Evaluation of a combination of NIR micro-spectrometers to predict chemical properties of sugarcane forage using a multi-block approach

Maxime Ryckewaert, Gilles Chaix, Daphné Héran, Abdallah Zgouz, Ryad Bendoula

Evaluation of a combination of NIR micro-spectrometers to predict chemical properties of sugarcane forage using a multi-block approach

Maxime Ryckewaert[a,d,e], Gilles Chaix[b,c,d,e], Daphné Héran[a], Abdallah Zgouz[b,c], Ryad Bendoula[a,e]

[a] *ITAP, Univ Montpellier, INRAE, Institut Agro, Montpellier, France*
[b] *CIRAD, UMR AGAP Institut, F-34398 Montpellier, France*
[c] *UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, F-34398 Montpellier, France*
[d] *ChemHouse Research Group, Montpellier, France*
[e] *HélioSPIR,361 Rue Jean-François Breton, 34196 Montpellier, France*

Abstract

Forage quality is essential in livestock farming and has an important role in the functioning of agricultural farms.

Access to biochemical variables provides an estimation of the feed value of crop for animal feed at harvest. Near infrared (NIR) spectroscopy provides measurements indirectly related to biochemical variables. In recent years, several micro-spectrometers have been developed that offer the opportunity to predict such biochemical variables at low cost. In this study, the potential of a combination of micro-spectrometers is evaluated to predict crude protein (CP) and total sugar content (TS) of sugarcane. First, each micro-spectrometer with optimal pretreatments was individually compared to a reference laboratory spectrometer. Then, a combination of micro-spectrometers is proposed and prediction models were established by a multi-block method from data fusion called Sequential and Orthogonalised - Partial Least Squares

(SO-PLS). For CP, the combination of micro-spectrometers provides model (sep=0.69%; bias=0.15%; R^2_{test}=0.910) close to those obtained with the reference spectrometer (sep=0.56%; bias=-0.13%; R^2_{test} =0.935). For TS, the results obtained with this combination of micro-spectrometers (sep=2.38%; bias=-0.52%; R^2_{test}=0.983) are better than those obtained with the reference spectrometer (sep=2.59%; bias=0.41%; R^2_{test}=0.978). For both chemical variables, the combination of the micro-spectrometers significantly increases the performance of the predictive models compared to the models obtained with the micro-spectrometers independently. Using several low-cost micro-spectrometers, combined with a multi-block method would give results as good as a single laboratory spectrometer with a lower cost.

Keywords: Food control, Micro-spectrometer, Spectroscopy, Data fusion, Forage, multi-block regression, Multivariate Data Analysis

---

1. Introduction

Forage quality is essential in livestock farming and has an important role in agricultural farm management (Ball et al., 2001; Collins and Fritz, 2003). Forage must respond to a set of constraints related to farms, production costs, animal requirements and environment (Wilkins, 2000). Forage feed value, including energy value or protein and mineral contents, ensures a nutritional quality for a good metabolic development of animals. Information such as protein or sugar content represents a major interest to estimate feed value of crops at harvest time. Accessing these parameters is possible by direct destructive laboratory measurements (Ball et al., 2001). However, these measurements have time and cost constraints.

In laboratory, near-infrared (NIR) spectroscopy is an alternative to access these parameters in an indirect and non-destructive manner (Stuth et al., 2003; Deaville and Flinn, 2000; Barton II and Windham, 1988). In the NIR range, spectral bands are related to harmonics and combinations of fundamental molecular vibrations, in particular stretching, bending and some deformations (Siesler et al., 2008; Workman and Springsteen, 1998).

In recent years, several micro-spectrometers have been developed (Yang et al., 2021). These micro-spectrometers provide fast and non-destructive measurements with a very low cost compared to laboratory spectrometers. With this technology, the increased use of NIR spectroscopy is expected to lead to new applications (Yan and Siesler, 2018; Wiesner et al., 2014; Siesler et al., 2008) directly accessible to crop producers.

To this end, micro-spectrometers are expected to be widely used. Hence, appropriate multivariate data analysis methods must be proposed to exploit spectral data from NIR spectroscopy (Wiesner et al., 2014). The reference method is Partial Least Squares Regression (PLS-R)(Wold et al., 2001) which is a bilinear regression method that allows to predict biochemical variables from spectral data.

Generally, prediction quality of regression models can be improved by choosing the best pretreatment according to variables to be predicted, spectral region considered and undesired spectra to be corrected (Engel et al., 2013; Rinnan et al., 2009). Another way to increase predictive capabilities is to predict a response variable from several complementary data sources (blocks), using so-called multi-block methods (Mishra et al., 2021). Recently, Sequential and Orthogonalised - Partial Least Squares (SO-PLS) was proposed as an extension of PLSR (Naes et al., 2011) involving an orthogonalisation procedure to sequentially capture additional information from different blocks. This category of methods offers the possibility to predict a response variable from a combination of several blocks such as NIR measurements combining with physicochemical parameters for the monitoring of anaerobic digestion (Awhangbo et al., 2020).

In this study, the potential of a combination of micro-spectrometers is evaluated to predict chemical variables of sugarcane for forage application. The main objective is to study the contribution of a multi-block method to exploit spectra resulting from a combination of a set of micro-spectrometers. In a first step, PLS models are established for each micro-spectrometer with optimal pretreatments. These models are compared to a model from a reference laboratory spectrometer. In a second step, a combination of micro-spectrometers is proposed and SO-PLS is used to build a prediction model.

2. Materials and methods

2.1. Sample preparation and reference analysis

A set of sixty sugarcane samples from different plant parts (leaf, stem or whole aerial part) were collected in the French West Indies (Guadeloupe) (Zgouz et al., 2020). Before chemical analysis, the samples were dried for 72h at 85°C, milled with a Retsch SM100 mill (Retsch GmBH,Germany) with a 1 mm exit sieve and analysed in the Cirad Selmet feed laboratory (Montpellier, France) to determine total sugar content (TS) and Crude Protein content (CP). CP content was estimated from the total nitrogen content (N) measured by Kjeldahl method, with the relationship CP = N * 6.25 and TS content was determined by adapted Luff-Schoorl method (noa, 1997). CP and TS are expressed as a percentage (%) of Dry Matter (DM).

2.2. Spectral measurement protocols

All samples were measured by a laboratory spectrometer used as reference: LabSpec 4 (ASD, Boulder, CO, USA).

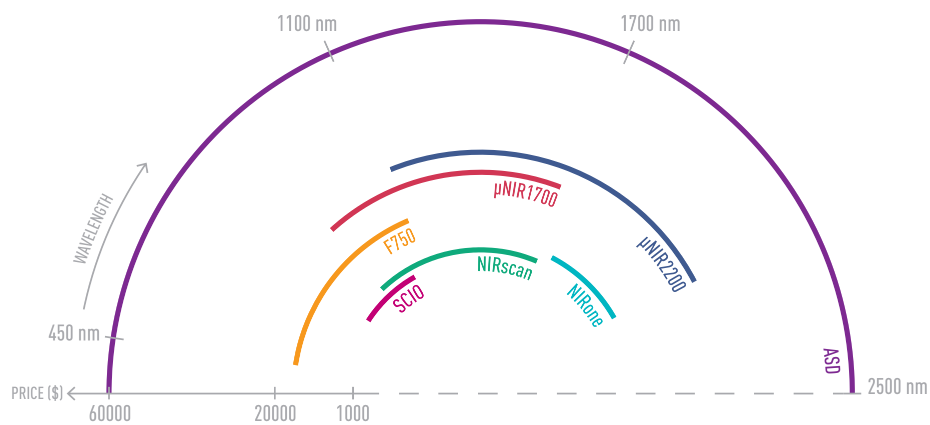


Figure 1: Spectral range according to approximate price for all spectrometers

Spectral acquisitions were also performed on the same samples (in a same FOSS cup) using different micro-spectrometers: SCIO (Consumer Physics, Israel), F750 (Felix Instrument, Camas, USA), µNIR1700 (Viavi, Solution-Milpitas, USA), DLP NIRscan Nano (Texas Instrument s Inc., Texas, USA), µNIR2200 (Viavi, Solution-Milpitas, USA) and NIRONE 2.2 (Spectral Engines, Finland). Fifty spectra were averaged for each spectrometer acquisition.

These micro-spectrometers, covering different visible and near-infrared spectral ranges (see Table 1), were used to establish predictive models and were compared with the ASD Labspec 4 which covers a much larger spectral range. Approximate price values of all spectrometers are displayed for comparison purposes (see Fig. 1). The lowest priced micro-spectrometers were the NIRone, NIRscan and SCIO with values close to 1000$. The reference spectrometer (ASD) was about 60x more expensive compared to these micro-spectrometers. µNIR1700 and µNIR2200 have intermediate values corresponding to 20x more expensive than NIRone, NIRscan and SCIO. The F750 is a compromise (8000$) between the micro-spectrometers in terms of price.

Table 1: Detail for each spectrometer/micro-spectrometers: acronym used on the document, spectral range and spectral resolution of the NIR spectrometers

Acronym	Manufacturer	Device model	Spectral range (nm)	Spectral resolution (at λ) (nm)
ASD	ASD Inc.	LabSpec 4	350 - 2500	3 (at 700 nm) - 10 (at 1300/2100 nm)
SCIO	Consumer Physics	SCIO	740 - 1070	not communicated
F750	Felix Instrument	F750	450 - 1140	8
µNIR1700	JDSU/VIAVI	MicroNIR1700	908 - 1676	10
µNIR2200	JDSU/VIAVI	MicroNIR2200	1150 - 2150	20
NIRscan	Texas Instrument	DLP NIRscan Nano	901 - 1701	10
NIRone	Spectral Engines	NIRone 2.2	1750 - 2150	20-26

2.3. Data analysis

2.3.1. Regression models

In the first section, Partial Least Squares Regression (PLSR) (Wold et al., 2001) was used to build models to predict chemical variables, from spectra, represented by a matrix $\mathbf{X}$. Each chemical variable is represented by a vector $\mathbf{y}$. A model was established for each micro-spectrometer where the final PLS equation can be established as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{r}_X \tag{1}$$

where $\mathbf{b}$ is a vector containing regression coefficients and $\mathbf{r}_X$ is corresponding to residuals of the model.

In the second section, Sequential and Orthogonalized Partial Least Squares (SO-PLS) (Naes et al., 2011) was used as a multi-block method to predict variables from multiple blocks. This method was used to evaluate a combination of micro-spectrometers, i.e. blocks corresponding to the spectra measured by these micro-spectrometers.

This method extracts the information sequentially from each data block. First, the SO-PLS algorithm started as PLS method with the first block containing spectral data, as previously described (eq. 1).

Then, an orthogonalisation procedure was performed to remove information (from the first regression) on the second block, defined by the matrix $\mathbf{Z}$. This orthogonalisation, providing $\mathbf{Z}_\perp$, can be written as follows:

$$\mathbf{Z}_\perp = \mathbf{Z} - \mathbf{T}(\mathbf{T}^t\mathbf{T})\mathbf{T}^t\mathbf{Z} \tag{2}$$

where $\mathbf{T}$ corresponds to the score matrix of $\mathbf{X}$ in a PLS procedure.

Then, a second PLS model is established between the residual vector, corresponding to the vector $\mathbf{r}_X$ (eq. 1) and the matrix $\mathbf{Z}_\perp$. This regression is established by following the same procedure as previously for the regression between $\mathbf{X}$ and $\mathbf{y}$ (eq. 1). At the end of this procedure, a vector $\mathbf{c}$ containing the regression coefficients is obtained. The final equation of the SO-PLS multi-block method can be written as follows:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_{\perp}\mathbf{c} + \mathbf{r}_{X,Z} \tag{3}$$

With $\mathbf{r}_{X,Z}$, the residual vector of the SO-PLS model.

2.3.2. Model evaluation

To evaluate model performances, a calibration set and a test set from available samples were defined by random selection. The calibration set was constituted by using two thirds of the samples, i.e. forty samples. Spectra corresponding to these 40 samples were used to build the prediction model. This step was performed in k-fold cross validation (five blocks) to select the relevant number of latent variables (LV).

Spectra from the 20 remaining samples, corresponding to the remaining third of the whole available samples, were used as an internal test set. Each model obtained in cross-validation procedures was applied to the test set.

To evaluate SO-PLS models, the order of the blocks was defined in the order of the spectral ranges (i.e. the lowest spectral range corresponds to the first block).

2.3.3. Pretreatments

Pretreatments commonly used in chemometrics were tested to establish the best prediction models: Standard Normal Variate (SNV) (Barnes et al., 1989), Variable Sorting for Normalization (VSN) (Rabatel et al., 2020) and Multiple Scatter Correction (MSC). These corrections were also combined with a Savitzky-Golay type smoothing (Savitzky and Golay, 1964) by varying the window size with common values corresponding to 20 nm to 400 nm as well as varying the polynomial order from 1 to 3. In addition, the first and second derivatives were also studied to these Savitzky-Golay smoothings.

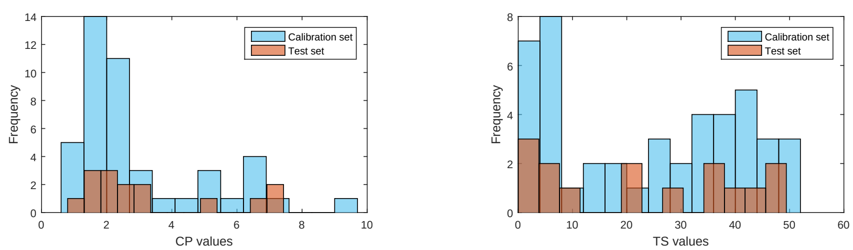
A choice on the best pretreatment (including no pretreatment) was made for each spectrometer according to several criteria from cross-validation procedure. More especially, this choice was motivated by minimizing validation errors while selecting a low number of latent variables.

3. Results and discussion

3.1. Data overview

3.1.1. Y values

For each variable describing the chemical properties, value distributions for the calibration set and the test set are displayed in figure 2.



(a) (b)

Figure 2: Histograms of calibration and test sets for each biochemical variable: (a) Crude protein content (CP, %DM), (b) Sugar content (TS, % DM)

For the calibration set, CP content (Fig. 2a) has values ranging from 1% to 10% of total dry matter with a distribution mainly between 1% and 4% and very few values above 4%. The calibration set value distribution of TS content (Fig. 2b) ranges from 0% to 50% with many observations having either low values between 0% and 5% or values around 40%.

For both chemical variables, test sets have similar value distributions than those of the corresponding calibration set. The visualization of value distributions confirms that the test set covers the whole value range.

3.1.2. Spectra analysis

Pseudo-absorbance spectra defined by log(1/R) (where R is the reflectance spectrum measured by a spectrometer) of all calibration and test sets are shown for each spectrometer (Fig. 3).

Spectra from ASD have a spectral range from 350 nm to 2500 nm (Fig. 3a). These spectra have shapes consistent with what is generally found in NIR spectroscopy of fruits and vegetables (Nicolaï et al., 2014): NIR spectra are dominated by water contribution. Two water-related absorption bands can be identified: 1436 nm, 1938 nm. A small peak at 1200nm can be observed. The same observations are made for the spectra measured in the near infrared range with the other spectrometers (Fig. 3d, 3e, 3f and 3g). A visible base line shift effect was observed in absorbance raw spectra, for all spectrometers. The increase in the optical path length traveled by the photons in a scattering medium reflects a multiplicative effect on the spectra (Osborne et al., 1993; Ryckewaert et al., 2020) resulting in baseline drifts of the ideal absorbance spectra. In the spectral range between 350 and 1000 nm, these scattering effects are much more dominant (Fig. 3a) and 3b) and result in a decreasing slope (Ishimaru, 1978). Weaker absorption peaks are present at 670 nm (Fig. 3a and 3b) and 1200 nm (Fig. 3a, 3d, 3e and 3f). In vegetable products, the absorption bands at 670 nm and 1200 nm have been

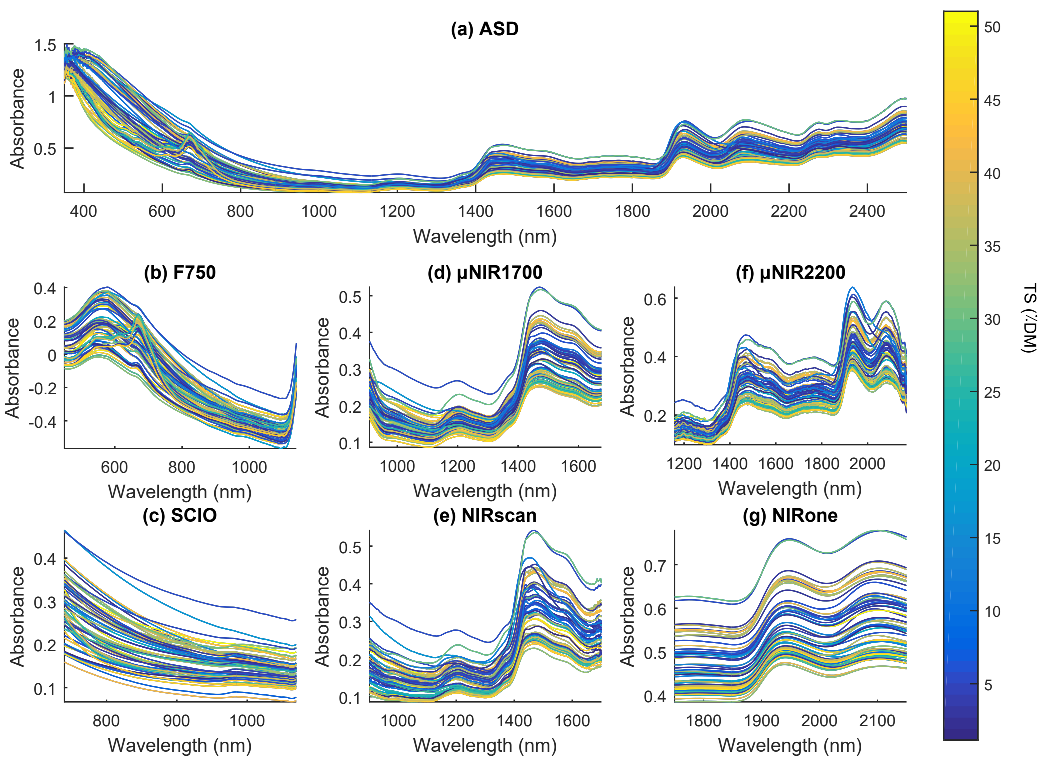


Figure 3: Absorbance spectra for each spectrometer: (a) ASD, (b) F750, (c) SCIO, (d) µNIR1700, (e) NIRscan, (f) µNIR2200, (g) NIRone

respectively associated with chlorophyll and sucrose (Osborne et al., 1993).

3.2. Regression model by spectrometer

3.2.1. Pretreatment and calibration model

The best models obtained by cross-validation to predict CP and TS variables by micro-spectrometers/spectrometer are presented in Table 2. For each variable (CP and TS), results are classified according to the error values obtained in cross-validation (*i.e.* secv).

Table 2: Calibration results for prediction of the two variables: CP and TS. The spectrometers are ranked in order of best to worst model based on SECV values. ; savgol (width (in nm) /deriv/order)

Y	Spectrometer	pretreatment	LV	sec (%DM)	secv (%DM)	R^2_{cv}
CP	μNIR2200	savgol(245/2/2) - SNV	9	0.29	0.51	0.942
	NIRone	SNV	9	0.45	0.55	0.929
	ASD	savgol(41/0/0)	10	0.40	0.57	0.926
	μNIR1700	savgol(185/1/2) - SNV	12	0.49	0.64	0.930
	F750	savgol(75/2/2)	3	0.65	0.73	0.87
	SCIO	savgol(90/1/1)	5	0.64	0.75	0.87
	NIRscan	savgol(53/2/2) - SNV	8	0.51	0.99	0.77
TS	ASD	savgol(41/0/0) - SNV	8	1.15	1.83	0.989
	μNIR2200	savgol(203/0/0) - SNV	9	1.48	2.46	0.981
	μNIR1700	savgol(155/1/1) - SNV	8	2.37	3.40	0.960
	NIRscan	savgol(88/2/2)	13	1.84	4.24	0.939
	NIRone	savgol(30/2/2) - SNV	9	3.00	6.31	0.87
	F750	savgol(105/1/1) - SNV	13	4.17	6.42	0.87
	SCIO	savgol(45/1/1)	4	7.79	8.33	0.76

For the CP variable, best models from cross-validation procedure are obtained with the μNIR2200 micro-spectrometers (secv=0.51%; R^2_{cv}=0.942), the NIRone (secv=0.55%; R^2_{cv} =0.929), the ASD (secv=0.57%; R^2_{cv} =0.926) and the μNIR1700 (secv=0.64%; R^2_{cv} =0.930). The models obtained with the NIRscan (secv=0.99%; R^2_{cv}=0.77), the SCIO (secv=0.75%; R^2_{cv} =0.87) and the F750 (secv=0.73%; R^2_{cv}=0.87) have higher errors and a lower R^2_{cv}.

For the TS variable, the best model from the cross-validation is obtained with the ASD (secv=1.83%; R^2_{cv} =0.989). The best models using micro-spectrometers are obtained with the μNIR2200 (secv=2.46%; R^2_{cv} =0.981) and the μNIR1700 (secv=3.40%; R^2_{cv} =0.960). The worst models are obtained with the SCIO (secv=8.33%; R^2_{cv} =0.76), the F750 (secv=6.42%; R^2_{cv} =0.87) and the NIRone (secv=6.31%; R^2_{cv} =0.87) The NIRscan (secv=4.24%; R^2_{cv} =0.939) shows intermediate results.

In most cases, best models are obtained with savgol smoothing of different window sizes depending on the spectrometer used and their respective spectral resolution. The VSN and MSC pretreatments did not give optimal results and do not appear in Table 2. This smoothing is combined with SNV pretreatment for μNIR2200, μNIR1700 and NIRscan in the case of CP prediction. SNV is also combined with smoothing on the μNIR2200 , μNIR1700 and NIRone and F750 for the TS variable.

Different pretreatments are obtained for micro-spectrometers, even to predict the same response variable. This is due to different characteristics of the spectrometer used (spectral region, spectral resolution, noise). For a given spectrometer, pretreatments can be different depending on the variable to be predicted. Indeed, two chemical variables of different nature will impact differently reflectance spectrum shape.

3.2.2. Model evaluation

Once optimal pretreatment and LV number are chosen during the cross-validation procedure, models are calibrated with the entire calibration set samples. (Tab. 2). These models are now applied to the test set. For the two chemical variables, performances of the prediction models are evaluated through sep and R^2_{test} values. These values are obtained for each spectrometer or micro-spectrometer and can be seen in table 3.

Table 3: Evaluation of prediction models on a test set. Spectrometers are ranked in order of best to worst model based on SEP value

Y	Spectrometer	sep (%DM)	bias (%DM)	R^2_{test}
CP	ASD	0.56	-0.13	0.935
	µNIR1700	0.58	0.18	0.932
	µNIR2200	0.67	0.20	0.908
	SCIO	0.72	-0.31	0.902
	NIRone	0.73	-0.027	0.88
	NIRscan	0.78	0.28	0.88
	F750	0.86	-0.30	0.87
TS	ASD	2.59	0.41	0.978
	µNIR2200	2.99	0.31	0.970
	µNIR1700	3.01	0.81	0.969
	NIRscan	4.40	0.55	0.939
	NIRone	8.04	-1.64	0.78
	F750	9.37	0.49	0.73
	SCIO	11.05	3.80	0.61

CP. For CP variable, R^2_{test} values range from 0.87 to 0.94 and error values range from 0.56 to 0.86%. The range of prediction error values needs to be compared with the protein content values, which range from 1 to 7%, but with the majority of values between 1 and 3% (Fig. 2a). The best model is obtained with the ASD (sep = 0.56%; bias=-0.13%; R^2_{test} = 0.935). The micro-spectrometers with values close to those obtained with the ASD are the µNIR1700 (sep =0.58%; bias=0.18%; R^2_{test}=0.932) and the µNIR2200 (sep =0.67%; bias=0.20%; R^2_{test}=0.908). Models with intermediate results are obtained with SCIO (sep =0.72%; bias=-0.31%; R^2_{test}=0.902), NIRone (sep =0.73%; bias=-0.0027%; R^2_{test}=0.88) and NIRscan (sep =0.78%; bias=-0.28%; R^2_{test}=0.88). The worst performance is obtained for the model using the F750 (sep =0.86%; bias=-0.30%; R^2_{test}=0.87)

The µNIR1700 has a spectral region defined between 908 nm and 1676 nm. This spectral region contains a part related to overtones of the C-H, C-N, N-H bonds present between 1600-1700 nm and related to protein (Clark and Lamb, 1991). Besides, predictive quality obtained with the NIRone remains satisfactory despite a very small spectral range defined between 1750 nm and 2150 nm. Nitrogen-Hydrogen (N-H) bonds absorb at 2055 nm and 2180 nm (Wetzel, 1983). In addition, proteins contain mostly amide structures that possess nitrogen-hydrogen (N-H) bonds. The NIRone spectral region is therefore suitable to predict protein content, as absorption peak at 2055 nm can be observed as well as the beginning of the absorption peak at 2180 nm.

Almost equivalent results are reached with SCIO despite a restricted spectral range from 740 nm to 1070 nm. In this range, information is related to protein at 1007 nm or to primary amines at 1000 nm and 1020 nm (Workman and Springsteen, 1998) .

TS. For TS (Fig. 3), R^2_{test} and sep values have a wider range. These values range from 0.61 to 0.98 for R^2_{test} and from 1.81 to 11.05% for sep. Best models on the test set were obtained for the ASD (sep =2.59%; bias=0.41%; R^2_{test}=0.978) and the µNIR2200 (sep =2.99%; bias=0.31%; R^2_{test}=0.970). Intermediate results were obtained for the µNIR1700 (sep=3.01%; bias=0.81%; R^2_{test}=0.969) and the NIRscan (sep=4.40%; bias=0.55%; R^2_{test}=0.939). On the other hand, models derived from the spectra measured by NIRone (sep=8.04%; bias=-1.64%; R^2_{test}=0.78), F750 (sep=9.37%; bias=0.49%; R^2_{test}=0.73) and SCIO (sep=11.05%; bias=3.80%; R^2_{test}=0.61) have a poor predictive quality.

Good and intermediate results (μNIR2200, ASD, μNIR1700 and NIRscan) are reached with spectrometers whose spectral ranges cover the region around O-H bond from sugar as crystalline sucrose (Kays et al., 1997) around 1441 nm (Workman and Springsteen, 1998). This is not the case for the NIRone, F750 and SCIO spectrometers (see 1), and may explain the high error values in TS prediction.

Best results are obtained with ASD and the μNIR2200. These spectrometers cover both spectral regions around 1441 nm, as well as around 2100 nm. At 2100 nm, O-H bending and C-O stretching combination can be observed and can be related to sugar content (Workman and Springsteen, 1998) and could explain the good results obtained.

3.3. Regression model from a combination of spectrometers

Table 4 shows prediction model evaluations for CP and TS using ASD and using a combination of micro-spectrometers. Micro-spectrometers retained for the combination are the following: SCIO, NIRscan and NIRone. These micro-spectrometers were chosen because their combination covers approximately the same spectral range as the ASD while minimising the cost compared to other micro-spectrometers (see fig. 1). Pretreatments used for each micro-spectrometer correspond to those identified by the micro-spectrometers independently (see table 2).

Table 4: Evaluation of prediction performance for all variables using ASD and using the combination of three micro-spectrometers (SCIO; NIRscan; NIRone). The latent variables displayed in the format '../../..' correspond to the latent variables of the micro-spectrometers in the order: SCIO/NIRscan/NIRone

Y	Spectrometer(s)	pretreatment	LV	sec	secv	sep	bias	R^2_{cv}	R^2_{test}
CP content	ASD	savgol(40/0/0)'	10	0.40	0.57	0.56	-0.13	0.926	0.935
	Combination	Best each	3/2/10	0.29	0.47	0.69	0.15	0.952	0.910
TS content	ASD	savgol(40/0/0)	8	1.15	1.83	2.59	0.41	0.989	0.978
	Combination	Best each	0/6/6	1.67	2.56	2.38	-0.52	0.978	0.983

Results obtained previously with ASD are reported here as reference values. For CP variable, the optimal prediction model remains the one using spectra obtained with ASD (sep=0.56%; bias=-0.13 %; R^2_{test} =0.935). Nonetheless, the micro-spectrometer combination provides a model with performances (sep=0.69%; bias=0.15 %; R^2_{test}=0.910) close to those obtained with ASD.

For TS, results obtained with this micro-spectrometer combination (sep=2.38%; bias=-0.52%; R^2_{test}=0.980) are better than those obtained with ASD (sep=2.59%; bias=0.41%;R^2_{test}=0.978). Combining the three micro-spectrometers leads to lower prediction errors than using each micro-spectrometer separately, for both variables tested. Indeed, the best performances previously obtained for CP prediction were sep=0.72% and R^2_{test}=0.88 with the NIRone, whereas TS prediction performances were sep=4.40% and R^2_{test}=0.939 with the NIRscan. The combination of sensors greatly improves the predictive qualities for the variables studied. However, this proposed micro-spectrometer combination does not reach the performances obtained with µNIR2200 for CP prediction (sep =0.65%; R^2_{test}=0.926) as well as TS prediction (sep =2.36%; R^2_{test}=0.985) (see Tab. 3). Nevertheless, the three-spectrometer combination has a lower cost than that of the μNIR2200 alone.

In multi-block methods, the number of latent variables is defined by cross-validation for each of the blocks (i.e. micro-spectrometers). This number varies according to the relative importance of each block to predict a given variable. Visualising the number of latent variables helps understanding the relevance of each micro-spectrometer and would be a guided way to select the best combination of micro-spectrometers according to predictive capabilities. Indeed, if the number of latent variables is equal to zero, this means that the micro-spectrometer is not considered in the multi-block model. This is the case for the SCIO micro-spectrometer which is not used for TS prediction (table 4). In this case study, the combination of only two micro-spectrometers (NIRscan and NIRone) would be sufficient to predict TS at a considerably lower cost.

Here we have chosen the pretreatments defined separately for these three identified micro-spectrometers. However, it is recommended to integrate the pretreatment choice into the cross-validation procedure of SO-PLS to ensure better complementarity between blocks and thus improve the prediction capabilities. An alternative is to systematically add blocks corresponding to relevant pretreatments for each micro-spectrometer. This alternative would have the capacity to be automatic but would impose new constraints in terms of computing time and memory space.

4. Conclusion

In this study, micro-spectrometers were evaluated individually to predict Crude Protein (CP) and sugar content (TS) on sugarcane forage samples. Optimal pretreatments were identified. For a micro-spectrometer, resulting pretreatment may differ according to the chemical variable to be predicted and depends on the measured phenomena. In a second step, a combination of three micro-spectrometers (SCIO, NIRscan and NIRone) was proposed.

Model performances were compared to those obtained with the laboratory spectrometer (ASD). Some models built from a single micro-spectrometer (the most expensive) gave similar performances as the laboratory spectrometer. For CP, the combination of micro-spectrometers gave a prediction performance (sep=0.69%; bias=0.15%; R^2_{test}=0.910) close to that obtained with the laboratory spectrometer (sep=0.56%; bias=-0.13%; R^2_{test} =0.935). For TS, the results obtained with this combination of micro-spectrometers (sep=2.38%; bias=-0.52%; R^2_{test}=0.983) are better than those obtained with the laboratory spectrometer (sep=2.59%; bias=0.41%; R^2_{test}=0.978). For both chemical variables, the combination of the micro-spectrometers significantly increases the performance of the predictive models compared to the models obtained with the micro-spectrometers independently.

Using several low-cost micro-spectrometers, combined with a multi-block method gave results as good as a single laboratory spectrometer. The overall cost can be lower than a reference spectrometer. In this study, the SO-PLS multi-block method with the number of latent variables selected per block shows the usefulness of micro-spectrometers in the prediction results. It guides the choice of the combination of micro-spectrometers by the variable(s) to be predicted. A further study could thus define the cost benefit versus the measurement efficiency. Trade-offs between prediction quality and device cost can then be defined according to the objective and constraints of the application, particularly as is the case with on-line monitoring applications or outdoor measurements.

Acknowledgement

We thank the French near infrared spectroscopy scientific network HelioSPIR for financial support, the CRA-w, Ird B. Barthes, Fondis electronics for the loan of equipment and the involvement of Denis Bastianelli and Laurent Bonnal from UMR ARE - Cirad. Finally we thank Abdellah Zgouz for his huge involvement in acquiring these data.

References

Donald M. Ball, Mike Collins, G. D. Lacefield, N. P. Martin, D. A. Mertens, K. E. Olson, D. H. Putnam, D. J. Undersander, and M. W. Wolf. Understanding forage quality. *American Farm Bureau Federation Publication*, 1 (01), 2001.

Michael Collins and J. Fritz. Forage quality. *Forages. Vol. 1. An Introduction to Grassland Agriculture*, pages 363–390, 2003. Publisher: Iowa State Press Ames, IA.

R. J. Wilkins. Forages and their role in animal systems. *Forage Evaluation in Ruminant Nutrition, CAB International*, pages 1–14, 2000.

Jerry Stuth, Abdi Jama, and Doug Tolleson. Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy. *Field Crops Research*, 84(1):45–56, October 2003. ISSN 0378-4290. doi: 10.1016/S0378-4290(03)00140-0. URL https://www.sciencedirect.com/science/article/pii/S0378429003001400.

E. R. Deaville and P. Ch Flinn. Near infrared (NIR) spectroscopy: an alternative approach for the estimation of forage quality and voluntary intake. *Forage evaluation in ruminant nutrition*, pages 301–320, 2000. Publisher: CABI Publishing Wallingford.

Franklin E Barton II and William R Windham. Determination of Acid-Detergent Fiber and Crude Protein in Forages by Near-Infrared Reflectance Spectroscopy: Collaborative Study. *Journal of Association of Official Analytical Chemists*, 71(6):1162–1167, November 1988. ISSN 0004-5756. doi: 10.1093/jaoac/71.6.1162. URL https://doi.org/10.1093/jaoac/71.6.1162.

Heinz W. Siesler, Yukihiro Ozaki, Satoshi Kawata, and H. Michael Heise. *Near-infrared spectroscopy: principles, instruments, applications*. John Wiley & Sons, 2008.

Jerry Workman and Art W. Springsteen, editors. *Applied spectroscopy: a compact reference for practitioners*. Academic Press, San Diego, 1998. ISBN 978-0-12-764070-9.

Zongyin Yang, Tom Albrow-Owen, Weiwei Cai, and Tawfique Hasan. Miniaturization of optical spectrometers. *Science*, 371(6528):eabe0722, January

2021. doi: 10.1126/science.abe0722. URL https://www.science.org/doi/10.1126/science.abe0722. Publisher: American Association for the Advancement of Science.

Hui Yan and Heinz W Siesler. Hand-held near-infrared spectrometers: State-of-the-art instrumentation and practical applications. *NIR news*, 29(7):8–12, November 2018. ISSN 0960-3360, 1756-2708. doi: 10.1177/0960336018796391. URL http://journals.sagepub.com/doi/10.1177/0960336018796391.

Kerstin Wiesner, Karen Fuchs, Alexander M. Gigler, and Remigiusz Pastusiak. Trends in Near Infrared Spectroscopy and Multivariate Data Analysis From an Industrial Perspective. *Procedia Engineering*, 87:867–870, 2014. ISSN 18777058. doi: 10.1016/j.proeng.2014.11.292. URL https://linkinghub.elsevier.com/retrieve/pii/S1877705814024072.

Svante Wold, Michael Sjöström, and Lennart Eriksson. PLS-regression: a basic tool of chemometrics. *Chemometrics and intelligent laboratory systems*, 58(2):109–130, 2001.

Jasper Engel, Jan Gerretzen, Ewa Szymańska, Jeroen J. Jansen, Gerard Downey, Lionel Blanchet, and Lutgarde M.C. Buydens. Breaking with trends in pre-processing? *TrAC Trends in Analytical Chemistry*, 50:96–106, October 2013. ISSN 01659936. doi: 10.1016/j.trac.2013.04.015. URL http://linkinghub.elsevier.com/retrieve/pii/S0165993613001465.

Asmund Rinnan, Frans van den Berg, and Soren Balling Engelsen. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10):1201–1222, November 2009. ISSN 01659936. doi: 10.1016/j.trac.2009.07.007. URL http://linkinghub.elsevier.com/retrieve/pii/S0165993609001629.

Puneet Mishra, Jean-Michel Roger, Delphine Jouan-Rimbaud-Bouveresse, Alessandra Biancolillo, Federico Marini, Alison Nordon, and Douglas N. Rutledge. Recent trends in multi-block data analysis in chemometrics for multi-source data integration. *TrAC Trends in Analytical Chemistry*, 137:116206, April 2021. ISSN 0165-9936. doi: 10.1016/j.trac.2021.116206. URL https://www.sciencedirect.com/science/article/pii/S0165993621000285.

T. Naes, O. Tomic, B.-H. Mevik, and H. Martens. Path modelling by sequential PLS regression. *Journal of Chemometrics*, 25(1):28–40, January 2011. ISSN 08869383. doi: 10.1002/cem.1357. URL http://doi.wiley.com/10.1002/cem.1357.

L. Awhangbo, R. Bendoula, J.M. Roger, and F. Béline. Multi-block data analysis for online monitoring of anaerobic co-digestion process. *Chemometrics and Intelligent Laboratory Systems*, 205:104120, October 2020. ISSN 01697439. doi: 10.1016/j.chemolab.2020.104120. URL https://linkinghub.elsevier.com/retrieve/pii/S0169743920300198.

Abdallah Zgouz, Daphné Héran, Bernard Barthès, Denis Bastianelli, Laurent Bonnal, Vincent Baeten, Sebastien Lurol, Michael Bonin, Jean-Michel Roger, Ryad Bendoula, and Gilles Chaix. Dataset of visible-near infrared handheld and micro-spectrometers – comparison of the prediction accuracy of sugarcane properties. *Data in Brief*, 31:106013, August 2020. ISSN 2352-3409. doi: 10.1016/j.dib.2020.106013. URL http://www.sciencedirect.com/science/article/pii/S2352340920309070.

SMRIDeterminationofthe reducing sugars in raw sugar bythe Luff Schoorl method. *SMRITest methods, TM050.*, 1997. URL http://www.smri.org/rawsugar.php.

R. J. Barnes, Mewa Singh Dhanoa, and Susan J. Lister. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied spectroscopy*, 43(5):772–777, 1989. Publisher: SAGE Publications Sage UK: London, England.

Gilles Rabatel, Federico Marini, Beata Walczak, and Jean-Michel Roger. VSN: Variable sorting for normalization. *Journal of Chemometrics*, 34(2), February 2020. ISSN 0886-9383, 1099-128X. doi: 10.1002/cem.3164. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/cem.3164.

Abraham Savitzky and Marcel JE Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8): 1627–1639, 1964.

Bart M. Nicolaï, Thijs Defraeye, Bart De Ketelaere, Els Herremans, Maarten L.A.T.M. Hertog, Wouter Saeys, Alessandro Torricelli, Thomas Vandendriessche, and Pieter Verboven. Nondestructive Measurement of Fruit and Vegetable Quality. *Annual Review of Food Science and Technology*, 5(1):285–312, February 2014. ISSN 1941-1413, 1941-1421. doi: 10.

1146/annurev-food-030713-092410. URL http://www.annualreviews.org/doi/10.1146/annurev-food-030713-092410.

B. G. Osborne, T. Fearn, and P. T. Hindle. *Practical NIR spectroscopy with applications in food and beverage analysis*. Addison-Wesley Longman Ltd, Harlow UK, 1993. ISBN 978-0-582-09946-3. URL http://discovery.ucl.ac.uk/267166/.

Maxime Ryckewaert, Nathalie Gorretta, Fabienne Henriot, Federico Marini, and Jean-Michel Roger. Reduction of repeatability error for analysis of variance-Simultaneous Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample. *Analytica Chimica Acta*, 1101:23–31, March 2020. ISSN 00032670. doi: 10.1016/j.aca.2019.12.024. URL https://linkinghub.elsevier.com/retrieve/pii/S0003267019314606.

Akira Ishimaru. *Wave propagation and scattering in random media*, volume 2. Academic press New York, 1978.

David H. Clark and Robert C. Lamb. Near Infrared Reflectance Spectroscopy: A Survey of Wavelength Selection To Determine Dry Matter Digestibility1, 2, 3. *Journal of Dairy Science*, 74(7):2200–2205, July 1991. ISSN 0022-0302. doi: 10.3168/jds.S0022-0302(91)78393-8. URL https://www.sciencedirect.com/science/article/pii/S0022030291783938.

David L. Wetzel. Near-infrared reflectance analysis. *Analytical chemistry*, 55 (12):1165A–1176A, 1983. Publisher: ACS Publications.

Sandra E. Kays, Franklin E. Barton, William R. Windham, and David S. Himmelsbach. Prediction of Total Dietary Fiber by Near-Infrared Reflectance Spectroscopy in Cereal Products Containing High Sugar and Crystalline Sugar. *Journal of Agricultural and Food Chemistry*, 45(10): 3944–3951, October 1997. ISSN 0021-8561, 1520-5118. doi: 10.1021/jf9703260. URL https://pubs.acs.org/doi/10.1021/jf9703260.