Partial Least Squares Estimation of Crop Moisture and Density by Near-Infrared Spectroscopy

# Partial Least Squares Estimation of Crop Moisture and Density by Near-Infrared Spectroscopy

Davide Cassanelli, Nicola Lenzini, Luca Ferrari and Luigi Rovati, *Member, IEEE,*

***Abstract*—Optical methods can provide measurements without coming into contact with the sample. In the agrifood sector, this feature can be exploited to measure physical properties of crops. In particular, we focused our research on moisture content and density estimation. These two physical quantities of the crop are extremely important not only to determine future treatments to be performed, e.g. drying methods and processes but, also for estimating the value of the product. In this article, we propose a new model for simultaneous estimation of crop moisture content and density, using Fourier transform near-infrared spectroscopy combined with partial least square multivariate methods. The model has been developed considering 140 fresh Medicago sativa samples properly harvested. Moisture content ranged from 9.4% to 83.9% whereas density from 46 kg/m$^3$ to 236 kg/m$^3$. Reference MC was computed according to the American Society of Agricultural and Biological Engineers standard whereas reference density was determined estimating the volume of a sample of known mass. The obtained results indicated that crop moisture content and density information can be recovered from the near-infrared absorption spectrum of the sample with coefficients of determination $R^2$ = 0.925 and $R^2$ = 0.681 for the moisture content and density, respectively. Mean root mean square relative errors of the estimation were 13.8% and 14.4% for the moisture content and density, respectively.**

***Index Terms*—Near Infrared Spectroscopy, Precision agriculture, Crop moisture content measurement, Crop density measurement, Partial Least Squares estimation, Multivariate statistics.**

## I. Introduction

MODERN agriculture is continuing to develop and search new analytical techniques and methodologies to evaluate the physical properties of foods and crops [1], [2], [3]. Research is looking at fast, nondestructive techniques, among which, Near Infrared Spectroscopy (NIRS) is one of the most interesting option [4], [5]. NIRS is a well known optical technique that can provide the physical properties of a sample without contact. In particular, organic samples can be investigated exploiting absorption peaks included in the near-infrared spectral region [6] [7].

An interesting application of NIRS is the determination of the water content in the crop, also known as moisture content (MC) [8]. Crop sample is heterogeneous being composed by stems, leaves and air interstices. Therefore, the information of interest is the total moisture that includes both the water content in the solid part of the sample and the moisture deposited on it.

Author Davide Cassanelli and Luigi Rovati are with the Department of Engineering "Enzo Ferrari", University of Modena and Reggio Emilia, Via Vivarelli 10, 41125, Modena, Italy.. Nicola Lenzini and Luca Ferrari are with CNH Industrial Italia Spa, Viale delle Nazioni, 55, 41122 Modena, Italy

Manuscript received XXX, XXX. This work was funded in part by the XXXX Grant XXX.

Another important physical quantity that characterizes this heterogeneous sample is its density. During the harvesting and processing of the crop, different compressions exerted can reduce or increase the volume occupied by the air thus changing the ratio between solid and gaseous part (without any economic value) of the sample.

Moisture and density are crucial for the correct procedure to perform on the alfalfa to make the bale useful for agriculture. Moreover these properties influences the cost of the crop. The ultimate goal is to determine MC and density directly during the harvest by equipping the harvesting machine with a NIRS probe. Unfortunately, the influence quantities that affect the relationship between MC and NIR spectrum are numerous and conditions in which the measurement should be performed during the harvest require the development of an optical system compact, robust and capable of operating in extreme environmental conditions.A particularly important influence quantity is the density of the crop [9]. This quantity heavily affects the propagation of photons within the sample going to directly influence the scattering coefficient [10]. Another critical aspect for the NIRS measurement is the variability of the sample. Different harvesting points offer crops with different characteristics, e.g. in terms of leaf-to-stem ratio, due to different soil compositions or sun exposures. Therefore, a model for the MC and density estimation must necessarily start from statistical analysis on numerous samples. As proposed by other researchers [11][12], multivariate statistics is a suitable tool to perform such analysis [13].

Among different multivariate techniques, Partial Least Squares (PLS) regression represents a promising approach to analyze the acquired NIR spectra. PLS finds the best correlation of two matrices, i.e. response variables and predictors, even if the predictors are noisy or collinear [14]. Nevertheless, to increase the signal-to-noise ratio and enhance the correlation between the spectral data and the characteristic of interest, the acquired spectra need to be pre-processed before the regression [15]. Processing proposed for this aim is scaling, first derivative, second derivative, and filtering such as the Savitzky-Golay filter [16]. To improve the prediction capability of the PLS model, only specific bands of the spectrum are selected. This selection determines the quality of the fitting procedure and is usually based on the knowledge of the sample optical properties and analysis of the correlation between the reference and predictor variables [16].

In this paper, using Fourier Transform Near-Infrared Spectroscopy (FT-NIRS) combined with partial least square (PLS) multivariate methods, a new model for simultaneous estimation of crop MC and density is proposed. The model has been

developed considering fresh alfalfa (Medicago sativa L.)–grass samples harvested with a conventional harvesting machine.
In the following, section II presents the materials and methods used to develop the proposed model. Measuring principle, experimental setup, sample collection, reference methods, measurement procedure and data analysis are described in this section. The obtained results are reported in section III and conclusions are drawn in section IV.

## II. MATERIALS AND METHODS

### A. Measurement principle

NIR spectral absorption and scattering at specific wavelengths are exploited to estimate the crop moisture and density. Since the propagation of photons within the sample under test (SUT) is governed by these two phenomena, the modified Lambert Beer's law quantifies the loss of radiation intensity for beam travelling through the SUT.

In the near infrared spectral range, many SUT constituents contribute to the acquired spectrum by absorbing and scattering photons. Nevertheless, during the crop harvesting and processing, the concentration of these constituents can be assumed to be constant, with the exception of the water content and the density. Therefore, it is reasonable to assume that the changes observed in the acquired spectrum can be attributed to variations in moisture and density. The measurement of these physical quantities could thus be recovered from the spectrum after an accurate calibration. This approach is well known in the literature for example for the estimation of MC considering the water absorption peaks [17]. As an example, Fig.1 shows the spectra acquired on the same crop sample at two different moisture contents. Note the large variation in absorbance at the water absorption peaks, i.e. 1450 nm and 1940 nm.

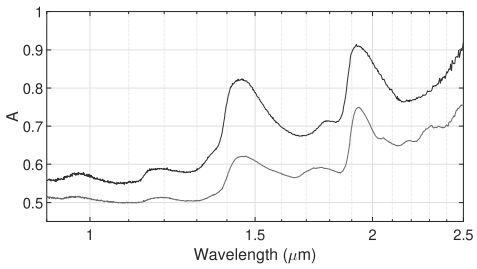



Fig. 1. Absorption spectra acquired from the same crop sample at two different moisture contents: MC=43.6% (•) and MC= 83.9% (•). Note the large variation in absorbance at the water absorption peaks, i.e. 1450 nm and 1940 nm.

Lambert Beer's law states that the outcoming light intensity from SUT is:

$$I = I_0 \cdot e^{-(c_W \cdot \epsilon \cdot L \cdot DPF + G)}, \tag{1}$$

where $I_0$ is the incident beam intensity, $c_W$ the water molar concentration, $\epsilon$ the water molar extinction coefficient [18], L the geometrical pathlength of photons within the SUT, DPF the Differential Pathlength Factor accounting for the increasing photon pathlength due to scattering, and G takes into account the losses due to scattering. Since our setup includes an integrating sphere (see next section) the last term can be neglected assuming an efficient collection of the scattered light. Therefore, the light attenuation, also called absorbance, is given by:

$$A = -\log_{10}\frac{I}{I_0} = -\frac{1}{2.303}\ln\frac{I}{I_0} = \frac{1}{2.303}\cdot c_W \cdot \epsilon \cdot L \cdot DPF. \tag{2}$$

From this simple equation, we can deduce how the water content in the SUT, i.e. $c_W$, can be derived by the absorbance measurement. Maximum sensitivity can be obtained by analyzing the absorbance in the spectral regions where the water molar extinction coefficient $\epsilon$ is maximum. On the other hand, since the scattering effect is described by DPF, we can suppose that higher density samples exhibit a greater scattering thus a larger DPF. However, DPF also depends on the photon absorption in the SUT and this makes estimating moisture and density from (2) non-trivial.
Note as the Lambert-Beer's law is fundamentally defined for a single absorbing material uniformly concentrated. Even if the modified law takes into account the phenomena of scattering, (1) assumes a gross simplification of the radiometric process underlying the spectral measurement. Nevertheless, this simplification is well known and have been employed with success over the last three decades to probe equally complex and heterogeneous samples such as highly scattering biological tissues, e.g. human brain, muscle and breast [19], [20], [21]. The basis of the modified Lambert-beer's law paradigm is that it is possible to derive changes in sample optical properties based on diffuse optical intensity measurements. The approach relates differential light diffusion changes (regardless of the considered geometry) to differential changes in tissue absorption [22].
In this paper, in the attempt to unravel the variables of interest and apply a linear regression model, we analyze the logarithmic transformation of the absorbance spectra:

$$LA = \ln A = \ln\left[\frac{1}{2.303}\cdot c_W \cdot \epsilon\right] + \ln\left[L \cdot DPF\right]. \tag{3}$$

The first addend in (3) is closely dependant to the SUT MC while the second to its density. To separate the contributions, we analyze the absorbance in two different spectral bands B1 and B2:
(B1) the spectral regions where the water molar extinction coefficient is maximum; this to maximize the contribution of the first term and neglecting the second addend in 3, thus deriving MC form the following approximation:

$$LA_{B1} \cong \ln\left[\frac{1}{2.303}\cdot c_W \cdot \epsilon\right]. \tag{4}$$

(B2) is the spectral regions where the water molar extinction coefficient is minimum, this to maximize the contribution of the second term neglecting the first addend in (3), thus deriving the density form the following approximation:

$$LA_{B2} \cong \ln\left[L \cdot DPF\right]. \tag{5}$$

We also assume that, in the spectral regions B2, the effect of the water content on the DPF is minimal, i.e. the propagation of photons is mainly governed by scattering. In bands B1 and

B2, this approximations allow us to separate the contribution of absorption, thus deriving MC, from that of scattering,thus deriving the density.
The approximations described in 4 and 5 are as valid as the SUT shows substantial absorption differences in B1 and B2 as the MC varies. In the case of the considered crop, as described in III, B1 can be chosen for example around one of water absorption peaks, e.g. $\lambda = 1450$ nm, while B2 near the minimum wavelengths analyzed, e.g. $\lambda = 900$ nm. Considering the absorption spectra shown in Fig. 1, a variation in MC of about 40% induces an absorbance variation at $\lambda = 1450$ nm of 0.2 units; this means that the attenuation of the light intensity goes from $\frac{I}{I_0} = (6.3 \cdot 10^{-5})$ at MC = 43.6% to $\frac{I}{I_0} = (2.5 \cdot 10^{-7})$ at MC = 83.9%. At the wavelength of 900 nm, the sensitivity of the absorbance to MC variations is much lower; indeed the attenuation varies by only 0.05 units, which corresponds to a variation of the $\frac{I}{I_0}$ ratio from $(6.3 \cdot 10^{-4})$ to $(2.5 \cdot 10^{-4})$. As it seems reasonable to assume, MC affects photon propagation at wavelengths close to the water absorption peak $\lambda = 1450$ nm while minimally perturbing photons at short wavelengths. These considerations lead us to assume that the approximations in 4 and 5 are reasonable even if the marginal correlation between MC and density will contribute to increasing the uncertainty on their estimations. The use of the logarithmic transformation of absorbance offers another advantage. Linear regression models try to keep constant the absolute fitting error over the considered range. However, as discussed in the introduction, MC and density are useful to determine the real economical value of the crop. It is clear that, for this purpose, having a constant absolute estimation error would lead to unacceptable errors in the price for small quantities of the product. The logarithmic transformation of both the predictors and outcomes can be interpreted as a shift from absolute differences to relative differences, thus, the estimation relative error of the linear regression models is kept constant. A constant relative error estimation therefore guarantees a fair economic value even for a small quantity of product.

### B. Experimental setup

The measurement information relates to the diffused light from the SUT. To acquire the greatest amount of information, the total back-scattered light intensity was measured. For this purpose, the illumination of SUT and collection of the diffused light were performed using an integrating sphere (ARCSphere-50-HAL, ARCoptix S.A., Switzerland). This integrating sphere has an internal diameter of 50 mm and a sample port of 10 mm with a sapphire window. Additional SMA fiber port allowed us to connect an FT-NIR spectrometer (FT-NIR Rocket, ARCoptix S.A., Switzerland) to acquire the diffused light spectrum. SUT was held on the sample port by a specimen holder that allowed changing its density keeping the moisture constant. Fig. 2 shows the scheme of the optical setup, whereas a picture of the whole system including the sample holder is shown in Fig. 3.

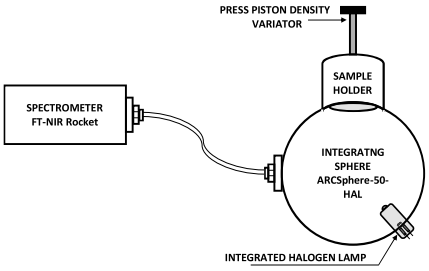



Fig. 2. Block diagram of the experimental setup used to perform the NIRS measurements.

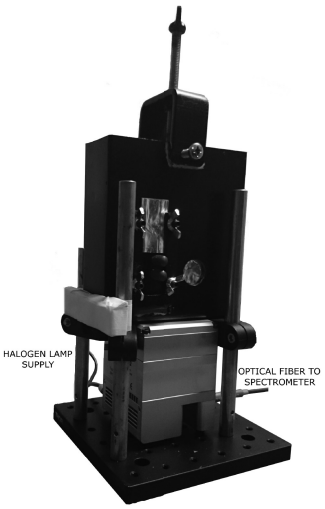



Fig. 3. Picture of the experimental setup. The sample holder allows us to adjust the volume and thus the density of SUT.

### C. Sample collection

Alfalfa grass samples, sown among June, September and October, were obtained from farmland in Modena (Italy). Samples were cut at a length of 2 cm. They were harvested in the days following the cut: the first set of samples was immediately collected, while the other were collected subsequently. The time interval between different samples harvesting was a function of the weather and farmland conditions. This procedure allowed obtaining naturally different MCs ranging from (9.4 $\pm$ 0.4)% to (83.9 $\pm$ 0.4)%. The MC content uncertainty computation will be discussed in the next section. Moreover, the samples were collected from two different fields into the farmland to increase the MC variability. After harvesting, the alfalfa samples were held in a hermetically sealed plastic bag, to preserve MC during the transport to the laboratory. The single SUT was obtained weighting 4.00 g of alfalfa.

### D. Reference methods

Reference gravimetric water content MC was computed according to the American Society of Agricultural and Biological Engineers standard [8]. The initial weight of the SUT $m_{SUT}$ was measured with an electronic weighing balance

(PCB 3500-2 Kern, Balingen, Germany) with resolution r = 0.01 g. Afterwards, SUT was dried at 103 °C for 24 hours and then weighed again to obtain the SUT dried weight $m_d$. The entire procedure was performed in a chemistry laboratory whose ambient humidity and temperature are kept controlled. After sample drying, rehydration of the sample (in particular on the surface of the solid part of the sample) can be neglected having performed the second weighing immediately after the drying operation.
Thus, the gravimetric water content MC was computed as[23]:

$$MC = \frac{m_w}{m_{SUT}}, \tag{6}$$

where $m_w = m_{SUT} - m_d$ is the mass of water in the SUT. The value of MC is given with its relative uncertainty of 0.4%, which is derived from the weight scale accuracy.
The SUT is heterogeneous being composed by stems, leaves and air interstices. This is the normal condition of samples during harvesting. For this heterogeneous sample the density can vary due to the manufacturing processes which differently comprises the crop, reducing or increasing the volume occupied by the air, thus changing the ratio between solid and gaseous part (without any economic value) of the sample. Therefore, reference SUT density was determine estimating the volume of a sample of known mass. The sample was inserted in a varying volume holder; this holder allowed us to adjust the volume occupied by the SUT through a press. The press piston, moved by a screw spindle with indicating gage, was able to reduce the height of the sample room at a defined value h. The procedure was performed considering total volume of the holder including air and solid part of the SUT.
Thus, the SUT density was computed as [24]:

$$\rho = \frac{m_{SUT}}{A_{base} \cdot h}, \tag{7}$$

where $A_{base}$ is the area of the sample holder base. Using this sample holder, the density of the samples was adjusted from 46 kg/m$^3$ to 236 kg/m$^3$ to simulate the range of density values performed by the forage harvesting machines (personal communication from CNH Industrial).

### *E. Measurement procedure*

After the spectrometer warm-up, the dark spectrum was acquired. Afterwards, the halogen lamp was turned on and a 99% NIST Standard Diffuser (Labsphere) was positioned on the sample port. After the lamp warm-up time, the reference spectrum was acquired. Finally, each SUT was placed into the sample holder, its volume was adjusted, and the absorbance spectrum acquired.
The logarithmic transformation of the absorbance spectrum was calculated for each SUT as:

$$\begin{aligned} A(\lambda_i) &= -\log_{10} \frac{I_{SUT}(\lambda_i) - I_D(\lambda_i)}{I_{Ref}(\lambda_i)/0.99 - I_D(\lambda_i)}, \\ LA(\lambda_i) &= \ln A(\lambda_i), \end{aligned} \tag{8}$$

where $I_{SUT}(\lambda_i)$, $I_D(\lambda_i)$ and $I_{Ref}(\lambda_i)$ refer respectively to the scattered spectrum by the SUT, dark spectrum acquired with the halogen lamp turned off and spectrum acquired from 99% NIST Standard Diffuser. $\lambda_i$ ranges between 0.9 $\mu$m and 2.5 $\mu$m with i$\in$(1,..., 1024). Spectral resolution is 8 cm$^{-1}$ in term of wavenumber thus ranges from about 1 nm at 1 $\mu$m to about 5 nm at 2.5 $\mu$m.

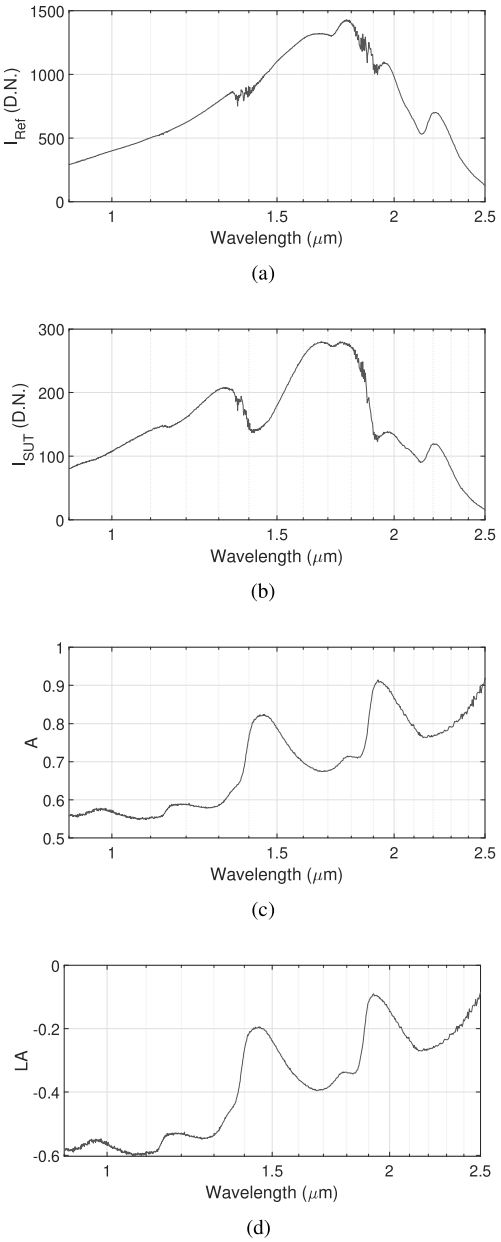



Fig. 4. Reference spectrum acquired from 99% NIST Standard Diffuser (a) and typical SUT spectrum (b). The corresponding absorbance spectrum and its logarithmic transformation are shown in (c) and (d), respectively.

### *F. Data analysis*

Data handling and PLS analysis of the acquired spectra were performed using MATLAB (MathWorks, Sherborn, MA, USA). Acquired spectra were first pre-processed to improve the prediction capability of the regression model.
For the moisture predictor, baseline effects were cancelled calculating the first derivative of the logarithmic transformation

of the absorbance [15][16]:

$$FDLA(\lambda) = \frac{dLA(\lambda)}{d\lambda} = \frac{d\ln A(\lambda)}{d\lambda}. \quad (9)$$

Derivative increased largely the high frequency noise, thus, the Savitzky-Golay (SG) spectral smoothing filter of the second degree was the then applied to maintain the signal-to-noise ratio to a reasonable value. This filtering function was practically performed through the MATLAB command **smooth(FDLA,'sgolay',2)**, where FDLA is the first derived spectrum, 'sgolay' is the Savitzky-Golay method and 2 is the degree of the filter. The resulting spectrum was then normalized to its maximum.

For the density predictor, after several tests, we observed that the first derivative did not improve the quality of the PLS results, and for that reason, we decided to apply the SG filter and the normalization directly to LA.

As an example of data pre-processing, Fig. 5 (a) and (b) show a typical logarithm transformation of the SUT absorbance and the corresponding predictor calculated by SG filtering. The spectral band B2 of this predictor was used to estimate SUT density. Typical FDLA and the corresponding predictor are shown in Fig. 5 (c) and (d), respectively. The spectral bands B1 of this predictor were used to estimate SUT MC.

As described in section II-A, bands B1 and B2 were selected, among the whole spectrum, taking into account the water molar extinction coefficient spectrum [18]. The band B1 considers the two main water absorption peaks in the measured spectral region, 1.38 $\mu$m and 1.88 $\mu$m, whereas B2 refers to the region with minimum absorption $(0.9 - 1.1)\mu$m.

PLS regression analyses in the spectral band B1 and B2 were performed to estimate SUT MC and density $\rho$, respectively. Within spectral bands B1 and B2, the wavelengths of interest were selected analyzing the correlation coefficients between reference and predictor variables calculated as:

$$C(\lambda_j) = \frac{1}{k-1}\sum_{i=1}^{k}\left(\frac{Y_i - \mu_Y}{\sigma_Y}\right)\left(\frac{X_i(\lambda_j) - \mu_{X(\lambda_j)}}{\sigma_{X(\lambda_j)}}\right), \quad (10)$$

where $\mu_Y$ and $\mu_{X(\lambda_j)}$ are the mean, calculated over the k response variables considered, of the response and predictor variable, respectively, whereas $\sigma_Y$ and $\sigma_{X(\lambda_j)}$ are the corresponding standard deviations. The correlation coefficients computation was performed through the MATLAB command **corrcoef(M)**, where M is a matrix composed by the reference values array, MC or $\rho$, in the first column and the spectral values, LA, at each wavelength in the other columns.

The linear regression model used, i.e. Partial Least Square model, is described by the following matrix equation:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ . \\ Y_k \end{bmatrix} = \begin{bmatrix} X_1(\lambda_1) & X_1(\lambda_2) & . & X_1(\lambda_n) \\ X_2(\lambda_1) & X_2(\lambda_2) & . & X_2(\lambda_n) \\ . & & & . \\ . & & & . \\ . & & & . \\ X_k(\lambda_1) & X_k(\lambda_2) & . & X_k(\lambda_n) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ . \\ . \\ . \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_k \end{bmatrix}, \quad (11)$$

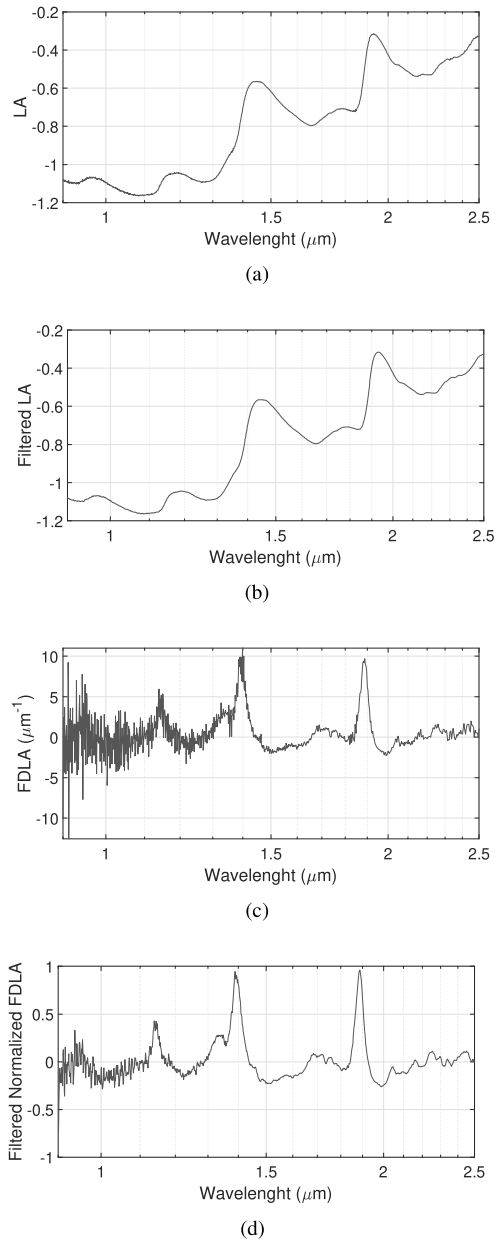



Fig. 5. Example of data pre-processing: logarithm transformation of the SUT absorbance and the corresponding predictor calculated by filtering, (a) and (b), respectively; first derivative of the logarithm transformation of the SUT absorbance and the corresponding predictor calculated by normalization and filtering, (c) and (d), respectively.

where $Y_i$, i.e. response variable, is the $i^{th}$ logarithmic transformation of the reference value of MC or $\rho$ determined according the procedure described in Section II-D. Predictor variable $X_i(\lambda_j)$ is the $i^{th}$ pre-treated FDLA in the case of MC estimation or simply the $i^{th}$ pre-treated LA in the case of $\rho$ estimation; in both the cases, the predictor variable is considered at the selected wavelength $\lambda_j$ at which there is a good correlation between predictor and response variables. $\beta_i$ is the estimated weighted regression coefficient whereas the $\varepsilon$ elements represent the residuals, i.e. the difference between the reference and predicted response. To avoid model over-fitting, regression of (11) is performed on a set of uncorrelated latent

variables obtained finding combinations of original variables that have a large covariance with the response values. This operation is performed thanks to the MATLAB PLS regression algorithm called through the command **plsregress(x,y,n)**, where x is the predictor variable, y is the response variable and n the number of PLS components, i.e. the number of latent variables.
The fittings are evaluated taking into account the statistics: (i) calibration and validation root mean square error (C-RMSE and V-RMSE, respectively), and (ii) the coefficient of determination, $R^2$ (C-$R^2$ and V-$R^2$, respectively).
RMSE is calculated as follows:

$$RMSE = \frac{1}{k}\sum_{i=1}^{k}\sqrt{\sum_{j=1}^{q}\frac{\varepsilon_i^2\mid_j}{q}}, \tag{12}$$

where q is the number of predictors associated to the i-th reference value and $\varepsilon_i^2\mid_j$ is the error between the i-th reference value and the j-th response.
$R^2$ is calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{k}\sum_{j=1}^{q}(Y_i\mid_j - Yref_i)^2}{\sum_{i=1}^{k}(\bar{Y} - Yref_i)^2}, \tag{13}$$

where $Yref_i$ is the reference value associated to the response $Y_i\mid_j$ and $\bar{Y}$ is the average of the $q \times k$ reference values.
After the application of the models, the logarithmic estimations are converted into linear MC and density, and the root mean square relative error between estimated and reference values is calculated for each reference value:

$$\begin{aligned}\varepsilon_r(MC)\% &= \frac{1}{q}\sqrt{\sum_{j=1}^{q}\left(\frac{MC_{est}\mid_j - MC_{ref}}{MC_{ref}}\right)^2}\times 100,\\ \varepsilon_r(\rho)\% &= \frac{1}{q}\sqrt{\sum_{j=1}^{q}\left(\frac{\rho_{est}\mid_j - \rho_{ref}}{\rho_{ref}}\right)^2}\times 100,\end{aligned} \tag{14}$$

where $est$ and $ref$ indicate the estimated and reference values, respectively.

## III. RESULTS

A total of 40 cuttings at different MCs were collected according to the procedure reported in section II-C.From these 40 alfalfa cuttings we extracted 160 different SUTs, which will be compressed differently with the sample holder to obtain, during the measurements, 160 nonidentical and practically independent due to the poor uniformity of the harvested samples.
In the spectral measurement range of the spectrometer, bands B1 was selected including the two main water absorption peaks: (1.3-1.6) $\mu$m and (1.8-2.1) $\mu$m, whereas, for B2, we considered the lowest wavelengths, i.e. (0.9-1.1) $\mu$m, avoiding the first peak of water absorption (1.2 $\mu$m); in this spectral band, water absorption is minimum. The wavelengths of interest were selected analyzing the correlation coefficients shown in Fig. 6.
The number of latent variables, n, chosen for the regression is initially set to 10. The definitive number is defined after several simulation, optimizing the RMSE and the $R^2$ statistics. This process is performed to increase the model accuracy, avoiding the over-fitting. The final number of latent variables is set to 12.

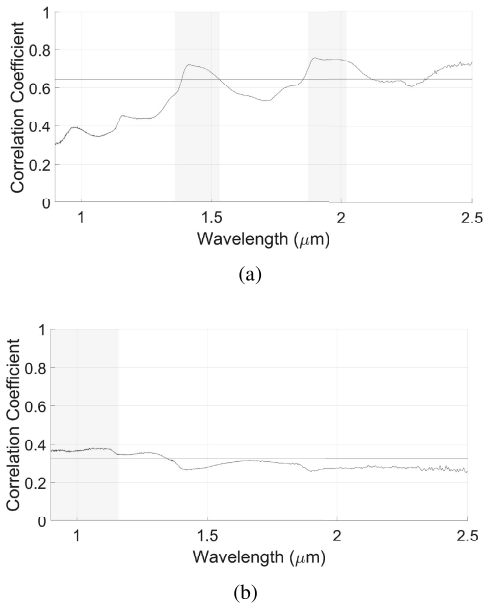



Fig. 6. Correlation coefficients between reference and predictor variables. SUT Moisture content (a) and density (b). Green regions represent the spectral bands B1 (a) and B2 (b). The red lines represent the correlation thresholds used to select the wavelengths. Only the wavelengths in B1 and B2 for which the correlation coefficient is greater than these thresholds are considered in the PLS analysis.

Within the bands B1 and B2, we considered the wavelengths for which the correlation coefficients is grater then 85% of its maximum value, which means 0.65 for MC and 0.4 for density. As expected, in absolute terms, the correction with MC is definitely higher, thanks to the huge absorption contribution due to water. On the other hand, it is known that the density estimation is more critical as it is correlated to the mean scattering coefficient in a highly heterogeneous medium. The low value of the correction coefficient reflects this critical aspect.
Afterwards, a final wavelengths refinement was performed by adding or removing some wavelengths to reduce the prediction errors.
As shown in Fig. 6, some wavelengths outside B1 and B2 could be considered acceptable according to our criterion, i.e. correlation coefficients greater than 85% of its maximum value. These values were not taken into consideration since the correlation is not supported by physical reasons; behaviors of this type are usually not very repeatable and closely linked to the selected set of samples.
Final selected wavelengths for the PLS analysis were;

$$\begin{aligned}B1:\lambda_i &\in (1.37-1.53)\cup(1.87-2.02)\mu m\ i=1...152\\ B2:\lambda_i &\in (0.90-1.16)\mu m\ i=1...310\end{aligned} \tag{15}$$

### A. Crop moisture estimation

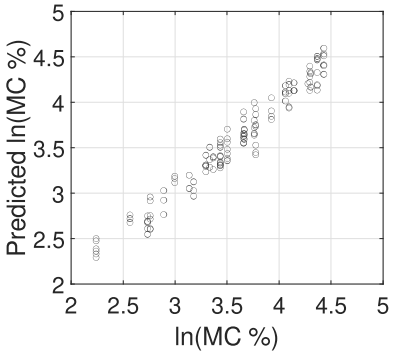
MC reference values range from 9.4% to 83.9%. For each of the k = 40 samples collected, q = 4 spectra at different densities

were acquired, from differently compressed SUTs. Therefore, MC estimation was performed analysing 160 independent predictor variables.

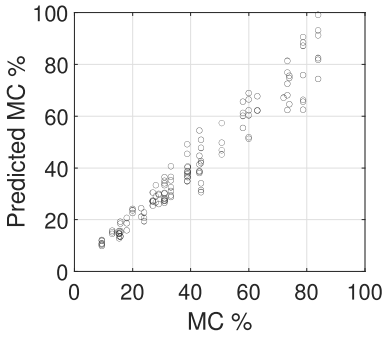
Calibration data set consisted in 128 predictor variables randomly selected. Fig. 7 shows the PLS response variables obtained from calibration as a function of the MC reference values. Validation data set consisted in the remaining 32 predictor variables. Fig. 8 shows the PLS response variables obtained from validation as a function of the MC reference values. Statistics of the developed PLS model are reported in Table I whereas Fig. 9 shows the relative error $\varepsilon_r(MC)\%$, calculated according to Eq.14, between the estimated and reference of MC.

| PARAMETER | FDLA |
|---|---|
| C-RMSE | 0.1230 |
| C-$R^2$ | 0.9581 |
| V-RMSE | 0.1599 |
| V-$R^2$ | 0.9249 |

TABLE I
STATISTICS RELATED TO THE MOISTURE ESTIMATION MODEL PERFORMED USING THE FDLA SIGNALS. C-RMSE AND C-$R^2$ REFER TO THE CALIBRATION MODEL, WHILE V-RMSE AND V-$R^2$ REFER TO THE VALIDATION MODEL.




(a)

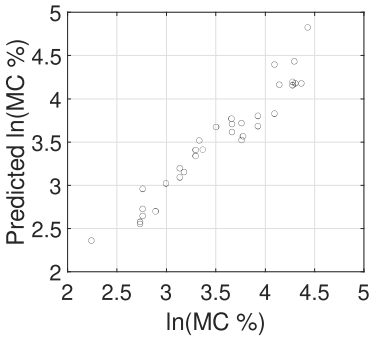



(b)

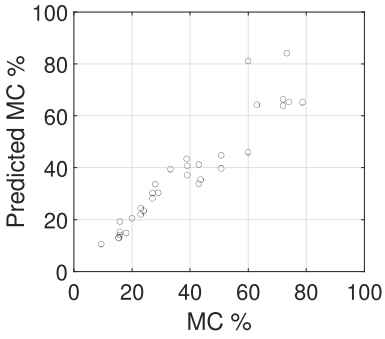
Fig. 7. Moisture content PLS model: calibration response variables as a function of the reference values (a). Predicted vs reference moisture content for the calibration data set (b).

### B. Crop density estimation

Starting from the same 40 samples collected according the procedure reported in section II-C, a new data set was

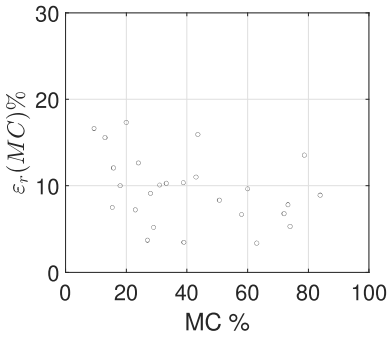



(a)

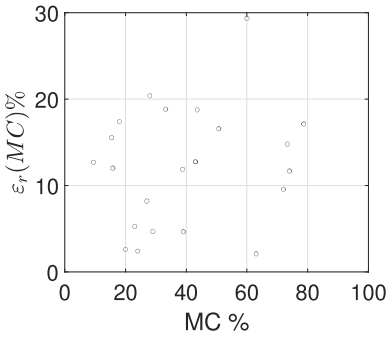



(b)

Fig. 8. Moisture content PLS model: validation response variables as a function of the reference values (a). Predicted vs reference moisture content for the validation data set (b).




(a)




(b)

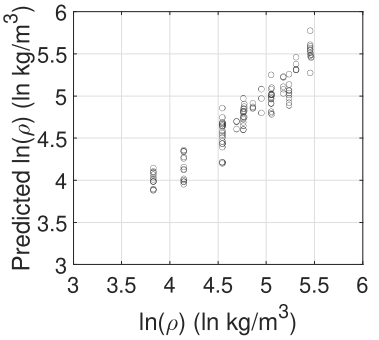
Fig. 9. Root mean square relative error between the estimated and reference values of moisture content. Results obtained for the calibration (a) and validation (b) SUTs.

generated. This set was composed by samples with k = 15 different densities with q moisture contents. q was a variable number depending on the density value, i.e. q = 6,10,15,30. The resulting number of predictor variables was 160. The density values ranged from 46 kg/m$^3$ to 236 kg/m$^3$. Calibration data set was composed by 128 predictor variables randomly selected. PLS response variables obtained from calibration as a function of the density reference values are shown in Fig. 10. As for moisture, the remaining 32 predictor variables composed the validation data set. Fig. 11 shows the PLS response variables obtained from validation as a function of the density reference values.

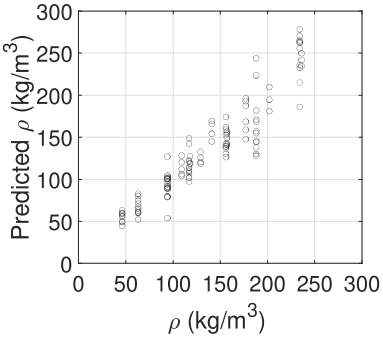
In Table II, the statistics for density estimation are presented whereas Fig. 12 shows the relative error $\varepsilon_r(\rho)\%$, calculated according to Eq.14, between the estimated and reference values of density.

| PARAMETER | LA |
|---|---|
| C-RMSE | 0.1240 |
| $C - R^2$ | 0.9008 |
| V-RMSE | 0.1791 |
| $V - R^2$ | 0.6806 |

TABLE II
STATISTICS RELATED TO THE DENSITY ESTIMATION MODEL PERFORMED USING THE LA SIGNALS. C-RMSE AND C-$R^2$ REFER TO THE CALIBRATION MODEL, WHILE V-RMSE AND V-$R^2$ REFERS TO VALIDATION.

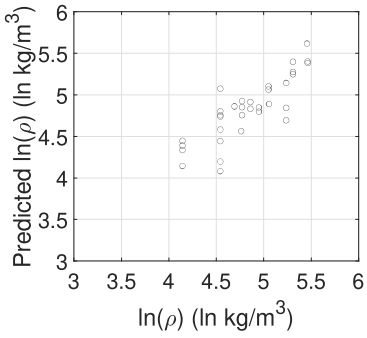



(a)

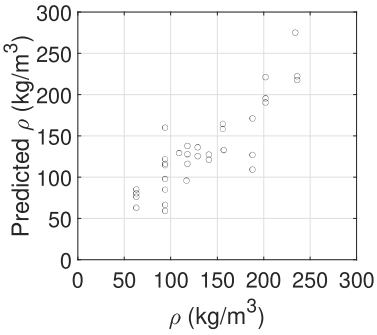



(b)

Fig. 10. Density PLS model: calibration response variables as a function of the reference values (a). Predicted vs reference density for the calibration data set (b).

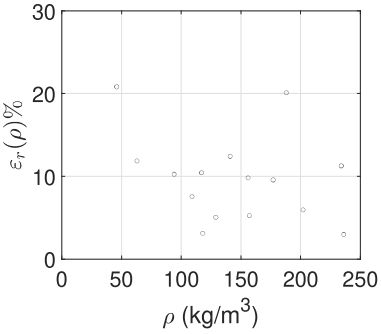



(a)

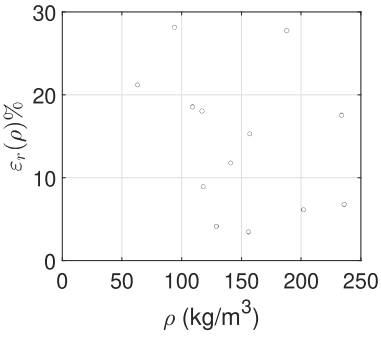



(b)

Fig. 11. Density PLS model: validation response variables as a function of the reference values (a). Predicted vs reference density for the validation data set (b).




(a)




(b)

Fig. 12. Root mean square relative error between the estimated and reference values of density. Results obtained for the calibration (a) and validation (b) SUTs.

## IV. DISCUSSION AND CONCLUSIONS

Determination of the moisture content and density of crop directly during the harvesting remains an ambitious goal in modern agriculture. Numerous critical aspects of the measurement make this objective non-trivial. Contactless NIRS measurements appear to be the most promising approach. However, this measurement technique suffers from numerous influence quantities, among which, the most important is the dependence of the MC information on the density of the sample. The crop density alters directly the photon propagation into the sample and, thus, the acquired spectrum.

The findings presented in this paper demonstrate the feasibility of a PLS multivariate method for simultaneous estimation of crop MC and density. As reported in Table I and II, results obtained on a fresh alfalfa grass samples show a good correlation between estimated and reference quantities, i.e. moisture content and density.

In this paper we introduce, for the first time in this application to our knowledge, the logarithmic transformation of the original variables to separate the spectral contributions due to MC and density. The use of the log transformation of the absorbance offers a second advantage: it makes the estimation relative error quite constant over the fitting range. As discussed in the introduction, moisture and density information are often used to estimate the economic value of the crop. It is clear that, for this purpose, having a constant absolute estimation error would lead to unacceptable errors in the price for small quantities of the product. Our approach overcome this limitation. In fact, as clearly visible for example in Fig. 7, the estimation of MC and density entails a greater spreading of the data at the upper end of the reference quantities range. As shown in Figs. 9 and 12, the PLS root mean square relative errors are practically independent of the reference value and have average values of 13.82% and 14.44% for the estimated MC and density, respectively.

As shown by the experimental results, the model offers better performance in estimating MC. This feature could basically be due to three aspects: (i) lack of homogeneity of the samples that lead to scattering phenomena very dependent on the measurement position, (ii) absorption in B2 dependent not only on density but also on other constituents of the crop [25], (iii) higher uncertainty of the method used to measure the reference density. Note that these results were obtained for a specific type of crop. Considering the simplifications introduced by the modified Lambert Beer's paradigm, the operations described in II-F must be repeated if other crops are considered. In particular, the calibration operations must be performed on a sufficient number of known and independent new samples.

Our ongoing activities are aimed at improving these four aspects. In particular, we are: (i) extending the measurement area to be less sensitive to non-uniformity of the sample, (ii) integrating the B2 band with other spectral bands and (iii) improving the density measurement methods and (iv) increasing the number of samples and density values in order to realize a model able to improve the MC and density estimations by solving the interdependence between these two physical quantities.

Much more work will be needed to refine this model but these first results confirm that NIRS is an excellent candidate for the ambitious goal of realizing a measurement system able of determining crop MC directly during the harvesting.


## ACKNOWLEDGMENT

The authors would like to thank Jacopo Cunial and Luisella Sistu for their contribution during the experimental activities and, Prof. Giovanni Molari for his helpful advice and discussions.


## REFERENCES


[1] J. F. A. F. L.d.S. Ribeiro, F.A. Gentilin and M. Franc¸a, "Development of a hardware platform for detection of milk adulteration based on near-infrared diffuse reflection," *IEEE Transactions on Instrumentation and Measurement*, vol. 65, no. 7, pp. 1698–1706, 2016.

[2] A. R. G. M. Larrain and E. Agosin, "A multipurpose portable instrument for determining ripeness in wine grapes using nir spectroscopy," *IEEE Transactions on Instrumentation and Measurement*, vol. 57, no. 2, pp. 294–302, 2008.

[3] J. Li and L. Sun, "Study on detection methods for frying times of soybean oil based on nirs," in *2019 4th International Conference on Measurement, Information and Control (ICMIC)*, 2019, pp. 83–88.

[4] M. Grossi, E. Valli, V. T. Glicerina, P. Rocculi, T. G. Toschi, and B. Ricc'o, "Practical determination of solid fat content in fats and oils by single-wavelength near-infrared analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 2, pp. 585–592, 2020.

[5] L. Zou, X. Yu, M. Li, M. Lei, and H. Yu, "Nondestructive identification of coal and gangue via near-infrared spectroscopy based on improved broad learning," *IEEE Transactions on Instrumentation and Measurement*, pp. 1–1, 2020.

[6] L. da Silva Dias, J. C. da Silva Junior, A. L. de Souza Maudeira Felìcio, and J. A. de Franc¸a, "A NIR Photometer Prototype With Integrating Sphere for the Detection of Added Water in Raw Milk," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 12, pp. 2812– 2819, Dec. 2018.

[7] G. Salvatori, K. Suh, R. Ansari, and L. Rovati, "Instrumentation and calibration protocol for a continuous wave near infrared hemoximeter," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 4, pp. 1368–1376, Aug. 2006.

[8] "Moisture Measurement - Forages." [Online]. Available: http://elibrary.asabe.org/abstract.asp?aid=41661&t=2

[9] N. Lenzini, L. Rovati, and L. Ferrari, "Effects of the density and homogeneity in NIRS crop moisture estimation," P. Lehmann, W. Osten, and A. Albertazzi Gonc¸alves, Eds., Munich, Germany, June 2017, p. 103294O. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2272443

[10] G. Zaccanti, S. D. Bianco, and F. Martelli, "Measurements of optical properties of high-density media," *Applied Optics*, vol. 42, no. 19, pp. 4023–4030, July 2003. [Online]. Available: https://www.osapublishing.org/ao/abstract.cfm?uri=ao-42-19-4023

[11] E. Shawky and D. A. Selim, "NIR spectroscopy-multivariate analysis for discrimination and bioactive compounds prediction of different Citrus species peels," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 219, pp. 1–7, Aug. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1386142519304019

[12] H. Zhang, Y. Li, and L. Jiang, "Multivariate Modeling Dahurian Larch Plantation Wood Density Based on Near Infrared Spectroscopy," in *2010 International Conference on Measuring Technology and Mechatronics Automation*. Changsha City, China: IEEE, Mar. 2010, pp. 748–751. [Online]. Available: http://ieeexplore.ieee.org/document/5459773/

[13] *P. Fornasini, The Uncertainty in Physical Measurements*. New York, NY: Springer New York, 2008. [Online]. Available: http://link.springer.com/10.1007/978-0-387-78650-6

[14] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, Oct. 2001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0169743901001551

[15] E. Szłyk, A. Szydłowska-Czerniak, and A. Kowalczyk-Marzec, "NIR Spectroscopy and Partial Least-Squares Regression for Determination of Natural -Tocopherol in Vegetable Oils," *Journal of Agricultural and Food Chemistry*, vol. 53, no. 18, pp. 6980–6987, Sept. 2005. [Online]. Available: https://pubs.acs.org/doi/10.1021/jf050672e

[16] F. Garcìa-Sànchez, L. Galvez-Sola, J. J. Martìnez-Nicolàs, R. Muelas-Domingo, and M. Nieves, "Using Near-Infrared Spectroscopy in Agricultural Systems," in *Developments in Near-Infrared Spectroscopy*, K. G. Kyprianidis and J. Skvaril, Eds. InTech, Mar. 2017. [Online]. Available: http://www.intechopen.com/books/developments-in-near-infrared-spectroscopy/using-near-infrared-spectroscopy-in-agricultural-systems

[17] M. Mesic, V. Corluka, and Z. Valter, "Analysis of some parameters influencing moisture quantity measurements in wheat with NIR technique," in *2005 18th International Conference on Applied Electromagnetics and Communications*, Oct. 2005, pp. 1–4.

[18] K. F. Palmer and D. Williams, "Optical properties of water in the near infrared," *J. Opt. Soc. Am.*, vol. 64, pp. 107–110, 1974.

[19] F. Crespi, A. Bandera, M. Donini, C. Heidbreder, and L. Rovati, "Noninvasive in vivo infrared laser spectroscopy to analyse endogenous oxyhaemoglobin, deoxy-haemoglobin, and blood volume in the rat cns," *Journal of Neuroscience Methods*, vol. 145, no. 1-2, pp. 11–22, 2005.

[20] T. Saitoh, A. Ooue, N. Kondo, K. Niizeki, and S. Koga, "Active muscle oxygenation dynamics measured during high-intensity exercise by using two near-infrared spectroscopy methods," *Adv Exp Med Biol.*, vol. 662, pp. 225–230, 2010.

[21] W. Baker, A. Parthasarathy, D. Busch, R. Mesquita, and A. Greenberg, JHand Yodh, "Modified beer-lambert law for blood flow," *Biomed Opt Express*, vol. 5, no. 11, pp. 4053–4075, 2014.

[22] L. Rovati, A. Bandera, M. Donini, G. Salvatori, and L. Pollonini, "Design and performance of a wide-bandwidth and sensitive instrument for near-infrared spectroscopic measurements on human tissue," *Review of Scientific Instruments*, vol. 75, no. 12, pp. 5315–5325, 2004. [Online]. Available: https://doi.org/10.1063/1.1818588

[23] R. B. McIntosh and M. E. Casada, "Fringing Field Capacitance Sensor for Measuring the Moisture Content of Agricultural Commodities," *IEEE Sensors Journal*, vol. 8, no. 3, pp. 240–247, Mar. 2008, conference Name: IEEE Sensors Journal.

[24] W. Meyer and W. Schilz, "Feasibility Study of Density-Independent Moisture Measurement with Microwaves," *IEEE Transactions on Microwave Theory and Techniques*, vol. 29, no. 7, pp. 732–739, July 1981, conference Name: IEEE Transactions on Microwave Theory and Techniques.

[25] N. Brogna, M. T. Pacchioli, A. Immovilli, F. Ruozzi, R. Ward, and A. Formigoni, "The use of near-infrared reflectance spectroscopy (nirs) in the prediction of chemical composition and in vitro neutral detergent fiber (ndf) digestibility of italian alfalfa hay," *Italian Journal of Animal Science*, vol. 8, no. sup2, pp. 271–273, 2009. [Online]. Available: https://doi.org/10.4081/ijas.2009.s2.271