**ORIGINAL PAPER** 



# FTIR spectral analysis combined with chemometrics in evaluation of composite mixtures of coconut testa flour and wheat flour

Rasika Gunarathne<sup>1,3</sup> · Nazrim Marikkar<sup>1</sup> · Chandhi Yalegama<sup>2</sup> · Eresha Mendis<sup>3</sup>

Received: 24 September 2021 / Accepted: 30 December 2021 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

#### Abstract

Fourier transform infrared (FTIR) authentication coupled with chemometric evaluation is widely used in qualitative and quantitative analysis of food and agricultural products. This study aimed to characterize coconut testa flour (CTF) of five different Sri Lankan cultivars and to develop statistical models between FTIR data and parameters of CTF, wheat flour (WF) and their admixtures. Flour samples were subjected to FTIR analysis within the range of 4000–500 cm<sup>-1</sup> and proximate data were recorded. Principal component analysis (PCA) was applied for FTIR data and models were developed using both univariate and partial-least-squares (PLS) regression to predict fat, protein, carbohydrates, moisture and level of admixture from FTIR spectra. Results showed that the spectral bands of both CTF and WF were indicative of organic functional groups associated with fat (~1745 cm<sup>-1</sup>), protein (~1655,~1550 cm<sup>-1</sup>), carbohydrates (~3300,~1065 cm<sup>-1</sup>) and moisture (~3300 cm<sup>-1</sup>). Spectral studies on the admixtures confirmed that proportional differences of the biomolecules had a direct influence on the variations occurring in the spectral bands. Application of PCA resulted in a clear discrimination among flour samples with a total variance of 97%. PLS regression showed more precise results than univariate regression. The precision of the developed PLS-regression predictive models was found to increase with the number of selected relevant spectral points and these models resulted in a good R<sup>2</sup> value and low RMSEC and RMSEP values. The results of this study would be helpful for further chemical studies on authentication, product development and nutrition of CTF and WF.

Keywords Coconut testa flour · Flour admixtures · FTIR spectroscopy · Chemometric analysis · PLS regression · PCA

# Introduction

Coconut (*Cocos nuciefera* L.) is one of the major plantation crops of several countries in the Asia–Pacific region. Flesh of coconut fruit is the most important part that gives several edible products that are of high nutritional value. During the industrial processing of the coconut flesh into products, the brown color outer layer is removed as a way to prevent unappealing brown color on finished-products. The removed outer layer of the flesh traditionally known as testa is an underutilized byproduct, which is about 18%

Nazrim Marikkar nazrim.ma@nifs.ac.lk

(w/w) of the coconut flesh. A number of investigations conducted in the recent past suggested its potential use as a raw material for bioactive natural products [1-5]. Coconut testa flour (CTF) obtained by crushing the defatted residue of testa can be a raw material for functional food preparation. A preliminary study on the composition of CTF by Marasinghe et al. [6] indicated that it had  $\sim 59.24\%$ carbohydrates, ~ 32.22% protein and ~ 23.49% fat. Being a non-cereal flour, CTF would differ in composition from all-purpose wheat flour (WF) that is used as staple food in many countries. A comparative study between CTF and WF showed that CTF had ~ 24% of protein and ~ 10% of fat, while WF had ~ 12% of protein and ~ 2% of fat [7]. Despite this, several other studies carried out to evaluate the composition of different types of WF indicated that there were variations based on cultivar differences [8-10]. Particularly, the studies by Majzoobi et al. [8]. and Sujka et al. [11], clearly explicated the possible compositional variations among different wheat cultivars. The results of the WF composition obtained by Zucco et al. [9] were

<sup>&</sup>lt;sup>1</sup> National Institute of Fundamental Studies, Hantana Road, Kandy, Sri Lanka

<sup>&</sup>lt;sup>2</sup> Coconut Research Institute, Lunuwila, Sri Lanka

<sup>&</sup>lt;sup>3</sup> Postgraduate Institute of Agriculture, University of Peradeniya, Peradeniya, Sri Lanka

similar to those reported by Marikkar et al. [7]. From the product development perspective, supplementing WF with CTF is a feasible way for novel products by considering the changing proportions of fat, protein, carbohydrate etc. This warrants analysis and authentication of WF, CTF and their admixtures by spectroscopic techniques such as FTIR.

FTIR is a convenient and rapid analytical tool that is used for authentication of various agricultural and food products [12]. It is reckoned as an excellent method for quantitative analysis of foods as the intensity variations of the spectral bands are tightly associated with the concentration of a particular bio-molecule or compound [13]. The multivariate data analysis is a widely applied tool in FTIR data processing for qualitative and quantitative applications [14] and amongst, PCA and PLS regression being frequently used in food related analyses. However, previous investigations on mid-infra-red spectral characterization of CTF or its constituent biomolecules such as fat, protein, carbohydrate, etc. is scanty. Although Amir et al. [15] demonstrated the use of mid FTIR spectroscopy in determining some chemical parameters of WF, spectroscopic comparison dealing with CTF, WF and their admixtures were not undertaken. Hence, this study aims to characterize CTF, WF and their admixtures in terms of the bio molecular functional groups present in them using FTIR spectroscopy and focus on employing the spectral data in such a way to develop qualitative and quantitative statistical interpretations. Thereby, it will help develop a rapid quantitative detection procedure for major biomolecules present in CTF, WF and their admixtures.

# **Materials and methods**

## Materials

Coconuts of 12-month maturity were collected from five different local cultivars [i.e. Gonthembili (GT), Ranthembili (RT), San Raman (SR), TallxTall (TT), and commercial hybrid (COM)] maintained at the varietal blocks of Coconut Research Institute, Lunuwila, Sri Lanka during the period January 2019 to March 2019. CTF was obtained according to the method described by Marasinghe et al. [6]. Briefly, nuts were de-husked, deshelled and the testa were peeled off followed by disintegration and drying at 70 °C. Cold press oil extraction was carried out for dried coconut testa of each cultivar and the remnant was ground into fine particles to prepare coconut testa flour (CTF). Six samples of commercial wheat flour (WF1, WF2, WF3, WF4, WF5, WF6) were obtained from Serendib Flour Mills (Pvt) Ltd. WF admixtures containing 10%, 25%, 50%, 75% of CTF (w/w) were prepared by mixing CTF of COM variety and WF1.

#### **FTIR measurements**

FTIR analysis of the flour samples was done adopting to KBr pallet method where each flour samples were mixed with KBr (FT-IR grade,  $\geq$  99% trace metals basis, Sigma Aldrich) at a 1:90 ratio and pressed into a pallet [16]. FTIR spectra were collected in the mid-infrared region of 4000–500 cm<sup>-1</sup> by co-adding at 64 scans, resolution of 8 cm<sup>-1</sup> using an FTIR Nicolet iS50 spectrometer (Thermo Nicolet, Madison, WI) equipped with deuterated triglycine sulphate (DTGS) KBr detector and KBr beam splitter. Spectra were recorded as absorbance values at each data point in 4 replicates.

# **Spectral analysis**

The manufacturer's software (OMNIC operating system, version 7.0 Thermo Nicolet) program was used to pre-process the raw spectra and to obtain the peak parameters. The raw spectra were baseline corrected and subjected to scale normalization. The processed spectra were used to collect the peak height and peak area of the major peaks that appeared within  $4000-1500 \text{ cm}^{-1}$ .

#### Analysis of proximate composition

Moisture, crude fat, crude protein and ash contents of coconut testa flour were determined according to standard methods described in AOAC (2005) manual [17]. The carbohydrate content of the flour was calculated by difference [100 - (crude protein + crude fat + ash + moisture)].

## **Statistical analysis**

The proximate data were statistically analyzed by oneway analysis of variance (ANOVA) using Tukey's test of MINITAB (version 17) statistical package at 0.05 probability level. The predictive models for spectral parameters (peak area and peak height) of each major peak up to fingerprint region (4000–1500 cm<sup>-1</sup>) with admixture levels and proximate data were developed using univariate regression analysis procedure in Minitab at 0.05 of probability level.

#### Multivariate data analysis

Chemometric analysis of flour samples was performed using the Unscrambler 9.7 (Camo, USA) software. PCA was carried out for the spectral data ranged from 4000 to 900 cm<sup>-1</sup>. The PLS regression analysis was performed following the method described by Sujka et al. [11]. Thus, different spectral ranges (A; 3700–3000 cm<sup>-1</sup>, B; 3000–2800 cm<sup>-1</sup>, C; 1800–1710 cm<sup>-1</sup>, D;

 $1730-1490 \text{ cm}^{-1}$ , E; 1490- 1210 cm<sup>-1</sup>, F; 1210-900 cm<sup>-1</sup>) were used in developing predictive models. The collected spectra of four replicates of 15 flour samples were randomly grouped into two sets, consisting of 40 and 20 elements, in the way of including at least one replicate of each sample in 20 element set. The set with 40 elements was used for calibration and cross-validation while the remaining other set of 20 elements was used as the unknown independent samples to test the predictive models. The model parameters; coefficient of determination of calibration  $(R_{cal}^{2})$ , coefficient of determination of prediction  $(R_{pre}^{2})$ , root mean square errors of calibration (RMSEC), root mean square errors of prediction (RMSEP), root mean square errors of cross validation (RMSECV) were compared to identify the best predictive model.

# **Results and discussion**

#### **Characterization of FTIR spectra**

Mid-infrared spectral bands of flour substances depict the molecular vibrations of organic functional groups associated with carbohydrates, lipids, proteins, moisture etc. Analyzing the spectral data of samples as illustrated in Figs. 1 and 2, different regions would provide useful information about biomolecules present in CTF and WF. The proximate composition of CTF, WF and their admixture are shown in Table 1. The spectral layout shown in Fig. 1 compares the mid-infrared spectral characteristics of CTF obtained from five local coconut cultivars. By overall, the spectral features of GT, RT, SR, TT and COM varieties were roughly similar, but some important variations were noticed in certain regions due to differences in their proximate compositions.

Fig. 1 FTIR spectral differences of coconut testa flour among coconut cultivars; Gon Thembili (GT), Ran Thembili (RT), San Raman Tall (SR), Tall x Tall (TT), Commercial hybrid (COM)



**Fig. 2** FTIR spectral of coconut testa flour of commercial hybrid (CTF\*), wheat flour of WF1 (WF\*), and admixture containing 10 % CTF\*, 25% CTF\*, 50% CTF\*, 75% CTF\* within the range of 4000–500 cm<sup>-1</sup>

 Table 1
 Proximate composition

 of coconut testa flour, wheat
 flour and admixtures

Flour type	Moisture (%)	Crude protein (%)	Crude fat (%)	Ash (%)	Carbohydrate (%)	
CTF						
COM	$5.41 \pm 0.87^{\rm b}$	$24.39 \pm 0.86^{a}$	$7.82\pm0.06^{\rm b}$	$4.65 \pm 0.49^{a}$	$58.74 \pm 0.88^{d}$	
GT	$2.91\pm0.49^{\rm a}$	$32.36 \pm 0.59^{d}$	$6.90 \pm 0.03^{a}$	$5.26 \pm 0.80^{\rm b}$	$52.83 \pm 0.63^{b}$	
RT	$7.41 \pm 0.52^{\circ}$	$28.39 \pm 1.18^{\circ}$	$10.51 \pm 0.05^{\circ}$	$5.23 \pm 0.38^{\rm b}$	$48.10 \pm 1.25^{a}$	
SR	$3.19\pm0.19^{\rm a}$	$26.12 \pm 0.64^{b}$	$11.77 \pm 0.27^{d}$	$6.51 \pm 0.30^{\circ}$	$52.56 \pm 0.54^{b}$	
T×T	$3.00\pm0.32^a$	$27.39 \pm 0.96^{\circ}$	$6.65\pm0.02^{a}$	$6.26 \pm 0.14^{\circ}$	$56.42 \pm 0.56^{\circ}$	
WF						
WF-1	$9.74 \pm 1.16^{bc}$	$12.61 \pm 0.27^{b}$	$0.94 \pm 0.01^{\circ}$	$0.62 \pm 0.14^{a}$	$76.40 \pm 1.00^{b}$	
WF-2	$8.54 \pm 0.83^{ab}$	$13.69 \pm 0.23^{\circ}$	$0.77\pm0.00^{\rm b}$	$0.67\pm0.12^{ab}$	$76.59 \pm 0.96^{b}$	
WF-3	$10.34 \pm 0.11^{\circ}$	$14.55 \pm 0.08^{d}$	$0.78\pm0.00^{\rm b}$	$0.89 \pm 0.16^{\rm b}$	$73.42 \pm 0.27^{a}$	
WF-4	$8.04\pm0.81^{a}$	$12.50 \pm 0.46^{b}$	$0.01 \pm 0.00^{a}$	$0.93\pm0.05^{\rm b}$	$78.61 \pm 0.98^{\circ}$	
WF-5	$10.98 \pm 0.35^{\circ}$	$11.38 \pm 0.55^{a}$	$0.01\pm0.00^{\rm a}$	$0.85 \pm 0.06^{ab}$	$76.47 \pm 0.22^{b}$	
WF-6	$10.41 \pm 0.67^{\circ}$	$16.01 \pm 0.02^{e}$	$0.01 \pm 0.00^{a}$	$1.85 \pm 0.21^{\circ}$	$71.93 \pm 0.76^{a}$	
Admixture series						
WF-1	$9.74 \pm 1.16^{b}$	$12.61 \pm 0.27^{a}$	$0.94 \pm 0.01^{a}$	$0.62 \pm 0.14^{a}$	$76.40 \pm 1.00^{e}$	
CTF 10%	$10.93\pm0.60^{\rm b}$	$17.18\pm0.15^{\rm b}$	$1.76 \pm 0.01^{b}$	$1.59\pm0.27^{a,b}$	$70.01 \pm 2.96^{d}$	
CTF 25%	$10.30 \pm 0.65^{b}$	$19.77 \pm 0.17^{\circ}$	$2.55 \pm 0.01^{\circ}$	$2.36\pm0.38^{\rm b}$	$65.12 \pm 0.42^{\circ}$	
CTF 50%	$9.81 \pm 0.52^{\rm b}$	$21.14 \pm 0.04$ <sup>cd</sup>	$5.57\pm0.02^d$	$4.25 \pm 0.62^{\circ}$	$59.06 \pm 0.56^{a}$	
CTF 75%	$9.47 \pm 0.40^{\rm b}$	$21.45 \pm 0.04$ <sup>cd</sup>	$6.32 \pm 0.02^{e}$	$5.28 \pm 0.88^{\circ}$	$58.35 \pm 1.13^{a}$	
CTF	$5.41\pm0.87^{a}$	$24.39 \pm 0.86^d$	$7.82\pm0.06^{\rm f}$	$4.65 \pm 0.49^{\circ}$	$5874 \pm 0.88^{a}$	

Each value in the table represents the mean of three replicates. Means that do not share a similar superscription letter are significantly different at 95% confidence ( $\alpha$ =0.05)

COM commercial hybrid; GT Gon Thembili; RT Ran Thembili; SR San Ramon; TT TallxTall; WF wheat flour, CTF coconut testa flour

## Region from 4000 to 2000 $cm^{-1}$

The spectra of CTF in Fig. 1 shows the absorption band (a) occurring at around  $3600 - 3100 \text{ cm}^{-1}$  had a significant contour where the peak was broad and blunt. In the analysis of flour samples, this is generally attributed to stretching vibration of O-H group associated with soluble and non-soluble carbohydrates [15]. Previously, Hospodarova et al. [18] assigned the broad band at 3331 cm<sup>-1</sup> to hydroxyl groups present in cellulosic fibre. In another study, Ibrahim et al. [19] assigned the broad-blunt band in the region 3876–3100 cm<sup>-1</sup> to O–H stretching vibrations of both D-glucose and D-fructose. The exact frequency of this band varies from 3303 to 3374 cm<sup>-1</sup> among different cultivars. This kind of changing nature have been already noticed by other researchers who dealt with varietal differences of WF [15]. Researchers commented that the presence of moisture and other molecules like alcohols, phenols and hydro peroxides present in samples might also contribute to the absorption peaks at ~  $3300 \text{ cm}^{-1}$  based on the O-H stretching vibrations [15, 20]. However, it is reasonable to assume that the highest contribution to band (a) comes from O-H stretching vibrations of carbohydrate as the total carbohydrate contents of these samples far exceeded their moisture content (Table 1). A spectral comparison among the samples namely, WF1 (WF\*), CTF of COM cultivar (CTF\*) and their admixtures are shown in Fig. 2. According to Fig. 2, WF\* exhibited a similar O–H stretching vibration peak corresponding to starch and other carbohydrate components at  $3372 \text{ cm}^{-1}$ . However, the intensity of the peak corresponding to WF\* was comparatively higher than that of CTF\* because of the higher proportion of total carbohydrate content in WF\* (Table 1). According to the spectral overlay (Fig. 2), all admixtures had this peak but their exact frequencies and intensities were slightly varied due to compositional differences based on WF\*/CTF\* mixing ratio (Table 1).

The spectra illustrated in Fig. 1 shows the absorption bands (b) and (c), which were generally attributed to asymmetric stretching vibrations of C–H (CH<sub>2</sub>) and symmetrical stretching vibrations of C–H (CH<sub>2</sub>), respectively [21]. In fact, CH<sub>2</sub> groups present in the aliphatic chain of protein and fatty molecules are commonly responsible for the appearance of these two peaks (b and c). Previous studies showed that strong peaks of (b) and (c) appearing in Mid-IR spectra of most plant oils and animal fats were due to methyl and methylene groups present in the aliphatic chain of fatty molecules [22, 23]. According to Fig. 1, all cultivars displayed the most prominent-sharp peak (b) at 2924 cm<sup>-1</sup> and the adjacent peak (c) with comparatively low intensity at around 2854 cm<sup>-1</sup>. However, with regard to these two peaks (b and c) differences were seen between CTF\* and WF\* as shown in Fig. 2. When compared to CTF\*, peak (c) of WF\* at ~ 2852 cm<sup>-1</sup> was almost absent while peak (b) resulting at 2930  $\text{cm}^{-1}$  appeared as a medium size peak with low-intensity and sharpness. Interestingly, a similar feature was previously noticed in the mid-IR spectrum of spelt wheat flour samples studied by Sujka et al. [11]. This could be because the fat content of WF\* was very low when compared to that of CTF\* as shown in Table 1. It was of an interest to see the effect of changing proximate composition of admixtures on the spectral characteristics. According to the Fig. 2, the peak (c) started to appear at ~ 2856 cm<sup>-1</sup> in the 10% level of admixture, and there were continual improvements in the sharpness and intensity of these peaks (b and c) at 25%, 50% and 75% admixture levels, indicating the increasing proportion of fat in the admixtures (Table 1).

## Region from 2000 to 1000 $cm^{-1}$

The spectra illustrated in Fig. 1 shows the absorption band (d) with frequency at  $1745 \text{ cm}^{-1}$ , which is due to the stretching vibration of carbonyl ester associated with fatty molecules of CTF. According previous studies, this peak is generally taken as a reference for fatty molecules in biological systems as it is the most prominent peak in both plant oils and animal fats [22, 23]. Some studies even investigated the correlation of this peak to saponification values of oils and fats occurring in the nature. The spectral overlay presented in Fig. 1 shows that the exact frequencies of this peak (d) displayed by all coconut cultivars were exactly 1745 cm<sup>-1</sup>. According to Kamara et al. [24], the peaks corresponding to stretching vibration of carbonyls were hardly seen in the white and yellow forms of defatted foxtail millet flour, confirming that fats occur only at trace level in these two flour types. While studying spelt flour, Sujka et al. [11] also observed a very weak C=O stretching vibration at the relevant wave number  $(1745 \text{ cm}^{-1})$  due to low-fat content. A similar observation was evidently seen in the present study. According to the spectral comparison between WF\* and CTF\* as shown in Fig. 2, peak (d) was almost absent in WF because of the low amount of fat in it (Table 1). Further, the peak (d) was found to emerge gradually with the increasing amount of CTF\* due to increasing fat content. According to Fig. 1, peak (d) started to appear at ~  $1744 \text{ cm}^{-1}$  in the 10% level of admixture and its sharpness and intensity continued to improve as noticed at 25%, 50% and 75% admixture levels.

The spectra illustrated in Fig. 1 shows the absorption bands (e) and (f), which were generally referred to as amide band I and amide band II confirming the presence of proteins in flour samples [25]. In a previous study involving different wheat flour types, an estimated content of 7 to 13% protein was found among different wheat varieties [8].

The prominent peak (e) resulted around 1652-1658 cm<sup>-1</sup> was due to the presence of amide band-I attributing to the C=O stretching of amides while the adjacent small peak (f) around 1548–1551 cm<sup>-1</sup> was ascribed to amide band-II which has primarily derived from N-H bending (40-60%) and secondarily from C-N stretching (18-40%) vibration respectively [15, 25]. The spectral comparison between WF\* and CTF\* in Fig. 2 shows the presence of both peaks (e and f) as these two types of flour are rich in protein (Table 1). In the spectra of flour admixtures, these two peaks appeared with some minor variations in their exact frequencies. The spectra illustrated in Fig. 1 demonstrated that the peaks (g) appeared around ~  $1450 \text{ cm}^{-1}$  was due to bending vibrations of aliphatic hydrocarbons of CH<sub>2</sub> and CH<sub>3</sub> while the peak (h) appeared around ~ 1375 cm<sup>-1</sup> were due to the bending vibrations of aliphatic hydrocarbons of CH<sub>3</sub> [16, 21, 26]. Previously Coates [27] stated that the occurrence of a strong methylene/methyl band at 1470 cm<sup>-1</sup> and a weak methyl band at 1380 cm<sup>-1</sup> are indicative of a long-chain linear aliphatic structures.

The spectra illustrated in Fig. 1 shows two absorption peaks around 1236  $\text{cm}^{-1}$  (i) and 1159  $\text{cm}^{-1}$  (k), which were previously ascribed to the C-O stretching vibration of esters [21]. This is further confirmed by Ain Najwa et al. [28] who also assigned peaks "i" and "k" to C-O stretching vibration of ester groups associated with fatty molecules. As shown in Fig. 1, the blunt-broad peak (m) appearing at 1064–1067 cm<sup>-1</sup> represented the characteristic fingerprint of carbohydrate molecules such as starch, pectin etc. According to several previous reports, carbohydrates generally show their characteristic bands in this region [19, 29]. Previously Ferro et al. [29] stated that the vibrational peaks corresponding to carbohydrates were found at around 960-1130 cm<sup>-1</sup> which is a complex region of the spectra with overlapping bands due to carbohydrates. A similar situation was seen previously in the case of spectra recorded for spelt-flour, which is a sub-species of wheat [11]. According to Fig. 2, the sharp high-intense peak was found at 1020 cm<sup>-1</sup> for WF\* while this peak appeared as a low-intense-weak peak at 1066  $\text{cm}^{-1}$  in the case of CTF\*. This could be probably due to high carbohydrate content of WF\* when compared to CTF\* as seen from Table 1. In the spectral overlay of admixtures depicted in Fig. 2, the sharpness of the peak (m) was found to decline gradually with the increasing proportion of CTF\* in the admixtures. This feature has resulted due to the variations in total carbohydrate contents in the admixture after mixing with CTF\* (Table 1).

#### **Regression analysis with peak parameters**

The results of the linear regression analysis carried out using proximate analysis data with the obtained area and height of peak maxima are shown in Table 2. The results showed that

	Peak	Parameter	Regression equation	$\mathbb{R}^2$	SE
Protein	Е	Н	P=-15.23+83.80 H	0.67	7.77
		А	P = -10.98 + 0.9994 A	0.74	0.08
	F	Н	P=3.39+133.79 H	0.78	9.41
		А	P=8.202+1.909 A	0.80	0.13
	В	Н	P=8.003+29.55 H	0.80	1.94
		А	P = 2.87 + 0.8475 A	0.61	0.09
	С	Н	P=13.965+51.09 H	0.78	3.53
		А	P = 14.240 + 2.133 A	0.80	0.14
Fat	D	Н	F=0.408+18.275 H	0.89	0.84
		А	F=0.521+0.8317 A	0.95	0.03
	В	Н	F = -3.364 + 18.910 H	0.91	0.79
		А	F = -6.899 + 0.5550 A	0.72	0.05
	С	Н	F=0.491+32.35 H	0.87	1.64
		А	F=0.612+1.3708 A	0.92	0.05
Carbohydrate	А	Н	C=114.37 - 63.27 H	0.74	4.98
		А	C = 124.48 - 0.1920 A	0.73	0.02
	В	Н	С=83.79 – 47.33 Н	0.84	2.77
		А	C=91.83 - 1.349 A	0.63	0.14
	С	Н	С=74.037 – 80.01 Н	0.78	5.55
		А	C=73.697 - 3.375 A	0.82	0.21
Moisture	А	Н	M=17.55 – 12.27 H	0.34	2.25
		А	M=18.47 - 0.03386 A	0.28	0.01

*P* protein%, *F* fat%; *C* carbohydrate%; *M* moisture%; *A* area of the peak, *H* height if the peak,  $R^2$  coefficient of determination; *SE* standard error

\*All regressions were significance (p=0.00) at 95% confidence level ( $\alpha=0.05$ )

the protein content has a moderate prediction with the height and area of peaks; e, f, b, c as the maximum coefficient of determination  $(R^2)$  of the regression falls below 0.8 while the area of peak-c showed the highest  $R^2$  value (0.8) with a low standard error-coefficient (SE) of 0.14. However, the regression between fat content and the area of peak-d was found to have the highest  $R^2$  (0.95) value with lower SE (0.03) followed by the area of peak-c (R<sup>2</sup>=0.92, SE=0.05) and height of peak-b ( $R^2 = 0.91$ , SE = 0.79) when compare the overall regression results. The regression model between the carbohydrate content and the area and height of the selected peaks; a, b, c showed a moderate fit ( $R^2 < 0.84$ ) with the data. Meanwhile, the  $R^2$  of the regressions of moisture content was found fall below 0.34 exhibiting the poor fitting of the data with the developed models. The obtained results explained that each biomolecule present in the flour samples could represent several peaks in the spectra. Further, major biomolecules namely, protein, carbohydrate and fat showed a considerably good relationship with both peaks b and c.

The linear regression analysis results obtained for the percentage of admixture with the spectral parameters are shown in Table 3. According to the data, both height and peak showed high  $R^2$  values for the produced regressions. However, the obtained SE for the regression between the

Table 2Results of linearregression analysis of proximatedata and FTIR spectral data

 
 Table 3
 Results of linear regression analysis of percentage of admixture and FTIR spectral data of CTF

Peak	Parameter	Regression equation*	R <sup>2</sup>	SE
A	Н	AD=-197.2+291.3 H	0.90	20.5
	А	AD=-273.1+0.9641 A	0.90	0.07
В	Н	AD = -60.48 + 293.8 H	0.95	13.8
	А	AD = -120.75 + 9.621 A	0.98	0.28
С	Н	AD = -0.49 + 474.2 H	0.96	20.3
	А	AD=1.59+20.821 A	0.96	0.89
D	Н	AD = -4.66 + 283.3 H	0.93	16.0
	А	AD = -1.26 + 12.110 A	0.98	0.40
Е	Н	AD = -283.5 +766.9 H	0.88	60.1
	А	AD = -233.0 + 8.872 A	0.94	0.49
F	Н	AD = -124.4 + 1545 H	0.90	40.8
	А	AD=-98.42+25.787 A	0.98	0.77

AD percentage of admixture, A area of the peak, H height if the peak,  $R^2$  coefficient of determination, SE standard error

\*All regressions were significance (p=0.00) at 95% confidence level ( $\alpha$ =0.05)

percentage of admixture and the height was noticeably higher which may affect the precision of the regression. The regressions of the percentage of CTF showed a strong fit with the area of peaks with higher  $R^2$  and low SE for all selected peaks; peak-a ( $R^2=0.9$ , SE=0.07), peak-b ( $R^2=0.98$ , SE=0.28), peak-c ( $R^2=0.96$ , SE=0.89), peak-d ( $R^2=0.98$ , SE=0.40), peak-e ( $R^2=0.94$ , SE=0.49) and peak-f ( $R^2=0.98$ , SE=0.77).

### Multivariate data analysis

## PCA

PCA was applied for discrimination of flour samples using the FTIR spectra within the range of 900–4000 cm<sup>-1</sup>. PCA analysis is useful in identifying patterns within data sets focusing on their alikeness and differences in the form of a score plot which is defined by PC1 and PC2 that explain the highest variation and second highest variation, respectively [30]. The score plot of PCA analysis in Fig. 3 depicts a good separation along the principal component 1 (PC1) with an 89% of variance and 8% variance for principal component 2 (PC2), making up 97% of total variance. According to Fig. 3, the CTF of all five cultivars is located on the negative side of PC1 while all the wheat flour samples are located on the positive side of PC1. In this, CTF of different coconut cultivars showed a good separation along PC1 and PC2 emphasizing their varietal differences. Further, cultivars namely RT, SR and TT were remained in the left upper quadrant while COM and GT were left in the lower quadrant. The WF samples namely W2, W3, W4, W6 were located in the right upper quadrant while W1 and W5 remained in the right lower quadrant. However, the separation among WF was found to be minimal for samples such as W2, W3, W4 and W6. In the case of admixture samples, both 10% and 25% CTF\* admixtures were found to remain in the right lower quadrant where 10% CTF\* showed minimal separation from W1 while 25% CTF\* located adjoining to both W1 and 10% CTF\* samples. The 75% CTF\* and 50% CTF\* were positioned in the left lower quadrant and 75% CTF\* was found in a position where it was in close proximity to COM cultivar exhibiting a low variation in separation. The 50% CTF\* admixture, however, was found far from CTF, W1 and all other admixtures, positioning very much closer to the origin of PC1. Thereby, the distribution of data in the score plot explained the changes that take place in spectral attributes in flour admixtures with the gradual increment of incorporated CTF content. These gradual variations could be attributed to the changes in



**Fig.3** Score plot of PCA obtained for coconut testa flour, wheat flour and admixtures. *COM* commercial hybrid; *GT* Gon Thembili; *RT* Ran Thembili; *SR* San Ramon; *TT* TallxTall; *W1–5* wheat flour types

the intensity of different functional groups peaks that are accompanied by the compositional changes of the flour mixtures as shown in Figs. 1, 2 and Table1.

#### **PLS regression analysis**

Table 4Statistical parametersof PLS models developedfor proximate data and FTIRspectral data of flour samples

Table 4 depicts the results of PLS regression carried out using calculated proximate data and FTIR spectral data. According to the data, the model P3 was identified as the best model among the other models developed for protein content as it exhibited the lowest RMSEC (1.81), RMSECV (2.11), RMSEP (1.94) with the highest  $R_{cal}^2$  (0.92) and  $R_{nre}^2$ (0.90). As for the fat content, F3 model showed the minimum RMSEC (0.76), RMSECV (0.85), RMSEP (0.90) with the highest  $R_{cal}^{2}$  (0.96) and  $R_{pre}^{2}$  (0.95). Among the three models developed for carbohydrate content, C3 model was best as it resulted in the lowest RMSEC (1.54), RMSECV (2.3), RMSEP (1.48) with the highest  $R_{cal}^2$  (0.98) and  $R_{pre}^2$ (0.98). The two models developed for the moisture content showed equal  $R_{cal}^{2}$  values (0.88) and RMSEC (0.96), however, the M2 model showed a lower RMSECV (1.33), RMSEP (0.95) and highest  $R_{pre}^2$  (0.91) when compared to the M1 model.

The PLS models constructed using CTF admixture levels and FTIR spectral data are shown in Table 5. According to the results, all models showed higher  $R_{cal}^2$  and  $R_{pre}^2$  values which are closer to 1 and low RMSEC, RMSECV and RMSEP values. Hence, these all models that were developed using the selected spectral regions which represent different biomolecules could be useful in predicting the percentage of CTF in a WF and CTF admixture. However, considering the lowest RMSEC, RMSECV and RMSEP values of the models, P\*3 ( $R_{cal}^2 = 1$ ,  $R_{pre}^2 = 0.99$ , RMSEC = 1.96, RMSECV = 3.00, RMSEP = 2.53) was noted as the best model. Therefore, it is fair to say that the changes in spectral regions (B, D and E) that are known to represent protein molecules may have well represented the compositional changes of the CTF and WF admixtures.

Overall, the PLS regression models were found to serve with more precision than the models developed by linear regression considering a single factor; peak height/peak area (Tables 2, 3). This fact was further confirmed by Sujka et al. [11], who reported the results of both classic square regression and PLS of different flour types. Further, it could be seen that the precision of the PLS model was also positively influenced by the number of relevant spectral points used in developing the model. More specifically, the inclusion of the spectral ranges of B and E when calibrating the models for protein, fat and carbohydrate has increased the precision of the models (Tables 4) exhibiting a certain correlation with C–H stretching region and some of the minor peaks found in the fingerprint region (Fig. 1).

Previously, PLS regression has been applied for flours obtained from different grain types to come up with statistical models for proximate parameters using FTIR spectral data, which resulted in precise models for protein, fat and moisture. In this study of Sujka et al. [11], a wide range of spectral regions which showed distinctive changes among

Parameter	Model	Spectral range	R <sub>cal</sub> <sup>2</sup>	RMSEC	RMSECV	R <sub>pre</sub> <sup>2</sup>	RMSEP	Factors number
Protein	P 1	D	0.91	1.99	2.26	0.87	2.26	3
	P 2	B, D	0.91	1.94	2.32	0.87	2.22	4
	P 3	B, D, E	0.92	1.81	2.11	0.90	1.94	3
Fat	F 1	С	0.94	0.95	0.99	0.94	0.99	2
	F 2	B, C	0.95	0.90	0.95	0.94	0.99	2
	F 3	B, C, E	0.96	0.76	0.85	0.95	0.90	2
Carbohydrate	C 1	А	0.95	2.27	3.36	0.94	2.40	6
	C 2	A, F	0.94	2.61	3.17	0.92	2.62	4
	C 3	A, B, E, F	0.98	1.54	2.30	0.98	1.48	6
Moisture	M 1	А	0.88	0.96	1.22	0.88	1.10	5
	M 2	A, E, F	0.88	0.96	1.13	0.91	0.95	4

 $R_{cal}^2$  coefficient of determination of calibration;  $R_{pre}^2$  coefficient of determination of prediction; *RMSEC* root mean square errors of calibration; *RMSEP* root mean square errors of prediction; *RMSECV* root mean square errors of cross validation

A: 3700–3000 cm<sup>-1</sup>

B: 3000–2800 cm<sup>-1</sup>

C:  $1800-1710 \text{ cm}^{-1}$ 

D: 1730–1490 cm<sup>-1</sup>

E: 1490–1210 cm<sup>-1</sup>

F: 1210–900 cm<sup>-1</sup>

 Table 5
 Statistical parameters

 of PLS models developed for
 the percentage of admixture and

 FTIR spectral data of CTF
 FTIR

Parameter	Model	Spectral range	R <sub>cal</sub> <sup>2</sup>	RMSEC	RMSECV	R <sub>pre</sub> <sup>2</sup>	RMSEP	Factors number
Protein	P*1	D	0.99	2.80	3.91	0.99	4.06	2
	P*2	B, D	1	2.44	3.63	0.99	3.40	2
	P*3	B, D, E	1	1.96	3.00	0.99	2.53	3
Fat	F*1	С	0.98	5.27	6.08	0.98	5.11	2
	F*2	B, C	0.98	5.09	6.04	0.98	4.54	2
	F* 3	B, C, E	0.99	3.22	4.40	0.99	3.08	2
Carbohydrate	C*1	А	0.99	3.82	6.43	0.91	9.84	4
	C*2	A, F	0.99	3.2	4.66	0.98	4.58	3
	C*3	A, B, E, F	0.99	2.91	3.94	0.99	3.41	2
Moisture	M*1	А	0.99	3.82	6.43	0.91	9.82	4
	M*2	A, E, F	0.99	3.34	4.29	0.98	4.16	2

 $R_{cal}^2$  coefficient of determination of calibration;  $R_{pre}^2$  coefficient of determination of prediction; *RMSEC* root mean square errors of calibration; *RMSEP* root mean square errors of prediction; *RMSECV* root mean square errors of cross validation

A: 3700-3000 cm<sup>-1</sup>

B: 3000–2800 cm<sup>-1</sup>

C: 1800–1710 cm<sup>-1</sup>

D: 1730–1490 cm<sup>-1</sup>

E: 1490-1210 cm<sup>-1</sup>

F: 1210-900 cm<sup>-1</sup>

grain types was employed in constructing the models. The success of PLS regression based model development has been evidently seen in several other studies. For instance, PLS regression was successfully used in both FT-IR and FT-NIR spectral analyses of quantitative authentication studies carried out for Chinese and Korean soybeans [12], corn [31], lard adulteration in cake formulation [32], ternary mixtures of red fruit oil, corn oil and soybean oil [22], virgin olive oil [13], canola oil and virgin coconut oil mixtures [33], etc.

In PLS regression analysis, the R<sup>2</sup> is used to elaborate the accuracy of the developed PLS model while RMSEC and RMSECV RMSEP are used to interpret its precision [34]. A model with low RMSEC, RMSEP and RMSECV values and higher R<sup>2</sup> value is reckoned as an acceptable model with high precision and accuracy [11, 14, 22]. Thus,  $R_{pre}^2$  and RMSEP can employ to determine the predictability of the developed model where high  $R_{pre}^2$  and low RMSEP values would indicate high predictability [11, 31, 33]. Selection of the appropriate number of factors is acting as a key point to eliminate the possibilities of developing an over-fit or underfit model which take place due to the addition of too many factors and too few factors respectively [11].

# Conclusion

This study presented a detailed analysis of the mid-infrared spectral bands of CTF of different cultivars and WF as a way to distinctly differentiate them using carbohydrates, proteins, and fats as component biomolecules. The spectral contour of all CTF cultivars was more or less similar to each other despite some minor differences in their peak intensities. However, remarkable spectral differences were noticed between CTF and WF owing to their differences in their proximate compositions. PCA was successfully employed in qualitative discrimination of CTF and their admixtures with a high variance. According to the results of quantitative analysis, it can be seen that PLS regression is more applicable for FTIR spectral analysis rather than univariate regression pertaining to the precision of the results. PLS regression was served as a great quantitative tool for developing statistically significant models for different parameters of studied flours. Therefore, the results of this study convinced that the FTIR coupled with multivariate analysis could serve as an excellent tool that can be used in authentication and quantification of CTF and its admixtures.

Author contributions RG performed the experiment, analyzed, interpreted the data and prepared the manuscript. NM provided the concept, design of the experiment and edited the manuscript. EM and CY reviewed and edited the manuscript.

Acknowledgements Authors gratefully acknowledge the financial support from Inter-Institutional Multidisciplinary Research Grant (Grant No. IIMD/2021/CRI/01) Scheme awarded by Sri Lanka Council of Agricultural Research Policy. This study was partly funded by the National Institute of Fundamental Studies, Hanthana Road, Kandy, Sri Lanka.

#### Declarations

**Conflict of interest** The authors have no conflicts of interest to declare that are relevant to the content of this article.

# References

- K.A. Adekola, A.B. Salleh, U.H. Zaidan, A. Azlan, E. Chiavaro, M. Paciulli, J.M.N. Marikkar, Total phenolic content, antioxidative and antidiabetic properties of coconut (*Cocos nuceifera* L.) testa and selected bean seed coats. Ital. J. Food Sci. 29, 742–753 (2017). https://doi.org/10.14674/IJFS-941
- P. Appaiah, L. Sunil, P.K. Prasanth Kumar, A.G. Gopala Krishna, Composition of coconut testa, coconut kernel and its oil. J. Am. Oil Chem. Soc. 91, 917–924 (2014). https://doi.org/10.1007/ s11746-014-2447-9
- M. Arivalagan, T.K.A. Royb, M. Yasmeena, K.C. Pavithrab, P.N. Jwalaa, K.S. Shivasankarab, M.R. Manikantana, K.B. Hebbara, S.R. Kanadec, Extraction of phenolic compounds with antioxidant potential from coconut (*Cocos nucifera* L.) testa and identification of phenolic acids and flavonoids using UPLC coupled with TQD-MS/MS. LWT - Food Sci. Technol. **92**, 116–126 (2018). https:// doi.org/10.1016/j.lwt.2018.02.024
- V. Geetha, K.P. Bhavana, R. Chetana, A.G. Gopala Krishna, G. Suresh Kumar, Studies on the composition and in-vitro antioxidant activities of concentrates from coconut testa and tender coconut water. J. Food Process. Technol. 7, 88–95 (2016)
- S.V. Ramesh, R. Pandiselvam, P.P. Shameena Beegum, R.M. Saravana Kumar, M.R. Manikantan, K.B. Hebber, Review of *Cocos nucifera* L. testa-derived phytonutrients with special reference to phenolics and its potential for encapsulation. J. Food Sci. Technol. (2021). https://doi.org/10.1007/s13197-021-05310-2
- S.S.K. Marasinghe, J.M.N. Marikkar, C. Yalegama, S. Wimalasiri, G. Seneviratne, R. Weerasooriya, R. Liyanage, Comparison of inter varietal differences in chemical composition and nutritional properties of coconut testa flour. J. Natn. Sci. Found. Sri Lanka 47, 351–358 (2019). https://doi.org/10.4038/jnsfsr.v47i3.8699
- J.M.N. Marikkar, R. Nagaraja, K.M.S. Somawathie, H.P.T.D. Hewapathirana, C. Yalegama, P. Littardi, E. Chiavaro, Effect of coconut testa flour on cookie characteristics. Ital. J. Food Sci. 32, 459–471 (2020)
- M. Majzoobi, A. Farahnaky, S. Amiri, Physicochemical characteristics of starch component of wheat flours obtained from fourteen Iranian wheat cultivars. Int. J. Food Prop. 14, 685–696 (2011). https://doi.org/10.1080/10942910903367613
- F. Zucco, Y. Borsuk, S.D. Arntfield, Physical and nutritional evaluation of wheat cookies supplemented with pulse flours of different particle sizes. LWT - Food Sci. Technol. 44, 2070–2076 (2011). https://doi.org/10.1016/j.lwt.2011.06.007
- K.D.P.P. Gunathilake, C. Yalegama, A.A.N. Kumara, Use of coconut flour as a source of protein and dietary fiber in wheat bread. Asian J. Food Agro-Indus. 2, 382–391 (2009)
- K. Sujka, P. Koczon, A. Ceglinska, M. Reder, H. Ciemniewskagytkiewicz, The application of FT-IR spectroscopy for quality control of flours obtained from Polish producers. J. Anal. Methods Chem. (2017). https://doi.org/10.1155/2017/4315678
- B.J. Lee, Y. Zhou, J.S. Lee, B.K. Shin, J.A. Seo, D. Lee, Y.S. Kim, H.K. Choi, Discrimination and prediction of the origin of Chinese and Korean soybeans using Fourier transform infrared spectrometry (FT-IR) with multivariate statistical analysis. PLoS ONE 13, 1–16 (2018). https://doi.org/10.1371/journal.pone.01963 15

- A. Hirri, M. Bassbasi, S. Platikanov, R. Tauler, A. Oussama, FTIR Spectroscopy and PLS-DA classification and prediction of four commercial grade virgin olive oils from Morocco. Food Anal. Methods 9, 974–981 (2016). https://doi.org/10.1007/ s12161-015-0255-y
- L. Wulandari, Y. Retnaningtyas, H.L. Nuri, Analysis of flavonoid in medicinal plant extract using infrared spectroscopy and chemometrics. J. Anal. Methods Chem. (2016). https://doi.org/10.1155/ 2016/4696803
- R.M. Amir, F.M. Anjum, M.I. Khan, M.R. Khan, I. Pasha, M. Nadeem, Application of Fourier transform infrared (FTIR) spectroscopy for the identification of wheat varieties. J. Food Sci. Technol. 50, 1018–1023 (2013). https://doi.org/10.1007/ s13197-011-0424-y
- A.N. NinaNaqiyah, J.M.N. Marikkar, M.E.S. Mirghani, A.F. Nurulhidaya, N.A.M. Yanty, Differentiation of fractionated components of lard from other animal fats using different analytical techniques. Sains Malaysiana. 46, 209–216 (2017). https://doi.org/ 10.17576/jsm-2017-4602-04
- 17. AOAC, Official Methods of Analysis of the Association of Official Analytical Chemists, 18th edn. (Association of Official Analytical Chemists, Gaithersburg, 2005)
- V. Hospodarova, E. Singovszka, N. Stevulova, Characterization of cellulosic fibers by FTIR spectroscopy for their further implementation to building materials. Am. J. Anal. Chem. 9, 303–310 (2018). https://doi.org/10.4236/ajac.2018.96023
- M. Ibrahim, M. Alaam, H. El-Haes, A.F. Jalbout, A. de Leon, Analysis of the structure and vibrational spectra of glucose and fructose. Eclet. Quím. 3, 15–21 (2006). https://doi.org/10.1590/ S0100-46702006000300002
- R. Thakur, Determination of grain moisture content using FTIR Spectroscopy. Indian J. Res. 4, 34–35 (2015)
- S.Y. Lim, M.S.A. Mutalib, H. Khaza'ai, S.K. Chang, Detection of fresh palm oil adulteration with recycled cooking oil using fatty acid composition and FTIR spectral analysis. Int. J. Food Prop. 21, 2428–2451 (2018). https://doi.org/10.1080/10942912.2018. 1522332
- A. Rohman, Infrared spectroscopy for quantitative analysis and oil parameters of olive oil and virgin coconut oil: A review. Int. J. Food Prop. 20, 1447–1456 (2017). https://doi.org/10.1155/2014/ 785914
- N. Vlachos, Y. Skopelitis, M. Psaroudaki, V. Konstantinidou, A. Chatzilazarou, E. Tegou, Applications of fourier transforminfrared spectroscopy to edible oils. Anal Chim. Acta. 573–574, 459–465 (2006). https://doi.org/10.1016/j.aca.2006.05.034
- M.T. Kamara, Z. Huiming, Z. Kexue, I. Amadou, F. Tarawalie, Comparative study of chemical composition and physicochemical properties of two varieties of defatted foxtail millet flour grown in China. Am. J. Food Technol. 4, 255–267 (2009). https://doi.org/ 10.3923/ajft.2009.255.267
- S. Diblan, P. Kadiroğlu, L.Y. Aydemir, FT-IR spectroscopy characterization and chemometric evaluation of legumes extracted with different solvents. Food Health 4, 80–88 (2018). https://doi. org/10.3153/FH18008
- Y. Lu, C. Du, Y. Shao, J. Zhou, Characterization of rapeseed oil using FTIR-ATR spectroscopy. J. Food Sci. Eng. 4, 244–249 (2015). https://doi.org/10.17265/2159-5828/2014.05.004
- 27. J. Coates, Interpretation of infrared spectra: A practical approach, in *Encyclopedia of Analytical Chemistry*. ed. by R.A. Meyers (Wiley, Hoboken, 2004)
- M.N. Ain Najwa, S. Abbasiliasi, M.N. Marikkar, A. Ariff, M. Amid, D.U. Lamasudin, M.Y. Abdul Manap, S. Mustafa, Defatted coconut residue crude polysaccharides as potential prebiotics: study of their effects on proliferation and acidifying activity of probiotics in vitro. J. Food Sci. Technol. 54, 164–173 (2017). https://doi.org/10.1007/s13197-016-2448-9

- L. Ferro, Z. Gojkovic, A. Gorzsás, C. Funk, Statistical methods for rapid quantification of proteins, lipids, and carbohydrates in Nordic microalga species using ATR-FTIR spectroscopy. Molecules 24, 3237 (2019). https://doi.org/10.3390/molecules24183237
- J.M.N. Marikkar, N.A.M. Yanty, Principal component analysis of fatty acid data to detect virgin coconut oil adulteration by palm olein. Int. J. Coco Res. Dev. 34, 30–38 (2018)
- N. Zhao, Z.S. Wu, Q. Zhang, X.Y. Shi, Q. Ma, Y.J. Qiao, Optimization of parameter selection for partial least squares model development. Sci. Rep. 5, 1–10 (2015). https://doi.org/10.1038/ srep11647
- Z.A. Syahariza, Y.B. Che Man, J. Selamat, J. Bakar, Detection of lard adulteration in cake formulation by Fourier transform infrared (FTIR) spectroscopy. Food Chem. 92, 365–371 (2005). https://doi. org/10.1016/j.foodchem.2004.10.039

- Y.B. Che Man, A. Rohman, Analysis of canola oil in virgin coconut oil using FTIR spectroscopy and chemometrics. J. Food Pharm. Sci. 1, 5–9 (2013)
- R.B. Pebriana, A. Rohman, E. Lukitaningsih, Development of FTIR spectroscopy in combination with chemometrics for analysis of rat meat in beef sausage employing three lipid extraction systems. Int. J. Food Prop. 20, 1995–2005 (2017). https://doi.org/ 10.1080/10942912.2017.1361969

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.