LWT - Food Science and Technology 154 (2022) 112760

Contents lists available at ScienceDirect

LWT

journal homepage: www.elsevier.com/locate/lwt

Authentication of cinnamon spice samples using FT-IR spectroscopy and chemometric classification

Panagiota Lixourgioti [a], Kirstie A. Goggin [b], Xinyu Zhao [a], Denis J. Murphy [b], Saskia van Ruth [c], Anastasios Koidis [a,*]

[a] Institute for Global Food Security, Queen's University, Belfast, BT9 5DJ, United Kingdom
[b] School of Applied Sciences, University of South Wales, Pontypridd, CF37 1DL, United Kingdom
[c] Food Quality and Design Group, Wageningen University and Research, P.O. Box 17, 6700 AA, Wageningen, the Netherlands

ARTICLE INFO

Keywords:
Cinnamon
Spice
Infrared spectroscopy
Food authentication
Multivariate classification

ABSTRACT

Cinnamon is a popular spice with a lengthy overseas supply chain. *C. cassia* is commonly traded as cinnamon, but the use of rapid methods to detect its adulteration has not yet been fully addressed. This work explores the use of FT-IR spectroscopy for the detection of adulteration in the cinnamon supply chain by several lower value ingredients. Two species of cinnamon (*C. verum* and *C. cassia*) and an adulterant (cinnamon spend, n = 2) were used to create 110 different in-house admixtures. Two different replacement fraud experiments were designed: *C. cassia* replaced with spend (Scenario A) and C. *verum* replaced with both *C. cassia* and spend (Scenario B). Initial analysis by GC-IMS showed promising differences between samples. The FT-IR spectra confirmed significant raw differences in absorbance. PCA for Scenario A demonstrated better separation than in Scenario B. The detection of adulteration of *C. cassia* (Scenario A) and *C. verum* (Scenario B) were equality accurate. Classification results showed that the PLS-DA technique was superior to SIMCA for both types of adulteration (PLS-DA: 94-90%; SIMCA: 83-79%, respectively). This demonstrates the potential of FT-IR as a screening method to identify cinnamon adulteration in supply chains and to provide accurate and rapid results without sample preparation.

1. Introduction

Cinnamon (*Cinnamomum* spp.) is one of the most widely used spices due to its attractive flavour and medicinal value (Rahadian et al., 2017). The most common types of cinnamon are *Cinnamomum verum* (also known as *C. verum* or "true" cinnamon) and *Cinnamomum cassia. C. verum* is mostly cultivated in Sri Lanka whereas *C. cassia*, (C. *cassia* bark or Chinese cinnamon bark) is cultivated in China and across Southeast Asia (Thomas & Kuruvilla, 2012, pp. 182–196). The type of cinnamon spice varies according to its geographic origin, soil type, climate, and cultivation methods, all of which can modulate its chemical composition (Avula et al., 2014).

The appearance of the bark varies for individual cinnamon quills. The outer layer of C. cassia barks is thicker and harder, and their quills curl inward from both sides. In contrast, the outer layer of C. verum, is thinner and softer with quills curled in a telescopic arrangement. Quality control involves evaluating the size, the odour, the thickness of the quill and the location of the bark (e.g., inner vs outer) (He et al., 2005).

Many industries avoid disclosing the geographic provenance of the cinnamon spice, especially when it is a mixture of many types or origins (Woehrlin et al., 2010). Due to consumers' increased awareness of the role that food plays in the diet, its safety and quality is now subject to increasing scrutiny. The increase in reported food fraud incidents published in the media has underpinned requests for stricter quality control regulations (Bansal et al., 2018). Economically motived adulteration such as the mixing high-quality with low-quality cinnamon has been regularly occurring in the food supply chain (Swetha et al., 2014). Therefore, quality control and early detection of potential adulterants in cinnamon, is of great significance for stakeholders of the supply chain and consumers alike. Several high-end analytical methods based on determination of marker compounds have been developed.

Cinnamaldehyde is the main active ingredient of the essential oil that gives cinnamon its unique combination of sweet and spicy flavours. Cinnamon also contains about 2% of other volatile oils, such as eugenol and linalool, as well as various polysaccharides, and polyphenols (Yeh et al., 2014). Current targeted detection methods are based on liquid

* Corresponding author.Institute for Global Food Security, School of Biological Sciences. Queen's University Belfast, 19 Chlorine Gardens, Belfast, BT9 5DJ. UK.
E-mail address: t.koidis@qub.ac.uk (A. Koidis).

https://doi.org/10.1016/j.lwt.2021.112760
Received 31 May 2021; Received in revised form 29 October 2021; Accepted 1 November 2021
Available online 9 November 2021

(Woehrlin et al., 2010) and gas chromatography (Li et al., 2013), and direct analysis real time mass spectrometry (Avula et al., 2014). DNA barcoding has also been used to detect adulteration of cinnamon and identify high-quality species for potential breeding purposes (Swetha et al., 2014). Although these methods have proven accurate and fit-for-purpose in the lab, their high cost and time of analysis have prevented them from wide adoption for routine deployment in supply chains, by both industry and regulators. Unfortunately, this means that no actual tests are being carried out in the supply chain and, intentional or unintentional, fraud still takes place.

For authentication of herbs and spices, especially in the ground form, fingerprinting approaches have proven to be effective providing fast and reliable results (Galvin-King et al., 2018). Many vibrational spectroscopy methods have been used to examine the quality of the product and provide reliable food fingerprinting (Ellis et al., 2012; Haughey et al., 2015; Lim et al., 2021). Whilst much knowledge has been accumulated about other spices such as oregano (Black et al., 2016), only a limited range of techniques have so far been developed to accurately screen for cinnamon adulteration.

In terms of vibrational spectroscopy methods, Fourier-transform Infrared (FT-IR) spectroscopy in the mid IR range, 450 - 4000 cm^{-1} provides the most value as it offers detailed structural information about the molecular makeup of the samples. Li et al. (2013) focused on cinnamon essential oil and used FT-IR to obtain fingerprints to identify different cinnamon species and determine origin of pure samples. In the only study of its kind, Yasmin et al. (2019) used FT-NIR and FT-IR methods to detect adulteration of cinnamon powder. However, although a large dataset was used, the entire range of cinnamon adulteration was not addressed in their study. In addition, the multivariate analysis method chosen shows that the emphasis was given on precise quantification of a specific type of fraud (i.e., *C. cassia* in *C. verum*) rather than on the classification of samples where origin, species and admixtures ratios can vary. To date, there have been no studies that have factored in adulteration by the addition of cinnamon spend, the by-product of cinnamon essential oil extraction, as a bulking agent in ground *C. cassia*, or studies looking at both volatile and non-volatile components at the same time.

The aim of this study was to investigate the use of FT-IR coupled with appropriate multivariate analysis to develop a universal chemometric model to enable rapid authenticity of cinnamon from a variety of different adulteration types and sources. The experimental design calls for two different adulteration scenarios, one focused on common ground cinnamon (Scenario A) and one on premium ground cinnamon (Scenario B). In both cases, cinnamon spend was included as adulterant to make the study more robust. To confirm and characterise the changes between different volatile profiles of the cinnamon samples in this study, Gas Chromatography - Ion Mobility Spectroscopy (GC-IMS) was employed.

2. Materials and methods (1200)

2.1. Sample preparation and admixture design

A total of 25 samples were used (*C. cassia, n=15; C. verum*, n = 10 and of cinnamon essential oil spend, n = 2). *C. cassia* samples were from Indonesia (n = 4), China (n = 4), India (n = 3), Sri Lanka (n = 2), Vietnam (n = 1) and Seychelles (n = 1). All *C. verum* samples were from Sri Lanka except one that was from Seychelles. The two cinnamon spend samples were from Sri-Lanka. The samples were sourced from trusted sources of the supply chain. All cinnamon samples arrived in the form of bark or quills (or wood husk pieces for the case of the spend, directly shipped from Sri Lanka) and were processed immediately after reception. Initially, all samples were partially grinded in small batches (50 g) using a typical blade grinder for 1 min at moderate speed (2000 rpm), followed by milling in a laboratory ball mill grinder (Retch PM100, Fisher Scientific) at 500 rpm for 5 min to achieve a fine and uniform grinding size (<10 μm). The cinnamon particles were passed from a 1000, 100 and a 10 μm micron sleeve to ensure uniformity. The time between each milling cycle was kept at 5 min to avoid overheating. The temperature was checked with a probe and never exceeded 28 °C. In-house admixtures were prepared in a weigh/weight basis. Two experiments were devised to simulate two different adulteration scenarios of cinnamon fraud in the supply chain. Scenario A was designed to detect *C. cassia* adulteration while Scenario B was designed to detect *C. verum* adulteration. More specifically:

Scenario A: *C. cassia* samples admixed at levels of 1 g/100 g–99 g/100 g (or 1–99%), with the cinnamon spend as the adulterant.

Scenario B: *C. verum* samples admixed at spend 1 g/100 g–99 g/100 g (or 1–99%), with both *C. cassia* and spend as the adulterants.

A complete range of cinnamon admixtures was used (1–99%) to cover all likely adulteration ratios. Different cinnamon samples were randomly assigned to different admixtures and the samples were cycled to maximise the possibilities of mixing different samples and therefore maximise variation in the data. The sample set was randomly divided into two groups: calibration (70% of the samples for each class) and validation (30% of the sample). For each experiment (Scenario A and B), a separate calibration and validation dataset were used. These two sets remained completely independent, i.e.the pure cinnamon samples present in the calibration set were not used in any of the admixtures of the validation set, and vice versa, to ensure no bias within the model. The experimental design is illustrated in more detail in Fig. 1. In total, 88 admixture powder samples were prepared and together with the pure samples were stored under vacuum at −20 °C until analysis.

2.2. Gas chromatography Ion Mobility Spectroscopy (GC-IMS)

GC-IMS analysis was performed using a FlavourSpec instrument (G. A.S., Dortmund Germany). Briefly, 0.500 g of each sample type (*C. cassia* inner bark, *C. cassia* outer bark, *C. verum*, cinnamon spend) was weighed out and placed into 20 mL glass headspace vials then secured with a magnetic screw cap and sealed with a PTFE-silicone septum. To note, initially such distinction between outer and inner *C. cassia* bark was attempted, but as seen later in the text, it was decided to merge them into one *C. cassia* class. Each sample was prepared in triplicate. In addition, three admixtures of 50:50 (*C. cassia: C. verum; C. cassia; C.* spend; *C. verum: C.* spend) were prepared. Samples were analysed on an optimised FlavourSpec programme with the following variables: IMS temperature: 45 °C, GC column temperature: 60 °C, IMS flow rate set at 150 mL/min. The GC flow rate was ramped: linear increase of 0–30 mL/min within 10 min and then 30–150 L/min for a further 5 min and then kept on a stable flow rate to ensure heavy volatile organic compounds (VOCs) elute from the column by the end of analysis (t = 23 min). The GC capillary column used was an FS-SE54, 15 m, 0.53 mm ID, 1 μm particle size (Supelco, Dortmund, Germany, and the carrier gas used was nitrogen.

2.3. Spectral data acquisition using fourier-transform infrared (FT-IR)

FT-IR spectra were collected on a Thermo Nicolet iS5 spectrometer (Thermo Scientific, Dublin, Ireland) equipped with ATR iD5 diamond crystal and ZnSe lens and DTGS KBr detector. Approximately 200 mg of pure, finely grinded sample was placed in the ATR diamond crystal sampling plate of the instrument. Each spectrum was acquired in the 4000 to 600 cm^{-1} range at room temperature using standard parameters (4 cm^{-1}, 32 number of scans) resulting in 1.928 cm^{-1} data spacing (Osorio et al., 2015).

2.4. Spectral region selection, pre-processing, and multivariate modelling

2.4.1. FT-IR

Standard normal variate (SNV), first order derivative, Savitsky-Golay smoothing (15 points in each sub-model) and Pareto scale pre-processing algorithms were included in the FT-IR spectra pre-

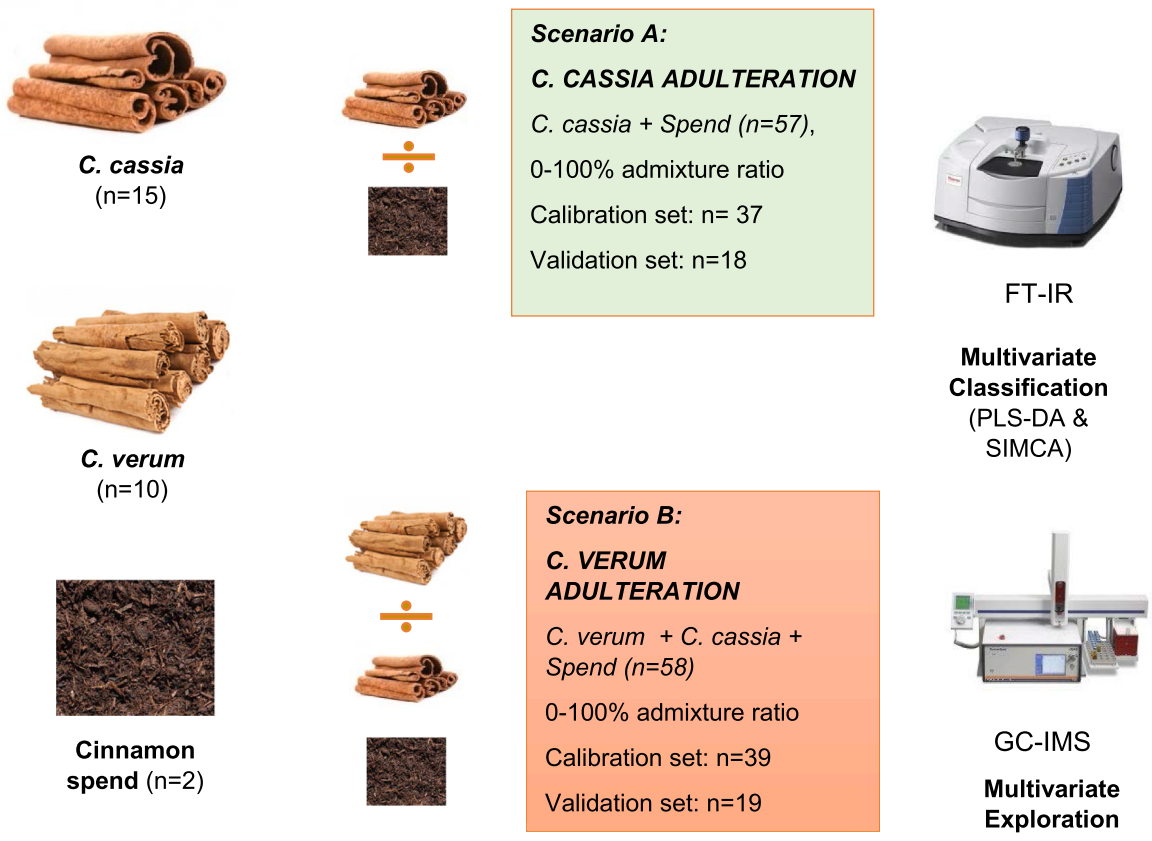

Fig. 1. The experimental design of the study.

processing. Three different FT-IR spectral regions were analysed, region 1 (3999.6–549.6 cm^{-1}) which included the entire spectrum, and two selective regions, region 2 (1801.2–601.7 cm^{-1}) and region 3 (3000.1–2800.1 cm^{-1} combined with 1801.2–601.7 cm^{-1}). The FT-IR spectral characteristics were using Principal Component Analysis (PCA). The prediction of the models was evaluated testing the validation set against calibrated models in different wavelength regions (Region 1: 1790 variables. Region 2: 623 variables, Region 3: 728 variables). Calibration of specific model classes was performed using two independent supervising classification techniques, Partial Least Square Discriminant Analysis (PLS-DA) and Soft Independent Model Class Analogy (SIMCA) using Umetrics SIMCA 15 (Umea, Sweden). Cross validation was also carried out in the calibration set in Umetrics SIMCA using the venetian blinds method with a 1/7th window (Osorio *et al.*, 2015). Based on the confusion matrix generated, recall, precision, and F1 values were calculated as performance metrics for model evaluation. Recall, also called 'sensitivity', is a measure of the ability of a classifier to predict samples of a certain class (target samples) from dataset and corresponds to the true positive rate. Precision or 'specificity' is a measure of prediction relevancy, or the ability to reject non target samples. F1 score is interpreted as a weighted average of precision and recall calculated by the following formula: F1 = 2* Precision*Recall/(Precision + Recall). Overall classification rate also considers both Recall and Precision and expressed as: [(correct classified target samples + correctly rejected target samples)/total class samples]. Model accuracy describes the overall correctness of the model and is calculated as the sum of correct samples divided by the total number of samples across all classes.

2.4.2. GC-IMS

GC-IMS signals with retention time from 60 to 1008 s and drift time from 7 to 15 ms (1201 × 4515 matrix) were processed. Origin software (Northampton, MA, USA) was used to create the three-dimensional GC-IMS spectral topographic map. PCA was used to show natural clustering of samples based on maximum peak intensity (peak-picking PCA). To conduct the multiway principal component analysis (M-PCA), R programming language powered with the RGCxGC toolbox (Quiroz-Moreno et al., 2020) was used. This contains smoothing, peak aliment and baseline correction to pre-process the signal followed by the M-PCA algorithm.

3. Results and discussion

3.1. Volatile organic compounds in cinnamon

Ion mobility spectroscopy (IMS) is an analytical technique for characterising chemical substances, and in projection, different samples, based on the difference in the migration speed of different gas phase ions in an electric field (Borsdorf et al., 2011). Here IMS is used to establish initial differences between the different cinnamon types of the study. Results of the GC-IMS VOC dataset of cinnamon samples show that is both multi-dimensional and highly complex. Differences are more readily visible with the three-dimensional spectral topographic map for the different types of pure cinnamon samples (Fig. 2) and for the 50:50 binary admixtures (Fig. 3). In these maps, the X axis is the drift time, the Y axis is the retention time, and the Z axis is the ion signal intensity. Each data point represents volatile compounds, and the intensity information is colour graded. The line parallel to the y-axis with a ratio of 1.0 on the X-axis represents the reactive ion peak. As indicated with initial olfactory inspection (GC-IMS), *C. verum* contained a greater number of VOCs than the other three types (Fig. 2A). The types of VOCs contained in *C. cassia* outer bark (Fig. 2B) and *C. cassia* inner bark (Fig. 2C) were very similar. As expected, cinnamon spent, the waste left after vapour-extracting cinnamon essential oil from outer/inner barks (Fig. 2D) contains the least VOCs. Comparing that with the original spectrum (a1), only *C. verum contains* some unique VOCs (A vs a.1). In

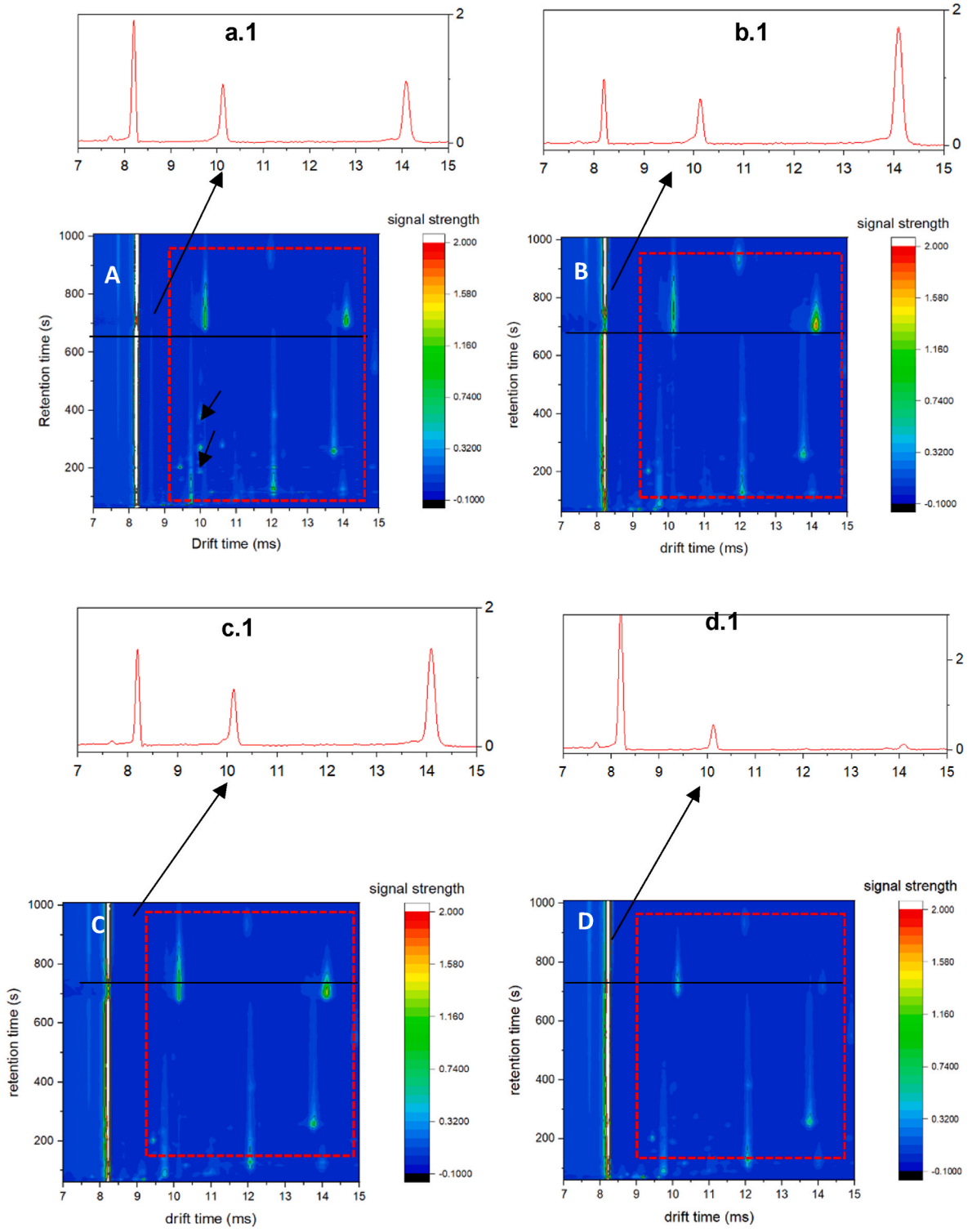

Fig. 2. Topographic plot of GC-IMS spectrum of cinnamon pure samples. (A) *C. verum sample*, (B) *C. cassia* outer bark sample, (C) *C. cassia* inner bark sample, (D) cinnamon Spent sample. Spectra at 711.6 s (a.1, b.1, c.1, and d.1).

the topographic map of in-house admixtures (Fig. 3), although there are certain similarities in the types of VOCs present, their concentration in the cinnamon admixture can vary to a point where further observations are difficult.

Therefore, two different types of PCA were employed to observe the VOC differences between different types of cinnamon. In the more traditional approach, 23 characteristic peaks were selected to characterise the unique quality information of different cinnamon samples at their maximum peak intensity value (peak picking PCA). Good clustering was achieved with PC1 and PC2 accounting for 65% of variation (Fig. 4A). The PCA showed good distinction between *C. verum* and cinnamon spend, also from *C. cassia* (inner and outer). Inner and outer *C. cassia* signals were very similar to each other and hard to distinguish in accordance with previous findings. In this case, however, the in-house admixtures were projected well within the orthogonal space of the pure samples in coordinates that indicate their origin (e.g., the 50:50

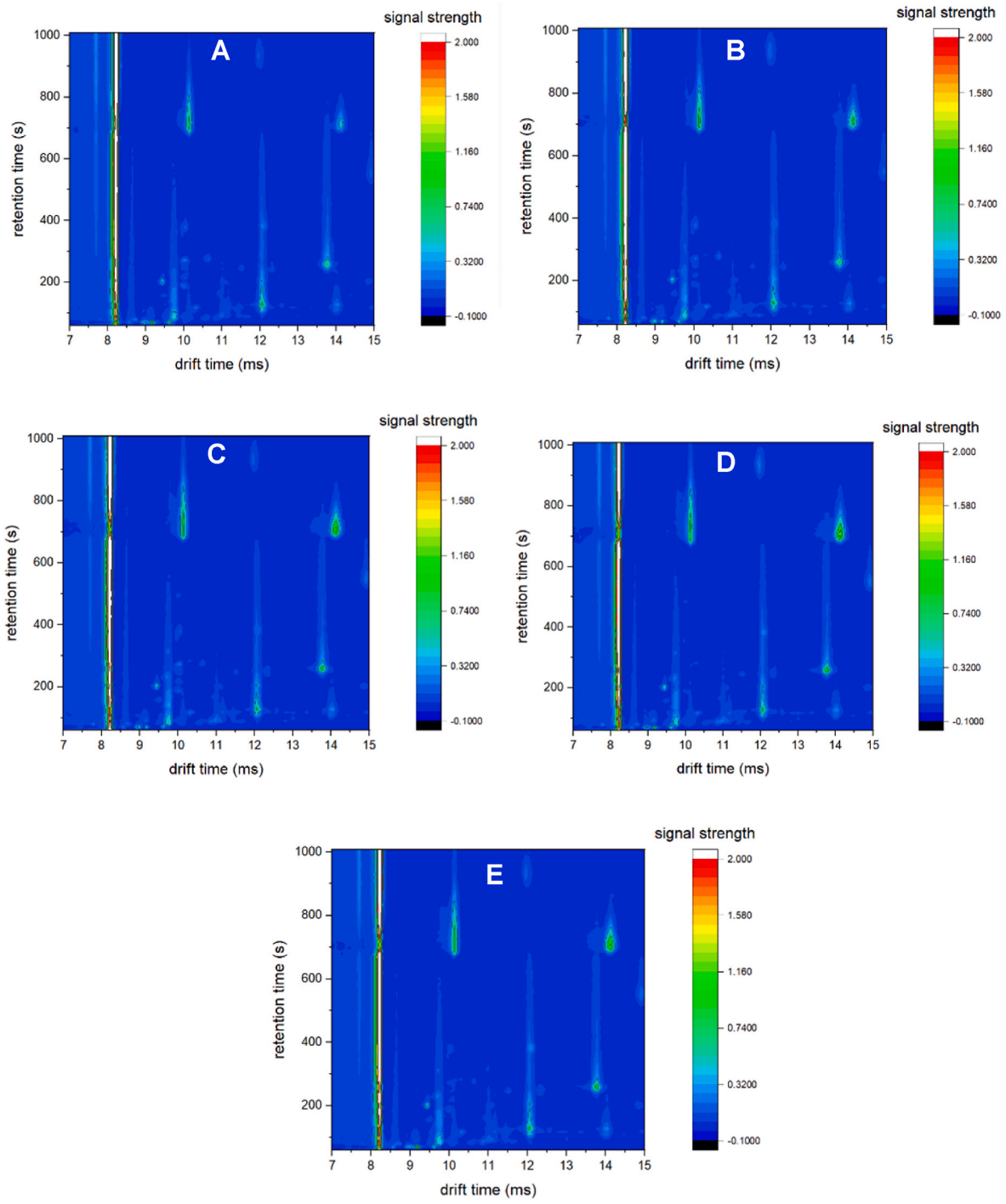

Fig. 3. Topographic plot of GC-IMS spectra for 50:50 binary admixture cinnamon samples. (A) "*C. verum*: *C. spent*", (B) "C. verum: *C. cassia* outer bark", (C) "*C. verum*: C. *cassia* inner bark", (D) "*C. cassia* outer bark: *C. cassia* inner bark", and (E) "*C. cassia* inner bark: cinnamon spend".

C. verum: cinnamon spend admixture is projected in the middle between the two pure classes of *C. verum* and cinnamon spend). To achieve an even higher variation, M-PCA was followed. Compared with traditional PCA, MPCA is based on a 2-D matrix instead of 1-D vector method (Abdi & Williams, 2010; Garrido-Delgado, delMarDobao-Prieto, Arce, & Valcárcel, 2015). As shown in Fig. 4B, the first two principal components explain 86% of the variation indicating that the covariance matrix is more accurate, although the separation was not visibly improved from the peak picking PCA (Fig. 4A). This experiment also showed that there is not meaningful distinction between cassia outer bark and inner bark, and the samples were combined for the rest of the experimental procedure (FT-IR).

For reference, a preliminary analysis was conducted to identify the main cinnamon VOCs based on normalisation to a standard ketone mix. To demonstrate the robustness of the system, repeatability and reproducibility were calculated using the intensity of the hexanal dimer in the standard ketone mixture. Repeatability (intra-day variation) was determined by analysing the same sample ten times in one day. Reproducibility (inter-day variation) was determined by measuring samples once a day over a period of three days. Values were calculated as Relative Standard Deviations (RSD). Repeatability was calculated as 1.62%, whilst reproducibility was calculated as 2.45%. GC-IMS analysis

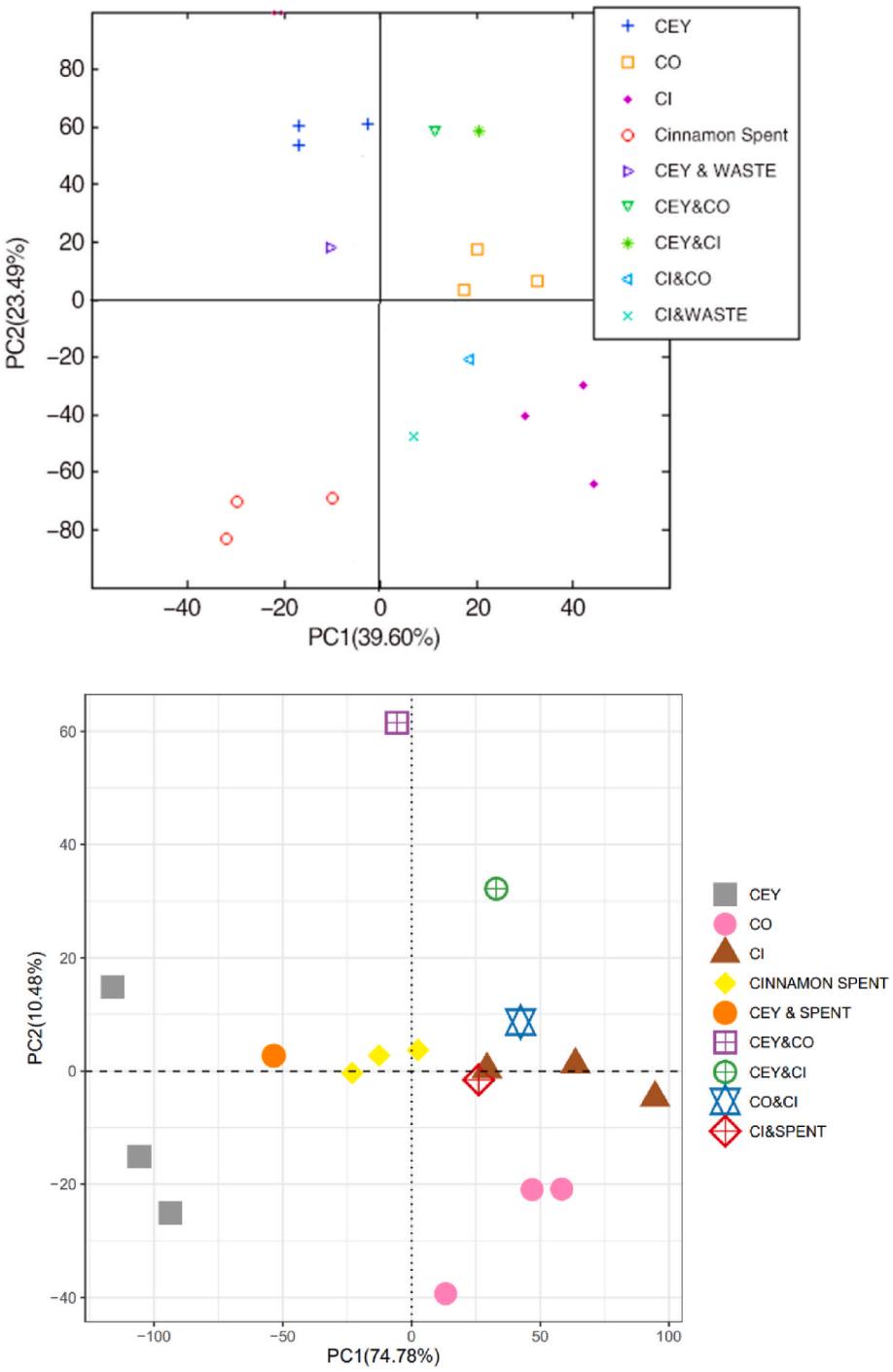

Fig. 4. PCA score plot based on cinnamon samples analysed by GC-IMS a) standard PCA based on 23 characteristic VOCs maximum intensity peaks b) MPCA (CEY: *C. verum* cinnamon; CO: *C. cassia* outer bark cinnamon; CI: *C. cassia* inner bark cinnamon; WASTE: cinnamon spent; CEY & CO: admixture of *C. verum* and *C. cassia* outer bark; CEY& WASTE, CEY&CI, CI&CO, CI&WASTE: admixtures with the above mentioned ingredients).

identified the following potential volatiles in the VOC profiles: linalool, n-nonanal, benzaldehyde, heptanal, hexanal, 1-butanol, and pentanal (see Supplementary material). These appear commonly in all the samples, and it is the first time reported in the literature. More work is needed however, to confirm these findings.

3.2. Characterisation of the FT-IR spectra of the different cinnamon samples

Fig. 5 shows the averaged FT-IR spectra of pure *C. cassia*, *C. verum* and cinnamon spend samples. The peaks correspond to bending and stretching vibrations between the molecules of the samples. Differences between the samples are mainly observed in the fingerprint region between the range wavenumbers 1800 - 600 cm^{-1} and in the hydroxyl/C–H region at 3500-2800 cm^{-1}. There was a high signal to noise ratio observed within the spectral region 2500 to 1800 cm^{-1} indicating that this region can be excluded in the subsequent analysis in line with the literature (Rodriguez *et al.*, 2019).

The peak at 1727 cm^{-1} represents the carbonyl bond and it is attributed mainly to the aldehydes of saturated fatty acid signatures in

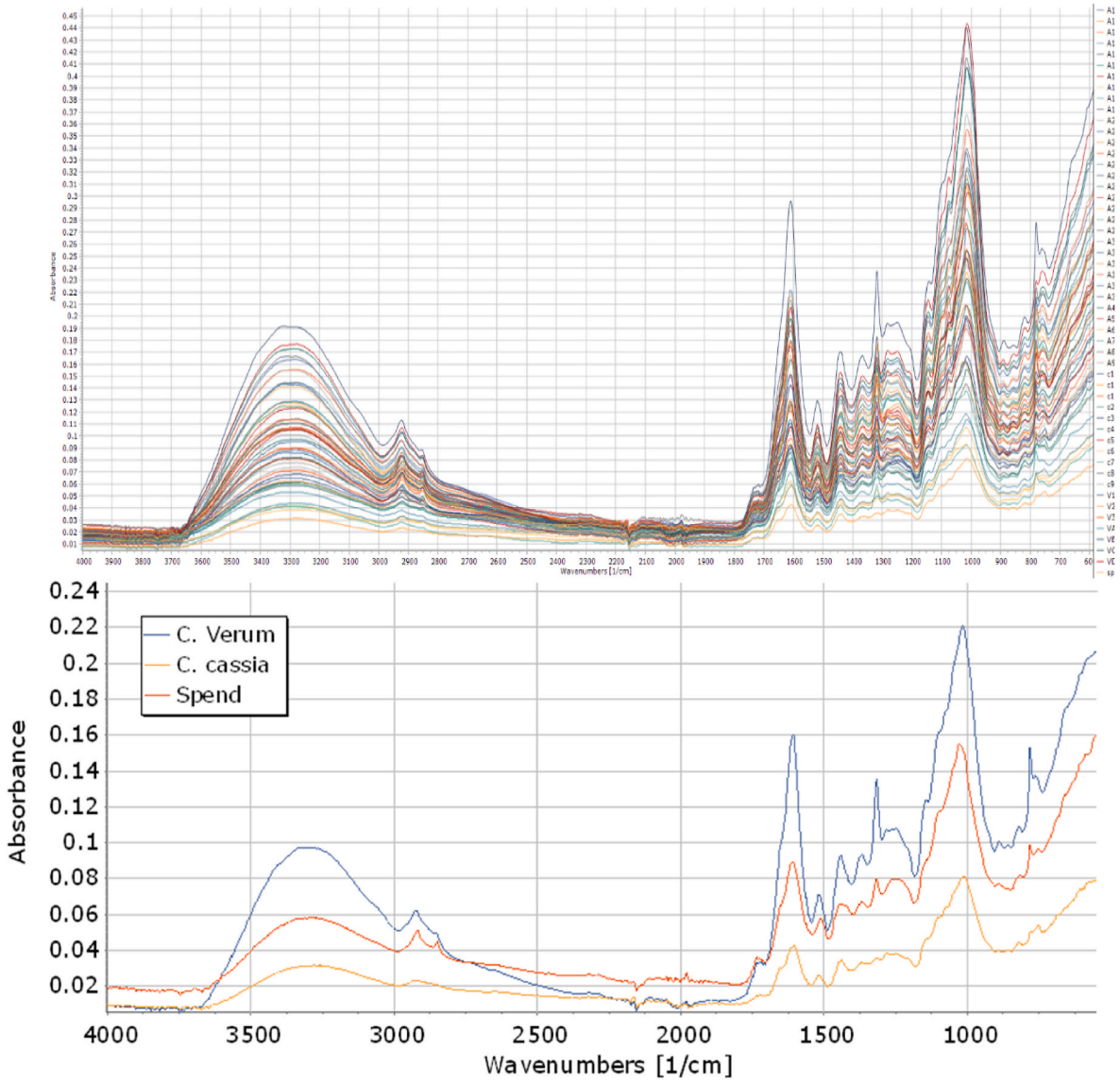

Fig. 5. FT-IR spectra of pure C. cassia (yellow line), C. verum (blue line), cinnamon spend (orange line) and their admixtures: (A) all samples, (B) peak characterisation of three sample types (averaged spectra). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the sample. The peaks at 1679 and 1626 cm^{-1} are attributed to the stretching vibration of an aldehyde carbonyl C═O, which corresponds to aldehydes such as coumarin (Lin *et al.*, 2017). Li *et al.* (2013) associated the appearance of these peaks with cinnamaldehyde and other aldehydes founded in high levels in cinnamon barks. The peak at 1573 cm^{-1} corresponds to the skeleton vibration C═C of an aromatic ring, normally associated with eugenol, a phenol present in cinnamon oil. Previous studies showed that eugenol is found in a greater amount in *C. verum* and coumarin is a component of *C. cassia* VOCs (Li et al., 2013). The peak at 1450 cm^{-1} is described by the bending vibration absorption of an alcohol C–OH. The peak at 1294 cm^{-1} corresponds to the bending absorption of the aromatic ring ═C–H and $–CH_2$ alkanes. The peak at 1248 cm^{-1} is due to the symmetric expansion of the aromatic acid ester C–O–C and the >C–OH stretching vibration of phenols. These peaks collectively represent characteristics eugenol and of the esters in cinnamon barks (Li *et al.*, 2013; Lin *et al.*, 2017). On the other hand, in the region 3500-2800 cm^{-1} (C–H and O–H region) the peak in 2920 cm^{-1} can be assigned to symmetric C–H stretching vibrations that are caused by cinnamon lipids (essential oil) and the peaks at 3300 cm^{-1} are O–H stretching vibration absorption peaks. These are due to the presence of various alcohols including eugenol, with both carbonyl and hydroxy function groups and small quantities of proteins present in the sample. Minute contents of water (moisture) could be naturally present in the samples and affect this area of the signal, despite the use of desiccators to handle samples during the analysis to provide a zero or at least a low and stable moisture content (Lohumi *et al.*, 2014).

Overall, the spectral fingerprints appeared similar for *C. cassia* and *C. verum* spectra, containing 15–16 peaks each. The spectra of the cinnamon spend had a reduced number of peaks compared to *C. verum* and *C. cassia* (Fig. 2A). The spectra of the in-house admixtures were more complex because of the breadth of mixing ratio used. The very small differences in raw data fingerprinting can be harnessed, however, in combination with multivariate analysis to build appropriate multivariate classification models.

3.3. Wavelength selection and spectral pre-processing

FT-IR spectral data were pre-processed to remove undesirable systematic variation within the data to increase predictive accuracy of calibration models. To test the appropriateness of the various pre-

processing combinations, the R2 value was used to describe the variation among all the components of the calibration set explained by the model and the Q2 informs about the predictability of the model using the cross validation was used (Table 1). Region 1 (all 1790 variables) seems to produce the best models overall compared to those of Region 2 (carbonyl functional group focused 623 variables), and Region 3 (fingerprint zone: 728 variables). This was despite the full spectrum containing zones with zero or weak vibrations that would theoretically add noise to the data if included in the modelling. This reiterates that unperceivable to the naked eye spectral details are sometimes responsible for the projection and classification results of these types of multivariate models (Ellis *et al.*, 2012). After several iterations, the most suitable pre-processing was pareto scaling with standard normal variate couple and Exponentially Weighted Moving Average (EWMA), which provides a different type of smoothing than the typical Savitzky-Golay smoothing filter (Al-Mbaideen, 2019).

3.4. Principal component analysis (PCA) for FT-IR spectra

Fig. 6 shows PCA models obtained for both Scenario A, *C. cassia* adulteration with spend (Fig. 6A), and Scenario B, *C. verum* adulteration with cassia and spend (Fig. 6B) from FT-IR spectra (Region 1 - all variables). The separation in Fig. 6A is slightly better than that of Fig. 6B, which shows the *C. verum* class ("*verum*") overlapping the "admixture" class. Even if there is some within class clustering in the cassia group in Scenario A (Fig. 6A), as mentioned earlier, the *C. cassia* inner and outer bark were merged due to their difficulty to authenticate in the pure samples of the study, and their very minor compositional differences in the VOC profiles (Figs. 2 and 4). Also, strictly speaking, the commercial/edible part of cinnamon is the dry inner bark (Yeh et al., 2014).

In both cases, the admixtures with low adulteration percentages (<10%) overlap with the parent class, indicating the difficulty of the analytical problem and the limit of the spectroscopic technique in such small mixing ratios. To truly test the efficiency of the models, proper classification with external validation was conducted (see 3.5). Further exploration based on the origin of the samples within the same class (separate PCAs of the *C. cassia* class and of the *C. verum* class), and even following the same design but using higher order PCs, did not reveal any obvious clustering of geographical locations or any other patterns that would contradict the previous findings (data not shown).

3.5. Classification models to predict adulteration

Soft Independent Modelling of Class Analogy (SIMCA) and Partial Least Squares Discriminant Analysis (PLS-DA) are two supervised linear classification techniques which are widely used in qualitative fingerprint analysis of foodstuffs, and especially in herbs and spices (Haughey et al., 2015; Luna et al., 2017). These techniques have been used to investigate prediction accuracy of models developed, containing FT-IR spectra of cinnamon samples. SIMCA creates independent models for each class (in this study; different types of cinnamon) using PCA. It then uses linear algebra (Euclidean distance) to measure the distance of the unknown samples to the PCA space of every class. On the other hand, PLS-DA is based on PLS regression with the Y-variable generated for the categorical variable (here, the class type) mapped into a linear space. As in any regression, the reduced X–Y space is generated while preserving the maximum linear correlation between the samples' variables (X) and the class type (Y) (Reinholds *et al.*, 2015).

Table 1
Measure of fit (R2), prediction ability (Q2) and optimal number of factors of PCA models for *C. cassia* and *C. verum* adulteration (Scenario A, B).

Adulteration Scenario	PCA model	R2X (cum)	Q2 (cum)	Latent variables
A: *C. cassia* replacement with cinnamon spent	**Region 1**: 4000–600 cm^{-1} (1790 variables	**0.985**	**0.934**	**7**
	Region 2: 1800–600 cm^{-1} (623 variables)	0.904	0.781	9
	Region 3: 3000-2800 & 1800- 600 cm^{-1} (728 variables)	0.865	0.769	12
B: *C. verum* replacement with *C. cassia* and cinnamon spend	**Region 1**: 4000–600 cm^{-1} (1790 variables)	**0.979**	**0.952**	**6**
	Region 2: 1800–600 cm^{-1} (623 variables)	0.904	0.884	15
	Region 3: 3000-2800 & 1800–600 cm^{1} (728 variables)	0.865	0.823	11

Table 2 shows the classification performance of the PLS and SIMCA models in the two adulteration scenarios. Overall, when *C. cassia* is adulterated with spend (Scenario A), both PLS-DA and SIMCA are consistently good at detecting if there is an admixture (89.4% and 83.3% overall classification rate, respectively). When *C. verum* is adulterated with *C. cassia* and cinnamon spend (Scenario B), the admixture is detected equally good with PLS-DA (89.9%) and slightly worse with SIMCA (73.7%). These numbers consider both the recall (the correct classification rate of target samples) and precision (the specificity of the class, i.e., the ability to reject non target samples).

A complete model, however, accounts not only for the detection of the admixture class, but also for the detection of the pure samples (3 classes in total for Scenario A and 4 for Scenario B). In this case, the total accuracy of the model is a well-documented metric for comparison (Rodriguez et al., 2019). The classification results between scenarios indicate a balanced picture, especially given that Scenario B has more classes, thus is slightly more disadvantaged in the classification. This, in line with the difficulty of the analytical problem, namely that the distinction between *C. cassia* and cinnamon spend, is an equally challenging task to the distinction between *C. verum cinnamon* and cassia/spend when used together as the adulterants. This is down to two major factors: i) the differences in chemical composition (Chen et al., 2014) and in the volatile profile (as seen in GC-IMS experiment, see 3.1) of the two cinnamons (*C. verum vs C. cassia*) that favours the detection of *C. verum* (Scenario B), ii) the spend contains almost no essential oil, which is a key factor for discrimination in the case where it is the only adulterant (Scenario A). The interplay of these two factors accounts for the balance in the classification results, especially when both classifiers are considered.

Overall, the PLS-DA models were more efficient in comparison with SIMCA models. The PLS-DA overcomes the SIMCA classifier in both scenarios (in recall, overall classification rate and total accuracy). This is also a common observation reported in the literature (Jiménez-Carvelo *et al.*, 2021; Min et al., 2021). By design, the two methods use different criterion to build the multivariate models. PLS-DA does not allow classifying a sample to other groups than the ones originally assigned. As a result, all measured variables play the same role in class prediction. In SIMCA, each category is modelled separately focusing on the analogies among the class elements rather than on discriminating among the different categories. Therefore, while PCA class models are calculated in SIMCA with the goal of capturing variations within each class, PLS-DA works more directly by identifying directions in the data space that discriminate classes and forces membership. This makes it more efficient to use PLS-DA for these applications, rather than the SIMCA classifier (Galtier *et al.*, 2011).

The FT-IR/PLS-DA detection of the *C. cassia* adulteration exhibits 95% total accuracy, high R2 and Q2 scores and consistency high F1, recall and precision and overall classification rate for all classes and it is the best model overall. The FT-IR/PLS-DA detection of *C. verum* adulteration (Scenario B) exhibits 90% total accuracy, especially in detecting admixtures and pure cassia, with the caveat of imperfect detection of pure spend and pure *C. verum* (which was also equally challenging with

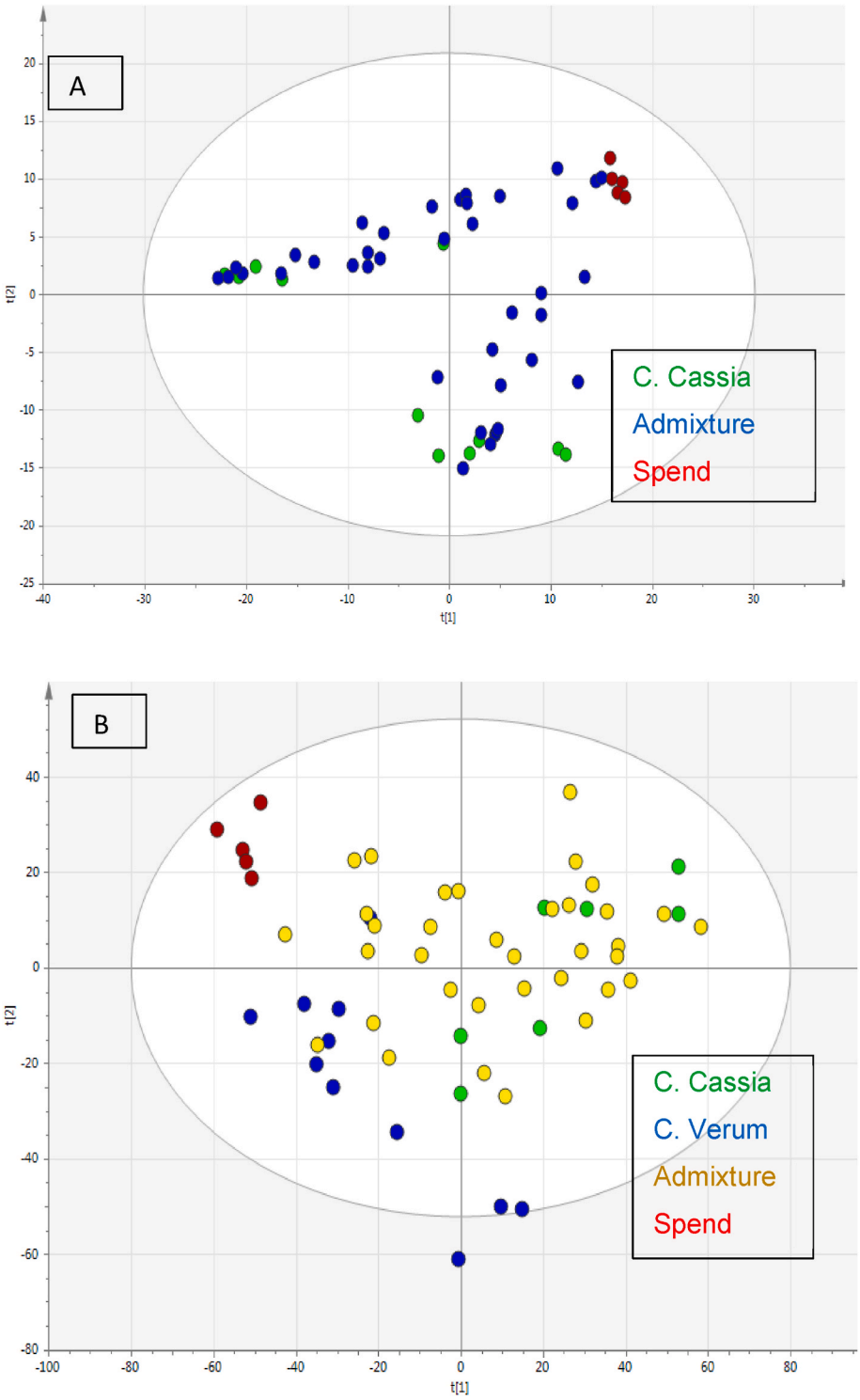

Fig. 6. PCA score plots of the FT-IR data from adulteration Scenario A (*C. cassia* adulteration with spend) and Scenario B (*C. verum* adulteration with cassia and cinnamon spend*). C. Cassia*: green*; C. verum*: blue; *C. spent*: red colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the SIMCA classifier). Arguably, there is an element of overtraining of the pure classes, especially the spend class, because it contains only two members. Class membership is not well balanced, so that ideally more pure samples of *C. cassia* and *C. verum* are required to increase the overall reliability of the calibration model and balance the class membership, including more of the various cinnamon species (for example, *C. burttii, C. mannii, and C. leiandra*).

Examination of the misclassification results in the context of precision, demonstrated it was not possible to identify samples which have been adulterated by less than 20% of cinnamon spend in both scenarios. In fact, the model led to a better classification of the spend samples and admixtures between the percentage of adulteration was between 20% and 80%. This is based on pure observation of the classification results using the external validation set. The more formal of process for calculating the limit of detection in a classification study was not followed because it is a different experiment (Downey & Kelly,

Table 2
Classification performance in cinnamon adulteration based on FT-IR spectral fingerprints. Scenario A (*C. cassia* adulteration with spend) and Scenario B (*C. verum* adulteration with *C. cassia* and spend).

Model design		Class	F1	Recall (%)	Precision (%)	Overall classification rate (%)
Scenario A	*PLS-DA*	*C. cassia*	0.86	75	100	94.4
		Admixture	0.91	**100**	**83.4**	**89.4**
		Spend	1.00	100	100	100.0
		Accuracy: 94.4% R2: 0.99 Q2:0.47				
	SIMCA	*C. cassia*	0.40	25	100	83.3
		Admixture	0.67	**100**	**50**	**83.3**
		Spend	1.00	100	100	100.0
		Accuracy: 83.4% R2: 0.99 Q2:0.98				
Scenario B	*PLS-DA*	*C. verum*	0.67	50	100	94.7
		C. cassia	1.00	100	100	100.0
		Admixture	0.86	**100**	**75**	**89.9**
		Spend	0.67	50	100	94.7
		Accuracy: 89.5% R2: 0.93 Q2: 0.72				
	SIMCA	*C. verum*	0.62	50	82.4	94.7
		C. cassia	1.00	100	100	100
		Admixture	0.81	**72.9**	**90.9**	**73.7**
		Spend	1.00	100	100	100.0
		Accuracy: 78.9% R2: 0.99 Q2:0.97				

2004) and was not the objective of this study.

A direct comparison of the above results with the study of Yasmin *et al.* (2019) is not possible because they used PLS-Regression instead of PLS-DA classification and a simpler analytical problem (binary) focused on quantification of *C. cassia* in *C. verum*. They also used both FT-NIR an FT-IR and a much narrower mixing ratio (5–50 wt %). Despite the different methodology, their error of the prediction (~4%), limit of detection (~6.2%) and limit of quantification (22.8%) using FT-IR were within the range of this study. Compared with similarly designed studies on other herbs and spices, such as sage (Galvin-King et al., 2019) and paprika (Lohumi *et al.*, 2017), that were focused on classification classifiers rather than binary regression, the results of the present study demonstrate similar, or better results.

Compared to other types of methods (such as the chromatographic methods as seen in the studies of Woehrlin et al. (2010) and Li et al. (2013) where targeted analysis is applied (coumarin or cinnamaldehyde), the spectroscopic determination is significantly faster (1–3 min) compared to 60 min on average including sample preparation, and arguably much more accurate, although the comparison is not fair due to the different use case of the methods (screening *vs* confirmation) (**Osorio *et al.***, 2015).

4. Conclusions

There are few studies related to FT-IR spectroscopy and cinnamon analysis in the context of detecting food fraud in supply chains. This is the first study that has explored the use of FT-IR spectroscopy to detect cinnamon adulteration rapidly and reliably from two pure varieties and from cinnamon waste. Although a small sample set was used from limited geographical locations that does not fully represent the variability of the entire global supply, our results nevertheless show that FT-IR can be used as a routine screening method to discriminate between pure and adulterated samples. Apart from the obvious benefits of speed and cost of analysis compared to wet chemistry methods such as LC or GC, this test can also be implemented in a portable FT-IR setup for field measurements outside a well-equipped lab. In fact, recent advances in miniaturisation already allow for small, battery-operated instruments, ideal for supply chain quality controls for cinnamon and potentially other herbs and spices.

In terms of prediction performance, the correct classification rate demonstrated between 94.4% and 78.9% and was consistently above 90% when PLS-DA was used as a classifier. Both *C. verum* and *C. cassia* adulteration were tackled effectively, and the respective calibration models developed both showed high prediction performance. Overall, the results demonstrate very good potential for this method to identify fraud in ground dried cinnamon. The method is highly sensitive, rapid, and provides low-cost results while requiring no sample preparation, especially if the admixture ratio is above the 20% threshold. These factors mean that reliable and robust screening protocols based on this methodology could potentially be routinely deployed in non-lab field locations and implemented by relatively non-specialised staff to allow for rapid and cost-effective screening of large volumes of samples in food supply chains.

For future studies, increasing the number of samples in pure class groups could enhance the training of the calibration model, balance the class membership, and add variability to the data allowing more unique chemical characteristics to be recognised within each pure species and potentially lowering the detection threshold to more sensitive levels for the admixtures. Apart from further investigating GC-IMS potential in the identification of cinnamon adulteration using VOC profiles, other future work could include NIR spectroscopy combined with FT-IR to provide a better compositional and structural analysis of cinnamon samples, as well as possibly linking their mineral components with their place of origin, using techniques such as GC-IMS or XRF for elemental and trace analysis.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Panagiota Lixourgioti: Investigation, Writing – original draft, Formal analysis, Visualization. **Kirstie A. Goggin:** Investigation, Formal analysis, Visualization, Writing – review & editing. **Xinyu Zhao:** Writing – original draft, Formal analysis, Visualization. **Denis J. Murphy:** Supervision, Writing – review & editing. **Saskia van Ruth:** Conceptualization, Writing – review & editing. **Anastasios Koidis:** Conceptualization, Methodology, Supervision, Writing – review & editing, Resources.

Acknowledgements

The authors would like to thank Miss Rachel Moore, BSc student at the School of Biological Sciences, Queen's University Belfast, for her kind help milling the samples.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lwt.2021.112760.

References

Abdi, H., & Williams, L. J. (2010). Principal component analysis. ' *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(4), 433–459.

Al-Mbaideen, A. A. (2019). Application of moving average filter for the quantitative analysis of the NIR spectra. *Journal of Analytical Chemistry, 74*(7), 686–692.

Avula, B., Smillie, T., Wang, Y., Zweigenbaum, J., & Khan, I. (2014). Authentication of true cinnamon (*Cinnamon verum*) utilising direct analysis in real time (DART)-QToF-MS. *Food Additives & Contaminants: Part A, 32*(1), 1–8.

Bansal, S., Thakur, S., Mangal, M., Mangal, A. K., & Gupta, R. K. (2018). DNA barcoding for specific and sensitive detection of *Cuminum cyminum* adulteration in *Bunium persicum*. *Phytomedicine, 50*, 178–183.

Black, C., Haughey, S., Chevallier, O., Galvin-King, P., & Elliot, C. (2016). A comprehensive strategy to detect the fraudulent adulteration of herbs: The oregano approach. *Food Chemistry, 201*, 551–557.

Borsdorf, H., Mayer, T., Zarejousheghani, M., & Eiceman, G. A. (2011). Recent developments in ion mobility spectrometry. *Applied Spectroscopy Reviews, 46*(6), 472–521.

Chen, P., Sun, J., & Ford, P. (2014). Differentiation of the four major species of cinnamons (*C. burmannii, C. verum, C. cassia,* and *C. loureiroi*) using a flow injection mass spectrometric (FIMS) fingerprinting method. *Journal of Agricultural and Food Chemistry, 62*(12), 2516–2521.

Downey, G., & Kelly, J. D. (2004). Detection and quantification of apple adulteration in diluted and sulfited strawberry and raspberry purées using visible and near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry, 52*, 204–209.

Ellis, D., Brewster, V., Dumn, W., Allwood, Golonavov, A., & Godacre, R. (2012). Fingerprinting food: Current technologies for the detection of food adulteration and contamination. *Chemical Society Reviews, 41*, 5706–5727.

Galtier, O., Abbas, O., Le Dréau, Y., Rebufa, C., Kister, J., Artaud, J., & Dupuy, N. (2011). Comparison of PLS1-DA, PLS2-DA and SIMCA for classification by origin of crude petroleum oils by MIR and virgin olive oils by NIR for different spectral regions. *Vibrational Spectroscopy, 55*(1), 132–140.

Galvin-King, P., Haughey, S. A., & Elliott, C. T. (2018). Herb and spice fraud; the drivers, challenges and detection. *Food Control, 88*, 85–97.

Galvin-King, P., Haughey, S. A., Montgomery, H., & Elliott, C. T. (2019). The rapid detection of sage adulteration using fourier transform infra-red (FT-IR) spectroscopy and chemometrics. *Journal of AOAC International, 102*(2), 354–362.

Garrido-Delgado, R., del Mar Dobao-Prieto, M., Arce, L., & Valcárcel, M. (2015). Determination of volatile compounds by GC–IMS to assign the quality of virgin olive oil. *Food Chemistry, 187*, 572–579.

Haughey, S., Galvin-King, P., Ho, Y., Bell, S., & Elliott, C. (2015). The feasibility of using near infrared and Raman spectroscopic techniques to detect fraudulent adulteration of chill powders with Sudan dye. *Food Control, 48*, 181–186.

He, Z. D., Qiao, C. F., Han, Q. B., Cheng, C. L., Xu, H. X., Jiang, R. W., But, P. H., & Shaw, P. C. (2005). Authentication and quantitative analysis on the chemical profile of Cassia bark (Cortex *Cinnamomum*) by high-pressure liquid chromatography. *Journal of Agricultural and Food Chemistry, 53*(7), 2424–2428.

Jiménez-Carvelo, A. M., Tonolini, M., McAleer, O., Cuadros-Rodríguez, L., Granato, D., & Koidis, A. (2021). Multivariate approach for the authentication of vanilla using Infrared and Raman spectroscopy. *Food Research International, 141*, 110196.

Li, Y. Q., Kong, D. X., & Wu, H. (2013). Analysis and evaluation of essential oil components of cinnamon barks using GC–MS and FT-IR spectroscopy. *Industrial Crops and Products, 41*, 269–278.

Lim, C. M., Carey, M., Williams, P., & Koidis, A. (2021). Rapid classification of commercial teas according to their origin and type using elemental content with X-ray fluorescence (XRF) spectroscopy. *Current Research in Food Science*. https://doi.org/10.1016/j.crfs.2021.02.002

Lin, R., Mohamed, M., Chen, T., & Kuo, S. (2017). Coumarin- and carboxyl-functionalized supramolecular polybenzoxazines form miscible blends with polyvinylpyrrolidone. *Polymers, 9*, 146.

Lohumi, S., Joshi, R., Kandpal, L. M., Lee, H., Kim, M. S., Cho, H., Mo, C., Seo, Y. W., Rahman, A., & Cho, B. K. (2017). Quantitative analysis of Sudan dye adulteration in paprika powder using FT-IR spectroscopy. *Food Additives & Contaminants: Part A, 34*(5), 678–686.

Lohumi, S., Sangdae, L., Lee, W., Kim, M., Mo, C., Bae, H., & Cho, B. (2014). Detection of starch adulteration in onion powder by FT-NIR and FT-IR spectroscopy. *Journal of Agricultural and Food Chemistry, 62*(38), 9246–9251.

Luna, A., da Silva, A., Alves, A., Rocha, R., Lima, I., & de Gois, J. (2017). Evaluation of chemometrics methodologies for the classification of *Coffea canephora* cultivars via FT-NIR spectroscopy and direct sample analysis. *Analytical Methods, 9*(29), 4255–42609.

Osorio, M. T., Haughey, S. A., Elliott, C. T., & Koidis, A. (2015). *Identification of vegetable oil botanical speciation in refined vegetable oil blends using an innovative combination of chromatographic and spectroscopic techniques*.

Quiroz-Moreno, C., Furlan, M. F., de Souza, J. R. B., Augusto, F., Alexandrino, G. L., & Mogollón, N. G. (2020). RGCxGC toolbox: An R-package for data processing in comprehensive two-dimensional gas chromatography-mass spectrometry. *Microchemical Journal*, 104830.

Rahadian, D., Muhammad, A., & Downtick, K. (2017). Cinnamon and its derivatives as potential ingredient in functional food—a review. *International Journal of Food Properties, 20*(2), 2237–2263.

Reinholds, I., Bartkevics, V., Silvis, I., van Ruth, S., & Esslinger, S. (2015). Analytical techniques combined with chemometrics for authentication and determination of contaminants in condiments: A review. *Journal of Food Composition and Analysis, 44*, 56–72.

Rodriguez, S., Rolandelli, G., & Buera, M. (2019). Detection of quinoa flour adulteration by means of FT-MIR spectroscopy combined with chemometric methods. *Food Chemistry, 274*, 392–401.

Swetha, V. P., Parvathy, V. A., Sheeja, T. E., & Sasikumar, B. (2014). DNA barcoding for discriminating the economically important Cinnamomum verum from its adulterants. *Food Biotechnology, 28*(3), 183–194.

Thomas, J., & Kuruvilla, K. M. (2012). *Cinnamon. Handbook of herbs and spices*. Woodhead Publishing.

Woehrlin, F., Fry, H., Abraham, K., & Preiss-Weigert, A. (2010). Quantification of flavoring constituents in cinnamon: High variation of coumarin in Cassia bark from the German retail market and in authentic samples from Indonesia. *Journal of Agricultural and Food Chemistry, 58*(19), 10568–10575.

Yasmin, J., Ahmed, M. R., Lohumi, S., Wakholi, C., Lee, H., Mo, C., & Cho, B. K. (2019). Rapid authentication measurement of cinnamon powder using FT-NIR and FT-IR spectroscopic techniques. *Quality Assurance and Safety of Crops & Foods, 11*(3), 257–267.

Yeh, T., Lin, C., & Chang, S. (2014). A potential low-coumarin cinnamon substitute: *Cinnamomum osmophloeum* leaves. *Journal of Agricultural and Food Chemistry, 62*(7), 1706–1712.