



## Research Article

# Data Fusion Approach Improves the Prediction of Single Phenolic Compounds in Honey: A Study of NIR and Raman Spectroscopies

Haroon Elrasheid Tahir<sup>1</sup>, Muhammad Arslan<sup>1</sup>, Gustav Komla Mahunu<sup>2</sup>, Jiyong Shi<sup>1\*</sup>, Xiaobo Zou<sup>1\*</sup>,  
Mohammed Abdalbasit Ahmed Gasmalla<sup>1,3</sup>, Abdalbasit Adam Mariod<sup>4</sup>

<sup>1</sup>School of Food and Biological Engineering, Jiangsu University, 301 Xuefu Rd., Zhenjiang, Jiangsu 212013, China

<sup>2</sup>Department of Food Science and Technology, Faculty of Agriculture, University for Development Studies, Tamale, Ghana

<sup>3</sup>Department of Nutrition and Food Technology, Omdurman Islamic University, Omdurman, Sudan

<sup>4</sup>College of Sciences and Arts-Alkamil, University of Jeddah, P.O. Box 110, Alkamil 2193, Saudi Arabia

## ARTICLE INFO

### Article History

Received 17 May 2019

Accepted 10 July 2019

### Keywords

Honey  
phenolic compounds  
data fusion  
spectroscopy  
partial least square (PLS)

## ABSTRACT

The combination of Near-Infrared Spectroscopy (NIR) and Raman Spectroscopy (RS) of 100 honey samples collected from different countries were used to develop the calibration model for determination of single phenolic compound. In high-performance liquid chromatography with diode-array detection analysis, 16 phenolic compounds were identified with *p*-hydroxybenzoic acid being the major compound detected in all honey samples. Thus, *p*-hydroxybenzoic acid was used for developing prediction models. Spectral data were modeled individually and using data fusion methodologies. The performance of the model based on RS spectra [ $R_p^2 = 0.9500$ , Root Mean Standard Error of Prediction (RMSEP) = 6.83] was higher than that based on the NIR spectra ( $R_p^2 = 0.8147$ , RMSEP = 13.80). The application of both low-level ( $R_p^2 = 0.9553$ , RMSEP = 6.59) and mid-level ( $R_p^2 = 0.9563$ , RMSEP = 7.95) data fusion together with Partial Least Squares (PLS) had effectively improved the prediction models of NIR but did not enhance prediction models based on RS technique. The results demonstrated that the NIR, RS, and data fusion approaches together with the PLS model could be used as alternative quantitative methods for determination of *p*-hydroxybenzoic acid in honey samples.

© 2019 International Association of Dietetic Nutrition and Safety. Publishing services by Atlantis Press International B.V.  
This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>)

## 1. INTRODUCTION

The growing market and consumer demand for quality foods with health benefits have generated the need for rapid and accurate analytical tools for the determination of bioactive compounds in honey and honey products. For the past years, several studies have reported on the determination of antioxidants in various food commodities. The conventional methods for analysis of phenolic compounds in honey are High-Performance Liquid Chromatography (HPLC) with Diode-Array Detection (DAD) or UV-Vis detector [1,2], whereas in some cases gas chromatography-mass spectrometry [3] and capillary electrophoresis have been used. Regardless of their precision, chemical assays are time-consuming, expensive, and require a large number of solvents [4]. Therefore, simple, rapid, and inexpensive instruments are needed. Spectroscopic technologies are widespread in the analysis of antioxidant in foods [5]. They have also become well-known in the investigation of honey quality because of their benefits (including speediness, directness, and

cost-effectiveness). The existence of novel multivariate tools to deal with several variables simultaneously is additional motivation for fast changes within the domain of measurement technology and the sensors. The current trend of using multiple spectra simultaneously is more promising than a single spectrum method because of a better representation of multivariate analysis. The combined use of nondestructive methodologies and data fusion techniques to predict the phenolic compounds in honey has been briefly described in this study. It is not aimed to provide complete detail of the conventional chemical methods, spectral interpretations, and spectroscopic analysis conditions, but rather to discuss all the main issues related to data fusion as well as the potential implementation of measuring of single phenolic compounds in honey. Data obtained from different kind of instruments usually depend on diverse physical phenomena, thus elucidations of findings are complicated. Some of the spectroscopies may be more accurate than others because these are more sensitive in certain circumstances [6]. Data fusion has been used in various areas: for instance, the combination of headspace-mass spectrometry, Fourier transform mid-infrared spectroscopy and UV-Visible spectrophotometry to predict virgin olive oil sensory descriptors [7], Near-Infrared (NIR), fluorescence, and nuclear magnetic resonance spectroscopy

\*Corresponding authors. Email: [shi\\_jiyong@ujs.edu.cn](mailto:shi_jiyong@ujs.edu.cn); [zou\\_xiaobo@ujs.edu.cn](mailto:zou_xiaobo@ujs.edu.cn)

Peer review under responsibility of the International Association of Dietetic Nutrition and Safety

to assess oil quality [8], electronic noses and spectroscopic instruments to identify the geographical olive oil [9], and electronic tongue and NIR for a rapid umami taste assessment [10]. Other promising methods, spectroscopic data, and chemometric modeling have been proposed for the quantification of antioxidant compounds in wine [11,12], honey [13,14], coffee [15], dark soy sauce [16], roselle tea [17], olive oil [18], chocolate [19], apples [20], raw propolis [21], and *Populus* spp. leaves [22]. In most of these studies, a single technique has been used for prediction of antioxidants.

The aim of the current experiment was to investigate the correlation between spectroscopy data and HPLC measurement. To improve the information quality and to assess the synergistic effect, low- and mid-level data fusion approaches were implemented to combine the data of NIR and Raman spectroscopy, and the outcome was compared and discussed. Up to now, there are no cases of the above-mentioned combined spectroscopy techniques for the prediction of single phenolic compounds in honey.

## 2. MATERIALS AND METHODS

### 2.1. Chemicals Reagents

All reagents, phenolic acids, and flavonoid used for HPLC analysis were of HPLC grade and the other chemicals were of analytical grade. Kaempferol, myricetin, rutin, caffeic acid, ferulic acid, *p*-hydroxybenzoic acid, catechol, chlorogenic acid, syringic acid, vanillic acid, and quercetin were from China Sigma-Aldrich (Shanghai, China). Catechin, chrysin, *p*-coumaric acid, gallic acid, and hydroxycinnamic acid were purchased from Altar Ink Quality Inspection Technology Co., Ltd. (Beijing, China). Milli-Q water purification system (Millipore Corp., Billerica, MA, USA) was used for water purification. Methanol was supplied by Sinopharm Chemical Reagent Co., Ltd Shanghai, China. Solid phase extraction (SPE) with C18 cartridges was supplied by Waters Corporation, Ireland.

### 2.2. Honey Samples

One hundred samples of honey of different botanical origins were obtained from different countries. The common names and botanical and geographical sources of honey samples are presented in Table 1.

### 2.3. Spectroscopic Analysis

For spectroscopic analysis, honey samples were warmed at 40°C and then mixed to dissolve the crystals to form homogenous samples before spectroscopic testing. Two nondestructive sensing techniques including NIR and Raman spectroscopies were obtained from our previous work [2,14].

#### 2.3.1. NIR spectroscopy

An Antaris II near-infrared spectrophotometer (Thermo Electron Co., USA) with a reflectance module was used to obtain the honey spectra. Each spectrum was the average of 32 scanning spectra. The range of spectra was from 10,000 to 4000/cm and each spectrum was measured according to 3.856/cm data intervals, thus, each spectrum resulted in 1557 variables.

**Table 1** | Honey samples

No.	Codes	Botanical origin	Common name	Geographical origin
1	H1	<i>Leptospermum scoparium</i>	Manuka honey	Australian ( $n = 7$ )
2	H2	Polyfloral	-	Australian ( $n = 5$ )
3	H3	Polyfloral	-	Germany ( $n = 4$ )
4	H4	<i>Citrus</i> spp.	Orange blossom	Spain ( $n = 8$ )
5	H5	<i>Eucalyptus</i> spp.	Eucalyptus honey	Spain ( $n = 7$ )
6	H6	<i>Rosmarinus officinalis</i> L.	Rosemary blossom	Spain ( $n = 7$ )
7	H7	<i>Acacia seyal</i>	Talih	Sudan ( $n = 8$ )
8	H8	<i>Ziziphus spina-christi</i>	Sider	Sudan ( $n = 7$ )
9	H9	Polyfloral	-	Sudan ( $n = 8$ )
10	H10	<i>Ziziphus spina-christi</i>	Sider	Pakistan ( $n = 7$ )
11	H11	<i>Vitex negundo</i> L.	Chaste honey	China ( $n = 7$ )
12	H12	<i>Litchi chinensis</i> Sonn.	Lychee honey	China ( $n = 9$ )
13	H13	Polyfloral	Wild flower	China ( $n = 8$ )
14	H14	Polyfloral	-	France ( $n = 4$ )
15	H15	Polyfloral	-	Cameroon ( $n = 3$ )

-, multispecies.

#### 2.3.2. Raman spectroscopy

The DXR Laser Raman Spectrometer (Thermo Fisher Scientific, USA) coupled with excitation laser of 532 nm, the spectrograph allows a resolution of 5/cm using 900 lines/mm, was used to obtain the honey spectra. Each honey was screened using a laser power at the sample of 10 mW. A spectrum from each sample was collected for 5 min using the continuous extended scan from 50 to 3500/cm.

### 2.4. Reference Analysis by HPLC-DAD

Honey samples (10 g) were diluted with 40 mL of purified water, adjusted to pH 2 with HCL (0.1%), and sonicated for 15 min at 25°C. Furthermore, honey solutions were filtered through cotton wool to remove any solid particles. An SPE with C18 cartridges (Waters Corporation, Ireland) was conditioned by passing 10 mL of methanol and 10 mL of purified water. The filtered honey solutions were loaded on the SEP cartridges and washed with 30 mL of acidified purified water (pH 2) to eliminate all sugars and other polar components of honey. The absorbed phenolic compounds were eluted with methanol (2 mL), and then the extracts were filtered through a 0.22- $\mu$ m membrane filter before being quantified by HPLC-DAD. Separation of phenolic compounds was performed on Zorbax SB-C18 column (250 mm, 149  $\times$  4.6 mm, 5  $\mu$ m particle size, Agilent Technologies, USA). The detailed analysis, condition, and validation parameters of HPLC-DAD investigation can be found in our previously published work [2].

### 2.5. Multivariate Data Analysis

The spectral preprocessing and partial least square regression (PLSR) model were computed using Matlab Version 7.1 (Mathworks Inc., Natick, MA, USA). The obtained spectra of NIR and Raman include noises as well as the honey information. Hence,

the pretreatment is important to reduce the interferences and undesirable information, and enhance the contribution of chemical components. On the basis of our previous work [2], the optimum pretreatments for Raman were Savitzky–Golay algorithm (15 data points and a second-order polynomial function) and Multiplicative Scatter Correction (MSC). Whereas for NIR, it was found that standard normal variate performed better as compared with another noted method.

Figure 1 shows the raw spectra of analyzed honey samples. In this work, the whole spectrum regions were (400–10,000/cm) for NIR and (1500–400/cm) for Raman.

From Figure 1, it is clear that honey samples presented characteristic bands and they have the potential to be used for detecting phenolic compound using chemometrics. The detailed interpretations of bands from the two spectroscopic techniques can be found in the literature [2,23–25].

For developing chemometrics model, the spectral data were classified into calibration and prediction sets as described in our previous study [14]. Briefly, all 100 samples were classified randomly based on their corresponding  $y$ -value (viz. the phenolic compound data). With respect to the classification of the calibration/prediction sets, 17 spectra of each 50 types of honey were located for the prediction set. Therefore, the calibration and prediction sets have 64 and 36 spectra, respectively. PLSR is generally used as statistical tool for constructing linear regression models based on the variable matrix  $X$  (spectroscopic data) and variable matrix  $Y$  ( $p$ -hydroxybenzoic

acid data) by guaranteeing that all latent variables were set on the basis of their relevance for predicting  $Y$  [26]. The Latent Variables (LVs) of the PLSR models were determined by leave-one-out cross-validation by mapping the number of factors against the Root Mean Squared Error Cross Validation (RMSECV). The optimum number of LVs was determined by the lowest value of RMSECV. The performance of the PLSR models was evaluated using  $R_c$ ,  $R_p$ , Root Mean Standard Error of Prediction (RMSEP), RMSECV, and the Ratio of Prediction to Deviation (RDP) values. In general, an excellent model must have high correlation coefficients and RDP value with low RMSECV and RMSEP [17,27,28].

## 2.6. Data Fusion

Data fusion can be performed at three levels: low, medium, and high. For low-level data fusion, the original data are simply merged after applying appropriate preprocessing methods. It is also pre-disposed to huge data computation, tedious and worse real-time; high-level data fusion (i.e., decision-making level) requires skills and immense data pretreatments [29]. Mid-level fusion includes extraction of features from each spectral source separately before data fusion then merge them into a single block, which will be processed by a calibration method. Generally, the highest scores from the Principal Component Analysis (PCA) or Partial Least Squares (PLS) regression were combined. In the present study, data compression using PCA were implemented. Briefly, PCA was performed on each data set individually and two approaches were

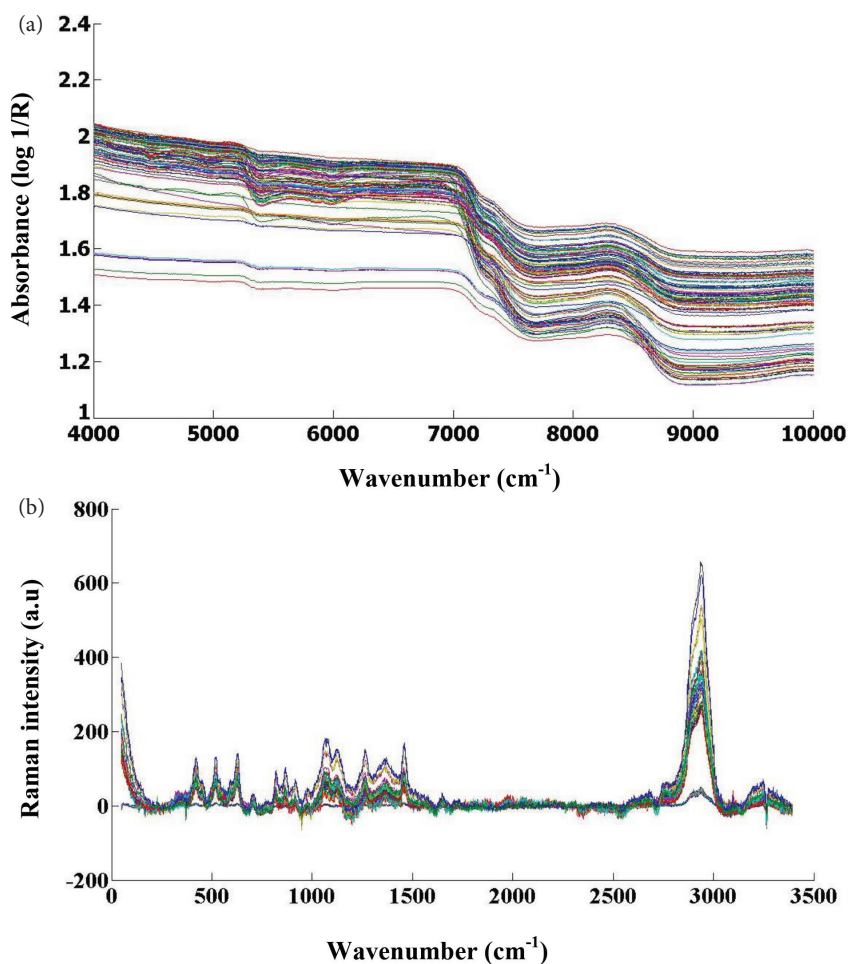


Figure 1 | Raw spectra of different honey varieties. (a) NIR spectra with a reflectance module (4000–10,000/cm) and (b) Raman spectra with Raman shift value (3500–50/cm).

attempted to select the number of PCs to be combined from the two data sets. In the first method (approach 1), a constant number of PCs – 10 were taken for each data set and merged in the fused matrices of the size  $100 \times 20$  (for the two data sets). In the second method (approach 2), a different number of PCs, corresponding to the highest total variance was selected for each data set and merged. Approach 1 (i.e., a fixed number of PCs) improved the prediction results, hence, it was adopted in this experiment.

## 3. RESULTS AND DISCUSSION

### 3.1. Phenolic Compound Profiles

In this study, 16 phenolic compounds have been identified in honey samples from various countries (Table S1). About 11 phenolic compounds have been identified in Australian Manuka honey with *p*-coumaric acid (64.08 mg/100 g) and catechin (60.71 mg/100 g) being the major compounds, whereas chrysin was at a lower level (0.81 mg/100 g). The concentrations of phenolic compounds in Manuka honey are higher than previously reported [30]. The higher concentration of phenolic compounds that were detected could be due to variations in the seasons and the effective approach (e.g., weight and extraction condition) adopted in this study. Similar differences in concentrations were reported in the literature [30–32]. All detected compounds have been previously reported in honey [2,30,33]. Polyfloral honey obtained from Australia showed a fewer number of phenolic compounds (syringic acid, *p*-hydroxybenzoic acid, and vanillic acid).

The honey richest in *p*-hydroxybenzoic acid was polyfloral samples collected from Germany, whereas Australian polyfloral and Rosemary blossom showed lower values. The highest concentrations of syringic acid were present in Cameroonian polyfloral honey (110.06 mg/100 g) and German honey (85.01 mg/100 g), whereas the lowest value was detected in Chinese chaste honey (1.28 mg/100 g). Among the analyzed honey samples, orange blossom showed the highest catechin content (161.81 mg/100 g) followed by Spanish eucalyptus honey (119.58 mg/100 g) and French polyfloral honey (118.74 mg/100 g). Vanillic acid was the major compound in Spanish eucalyptus honey and Pakistani Sider honey and Lychee honey showed the lowest value. Low concentration of gallic acid were detected in polyfloral honeys obtained from France and Germany but was not present in any other sample. The profile of phenolics of Sider honey obtained from Sudan and Pakistan were significantly different (Table S1). Caffeic acid and rutin were only detected in Sudanese Sider honey, whereas catechin and chrysin were only present in Pakistani Sider honey. The concentrations of phenolic compounds in both Sider honey samples varied from those reported in the literature [2,34,35].

The total amount of phenolic compounds measured in the analyzed honeys followed the order: H10 > H8 > H7 > H1 > H3 > H12 > H4 > H14 > H5 > H9 > H15 > H11 > H13 > H2 > H6. Interestingly, significant differences in phenolic compounds were observed among honey samples of the same botanical origin but of different geographical origins and different seasons [1,2]. Among the phenolic compounds analyzed, only *p*-hydroxybenzoic acid was identified in all honey samples examined; thus, this compound was selected for developing a prediction model. Table 2

**Table 2** | *p*-Hydroxybenzoic acid (mg/100 g) in honey samples from different geographical origins

Countries	Minimum	Maximum	Mean	SD	CV
China	23.14	98.15	50.67	33.04	65.25
Pakistan	47.98	49.45	48.68	0.43	0.88
Spain	8.35	45.20	22.11	13.04	58.98
Australian	2.99	25.46	15.50	11.17	72.11
Germany	120.21	120.51	120.32	0.11	0.09
France	28.18	29.60	29.07	0.49	1.68
Sudan	40.98	88.98	65.18	20.29	31.14
Cameroon	22.15	23.25	22.63	0.40	1.77

SD, standard deviation; CV, Coefficient of variation; CV =  $\{SD/Mean\} \times 100$ .

shows the *p*-hydroxybenzoic acid in honey samples from different geographical origins. This variation could be attributed to the differences in geographical and botanical origins of honey (Table 1).

The descriptive data presented in Table 3 indicate that the categories of honey studied significantly vary in *p*-hydroxybenzoic acid content in the calibration and prediction sets. From Table 3, it is clear that the range of reference chemical data in the calibration set cover the range of prediction sets and there are no significant differences between their standard deviations. Thus, the distribution of the samples is convenient both in the calibration and in the prediction sets.

### 3.2. NIR, Raman and Data Fusion Approaches

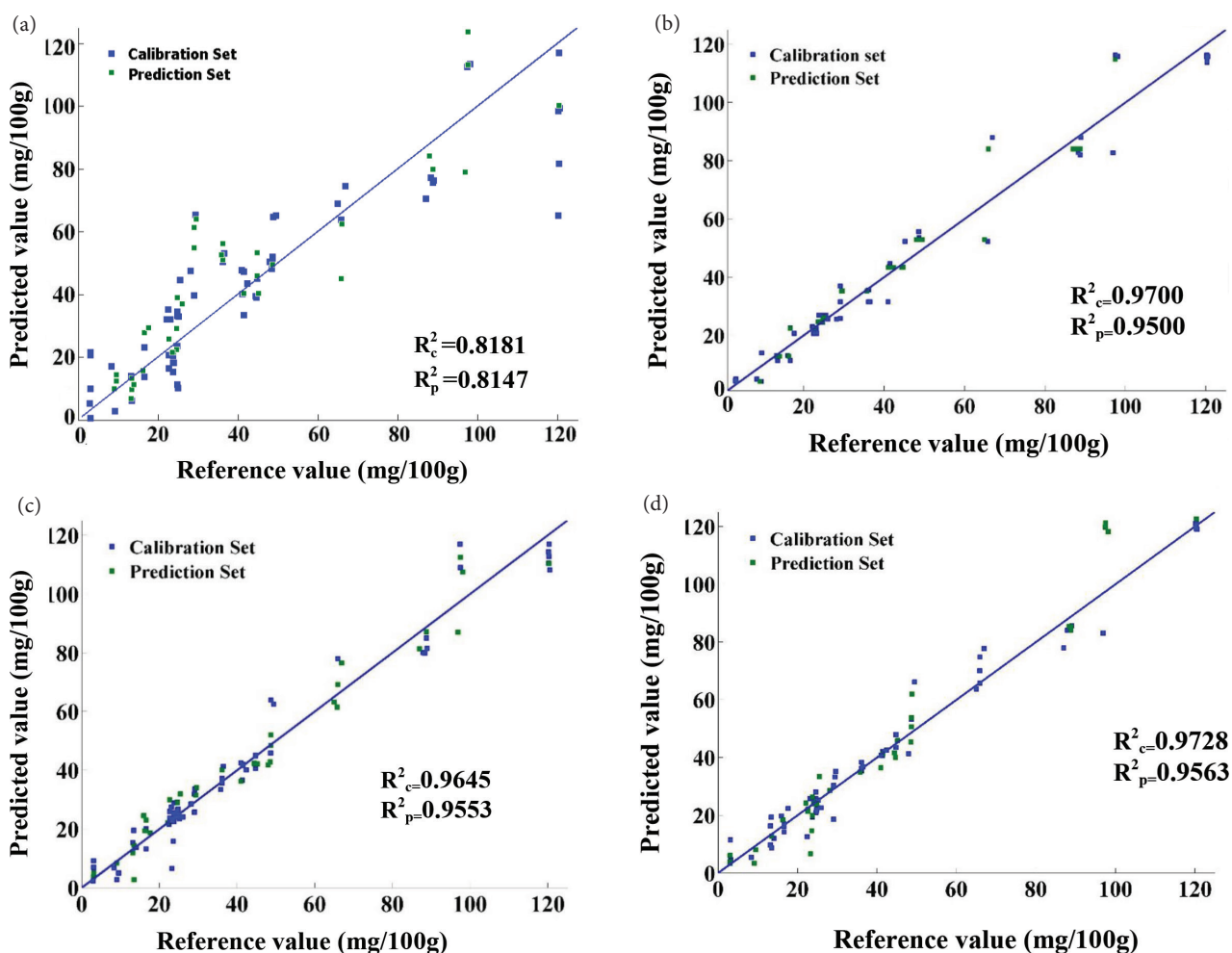
The prediction of *p*-hydroxybenzoic acid content was performed using a PLS model in which *p*-hydroxybenzoic acid was predicted from full NIR spectra (1557 spectra) and selected region of Raman (1142 spectra). Table 3 indicates the parameters used to assess the performance of the established PLS model for prediction of *p*-hydroxybenzoic acid in the analyzed honey samples. Scatter plots of experimental against predicted values for NIR and Raman have been indicated in Figure 2. Although both techniques showed good prediction results, the PLS model using the Raman data as X outperformed NIR, as the first technique obtained higher  $R^2$  and RDP values than when NIR spectra were used as X. However, the use of  $R^2$  and RDP to validate whether the fit model is satisfactory may not be the most suitable method. There are other parameters that could be evaluated. RMSECV, RMSEP, and LVs can provide information about the robustness and the global error of the calibration and prediction models. Raman spectroscopy obtained lower RMSECV and RMSEP and a smaller number of LVs. This was expected as NIR provided a large number of spectra (1557), however, not all of them contained useful information. To improve the results obtained by the NIR technique and to combine the benefit of the synergistic effect of different sources of information, two data fusion strategies were adopted.

Data fusion approaches are now widely used for prediction of phenolic compounds in foods [2,12,36–38]. In this experiment, the PLS model based on data fusion was established. For combining NIR and Raman, the spectra were scaled using simple mean centering and variance scaling before integrating spectra to give alike weights to the data obtained from techniques of different features. Subsequently, the entire preprocessed NIR spectra and the selected

**Table 3** | Descriptive statistics for *p*-hydroxybenzoic acid content used for developing PLSR model

	Calibration set ( <i>N</i> = 64)			Prediction set ( <i>N</i> = 36)		
	Range	Mean	SD	Range	Mean	SD
<i>p</i> -Hydroxybenzoic acid (mg/100 g)	2.99–120.51	41.92	33.06	9.10–120.32	40.12	30.15

SD, standard deviation.

**Figure 2** | Correlation plots for the prediction of *p*-hydroxybenzoic acid using PLS based on the (a) NIR and (b) Raman combined with (c) low-level data fusion and (d) mid-level data fusion.

Raman region merged in a single matrix had the number of columns equal to the number of honey samples and the number of rows equal to the total number of data (selected regions). The low-level data fusion substantiates the fact that it is beneficial to integrate NIR and Raman spectra, as the combined data model performed significantly better than PLS model based exclusively on NIR data (Figure 2). This could be due to the integration of the information from two technologies that could potentially provide comprehensive information and this can improve the results. In fact, it only makes sense to compare the result of low-level data fusion with the PLS models based on NIR spectra as Raman proved to be more adequate than low-level data fusion model and NIR based ones. Therefore, one is interested in investigating the effect of mid-level data fusion on the performance of the PLS model.

Table 4 presents the calibration and prediction results acquired using the two data sets. As already stated, the PLS model was developed using the most significant PC scores. Therefore, the original number of variables for NIR and Raman (2699) matrix was reduced to 20 PCs. As could be seen in Table 4, after applying data extraction and combination, in comparison with PLS model based on NIR, the model constructed for prediction of *p*-hydroxybenzoic acid based on mid-level data fusion delivered better results. Although low- and mid-level data fusion improves the PLS model based on NIR spectra, it does not improve PLS model based on Raman; indicating that PLS model based on the integrated information cannot effectively improve the prediction performance of PLS model based on the Raman spectra, not even after PC extraction, which was in agreement with the literature [39]. Table 4 shows that the result of

**Table 4** | Summary of PLS prediction results for *p*-hydroxybenzoic acid in honey using NIR and Raman spectroscopies

Spectroscopic techniques	LVs	Calibration set ( <i>N</i> = 64)		Prediction set ( <i>N</i> = 36)		RDP
		$R_c^2$	RMSECV	$R_p^2$	RMSEP	
NIR	10	0.8181	14.00	0.8147	13.80	2.18
Raman	3	0.9700	6.11	0.9500	6.83	4.41
Low-level data fusion	10	0.9645	6.08	0.9553	6.59	4.58
Mid-level data fusion	10	0.9728	5.31	0.9563	7.95	3.79

LVs, latent variables; RDP, the ratio of prediction to deviation [SD/RMSEP]; RMSECV, root mean squared error cross validation; RMSEP, root mean squared error of prediction.

mid-level data fusion is closely related to the low-level data fusion in terms of correlation coefficient values. However, a greater scattering of data for the mid-level data fusion was observed. This greater dispersion is in relation to RMSEP value as indicated in Table 4. The determination of coefficient  $R^2$  obtained for the NIR, Raman, and the two data fusion approaches together with the PLS model could be used as alternative quantitative method for determination of *p*-hydroxybenzoic acid in honey samples.

Finally, this present article is the continuation of our previous feasibility study on honey collected from Sudan [2]. We have tried to investigate the feasibility of generalization of similar studies in honey samples collected from different countries. Both studies have demonstrated the feasibility of detecting the single phenolic compound in honey; however, it was also a challenge to find predominant single compounds in honey from the different countries.

## 4. CONCLUSION

This study compared NIR, Raman, and the two data fusion approaches in the quantification determination of *p*-hydroxybenzoic acid in honey and all methodologies could be used for the prediction of single phenolic acid in honey. The RMSEP of low-level data fusion (6.59) was lower than that of mid-level data fusion (7.95) and  $R_p^2$  of low-level data fusion (0.9553). Whereas the mid-level data fusion (0.9563) was significantly higher than the NIR model (0.8147) but no significant effect was observed for Raman (0.9500). For the application of mid-level data fusion, PCA was applied for extracting information. This reduced the number of variables, but it did not improve the results of Raman, demonstrating that the added NIR data is relatively not used. The advantages of the data fusion in the present paper are obvious for solo NIR models because the prediction results of the fused data are higher than 0.95, hence indicate that the data obtained from the two technologies have a synergistic effect. From a practical viewpoint, the potential of determining single phenolic compounds by single or multi-spectroscopic techniques is essential evidence that provides the opportunity to extend the application of these technologies to a larger number of laboratories. In fact, just by low-level data fusion it will be possible to answer the question: “Does data fusion approach improve the prediction of single phenolic compounds in honey?”

## CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

## AUTHORS' CONTRIBUTION

HET and MA performed some experiments. HET wrote the manuscript and obtained funding for the study. GKM helped in experiment design and discussion. JS contributed to discussion of the results. XZ designed and supervised the experiments. MAAG contributed in writing-revision. AAM contributed in data analysis.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support provided by the China Postdoctoral Science Foundation (Grant No. 2017M611736) and National Natural Science Foundation of China (Grant No. 31750110458).

## REFERENCES

- Can Z, Yildiz O, Sahin H, Akyuz Turumtay E, Silici S, Kolayli S. An investigation of Turkish honeys: their physico-chemical properties, antioxidant capacities and phenolic profiles. *Food Chem* 2015;180:133–41.
- Tahir HE, Xiaobo Z, Zhihua L, Jiyong S, Zhai X, Wang S, et al. Rapid prediction of phenolic compounds and antioxidant activity of Sudanese honey using Raman and Fourier transform infrared (FT-IR) spectroscopy. *Food Chem* 2017;226:202–11.
- Daher S, Gülaçar FO. Analysis of phenolic and other aromatic compounds in honeys by solid-phase microextraction followed by gas chromatography-mass spectrometry. *J Agric Food Chem* 2008;56:5775–80.
- Cigić IK, Prosen H. An overview of conventional and emerging analytical methods for the determination of mycotoxins. *Int J Mol Sci* 2009;10:62–115.
- Cozzolino D. Infrared spectroscopy as a versatile analytical tool for the quantitative determination of antioxidants in agricultural products, foods and plants. *Antioxidants (Basel)* 2015;4:482–97.
- Sandak J, Sandak A, Riggio M. Multivariate analysis of multi-sensor data for assessment of timber structures: principles and applications. *Construct Build Mater* 2015;101:1172–80.
- Borràs E, Ferré J, Boqué R, Mestres M, Aceña L, Calvo A, et al. Prediction of olive oil sensory descriptors using instrumental data fusion and partial least squares (PLS) regression. *Talanta* 2016;155:116–23.
- Godinho MS, Blanco MR, Gambarra Neto FF, Lião LM, Sena MM, Tauler R, et al. Evaluation of transformer insulating oil quality using NIR, fluorescence, and NMR spectroscopic data fusion. *Talanta* 2014;129:143–9.
- Casale M, Casolino C, Oliveri P, Forina M. The potential of coupling information using three analytical techniques for identifying the geographical origin of Liguria extra virgin olive oil. *Food Chem* 2010;118:163–70.
- Bagnasco L, Cosulich ME, Speranza G, Medini L, Oliveri P, Lanteri S. Application of a voltammetric electronic tongue and

- near infrared spectroscopy for a rapid umami taste assessment. *Food Chem* 2014;157:421–8.
- [11] Sen I, Ozturk B, Tokatli F, Ozen B. Combination of visible and mid-infrared spectra for the prediction of chemical parameters of wines. *Talanta* 2016;161:130–7.
- [12] Wu Z, Xu E, Long J, Pan X, Xu X, Jin Z, et al. Comparison between ATR-IR, Raman, concatenated ATR-IR and Raman spectroscopy for the determination of total antioxidant capacity and total phenolic content of Chinese rice wine. *Food Chem* 2016;194:671–9.
- [13] Escuredo O, Seijo MC, Salvador J, González-Martín MI. Near infrared spectroscopy for prediction of antioxidant compounds in the honey. *Food Chem* 2013;141:3409–14.
- [14] Tahir HE, Xiaobo Z, Tinting S, Jiyong S, Mariod AA. Near-infrared (NIR) spectroscopy for rapid measurement of antioxidant properties and discrimination of Sudanese honeys from different botanical origin. *Food Anal Methods* 2016;9:2631–41.
- [15] Páscoa RNMJ, Magalhães LM, Lopes JA. FT-NIR spectroscopy as a tool for valorization of spent coffee grounds: application to assessment of antioxidant properties. *Food Res Int* 2013;51:579–86.
- [16] Ouyang Q, Zhao J, Chen Q, Lin H, Sun Z. Rapid measurement of antioxidant activity in dark soy sauce by NIR spectroscopy combined with spectral intervals selection and nonlinear regression tools. *Anal Methods* 2012;4:940–6.
- [17] Tahir HE, Xiaobo Z, Jiyong S, Mariod AA, Wiliam T. Rapid determination of antioxidant compounds and antioxidant activity of Sudanese Karkade (*Hibiscus sabdariffa* L.) using near infrared spectroscopy. *Food Anal Methods* 2016;9:1228–36.
- [18] Condelli N, Caruso MC, Galgano F, Russo D, Milella L, Favati F. Prediction of the antioxidant activity of extra virgin olive oils produced in the Mediterranean area. *Food Chem* 2015;177:233–9.
- [19] Hu Y, Pan ZJ, Liao W, Li J, Gruget P, Kitts DD, et al. Determination of antioxidant capacity and phenolic content of chocolate by attenuated total reflectance-Fourier transformed-infrared spectroscopy. *Food Chem* 2016;202:254–61.
- [20] Schmutzler M, Huck CW. Simultaneous detection of total antioxidant capacity and total soluble solids content by Fourier transform near-infrared (FT-NIR) spectroscopy: a quick and sensitive method for on-site analyses of apples. *Food Control* 2016;66:27–37.
- [21] Revilla I, Vivar-Quintana AM, González-Martín I, Escuredo O, Seijo C. The potential of near infrared spectroscopy for determining the phenolic, antioxidant, color and bactericide characteristics of raw propolis. *Microchem J* 2017;134:211–17.
- [22] Fernández-Martínez J, Joffre R, Zacchini M, Fernández-Marín B, García-Plazaola JL, Fleck I. Near-infrared reflectance spectroscopy allows rapid and simultaneous evaluation of chloroplast pigments and antioxidants, carbon isotope discrimination and nitrogen content in *Populus* spp. leaves. *Forest Ecol Manage* 2017;399:227–34.
- [23] Fernández-Pierna JA, Abbas O, Dardenne P, Baeten V. Discrimination of Corsican honey by FT-Raman spectroscopy and chemometrics. *Biotechnol Agron Soc Environ* 2011;15:75–84.
- [24] Özbacı B, Boyacı İH, Topcu A, Kadılar C, Tamer U. Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks. *Food Chem* 2013;136:1444–52.
- [25] Anjos O, Campos MG, Ruiz PC, Antunes P. Application of FTIR-ATR spectroscopy to the quantification of sugar in honey. *Food Chem* 2015;169:218–23.
- [26] Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab Syst* 2001;58:109–30.
- [27] Kamruzzaman M, ElMasry G, Sun DW, Allen P. Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging. *Food Chem* 2013;141:389–96.
- [28] He HJ, Sun DW, Wu D. Rapid and real-time prediction of lactic acid bacteria (LAB) in farmed salmon flesh using near-infrared (NIR) hyperspectral imaging combined with chemometric analysis. *Food Res Int* 2014;62:476–83.
- [29] Borràs E, Ferré J, Boqué R, Mestres M, Aceña L, Busto O. Data fusion methodologies for food and beverage authentication and quality assessment – a review. *Anal Chim Acta* 2015;891:1–14.
- [30] Yao L, Datta N, Tomás-Barberán FA, Ferreres F, Martos I, Singanusong R. Flavonoids, phenolic acids and abscisic acid in Australian and New Zealand *Leptospermum* honeys. *Food Chem* 2003;81:159–68.
- [31] Weston RJ, Mitchell KR, Allen KL. Antibacterial phenolic components of New Zealand manuka honey. *Food Chem* 1999;64:295–301.
- [32] Weston RJ, Brocklebank LK, Lu Y. Identification and quantitative levels of antibacterial components of some New Zealand honeys. *Food Chem* 2000;70:427–35.
- [33] Chan CW, Deadman BJ, Manley-Harris M, Wilkins AL, Alber DG, Harry E. Analysis of the flavonoid component of bioactive New Zealand mānuka (*Leptospermum scoparium*) honey and the isolation, characterisation and synthesis of an unusual pyrrole. *Food Chem* 2013;141:1772–81.
- [34] Makawi SZA, Gadkariem EA, Ayoub SMH. Determination of antioxidant flavonoids in Sudanese honey samples by solid phase extraction and high performance liquid chromatography. *J Chem* 2009;6:S429–37.
- [35] Tahir HE, Xiaobo Z, Zhihua L, Yaodi Z. Comprehensive evaluation of antioxidant properties and volatile compounds of Sudanese honeys. *J Food Biochem* 2015;39:349–59.
- [36] Arslan M, Xiaobo Z, Tahir HE, Zareef M, Xuetao H, Rakha A. Total polyphenol quantitation using integrated NIR and MIR spectroscopy: a case study of Chinese dates (*Ziziphus jujuba*). *Phytochem Anal* 2019;30:357–63.
- [37] Huang X, Teye E, Sam-Amoah LK, Han F, Yao L, Tchabo W. Rapid measurement of total polyphenols content in cocoa beans by data fusion of NIR spectroscopy and electronic tongue. *Anal Methods* 2014;6:5008–15.
- [38] Rodríguez-Mendez ML, Apetrei C, Gay M, Medina-Plaza C, de Saja JA, Vidal S, et al. Evaluation of oxygen exposure levels and polyphenolic content of red wines using an electronic panel formed by an electronic nose and an electronic tongue. *Food Chem* 2014;155:91–7.
- [39] de Groot PJ, Swierenga H, Postma GJ, Melssen WJ, Buydens LMC. Effect on the partial least-squares prediction of yarn properties combining Raman and infrared measurements and applying wavelength selection. *Appl Spectrosc* 2003;57:642–8.

## SUPPLEMENTARY TABLE

Table S1 | The average of phenolic compounds identified in various honey samples

Codes	Phenolic compounds quantified (mg/100 g)														Total phenol		
	CAT	Catec	HydroxA	p-HydroxA	Rut	p-COUA	FA	Myr	Chry	QE	Kae	SA	VA	ChloA		GA	CA
H1	60.71	38.21	15.59	24.85	18.96	64.08	25.35	25.17	0.81	27.46	5.86	n.d	n.d	n.d	n.d	n.d	307.04 <sup>d</sup>
H2	n.d	n.d	n.d	3.05	n.d	n.d	n.d	n.d	n.d	n.d	n.d	52.99	11.57	n.d	n.d	n.d	67.61 <sup>n</sup>
H3	n.d	37.92	n.d	120.31	20.04	29.48	n.d	n.d	n.d	n.d	2.71	85.01	n.d	n.d	0.54	n.d	296.02 <sup>e</sup>
H4	161.81	n.d	n.d	15.03	38.57	n.d	n.d	n.d	n.d	n.d	n.d	n.d	31.52	12.61	n.d	2,7018	262.23 <sup>g</sup>
H5	119.58	n.d	n.d	44.76	n.d	8.43	0.78	n.d	n.d	n.d	n.d	n.d	44.22	n.d	n.d	n.d	217.78 <sup>h</sup>
H6	4.25	n.d	n.d	9.10	n.d	n.d	0.81	n.d	n.d	n.d	n.d	n.d	23.57	n.d	n.d	n.d	37.72 <sup>o</sup>
H7	98.13	n.d	n.d	65.91	37.51	2.69	172.96	n.d	n.d	n.d	n.d	n.d	n.d	22.82	n.d	n.d	400.02 <sup>c</sup>
H8	n.d	78.79	n.d	88.33	41.10	45.37	75.32	n.d	n.d	n.d	n.d	24.75	9.88	57.30	n.d	10.97	431.79 <sup>b</sup>
H9	n.d	n.d	n.d	41.45	39.86	29.24	2.75	n.d	n.d	n.d	n.d	n.d	2.94	61.55	n.d	n.d	177.78 <sup>r</sup>
H10	25.24	78.29	n.d	48.67	n.d	4.15	75.67	n.d	68.33	n.d	n.d	66.51	70.19	17.56	n.d	n.d	454.62 <sup>a</sup>
H11	7.41	76.42	11.60	23.76	n.d	n.d	n.d	n.d	n.d	n.d	n.d	1.28	n.d	n.d	n.d	n.d	120.46 <sup>f</sup>
H12	n.d	76.66	n.d	36.13	n.d	n.d	154.33	n.d	n.d	n.d	n.d	n.d	0.29	n.d	n.d	n.d	267.41 <sup>f</sup>
H13	n.d	n.d	n.d	97.46	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	n.d	11.46	n.d	n.d	108.92 <sup>m</sup>
H14	118.74	38.39	n.d	29.08	19.40	n.d	n.d	n.d	n.d	15.20	0.49	n.d	n.d	n.d	0.59	n.d	221.90 <sup>b</sup>
H15	n.d	n.d	n.d	22.64	n.d	24.43	n.d	n.d	n.d	n.d	n.d	110.06	n.d	n.d	n.d	n.d	157.12 <sup>k</sup>

n.d, not detected; CAT, catechin; Catec, catechol; HydroxA, hydroxycinnamic acid; p-HydroxA, p-hydroxybenzoic acid; Rut, rutin; p-COUA, p-coumaric acid; FA, ferulic acid; Myr, myricetin; Chry, chrysin; QE, quercetin; Kae, kaempferol; SA, syringic acid; VA, vanillic acid; ChloA, chlorogenic acid; GA, gallic acid; CA, caffeic acid.