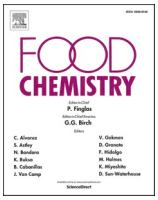
Food Chemistry 404 (2023) 134649

Contents lists available at ScienceDirect

Food Chemistry

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



Authentication of saffron using 60 MHz ^{1}H NMR spectroscopy

Yvonne Gunning, Kate S. Davies, E. Kate Kemsley *

Core Science Resources Group, Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, UK

ARTICLE INFO

Keywords:
Saffron
NMR
Spectroscopy
Benchtop
Authenticity
Adulteration
Outlier
Anomaly

ABSTRACT

60 MHz proton NMR spectroscopy was used to analyse extracts from saffron spice and a range of potential adulterants and mixtures. Using a simple extraction procedure, good quality spectra were obtained which contain peaks from the characteristic metabolites picrocrocin and crocins, fatty acids and kaempferol. The spectra of samples from trusted suppliers were used to train one-class classification models by SIMCA, nearest neighbour and isolation forest methods. Applying these to spectra of saffron samples purchased from the online marketplace, it was found that 7 out of 33 samples were highly anomalous. From comparison with the spectra of known mixtures and confirmatory spectral analysis using 600 MHz NMR, it is probable that these contain considerable amounts of undisclosed foreign matter.

1. Introduction

The culinary spice saffron is produced from the dried stigmas of *Crocus sativus L.* flowers, cultivated in many countries including Iran, Spain, Turkey, Morocco, Italy and Greece. Extensively used in the preparation of food, saffron is valued for its distinctive colour, flavour and taste. These are due to certain secondary metabolites found in the plant (Mykhailenko, Kovalyov, Goryacha, Ivanauskas, & Georgiyants, 2019), notably crocins, a family of water-soluble pigments that contribute to the red colour, and picrocrocin, a colourless glycoside which is largely responsible for the unique taste. Picrocrocin is a precursor for the monoterpene safranal, a volatile oil responsible for the characteristic aroma of saffron. In addition to its culinary use, saffron has been used as a dye in the textile and cosmetic industries, and in traditional medicine for the treatment of various diseases (Hosseinzadeh & Nassiri-Asl, 2013). There is increasing evidence that saffron contains clinically useful bioactive components related to the crocins, safranal and crocetin (Christodoulou, Kadoglou, Kostomitsopoulos, & Valsami, 2015; Moshiri, Vahabzadeh, & Hosseinzadeh, 2015). Clinical findings suggest that saffron is safe in therapeutic doses (Bostan, Mehri, & Hosseinzadeh, 2017; Modaghegh, Shahabian, Esmaelli, Rajbai, & Hosseinzadeh, 2008).

The production of saffron is highly labour-intensive, requiring hand-picking of the stigmas ('strands' or 'threads'), three per blossom (see Supplementary Fig. 1). Approximately 450,000 are needed to produce 1 kg of strands, making saffron the most expensive spice in the world. Its value further depends on quality aspects relating to the secondary metabolites. These are influenced by factors including climate, soil composition, drying and storage conditions. The limited production and very high market value make saffron vulnerable to fraud. Adulteration strategies used by counterfeiters commonly aim to increase the volume and weight of the crop (Negi, Pare, & Meenatchi, 2021). Typical bulking agents include: a range of mineral substances, various liquids (vegetable oil, honey) in which the stigmas are soaked before drying, different parts of the saffron plant, and other plant-based materials (Torelli, Marieschi, & Bruni, 2014). Other types of fraud aim to boost the perceived quality by adding natural or synthetic colourants.

Considerable effort has been directed towards developing methods for verifying the authenticity of saffron (Kumari, Jaiswal, & Tripathy, 2021). Untargeted methods that can detect a broad spectrum of unknown adulterants are especially sought after, in particular those that involve minimal sample preparation and are easily implemented. Infrared spectroscopy has been explored for detecting adulteration of saffron with various plant-based adulterants (Amirvaresi, Nikounezhad, Amirahmadi, Daraei, & Parastar, 2021; Petrakis & Polissiou, 2017) and for examining the quality of traded saffron (Ordoudi, Pascual, & Tsimidou, 2014). An untargeted metabolite fingerprinting method has been reported that uses UPLC-MS with multivariate data analysis to distinguish saffron samples according to their geographic origin (Rubert, Lacina, Zachariasova, & Hajslova, 2016).

Of interest in the present work is another untargeted profiling approach, nuclear magnetic resonance (NMR) spectroscopy. Various

* Corresponding author.
E-mail addresses: yvonne.gunning@quadram.ac.uk (Y. Gunning), kate.davies@quadram.ac.uk (K.S. Davies), kate.kemsley@quadram.ac.uk (E.K. Kemsley).

https://doi.org/10.1016/j.foodchem.2022.134649
Received 7 June 2022; Received in revised form 11 October 2022; Accepted 14 October 2022
Available online 18 October 2022

researchers have used high-field (400 MHz and above) NMR for investigating authentication and quality issues in saffron, often in conjunction with chemometrics. Reported studies include the detection of adulteration with plant-based materials (Musio et al., 2022; Petrakis, Cagliani, Polissiou, & Consonni, 2015) and with synthetic dyes (Dowlatabadi et al., 2017; Petrakis, Cagliani, Tarantilis, Polissiou, & Consonni, 2017). Yilmaz, Nyberg, Molgaard, Asili, and Jaroszewski (2010) sought to distinguish authentic Iranian saffron from authentic and adulterated samples from other countries. NMR has also been used to examine quality and shelf life issues: Ordoudi, Cagliani, Lalou, Naziri, Tsimidou, and Consonni (2015) proposed markers for detecting deterioration in quality, and Consonni, Ordoudi, Cagliani, Tsiangali, and Tsimidou (2016) tracked changes in saffron with storage.

These studies capitalise on a recognised advantage of high-field NMR spectroscopy: its ability to provide wide-ranging information on the soluble constituents of a sample prepared without any separation procedure, thereby preserving the original ratios of compounds present (Cagliani, Culeddu, Chessa, & Consonni, 2015). Less attractive are the capital and running costs of spectrometers that utilise cryogenic superconducting magnets and probes, along with specialist staff to operate and maintain the spectrometer. In the present work, we explore the potential of 60 MHz 'benchtop' proton (^{1}H) NMR spectroscopy for the inexpensive, rapid and robust screening of saffron. To the best of our knowledge, benchtop NMR has not previously been considered in this capacity, although there are reported uses of this emerging technology for studying secondary metabolites in extracts from plant material (Gunning et al., 2018; Pages, Gerdova, Williamson, Gilard, Martino, & Malet-Martino, 2014; Wu et al., 2021).

In the present work, we acquire 60 MHz benchtop ^{1}H NMR spectra from a collection of authentic saffron samples, as well as mixtures with a range of potential adulterants. Sample preparation is kept to a minimum, with extractions carried out using a low-cost, low-risk solvent, with the goal of achieving an analytical protocol suitable for translation into commercial settings. For data analysis, we employ one-class classification methods which model only the 'target' class ('authentic' saffron) and use a threshold to accept or reject further items. Such methods are intended for use with heavily unbalanced datasets (Santoyo-Ramon, Casilari, & Cano-Garcia, 2021; Seliya, Zadeh, & Khoshgoftaar, 2021). This approach is preferred in the present work, because it is difficult to obtain examples of proven fraudulent products in sufficient numbers to develop a binary 'authentic vs adulterated' classification model. Further, the wide range of possible adulterant materials means that simulating such samples is challenging; however, we show that for selected scenarios, the approach is capable of detecting adulteration at useful levels. Finally, using the established protocol, we conduct a survey of further saffron samples procured from suppliers in the international online marketplace. Our findings strongly suggest that there is an ongoing issue with adulteration in the sector.

2. Materials and methods

2.1. Samples

2.1.1. Saffron

52 retail packs of saffron were purchased at intervals over a period of two years from six different UK suppliers (labelled A1 – A6) with most from a retailer (A1) recognised for their exemplary supply chain responsibility (Spence & Bourlakis, 2009). All were sold as whole dried stigmas, which in practice means a mixture of intact strands as well as fragments (see Supplementary Fig. 1(c)). These samples were used as the 'authentic' reference collection. Additionally, saffron bulbs were obtained from a UK supplier and cold-house grown locally in Norwich, UK. This produced twelve saffron plants from which the stigmas were harvested and dried, giving approximately 50 mg of dried material. This sample was also included in the authentic collection, giving 53 biologically independent samples in total. These are summarised in Supplementary Table 1.

A further 33 packs of saffron, regarded as 'survey' samples, were purchased from online suppliers. The majority of these were sold as whole dried stigmas; three were sold as ground powder. In all cases, the product labelling claimed the contents to be wholly comprised of saffron. Where sufficient material was available, duplicate analyses were carried out, at the time of purchase and after a period in storage. In some cases, repeat purchases of the same brand were made at intervals over the two-year period of study. These details are given in Supplementary Table 2.

2.1.2. Potential adulterants

Seven materials that have been documented in the literature as possible adulterants of saffron were studied in this work. Dried petals of arnica (*Arnica montana*), calendula (*Calendula officinalis*, 'pot marigold') and safflower (*Carthamus tinctoris*), turmeric spice (the ground dried rhizome of *Curcuma longa*), cayenne pepper (ground dried fruits of *Capsicum annuum*) and sandalwood (ground wood of *Santalum* genus) were purchased from online UK retailers. Tartrazine ('acid yellow 23') analytical standard was purchased from Sigma-Aldrich (Merck Life Sciences, Gillingham, UK).

2.2. Sample preparation and extraction

2.2.1. Saffron

For the saffron samples, stigmas were ground in a pestle and mortar, and 50 mg of the ground material was placed in 1 mL of deuterated dimethyl sulfoxide (DMSO-d_6, 99.9 % D, 0.03 % (v/v) tetramethylsilane (TMS), Merck Life Sciences, Gillingham, UK). For the saffron samples purchased ready ground, 50 mg of powder was weighed directly into an Eppendorf tube and 1 mL of DMSO-d_6 added. After 30 mins steeping at laboratory ambient temperature (21 ℃), the supernatant was filtered through cotton wool into an NMR tube (Aldrich® Colorspec™ disposable NMR tubes, diameter 5 mm, size 8in from Merck Life Sciences UK, Gillingham, UK). DMSO-d_6 was chosen as it is an effective solvent of the metabolites of interest in saffron, giving simultaneous access to both lipophilic and hydrophilic compounds (Cagliani et al., 2015). Further, it is aprotic, relatively inert and also nontoxic, making it amongst the safest of NMR solvents.

2.2.2. Mixtures of saffron with adulterants

36 mixture samples were prepared by combining different proportions by weight of ground saffron with each adulterant (nominally 10, 20, 30, 40 and 50 % w/w of adulterant, with additionally the level 5 % w/w of tartrazine). The dried petals of arnica, calendula and safflower were ground in a pestle and mortar before weighing and mixing with the ground saffron. Extraction of 50 mg of each ground mixture was carried out as described for saffron above. The saffron samples used in preparing the mixture series were taken from the authentic collection (see Supplementary Table 3 for details).

2.3. NMR spectral acquisition

2.3.1. 60 MHz ^{1}H NMR spectroscopy

60 MHz ^{1}H NMR spectra were acquired from all extracts using a Pulsar benchtop spectrometer (Oxford Instruments, Tubney Woods, Abingdon, Oxford, UK) running SpinFlow software (Oxford Instruments). The spectrometer magnet is maintained at 37 ℃ which gives a temperature inside the probe of ~ 35 ℃ during acquisition. The 90° pulse length was 14.4 μs as determined by the machine's internal calibration cycle. For each extract, 256 free induction decays (FIDs) were collected using a filter width of 5000 Hz, a scan time of approximately 6.55 s and recycle (inter-scan) delay of 2.0 s, resulting in a total acquisition time of approximately 36 min per extract. These parameters were chosen as we have found from previous studies of plant material extracts that they represent a good compromise between sample throughput and

spectral signal-to-noise. FIDs were zero-filled to give spectra containing 32,768 data points. The FIDs were Fourier-transformed, co-added and phase-corrected using SpinFlow and MNova software packages (Mestrelab Research, Santiago de Compostela, Spain) to present a single frequency-domain spectrum from each extract. The linewidth was checked daily using a sealed standard sample of TMS in chloroform-d, and shimming carried out as needed to keep the FWHM of the chloroform peak at ≤ 0.5 Hz. The chemical shift scale in all extract spectra was referenced to the DMSO-d_6 peak at 2.5 ppm.

2.3.2. 600 MHz ^{1}H NMR spectroscopy

Some selected extracts were also analysed by high-field, high-resolution NMR, using a Bruker Avance NEO 600 MHz spectrometer equipped with a triple resonance TCI cryoprobe. The console was controlled by Topspin 2.0 software. Spectra were recorded at 298 K with 16 FIDs co-added for each extract. The chemical shift scale was referenced to the DMSO-d_6 peak at 2.5 ppm.

2.4. Data analysis

All data analysis was carried out in Matlab (The Mathworks, Cambridge, UK) using functions from the associated Statistics and Machine Learning Toolbox, and from the publicly available DD-SIMCA Matlab tool (Zontov, Rodionova, Kucheryavskiy, & Pomerantsev, 2017). Three different one-class classification approaches were employed to model the target class (the spectra of the authentic reference collection). These were soft independent modelling of class analogy (SIMCA), one-class classification using nearest neighbours (OCC-NN), and isolation forests (IF).

Based upon principal component analysis (PCA), SIMCA was originally proposed in Wold (1976) and has since become a widely accepted method for tackling one-class classification. A recent proposed modification is data driven SIMCA (DD-SIMCA) which is described in detail by Zontov et al. (2017); this has been employed to provide a benchmark analysis in the present work.

OCC-NN are a family of density-based outlier detection algorithms (Tang & He, 2017). In its simplest form, OCC-NN examines the local neighbourhood of an observation by measuring the distance(s) to its nearest neighbour(s) within the target class. If this is higher than some threshold value, the observation is rejected as anomalous, otherwise it is accepted as a member of the target class. The threshold and number of nearest neighbours considered are adjusted to give an acceptable rate of type I errors ('false positives' or 'false alarms', expressed by the 'contamination fraction' hyperparameter). The implementation of OCC-NN in the present work additionally employs the ensemble techniques of 'feature' (spectral data point) sampling to generate diverse classifiers, and aggregation across the ensemble to obtain outcomes. Matlab code for carrying out the method is given in Supplementary Fig. 2.

An IF comprises an ensemble of binary partition trees constructed from the target class, each of which is grown until all items occupy a separate leaf node (Liu, Ting, & Zhou, 2012). To create diverse trees, there is random sampling of both features and observations. The premise is that outliers will be isolated sooner than typical observations, with shorter paths from root to leaf nodes compared with the average over all trees. The comparison threshold value is chosen to give a type I error rate consistent with the contamination fraction hyperparameter. Because the main aim is isolation of anomalous observations from the rest, the size of the target class is not a crucial factor in determining outcomes, thus IF is considered to be a suitable approach for small datasets. This contrasts with many multivariate methods where large numbers of observations are required. In the present work, IF has been implemented using the isolation forest algorithm from the Matlab Statistics and Machine Learning Toolbox.

For all the modelling methods, all hyperparameters were established at the training stage, making use of the target class spectra only. In DD-SIMCA, resampled k-fold cross-validation was used to select the number of PC dimensions to use without overfitting the model, whereas the two ensemble approaches inherently involve extensive resampling, cross-validation and aggregation, which boost accuracy whilst minimising the overfitting risk.

Once trained, the models were used to make predictions about further observations: the mixture and survey sample spectra. As examples of adulterated saffron, the mixture series are true members of the 'outlier' class. Although such samples play no role in training one-class models, they can usefully act as test samples for directly comparing different models' performances. (Note that the mixture compositions are not intended to represent all possible embodiments of adulterated saffron; indeed, comprehensive simulation of the outlier class would be an inordinate challenge, and unnecessary in the context of one-class classification.) The survey samples were all sold as 'pure saffron', so the *a priori* expectation is that they will be accepted as such by one-class models at the same rate as the training samples; obtaining more outliers than anticipated suggests the presence of compromised samples.

3. Results and discussion

3.1. Data exploration

The 60 MHz ^{1}H NMR spectrum of an extract of an authentic saffron is shown in Fig. 1, where it is compared with a spectrum of the same sample obtained using 600 MHz ^{1}H NMR. The region containing the DMSO-d_6 peaks at 2.5 ppm and the water peak at 3.4 ppm has been discarded, because in the 60 MHz spectrum, this can be highly variable (see the complete set of authentic saffron 60 MHz spectra given in Supplementary Fig. 3). This reflects both the absorption of water by DMSO-d_6 and the dehydration state of the samples. From a pattern recognition viewpoint, it is preferable to exclude regions with such confounding information content. This approach was taken in previously published work on seized controlled substances (Antonides et al., 2019) which have unknown and/or poorly controlled storage histories.

Some of the major features are marked, drawing on the comprehensive high field assignments by Cagliani et al (2015). These include several resonances arising from the key metabolites found in saffron. Picrocrocin gives rise to large peaks at 1.16, 1.18, 2.10 and 10.05 ppm, the latter a singlet from the aldehydic proton of the 2,6,6-trimethyl-1-cyclohexene-1-carboxaldehyde moiety. This feature is well resolved at both field strengths and occurs in a region of the NMR spectrum that is often sparsely populated, making it a useful visual marker for the presence of saffron in a sample.

More generally, the 60 MHz spectrum appears as a broad envelope of features that maps onto the much more resolved peaks in the 600 MHz spectrum. However, it is important to recognize that notwithstanding this lack of peak resolution, the low field benchtop spectra represent the same underlying chemical information in overlapped form. Multivariate statistical methods are useful for examining this type of information content. Principal Component Analysis (PCA) using the covariance matrix form was applied to the collection of authentic saffron spectra. The spectral regions utilised were as shown in Fig. 1. To reduce spectral variability caused by differences in extraction efficiency, the data were pre-treated using standard normal variate (SNV) correction. This scales each spectrum to have a mean of zero and variance of unity; it is a commonly used pre-treatment for mitigating unwanted spectral variation that is multiplicative in nature (Barnes, Dhanoa, & Lister, 1989) and has been found effective in the analysis of benchtop NMR spectra (Gunning et al., 2022).

The first 7 PCs accounted for nearly 95 % of the total dataset variance (see Supplementary Fig. 4), with the first two alone representing 75 %; the scores for these dimensions are plotted against one another in Fig. 2. For normally distributed scores, the squared Mahalanobis distances from the authentic group centre in this 2-d space are distributed as χ^2 with 2 degrees of freedom. The ellipse marked on the plot corresponds to the $p = 0.05$ critical value. This provides a simple classifier for accepting

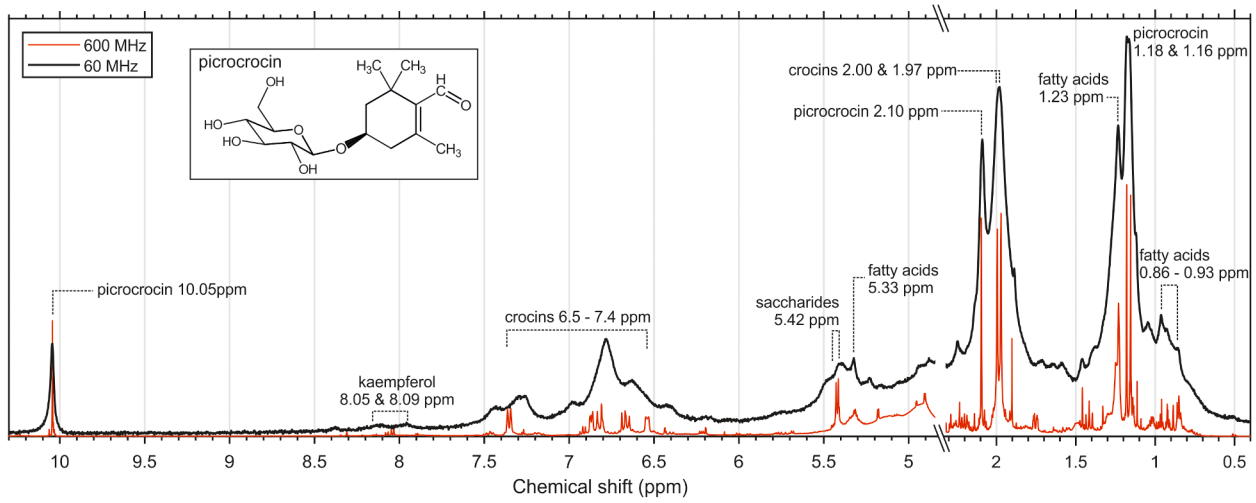


Fig. 1. Comparison of 60 MHz (benchtop) and 600 MHz (high field) ^{1}H NMR spectra obtained from an extract of saffron in DMSO-d_6, with annotations of the main features identified in the 60 MHz spectrum. The benchtop spectral profile is dominated by resonances attributed to picrocrocin, a major saffron metabolite, the structure of which is given in the inset panel. The isolated peak at 10.05 ppm is in a generally uncrowded region of the spectrum and provides a useful indicator of the presence of saffron in a sample.

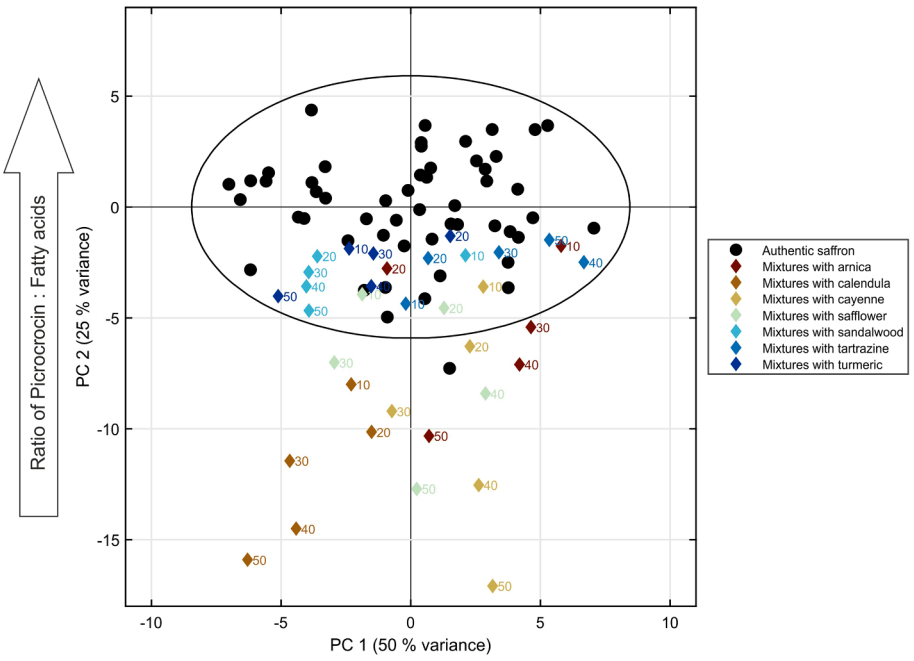


Fig. 2. First versus second Principal Component scores calculated from the collection of authentic saffron spectra. The ellipse corresponds to the $p = 0.05$ critical value for the χ^2 distribution with 2 degrees of freedom. 95 % of authentic samples are expected to be located within this boundary. Also shown transformed into the same PC space are the scores for the mixture samples, colour-coded and labelled with the %w/w added 'adulterant'. None of the mixtures with sandalwood, tartrazine or turmeric are detected by this classification method.

or rejecting new samples as authentic at the chosen probability level. To illustrate, also shown transformed onto the same axes are the scores for the extracts of mixtures with potential adulterants. The mixtures with calendula are all correctly rejected as inauthentic, as are many of the mixtures with arnica, cayenne and safflower, with scores along the second PC axis responsible for the discrimination. From examination of the PC loading (See Supplementary Fig. 4), this dimension is strongly associated with the ratio picrocrocin:fatty acids, in the direction as indicated by the arrow. This shows that as the adulterant content increases in the mixtures, so the proportion of picrocrocin decreases. Notably, the 60 MHz spectra of pure extracts of these particular adulterants have spectral profiles consistent with predominantly fatty acids. These are shown in Fig. 3, along with those of the other potential adulterants under study, and a saffron spectrum for comparison.

In contrast, all of the mixtures adulterated with sandalwood, tartrazine or turmeric are incorrectly accepted as authentic (type II errors). This is despite their 60 MHz spectra differing greatly from that of saffron (see Fig. 3). This is because the PCA is performed on the authentic class only, and regions in which there is low variance in saffron spectra exert relatively little influence on the transformation. For example, one of the strongest peaks in the turmeric spectrum is at 7.2 ppm, but this is one of the least variable regions in saffron (see Supplementary Fig. 3(b)). This applies similarly to the peaks in the tartrazine spectrum between 7.3 and 8.2 ppm, or in sandalwood between 6.3 and 9.5 ppm. Thus, the presence of these adulterants in a sample does not affect their scores on the first two PC axes sufficiently to place them outside the authentic class boundary. This shows that classification based on a subset of two PC scores alone is too simple to detect many kinds of compromised samples. The way forward is to increase the number of scores used and/or to make use of the remaining variance unexplained by the PC subset; this forms the basis of the DD-SIMCA method, discussed in the following section.

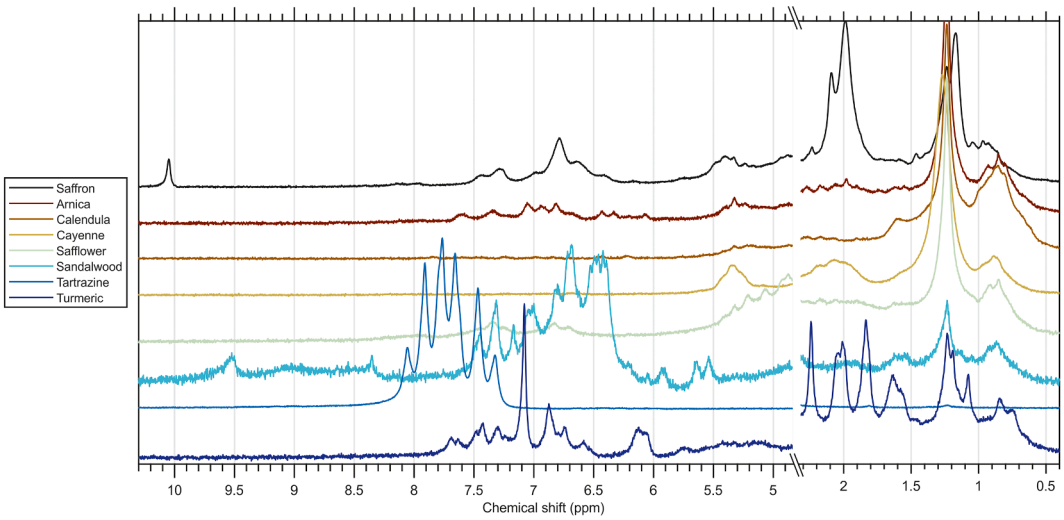


Fig. 3. Spectra of an extract of saffron (uppermost black trace) compared with those of a range of potential adulterants, shown with a vertical offset for clarity. The large methylene peak from fatty acids at 1.23 ppm is seen in the extracts from other plant materials and is the dominant feature in arnica, calendula, cayenne and safflower. The spectra are shown here after SNV-correction of the region of interest only (i.e., solvent region excised). Extraction efficiency varies markedly with substance, and this is responsible for the clear differences in spectral signal-to-noise.

3.2. One-class classification

Three different methods, DD-SIMCA, OCC-NN and IF, were used to carry out one-class classification. SIMCA and its many variants have been widely used for spectral analysis, including in authenticity studies (Horn, Esslinger, Fauhl-Hassek, & Riedl, 2021; Milani, Rossini, Catelani, Pezza, Toci, & Pezza, 2020). OCC-NN has also previously been used to analyse benchtop NMR data in an authenticity context (Gunning et al., 2022). There are no reported applications of IF to NMR spectra or food authenticity issues, but it is gaining traction as a fast, robust method for treated one-class problems in other fields (Alonso-Sarria, Valdivieso-Ros, & Gomariz-Castillo, 2019; Song, Aryal, Ting, Liu, & He, 2022).

As a well-established method, DD-SIMCA was used in the present work to provide a benchmark comparator. Making use of the full spectral range (excluding the solvent peak regions, as discussed above), a DD-SIMCA model was trained using the spectra of the authentic saffron extracts, using resampled k-fold cross-validation to determine the optimal number of PC dimensions (see Supplementary Fig. 5). This model tuning step is important, as SIMCA approaches are known to be highly sensitive to the choice of model dimensionality (Davies & Fearn, 2008). OCC-NN and IF models were likewise trained using the spectra of the authentic saffron extracts. Both these methods make use of machine learning principles, and as such, the final accept/reject decision for an observation is reported as the outcome of a majority vote by an ensemble of learners. For both methods, 300 learners were found to give stable outcomes (see Supplementary Fig. 6).

Shown in Fig. 4(a) are the results obtained from applying the one-class models to the mixture series spectra. All the adulterant types cause the mixtures to be flagged as anomalous at some concentration level, which varies with the modelling method and the material. All models report mixtures with tartrazine as outliers, even from the lowest level investigated of 5 %w/w. Unlike the other adulterants, this is a single compound added directly at the sample preparation stage. Mixtures with calendula are also flagged by DD-SIMCA and OCC-NN at all concentrations investigated, although only from 20 %w/w upwards by IF. For the remaining materials, the lowest adulterant level detected is $\geq$ 20 %w/w, depending on the substance and modelling method. Differing extraction efficiencies are likely responsible for much of the variation in outcomes. Notably, turmeric is not consistently detected by all methods at levels less than 40 %w/w; this is not as good as achieved by Musio et al (2022) who took advantage of the greater sensitivity of 400 MHz NMR. Indeed, the detection limits demonstrated here are clearly insufficient for spotting low level contamination. However, fraudulent substitution for economic gain likely involves substantial amounts of foreign material, so the approach may yet offer promise for broad, untargeted screening.

Focusing now on differences in the modelling results, Fig. 4(b) presents a confusion matrix for each method. These summarise the outcomes for the mixture series and authentic samples (the latter from the training phase validation folds only) and are useful for making a direct comparison of model performances on the exact same set of observations. OCC-NN has the best performance overall, with a type I error rate (2 out of 53) consistent with the contamination fraction hyperparameter, and the highest sensitivity (26 out of 36 mixtures correctly flagged as outliers; note, though, that the sensitivity value is meaningful only for the sample types under consideration, not a potentially much wider outlier class). This contrasts with DD-SIMCA and IF, both of which flag 3 authentic samples and 25 mixtures as outliers, although close examination of Fig. 4(a) and Supplementary Figs. 5 and 6 shows that the misclassified samples are not exactly the same by each method.

These performance differences are likely due to the underlying assumptions about the target class structure, which differ considerably between methods: DD-SIMCA employs a classical χ^2 test to define the class boundary; by focussing only on local inter-observation distances, OCC-NN allows for an arbitrarily shaped class; and IF makes no assumptions about the target class size or shape at all, concerning itself only with those observations found to be extremes. However, the performance variations overall are not large, and are inevitably somewhat exaggerated by the use of hard decision boundaries. The correspondence between methods can be explored by making a more nuanced comparison; this is illustrated for the mixture samples in Supplementary Fig. 7. The agreement between outcomes is also evident from the analysis of the survey samples, discussed in the next section.

3.3. Survey of samples purchased online

Turning now to the online purchased survey samples, eight of the 33 unique packs were identified as outliers by all one-class modelling methods (Fig. 4(c)). Samples 7, 17 and 30 were repeat purchases from brand 7 made at different timepoints spanning nearly two years, suggesting a long-term integrity issue with this supplier. Survey samples 3, 15, 16 and 28 all contained sufficient material for repeat extractions,

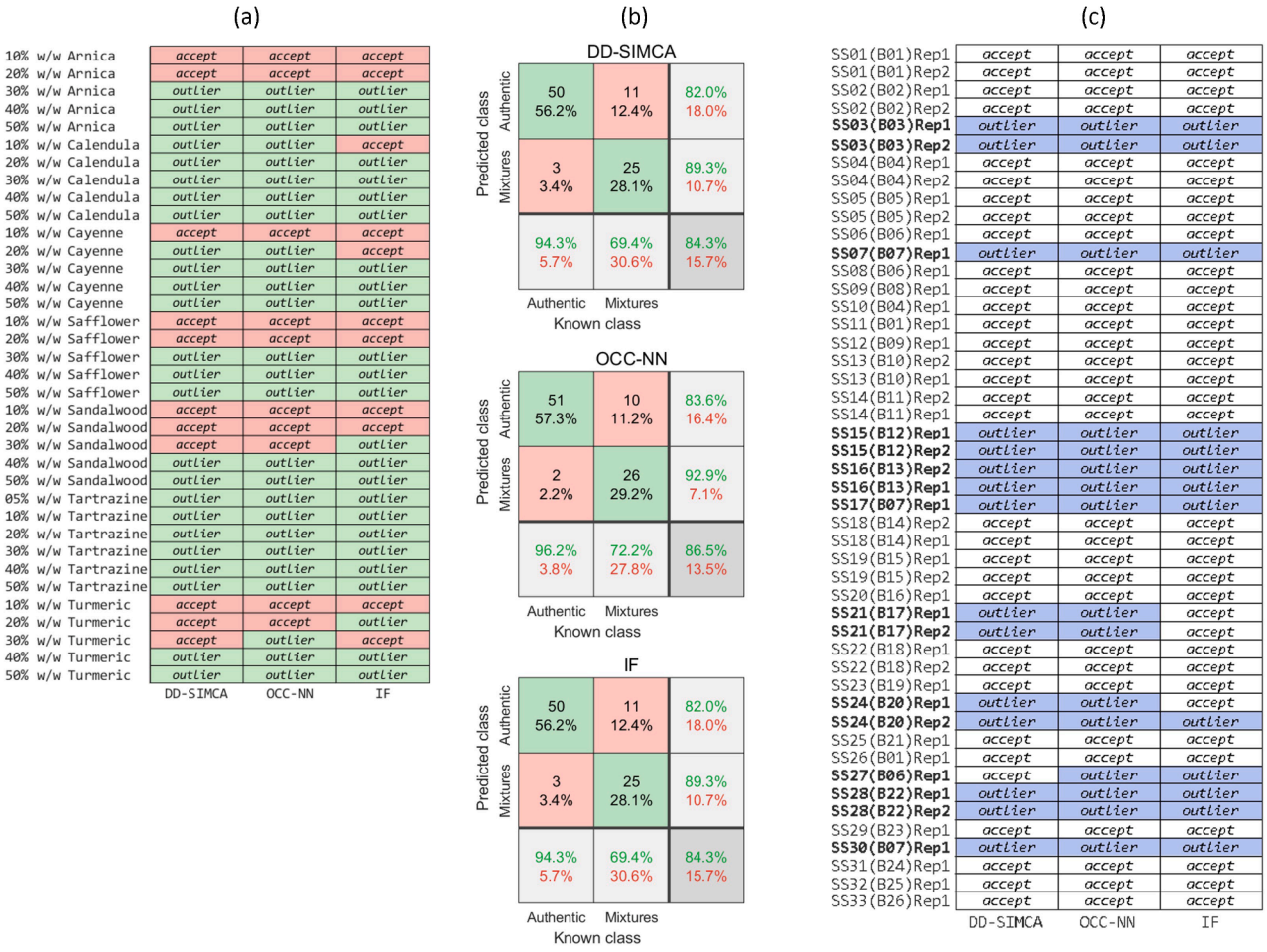


Fig. 4. Panel (a) shows the outcomes by each of the modelling methods for the series of mixtures with known amounts of different potential adulterants. Green cells indicate samples correctly flagged as outliers; red cells indicate samples incorrectly accepted as authentic. Panel (b) brings in the results from the training phase to produce a confusion matrix for each method, which summarises the correct and incorrect classifications for both the authentic and mixture samples. The upper and lower percentages in the bottom left cells in each case are respectively the specificity and the type I error rate (true negative and false positive rates); the latter is seen to be broadly consistent with the contamination fraction hyperparameter for all methods. Panel (c) shows the outcomes by each modelling method for the survey samples. Rows in the table are labelled with the sample code, the brand code in brackets, and the extraction replicate (chronological, run 1 or 2). Samples flagged as outliers are indicated by blue cells and bold text sample codes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

which were carried out immediately upon purchase and again after a year in storage; in all cases, the spectra from both analyses were flagged as anomalous. One of the extracts from a further sample, 24, was found anomalous by all methods, but the other only by DD-SIMCA and OCC-NN. A further two samples, 21 and 27, were each flagged as anomalous by only two of the three methods. All the suspicious samples were purchased as whole saffron strands. Nothing unusual was noticed about them upon inspection, although the samples from brand 7 were somewhat harder to grind and produced noticeably paler extracts than all other samples (see Supplementary Fig. 8).

Visual examination of the spectra found that in many of the suspicious cases, the internal ratio of the fatty acid peak at 1.23 ppm to any of the characteristic saffron peaks (e.g from picrocrocin at 1.16 & 1.18 ppm) is significantly higher than for authentic samples. This can be discerned in the heatmap of the complete spectral collection of saffron (authentic and survey) shown in Fig. 5. Each row corresponds to a spectrum, identified by the labels on the right-hand side; the ten anomalous samples discussed above are indicated by bold text. Hierarchical clustering applied to this spectral collection found that the six samples with the highest 'outlier scores' by one-class modelling (see Supplementary Fig. 9), along with sample 21, formed clusters that were well-separated from other samples. This correspondence across the various data analysis approaches strongly suggests that these seven samples are compromised in some way. Further, from the comparison of their outlier scores with those of the mixture series (see Supplementary Fig. 7), it seems likely that significant amounts of foreign material(s) are involved.

The two extracts from sample 24 formed a small isolated cluster; unusually, these had atypically large ratios of the picrocrocin:fatty acids. In terms of authenticity, it is not clear what this type of anomalous composition might indicate. The remaining suspicious samples, 27 and 28, are found to cluster together with other survey samples accepted as authentic. Along with their outlier scores, which are amongst of the lowest of the flagged samples (see Supplementary Fig. 9), this suggests it is reasonable to treat these sample as type I errors, which should be expected at a rate consistent with the contamination fraction used in the model training.

In the 60 MHz spectra, no obvious unexpected peaks were observed for any of the anomalous samples. However, when 600 MHz spectra were acquired from these samples, it became clear that they were anomalous in more than just the picrocrocin:fatty acid ratio. An illustrative example is given in Fig. 6 for the spectra from survey sample 3. Fig. 6(a) compares the 60 MHz spectra of the duplicate extracts with the mean of the authentic saffron spectra. Other than the internal ratio

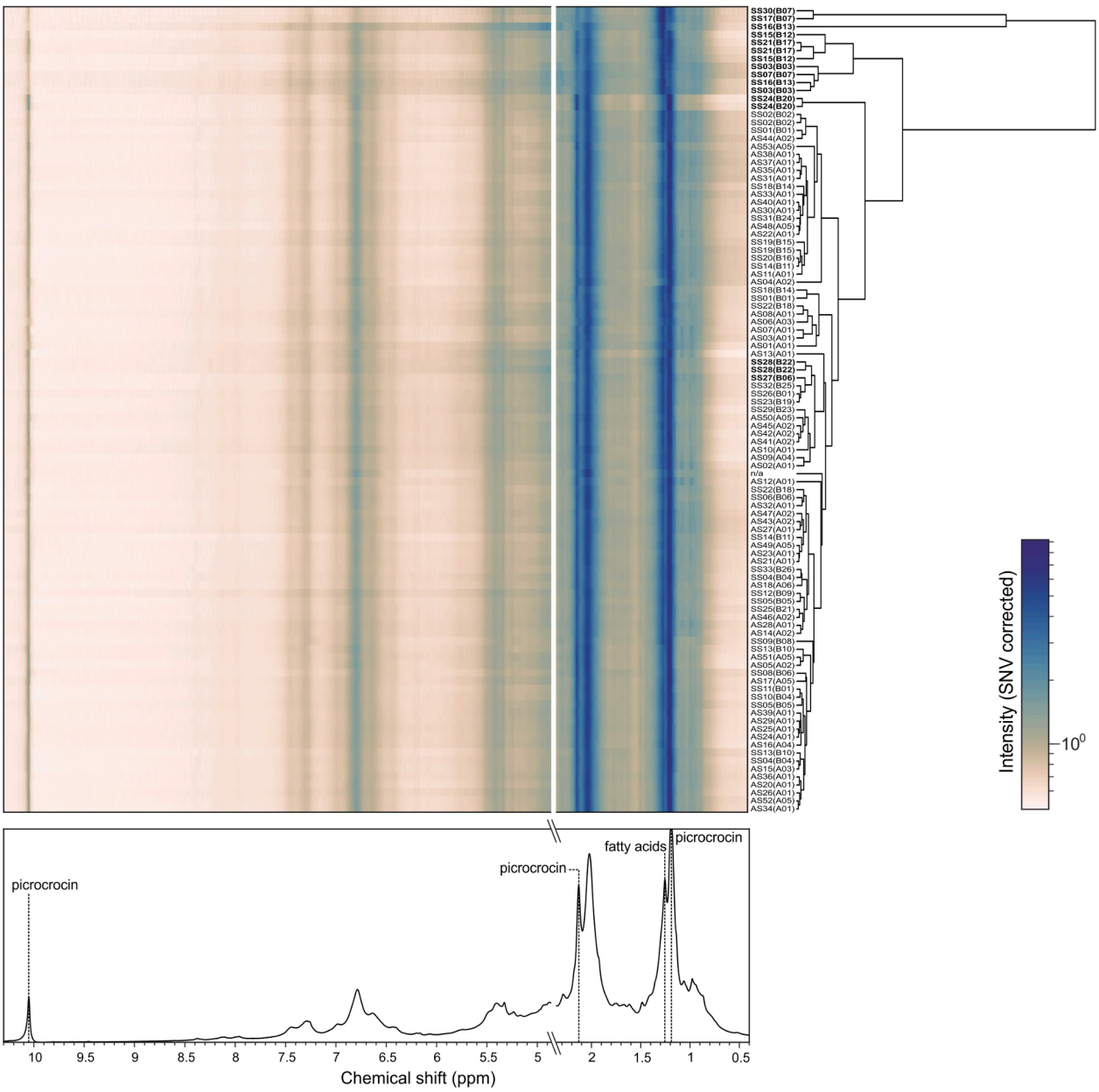


Fig. 5. The main panel shows a heatmap of the complete spectral collection of saffron samples (authentic and survey). To aid interpretation, the mean spectrum of just the authentic saffron collection is shown in the panel below. A hierarchical cluster tree (correlation distances, unweighted average distance) was calculated from the spectra and is shown on the right-hand side. Labels in bold face indicate those samples found anomalous by one class modelling.

difference already noted, the only other hint of a spectral profile difference occurs at ~ 2.2 ppm, although at 60 MHz this is somewhat close to the solvent signal to be certain. However, in Fig. 6(b) the 600 MHz spectrum of the same survey sample is compared with that of a typical authentic saffron; several multiplets can be seen which are not expected in saffron, including a possible doublet in the region ~ 2.2 ppm (marked 'A' on the figure). This confirms the presence of an unidentified substance due to probable adulteration of the saffron sample. Establishing the identity of these unknown material(s) requires detailed forensic analysis and will be the subject of future work.

4. Conclusions

Using a simple extraction procedure, samples of saffron can be successfully analysed using 60 MHz benchtop NMR, yielding spectra with clear features attributable to the characteristic secondary metabolites of picrocrocin and crocins. The sample preparation method is low-cost, low-risk and suitable for routine use as an authenticity screen. Spectra were obtained from a collection of saffron samples from trusted UK suppliers, along with serial mixtures of some of these samples with known possible adulterants of saffron; these were used to demonstrate the potential of the screening method to detect foreign substances.

To provide an objective means of detecting anomalous samples, one-class modelling was carried out using three algorithmically very different approaches, DD-SIMCA, OCC-NN and IF. The ensemble methods offer the advantages of a straightforward training phase with inherent protection against overfitting and only one main hyper-parameter, the contamination fraction, which sets the type I error rate. Models were trained on the saffron samples from trusted suppliers, which formed a well-behaved target class. The performance of all approaches was found to be very similar, with OCC-NN giving the best sensitivity and specificity, although only by a small margin. All models

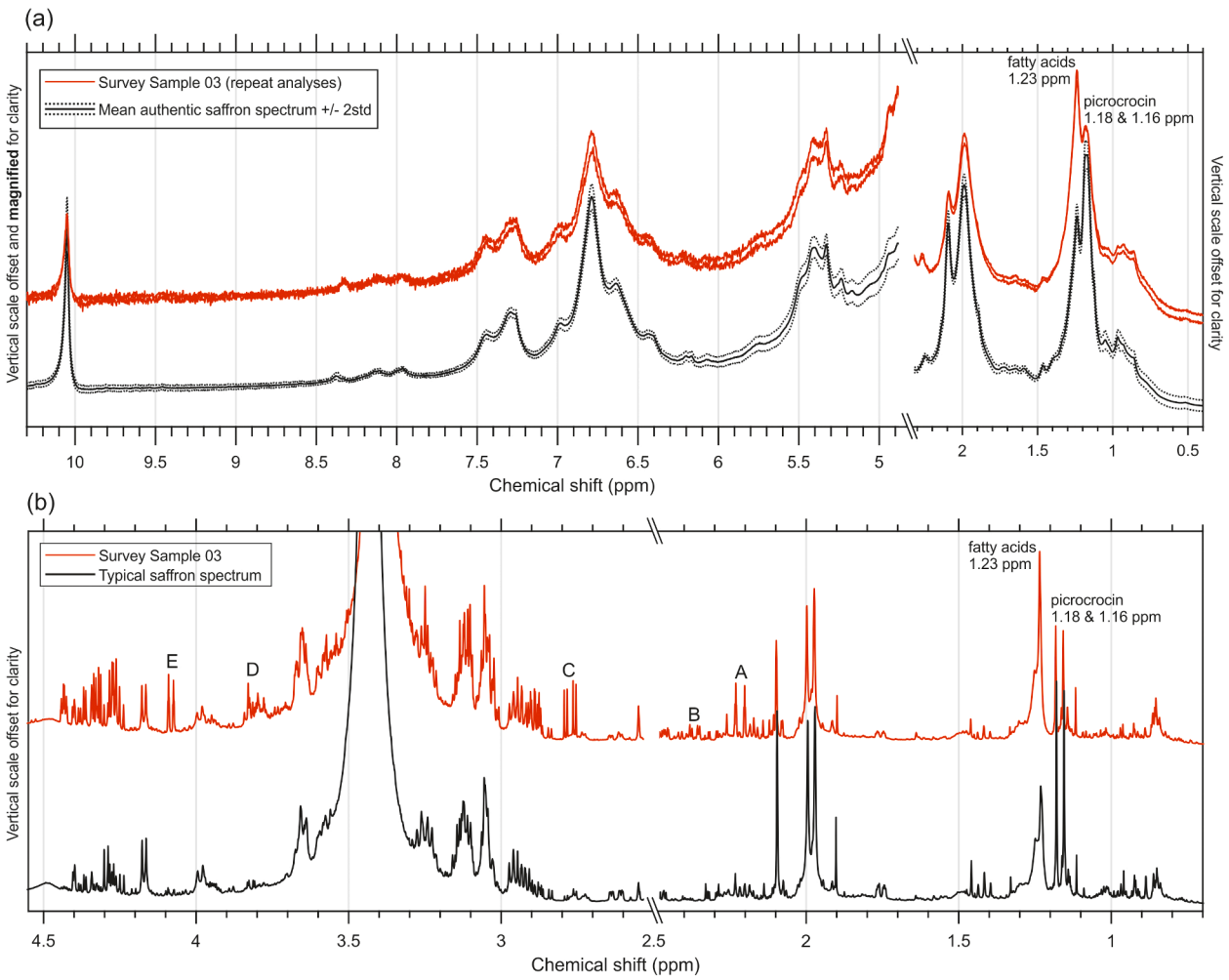


Fig. 6. Spectra from survey sample 3 acquired (a) at 60 MHz field strength, where duplicate analyses are compared with the mean authentic saffron spectrum +/- the standard deviation to indicate variation across the collection; and (b) at 600 MHz field strength, where the comparison is with a typical spectrum from the authentic collection. In both cases, the most obvious difference is in the ratio of the picrocrocin peak at 1.18 ppm to the neighbouring fatty acid peak at 1.23 pm; this is always > 1 for authentic saffron. It is substantially lower in the survey sample, indicating proportionally a much lower picrocrocin content. In the 600 MHz spectrum, additional unexplained peaks and multiplets are also seen, e.g. as marked at A, B, C, D and E.

were able to flag as anomalous the serial mixtures, at concentration levels that varied with the adulterant substance. Tartrazine had the lowest detection limit of 5 %w/w, whereas some of the dried plant materials needed 30 %w/w to be flagged as suspicious. These detection limits are well above contamination levels, but arguably are typical of fraudulent substitution for economic gain.

Using the same approach, a survey was conducted of 33 packs of saffron purchased from the international online marketplace from 26 individual brands. Seven of the packs were found to be anomalous by application of the one-class models, corroborated by hierarchical cluster analysis; these are believed to be compositionally compromised in some way. Since the benchtop NMR method is not sensitive enough to detect very low levels of adulteration, it is likely that these samples contain quite high levels of foreign material, beyond what might be commensurate with adventitious contamination. From sampling theory, we estimate the prevalence of fraud in the sector to be between 12 and 35 % (see Supplementary Fig. 10), similar to the rate found in the study by Torelli et al (2014). Notably, three of the survey samples were repeat purchases of a single brand made years apart. This suggests ongoing, deliberate fraud by a bad actor in the supply chain rather than an isolated incident.

Examination of the 60 MHz NMR spectra showed that most samples flagged as suspicious had unusual internal ratios of the picrocrocin:fatty acid peaks. This was confirmed by 600 MHz NMR analysis, which also revealed additional peaks in the spectrum not expected in authentic saffron. Determining the exact nature of the adulterant(s) is beyond the scope of the present study and will be the subject of future investigation.

CRediT authorship contribution statement

Yvonne Gunning: Methodology, Investigation, Supervision, Writing – review & editing. **Kate S. Davies:** Investigation. **E. Kate Kemsley:** Formal analysis, Software Conceptualization, Writing – original draft.

Declaration of Competing Interest

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.

Data availability

Code for carrying out the data analysis method of interest is supplied in the supplementary information.

Acknowledgments

The authors gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); this research was funded by the BBSRC Core Capability Grant number BB/CCG1860/1. We would also like to thank our colleagues in the Core Science Resources group at Quadram Institute Bioscience for their help in cultivating saffron plants for use in this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2022.134649.

References

Alonso-Sarria, F., Valdivieso-Ros, C., & Gomariz-Castillo, F. (2019). Isolation Forests to Evaluate Class Separability and the Representativeness of Training and Validation Areas in Land Cover Classification. *Remote Sensing, 11*(24). https://doi.org/10.3390/rs11243000

Amirvaresi, A., Nikounezhad, N., Amirahmadi, M., Daraei, B., & Parastar, H. (2021). Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy based on chemometrics for saffron authentication and adulteration detection. *Food Chemistry, 344*, 9. https://doi.org/10.1016/j.foodchem.2020.128647

Antonides, L. H., Brignall, R. M., Costello, A., Ellison, J., Firth, S. E., Gilbert, N., … Mewis, R. E. (2019). Rapid Identification of Novel Psychoactive and Other Controlled Substances Using Low-Field(1)H NMR Spectroscopy. *Acs Omega, 4*(4), 7103–7112. https://doi.org/10.1021/acsomega.9b00302

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy, 43*(5), 772–777. https://doi.org/10.1366/0003702894202201

Bostan, H. B., Mehri, S., & Hosseinzadeh, H. (2017). Toxicology effects of saffron and its constituents: A review. *Iranian Journal of Basic Medical Sciences, 20*(2), 110–121. https://doi.org/10.22038/ijbms.2017.8230

Cagliani, L. R., Culeddu, N., Chessa, M., & Consonni, R. (2015). NMR investigations for a quality assessment of Italian PDO saffron (Crocus sativus L.). *Food Control, 50*, 342–348. https://doi.org/10.1016/j.foodcont.2014.09.017

Christodoulou, E., Kadoglou, N. P. E., Kostomitsopoulos, N., & Valsami, G. (2015). Saffron: A natural product with potential pharmaceutical applications. *Journal of Pharmacy and Pharmacology, 67*(12), 1634–1649. https://doi.org/10.1111/jphp.12456

Consonni, R., Ordoudi, S. A., Cagliani, L. R., Tsiangali, M., & Tsimidou, M. Z. (2016). On the Traceability of Commercial Saffron Samples Using H-1-NMR and FT-IR Metabolomics. *Molecules, 21*(3), 13. https://doi.org/10.3390/molecules21030286

Davies, A. M. C., & Fearn, T. (2008). Back to basics: Multivariate qualitative analysis, "SIMCA". *Spectroscopy Europe, 20*, 15–19. https://doi.org/10.1255/sew.2008.a1

Dowlatabadi, R., Farshidfar, F., Zare, Z., Pirali, M., Rabiei, M., Khoshayand, M. R., & Vogel, H. J. (2017). Detection of adulteration in Iranian saffron samples by H-1 NMR spectroscopy and multivariate data analysis techniques. *Metabolomics, 13*(2), 11. https://doi.org/10.1007/s11306-016-1155-x

Gunning, Y., Defernez, M., Watson, A. D., Beadman, N., Colquhoun, I. J., Le Gall, G., & Kemsley, E. K. (2018). 16-O-methylcafestol is present in ground roast Arabica coffees: Implications for authenticity testing. *Food Chemistry, 248*, 52–60. https://doi.org/10.1016/j.foodchem.2017.12.034

Gunning, Y., Taous, F., El Ghali, T., Gibbon, J. D., Wilson, E., Brignall, R. M., & Kemsley, E. K. (2022). Mitigating instrument effects in 60 MHz H-1 NMR spectroscopy for authenticity screening of edible oils. *Food Chemistry, 370*. https://doi.org/10.1016/j.foodchem.2021.131333

Horn, B., Esslinger, S., Fauhl-Hassek, C., & Riedl, J. (2021). H-1 NMR spectroscopy, one-class classification and outlier diagnosis: A powerful combination for adulteration detection in paprika powder. *Food Control, 128*. https://doi.org/10.1016/j.foodcont.2021.108205

Hosseinzadeh, H., & Nassiri-Asl, M. (2013). Avicenna's (Ibn Sina) the Canon of Medicine and Saffron (Crocus sativus): A Review. *Phytotherapy Research, 27*(4), 475–483. https://doi.org/10.1002/ptr.4784

Kumari, L., Jaiswal, P., & Tripathy, S. S. (2021). Various techniques useful for determination of adulterants in valuable saffron: A review. *Trends in Food Science & Technology, 111*, 301–321. https://doi.org/10.1016/j.tifs.2021.02.061

Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation-Based Anomaly Detection. *ACM Transactions on Knowledge Discovery from Data, 6*(1), 39. https://doi.org/10.1145/2133360.2133363

Milani, M. I., Rossini, E. L., Catelani, T. A., Pezza, L., Toci, A. T., & Pezza, H. R. (2020). Authentication of roasted and ground coffee samples containing multiple adulterants using NMR and a chemometric approach. *Food Control, 112*. https://doi.org/10.1016/j.foodcont.2020.107104

Modaghegh, M. H., Shahabian, M., Esmaelli, H. A., Rajbai, O., & Hosseinzadeh, H. (2008). Safety evaluation of saffron (Crocus sativus) tablets in healthy volunteers. *Phytomedicine, 15*(12), 1032–1037. https://doi.org/10.1016/j.phymed.2008.06.003

Moshiri, M., Vahabzadeh, M., & Hosseinzadeh, H. (2015). Clinical Applications of Saffron (Crocus sativus) and its Constituents: A Review. *Drug Research, 65*(6), 287–295. https://doi.org/10.1055/s-0034-1375681

Musio, B., Todisco, S., Antonicelli, M., Garino, C., Arlorio, M., Mastrorilli, P., … Gallo, V. (2022). Non-Targeted NMR Method to Assess the Authenticity of Saffron and Trace the Agronomic Practices Applied for Its Production. *Applied Sciences-Basel, 12*(5), 15. https://doi.org/10.3390/app12052583

Mykhailenko, O., Kovalyov, V., Goryacha, O., Ivanauskas, L., & Georgiyants, V. (2019). Biologically active compounds and pharmacological activities of species of the genus Crocus: A review. *Phytochemistry, 162*, 56–89. https://doi.org/10.1016/j.phytochem.2019.02.004

Negi, A., Pare, A., & Meenatchi, R. (2021). Emerging techniques for adulterant authentication in spices and spice products. *Food Control, 127*, 14. https://doi.org/10.1016/j.foodcont.2021.108113

Ordoudi, S. A., Pascual, M. D., & Tsimidou, M. Z. (2014). On the quality control of traded saffron by means of transmission Fourier-transform mid-infrared (FT-MIR) spectroscopy and chemometrics. *Food Chemistry, 150*, 414–421. https://doi.org/10.1016/j.foodchem.2013.11.014

Ordoudi, S. A., Cagliani, L. R., Lalou, S., Naziri, E., Tsimidou, M. Z., & Consonni, R. (2015). H-1 NMR-based metabolomics of saffron reveals markers for its quality deterioration. *Food Research International, 70*, 1–6. https://doi.org/10.1016/j.foodres.2015.01.021

Pages, G., Gerdova, A., Williamson, D., Gilard, V., Martino, R., & Malet-Martino, M. (2014). Evaluation of a Benchtop Cryogen-Free Low-Field H-1 NMR Spectrometer for the Analysis of Sexual Enhancement and Weight Loss Dietary Supplements Adulterated with Pharmaceutical Substances. *Analytical Chemistry, 86*(23), 11897–11904. https://doi.org/10.1021/ac503699u

Petrakis, E. A., Cagliani, L. R., Polissiou, M. G., & Consonni, R. (2015). Evaluation of saffron (Crocus sativus L.) adulteration with plant adulterants by H-1 NMR metabolite fingerprinting. *Food Chemistry, 173*, 890–896. https://doi.org/10.1016/j.foodchem.2014.10.107

Petrakis, E. A., Cagliani, L. R., Tarantilis, P. A., Polissiou, M. G., & Consonni, R. (2017). Sudan dyes in adulterated saffron (Crocus sativus L.): Identification and quantification by H-1 NMR. *Food Chemistry, 217*, 418–424. https://doi.org/10.1016/j.foodchem.2016.08.078

Petrakis, E. A., & Polissiou, M. G. (2017). Assessing saffron (Crocus sativus L.) adulteration with plant-derived adulterants by diffuse reflectance infrared Fourier transform spectroscopy coupled with chemometrics. *Talanta, 162*, 558–566. https://doi.org/10.1016/j.talanta.2016.10.072

Rubert, J., Lacina, O., Zachariasova, M., & Hajslova, J. (2016). Saffron authentication based on liquid chromatography high resolution tandem mass spectrometry and multivariate data analysis. *Food Chemistry, 204*, 201–209. https://doi.org/10.1016/j.foodchem.2016.01.003

Santoyo-Ramon, J. A., Casilari, E., & Cano-Garcia, J. M. (2021). A Study of One-Class Classification Algorithms for Wearable Fall Sensors. *Biosensors-Basel, 11*(8), 20. https://doi.org/10.3390/bios11080284

Seliya, N., Zadeh, A. A., & Khoshgoftaar, T. M. (2021). A literature review on one-class classification and its potential applications in big data. *Journal of Big Data, 8*(1). https://doi.org/10.1186/s40537-021-00514-x

Song, X. Y., Aryal, S., Ting, K. M., Liu, Z., & He, B. (2022). Spectral-Spatial Anomaly Detection of Hyperspectral Data Based on Improved Isolation Forest. *IEEE Transactions on Geoscience and Remote Sensing, 60*. https://doi.org/10.1109/tgrs.2021.3104998

Spence, L., & Bourlakis, M. (2009). The evolution from corporate social responsibility to supply chain responsibility: The case of Waitrose. *Supply Chain Management-an International Journal, 14*(4), 291–302. https://doi.org/10.1108/13598540910970126

Tang, B., & He, H. B. (2017). A local density-based approach for outlier detection. *Neurocomputing, 241*, 171–180. https://doi.org/10.1016/j.neucom.2017.02.039

Torelli, A., Marieschi, M., & Bruni, R. (2014). Authentication of saffron (Crocus sativus L.) in different processed, retail products by means of SCAR markers. *Food Control, 36*(1), 126–131. https://doi.org/10.1016/j.foodcont.2013.08.001

Wold, S. (1976). Pattern recognition by means of disjoint principal components models. *Pattern Recognition, 8*(3), 127–137.

Wu, N., Balayssac, S., Assemat, G., Danoun, S., Dejean, S., Malet-Martino, M., & Gilard, V. (2021). Evaluation of low-field versus high-field proton NMR spectroscopy for quality control of cinnamon samples. *Journal of Food Composition and Analysis, 96*. https://doi.org/10.1016/j.jfca.2020.103706

Yilmaz, A., Nyberg, N. T., Molgaard, P., Asili, J., & Jaroszewski, J. W. (2010). H-1 NMR metabolic fingerprinting of saffron extracts. *Metabolomics, 6*(4), 511–517. https://doi.org/10.1007/s11306-010-0221-z

Zontov, Y. V., Rodionova, O. Y., Kucheryavskiy, S. V., & Pomerantsev, A. L. (2017). DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach. *Chemometrics and Intelligent Laboratory Systems, 167*, 23–28. https://doi.org/10.1016/j.chemolab.2017.05.010