

available at www.sciencedirect.comjournal homepage: www.elsevier.com/locate/aca

Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder

Anežka Veselá^a, António S. Barros^b, Andriy Synytsya^a,
Ivonne Delgadillo^b, Jana Čopíková^a, Manuel A. Coimbra^{b,*}

^a Institute of Chemical Technology, Department of Carbohydrate Chemistry and Technology, Technická 1905, 166 28 Prague 6, Czech Republic

^b Universidade de Aveiro, Departamento de Química, 3810-193 Aveiro, Portugal

ARTICLE INFO

Article history:

Received 14 March 2007

Received in revised form

14 August 2007

Accepted 20 August 2007

Published on line 26 August 2007

Keywords:

Cocoa powder

Infrared spectroscopy

Near infrared

Fourier-transform infrared

Fat

Nitrogen

Moisture

Outer product-principal component

transform-partial least squares

regression

ABSTRACT

The combination of the near infrared (NIR) and Fourier-transform infrared (FTIR) absorbance spectra (1100–2500 nm and 4000–600 cm⁻¹) of 100 cocoa powder samples was used to build calibration models for the determination of the content of fat, nitrogen, and moisture. The samples that comprised the dataset had an average composition of 13.51% of fat, 3.77% nitrogen, and 3.98% moisture. The fat content ranged from 2.42 to 22.00%, the nitrogen from 0.88 to 4.48%, and moisture from 1.60 to 7.80%. For NIR, the relative root mean square error of cross-validation (RMSECV) was 7.0% ($R^2 = 0.96$) for fat, 1.7% ($R^2 = 0.98$) for nitrogen, and 5.2% ($R^2 = 0.94$) for moisture. For FTIR, the relative RMSECV was 10.4% ($R^2 = 0.94$) for fat and 3.9% ($R^2 = 0.95$) for nitrogen. However, for moisture, it was not possible to build a calibration model with suitable predictability. The combination of the NIR and FTIR domains (data fusion) by outer product analysis PLS1 allowed to predict these parameters and to characterise frequencies in one domain based on the information of the other domain. This work allows to conclude that the second derivative of NIR is the recommended procedure to quantify fat, nitrogen, and moisture content in cocoa powders by infrared spectroscopy.

© 2007 Elsevier B.V. All rights reserved.

1. Introduction

Cocoa powder is formed from the cocoa mass where presses are used to remove some of the fat, leaving a solid material called cocoa press cake. These cakes are then crushed to form cocoa powder. The processing can be altered to produce cocoa powders of different composition and with different fat levels. The content of main components in cocoa powder, given

by the International Cocoa Organization, is summarized in Table 1 [1,2]. Cocoa bean products such as chocolate tablets, chocolate confectioneries, chocolate cookies, cocoa powder, instant chocolate drinks, etc., are very popular commodities.

Infrared spectroscopy offers a number of important advantages over traditional chemical methods. It is a physical, non-destructive and non-invasive method, requiring minimal or no sample preparation and its precision can be high. It is a

* Corresponding author. Tel.: +351 234 370706; fax: +351 234 370084.

E-mail address: mac@ua.pt (M.A. Coimbra).

0003-2670/\$ – see front matter © 2007 Elsevier B.V. All rights reserved.

doi:10.1016/j.aca.2007.08.039

Table 1 – Main components in cocoa powder [1,2]

Component	Content (% w/w)
Carbohydrates	41
Cocoa butter	11
Ash	5.5
Total nitrogen	4.3
Moisture	3.0
Theobromine	2.8
Water soluble ash	2.2
Shell (calculated to unalkalised nib)	1.4
Ash insoluble in 1:1 (w/w) HCl	0.08

multi-analytical technique, that is, several determinations can be made simultaneously. The method offers the possibility of measuring physical and chemical properties. Once calibrated, the NIR and mid-infrared (MIR) spectrometers are simple to operate. The main disadvantages are the dependence on time-consuming and laborious calibration procedures and the complexity in the choice of data treatment.

NIR spectroscopy has been shown a rapid non-destructive, non-invasive method for the characterisation of chocolate [3] and cocoa powder [4]. The preliminary studies were performed on the set of 20 samples of cocoa powder with different, predetermined compositional values. Experiments showed satisfactory accuracy for fat, protein, and carbohydrate prediction in cocoa powder [5]. Analysis of powdered cocoa products showed that moisture, fat, and sucrose could be analyzed in such products by NIR spectroscopy because good correlation coefficients and low standard errors of prediction were achieved [6]. Fourier-transform near infrared (FT-NIR) spectroscopy [7] was used to establish calibration models with the aim of determining sucrose, lactose, fat, and moisture in chocolate. The validation results proved that all these parameters, with exception of moisture, can be predicted with sufficient accuracy. NIR spectroscopy technique has also been developed for the prediction of procyanidins in cocoa beans. In addition, very robust calibration models were obtained for the prediction of the total procyanidin oligomers [8].

The MIR spectroscopy has been commonly applied to the structural identification or qualitative determination of the “fingerprint” of organic compounds, as some groups of atoms display characteristic vibrational absorption. Moreover, MIR spectroscopy is amenable for quantitative analysis applications, as the intensities of the bands in the spectrum are proportional to the concentration of their respective functional groups, although the NIR is the preferred method as it is more robust with respect to the MIR/FTIR approaches. Foodstuff is a subject that may strongly profit from FTIR spectroscopic technique because foods have compounds with characteristic absorption bands in this region of the electromagnetic spectrum, namely, lipids, proteins, carbohydrates, and water. FTIR spectroscopy has been proposed for analysis of potential lard adulteration in chocolate and chocolate products [9], as a tool for the rapid detection of other vegetable fats mixed in cocoa butter [10], and for routine analysis of chocolate milk fat, sucrose, lactose, and total solids using an automated heated flow cell [11].

The aim of this work is twofold: (i) to use the reflectance NIR and attenuated single total reflectance FTIR spectroscopy in

order to build calibration models for the analysis of fat, nitrogen, and moisture in cocoa powder samples; (ii) to combine both spectroscopic domains, by means of outer product analysis (data fusion), in tandem with partial least squares (PLS1) regression [12–16] for a better characterisation of the samples as a function of the parameters under study. Outer product analysis (OPA) is a method which makes it possible to emphasise co-evolutions of spectral regions in signals acquired in two different domains (heterospectral) or even for the same domain (homospectral) [17–25].

2. Experimental

2.1. Cocoa powder samples

The sample set comprised 100 samples of cocoa powder of different brands obtained mainly in Czech Republic and in different other countries. Samples of cocoa powder and sucrose mixture, cocoa fibre, and cocoa powder containing dried milk and wheat starch were also included.

2.2. Fat determination

A conventional Soxhlet extraction with petroleum ether was used for determination of the fat content [26]. The difference between two parallel determinations was less than 0.20% (w/w).

2.3. Nitrogen determination

Nitrogen content determination according to Kjeldahl was used [27]. The determination was performed on a Kjeltect System (Kjeltec, Sweden). The results were expressed as percentage of nitrogen. The difference between two parallel determinations was less than 0.10% (w/w).

2.4. Moisture determination

Samples were dried to constant weight in platinum dish in an aerated oven at 100 °C. Loss in weight was reported as H₂O [28]. The difference between two parallel determinations was less than 0.20% (w/w).

2.5. NIR spectroscopy

Near infrared spectra were acquired on a NIR Systems 6500 (Perstop Analytical Company, USA) with software NIR3 version 3.10 Infrasoft International (ISI). Wavelengths are specified by grating monochromator with the range of 400–2500 nm. Tungsten filament lamp was used as a source of radiation. Spectra were recorded at the reflectance mode from 1100 to 2500 nm at the resolution of 2 nm. Three replicated spectra (36 co-added scans), in the small ring cup, were collected for every sample. A ceramic standard was used as a photometric standard. The scanning speed to obtain one replicate was 62 s. No spectral changes were detected during the replicate measurements. The background spectrum was obtained by an internal reference.

2.6. FTIR spectroscopy

Infrared spectroscopy in mid region was performed on FTIR spectrometer Bruker IFS-55 (Bruker, Germany) with a single reflection ATR accessory, with a diamond cell (Golden Gate) and a DTGS detector and triangular apodization function. Spectra were recorded at the absorbance mode from 4000 to 400 cm^{-1} at the resolution of 8 cm^{-1} . Five replicated spectra (128 co-added scans) were collected for every sample pressed on the ATR crystal. The background spectrum was obtained against the air.

2.7. Outer product analysis (OPA)

The outer product matrix for two spectra (vectors) is defined as all possible product combinations of the vector variables. The resulting matrix represents all variable's combinations between two domains (homospectral or heterospectral). The outer product was used to combine the FTIR in the 4000–600 cm^{-1} (848 variables) and NIR in the 1100–2500 nm

(700 variables) spectral regions [17]. For each sample, the outer product of these two regions gave a matrix with $(848 \times 700 = 593,600)$ variables). Then, this sample matrix was unfolded to give a vector with a size of $(1 \times 593,600)$. Once this process is performed for each sample, then the correspondent vectors are concatenated by row to give a matrix $\mathbf{K}_{(n, 593,600)}$, where n is the number of spectra. This matrix was then used in PLS1 regression to model the fat, nitrogen, and moisture contents in cocoa powder samples as a function of the relationships between both domains. The obtained \mathbf{b} vector, which established the relationship between the \mathbf{K} variables and the \mathbf{y} vector, was therefore composed of 593,600 values. In order to facilitate the interpretation of this vector, it was folded back to give a \mathbf{B} (848×700) matrix which highlighted the links between the FTIR and NIR variables (wavenumber and wavelength).

2.8. Calibration models building framework

In order to build calibration models for the quantification of fat, nitrogen, and moisture using NIR and FTIR, a Monte Carlo

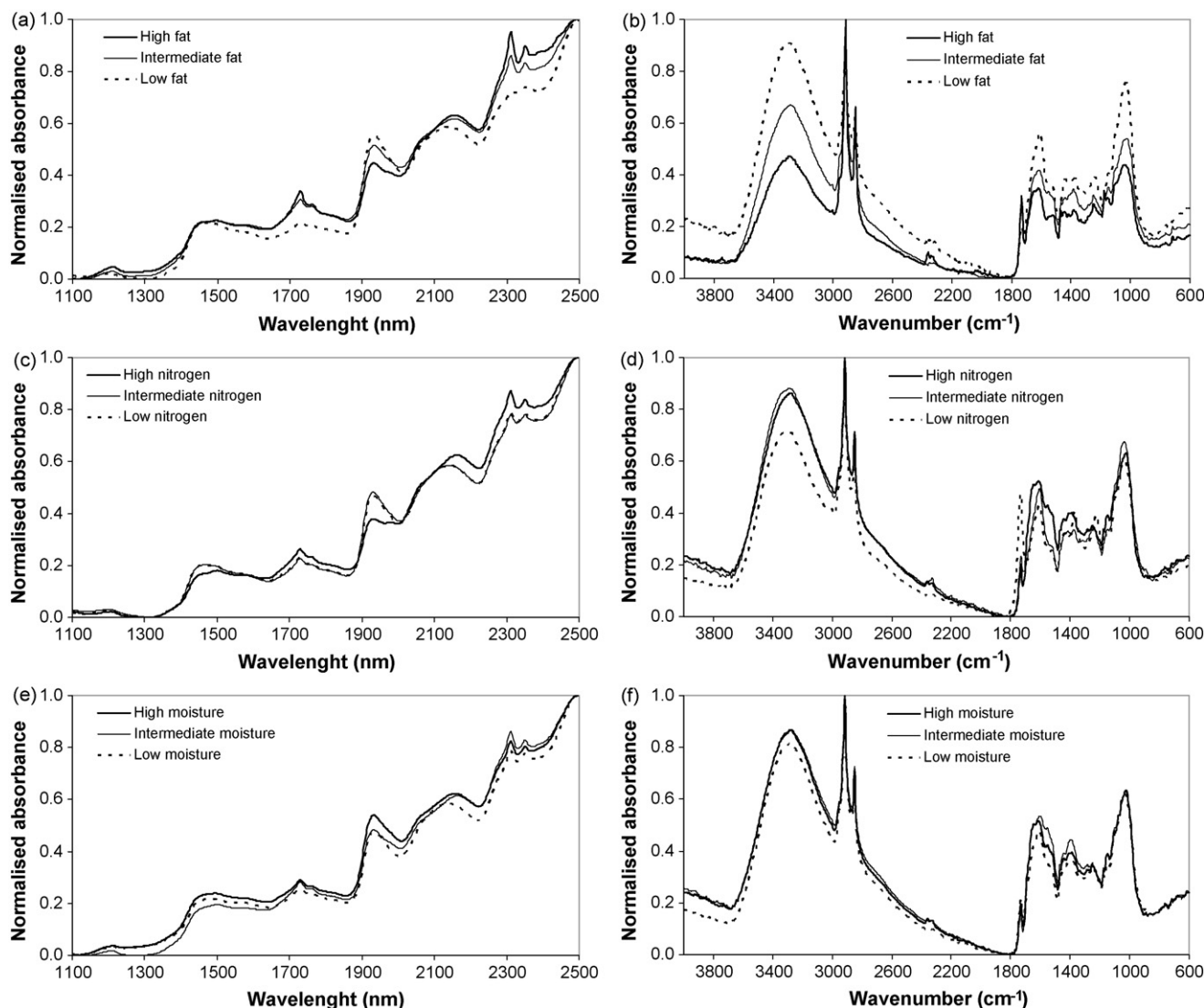


Fig. 1 – Characteristic NIR (left) and FTIR (right) spectra of cocoa powder containing different contents of: (a) and (b) fat; (c) and (d) nitrogen; (e) and (f) moisture.

cross-validation [29] framework was used. Each dataset was split into a calibration (learning set) and a validation (external) set to assess the predictive power of each model. The splitting process consisted in sorting the fat, nitrogen, and moisture values and then, randomly, selected 40% of the samples, where replicates are considered a sample, were used as validation set. The remaining 60% of the samples, with replacement, were used as calibration set. This means that before doing the models building one selects again randomly (uniform distribution) samples from the 60% samples used in the calibration set and replicate them to give a set with 100% samples. The external validation set remains with the same 40% samples. This procedure is used to give an unbiased estimation of the model relative to the size of the initial dataset which will provide a more robust building of the models. It was repeated several times (iterations). The dimensionality (number of latent variables) of the calibration model was assessed by internal cross-validation at each iteration level, with leave-three-out for NIR dataset and leave-five-out for FTIR, in accordance to the number of replicates. Then these models were used to predict the parameters of interest from the external set, expressed as the relative root mean square error of prediction (RMSEP). Several data pre-treatment procedures were tested and the optimal ones were selected for each domain and parameter to quantify.

Since this approach is very computational demanding, the principal component transform PLS1 (PCT-PLS1) [30] was used to build the calibration models in order to accelerate the Monte Carlo cross-validation process. The PCT-PLS1 procedure is based on a full rank eigen decomposition (NIPALS) of the X matrix before proceeding to the PLS1 regression. This approach dramatically accelerates the cross-validation of the calibration models and at the same time being parsimonious on computer memory requirements. This is most noticeable for huge data sets or when many iterations in a given set of data is need (like Monte Carlo cross-validation). This procedure preserves all the PLS1 modelling properties, like regression vectors interpretation, which facilitates the application of this procedure to build calibration and prediction models.

3. Results and discussion

3.1. Characterisation of cocoa powder samples

The 100 samples of cocoa powders that comprised the dataset under study had an average composition of 13.51% of fat, 3.77% of nitrogen, and 3.98% of moisture. The fat content in the dataset ranged from 2.42 to 22.00% ($s=0.49\%$ w/w), the nitrogen from 0.88 to 4.48% ($s=0.02\%$ (w/w)), and moisture from 1.60 to 7.80% ($s=0.19\%$ (w/w)).

The samples can be split into three groups according to their fat content. The group with high fat content included 37 samples and had an average fat content of 19.3%, ranging from 15.2 to 22.0%; the intermediate fat content group included 57 samples and had an average fat content of 10.4%, ranging from 8.6 to 13.9%; and the group with lower fat content included six samples and had an average fat content of 5.1%, ranging from 2.4 to 7.3%. According to nitrogen content,

the samples can be divided into two groups. The group with higher nitrogen content included 93 samples and had an average nitrogen content of 3.9%, ranging from 3.4 to 4.5%, and the group with lower nitrogen content included seven samples, with an average content of 2.0%, ranging from 0.88 to 2.8%.

Typical NIR and FTIR spectra of cocoa powder samples containing high (20.82%), intermediate (13.78%) and low (3.38%) content of fat is shown in Fig. 1a and b, respectively. Typical NIR and FTIR spectra of cocoa powder samples containing, for the same amount of fat (9%), high (4.30%), intermediate (2.54%), and low (0.88%) content of nitrogen is shown in Fig. 1c and d, respectively. Typical NIR and FTIR spectra of cocoa powder samples containing, for the same amount of fat (10%), high (6.41%), intermediate (4.71%), and low (2.88%) content of nitrogen is shown in Fig. 1e and f, respectively. The spectra show mainly high absorbance bands characteristic for fat, with differences observed in several spectral regions. In order to assess if these spectral variances are suitable to build calibrations models for fat, nitrogen, and moisture, the PCT-PLS1 regression procedure was applied to the NIR and FTIR spectra of the 100 samples of cocoa powder dataset.

3.2. Calibration models for NIR

Using the 1100–2500 nm region of the NIR spectra, the PCT-PLS1 regression method was applied to the cocoa powder

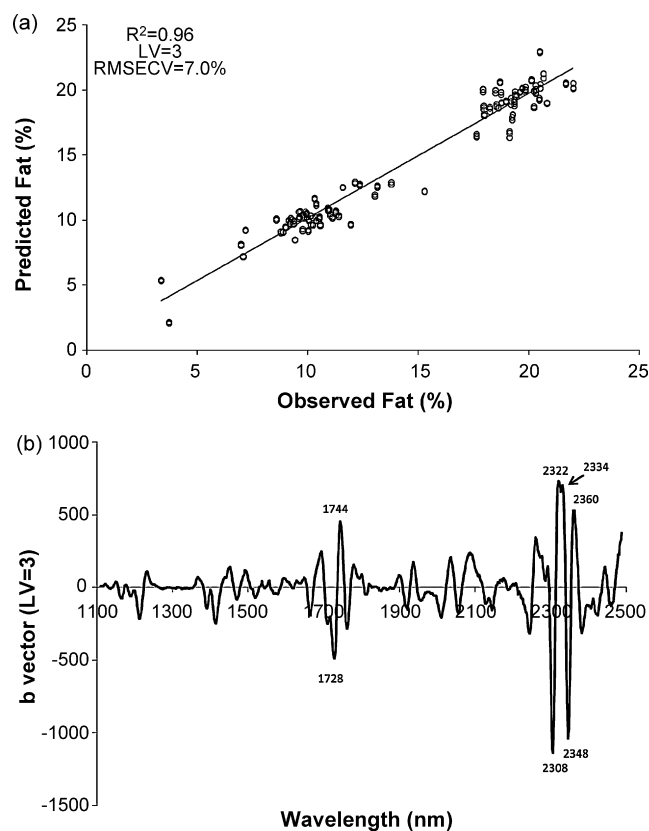


Fig. 2 – PCT-PLS1 regression model of the NIR second derivative spectra for the estimation of fat in cocoa powder samples (three LVs). (a) Relationship plot between observed vs. predicted amount of fat; (b) b vector profile.

Table 2 – Wavelength assignments of NIR spectra of cocoa powder samples [32–34]

Wavelength (nm)	Assignment
2360	$\nu + \delta(\text{CH}, \text{CC})$
2348	$\nu_s + \delta(\text{CH}_2)$ polysaccharides
2334	$\nu + \delta(\text{CH}_2)$ fat
2322	$\nu + \delta(\text{CH}_2)$ fat
2308	$\nu_{as} + \delta(\text{CH}_2)$ polysaccharides
2040	$\nu_{as} + \delta(\text{H}_2\text{O})$, very strongly bounded water, proteins, amines, acids
2292	$\nu(\text{CH}) + \nu(\text{C}=\text{C})$
2274	$\nu(\text{NH}) + \nu(\text{C}=\text{N})$
2078	Protein
1972	$\nu_{as}(\text{NH}) + \text{amide II}$
1938	Second overtone $\nu(\text{C}=\text{O})$
1934	$\nu_{as} + \delta(\text{H}_2\text{O})$ weakly bounded water, proteins, aromatics
1884	$\nu_{as} + \delta(\text{H}_2\text{O})$ very weakly bounded water fat
1744	First overtone $\nu(\text{CH})$
1728	First overtone $\nu_{as}(\text{CH}_2)$
1674	First overtone $\nu(\text{CH})$ aromatics
1444	$\nu_{as} + \nu_s(\text{H}_2\text{O})$, non-bounded water
1414	$\nu_{as} + \nu_s(\text{H}_2\text{O})$, weakly bounded water, proteins, aromatics

ν : stretching; ν_s : symmetric stretching; ν_{as} : asymmetric stretching; δ : rocking.

dataset for the estimation of their fat, nitrogen, and moisture contents.

Concerning the fat content, it was found that the optimal calibration model was based on the second derivative, using Savitzky–Golay [31] procedure with a second degree polynomial with a 10 points window. Three latent variables (LVs), giving a relative root mean square error of calibration (RMSECV) of 7.0% with a coefficient of determination (R^2) of 0.96 and a RMSEP of 1.2%, were obtained. The relationship between observed and predicted values for fat based on this PCT-PLS1 model is shown in Fig. 2a. Furthermore, the correspondent **b** vector profile (Fig. 2b), shows that the most important bands related with fat variation are located at 1744, 2322, 2334, and 2360 nm, which increase as the fat content raises. According to Table 2, these variations are mainly characterised by the symmetric vibration of C–H, C=O second overtone and stretching and rocking vibration of CH_2 of fat. On opposite direction, the bands located at 1728, 2308, and 2348 nm decrease as the fat content increases. These frequencies are most related to stretching and rocking vibrations of CH_2 of polysaccharides.

Regarding nitrogen, and once more, to have a predictive power, the second derivative of the NIR spectra in tandem with seven LVs give the optimal calibration model. The relationship plot between the observed and predicted nitrogen content is depicted in Fig. 3a. The relative RMSECV obtained was 1.7%, with a R^2 of 0.98 and a relative RMSEP of 0.09%. The **b** vector profile of the calibration model (Fig. 3b) shows the most important wavelengths related to the variability of nitrogen and, according to Table 2, one can find the stretching modes of N–H and C=N, as well as the wavelength related to protein (2078 nm), as the ones that are positively related to the increase of the nitrogen content.

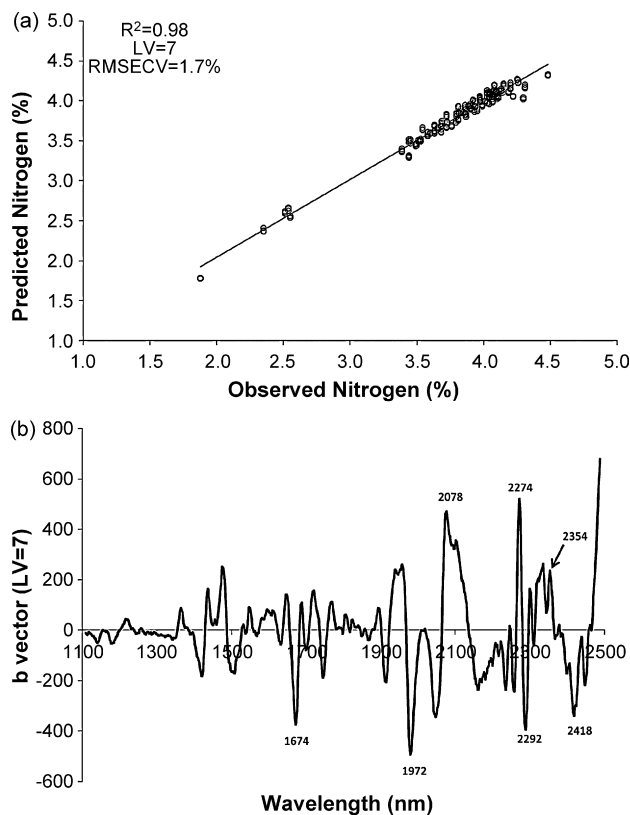


Fig. 3 – PCT-PLS1 regression model of the NIR second derivative spectra for the estimation of nitrogen in cocoa powder samples (seven LVs). (a) Relationship plot between observed vs. predicted amount of nitrogen; (b) **b vector profile.**

In relation to moisture, it was found that the optimal model needed six LVs and was based on the second derivative of NIR spectra. The relationship between the observed and the predicted values is plotted in Fig. 4a. The relative RMSECV obtained was 5.2%, with a R^2 of 0.94 and a relative RMSEP of 0.4%. The **b** vector profile of the calibration model (Fig. 4b) shows that the most important wavelengths related to the moisture content are mainly related to the frequencies linked to bounded and unbounded water molecules (cf. Table 2).

3.3. Calibration models for FTIR

For FTIR it was found, by model testing, that the standard normal deviates (SNV) data pre-treatment was the one that gave the most robust results. Using the SNV of the $4000\text{--}600\text{ cm}^{-1}$ region of the FTIR spectra, the PCT-PLS1 regression procedure was applied to the cocoa powder dataset for the estimation of their fat, nitrogen, and moisture content. For fat, it was found that it was possible to build a calibration model with seven LVs. The relative RMSECV obtained was 10.4% with a R^2 of 0.94 and a relative RMSEP of 1.4%. The relationships between the actual and observed fat values is represented in Fig. 5a, showing a linear relationship between 2.4 and 22.0% of fat. The correspondent **b** vector profile of the calibration model (Fig. 5b) shows that the most important wavenumbers related

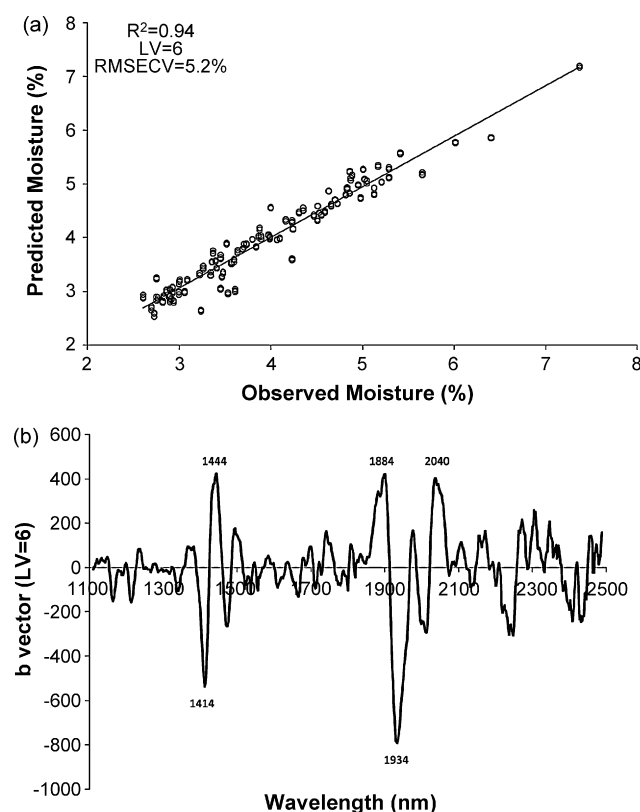


Fig. 4 – PCT-PLS1 regression model of the NIR second derivative spectra for the estimation of moisture in cocoa powder samples (six LVs). (a) Relationship plot between observed vs. predicted amount of moisture; (b) *b* vector profile.

to the variability of fat were the bands located at 2916, 2850, 1655, 1469, and 1176 cm^{-1} , which increase as the fat content increases, and the bands located at 1747, 1365, and 1018 cm^{-1} , which decrease as the fat content increases. According to Table 3, the bands that are positively correlated with fat are those related with CH_2 symmetric and asymmetric stretching and aromatic compounds possibly soluble in fat. The bands that are negatively correlated with fat are those mainly related with C–N aromatic bonds, most likely from proteins.

Concerning nitrogen, it was found that it was required seven LVs in order to have predictive power. The relative RMSECV obtained was 3.9% with a R^2 of 0.95 and a relative RMSEP of 0.14%. The observed versus predicted nitrogen relationships is described in Fig. 6a, showing a linear relationship between 0.88 and 4.48% of nitrogen in cocoa powder samples. The *b* vector profile of the calibration model (Fig. 6b) shows that the most important wavenumbers related to the variability of nitrogen were the bands located at 1647, 1516, and 1176 cm^{-1} , which increase as the nitrogen content increases, and the bands located at 2924, 2858, 1747, 1574, 1404, and 1026 cm^{-1} , which increase as the nitrogen content decreases. The former wavenumbers are mainly ascribed to vibrations of amides I and II as well as stretching modes of aromatic C–C and C–N, the latter ones mainly attributed to symmetric and asymmetric stretching modes of CH_2 and C=O.

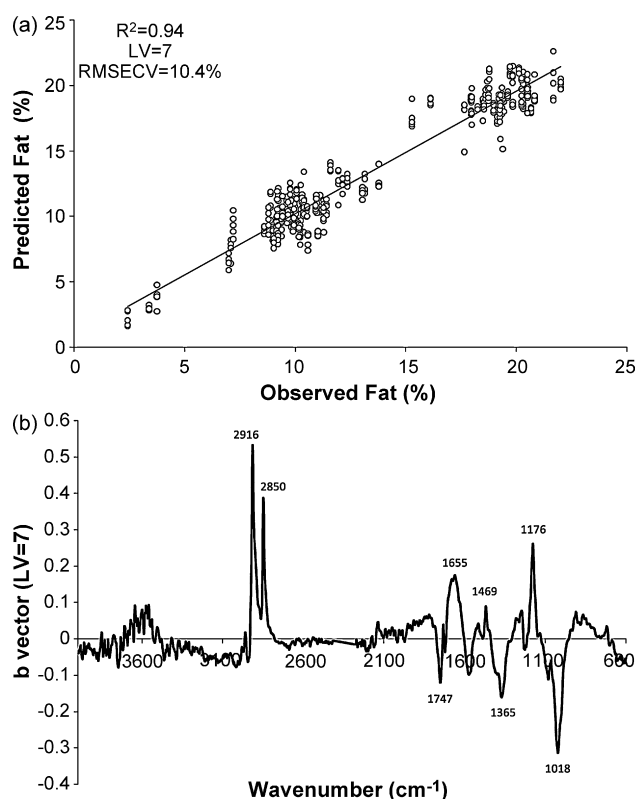


Fig. 5 – PCT-PLS1 regression model of the FTIR spectra for the estimation of fat in cocoa powder samples (seven LVs). (a) Relationship plot between observed vs. predicted amount of fat; (b) *b* vector profile.

A tentative approach was taken for the quantification of the moisture content using PCT-PLS1, but it was not possible to achieve a model with a robust predictive power.

Table 3 – Wavenumber assignments of FTIR spectra of cocoa powder samples [35–37]

Wavenumber (cm^{-1})	Assignment
2924	$\nu_{\text{as}}(\text{CH}_2)$
2916	$\nu_{\text{as}}(\text{CH}_2)$
2858	$\nu_{\text{s}}(\text{CH}_2)$
2850	$\nu_{\text{s}}(\text{CH}_2)$
1747	$\nu(\text{C}=\text{O})$ ester, acid
1744	$\nu(\text{C}=\text{O})$ ester, fat
1647	$\delta(\text{H}_2\text{O})$, amide I (typical 1660)
1574	Amide II (typical 1540), $\nu(\text{CC})$ aromatic
1516	Amide II (typical 1540), $\nu(\text{CC})$ aromatic
1466	$\delta(\text{CH}_2)$, fat
1404	$\nu_{\text{s}}(\text{COO}^-)$
1365	$\nu_{\text{s}}(\text{CH}_3)$
1180	$\nu(\text{CC})$, $\nu(\text{CN})$, $\nu(\text{CO})$
1176	$\nu(\text{CC})$, $\nu(\text{CN})$, $\nu(\text{CO})$
1026	$\nu(\text{CC})$, $\nu(\text{CN})$, $\nu(\text{CO})$
1018	$\nu(\text{CC})$, $\nu(\text{CN})$, $\nu(\text{CO})$

ν : stretching; ν_{s} : symmetric stretching; ν_{as} : asymmetric stretching; δ : rocking.

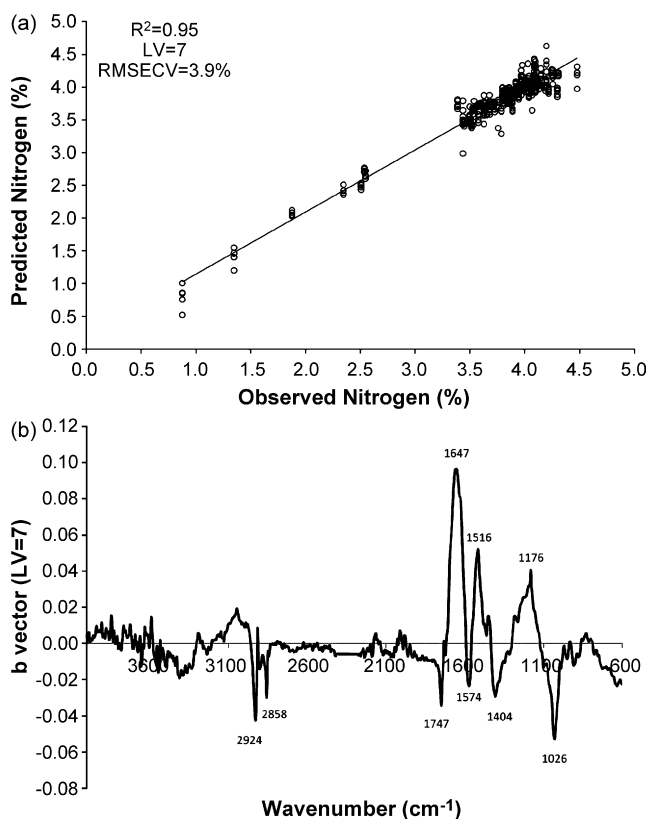


Fig. 6 – PCT-PLS1 regression model of the FTIR spectra for the estimation of nitrogen in cocoa powder samples (seven LVs). (a) Relationship plot between observed vs. predicted amount of nitrogen; (b) b vector profile.

3.4. Outer product-principal component transform-partial least squares regression (OP-PCT-PLS1)

The outer product matrix between NIR and FTIR was built by the combination of the second derivative of NIR spectra (1100–2500 nm) and the SNV data processed FTIR spectra (4000–600 cm^{-1}). The outer product between the 100 samples of both domains (NIR and FTIR) gave a matrix with 300 rows (three replicates) and 593,600 columns. This matrix was then used to build calibration models by means of PCT-PLS1 using the Monte Carlo cross-validation approach to determine fat, nitrogen, and moisture contents.

Concerning the modelling of fat content by OP-PCT-PLS1 it was found that five LVs were needed to obtain a predictive power with a relative RMSECV of 7.6% and a relative RMSEP of 1.2% (cf. Fig. 7a). The folded b vector is shown in Fig. 7b as a 2D false colour map. This map allows to characterise the relationships between the frequencies of both domains, with the most important links between variables in the two domains represented by blue (minimum) and red (maximum) spots. The positive relationships are related to those interactions that characterise the increase of fat content and, conversely, the negative ones are anti-related to the fat amount. In order to facilitate the 2D map analysis, the most interesting FTIR profiles are recovered as a function of NIR dimension. Hence, it was found three important FTIR profiles located at 2919, 2850,

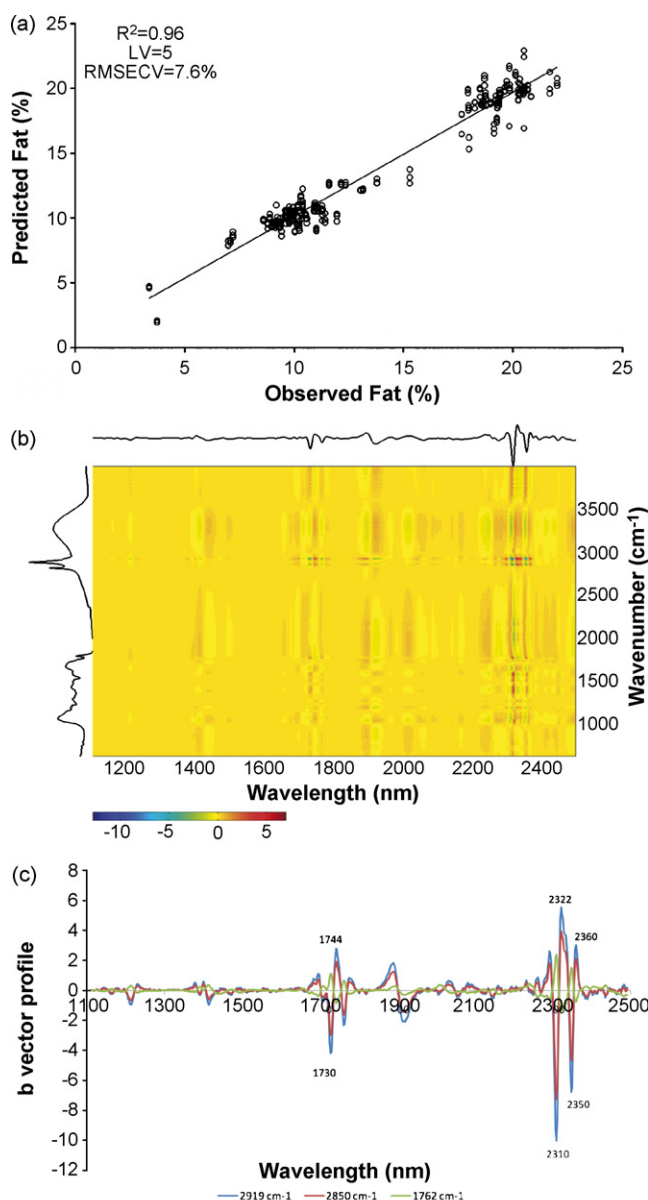


Fig. 7 – OP-PCT-PLS1: (a) relationship plot between observed vs. predicted amount of fat; (b) b vector 2D (spectral average profiles of NIR and FTIR are shown, respectively, in the top and left sides); (c) b vector profiles at 2919, 2850, and 1762 cm^{-1} .

and 1762 cm^{-1} that is shown in Fig. 7c. The first two bands, ascribed to asymmetric and symmetric stretching modes of CH_2 , are positively related to the 1744, 2322, and 2360 nm, assigned to the first overtone of CH (stretching) and to stretching and rocking modes of CH_2 and CH groups. The profile of the band around 1762 cm^{-1} , attributed to stretching mode of the C=O group, is approximately symmetrical to the 2919 and 2850 cm^{-1} profile bands (Fig. 7c), meaning that they are anti-correlated and thus, the correspondent NIR bands are also anti-correlated (cf. Table 4).

For the nitrogen content, the obtained model was more complex. It was required 11 LVs to model the interactions

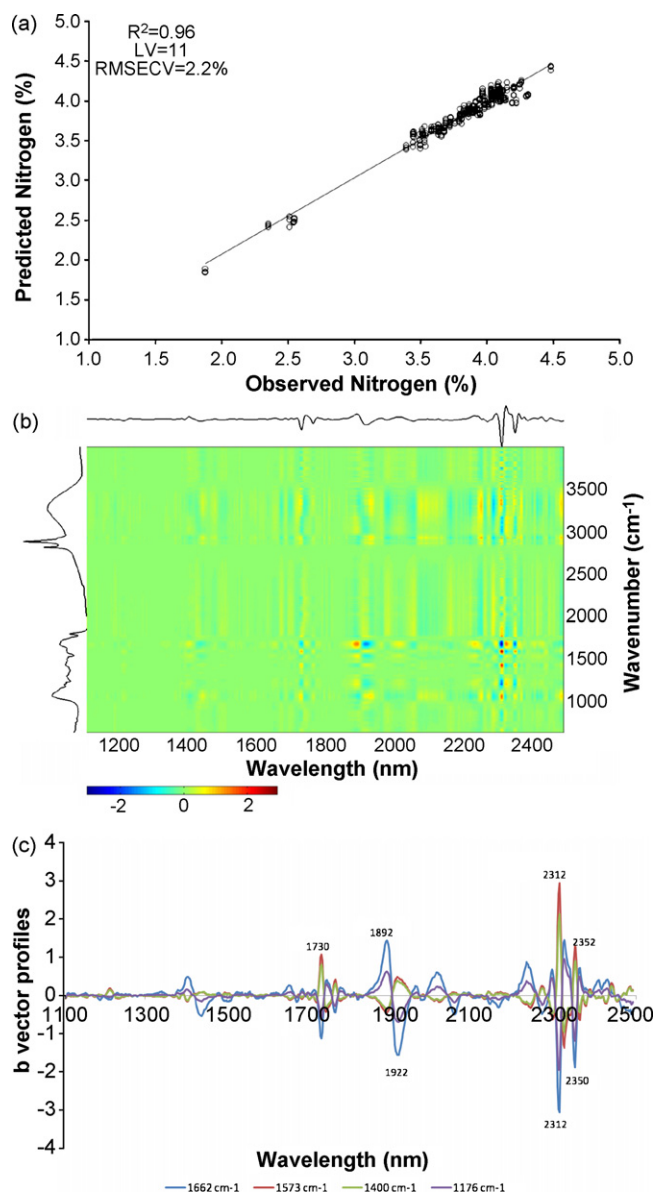


Fig. 8 – OP-PCT-PLS1: (a) relationship plot between observed vs. predicted amount of nitrogen; (b) b vector 2D (spectral average profiles of NIR and FTIR are shown, respectively, in the top and left sides); (c) b vector profiles at 1662, 1573, 1400, and 1176 cm^{-1} .

between FTIR and NIR as a function of the nitrogen content with a cross-validation error (relative RMSECV) of 2.2% and a relative RMSEP of 0.10%. The relationships between both domains, as a function of the nitrogen content, are depicted in the unfolded b vector false colour map (Fig. 8b). From this map several important FTIR profiles were recovered and plotted in Fig. 8c, the 1662, 1573, 1400, and 1176 cm^{-1} . These interaction profiles show important relationships between amides I and II, and stretching of C–C and C–N with different stretching vibration modes of CH_2 (cf. Table 4).

Regarding the moisture content, it was possible to find a suitable calibration model using six LVs with a calibration error (relative RMSECV) of 5.9% and a relative RMSEP of 0.40% and

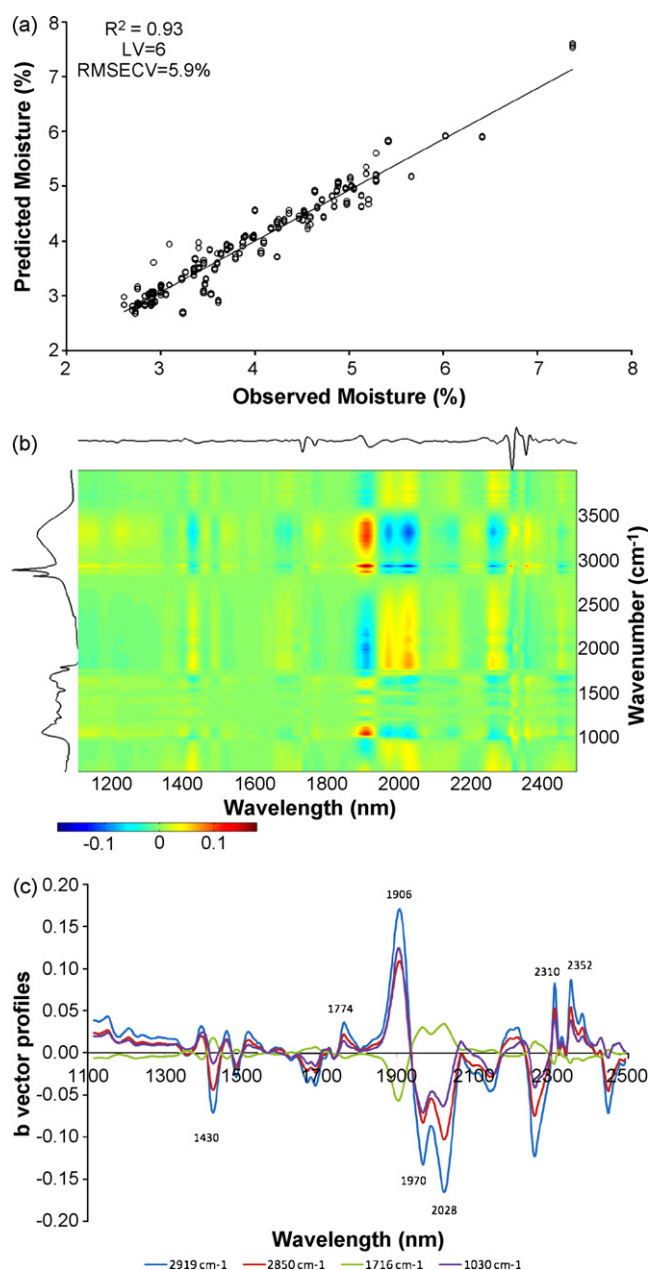


Fig. 9 – OP-PCT-PLS1: (a) relationship plot between observed vs. predicted amount of moisture; (b) b vector 2D (spectral average profiles of NIR and FTIR are shown, respectively, in the top and left sides); (c) b vector profiles at 2919, 2850, 1716, and 1030 cm^{-1} .

hence, to establish a relationship between the NIR and the FTIR variations as a function of the moisture content. This result was interesting as it was not possible to calibrate the FTIR for the moisture content, as seen in the previous section. Therefore, using the information present in NIR one could, in turn, extract the variation source concerning moisture content in the FTIR domain. In fact, the unfolded b vector relating to moisture is shown in Fig. 9b, and one can see important inter-relationship regions. The most important interactions were extracted as profiles and are plotted in Fig. 9c. As one can see from Fig. 9c and Table 4, the profiles indicated that

Table 4 – b vector variables—main interactions as a function of fat, nitrogen, and moisture

FTIR (wavenumber, cm ⁻¹)	NIR (wavelengths, nm)
Fat	
2919 and 2850 (ν_{as} and ν_s of CH ₂)	Positive link—1744, 2322, and 2360: first overtone ν CH; ν and δ of CH ₂ ; ν and δ of CH and CC. Negative link—1728, 2308, and 2348: first overtone ν_{as} of CH ₂ ; $\nu_{as} + \delta$ (CH ₂) polysaccharides; $\nu_s + \delta$ (CH ₂) polysaccharides
1762 (ν (C=O))	Positive link—1728, 2308, and 2348: first overtone ν_{as} of CH ₂ ; $\nu_{as} + \delta$ (CH ₂) polysaccharides; $\nu_s + \delta$ (CH ₂) polysaccharides. Negative link—1744, 2322, and 2360: first overtone ν CH; ν and δ of CH ₂ ; and ν and δ of CH and CC
Nitrogen	
1666: amide I	Positive link—1892 and 2324: mainly stretching vibration of CH ₂ . Negative link—1730, 1920, 2312, and 2350: first overtone of ν_{as} (CH ₂), ν_{as} typical of proteins, and ν_s and ν_{as} of CH ₂
1566: amide II, 1400: ν_s (COO ⁻)	Positive link—1730, 2312, and 2352: first overtone of ν_{as} (CH ₂), ν_{as} typical of proteins, and ν_s and ν_{as} of CH ₂
1180: ν (CC), ν (CN), ν (CO)	Positive link—1892 and 2324: mainly stretching vibration of CH ₂ . Negative link—1730, 2312, and 2352: first overtone of ν_{as} (CH ₂), ν_{as} typical of proteins, and ν_s and ν_{as} of CH ₂
Moisture	
2919: ν_{as} (CH ₂), 2850: ν_s (CH ₂), 1030: ν (CC), ν (CN), ν (CO)	Positive link—1906: free -OH (water). Negative link—1430, 1970, and 2028: $\nu_{as} + \nu_s$ (H ₂ O), weakly bounded water; ν_{as} (NH) + amide II and $\nu_{as} + \delta$ (H ₂ O), very strongly bounded water
1716: ν (C=O)	Negative link—1430 and 1906: $\nu_{as} + \nu_s$ (H ₂ O), weakly bounded water and free -OH (water)

ν : stretching; ν_s : symmetric stretching; ν_{as} : asymmetric stretching; δ : rocking.

Table 5 – Regression models statistical summary

	LV	Relative RMSECV (%)	R ²	Relative RMSEP (%)	Pre-treatment	R.S.D. (%)		
						Min.	Max.	Average
NIR								
Fat	3	7.0	0.96	1.20	Second derivative	0.10	2.5	0.54
Nitrogen	7	1.7	0.98	0.09	Second derivative	0.05	1.1	0.32
Moisture	6	5.2	0.94	0.40	Second derivative	0.01	1.40	0.43
FTIR								
Fat	7	10.4	0.94	1.40	SNV	0.4	26	4.7
Nitrogen	7	3.9	0.95	0.138	SNV	0.3	9.2	1.7
Outer product between NIR and FTIR								
Fat	5	7.6	0.96	1.2	Second derivate (NIR) SNV (FTIR)	0.10	16	2.2
Nitrogen	11	2.2	0.96	0.10	Second derivate (NIR) SNV (FTIR)	0.12	3	0.73
Moisture	6	5.9	0.93	0.40	Second derivate (NIR) SNV (FTIR)	0.02	12	0.99

there is a link between symmetric and asymmetric stretching modes of CH₂ and free -OH (water) as well as strong and weakly bounded water molecules. This seems to indicate that those relationships are mainly provided by hydrogen bonding.

4. Concluding remarks

The results presented in this work show that NIR and FTIR can be used as quick tools to predict the content of fat, nitrogen, and moisture in cocoa powder. Furthermore, this work introduced outer product analysis as data fusion technique for the analysis of cocoa powder samples. It was shown that outer product can be useful for two main reasons: (i) prediction

of the parameters of interest using the information present in both spectral domains; (ii) finding relationships between both domains as a function of the parameters of interest. This approach allowed characterising frequencies in one domain based on the information of the other domain. According to the regression models statistical summary given in Table 5. The use of the second derivative of NIR is the recommended procedure to quantify fat, nitrogen, and moisture content in cocoa powders by infrared spectroscopy.

Acknowledgements

Publication of this manuscript was supported by grant MSM 6046137305 (Ministry of Education, Youth and Sports, Czech

Republic) and Program Erasmus/Socrates (Portugal/Czech Republic 2004). Thanks are also due to FCT, Portugal, for funding the Research Unit 62/94 “Química Orgânica, Produtos Naturais e Agro-Alimentares”.

REFERENCES

- [1] <http://www.icco.org>.
- [2] <http://www.finel.fi>.
- [3] J. Moros, F.A. Iñón, S. Garrigues, M. de la Guardia, *Anal. Chim. Acta* 584 (2007) 215.
- [4] A. Trilcová, J. Copiková, M.A. Coimbra, A.S. Barros, H. Kristkova, L. Egert, A. Synytsya, *Chem. Listy* 99 (2005) 821.
- [5] J. Kaffka, K.H. Norris, F. Kulcsár, I. Draskovits, *Acta Alimen.* 11 (1982) 271.
- [6] J.J. Permanyer, M.L. Perez, *J. Food Sci.* 54 (1989) 768.
- [7] J. Tarkosova, J. Copikova, *J. Near Infrared Spectrosc.* 8 (2000) 251.
- [8] E. Whitacre, J. Oliver, R. van den Broek, P. van Engelen, B. Kremers, B. van der Horst, M. Stewart, A. Jansen-Beuvink, *J. Food Sci.* 68 (2003) 2618.
- [9] Y.B. Che Man, Z.A. Syahariza, M.E.S. Mirghani, S. Jinap, J. Bakar, *Food Chem.* 90 (2005) 815.
- [10] R. Goodacre, E. Anklam, *J. Am. Oil Chem. Soc.* 78 (2001) 993.
- [11] R.A. Cocciardi, A.A. Ismail, F.R. van de Voort, J. Sedman, *Milchwissen. Milk Sci. Int.* 56 (2001) 690.
- [12] S. Wold, H. Martens, H. Wold, *Lecture notes in mathematics 1983*, in: A. Ruhe, B. Kägström (Eds.), *Proceedings of the Conference Matrix Pencils*, March 1982, Springer Verlag, Heidelberg, 1983, pp. 286–293.
- [13] H. Martens, T. Naes, *Multivariate Calibration*, 2nd ed., Chichester, Wiley, 1994.
- [14] S. Wold, J. Trygg, A. Berglund, H. Antti, *Chemom. Intell. Lab. Syst.* 58 (2001) 131.
- [15] S. Wold, M. Sjöström, L. Eriksson, *Chemom. Intell. Lab. Syst.* 58 (2001) 109.
- [16] H. Martens, *Chemom. Intell. Lab. Syst.* 58 (2001) 85.
- [17] A.S. Barros, M. Safar, M.F. Devaux, P. Robert, D. Bertrand, D.N. Rutledge, *Appl. Spectrosc.* 51 (1997) 1384.
- [18] A.S. Barros, I. Mafra, D. Ferreira, S. Cardoso, A. Reis, J.A. Lopes da Silva, I. Delgadillo, D.N. Rutledge, M.A. Coimbra, *Carbohydr. Polym.* 50 (2002) 85.
- [19] D.N. Rutledge, A.S. Barros, F. Gaudard, *Magn. Res. Chem.* 35 (1997) S13.
- [20] D.N. Rutledge, A.S. Barros, R. Giangiacomo, *Interpreting near infrared spectra of solutions by outer product analysis with time domain NMR*, in: G.A. Webb, P.S. Belton, A.M. Gil, I. Delgadillo (Eds.), *Magnetic Resonance in Food Science: A View to the Future*, Royal Society of Chemistry, 2001, pp. 179–192.
- [21] C. Di Natale, M. Zude-Sasse, A. Macagnano, R. Paolesse, B. Herold, A. D’Amico, *Anal. Chim. Acta* 459 (2002) 107.
- [22] M.H. Lopes, A.S. Barros, C. Pascoal Neto, D.N. Rutledge, I. Delgadillo, A.M. Gil, *Biopolymers* 62 (2001) 268.
- [23] B. Jaillais, R. Pinto, A.S. Barros, D.N. Rutledge, *Vib. Spectros.* 39 (2005) 50.
- [24] D.N. Rutledge, D.J.-R. Bouveresse, *Chemom. Intell. Lab. Syst.* 85 (2007) 170.
- [25] J. Forshed, R. Stolt, H. Idborg, S.P. Jacobsson, *Chemom. Intell. Lab. Syst.* 85 (2007) 179.
- [26] AOAC Official Method 963.15.
- [27] AOAC Official Method 970.22.
- [28] AOAC Official Method 931.04.
- [29] Q.-S. Xu, Y.-Z. Liang, *Chemom. Intell. Lab. Syst.* 56 (2001) 1.
- [30] A.S. Barros, D.N. Rutledge, *Chemom. Intell. Lab. Syst.* 73 (2004) 245.
- [31] A. Savitzky, J. Golay, *Anal. Chem.* 36 (1964) 1627.
- [32] P. Williams, K. Norris, in: I. Murray, P.C. Williams (Eds.), *Near-Infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists, Inc., 1987, pp. 29–33 (Chapter 2).
- [33] S. Šašić, Y. Ozaki, *Appl. Spectrosc.* 54 (2000) 1327.
- [34] P.B. Dickens, S.H. Dickens, *J. Res. Natl. Inst. Stand. Technol.* 104 (1999) 173–197.
- [35] Y.B. Che Man., Z.A. Syahariza, M.E.S. Mirghani, S. Jinap, J. Bakar, *Food Chem.* 90 (2005) 815.
- [36] S. Gunasekaran, G. Sankari, S. Ponnusamy, *Spectrochim. Acta A* 61 (2005) 117.
- [37] R.M. Silverstein, G.C. Bassler, T.C. Morrill, *Spectrometric Identification of Organic Compounds*, John Wiley & Sons, Inc., 1991, pp. 158–162 (Chapter 3).