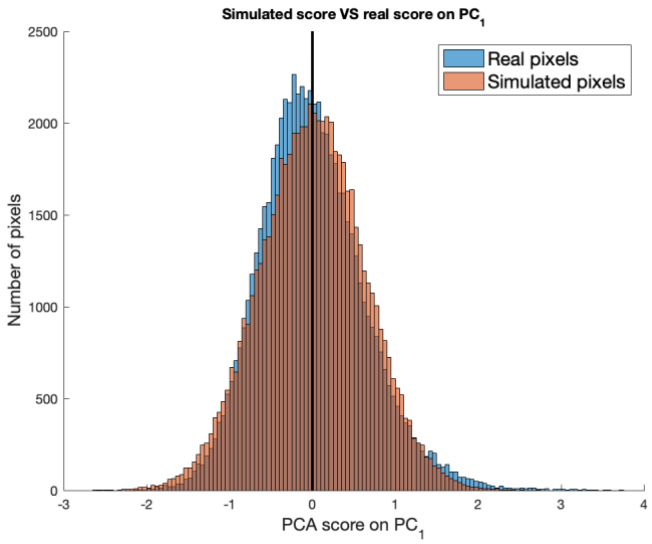

Figure 39: The distribution of simulated scores (PC 1) compared to the distribution of the scores of real pixels.

C. The particle detection in hyperspectral images

The particle agglomeration

The detection results shown on the maps (Figure 24 and Figure 34) are useful to prove the agglomeration of peanut particles in the sample mixtures. This phenomenon is also reported when studying the particle size of the pure samples (Figure 35). The presence of particles with a diameter between 600 µm and 1 mm for peanut flour can only be explained by their agglomeration. It is referred to as the particle stickiness problem [118]. For microscopic particles, a cohesion force exists and maintains them together even during the mixing. We show that the neighbor pixels that were detected as containing peanut flour were likely to show peanut particle clusters. Figure 34 also shows that for the high peanut concentrated samples, the detected clusters have larger pixel neighborhoods. This can be explained by the fact a larger quantity of peanut flour was introduced in the mixing. Hence, there is more chance that big clusters were incorporated. On the other hand, when smaller quantities were involved, big flour clusters had to be crushed to satisfy the mass of peanut introduced.

Thanks to these observations, the detection of pixel clusters was an argument for more credible results. Also, the inspection of the detector statistic shows the variation of signal intensity around a cluster of detected pixels (Figure 40).

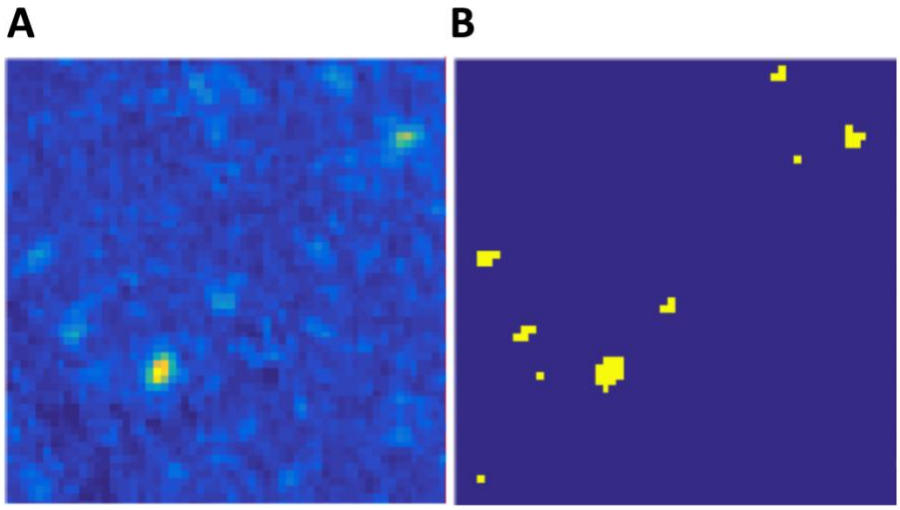

Figure 40: (A) the statistic of the MSD with the colormap highlighting the intensity of the detection statistic: high detection for yellow and low detection for blue. (B) is the corresponding detection map with detected pixels in yellow.

Besides the validation of the detection results, particle clustering could be used as an a priori knowledge of the problem. The principle is that there is more chance a peanut particle has been measured if several neighbor pixels show a high detection score. It could happen that this detection score is not sufficient so that each pixel individually is not detected by the algorithm. However, it may be relevant to consider that high scores of neighbor pixels should be combined to detect the area as potentially contaminated. This kind of method could be implemented using image processing morphology, as described in the HSI-MCR-ALS method [113].

Finally, it may be essential to consider the particle clustering effect due to the stickiness in the case of involuntary adulteration in food powder. Such contamination may happen because of particle sticking onto the equipment [36]. Following this hypothesis, one may assume that those particles stay together because of stickiness, even during the mixing process. Therefore, the detection of powder contamination could be simpler than being able to detect a single particle. Some investigations and empirical studies would be essential to understand how involuntary contamination are dispersed across food powder during the process.

The number of detected particles

One first idea to investigate the relevancy of detections in the mixed samples is to compare it to the theoretical global adulteration rate of the samples. We assume the number of detected pixels increases as the adulterant concentration increases, and that both are correlated. However, we showed that as the adulterant concentration increases, this relation is no longer linear (Figure 26). The simulation described in [67] explains how this non-linearity arises. This phenomenon happens because we consider a pixel entirely as adulterated as soon as it contains enough adulterant spectral

signature to be detected. Therefore, if the detector sensitivity is 20 %, all the pixels containing at least 20 % of adulterant will be detected. When the target concentration is low, its particles are sparsely distributed on the image. Hence, only a few detected pixels appear. As the adulterant concentration rises, there is more chance that a pixel contains some of its particles. It could be that underneath adulterant particles influence the pixel signal. As a result, many more pixels will be detected as being influenced by the adulterant chemical. However, since each pixel counts for one detection, the number of detected pixels increases much faster than the concentration.

Instead of counting the number of detections, we could count the predicted contribution of the adulterant in the pixel spectrum. The relationship between this contribution and the global target concentration should be more linear. However, this contribution is difficult to assess and we may end up with the same kind of difficulty. Indeed, in both methods used for peanut detection, the peanut contribution or the test statistic is never zero for non-contaminated pixels. This is visible for pure samples, and this is due to the ambiguity of the unmixing process. Hence, such a process would lead to an overestimated overall contribution of the adulterant.

The way the pixels are counted may lead to misleading results. If we assume a sample contains 10 % of peanut flour and the particles are homogenously distributed in the pixels, each pixel should contain a contribution of 10 % of peanut. Then, assuming the detection system has a 20 % sensitivity, it means that no pixels could be detected from the sample. It points out an essential assumption in our methodology. We actually assume the sample is not perfectly homogeneous so that we expect to detect several particles in one pixel to have the chance to detect it. The description of such a system does not hold in our case because the particles are not small enough compared to the pixels (Figure 38). This observation refers to the concept of the scale of scrutiny, which is very important in powder quality inspection in pharmaceuticals [119]. That is the sample size for which powder properties like homogeneity have to be ensured. In our case, we could say the scale of scrutiny is the pixel field of view. The property to be ensured is that the particle size is large enough to guarantee powder heterogeneity. In other words, we expect that if an adulterant particle is in the pixel, it has a chance to be detected. Either because its size is big enough or either because it is not the only one in the pixel field of view.

Detecting a global adulteration

In the food industry, the problem of food contamination could be stated differently than what was considered up to now. The question should be the following: assuming a given quantity of powder has contaminated a sample, how many pixels should be investigated to detect at least one particle?

We assume a binomial distribution can rule the content of a pixel field of view. A pixel contains n particles that are either standard particles or adulterant with a

probability P related to the global adulteration rate. Hence, the probability of finding at least one particle in the pixel is given by:

$$P_{\text{pixel}}\left(N_{particle\ in\ pixel} > 1\right) = 1 - (1 - P)^n$$

Equation 19: The probability of finding at least one particle in one pixel for a binomial distribution.

Assuming the system is able to detect a pixel as soon as it contains one particle, the probability that one pixel is detected in an entire image of N pixels is:

$$P_{\text{image}}\left(N_{detected\ pixel\ in\ image} > 1\right) = 1 - \left(1 - P_{\text{pixel}}\left(N_{particle\ in\ pixel} > 1\right)\right)^N$$

Equation 20: The probability of finding at least one particle one image of N pixels for a binomial distribution.

From this calculation, we can infer the number of pixels one should investigate to have a high probability $P_{\text{detection}}$ to detect at least one pixel in the image. We may assume this is a quantity of interest for the industry because as soon as an adulterant particle is detected, the sample should be considered for removal. Table 5 gives the result of such a calculation by setting $P_{\text{detection}} = 0.99$ and using the median particle size obtained in our experiment (137 µm for peanut and 80 µm for wheat). For instance, it shows that at least 77 pixels should be investigated to ensure one detection in a sample with 1 % of adulterant. In practice, a hyperspectral imaging system generally measures several hundreds of pixels for each acquisition band. Hence, it is possible to detect adulteration of 10 % to 0.1 % with a few hyperspectral measurements. Even if it is relevant regarding the detection results obtained in our experiment, these estimations should be improved. Many assumptions were made and could be refined.

Table 5: Minimal number of pixels to investigate to ensure at least one particle is detected (99 % of chance) for three different levels of adulterant concentration (10 %, 1 % and 0.1 %).

Particle type		Concentration of adulterant		
Standard	Adulterant	10 %	1 %	0.1 %
Wheat	*Peanut*	*8*	*77*	*764*
Chocolate	*Peanut*	*7*	*69*	*688*

First, the assumption that one particle could be detectable in one pixel field of view is difficult to generalize. In theory, if the particle is at the sample surface, the detectability could be assessed considering its size. However, if the particle is in the pixel depth, its detectability is much more difficult to assess. There is a need for more empirical study regarding the detection of underlying targets in food powders using NIR HSI.

Some assumptions of the statistical model may be improved as well. The Bernoulli process for choosing the particles in each pixel may be changed. The particle clustering could be taken into account by considering there is more chance to get the same particle type for succeeding random draws, i.e., the trials are not completely independent. Moreover, the number of particles in the pixel field of view should be updated according to the particle choice because they do not have the same volume. Finally, it is assumed that each pixel of the image is an independent experience. However, this is not the case because of particle clusters that are larger than a pixel surface. The improvement of this hypothesis will likely lead to less optimistic prediction than Table 5 shows.

5. Conclusion

The capability of near-infrared hyperspectral imaging supported with the chemometric method MCR-ALS to detect peanut flour in chocolate powder was demonstrated.

Detection of adulterated chocolate powder pixels with peanut flour could not be achieved with PCA due to the intrinsic complexity of the problem, as the spectra of chocolate powder and peanut flour are very similar and peanut adulteration occurred at the subpixel level. To cope with this situation, MCR-ALS was used instead. Specifically, we tested two different MCR-ALS methods: a method that incorporates a selectivity constraint and another that does not. The best results were obtained for the constrained MCR-ALS, highlighting that the detection of peanut adulteration in chocolate powder is a very challenging problem.

MCR-ALS results were used to build a metric for assessing peanut adulteration. With this metric, we were able to detect peanut adulteration in all the contaminated samples. On the other hand, a selectivity and sensitivity of 1 were obtained on the pure samples. Correlations of 0.92 and 0.93 were obtained between the number of adulterated pixels and the real concentration of mixed samples for MCR-ALS and MCR-ALS-CSEL respectively. This result supports the fact that the selectivity constraint was efficient to obtain spectral profiles closer to the real ones and to detect more adulterated pixels.

Due to its high performance, the method has the potential to be used for similar systems involving powder samples. Future work will be carried out to optimize the data acquisition and measurement, accounting by the sensitivity of the hyperspectral camera and the penetration depth of the near-infrared radiations. More advanced technique could also be used to optimize the MCR-ALS algorithm locally. The local rank analysis enables to estimate the complexity of a neighborhood of pixels to set the number of components [120]. Using this method, the MCR-ALS fitting may be improved locally.

Conclusion and future work

1. Conclusion

This thesis aims to study the detection of minority compounds in food powder using NIR HSI. It involves two scientific problems: how to determine the maximal depth from which the NIR signal comes from in a powder sample; and how to unmix the signal of a hyperspectral pixel for detecting minor compounds. These problems are applied to the case of contamination, which is characterized by the occurrence of foreign compounds in more or less high quantities.

The first part of the thesis focuses on the interactions of several layers of food powder and theirs contributions to the NIR signal. This subject is still a problem to solve theoretically and a few studies tackle the penetration depth of NIR radiations in an empirical way. An original approach is proposed using a sample holder made of PLA that mimics an underlying sample at different depths. We show that the NIR signals come from surface layers no deeper than 2 mm in powders like wheat flour.

The second part of the thesis deals with modelling the spectral variability within a pixel, which is essential when the particle size is smaller than the pixel field of view. The Linear Mixing Model is used to model the pixel signal, and detect minor compounds. A Matched Subspace Detector is designed to model the spectral variability and perform pixel detections. The unmixing and detection strategy requires the use of validation data, which are difficult to obtain in our context. In addition, the fact that food powders share some spectral pattern leads to an ambiguous problem for unmixing. A data simulation is used to overcome the first difficulty.

Then, the Multivariate Curve Resolution Alternating Least-Squares method is used to resolve the spectral ambiguity within the pixels. We proposed to use a selectivity constraint of the concentration matrix to reduce the ambiguity of the solution. The combination with an outlier detection strategy enables to improve the detection of peanut flour in chocolate powder.

The detection experiments of this thesis are performed using industrial food powders. The peanut flour is the sample to detect because it represents a harmful contaminant for allergic people. This thesis shows that ensuring the detection of such chemical in the food industry is a complex problem. Moreover, the quantity of allergen that can trigger a reaction may be very small. Such a constraint is not compatible with the detection restrictions implied by the technology. In addition, the food allergen detection using NIR suffers from a lack of definition of what is an allergen from the NIR point of view. As a conclusion, such a complex detection problem using NIR HSI must be addressed carefully.

2. Future works

In the first chapter, we used a sample holder as the target to be detected under layers of wheat flour. Thus, we studied the detection case of a plane surface of sample. Detecting smaller target spheres beneath controlled layers of wheat flour should be a case to investigate. First, it is a more realistic representation of particle detection within a powder. Then, comparing the results between this experience and the plane sample holder could provide insights regarding light scattering in complex media.

We studied the detection depth using diffuse reflectance lighting. The study of the optical light path can be carried out using spatially resolved spectroscopy (SRS). It enables to select specific longer or shorter light paths and measure them independently. This measurement technique is suitable with NIR HIS since each pixel can be considered as an independent sensor. We discussed the fact the detection depth is often limited by the surface signals in diffuse reflectance. The SRS enables to filter out these signals by measuring the reflectance at a given distance from the incident beam. Hence, an appropriate distance could provide better detection depth that the one obtained with diffuse reflectance. Such an experiment can be carried out using a white laser beam, a HSI and an appropriate sample holder.

In the second chapter, we studied the modelling of a pixel signal using linear models. As we discussed, such models are theoretically able to take the signal mixing in the sensor into account. However, the interactions between multiple layers provoke nonlinear effects in the signal. As a consequence, a more precise simulation method for spectral data could be based on nonlinear methods. Some multiplicative terms between the reflectance of each particle signal can be taken into account.

In this chapter, we developed the Matched Subspace Detector (MSD) using Principal Component Analysis. Other methods could be used to obtain the spectral subspaces. Independent Component Analysis (ICA) or Non-Negative Matrix Factorization (NMF) are two methods than could be compared to PCA. Another method could be to use multiplicative interactions of the compound's reflectance signatures to take their interactions into account in the detection method.

In the third chapter, Multivariate Curve Resolution Alternating Least-Squares (MCR-ALS) method was applied to unmix the pixel signal with a selectivity constraint. This method is used after the hyperspectral cube is unfolded and the MCR-ALS does not take spatial constraint as such into account. However, it is possible to reconstruct the cube structure in the ALS procedure to apply image filter and constraints. The combination of such a constraint could lead to a better spectral resolution and possibly to better detection performances.

Appendices

Appendix A: The Gaussian Mixture Model

The Gaussian Mixture Model probability density is given by [60]:

$$p(\boldsymbol{x}|\boldsymbol{\omega_i}, \boldsymbol{\mu_i}, \boldsymbol{\Sigma_i}) = \sum_{i=1}^{M} \boldsymbol{\omega}_i \; g(\boldsymbol{x}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma_i})$$

$$g(\boldsymbol{x}|\boldsymbol{\mu_i}, \boldsymbol{\Sigma}_i) = \frac{1}{(2\pi)^{m/2}|\boldsymbol{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu_i})^T \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i)\right)$$

Equation 21: The Mixture of Gaussian Model.

This model is a weighted sum – with the weights $\boldsymbol{\omega}_i$ – of Gaussian probability density functions that all have their parameters: the average $\boldsymbol{\mu}_i$ and the covariance matrices $\boldsymbol{\Sigma_i}$. The notation $|\boldsymbol{\Sigma}_i|$ refers to the matrix determinant. An example is the combination of two Gaussians to fit a bi-modal distribution (Figure 41).

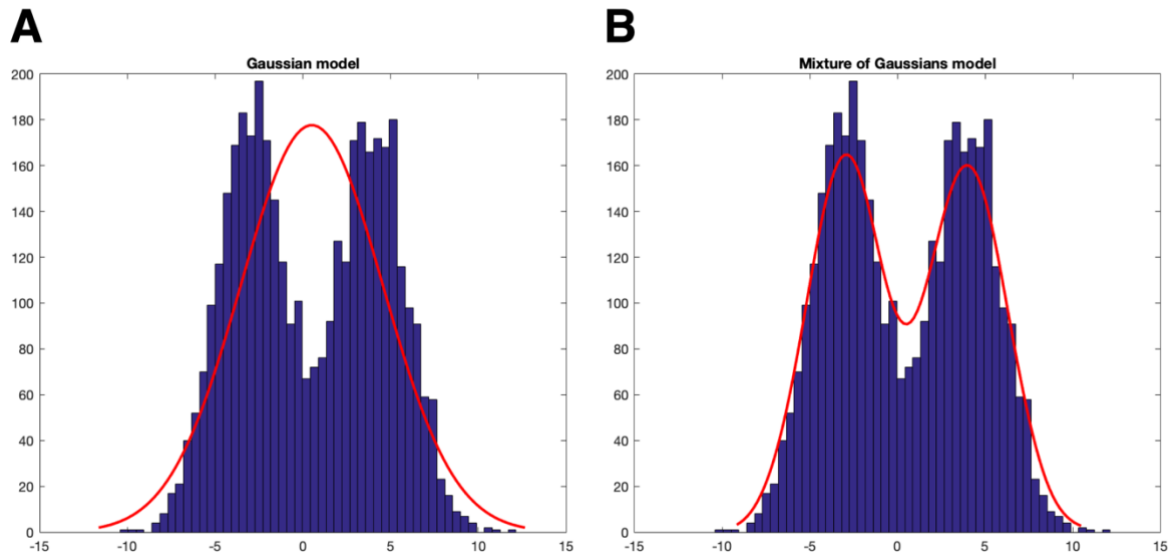

Figure 41: Example of a bi-modal distribution fitted with a single Gaussian model (A) and with GMM (B).

The GMM is obtained after estimating its parameters: $\boldsymbol{\omega}_i$, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma_i}$, the weights, the mean and, the covariance matrix for each Gaussian. This estimation is usually done using the maximum likelihood approach, which is maximizing the following quantity:

$$p(\mathbf{X}|\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{n} p(\boldsymbol{x}_i|\boldsymbol{\omega}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Equation 22: The quantity to maximize in the maximum likelihood approach.

In Equation 22, $\mathbf{X}$ is the generic notation for the matrix containing the training vectors that are assumed to be independent. In the case of peanut detection in chocolate powder, $\mathbf{X}$ is replaced by the concentration profile matrix $\mathbf{C}$ obtained by MCR-ALS. The quantity is commonly maximized using an iterative method called the Expectation-Maximization (EM) algorithm [60]. It aims to continuously increase the quantity Equation 22 by estimating the parameters at each step. For more details, the formula for the calculation of the parameters at each step are given in [60].

Appendix B: The Mahalanobis distance for outlier detection

The Mahalanobis distance is a measure of the distance between one observation and a distribution. It gives an estimation of how far the observation is from the mean of the distribution considering its variance [62]. The distance is calculated using Equation 23:

$$d_m{}^2 = (\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})$$

Equation 23 : The expression of the Mahalanobis distance.

$\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix of the distribution, and $\boldsymbol{x}$ is the observation. The Mahalanobis distance takes into account the variance of the distribution for the distance evaluation. Figure 42 shows that the Mahalanobis distance is more important for the outlier (red circle marker) than for extreme observations of the distribution (on the top-left of the left figure). This distance is equivalent to the Euclidian distance in the principal component space with the axis re-scaled to have unit variance. The right plot of Figure 42 shows it because the Mahalanobis distance has a circular geometry in the PC-space.

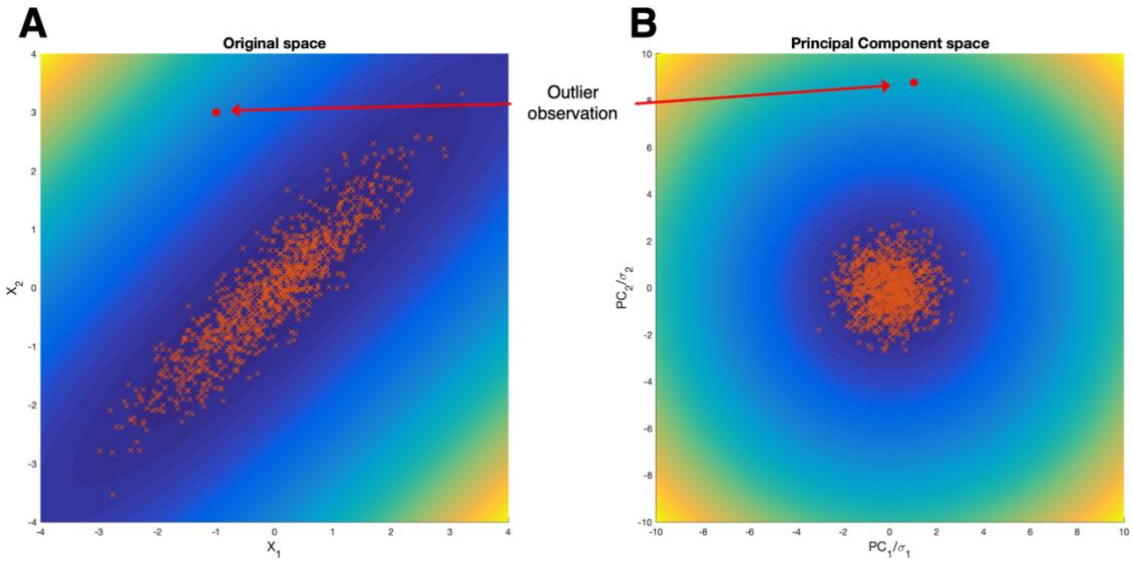

Figure 42: Randomly generated observations in the original space (A); and in the PC-space with the reduced by the PC standard deviation (B). The colormap, on both graphs, shows the Mahalanobis distance in the feature space.

References

[1] A.A. Gowen, C.P. O'Donnell, P.J. Cullen, G. Downey, J.M. Frias, Hyperspectral imaging - an emerging process analytical tool for food quality and safety control, Trends Food Sci. Technol. 18 (2007) 590–598. https://doi.org/10.1016/j.tifs.2007.06.001.

[2] P. Geladi, J. Burger, T. Lestander, Hyperspectral imaging : calibration problems and solutions, Chemom. Intell. Lab. Syst. 72 (2004) 209–217. https://doi.org/10.1016/j.chemolab.2004.01.023.

[3] R. Lu, Y.-R. Chen, Hyperspectral imaging for safety inspection of food and agricultural products, Pathog. Detect. Remediat. Safe Eat. 3544 (1999) 121–133. https://doi.org/10.1117/12.335771.

[4] Y. Feng, D. Sun, Application of Hyperspectral Imaging in Food Safety Inspection and Control : A Review, Crit. Rev. Food Sci. Nutr. 52 (2012) 1039–1058. https://doi.org/10.1080/10408398.2011.651542.

[5] B. Park, K.C. Lawrence, W.R. Windham, R.J. Buhr, Hyperspectral Imaging for Detecting Fecal and Ingesta Contaminants on Poultry Carcasses, Trans. Am. Soc. Agric. Eng. 45 (2002) 2017–2026. https://doi.org/10.13031/2013.11413.

[6] B. Park, K.C. Lawrence, W.R. Windham, D.P. Smith, Performance of hyperspectral imaging system for poultry surface fecal contaminant detection, J. Food Eng. 75 (2006) 340–348. https://doi.org/10.1016/j.jfoodeng.2005.03.060.

[7] S. Kang, K. Lee, J. Son, M.S. Kim, Detection of fecal contamination on leafy greens by hyperspectral imaging, Procedia Food Sci. 1 (2011) 953–959. https://doi.org/10.1016/j.profoo.2011.09.143.

[8] M.S. Kim, Y.-R. Chen, B.-K. Cho, K. Chao, C.-C. Yang, A.M. Lefcourt, D. Chan, Hyperspectral reflectance and fluorescence line-scan imaging for online defect and fecal contamination inspection of apples, Sens. Instrum. Food Qual. Saf. 1 (2007) 151–159. https://doi.org/10.1007/s11694-007-9017-x.

[9] K. Bhuvaneswari, P.G. Fields, N.D.G. White, A.K. Sarkar, C.B. Singh, D.S. Jayas, Image analysis for detecting insect fragments in semolina, J. Stored Prod. Res. 47 (2011) 20–24. https://doi.org/10.1016/j.jspr.2010.08.003.

[10] X. Fu, M.S. Kim, K. Chao, J. Qin, J. Lim, H. Lee, A. Garrido-Varo, D. Pérez-Marín, Y. Ying, Detection of melamine in milk powders based on NIR hyperspectral imaging and spectral similarity analyses, J. Food Eng. 124 (2014) 97–104. https://doi.org/10.1016/j.jfoodeng.2013.09.023.

[11] Y. Huang, K. Tian, S. Min, Y. Xiong, G. Du, Distribution assessment and quantification of counterfeit melamine in powdered milk by NIR imaging methods, Food Chem. 177 (2015) 174–181. https://doi.org/10.1016/j.foodchem.2015.01.029.

[12] M. Huang, M.S. Kim, S.R. Delwiche, K. Chao, J. Qin, C. Mo, C. Esquerre, Q. Zhu, Quantitative analysis of melamine in milk powders using near-infrared hyperspectral imaging and band ratio, J. Food Eng. 181 (2016) 10–19. https://doi.org/10.1016/j.jfoodeng.2016.02.017.

[13] J. Lim, G. Kim, C. Mo, M.S. Kim, K. Chao, J. Qin, X. Fu, I. Baek, B.K. Cho, Detection of melamine in milk powders using near-infrared hyperspectral imaging combined with regression coefficient of partial least square regression model, Talanta. 151 (2016) 183–191. https://doi.org/10.1016/j.talanta.2016.01.035.

[14] P. Mishra, A. Herrero-Langreo, P. Barreiro, J.M. Roger, Detection and quantification of

peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis, J. Near Infrared Spectrosc. 23 (2015) 15–22. https://doi.org/10.1255/jnirs.1142.
[15] P. Mishra, C.B.Y. Cordella, D.N. Rutledge, P. Barreiro, J.M. Roger, B. Diezma, Application of independent components analysis with the JADE algorithm and NIR hyperspectral imaging for revealing food adulteration, J. Food Eng. 168 (2016) 7–15. https://doi.org/10.1016/j.jfoodeng.2015.07.008.
[16] X. Zhao, W. Wang, X. Ni, X. Chu, Y.F. Li, C. Sun, Evaluation of near-infrared hyperspectral imaging for detection of peanut and walnut powders in whole wheat flour, Appl. Sci. 8 (2018). https://doi.org/10.3390/app8071076.
[17] J. Lim, A. Lee, J. Kang, Y. Seo, B. Kim, G. Kim, S. Kim, Non-destructive detection of bone fragments embedded in meat using hyperspectral reflectance imaging technique, Sensors (Switzerland). 20 (2020) 1–13. https://doi.org/10.3390/s20144038.
[18] Y. Zhang, X. Wang, J. Shan, J. Zhao, W. Zhang, L. Liu, F. Wu, Hyperspectral Imaging Based Method for Rapid Detection of Microplastics in the Intestinal Tracts of Fish, Environ. Sci. Technol. 53 (2019) 5151–5158. https://doi.org/10.1021/acs.est.8b07321.
[19] A.A. Gowen, M. Taghizadeh, C.P. O'Donnell, Identification of mushrooms subjected to freeze damage using hyperspectral imaging, J. Food Eng. 93 (2009) 7–12. https://doi.org/10.1016/j.jfoodeng.2008.12.021.
[20] J. Li, L. Chen, W. Huang, Q. Wang, B. Zhang, X. Tian, S. Fan, B. Li, Multispectral detection of skin defects of bi-colored peaches based on vis-NIR hyperspectral imaging, Postharvest Biol. Technol. 112 (2016) 121–133. https://doi.org/10.1016/j.postharvbio.2015.10.007.
[21] C.B. Singh, D.S. Jayas, J. Paliwal, N.D.G. White, Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging, J. Stored Prod. Res. 45 (2009) 151–158. https://doi.org/10.1016/j.jspr.2008.12.002.
[22] R. Grau, A.J. Sánchez, J. Girón, E. Iborra, A. Fuentes, J.M. Barat, Nondestructive assessment of freshness in packaged sliced chicken breasts using SW-NIR spectroscopy, Food Res. Int. 44 (2011) 331–337. https://doi.org/10.1016/j.foodres.2010.10.011.
[23] J. Jin, L. Tang, Z. Hruska, H. Yao, Classification of toxigenic and atoxigenic strains of Aspergillus flavus with hyperspectral imaging, Comput. Electron. Agric. 69 (2009) 158–164. https://doi.org/10.1016/j.compag.2009.07.023.
[24] A. Del Fiore, M. Reverberi, A. Ricelli, F. Pinzari, S. Serranti, A.A. Fabbri, G. Bonifazi, C. Fanelli, Early detection of toxigenic fungi on maize by hyperspectral imaging analysis, Int. J. Food Microbiol. 144 (2010) 64–71. https://doi.org/10.1016/j.ijfoodmicro.2010.08.001.
[25] P. Williams, M. Manley, G. Fox, P. Geladi, Indirect detection of Fusarium verticillioides in maize (Zea mays L.) kernels by near infrared hyperspectral imaging, J. Near Infrared Spectrosc. 18 (2010) 49–58. https://doi.org/10.1255/jnirs.858.
[26] D.A. Burns, E.W. Ciurczak, Handbook of Near-Infrared Analysis, 2001.
[27] S. Chandrasekhar, Radiative Transfer, New York, 1960.
[28] A. Schuster, No Radiation Through a Foggy Atmosphere, Astrophys. J. 21 (1905) 1.
[29] P. Kubelka, F. Munk, Ein Beitrag zur Optik der Farbanstriche, Zeitschrift Für Tech. Phys. 12 (1931) 593–601.
[30] O. Berntsson, T. Burger, S. Folestad, L.G. Danielsson, J. Kuhn, J. Fricke, Effective sample size in diffuse reflectance near-IR spectrometry, Anal. Chem. 71 (1999) 617–623. https://doi.org/10.1021/ac980652u.

[31] J. Kuhn, S. Korder, M.C. Arduinischuster, R. Caps, J. Fricke, Infraredoptical transmission and reflection measurements on loose powders, Rev. Sci. Instrum. 64 (1993) 2523–2530. https://doi.org/10.1063/1.1143914.
[32] S. Stolik, J.A. Delgado, A. Pérez, L. Anasagasti, Measurement of the penetration depths of red and near infrared light in human ”ex vivo” tissues, J. Photochem. Photobiol. B Biol. 57 (2000) 90–93. https://doi.org/10.1016/S1011-1344(00)00082-8.
[33] A. Ciani, K.U. Goss, R.P. Schwarzenbach, Light penetration in soil and particulate minerals, Eur. J. Soil Sci. 56 (2005) 561–574. https://doi.org/10.1111/j.1365-2389.2005.00688.x.
[34] J. Lammertyn, A. Peirs, J. De Baerdemaeker, B. Nicolaï, Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment, Postharvest Biol. Technol. 18 (2000) 121–132. https://doi.org/10.1016/S0925-5214(99)00071-X.
[35] M. Huang, M.S. Kim, K. Chao, J. Qin, C. Mo, C. Esquerre, S. Delwiche, Q. Zhu, Penetration depth measurement of near-infrared hyperspectral imaging light for milk powder, Sensors (Switzerland). 16 (2016) 1–11. https://doi.org/10.3390/s16040441.
[36] J.J. Fitzpatrick, L. Ahrné, Food powder handling and processing : Industry problems , knowledge barriers and research opportunities, Chem. Eng. Process. 44 (2005) 209–214. https://doi.org/10.1016/j.cep.2004.03.014.
[37] Joint FAO/WHO Codex Alimentarius Commission, Codex Standard 152–1985 “Wheat Flour.,” in: Codex Aliment., 1995.
[38] E.A. Ashton, A. Schaum, Algorithms for the detection of sub-pixel targets in multispectral imagery, Photogramm. Eng. Remote Sensing. 64 (1998) 723–731.
[39] M.A. Kolodner, An automated target detection system for hyperspectral imaging sensors, Johns Hopkins APL Tech. Dig. (Applied Phys. Lab. 27 (2007) 208–217.
[40] F. Zhu, Y. Wang, S. Xiang, B. Fan, C. Pan, Structured Sparse Method for Hyperspectral Unmixing, ISPRS J. Photogramm. Remote Sens. 88 (2014) 101–118.
[41] P. Vermeulen, J.A. Fernández Pierna, H.P. van Egmond, P. Dardenne, V. Baeten, Online detection and quantification of ergot bodies in cereals using near infrared hyperspectral imaging, Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess. 29 (2012) 232–240. https://doi.org/10.1080/19440049.2011.627573.
[42] D. Manolakis, G. Shaw, Detection Algorithms for Hyperspectral Imaging Applications, Signal Process. Mag. IEEE. 19 (2002) 29–43. http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=974724&url=http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=974724.
[43] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods. 6 (2014) 4964–4976. https://doi.org/10.1039/c4ay00571f.
[44] C. Ruckebusch, L. Blanchet, Multivariate curve resolution: A review of advanced and tailored applications and challenges, Anal. Chim. Acta. 765 (2013) 28–36. https://doi.org/10.1016/j.aca.2012.12.028.
[45] R. Tauler, Multivariate curve resolution applied to second order data, Chemom. Intell. Lab. Syst. 30 (1995) 133–146. https://doi.org/10.1016/0169-7439(95)00047-X.
[46] N. Omidikia, S. Beyramysoltan, J. Mohammad Jafari, E. Tavakkoli, M. Akbari Lakeh, M. Alinaghi, M. Ghaffari, S. Khodadadi Karimvand, R. Rajkó, H. Abdollahi, Closure constraint in multivariate curve resolution, J. Chemom. 32 (2018) 1–15. https://doi.org/10.1002/cem.2975.
[47] A.C. Olivieri, R. Tauler, The effect of data matrix augmentation and constraints in

extended multivariate curve resolution–alternating least squares, J. Chemom. 31 (2017) 1–10. https://doi.org/10.1002/cem.2875.
[48] R. Tauler, A. Smilde, B. Kowalski, Selectivity, local rank, three-way data analysis and ambiguity in multivariate curve resolution, J. Chemom. 9 (1995) 31–58. https://doi.org/10.1002/cem.1180090105.
[49] W.F. Cordeiro Dantas, J.C. Laurentino Alves, R.J. Poppi, MCR-ALS with correlation constraint and Raman spectroscopy for identification and quantification of biofuels and adulterants in petroleum diesel, Chemom. Intell. Lab. Syst. 169 (2017) 116–121. https://doi.org/10.1016/j.chemolab.2017.04.002.
[50] M. Boiret, A. de Juan, N. Gorretta, Y.M. Ginot, J.M. Roger, Setting local rank constraints by orthogonal projections for image resolution analysis: Application to the determination of a low dose pharmaceutical compound, Anal. Chim. Acta. 892 (2015) 49–58. https://doi.org/10.1016/j.aca.2015.08.031.
[51] D.G. Manolakis, G. Shaw, Detection algorithms for hyperspectral imaging applications, IEEE Signal Process. Mag. 19 (2002) 29–43. https://doi.org/10.1109/79.974724.
[52] B. Du, L. Zhang, Target detection based on a dynamic subspace, Pattern Recognit. 47 (2014) 344–358. https://doi.org/10.1016/j.patcog.2013.07.005.
[53] A. Laborde, B. Jaillais, R. Bendoula, J.M. Roger, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, D. Bertrand, A. Boulanger, C.B.Y. Cordella, A partial least squares-based approach to assess the light penetration depth in wheat flour by near infrared hyperspectral imaging, J. Near Infrared Spectrosc. (2019). https://doi.org/10.1177/0967033519891594.
[54] J.M. Olinger, P.R. Griffiths, T. Burger, Theory of Diffuse Reflection in the NIR Region, in: D.A. Burns, E.W. Ciurczak (Eds.), Handb. Near-Infrared Anal., Second Edi, Marcel Dekker, New York, 2001: pp. 19–51.
[55] N. Harnby, Characterization of powder mixtures, in: N. Harnby, M.F. Edwards, A.W. Nienow (Eds.), Mix. Process Ind., Second Edi, Butterworth Heinemann, Oxford, 1992: pp. 25–41.
[56] J.A. Fernández Pierna, D. Vincke, P. Dardenne, Z. Yang, L. Han, V. Baeten, Line scan hyperspectral imaging spectroscopyfor the early detection of melamineand cyanuric acid in feed, J. Near Infrared Spectrosc. 22 (2014) 103–112. https://doi.org/10.1255/jnirs.1109.
[57] P. Kubelka, New Contributions to the Optics of Intensity Light-Scattering Materials. Part I, J. Opt. Soc. Am. 38 (1948) 448–457.
[58] P. Kubelka, New Contributions to the Optics of Intensity Light-Scattering Materials. Part II: Nonhomogeneous Layers, J. Opt. Soc. Am. 44 (1954) 330–335.
[59] J.M. Olinger, P.R. Griffiths, Quantitative Effects of an Absorbing Matrix on Near-Infrared Diffuse Reflectance Spectra, Anal. Chem. 60 (1988) 2427–2435. https://doi.org/10.1021/ac00172a022.
[60] M.V. Padalkar, N. Pleshko, Wavelength-Dependent Penetration Depth of Near Infrared Radiation into Cartilage, HHS Public Access. 140 (2016) 2093–2100.
[61] T. Naes, C. Irgens, H. Martens, Comparison of Linear Statistical Methods for Calibration of NIR Instruments, Appl. Stat. 35 (1986) 195. https://doi.org/10.2307/2347270.
[62] P. Emmel, Nouvelle Formulation du Modèle de Kubelka et Munk avec Applications aux Encres Fluorescentes, Actes l'Ecole Printemps 2000, Le Pays d'Apt En Couleurs, Apt Roussillon, Fr. (2000) 87–96.
[63] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, J. Chanussot, Noise reduction in

hyperspectral imagery: Overview and application, Remote Sens. 10 (2018) 1–28. https://doi.org/10.3390/rs10030482.

[64] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 5 (2012) 354–379. https://doi.org/10.1109/JSTARS.2012.2194696.

[65] C.C. Borel, S.A.W. Gerstl, Nonlinear spectral mixing models for vegetative and soil surfaces, Remote Sens. Environ. 47 (1994) 403–416. https://doi.org/10.1016/0034-4257(94)90107-4.

[66] B. Somers, K. Cools, S. Delalieux, J. Stuckens, D. Van der Zande, W.W. Verstraeten, P. Coppin, Nonlinear Hyperspectral Mixture Analysis for tree cover estimates in orchards, Remote Sens. Environ. 113 (2009) 1183–1193. https://doi.org/10.1016/j.rse.2009.02.003.

[67] A. Laborde, B. Jaillais, J.M. Roger, M. Metz, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, C. Cordella, Subpixel detection of peanut in wheat flour using a matched subspace detector algorithm and near-infrared hyperspectral imaging, Talanta. 216 (2020) 120993. https://doi.org/10.1016/j.talanta.2020.120993.

[68] E.W. Lusas, Food uses of peanut protein, J. Am. Oil Chem. Soc. 56 (1979) 425–430. https://doi.org/10.1007/BF02671530.

[69] M. Wensing, A.H. Penninks, S.L. Hefle, S.J. Koppelman, C.A.F.M. Bruijnzeel-Koomen, A.C. Knulst, The distribution of individual threshold doses eliciting allergic reactions in a population with peanut allergy, J. Allergy Clin. Immunol. 110 (2002) 915–920. https://doi.org/10.1067/mai.2002.129235.

[70] A.E. Flinterman, S.G. Pasmans, M.O. Hoekstra, Y. Meijer, E. Van Hoffen, E.F. Knol, S.L. Hefle, C.A. Bruijnzeel-Koomen, A.C. Knulst, Determination of no-observed-adverse-effect levels and eliciting doses in a representative group of peanut-sensitized children, J. Allergy Clin. Immunol. 117 (2006) 448–454. https://doi.org/10.1016/j.jaci.2005.11.035.

[71] G. Elmasry, D. Sun, P. Allen, Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging, J. Food Eng. 117 (2013) 235–246. https://doi.org/10.1016/j.jfoodeng.2013.02.016.

[72] M. Kamruzzaman, Y. Makino, Assessment of Visible Near-Infrared Hyperspectral Imaging as a Tool for Detection of Horsemeat Adulteration in Minced Beef, Food Bioprocess Technol. 8 (2015) 1054–1062. https://doi.org/10.1007/s11947-015-1470-7.

[73] M. Kamruzzaman, Y. Makino, S. Oshita, Rapid and non-destructive detection of chicken adulteration in minced beef using visible near-infrared hyperspectral imaging and machine learning, J. Food Eng. 170 (2016) 8–15. https://doi.org/10.1016/j.jfoodeng.2015.08.023.

[74] W.H. Su, D.W. Sun, Fourier Transform Infrared and Raman and Hyperspectral Imaging Techniques for Quality Determinations of Powdery Foods: A Review, Compr. Rev. Food Sci. Food Saf. 17 (2018) 104–122. https://doi.org/10.1111/1541-4337.12314.

[75] D.I. Ellis, V.L. Brewster, W.B. Dunn, J.W. Allwood, A.P. Golovanov, D.I. Ellis, V.L. Brewster, W.B. Dunn, J.W. Allwood, A.P. Golovanov, R. Goodacre, Fingerprinting food: current technologies for the detection of food adulteration and contamination, Chem. Soc. Rev. 41 (2012) 5706–5727. https://doi.org/10.1039/c2cs35138b.

[76] L. Manning, J. Soon, Developing systems to control food adulteration, J. Food Policy. 49 (2014) 23–32. https://doi.org/10.1016/j.foodpol.2014.06.005.

[77] P. Vermeulen, M.B. Ebene, B. Orlando, J.A. Fernández Pierna, V. Baeten, Online detection and quantification of particles of ergot bodies in cereal flour using near-infrared hyperspectral imaging, Food Addit. Contam. - Part A Chem. Anal. Control. Expo. Risk Assess. 34 (2017) 1312–1319. https://doi.org/10.1080/19440049.2017.1336798.
[78] J.A. Fernández Pierna, D. Vincke, V. Baeten, C. Grelet, F. Dehareng, P. Dardenne, Use of a multivariate moving window PCA for the untargeted detection of contaminants in agro-food products, as exemplified by the detection of melamine levels in milk using vibrational spectroscopy, Chemom. Intell. Lab. Syst. 152 (2016) 157–162. https://doi.org/10.1016/j.chemolab.2015.10.016.
[79] S. Verdú, F. Vásquez, R. Grau, E. Ivorra, A.J. Sánchez, J.M. Barat, 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 (2016) 373–380. https://doi.org/10.1016/j.foodcont.2015.11.002.
[80] L.L. Scharf, B. Friedlander, Matched Subspace Detectors, IEEE Trans. Signal Process. 42 (1994) 2146–2157. https://doi.org/https://doi.org/10.1109/78.301849.
[81] M. Metz, A. Biancolillo, M. Lesnoff, J. Roger, A note on spectral data simulation, Chemom. Intell. Lab. Syst. 200 (2020) 103979. https://doi.org/10.1016/j.chemolab.2020.103979.
[82] A. Laborde, F. Puig-Castellví, D. Jouan-Rimbaud Bouveresse, L. Eveleigh, C. Cordella, B. Jaillais, Detection of chocolate powder adulteration with peanut using near-infrared hyperspectral imaging and Multivariate Curve Resolution, Food Control. 119 (2021) 107454. https://doi.org/10.1016/j.foodcont.2020.107454.
[83] D.A. Moneret-Vautrin, G. Kanny, Update on threshold doses of food allergens: Implications for patients and the food industry, Curr. Opin. Allergy Clin. Immunol. 4 (2004) 215–219. https://doi.org/10.1097/00130832-200406000-00014.
[84] A.J. van Hengel, Food allergen detection methods and the challenge to protect food-allergic consumers, Anal. Bioanal. Chem. 389 (2007) 111–118. https://doi.org/10.1007/s00216-007-1353-5.
[85] S.L. Hefle, S.L. Taylor, Food allergy and the food industry, Curr. Allergy Asthma Rep. 4 (2004) 55–59. https://doi.org/10.1007/s11882-004-0044-y.
[86] A.S. Chang, A. Sreedharan, K.R. Schneider, Peanut and peanut products: A food safety perspective, Food Control. 32 (2013) 296–303. https://doi.org/10.1016/j.foodcont.2012.12.007.
[87] S. Lohumi, S. Lee, H. Lee, B.K. Cho, A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration, Trends Food Sci. Technol. 46 (2015) 85–98. https://doi.org/10.1016/j.tifs.2015.08.003.
[88] M. Boiret, A. de Juan, N. Gorretta, Y.M. Ginot, J.M. Roger, Distribution of a low dose compound within pharmaceutical tablet by using multivariate curve resolution on Raman hyperspectral images, J. Pharm. Biomed. Anal. 103 (2015) 35–43. https://doi.org/10.1016/j.jpba.2014.10.024.
[89] E. Lancelot, D. Bertrand, M. Hanafi, B. Jaillais, Near-Infrared hyperspectral imaging for following imbibition imbibition of single wheat kernel sections, Vib. Spectrosc. 92 (2017) 46–53. https://doi.org/10.1016/j.vibspec.2017.05.001.
[90] INRA, ANSES, Étude des allergènes dans les produits transformés disponibles sur le marché français entre 2008 et 2012, 2015.
[91] H. Jin, Y. Ma, L. Li, J.H. Cheng, Rapid and Non-destructive Determination of Oil Content of Peanut (Arachis hypogaea L.) Using Hyperspectral Imaging Analysis, Food Anal.

Methods. 9 (2016) 2060–2067. https://doi.org/10.1007/s12161-015-0384-3.
[92] P.A. da Costa Filho, Rapid determination of sucrose in chocolate mass using near infrared spectroscopy, Anal. Chim. Acta. 631 (2009) 206–211. https://doi.org/10.1016/j.aca.2008.10.049.
[93] V. Olmos, L. Benítez, M. Marro, P. Loza-Alvarez, B. Piña, R. Tauler, A. de Juan, Relevant aspects of unmixing/resolution analysis for the interpretation of biological vibrational hyperspectral images, TrAC - Trends Anal. Chem. 94 (2017) 130–140. https://doi.org/10.1016/j.trac.2017.07.004.
[94] A. de Juan, Multivariate curve resolution for hyperspectral image analysis, 2020. https://doi.org/10.1016/B978-0-444-63977-6.00007-9.
[95] R. Tauler, B.R. Kowalski, MCR applied to spectral data from multiple runs of an industrial process, Anal. Chem. 65 (1993) 2040–2047. https://pubs.acs.org/sharingguidelines.
[96] G.H. Golub, C. Reinsch, Comparacion De Inhibicion De La Prueba Cutanea De La Histamina Con Astemizole, Loratadina Y Terfenadina, Numer. Math. 14 (1970) 403–420.
[97] A. de Juan, J. Jaumot, R. Tauler, Multivariate Curve Resolution (MCR). Solving the mixture analysis problem, Anal. Methods. 6 (2014) 4964–4976. https://doi.org/10.1039/c4ay00571f.
[98] W.H. Lawton, E.A. Sylvestre, Self Modeling Curve Resolution, Technometrics. 13 (1971) 617–633. https://doi.org/10.1002/0471667196.ess0035.pub2.
[99] M. Yang, N. Ahuja, Gaussian Mixture Model for Human Skin Color and Its Applications in Image and Video Databases_Image, Proc. Vol. 3656, Storage Retr. Image Video Databases VII. 3656 (1998) 458–466.
[100] C.B.Y. Cordella, D. Bertrand, SAISIR: A new general chemometric toolbox, TrAC - Trends Anal. Chem. 54 (2014) 75–82. https://doi.org/10.1016/j.trac.2013.10.009.
[101] J. Jaumot, A. de Juan, R. Tauler, MCR-ALS GUI 2.0: New features and applications, Chemom. Intell. Lab. Syst. 140 (2015) 1–12. https://doi.org/10.1016/j.chemolab.2014.10.003.
[102] J. Workman Jr., L. Weyer, Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy, CRC Press, CRC Press, Boca Raton, 2012.
[103] X. Zhang, R. Tauler, Measuring and comparing the resolution performance and the extent of rotation ambiguities of some bilinear modeling methods, Chemom. Intell. Lab. Syst. 147 (2015) 47–57. https://doi.org/10.1016/j.chemolab.2015.08.005.
[104] D.N. Rutledge, D. Jouan-Rimbaud Bouveresse, Components Analysis with the JADE algorithm, Trends Anal. Chem. 50 (2013) 22–32.
[105] B. Jaillais, K. Meghar, L. Nus, S. Miron, D. Brie, K. Meghar, L. Nus, S. Miron, D. Brie, Unsupervised processing of hyperspectral images, in: Chim. XIX, Paris, France, 2018.
[106] J. Li, A. Agathos, D. Zaharie, J.M. Bioucas-Dias, A. Plaza, X. Li, Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing, IEEE Trans. Geosci. Remote Sens. 53 (2015) 5067–5082. https://doi.org/10.1109/TGRS.2015.2417162.
[107] S. Jin, B. Wang, W. Xia, Target detection in hyperspectral imagery based on independent component analysis with references, Hongwai Yu Haomibo Xuebao/Journal Infrared Millim. Waves. 34 (2015) 177–183. https://doi.org/10.11972/j.issn.1001-9014.2015.02.010.
[108] R.J. Johnson, J.P. Williams, K.W. Bauer, AutoGAD: An improved ICA-based hyperspectral anomaly detection algorithm, IEEE Trans. Geosci. Remote Sens. 51 (2013) 3492–3503. https://doi.org/10.1109/TGRS.2012.2222418.
[109] J. Plaza, E.M.T. Hendrix, I. García, G. Martín, A. Plaza, On endmember identification in

hyperspectral images without pure pixels: A comparison of algorithms, J. Math. Imaging Vis. 42 (2012) 163–175. https://doi.org/10.1007/s10851-011-0276-0.
[110] C. Ruckebusch, A. De Juan, L. Duponchel, J.P. Huvenne, Matrix augmentation for breaking rank-deficiency: A case study, Chemom. Intell. Lab. Syst. 80 (2006) 209–214. https://doi.org/10.1016/j.chemolab.2005.06.009.
[111] D. Jouan-Rimbaud Bouveresse, A. Moya-González, F. Ammari, D.N. Rutledge, Two novel methods for the determination of the number of components in independent components analysis models, Chemom. Intell. Lab. Syst. 112 (2012) 24–32. https://doi.org/10.1016/j.chemolab.2011.12.005.
[112] J. Jaumot, R. Tauler, MCR-BANDS: A user friendly MATLAB program for the evaluation of rotation ambiguities in Multivariate Curve Resolution, Chemom. Intell. Lab. Syst. 103 (2010) 96–107. https://doi.org/10.1016/j.chemolab.2010.05.020.
[113] S. Hugelier, O. Devos, C. Ruckebusch, On the implementation of spatial constraints in multivariate curve resolution alternating least squares for hyperspectral image analysis, J. Chemom. 29 (2015) 557–561. https://doi.org/10.1002/cem.2742.
[114] M. Ghaffari, S. Hugelier, L. Duponchel, H. Abdollahi, C. Ruckebusch, Effect of image processing constraints on the extent of rotational ambiguity in MCR-ALS of hyperspectral images, Anal. Chim. Acta. 1052 (2019) 27–36. https://doi.org/10.1016/j.aca.2018.11.054.
[115] P. Firmani, S. Hugelier, F. Marini, C. Ruckebusch, MCR-ALS of hyperspectral images with spatio-spectral fuzzy clustering constraint, Chemom. Intell. Lab. Syst. 179 (2018) 85–91. https://doi.org/10.1016/j.chemolab.2018.06.007.
[116] S. Jia, Y. Qian, Spectral and spatial complexity-based hyperspectral unmixing, IEEE Trans. Geosci. Remote Sens. 45 (2007) 3867–3879. https://doi.org/10.1109/TGRS.2007.898443.
[117] J.M. Rodríguez Alves, J.M.P. Nascimento, J.M. Bioucas-Dias, A. Plaza, V. Silva, Sparse unmixing of hyperspectral data, IEEE Trans. Geosci. Remote Sens. 49 (2011) 2014–2039. https://doi.org/10.1109/IGARSS.2013.6723057.
[118] P. Boonyai, B. Bhandari, T. Howes, Stickiness measurement techniques for food powders: A review, Powder Technol. 145 (2004) 34–46. https://doi.org/10.1016/j.powtec.2004.04.039.
[119] H.J. Venables, J.I. Wells, Powder mixing, Drug Dev. Ind. Pharm. 27 (2001) 599–612. https://doi.org/10.1081/DDC-100107316.
[120] A. De Juan, M. Maeder, T. Hancewicz, R. Tauler, Local rank analysis for exploratory spectroscopic image analysis. Fixed Size Image Window-Evolving Factor Analysis, Chemom. Intell. Lab. Syst. 77 (2005) 64–74. https://doi.org/10.1016/j.chemolab.2004.11.006.

Figure index

Equation index

Table index

Résumé de la thèse par chapitre

Introduction

Ce projet de thèse est issu d'une collaboration entre l'AgroParisTech et l'ONIRIS dans le but de développer des solutions de contrôle de procédés de fabrication dans l'industrie agroalimentaire. En particulier, il s'agit de contrôler les processus industriels liés aux poudres telles que les farines à l'aide de l'imagerie hyperspectrale proche infrarouge. Cette technique permet, entre autres, d'analyser les poudres agroalimentaires et d'en détecter certains contaminants. Cependant, nous avons identifié que cette détection soulève deux problèmes techniques majeurs. Tout d'abord, le champ de vision des pixels est plus grand que la taille de la plupart des particules de poudre. Lorsque des composés de natures chimiques différentes sont mesurés au sein d'un même pixel, le spectre proche infrarouge est un mélange de signaux dits purs qu'il faut démélanger. Ensuite, les radiations proches infrarouges ne sont capables de pénétrer qu'une épaisseur limitée de matériau à cause du phénomène d'absorption. Ce phénomène limite la capacité de détection en profondeur d'un système proche infrarouge. Pour cette raison, il est important de pouvoir quantifier la profondeur de détection.

Dans cette thèse, nous proposons d'adresser les deux questions scientifiques qui en découlent : comment modéliser le mélange de signaux dans les pixels et détecter la présence de composés minoritaires ; comment déterminer la profondeur de détection d'un système d'imagerie proche infrarouge.

Les poudres sont largement utilisées dans l'industrie agroalimentaire et ceci pour de multiples raisons. En particulier, elles permettent un usage et un transport simplifiés, mais aussi une meilleure conservation de l'ingrédient grâce au retrait de l'eau. Ces poudres sont utilisées dans le cadre de recettes permettant la fabrication de produits agroalimentaires grâce aux différents procédés comme le mélange ou la cuisson. Or, les procédés complexes ainsi que les contraintes industrielles d'une recette impliquent le risque de sa contamination par un élément étranger. Une telle situation peut mener à la présence fortuite de substances non indiquées dans la liste d'ingrédients du produit. Il s'agit d'un véritable problème de santé publique qui concerne particulièrement les personnes ayant des allergies alimentaires. Un tel constat pose la question du contrôle des produits pulvérulents tout au long de la chaîne de traitement agroalimentaire. Il existe actuellement plusieurs méthodes permettant de limiter les risques de contaminations dans ce contexte. De plus, des méthodes de contrôle en temps réel se développent pour compléter les méthodes de contrôle traditionnelles qui consistent à réaliser des tests chimiques sur des échantillons prélevés. C'est le cas de la spectroscopie proche infrarouge (SPIR) qui permet l'analyse indirecte de la nature chimique d'un échantillon par l'étude de son interaction avec des radiations proche infrarouges.

L'imagerie hyperspectrale proche infrarouge est une technique qui utilise la combinaison de l'imagerie et la SPIR. La spectroscopie classique est capable, en fonction du dispositif de mesure utilisé, de produire une mesure sur une zone déterminée de l'échantillon. Le spectre mesuré est ensuite considéré comme représentatif de l'ensemble de la matrice étudiée. L'imagerie hyperspectrale permet de dresser une carte spectrale de l'échantillon puisque chaque pixel de l'image correspond à une mesure spectrale. Après transformation de la mesure spectrale en information d'intérêt, l'imagerie est donc capable de fournir une carte de l'échantillon pouvant mener à une étude de son hétérogénéité chimique. Cette approche est particulièrement intéressante pour mettre en avant la présence d'éléments anormaux dans une matrice, pourvu que la mesure spectrale les fasse ressortir par l'analyse chimiométrique.

L'imagerie hyperspectrale proche infrarouge a été utilisée pour de multiples applications en sécurité alimentaire. Quelques exemples sont la recherche de contaminations avec la détection de mélamine dans la poudre de lait ou de la cacahuète dans la farine de blé. Ces applications montrent l'utilisation de différentes méthodes chimiométriques appliquées à l'imagerie hyperspectrale. Les algorithmes de classification ou l'utilisation de métrique de similarité spectrale permettent d'assigner une classe à chaque pixel de l'image. Ces classes sont définies par des mesures des spectres purs des composés chimiques à identifier. Dans d'autres cas, des méthodes de quantification sont utilisées pour assigner une proportion à chaque pixel ou à l'ensemble de l'image.

Les méthodes de démélange telles que la *Multivariate Curve Resolution Alternating Least-Squares* ont été utilisées. Cette méthode permet de résoudre le problème du mélange linéaire au niveau de chaque pixel. En spectroscopie, elle est généralement utilisée pour des méthodes plus résolues que le proche infrarouge comme le Raman ou le moyen infrarouge dans un but de quantification. Dans cette thèse, nous proposons d'appliquer cette méthode dans un but de détection de composés minoritaires. Dans ce contexte, l'ambiguïté du modèle est grande car les spectres des composés purs sont proches. Par conséquent, nous proposons de mettre en place des contraintes particulières pour obtenir un démélange efficace.

Finalement, les méthodes de détecteur à sous-espace sont utilisées dans le domaine de l'observation terrestre pour la détection de cible. Ces méthodes permettent de prendre en compte la variabilité spectrale du contaminant et de l'échantillon séparément afin de construire une métrique de détection. Une difficulté de cette méthode réside dans le choix des paramètres du détecteur qui nécessite d'utiliser des valeurs de référence pour la validation. Ces données sont très difficiles à obtenir dans le cas de l'étude des produits agroalimentaires. Nous proposons de mettre en place ce type de détecteur en utilisant une stratégie de simulation de données spectrales pour sa validation.

La profondeur de pénétration des radiations du proche infrarouge est étudiée depuis le début du siècle dernier. Cependant, ces développements théoriques ne permettent pas toujours d'évaluer ce phénomène dans des cas pratiques beaucoup

plus complexes. Plusieurs auteurs ont participé au développement d'une approche empirique permettant de déterminer la profondeur de pénétration des radiations du proche infrarouge dans plusieurs matériaux comme des fruits ou des tissus humains. Cette méthode ne permet pas d'évaluer la capacité de détection d'un contaminant à travers une poudre. Une autre approche, empirique elle aussi, a été développée pour déterminer l'efficacité d'un modèle de détection de mélamine avec différentes couches de profondeur de poudre de lait. Cette méthode apporte un éclairage important sur le sujet mais ne permet d'évaluer avec précision la profondeur de détection du dispositif. Nous proposons une nouvelle méthode qui s'appuie sur l'utilisation d'un socle de mesure adapté et d'une méthode chimiométrique pour la détermination de la profondeur de détection d'un système.

Ce manuscrit est décomposé en trois chapitres portant sur trois travaux publiés dans des revues scientifiques.

Le premier chapitre est consacré au développement de la méthode de détermination de la profondeur de pénétration. La conception du socle de mesure et la méthode chimiométrique sont expliquées. La méthode est ensuite appliquée à une farine de blé.

Le deuxième chapitre est dédié à l'étude de la détection de composés minoritaires en imagerie hyperspectrale par l'utilisation d'un détecteur à sous-espace. Sa conception est détaillée et validée par une approche de simulation de données spectrales. Le détecteur est ensuite appliqué à la détection de farine de cacahuète dans la farine de blé.

Le troisième chapitre propose le développement d'une méthode de démélange du signal des pixels basé sur la *Multivariate Curve Resolution Alternating Least-Squares* (MCR-ALS). La mise en place d'une contrainte de sélectivité et la combinaison avec un algorithme de détection d'aberrations sont détaillées. Cette méthode est appliquée à la détection de farine de cacahuète dans une poudre de chocolat, mélange de cacao et de sucre.

La profondeur de détection d'un système de mesure hyperspectrale proche infrarouge

Introduction

L'industrie agroalimentaire cherche à garantir l'innocuité de ses matières premières qui sont souvent sous forme de poudre. La spectroscopie proche infrarouge permet d'étudier indirectement la chimie de ces produits sur une épaisseur finie à cause de la profondeur de pénétration des radiations.

Le problème de la profondeur de détection implique trois concepts : la profondeur de pénétration des radiations du proche infrarouge, la dynamique du capteur du système de mesures et la nature des signaux spectraux à détecter.

La théorie de Kubelka-Munk fournit une expression pour l'intensité du signal de réflexion obtenu pour un matériau d'épaisseur infinie ayant un coefficient d'absorption et de diffusion connu. Cette formule implique que le signal de réflexion converge en intensité pour une épaisseur donnée ; et que les couches plus profondes n'influencent pas le signal mesuré.

D'autres auteurs ont proposé une démarche empirique consistant à faire varier l'épaisseur de poudre pour des mesures successives de réflexion diffuse. Cette approche permet d'obtenir des profils de réflectance qui sont analysés pour mesurer la profondeur de pénétration des radiations. Ces travaux montrent une cohérence avec les résultats théoriques et permettent d'obtenir des valeurs pratiques de la profondeur de pénétration des radiations sur plusieurs gammes et dans plusieurs matériaux.

D'autres auteurs encore, ont analysé le problème de la profondeur de détection sous un angle empirique différent. La démarche consiste à déposer une couche de poudre d'épaisseur variable par-dessus une couche de contaminant à détecter. Un modèle chimiométrique développé en amont a pour but de détecter les spectres du contaminant. Ce dernier est appliqué sur les mesures réalisées avec une épaisseur de poudre qui augmente. La dégradation progressive des résultats de détection du modèle permet d'évaluer la profondeur de détection du dispositif.

Nous proposons une nouvelle approche qui introduit l'utilisation d'un socle de mesure en pente faisant varier progressivement l'épaisseur de poudre. Nous proposons une méthode pour l'évaluation de la profondeur de détection et nous analysons ces résultats en utilisant l'approche empirique des profils de réflectance.

Matériel et méthodes

Un socle de mesure adapté a été conçu dans le but d'étudier la profondeur de détection dans la farine de blé. Il se compose une cavité centrale en pente de 0.5 mm à 3.5 mm qui permet de faire varier l'épaisseur de farine. Le matériau du socle est de l'acide polylactique (PLA), un polymère ayant une absorption spectrale marquée à 1168

nm. Ce matériau sert de cible spectrale pour la détection dans la farine de blé. La connaissance de la géométrie du socle ainsi que la mesure hyperspectrale de ce dernier permettent d'attribuer, pour chaque pixel, la valeur d'épaisseur de farine correspondante.

Cette information est exploitée pour permettre l'extraction des profils de réflectance : la valeur de réflectance mesurée en fonction de l'épaisseur de farine. Ces profils sont obtenus pour chaque longueur d'onde du spectre. L'information de l'épaisseur de chaque pixel est aussi utilisée pour construire l'ensemble des données d'apprentissage de la régression *Partial Least-Square* (PLS). Cette méthode de régression permet de relier les niveaux d'absorbance spectrale aux valeurs d'épaisseur de la farine de blé.

Résultats et discussions

L'analyse des profils de réflectance montre deux phases. Dans la première, le niveau de réflectance augmente en fonction de l'épaisseur de la farine de blé. Ceci s'explique par le phénomène d'absorption du PLA qui s'atténue. Pour une profondeur de PLA donnée, celui-ci n'influence plus le signal de réflectance en surface ce qui explique que le profil de réflectance se stabilise à la valeur prévue par la théorie de Kubelka Munk.

Cette perte d'information dans le signal s'explique par le faible nombre de photons qui parviennent à traverser la couche de farine pour atteindre le PLA. Ces photons ont une influence très faible sur le signal. Or une telle influence est difficilement perceptible pour un capteur notamment à cause du phénomène de saturation et de la forte contribution de la réflexion surfacique.

La méthode de la régression PLS permet d'obtenir la profondeur de PLA perçue en fonction du spectre complet de réflectance. Comme pour le profil de réflectance, cette méthode met en évidence deux phases qui permettent de caractériser la profondeur de détection du PLA dans la farine de blé.

Ces deux approches montrent une cohérence dans les résultats obtenus et permettent de mettre en évidence la profondeur de détection du PLA d'environ 2 mm dans la farine de blé.

Conclusion du premier chapitre

La méthode proposée est applicable à d'autres couples d'échantillons dont on souhaite connaître la profondeur de détection. Notre étude montre qu'une cible spectrale (comme le PLA) ne peut être détectée qu'à une profondeur maximale de 2 mm. Cela signifie que la réflexion diffuse proche infrarouge n'est pas capable d'assurer l'innocuité d'un échantillon de farine de blé au-delà de 2 mm. Cette valeur dépend des espèces chimiques en jeu, aussi bien du point de vue de l'échantillon que de celui de la cible. Elle est aussi influencée par la granulométrie et le tassage de la poudre. Par

conséquent il s'agit d'une propriété difficile à prévoir qu'il est nécessaire d'étudier au préalable d'une démarche de détection.

La détection de farine de cacahuète par détecteur à sous-espace

Introduction

La cacahuète est une matière première utilisée dans l'industrie agroalimentaire. Il s'agit aussi d'un allergène alimentaire majeur. Quelques études scientifiques ont déjà démontré la possibilité de détecter des éclats de cacahuète dans la farine de blé par imagerie hyperspectrale proche infrarouge. Cependant, les particules de cacahuète étaient repérables de deux manières : leur taille plus importante que le pixel, et leur composante lipidique qui permet d'identifier leurs spectres. En revanche, certains contaminants comme la farine de cacahuète ne contiennent pas de composante lipidique et ont des particules plus petites que les pixels. Dans ce cas, les méthodes utilisées ne sont plus adaptées.

Le *Matched Subspace Detector* (MSD) est une méthode utilisée en imagerie hyperspectrale d'observation terrestre permettant la détection de cibles sous pixelliques. Cette méthode a l'avantage de prendre en compte la variabilité des mesures spectrales dans la métrique de détection. Cependant, sa conception nécessite le choix de paramètres nécessitant d'être validé par des données de référence très difficiles à obtenir dans le cas de poudres.

Matériel et méthodes

Des échantillons de farine de blé adultéré par de la farine de cacahuète en 8 proportions massiques différentes (de 20 % à 0.02 %) sont préparés et mélangés pour être disposés dans des socles de mesure. Ces derniers sont mesurés par un système d'imagerie hyperspectrale proche infrarouge.

Les mesures des échantillons purs de farine de blé et de cacahuète sont utilisées pour mettre en place une stratégie de simulation de données par Analyse en Composantes Principales (ACP). Cette méthode permet de définir les espaces de variabilité spectrale des mesures pour chacun des échantillons. Ensuite, le modèle de mélange linéaire est utilisé pour simuler des spectres de mélange en plusieurs proportions.

Le MSD est développé à partir d'un test d'hypothèse : dans la première, un pixel ne contient que des particules de farine de blé ; dans le deuxième, le pixel contient aussi des particules de cacahuète. Le modèle de mélange linéaire et les profils spectraux obtenus par ACP sont obtenus pour modéliser chacune des hypothèses. Dans chaque cas, le choix du nombre de profils utilisés pour décrire la farine de blé et la farine de cacahuète sont les paramètres à optimiser.

Résultats et discussions

Les résultats de la simulation spectrale montrent une cohérence dans la répartition des spectres sur un graphe de plan factoriel issu de l'ACP. Cela permet de valider leur usage pour la validation et le choix des paramètres du MSD.

Les trois choix de paramètres les plus prometteurs sont comparés en utilisant les données simulées pour validation. Les taux de détection des spectres simulés de 5 % à 20 % de cacahuète sont utilisés et montrent que deux choix de paramètres ont les meilleures performances.

Les détecteurs sont ensuite appliqués sur les mesures des échantillons réels. Les résultats montrent une cohérence avec les résultats simulés dans le nombre de détections obtenues. L'analyse des détections par comparaison des cartes permet de sélectionner les meilleurs paramètres qui détectent le plus de pixels.

Ces résultats montrent que les paramètres choisis correspondent au meilleur équilibrage du nombre de profils spectraux utilisés dans la conceptualisation du MSD. L'analyse du nombre de détections sur les échantillons réels montre qu'ils sont cohérents avec les concentrations en cacahuète introduites.

Conclusion du deuxième chapitre

Cette étude montre que l'utilisation de méthode de détection basée sur les détecteurs à sous-espace est pertinente pour des applications en sûreté alimentaire. Cependant, la validation des résultats dans le cas de détection de composés minoritaires dans les poudres est un problème technique complexe à résoudre.

La simulation spectrale permet d'apporter des données utiles à la validation du détecteur. Nous proposons une méthode dont les résultats montrent une cohérence avec les données réelles mesurées. De plus, nous montrons que le MSD est capable de détecter des contaminations globales d'échantillon de l'ordre de 0.02 %.

La détection de farine de cacahuète dans le chocolat en poudre par *Multivariate Curve Resolution Alternating Least-Squares*

Introduction

Le chocolat en poudre est un produit transformé de l'industrie agroalimentaire constitué majoritairement de cacao et de sucre. L'étude de l'adultération d'un tel produit par de la cacahuète est un sujet plus complexe que la farine. En effet, la poudre de chocolat est un mélange complexe qu'il est difficile de qualifier d'un point de vue

spectral. Il est donc avantageux d'utiliser une méthode de démélange par pixel qui permet de faciliter la procédure de détection par la suite.

La *Multivariate Curve Resolution Alternating Least-Squares* (MCR-ALS) est une méthode de démélange permettant de décomposer le signal spectral de chaque pixel en une somme de signaux purs. Cette méthode est donc pertinente pour traiter l'adultération du chocolat en poudre par la cacahuète.

Matériel et méthodes

Les échantillons consistent en un mélange de farine de cacahuète avec de la poudre de chocolat industrielle en trois proportions différentes (de 10% à 0.01 %).

La MCR-ALS considère un modèle bilinéaire pour décomposer le signal de chaque pixel en une somme de profils spectraux et de leur coefficient de contribution. Un tel modèle est ambigu puisqu'une infinité de solutions peut exister. Pour obtenir celles adaptées à l'interprétation spectrale, il est donc nécessaire d'appliquer des contraintes.

Un premier groupe de contraintes correspond à la nature des signaux spectraux étudiés et est généralement utilisé dans le cadre des mesures spectroscopiques. Nous proposons un deuxième groupe qui y ajoute une contrainte de sélectivité sur les concentrations. La connaissance de la composition des images des échantillons purs permet de contraindre la valeur de leur concentration associée pour les profils spectraux correspondants.

Deux méthodes de démélange sont utilisées : la MCR-ALS utilisant le premier groupe de contrainte et la MCR-ALS-CSEL utilisant le deuxième. Par la suite, un algorithme de détection d'outlier basé sur un modèle de mélange de Gaussiennes est utilisé sur les profils de concentration obtenus. Cela permet d'obtenir des cartes de détection de la cacahuète dans la poudre de chocolat.

Résultats et discussions

Une ACP sur les données hyperspectrales montre la difficulté de démélanger les pixels issus des échantillons contaminés. L'analyse du plan factoriel montre une difficulté à interpréter chaque composante spectrale ce qui est problématique pour l'interprétation des détections.

La MCR-ALS permet une interprétation plus claire des composantes attribuées au sucre, au cacao et à la cacahuète. Cependant, nous remarquons une forte ambiguïté entre les deux derniers profils ce qui s'explique par la proximité chimique du cacao transformé et de la cacahuète dégraissée.

La MCR-ALS-CSEL permet de réduire l'ambiguïté entre ces deux profils spectraux grâce à l'introduction de la contrainte de sélectivité. La comparaison des cartes de détection obtenues grâce aux deux techniques de démélange montre une meilleure détection par la MCR-ALS-CSEL.

La position des pixels détectés en supplément par cette méthode montre une cohérence spatiale. Celle-ci est renforcée par l'analyse de la granulométrie des poudres qui montre un phénomène d'agglomération des particules de cacahuète. Bien que certains agglomérats de particules soient plus grands qu'un pixel de la caméra, leur contribution spectrale est aussi influencée par les couches inférieures de chocolat en poudre, ce qui justifie l'intérêt de la méthode de démélange.

Cette étude met en avant la difficulté de définir des profils spectraux dans le cas de mélanges complexes comme le chocolat en poudre. Dans le cas de produits transformés, la signature spectrale des matières premières peut être modifiée par les différents traitements. De ce fait, il est difficile de comparer les spectres de ces produits avec les spectres des matières premières non traitées ce qui complique l'interprétation.

Conclusion du troisième chapitre

Notre étude montre que la MCR-ALS est une méthode pouvant être utilisée pour la détection de particules plus petites que le pixel et même lorsqu'il existe une forte ambiguïté des signatures spectrales. L'introduction d'une contrainte de sélectivité sur les concentrations et la combinaison de la MCR-ALS avec un algorithme de détection d'aberrations permet de détecter jusqu'à une contamination globale de 0.1%.

Conclusion générale

Cette thèse propose d'étudier deux problématiques liées à la détection de composés minoritaires dans les poudres agroalimentaires par imagerie hyperspectrale. La première repose sur l'épaisseur limitée que les radiations du proche infrarouge sont capables d'inspecter. L'objectif du premier chapitre de cette thèse était de proposer une méthode pour la détermination de la profondeur de détection d'une cible spectrale sous une couche de poudre. Nous avons proposé un concept de socle de mesure pour réaliser ces mesures ainsi qu'une méthode de chimiométrie adaptée. Nous avons montré qu'un signal spectral pouvait être complètement atténué par une épaisseur supérieure à 2 mm de farine de blé, ce qui limite les applications de détection. En revanche, ces études ont été réalisées en utilisant un éclairage diffus. Il existe d'autres méthodes de mesure telle que la spectroscopie résolue spatialement qui permet de ne mesurer qu'une portion utile du signal. Nous pensons que cette technique a un potentiel très intéressant pour l'étude de la profondeur de détection.

La deuxième problématique abordée est celle de la détection de particules de taille inférieure à celle des pixels. Nous avons proposé une méthode de détection permettant de modéliser la variabilité des mesures spectrales grâce à au MSD. Nous avons montré l'efficacité de cette méthode dans le cas de la détection de cacahuète dans la farine de blé. À cause du manque de données de validation, une méthode de simulation de données peut être utile pour trouver les bons paramètres du détecteur.

Cependant, les méthodes de simulation linéaire négligent les interactions du signal au sein des poudres. Des méthodes non linéaires plus avancées pourraient être plus performantes afin d'obtenir une simulation plus proche de la réalité des mesures.

Dans un deuxième temps, nous avons étudié la contamination de chocolat en poudre par la farine de cacahuète. Nous avons proposé une méthode de démélange basé sur la MCR-ALS afin de traiter ce cas plus complexe. Le développement d'une contrainte et la combinaison de cet algorithme avec une détection d'aberrations ont permis de détecter des adultérations au niveau du pixel. Dans cette étude, la dimension spatiale des données hyperspectrales n'est pas exploitée. Or, nous avons montré qu'il existe une cohérence spatiale des résultats de détection qui pourrait faire l'objet d'une contrainte à l'aide de filtres spatiaux.

université PARIS-SACLAY

ECOLE DOCTORALE N°581

Agriculture, alimentation, biologie, environnement et santé (ABIES)

Titre : Détection de composants minoritaires dans les produits pulvérulents de l'industrie agro-alimentaire par imagerie hyperspectrale proche infrarouge

Mots clés : Imagerie hyperspectrale ; Spectroscopie proche infrarouge ; Profondeur de pénétration ; Démélange de spectres ; Poudre agro-alimentaire ; Détection

Résumé : L'imagerie hyperspectrale proche infrarouge (PIR) permet d'obtenir une carte spectrale d'un échantillon organique. La mesure d'un spectre pour chaque pixel de la caméra permet notamment la recherche de composés minoritaires dans les poudres agroalimentaires. Cependant, l'analyse spectrale PIR est limitée à une couche de profondeur donnée. De plus, la taille des particules associée à une résolution insuffisante des caméras PIR actuelles induisent un mélange des signaux spectraux dans les pixels de l'image. Ces deux problèmes sont une limitation pour l'analyse des composés minoritaires dans les poudres agroalimentaires.

Nous proposons une méthode permettant de déterminer la profondeur de détection d'une cible composite placée dans un produit pulvérulent tel que la farine de blé. Basée sur une régression par projection sur les structures latentes, cette méthode permet d'appréhender l'atténuation du signal PIR lorsque la couche de poudre augmente, et ce malgré les problèmes inhérents à la détection en profondeur.

Deux stratégies de démélange de spectres sont proposées dans le but de détecter les pixels contenant des signatures de particules minoritaires. Le manque de valeurs de référence utilisées en tant que données de validation des algorithmes ainsi que l'ambiguïté des spectres des composés purs à démélanger sont deux difficultés majeures. Une première stratégie consiste à modélisation la variabilité des spectres étudiés via l'Analyse en Composantes Principales afin de construire un algorithme de détection performant. La deuxième stratégie, basée sur la Multivariate Curve Resolution Alternating Least-Squares permet le démélange des signaux par pixels dans un cas plus complexe.

Title : Detection of minor compounds in food powder using near infrared hyperspectral imaging

Keywords: Hyperspectral imaging ; Near-infrared spectroscopy ; Penetration depth ; Spectral unmixing ; food powder ; detection

Abstract: The near-infrared (NIR) hyperspectral imaging provides a spectral map for organic samples. Minor compounds in food powder can be looked for by analyzing the pixel spectra. However, the NIR spectral analysis is limited to a given depth. Besides, particles smaller than the pixel size induce a mixed spectral signature in the pixels. These two issues are an obstacle for the analysis of minor compounds in food powders.

We propose a method to determine the detection depth of a composite target under a layer of powder such as wheat flour. It is based on the Partial Least Squares regression and provides an understanding of how the NIR signal is attenuated when the layer of powder increases despite the penetration depth issues.

Two spectral unmixing strategies are proposed to detect pixel with minor compound NIR signatures. The lack of reference values to validate the model and the ambiguity of the spectral signature to unmix are two major difficulties. The first method models the spectral variability using Principal Component Analysis to design a performant detection algorithm. Then, for a more complex situation, the Multivariate Curve Resolution Alternating Least-Squares algorithm is used to unmix each pixel.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France