



**HAL**  
open science

# A partial least squares-based approach to assess the light penetration depth in wheat flour by near infrared hyperspectral imaging

Antoine Laborde, Benoît Jaillais, Ryad Bendoula, Jean-Michel Roger, Delphine Jouan-Rimbaud Bouveresse, Luc Eveleigh, Dominique Bertrand, Anthony Boulanger, Christophe Cordella

## ► To cite this version:

Antoine Laborde, Benoît Jaillais, Ryad Bendoula, Jean-Michel Roger, Delphine Jouan-Rimbaud Bouveresse, et al.. A partial least squares-based approach to assess the light penetration depth in wheat flour by near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy*, 2020, 28 (1), pp.25-36. 10.1177/0967033519891594 . hal-02623568

**HAL Id: hal-02623568**


**<https://hal.inrae.fr/hal-02623568v1>**

Submitted on 28 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A partial least squares-based approach to assess the light penetration depth in wheat flour by near infrared hyperspectral imaging

Journal of Near Infrared Spectroscopy  
2020, Vol. 28(1) 25–36  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/0967033519891594  
journals.sagepub.com/home/jns  


Antoine Laborde<sup>1,2</sup> , Benoît Jaillais<sup>3</sup> , Ryad Bendoula<sup>4</sup> ,  
Jean-Michel Roger<sup>4</sup>, Delphine Jouan-Rimbaud Bouveresse<sup>1</sup>, Luc Eveleigh<sup>5</sup>,  
Dominique Bertrand<sup>6</sup>, Anthony Boulanger<sup>2</sup> and Christophe BY Cordella<sup>1</sup>

## Abstract

Near infrared hyperspectral imaging technique has been used for adulteration detection in food samples for several years. However, the sensor cannot screen beyond a certain material thickness. This work studies a method to determine the penetration depth of near infrared radiations in the context of detection. The case of wheat flour in a polylactic acid sample holder is investigated. A sample holder is specially designed to have the wheat flour thickness vary from 0.5 to 3.5 mm. Hyperspectral images are acquired and the partial least squares regression method is used to determine the amount of polylactic acid in each spectrum. Partial least squares prediction results are interpreted using sensor considerations and the Kubelka–Munk theory. Thereafter, the Kubelka–Munk model is fitted and the penetration depth is determined for each wavelength using the reflectance profiles. Similarities between these results and partial least squares regression coefficients lead to the conclusion that partial least squares, combined with the near infrared spectra, is able to characterize the detection depth. The value is calculated by fitting two linear models on the partial least squares prediction results. As a result, 1.80 mm is found to be the detection depth, defined as the maximum thickness of wheat flour to ensure the detection of polylactic acid through the background. Reflectance profiles also show that the penetration depth is highly dependent on the wavelength. This study aimed at providing a method that could be used to evaluate the penetration depth and the detection depth in other contexts than a polylactic acid target in wheat flour.

## Keywords

Hyperspectral imaging, near infrared spectroscopy, penetration depth, partial least squares regression, detection

Received 10 April 2019; accepted 3 October 2019

## Introduction

Near infrared (NIR) spectroscopy has been used for the last 20 years in many industries as a fast, cost-effective and non-destructive control method. The food industry uses this technology to determine products quality parameters at different stages of their process. Recent advances in NIR technology have enabled the combination of NIR spectroscopy with spatial imaging to create hyperspectral imaging (HSI) spectroscopy. Since food products may be chemically heterogeneous, HSI is particularly useful as it enables acquisition of chemical information. As a consequence, this technique has been recognized as a powerful tool to provide quality and safety analysis for the food industry during the last decade.<sup>1</sup> HSI technology has proven its efficiency for the detection of defects and adulterations. Spectroscopic measurement with a large point-source is not able to detect small amounts of contaminant in a food product.

Since HSI resolves the spectroscopic measurement to an image with hundreds of small pixels, there is a better chance to isolate the contaminant in a small group of pixels and to be able to detect its spectral signature. Indeed, HSI has been used for several food adulteration purposes: melamine in milk powder,<sup>2–4</sup> peanut in wheat flour,<sup>5,6</sup> sorghum oats and corn in

<sup>1</sup>UMR914 PNCA INRA/AgroParisTech/Université Paris-Saclay, Paris, France

<sup>2</sup>GreenTropism, Paris, France

<sup>3</sup>Unité de Statistiques, Sensométrie, Chimiométrie INRA/ONIRIS, Nantes, France

<sup>4</sup>IRSTEA, UMR ITAP, Montpellier University, Montpellier, France

<sup>5</sup>UMR 1145 Ingénierie Procédés Aliments INRA/AgroParisTech/Université Paris-Saclay, Paris, France

<sup>6</sup>Dataframe, Nantes, France

### Corresponding author:

Antoine Laborde, AgroParisTech, 16 rue Claude Bernard, Paris 75231, France.

Email: lab.antoine@gmail.com

wheat flour<sup>7</sup> and sugar syrup in honey<sup>8</sup> are some examples. In this context, there is considerable 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.<sup>9</sup> 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 NIR spectroscopy<sup>10</sup> 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<sup>2</sup> have two very distinct spectral signatures, but it is less true for wheat flour and peanut particles.<sup>5</sup> As a consequence, the required signal-to-noise ratio (SNR) 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 has 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 determine the maximal depth for which infrared radiations can be used to detect material in a detection purpose. Then, a better understanding of the phenomenon encompasses the physical phenomenon of light penetration, the sensor dynamic and the unmixing application case.

The Kubelka–Munk theory<sup>11–13</sup> provides some understanding about the penetration depth of N radiations. Using a model with two fluxes of photons, it shows the diffuse reflectance for an infinitely deep sample ( $R_\infty$ ) depends on the ratio between the scattering and absorption constants.<sup>9</sup> 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  $R_\infty$ . 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.<sup>9</sup> proposed an approach to determine the number of interrogated particles by comparing the base-line-corrected value of the pseudo-absorbance  $\log_{10}(1/R)$  to the absorbance per particle. The authors deduced that for an absorbing matrix like carbazole, the penetration depth is less than 1 mm. Berntsson et al.<sup>14</sup> have provided further understanding about the effect of sample thickness on diffuse reflectance measurements. 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 that defines the sample mass which is sufficient to reach 98% of the diffuse reflectance of a corresponding optically thick sample ( $R_\infty$ ). 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 and 2500 nm. For microcrystalline cellulose powder, the penetration depth shows a global decreasing behavior between 1000 and 2500 nm with penetration depth varying between 2.0 and 0.33 mm.

Stolik et al.<sup>15</sup> measured human tissues in transmittance mode in order to determine their penetration depth. 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.<sup>16</sup> used reflectance diffuse measurements on apple slices in order to determine the

penetration depth of NIR radiations. 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 and 4 mm in apples.

More recently, Padalkar and Pleshko<sup>17</sup> have worked on the light penetration depth into cartilage to ensure the signal is not corrupted by underlying subchondral bone. 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.<sup>18</sup> used a similar protocol by placing melamine under different thicknesses of milk powder. The authors showed that a partial least squares-discriminant analysis (PLS-DA) failed to detect melamine contribution for a thickness of milk powder larger than 2 mm.

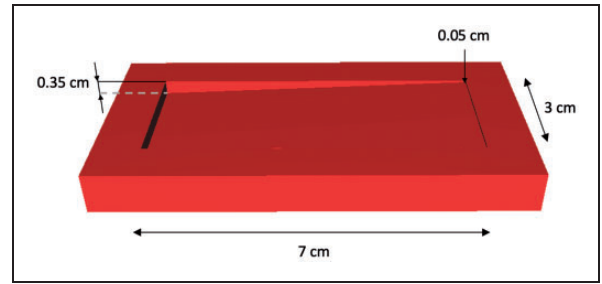
The literature shows that light penetration depth in spectroscopy can be studied through different underlying definition of the phenomenon. Some authors rely on theoretical models such as Kubelka–Munk or the diffusion model, whereas others use an empirical method that is specific to the application such as melamine and milk powder<sup>18</sup> or light penetration into cartilage.<sup>17</sup> 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 NIR radiation into wheat flour using a sample holder of polylactic acid (PLA). Hyperspectral imaging was used to acquire a great number of spectra with spatial information. Partial least squares (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.

## Material and methods

### Samples

A sample holder has been designed and manufactured for this experiment using a three-dimensional printer (Figure 1). 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 to 0.5 mm. The sample holder is made of PLA plastic, which has absorbance peaks in the NIR spectral



**Figure 1.** Scheme of the sample holder. The central pit enables is designed as a slope from a 0.05 cm depth down to 0.35 cm. The whole structure is made of red polylactic acid.

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.

### 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 line with a pixel size of 300 μm × 300 μm. 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.

### Data processing

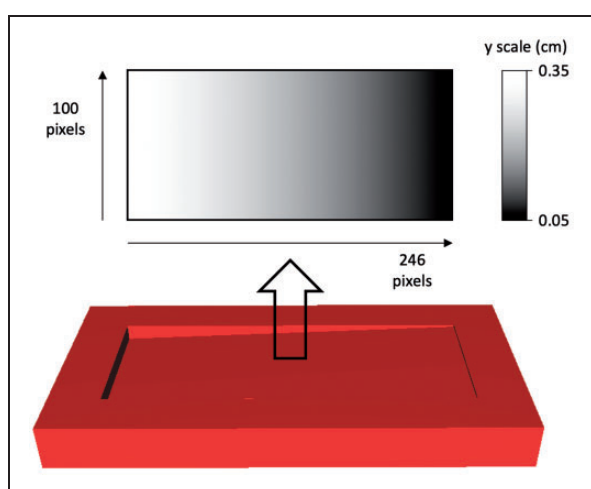
Images were cropped to focus on the central cavity of the sample holder leading to 100 × 246-pixels images. The white reference image was averaged to obtain one spectrum for every pixel of sensor's line ( $I_0$ ). The black measurement ( $I_B$ ) and the white reference were used to calculate the reflectance signal from the raw measurement ( $I$ ) using (equation 1)

$$R = \frac{I - I_B}{I_0 - I_B} \quad (1)$$

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

### 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  $y$  target value. This thickness depends on the sample holder geometry. As a consequence, the  $y$  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 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 two-dimensional mask that can be applied on the hyperspectral image (Figure 2). For spectral analysis, hyperspectral cubes are unfolded



**Figure 2.** Two-dimensional mask obtained for  $y$  target value using the sample holder geometry.

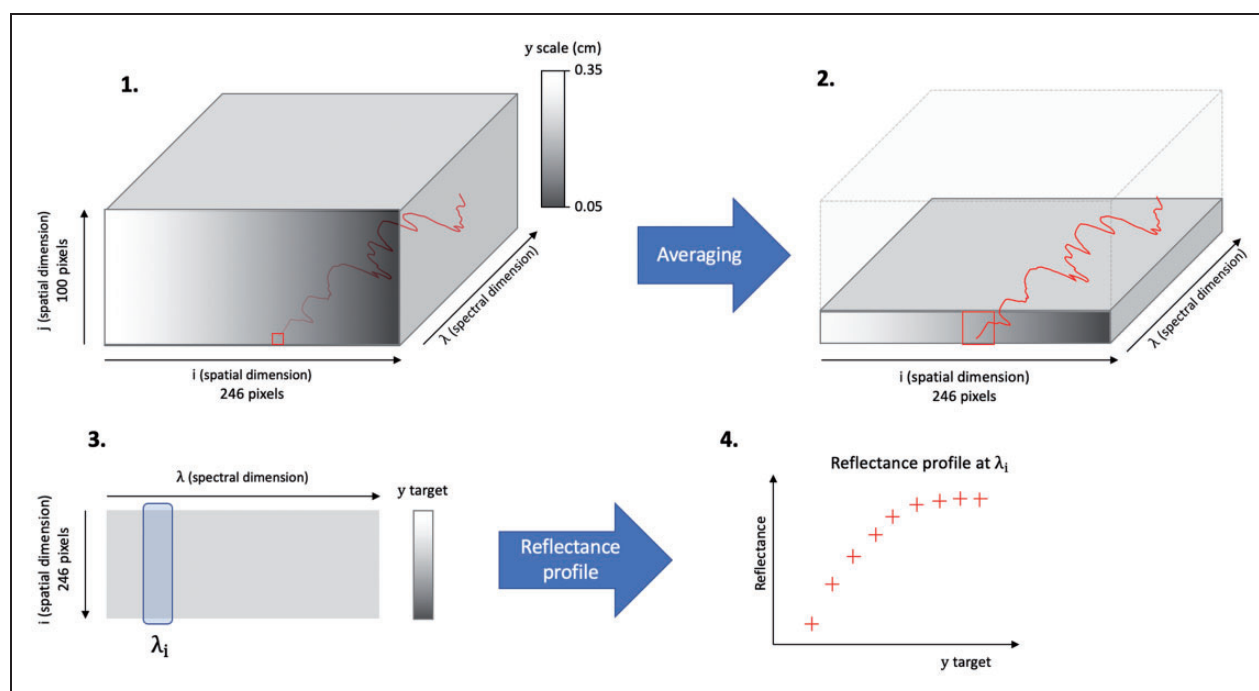
to obtain matrices of 24,600 lines and 256 columns. The two-dimensional mask for  $y$  values is unfolded in the same way so that each spectrum of the matrix is associated with the appropriate  $y$  target.

### Reflectance profile extraction

The reflectance profiles across the sample holder were extracted for each wavelength following the procedure described in Figure 3. All the pixels on the same vertical line were averaged in order to obtain one spectrum for each  $y$  thickness value (steps 1 to 2). As a result, a two-dimensional matrix was obtained as well as the corresponding vector of  $y$  target values. After selecting a wavelength band, all the corresponding reflectance values were extracted and plotted against the  $y$  thickness values giving the reflectance profile (steps 3 to 4).

### Partial least squares regression

Partial least squares regression is an algorithm used for predicting a target value  $y$  using predictors  $X$  with a linear relationship:  $y = X\beta$ . 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. Principal component analysis (PCA) is one method used for constructing such variables that are orthogonal and ranked according to the amount of variance they represent in  $X$ . PCR is



**Figure 3.** Procedure for obtaining reflectance profile from the original hyperspectral cube.

achieved by performing MLR on these new variables. However, PCR does not take into account the relationship with target values  $y$  in the construction of the orthogonal latent variables. PLS solves this problem by constructing latent variables based on the covariance between  $X$  and  $y$ .<sup>19,20</sup> PLS has been widely used in chemometrics as it is particularly suitable for NIR spectral data.<sup>21</sup> 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  $y$  thickness vector is used as target for the PLS calibration. The training was performed using cross-validation on the first sample replicate. About 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.

## Results and discussion

### Reflectance evolution for each wavelength

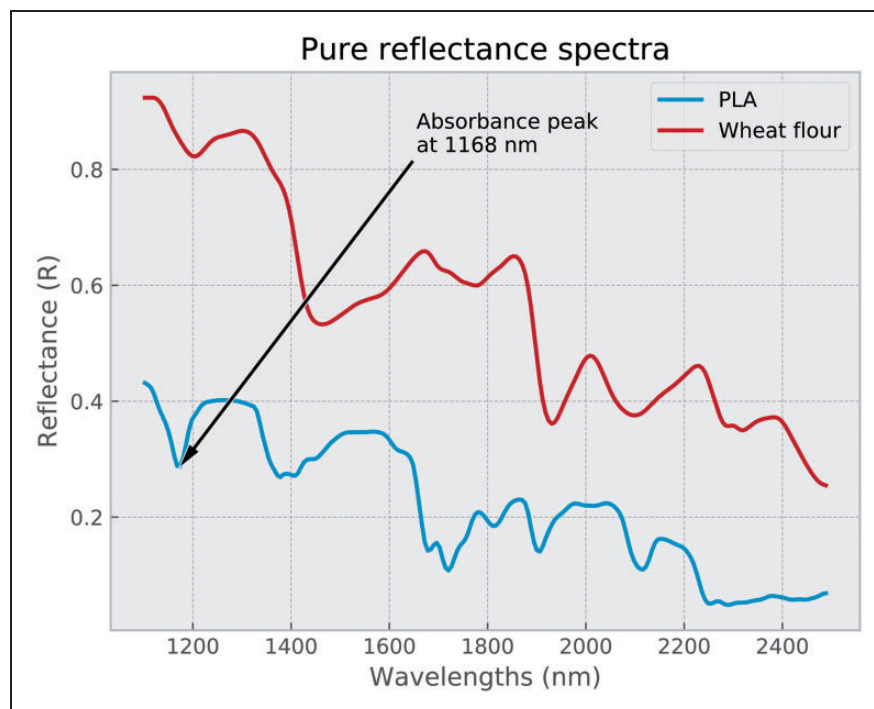
Figure 4 shows the reflectance spectral signatures of PLA and wheat flour. The spectrum of PLA exhibits high and resolute absorption peaks all along the near infrared range. The absorption peak at 1168 nm represents a high difference in reflectance between PLA and wheat flour. Figure 5 shows the reflectance 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 show an increasing reflectance profile. The second part shows a stabilization of the reflectance level for high thickness values.

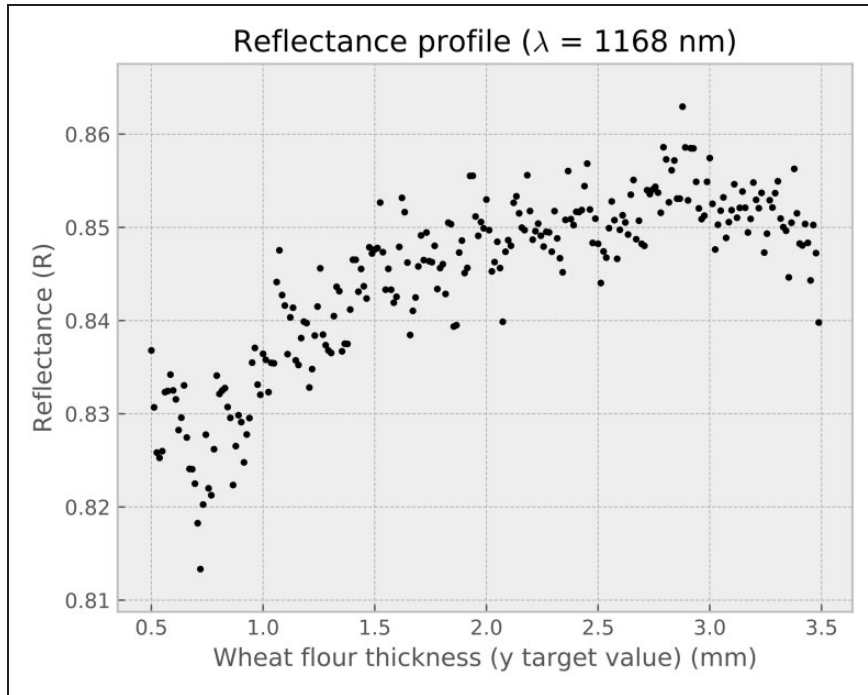
When the wheat flour thickness is low, the PLA plays an important role in the resulting diffuse reflectance signal. As it absorbs radiation around 1168 nm, the reflectance profile at this wavelength starts with low reflectance values. When the thickness increases, the role of wheat flour becomes more 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.

### Physical interpretation

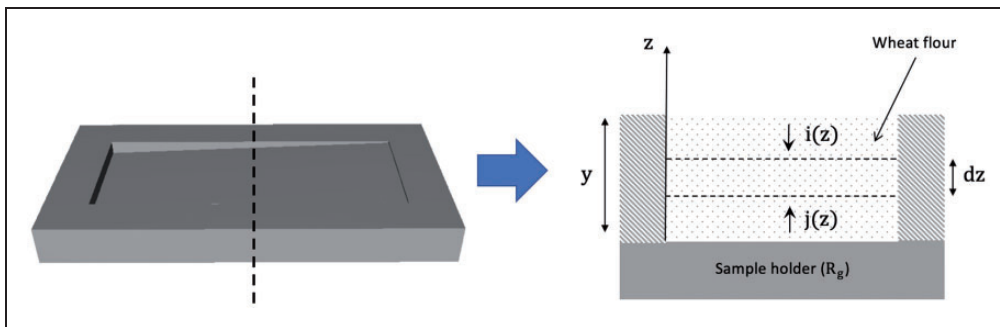
*The Kubelka–Munk model.* The Kubelka–Munk model developed on a monochromatic case will be applied to the following one. It is assumed the results are applicable to every wavelength of the detector range between 1100 and 1350 nm. Let us consider a layer of wheat flour of thickness  $y$  as shown in Figure 6. 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  $R_g$ . An infinitesimal layer of thickness  $dz$  is considered in the wheat flour at the height  $z$ . This layer is crossed by two



**Figure 4.** The pure reflectance spectra of wheat flour and polylactic acid.



**Figure 5.** The reflectance profile of near-infrared radiation at 1168 nm crossing increasing depths of wheat flour in the polylactic acid sample holder.



**Figure 6.** The Kubelka-Munk formalism applied on a slice of the sample holder. The borders of the sample holder are neglected.

fluxes: the descending flux  $i(z)$  and the ascending flux  $j(z)$ . 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

$$-di(z) = -Ki(z)dz - Si(z)dz + Sj(z) \quad (2)$$

$$dj(z) = -Ki(z)dz - Si(z)dz + Si(z) \quad (3)$$

These two equations result in a differential equation system

$$\begin{cases} \frac{di(z)}{dz} = (K + S)i(z) - Sj(z) \\ \frac{dj(z)}{dz} = -(K + S)i(z) + Si(z) \end{cases} \quad (4)$$

Kubelka and Munk proposed a solution to this system,<sup>22</sup> which only involves the coefficients  $K$  and  $S$  as well as the reflectance of the sample holder  $R_g$

$$R = \frac{1 - R_g(a - b \coth(bSy))}{a - R_g + b \coth(bSy)}, \quad \text{where} \quad (5)$$

$$a = \frac{K + S}{K} \quad \text{and} \quad b = \sqrt{a^2 - 1}$$

This solution (equation 5) 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.

The derivation of the Kubelka-Munk model shows the reflectance measurement ( $R$ ) approaches

a limit  $R_\infty$  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  $R_\infty$  only depends on  $K$  and  $S$ <sup>22</sup>

$$R_\infty = 1 + \frac{K}{S} - \sqrt{\frac{K^2}{S^2} + 2 \frac{K}{S}} \quad (6)$$

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

**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)<sup>23</sup> and lead 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 5) 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 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  $R_\infty$  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  $R_\infty$  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.

### Determination of the penetration depth

The reflectance profiles obtained for each wavelength (Figure 5) correspond to the context of the Kubelka–Munk theory. The experimental points are used to fit the model defined by equation (5) 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 5). This value is chosen to represent the spread of the experimental points around the fitted curve  $R(y)$ .
2. The value of  $R_\infty$  is calculated using equation (6).
3. The threshold reflectance value  $R_T$  is calculated as the difference between the reflectance limit  $R_\infty$  and the RMSE:  $R_T = R_\infty - \text{RMSE}$ .
4. The penetration depth  $y_p$  is determined such that  $R(y_p) = R_T$ .

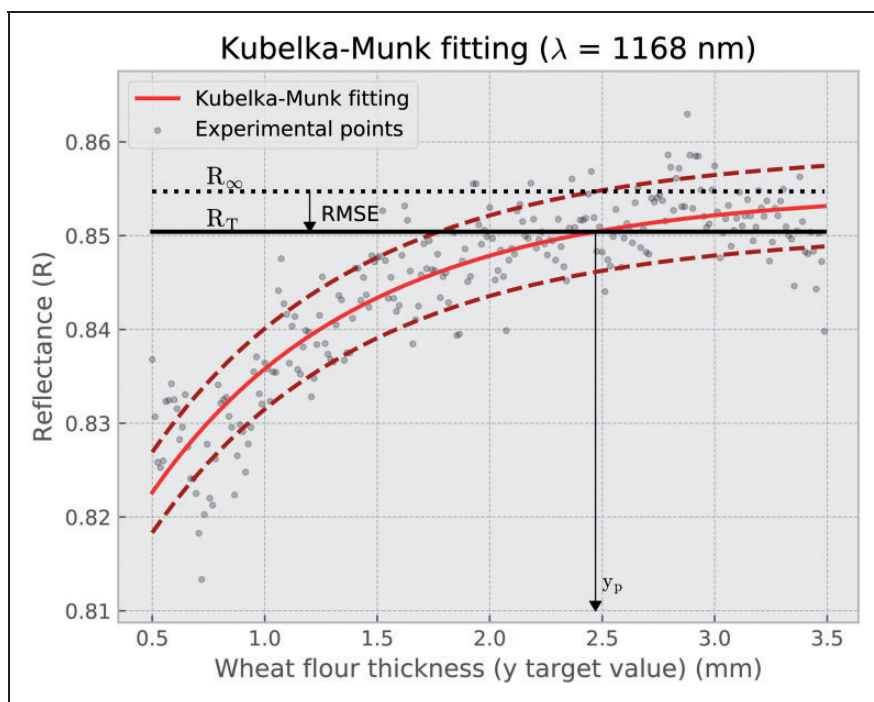
The penetration depth profile obtained is presented on the Figure 8. The profile is estimated on the range 1100–1350 nm because the Kubelka–Munk model fitting cannot be applied on higher wavelengths. In these conditions, the reflectance profile is flat because the PLA spectral signature is not visible 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 to 1150 nm). The profile can be explained by the absorption coefficients of the pure materials (Figure 4). 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 and 1170 nm can be explained by the strong absorption of PLA between 1123 and 1211 nm.

### PLS regression results

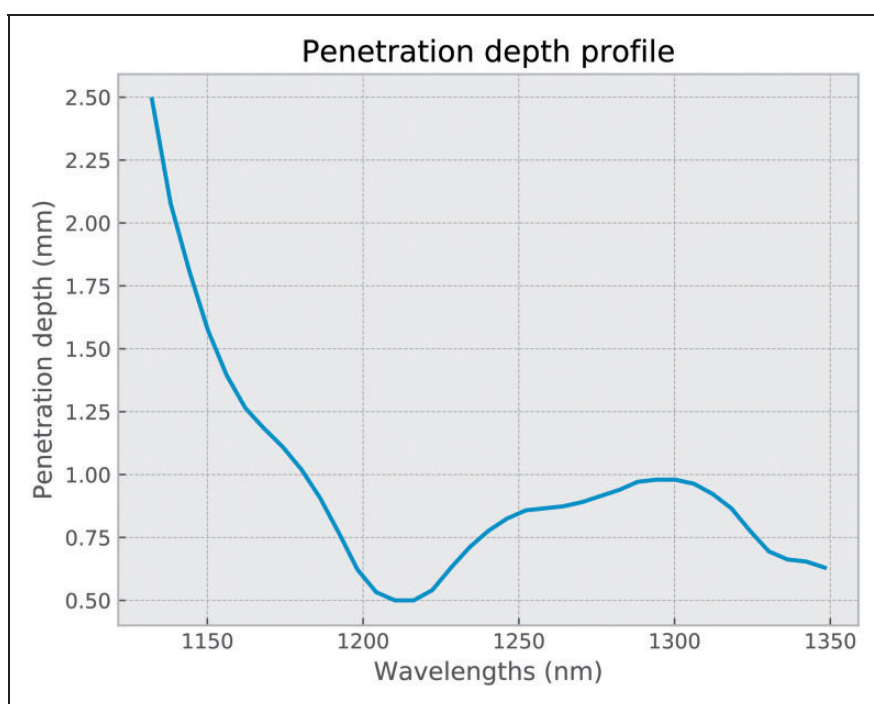
In this study, the use of PLS regression can be compared to unmixing problems. In this situation, the aim consists in finding the  $n$  pure spectral endmembers ( $\mathbf{s}_i$ ) associated with their proportions ( $c_i$ ) in a linear mixing model (equation 7) that decomposes the signal of the measured spectrum ( $\mathbf{x}$ ):

$$\mathbf{x} = \sum_{i=1}^n \mathbf{s}_i c_i + \mathbf{R} \quad (7)$$





**Figure 7.** Procedure for the calculation of the penetration depth at the wavelength band 1168 nm.



**Figure 8.** Penetration depth profile obtained from the reflectance profile measurements.

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 ( $\mathbf{R}$ ). However, solving the linear mixing model for each pixel requires some knowledge about the endmembers ( $s_j$ ). 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.<sup>24-26</sup> 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  $\hat{y}$  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  $\mathbf{y}$ .

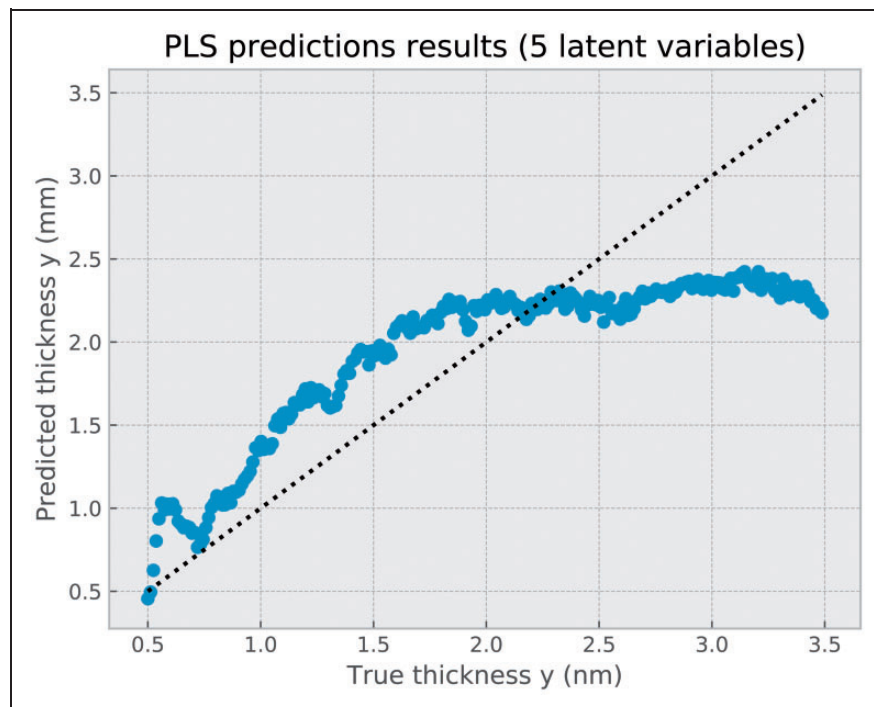
PLS model was trained on the spectra issued from the training images. Five latent variables were chosen as more variables did not improve the cross-validation error. The obtained model was then applied on the image of a different wheat flour sample. Each pixel gave a prediction leading to 24,600 prediction points. All predictions coming from pixels of the same line (same wheat flour thickness  $y$ ) were averaged to improve visualization. As a result, 246 prediction points were obtained and plotted on a graph (Figure 9) 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 and 1.5 mm). Thus, PLSR 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.

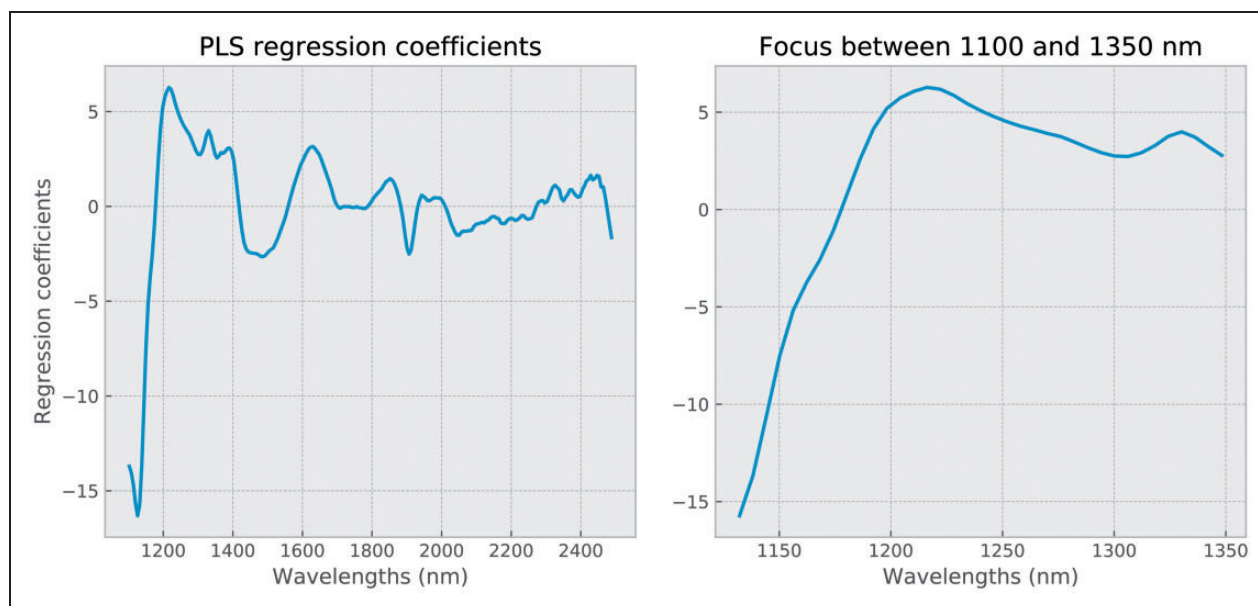
Figure 10 shows the PLS regression coefficients for each wavelength. The high weights attributed to the low wavelengths show the importance of these variables for quantifying the PLA spectral signature across the sample holder. The right part of Figure 10 focuses on the range 1100–1350 nm to highlight the similarities between the regression coefficients and the penetration depth results obtained (Figure 8).

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 9). Thus, by extrapolation, the Kubelka–Munk theory and the sensor considerations explain the behavior observed on the PLS prediction results (Figure 9).

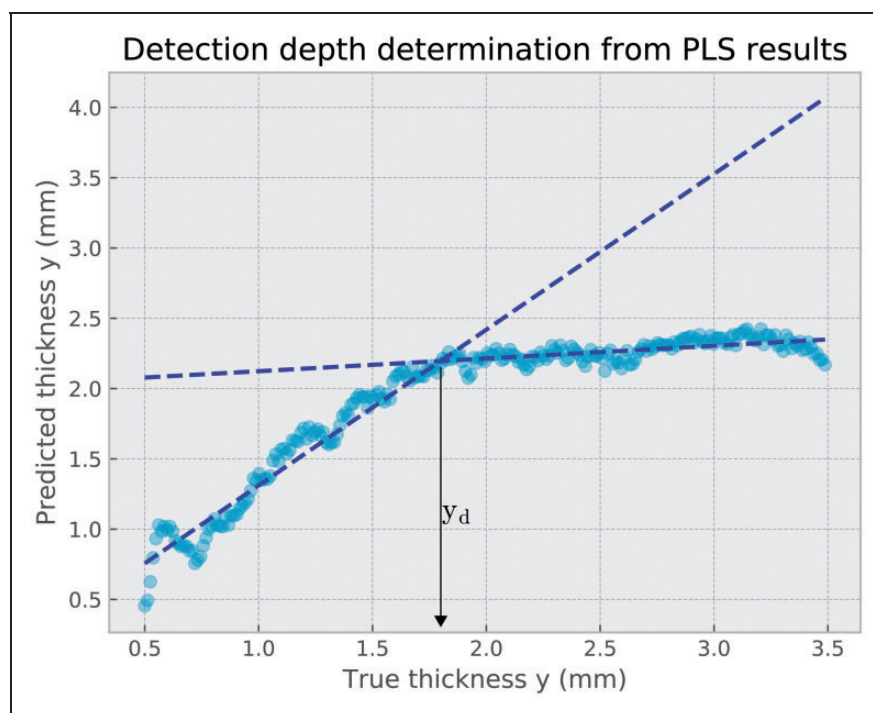
Indeed, for a given wheat flour thickness  $y$ , the signal of the PLA cannot be extracted from the diffuse reflectance spectrum. This may be an issue for detection problems when the target is buried under a layer of sample. For this reason, it is important to define the higher thickness  $y_d$  for which a detection is possible. This limit can be defined by using the PLS prediction results. In the context of the sample holder,



**Figure 9.** Partial least squares prediction results for wheat flour thickness across the sample holder.



**Figure 10.** Regression coefficients of the partial least squares model used for the prediction of the wheat flour thickness.



**Figure 11.** Determination method for the detection depth from the partial least squares prediction results.

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 11).

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.

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 was 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.

## Conclusions

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 until 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.




## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## ORCID iDs

Antoine Laborde  <https://orcid.org/0000-0003-2211-0094>  
Benoît Jaillais  <https://orcid.org/0000-0001-5196-6698>  
Ryad Bendoula  <https://orcid.org/0000-0002-2794-1252>

## References

1. Gowen AA, O'Donnell CP, Cullen PJ, et al. Hyperspectral imaging – an emerging process analytical tool for food quality and safety control. *Trends Food Sci Technol* 2007; 18: 590–598.
2. Pierna Fernández JA, Vincke D, Dardenne P, et al. Line scan hyperspectral imaging spectroscopy for the early detection of melamine and cyanuric acid in feed. *J Near Infrared Spectrosc* 2014; 22: 103–112.
3. Qin J, Kim M, Chao K, et al. Detection and quantification of adulterants in milk powder using a high-throughput Raman chemical imaging technique. *Food Addit Contam Part A* 2017; 34: 152–161.
4. Lim J, Kim G, Mo C, et al. Detection of melamine in milk powders using near-infrared hyperspectral imaging combined with regression coefficient of partial least square regression model. *Talanta* 2016; 151: 183–191.
5. Mishra P, Herrero-Langreo A, Barreiro P, et al. Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis. *J Near Infrared Spectrosc* 2015; 23: 15–22.
6. Mishra P, Cordella CBY, Rutledge DN, et al. Application of independent components analysis with the JADE algorithm and NIR hyperspectral imaging for revealing food adulteration. *J Food Eng* 2016; 168: 7–15.
7. Verdú S, Vásquez F, Grau R, et al. Detection of adulterations with different grains in wheat products based on the hyperspectral image technique: the specific cases of flour and bread. *Food Control* 2016; 62: 373–380.
8. Shafiee S, Polder G, Minaei S, et al. Detection of Honey Adulteration using hyperspectral imaging. *IFAC-PapersOnLine* 2016; 49: 311–314.
9. Olinger JM, Griffiths PR and Burger T. Theory of diffuse reflection in the NIR region. In: Burns DA and Ciurczak EW (eds) *Handbook of near-infrared analysis*. 2nd ed. New York: Marcel Dekker, 2001, pp. 19–51.
10. Harnby N, Characterization of powder mixtures. In: Harnby N, Edwards MF, Nienow AW (eds) *Mixing in the process industries*, 2nd ed. Oxford: Butterworth Heinemann, 1992, pp. 25–41.
11. Kubelka P and Munk F. Ein Beitrag zur Optik der Farbanstriche. *Zeitschrift für technische Physik* 1931; 12: 593–601.
12. Kubelka P. New contributions to the optics of intensity light-scattering materials. Part I. *J Optic Soc Am* 1948; 38: 448–457.
13. Kubelka P. New contributions to the optics of intensity light-scattering materials. Part II: nonhomogeneous layers. *J Optic Soc Am* 1954; 44: 330–335.
14. Berntsson O, Burger T, Folestad S, et al. Effective sample size in diffuse reflectance near-IR spectrometry. *Anal Chem* 1999; 71: 617–623.
15. Stolik S, Delgado JA, Pérez A, et al. Measurement of the penetration depths of red and near infrared light in human “ex vivo” tissues. *J Photochem Photobiol B Biol* 2000; 57: 90–93.
16. Lammertyn J, Peirs A, De Baerdemaeker J, et al., Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. 2000; 18: 121–132.

17. Padalkar MV and Pleshko N. Wavelength-dependent penetration depth of near infrared radiation into cartilage. *HHS Public Access* 2016; 140: 2093–2100.
18. Huang M, Kim M, Chao K, et al. Penetration depth measurement of near-infrared hyperspectral imaging light for milk powder. *Sensors (Switzerland)* 2016; 16: 1–11.
19. Helland IS. Partial least squares regression and statistical models. *Scand J Stat* 1990; 17: 97–114.
20. Geladi P and Kowalski BR. Partial least-squares regression: a tutorial. *Anal Chim Acta* 1986; 185: 1–17.
21. Naest T, Irgens C and Martens H. Methods of linear statistical comparison for calibration of NIR instruments. *J R Stat Soc Ser C* 2015; 35: 95–206.
22. Emmel P. Nouvelle formulation du modèle de Kubelka et Munk avec application aux encres fluorescentes. In: *Actes de l'Ecole de Printemps 2000 - Le Pays d'Apt en Couleurs*, 2000, 2000, pp. 87–96.
23. Rasti B, Scheunders P, Ghamisi P, et al. Noise reduction in hyperspectral imagery: overview and application. *Remote Sens* 2018; 10: 1–28.
24. Bioucas-Dias JM, Plaza A, Dobigeon N, et al. Hyperspectral unmixing overview: geometrical, statistical, and sparse regression-based approaches. *IEEE J Sel Top Appl Earth Obs Remote Sens* 2012; 5: 354–379.
25. Borel CC and Gerstl SAW. Nonlinear spectral mixing models for vegetative and soil surfaces. *Remote Sens Environ* 1994; 47: 403–416.
26. Somers B, Cools K, Delalieux S, et al. Nonlinear hyperspectral mixture analysis for tree cover estimates in orchards. *Remote Sens Environ* 2009; 113: 1183–1193.