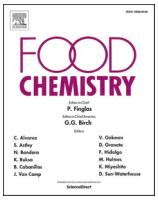
Food Chemistry 370 (2022) 131333

Contents lists available at ScienceDirect

Food Chemistry

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



Analytical Methods

Mitigating instrument effects in 60 MHz ^{1}H NMR spectroscopy for authenticity screening of edible oils

Yvonne Gunning [a], Fouad Taous [b], Tibari El Ghali [b], James D. Gibbon [c], E. Wilson [a], Rachel M. Brignall [c], E. Kate Kemsley [a,*]

[a] Quadram Institute Bioscience, Norwich Research Park, Colney, Norwich NR4 7UQ, UK
[b] Centre National de l'Energie des Sciences et des Techniques Nucléaires (CNESTEN) Rabat, Morocco
[c] Oxford Instruments, Tubney Woods, Oxford OX13 5QX, UK

ARTICLE INFO

Keywords:
Edible oil
Argan oil
Olive oil
Authenticity
Chemometrics
^{1}H NMR
Benchtop NMR spectroscopy

ABSTRACT

Low field (60 MHz) ^{1}H NMR spectroscopy was used to analyse a large (n = 410) collection of edible oils, including olive and argan, in an authenticity screening scenario. Experimental work was carried out on multiple spectrometers at two different laboratories, aiming to explore multivariate model stability and transfer between instruments. Three modelling methods were employed: Partial Least Squares Discriminant Analysis, Random Forests, and a One Class Classification approach. Clear inter-instrument differences were observed between replicated data collections, sufficient to compromise effective transfer of models based on raw data between instruments. As mitigations to this issue, various data pre-treatments were investigated: Piecewise Direct Standardisation, Standard Normal Variates, and Rank Transformation. Datasets comprised both phase corrected and magnitude spectra, and it was found that that the latter spectral form may offer some advantages in the context of pattern recognition and classification modelling, particularly when used in combination with the Rank Transformation pre-treatment.

1. Introduction

With a long history of use by research chemists, high-field NMR spectroscopy has been a gold standard in analysis for decades, although it is largely beyond the reach of all but high-end labs and large corporations. The arrival of 'benchtop' spectrometers in recent years has changed this landscape. Based on permanent rather than superconducting magnets, benchtop instruments work at much lower field strengths, typically in the range 40–100 MHz. They do not need any cryogens, are less complex to use and have a smaller physical footprint, with commensurate reductions in capital and maintenance costs. This has made the diagnostic strengths of NMR spectroscopy accessible to a far wider range of potential users.

Within the food sector, benchtop NMR is gaining traction in applications including quality control and authentication. In a recent review, van Beek (2021) discusses a range of these, also noting that chemometric handling of benchtop NMR data is advantageous. Its transformative potential is likewise noted in a review of targeted and untargeted NMR spectroscopy for food authenticity applications (Sobolev et al., 2019). Several of the studies discussed involve the analysis of oils and fats. This compound class is well-suited for analysis by benchtop NMR; indeed, the first published application in foods was on distinguishing olive from hazelnut oils (Parker, Limer, Watson, Defernez, Williamson, & Kemsley, 2014).

There is an extensive literature on authenticating edible oils using wider analytical techniques, with recent examples based on mass spectrometry (Quintanilla-Casas et al., 2021), infrared spectroscopy (Sota-Uba, Bamidele, Moulton, Booksh, & Lavine, 2021), chromatography (Jimenez-Carvelo, Perez-Castano, Gonzalez-Casado, & Cuadros-Rodriguez, 2017), NMR relaxometry (Ancora et al., 2021) and high-resolution NMR spectroscopy (Tang, Green, Wang, & Hatzakis, 2021). This reflects the economic importance of these products, their vulnerability to substitution or adulteration, and the difficulty in detecting such compromises. Since edible oils are largely (>95% w/w) composed of triglycerides, many of these methods are based on distinguishing the composition profile of the fatty acid chains in the product of interest from those of potential adulterants, in the presence of considerable natural variation. However, most published authenticity studies, using

* Corresponding author.
E-mail address: kate.kemsley@quadram.ac.uk (E.K. Kemsley).

https://doi.org/10.1016/j.foodchem.2021.131333
Received 30 April 2021; Received in revised form 1 October 2021; Accepted 3 October 2021
Available online 5 October 2021

whichever technique, are conducted on a single instrument at one laboratory. There are fewer reports of ring trials in which the results from different laboratories are compared, as for example in Köppel et al. (2020) and Velasco et al. (2021). An equally pertinent issue is the transfer of databases and accompanying data analysis to different instruments (Feundale, Woody, Tan, Myles, Brown, & Ferre, 2002). Such considerations need to be addressed in the deployment of a real-world analytical method (McVey, McGrath, Haughey, & Elliott, 2021). A major effort in setting up authenticity screening is obtaining the requisite collection of reference samples of known provenance. This is time-consuming and expensive, and the resulting data collection is thus a valuable resource. Ideally, such a reference database should be valid for an extended period of time, remaining useful after instrument refurbishment, or after transfer to a replacement or additional instrument.

Whilst NMR spectroscopy is valued for its inherent specificity, reproducibility and linear response with absorbing species, inter-spectrometer differences exist for a variety of technical and practical reasons (Alam, Alam, McIntyre, Volk, Neerathilingam, & Luxon, 2009). Even within production tolerances, there can be non-trivial variations in magnetic field homogeneity, probe sensitivity, temperature stability, shielding and shims that lead to database porting challenges. It is acknowledged that modelling conducted on data obtained from one spectrometer will not necessarily transfer without adaptation to a new setting (Galvan, Bona, Borsato, Danieli, & Killner, 2020). Further, any adverse impact will likely be greatest on untargeted chemometric approaches which seek to utilise a wide spectral range.

The present work reports a 60 MHz benchtop ^{1}H NMR screening method for authenticating olive and argan oils originating from Morocco, where they are major economic commodities. The development work was carried out in tandem on multiple spectrometers sited at two different laboratories, with the aim of examining the issue of classification model transfer between instruments. Intra- and inter-instrument variance is explored, and mitigations to these issues are considered, including a spectral transformation which to the best of our knowledge has not previously been used in the analysis of NMR spectra. Each authentication scenario is modelled as a two-group classification problem ('authentic' versus 'non-authentic') by two disparate multivariate methods and also as a one-class classification problem. The outcomes on various independent, inter-instrument test sets are compared, looking for robustness of performance in the presence of large systematic variances.

Conventional practice in NMR spectroscopy is to use the real component of the Fourier transformed FID, phase corrected to give a frequency-domain absorbance spectrum, as the primary measurement on a sample; ideally, this has the appearance of sharp Lorentzian peaks rising from a flat baseline. Phase correction may be conducted manually, or by using one of a range of automated algorithms; the latter approach is most appropriate for high-throughput studies, as it much faster and avoids operator subjectivity. The present study also explores the use of magnitude spectra. These are obtained from the real *and* imaginary parts of the Fourier-transformed FIDs and thus relate more directly to the induced magnetization. Although this represents something of a paradigm shift as far as the NMR community is concerned, the benefits of using this spectral form have been recognised (Harrington & Wang, 2017), and are confirmed in the work presented here.

2. Materials and methods

2.1. Samples

410 edible oils of known provenance were used to develop authentication models for screening olive and argan oils. These were provided by CNESTEN (Centre National de l'Energie des Sciences et des Techniques Nucléaires, Morocco), collected as part of a FAO/IAEA (Food and Agriculture Organization of the United Nations, International Atomic Energy Agency) Coordinated Research Project 'Field-deployable Analytical Methods to Assess the Authenticity, Safety and Quality of Food' (project code D52040). The eventual aim of the project is to develop approaches that can be employed by authorities and regulators in developing countries, to promote confidence in the food supply chain and help protect their important export commodities.

The sample types were as follows: olive oil (n = 204), argan oil (n = 120), and mixtures (n = 86) of olive and argan with soya or sunflower oil, with the latter two oils in the range 3 – 100% w/w. The olive and argan oils were obtained directly from producers, including many traditional cooperatives, in various different Moroccan regions. The olive oils were collected in 2018 and 2019, whereas the argan oils were mostly from 2017 and 2018, with a few dating back to 2016 or earlier. The number of samples from each province and the year of collection are listed in Supplementary Table 1, along with further details of oil production and storage.

An additional 10 oils (6 olive, 4 argan) were provided by CNESTEN that, from spectral analysis, were found to be degraded or contaminated. These samples were excluded from the developmental work, but their spectra were retained as examples of samples compromised other than by adulteration with other vegetable oils.

Two historic spectral collections ('Archive 1′ and '2′) were also used in this study to serve as additional inter-instrument test sets. Together these comprise almost 200 spectra from various edible oils and mixtures (detailed in Supplementary Table 2).

2.2. Sample batches and spectrometers

The 'CNESTEN' samples were delivered to Oxford Instruments ('Lab 1′) in two batches ('Batch 1′ and 'Batch 2′). ^{1}H spectra were acquired of all samples using 'Pulsar' benchtop NMR spectrometers (Oxford Instruments, Abingdon, UK) operating at a frequency of 60 MHz. Two instruments from the same hardware series were used, one for each batch.

After spectral acquisition at Lab 1, the samples were transferred to Quadram Institute Bioscience ('Lab 2′), where both batches were analysed using an earlier hardware version of the Pulsar. Details of the samples in each batch, dates of spectral analysis, and primary instrument settings are given in Table 1.

Archive 1 spectra were acquired on the same Lab 2 Pulsar as the

Table 1
CNESTEN samples.

Oil type	Number	Numbers delivered in each batch:	
Olive	204	Batch 1:	0
		Batch 2:	204
Argan	120	Batch 1:	43
		Batch 2:	77
Other *(3 – 90 %w/w soya in olive; 3 – 90 %w/w sunflower in olive; 3 – 90 %w/w soya in argan; 3 – 90 %w/w sunflower in argan; soya; sunflower)*	86	Batch 1:	38
		Batch 2:	48
Spectral acquisition dates:			
	Lab 1*	Lab 2**	
Batch 1	April 2019	June – July 2019	
Batch 2	September 2019	October 2019 – January 2020	

*At Lab 1, two different Pulsars from the same hardware series were used to analyse the CNESTEN samples, one for each Batch. Primary instrument-specific settings optimised during machine setup were the 90° pulse width (P90) and receiver attenuation (RA). For the spectrometer used in Batch 1 analysis, these values were P90 = 10.94 μs, RA = 33 dB, and for that used in Batch 2 analysis, P90 = 13.60 μs and RA = 34 dB.

**At Lab 2, the analysis of all CNESTEN samples was carried out on one Pulsar only. This was from an earlier hardware series than those used at Lab 1. For this spectrometer, P90 = 13.60 μs and RA = 30 dB.

CNESTEN samples, but three years earlier and before a major upgrade to the instrument's shielding. Archive 2 spectra were collected at Lab 2 in 2013 using a further spectrometer, a first-generation Pulsar.

2.3. Spectral acquisition

For all acquisitions, 0.6 ml of sample was pipetted into a standard disposable 5 mm NMR tube, with no other preparation step. For each spectrum, 32 FIDs were acquired, recording 32,768 data points over a 5000 Hz window (dwell time = 0.2 ms) with a relaxation delay of 2 s. Signal averaging of correlation aligned FIDs was conducted within the Spinflow user interface (Oxford Instruments, Abingdon, UK). This does not require the presence of a reference compound or frequency lock. The total acquisition time was less than 5 min, in line with our requirement for a high-throughput analysis. The spectral linewidth was maintained between 0.6 and 0.9 Hz, by daily measurement of the chloroform FWHM in a sealed standard, and shimming when necessary.

2.4. Data analysis

2.4.1. FID post-processing

All data analysis including FID post-processing was carried out using Matlab (The Mathworks Inc, Cambridge, UK) using bespoke scripts and built-in functions from the Statistics and Machine Learning Toolbox. Following acquisition, the signal-averaged FIDs were post-processed using linear back prediction and Fourier transformed into the frequency domain. The magnitude spectrum from each sample was retained. Phase corrected spectra were also generated using an automated script written in-house. The chemical shift scale was referenced to the glyceride peak which is found in all triglyceride spectra at 4.2 ppm. All spectra were truncated to the region 0–10.5 ppm, yielding 4267 data points per spectrum to pass into the data exploration and modelling.

2.4.2. Data exploration and authentication models

Inter-batch and inter-instrument variances were explored in the CNESTEN samples using Principal Component Analysis (PCA). This was applied to various subsets of phase corrected and magnitude spectra, raw and pre-treated with the following methods: piecewise direct standardisation ('PDS'); standard normal variate ('SNV') correction; and rank transformation ('RT').

Two unrelated multivariate methods, partial least squares discriminant analysis ('PLSDA') and random forests ('RF'), were used to train two-group ('authentic versus 'non-authentic') classification models. In both approaches, a score is obtained for each item analysed which, along with an appropriately chosen threshold, serves as a binary classifier. A third, previously disclosed (Gunning et al., 2020) one-class classification method was also employed, which obtains a model using the authentic class in each case only, accepting or rejecting items via thresholds on an ensemble of nearest neighbour distances ('NN-OCC').

All classification modelling was carried out separately for the 'positive' classes of olive and argan oils. In each case, the oil type not used as the positive class was considered 'non-authentic' along with the collection of 'other' spectra (see Table 1). To develop models, the CNESTEN samples were partitioned into training and test sets. Lab 1 spectra of the training samples only were used in model development. Technical details on the implementation of each modelling method are given in Supplementary Table 3, including criteria for predictor selection, model optimisation, the cross-validation scheme, and determination of the classifier threshold. Spectra were analysed raw, and after the various pre-treatments detailed above. All models were applied to the Lab 1 and Lab 2 test set spectra, as well as the Archive collections, and the rates of correct classifications compared.

Sample allocations to the training and test partitions were made randomly, subject to the constraints that within each sample type, approximately 2/3 were allocated to the training set and the remainder to the test set; further, the small number of samples for which spectra were not available from both labs were allocated such as to maximise the total size of the training and test sets (i.e. samples with a missing Lab 1 spectrum were placed in the test set, and with a missing Lab 2 spectrum in the training set). The entire partitioning and modelling procedures were repeated 100 times to enable examination of the robustness of classification outcomes with respect to the constitution of the training set.

3. Results and discussion

3.1. Data exploration

In high-throughput experiments involving hundreds of spectral acquisitions, it is likely that the quality of a small proportion will fall short. Examination of individual spectra is important for spotting obviously faulty measurements. In the present study, these have variously arisen from: insufficient sample to properly fill the NMR tube, trapped air bubbles, and loss of shim or other instrumental problems during data acquisition; these spectra were excluded from the subsequent data exploration and modelling work. Despite these issues, all CNESTEN samples were represented in at least one Lab's data collections, with the rejected spectra numbering fewer than 20 at each Lab.

Visual inspection of the large amounts of data in a high-throughput study is a challenge. Overlaid plots of large numbers of spectra are not always informative, especially when the differences between samples are small and the number of spectra is in the hundreds. This is exemplified in Fig. 1, where the spectra of olive and argan oils from Batches and Labs 1 and 2 are compared in a stacked plot. All are typical 60 MHz ^{1}H spectra of edible oils with large resonances arising predominantly from triglycerides, which make up around 95% of the oil composition. Spectral assignments of these features have been made elsewhere (Gerdova et al., 2015). The resonances are broad and overlapped in comparison with measurements made at higher field strengths. Some additional line broadening is caused by the sample viscosity, because the oils have been analysed undiluted in keeping with the desire for a high-throughput protocol with minimal use of solvents. This follows the precedent of a pilot study which demonstrated no analytical advantage from working with oils diluted in a solvent (Gunning et al., 2020).

It is possible to discern systematic differences between oils: for example, the internal ratio of the bis-allylic (2.6–3.0 ppm) to glyceride (4.2–4.4 ppm) peak areas is greater for argan than olive, consistent with the known composition of the two oil types.

Confounding differences can also be seen. The intensity of Lab 2 data is systematically greater than that of Lab 1. This can only be due to differences between the spectrometers involved, as the spectra were collected from replicate samples, using the same acquisition protocol at both sites, and with all FID post-processing carried out using the same software suite. Further, within Lab 1 data only, there is an intensity difference between Batch 1 and Batch 2; this must be caused by the differences between the two spectrometers used, since no equivalent batch effect can be seen within the spectra collected on the single instrument used at Lab 2. However, the Batch 2 spectra of both oil types from Lab 2 do exhibit somewhat more variation, most noticeably on either side of the large methylene resonance at 1.2 ppm. This may reflect some aspect of spectrometer drift during the comparatively longer period over which the data were acquired (see Table 1).

An alternative way of viewing the data is by principal component analysis (PCA), which provides an efficient means of exploring large collections of spectra. In PCA, the focus is on variance in the dataset, making it ideal for examining batch and instrument effects. Scores in the first two principal component dimensions are plotted against one another in Fig. 1(b). Each data point represents a spectrum in the PCA coordinate system, and is coded by oil type, laboratory and batch. The instrument and batch effects are clear, appearing in the first PC dimension and accounting for almost all the dataset variance. It is known that differences between vegetable oils can be detected by

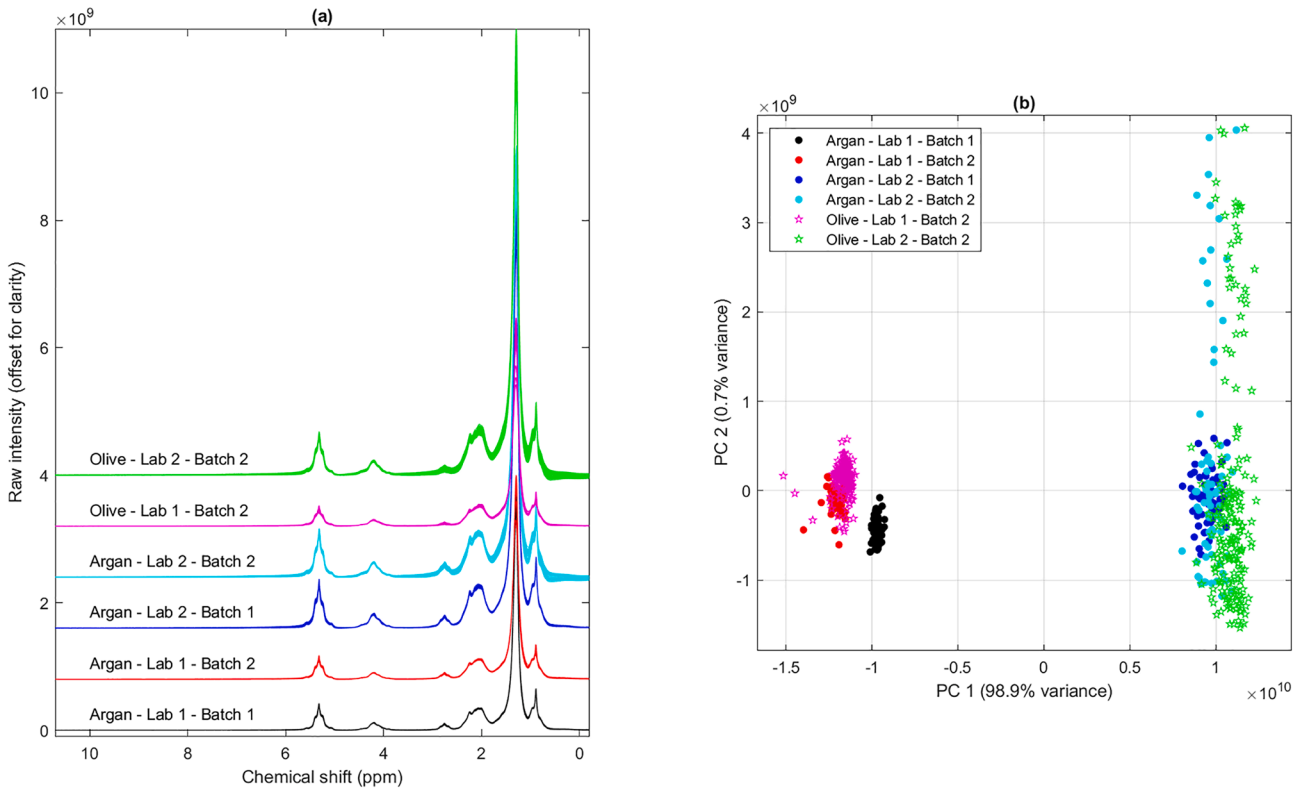


Fig. 1. (a): Stacked plot showing phase corrected spectra of the authentic argan and olive oils Spectra are grouped by oil type, Batch and Lab where acquisition took place. The numbers in each group are detailed in Table 1. (b) PC scores plot with symbols coded by the same groups. The clustering in the data by acquisition Lab is much greater than that by oil type. Further, within Lab 1 data, the clustering of the argan oils by Batch (black vs red filled circles) is greater than the difference between oil type (red filled circles vs magenta stars). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

benchtop NMR, but in this raw dataset, the batch and instruments differences are comparatively greater. The increased variability within Lab 2 Batch 2 data is also evident, appearing as the tail of highly scattered points towards the upper right corner of the plot.

The equivalent figure using the magnitude data is given in the Supplementary Fig. 1. Whilst this dataset is also dominated by the batch and instrument effects, and the PC variance profile is similar, it is noteworthy that the differences between the two oil types are much more clearly visible. The within-group variances are also less, in particular for Lab 2 data. This implies that the greater scatter in Fig. 1(b) is a consequence of phase correction rather than intrinsic measurement error. This is perhaps unsurprising, since universally applicable automated phase correction is notoriously difficult to achieve. Manual phase correction is also problematic due to operator subjectivity, especially in the case of high-throughput datasets. It has been recognised that for pattern recognition applications, working with NMR magnitude spectra instead of phase corrected may confer an advantage (Harrington & Wang, 2017; Williamson, Kemsley, Sutcliffe, & Mewis, 2019). All data analysis in the present work was therefore also applied to magnitude spectra.

3.2. Data pre-treatments

The adverse impact of systematic inter-instrument variances on multivariate modelling is well-recognised in many fields of spectroscopy. As uptake of NMR increases and progressively larger datasets are shared across labs, it is apparent that similar issues affect NMR data. The transfer of calibration models between different NMR instruments or configurations is explored by Alam et al. (2009), who also note that variances due to instrumental factors can compete in size with spectral changes of interest that occur for example in metabonomic studies.

Various techniques have been investigated to mitigate such effects (Fonollosa, Fernandez, Gutierrez-Galvez, Huerta, & Marco, 2016). There are two main strategies: the first is to map the data acquired on one spectrometer onto that from another. The latter instrument is often referred to as the 'primary' and is generally the spectrometer on which a calibration or reference collection was acquired; the former is referred to as the 'secondary', to which the model is to be transferred. Piecewise Direct Standardization ('PDS') is a well-established example of this approach (Bouveresse & Massart, 1996) and has been employed in the present work.

The second strategy is to attempt to remove unwanted systematic variation by various normalisations which may be applied to the whole dataset or to spectra individually. In the present work, two such methods were employed: standard normal variate ('SNV') correction and rank transformation ('RT'). SNV correction, in which the spectral mean is set to zero and variance to unity (Barnes, Dhanoa, & Lister, 1989), is used widely in the analysis of near-infrared spectra to mitigate unwanted variations. Rank transformation is well-known in traditional statistical analysis, underlying non-parametric methods such as the Kruskal-Wallis test by ranks, which has been reported in the analysis of 'binned' NMR data from metabolomics studies (Jiang, Wang, Zhang, Feng, Wang, & Zhu, 2014). However, to the best of our knowledge, rank transformation has never been used as a pre-treatment applied to whole NMR spectra.

Fig. 2 illustrates the effect of these pre-treatments on a subset of spectra from argan oils only, chosen because these are represented in good numbers in the collections from both Labs and Batches. SNV and RT were applied to the raw data from both Labs. PDS was used to treat the spectra from Lab 2 only, calculating the mapping matrix using 12 'transfer' spectra selected from the samples common to both Labs by the Kenning-Stone method (Bouveresse & Massart, 1996). In the figure, the pre-treated dataset is shown in the first and third columns, for respectively the phase corrected and magnitude spectral forms.

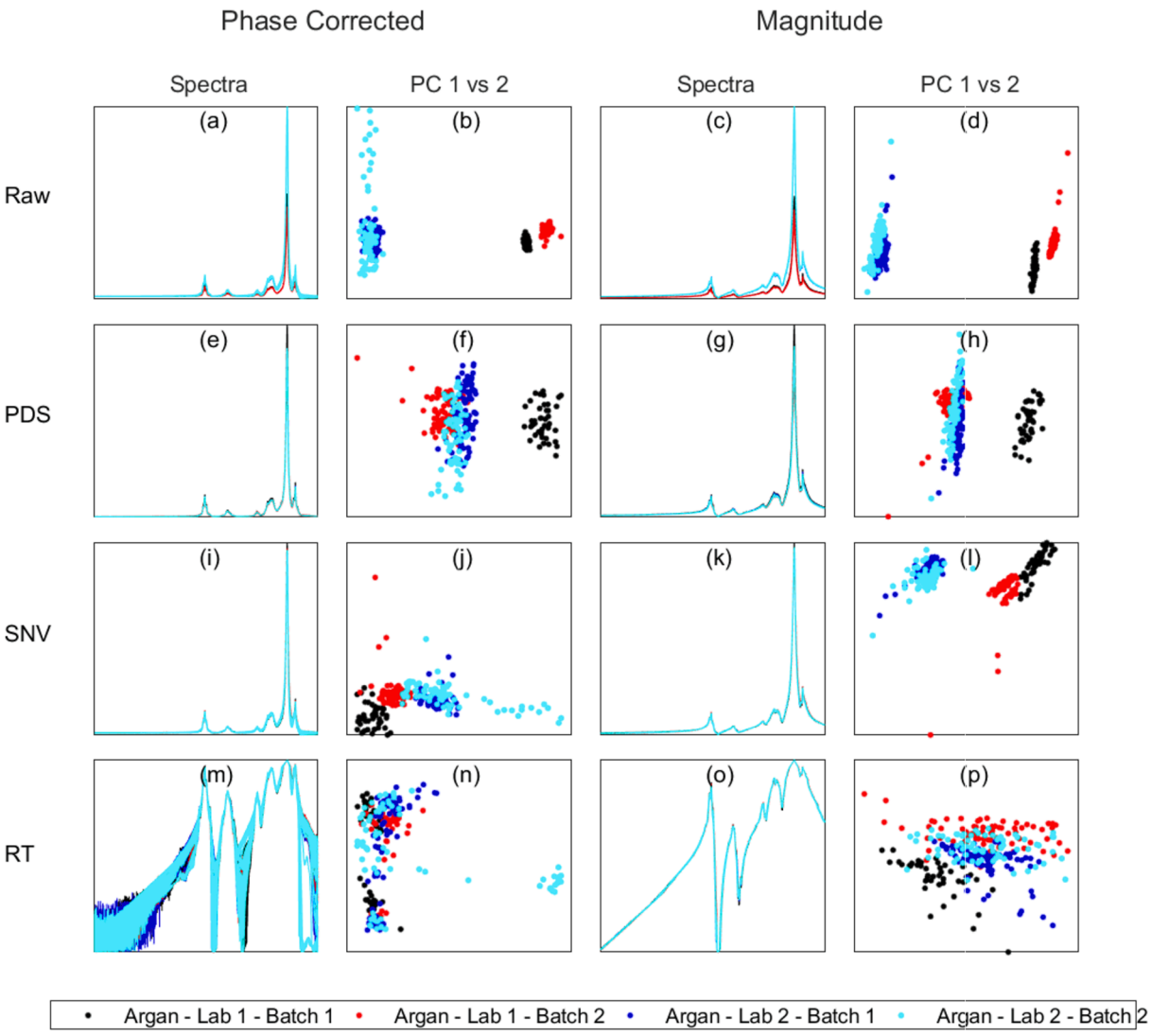


Fig. 2. In the grid of panels, the first and third columns show overlaid plots of respectively the phase corrected and magnitude spectra from argan oils only, pre-treated as indicated. Panels in the second and fourth columns show the first versus second PC scores plots from the neighbouring pre-treated datasets. The first PC score is plotted on the horizontal axis in each case, and the second on the vertical.

Panels in the second and fourth columns show PC scores plots (first versus second PCs) obtained from the adjacent dataset, and display the major effect of the pre-treatment. PDS is able to map the Lab 2 spectra nearer to Lab 1 data (panels (f) and (h)), although it cannot of course also remove the batch effect that exists within Lab 1 data. This would require a second, different PDS transform, raising the question of how multiple concurrent effects should be addressed. More generally, the performance of the PDS approach is heavily dependent on the number and nature of the transfer spectra. The method requires analogous samples from the two collections needing mapping, and after the event, these are not always available. This issue is considered again in the analysis of the historic datasets below.

SNV correction (panels (j) and (l)) mitigates the inter-instrument variation substantially but not entirely: clustering according to Lab is still seen in the first PC dimension for both spectral forms. In comparison, RT is more effective, particularly for magnitude data, where only the Lab 1 batch effect remains in the second PC dimension (panel (p)). Further pairwise scores plots up to the 5th PC dimension are given for this data subset in Supplementary Fig. 2, showing that this transformation relegates batch and instrument effects to a minor source of variance.

RT produces a dramatic change in the appearance of both phase corrected and magnitude data (compare panels (m) and (o) to the original, raw spectra of each form in panels (a) and (c)). It also emphasises differences between the two forms arising from their respective characteristics. In magnitude spectra, resonances are spread out across a wide frequency range, with the result that the whole spectrum generally contains signal as well as noise. In contrast, ideal phase corrected spectra, by definition of the phase correction routine, comprise as narrow as possible absorption bands separated by regions of flat baseline. It is the presence of the baseline which causes RT-transformed phase corrected spectra to become noisy. In an RT-treated spectrum, every data point, including those in the baseline, must take on a new value equal to its rank in the data vector. This is clearly undesirable if there are large regions of baseline. In contrast, because all the data points in a magnitude spectrum contain at least some signal, the ranking process remains meaningful throughout.

Rank transformation may offer an advantage from the perspective of visualisation, as has been recognised in other disciplines (Comins & Hussey, 2015). By swapping out intensity information for an integer rank, the maximum possible internal intensity ratio of any pair of features is fixed by the number of data points in the spectrum. Large peaks that are similar across the dataset are relatively suppressed in size, and smaller regions of dissimilarity are substantially amplified. This is

helpful for spotting small changes in the spectrum, such as the presence of a contaminant. Even a very small, unexpected peak is enough to disrupt the ranking order of spectral data points, not just at the peak position, but also at other locations in the spectrum of similar intensity. A compelling example of this in practice is given in Fig. 3, where the spectra of the 10 contaminated oils are compared with Lab 1 data, in raw phase corrected and RT-transformed magnitude forms.

3.3. Classification modelling

There are many challenges in modelling authentication scenarios connected to sample availability. Most straightforward is to define the problem as two-class classification ('authentic' versus 'non-authentic'). In this framework, plentiful samples are needed of both types. The authentic class must be large enough to capture expected variation, which can be considerable in the case of natural products. Further, real-world examples of fake or fraudulent non-authentic samples are often unavailable, so to model this class, simulated 'adulterated' samples need to be employed.

Alternatively, a one-class classification (OCC) approach can be used, in which model parameters are obtained from the positive class only, eliminating (at least mathematically) the need for abundant non-authentic samples. This is conceptually a more apposite approach for scenarios in which the authentic class is clearly defined but the non-authentic is open-ended, and in certain situations it may be the only way to proceed. Of course, in the absence of any non-authentic samples, it is impossible to estimate the false positive rates; and if these are available, it is often found that two-group approaches out-perform OCC in terms of classification success rates.

In the present study, there are two authentic classes of interest, as well as a disparate collection of mixtures used to simulate non-authentic samples. It was decided to treat the two authentication issues independently, developing separate classification models for the two 'positive classes', authentic olive and argan.

Three unrelated modelling methods were employed. These were partial least squares discriminant analysis ('PLSDA') and random forest ('RF') classification, both of which were implemented as two-class classification models; and a one-class classification method adapted for authentication scenarios based on nearest neighbour distances ('NN-OCC').

In common with most modelling methods, there are variations in how these methods can be implemented. In this study, PLSDA is used as a whole-spectrum method in common with many other reported applications of partial least squares regression (Boknaes, Jensen, Andersen, & Martens, 2002; Indahl, Sahni, Kirkhus, & Naes, 1999; Lindberg, Persson, & Wold, 1983). The RF method uses the original algorithm as proposed by Breiman (2001). In both these methods, the output for each item analysed is a score, which is used along with a threshold to ascribe it to either the positive or negative class. The NN-OCC method measures the multivariate distances of each positive class spectrum from the other authentic items, and uses a kernel density function to obtain an upper threshold for the nearest neighbour distance. Acceptance into the positive class is decided by pooling the outcomes made using an ensemble of different distance metrics. Further technical details of all the methods are given in the Supplementary Table 3.

All three approaches were used to obtain olive and argan oil authentication models from the CNESTEN sample collection. Phase corrected and magnitude datasets were analysed raw, as well as with PDS, SNV, and RT pre-treatments. Partitioning into training and test sets was carried out as detailed in the Materials and Methods section, and Lab 1 spectra from training samples only were used in model development. The established models were applied to each of the test sets (Lab 1, Lab 2, Archive 1 and Archive 2). The complete procedure was carried out 100 times using different randomly-generated partitions.

The results for the Lab 1 and Lab 2 test sets for the olive oil authentication models are shown respectively in Fig. 4(a) and (b) as heatmaps indicating the true positive rates (TPRs) and true negative rates (TNRs) obtained from the 100 repartitions. Comparison of the outcomes from these two test sets directly informs on the ability of models to transfer to a new instrument, since these contain spectra from

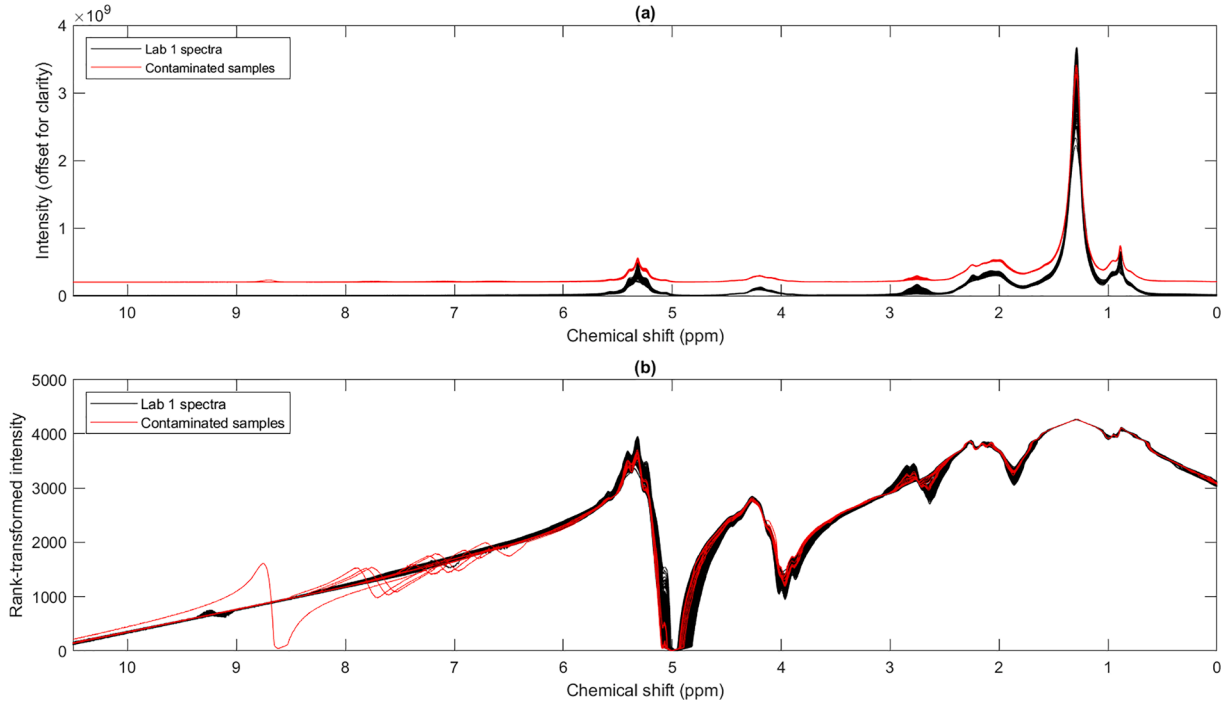


Fig. 3. Panel (a) shows the complete collection of phase corrected Lab 1 spectra (black traces) and the 10 contaminated samples (red traces), offset for clarity. It is difficult to see the contaminant peaks in a full spectral view like this, even with the offset. Panel (b) shows rank-transformed magnitude data from the same sample collection with the same colour-coding. The additional peaks from the contaminants have been amplified by the rank-transformation and are impossible to overlook. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

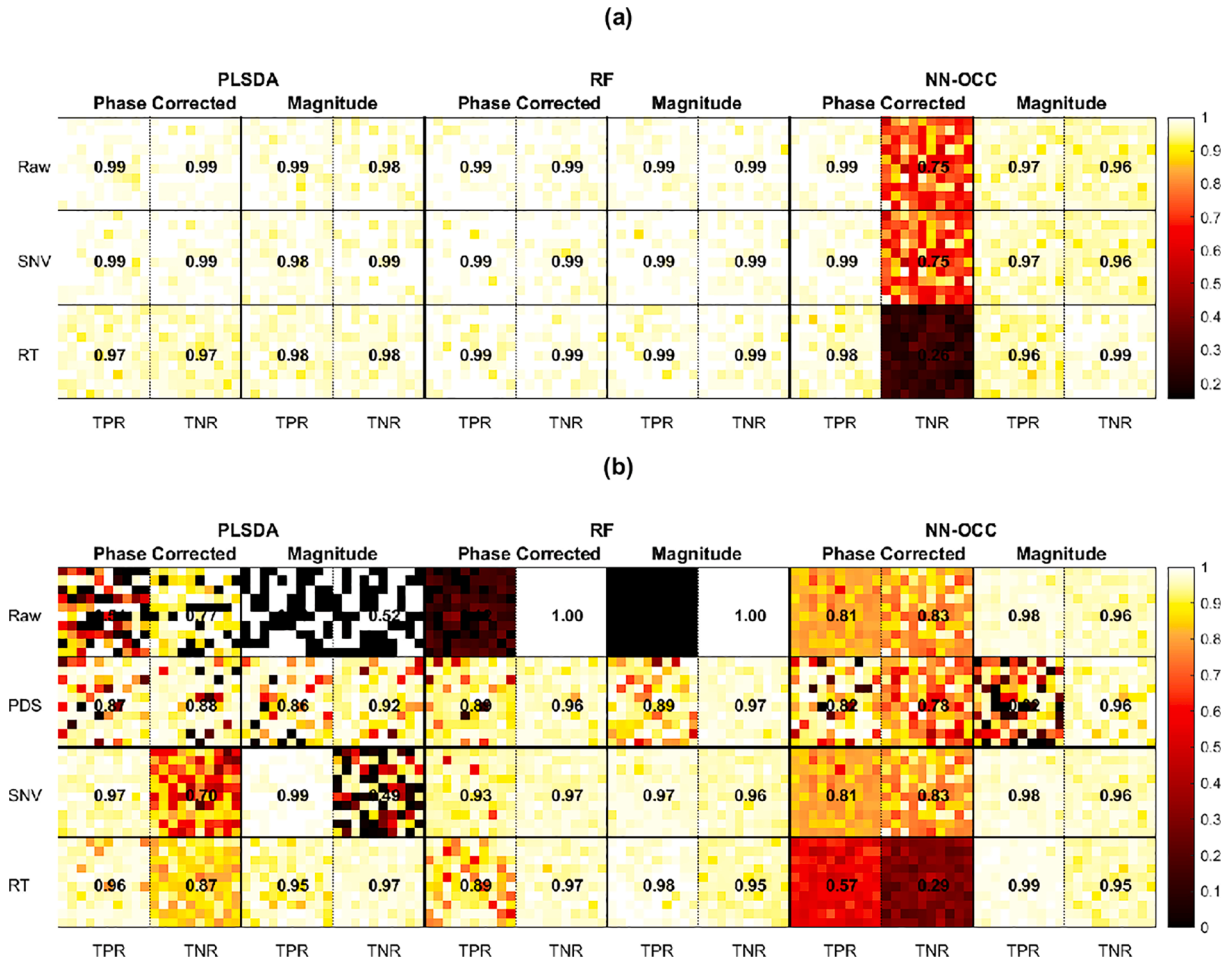


Fig. 4. Results from the olive oil authentication models calculated from Lab 1 training data, expressed as heatmaps showing all true positive rates (TPRs) and true negative rates (TNRs) from the 100 repartitions of the CNESTEN samples. The success rates from all repartitions are represented by small squares, colour-coded according to the scale on the right-hand side of the panels, and arranged into 10 × 10 blocks. Each column of blocks corresponds to a different combination of the modelling approach and spectral form used. Each row of blocks corresponds to a different data pre-treatment. The mean TPR and TNR is indicated by the value in the centre of each block. Panel (a) shows the results from application to the Lab 1 test set repartitions. Because this always comprises data from the same instrument that the model was trained with, PDS has not been applied. Panel (b) shows the results from application to the Lab 2 test set repartitions. Note that the same sample repartitions were used for both panels, thus the outcomes are directly comparable. It is clear that application of the Lab 1 models to Lab 2 data leads to a decline in classification performance by all combinations of pre-treatment/spectral form/method, ranging from small to substantial.

the same samples (with some very minor differences due to missing samples/spectra, see Materials and Methods).

Looking first at Fig. 4(a), which summarizes the results from model applications to the Lab 1 test set, it is seen that the outcomes by PLSDA and RF are very similar. Classification success rates are nearly perfect across the board by either method or pre-treatment, including none. This indicates that benchtop NMR can readily distinguish olive oil from other oil types including mixtures. Further, the specificity of NMR resonances leads to model coefficients that can be understood chemically. This is best illustrated in the whole spectrum PLSDA method, where the regression coefficients clearly map onto spectral features in the original assigned spectrum (see as an example Supplementary Fig. 3). This confirms that the discrimination between the authentic and non-authentic classes is based upon chemical information present in the NMR spectra.

From the constitution of the non-authentic samples that are misclassified, the detection limit for the adulterants under study is better than 10% w/w (see Supplementary Fig. 4). This is in agreement with multiple published studies demonstrating the potential of benchtop NMR for authenticating edible oils (Kim, Lee, Kwon, Chun, Ahn, & Kim, 2018; Riegel, 2015; Wang, Wang, Hou, & Nie, 2020; Zhu, Wang, & Chen, 2017). Rather than a limitation of benchtop NMR, this detection limit reflects the overlap in the naturally variable triglyceride compositions of different oil types.

The results for the NN-OCC method are somewhat different. With SNV or RT pre-treatment, and even no pre-treatment, it can achieve similar success rates to PLSDA and RF, but only from magnitude spectra. When using phase corrected data, the method struggles to reject non-authentic items. This is not altogether unexpected, since the approach was developed and optimised for magnitude data. It includes the use of Spearman's correlation coefficient which essentially measures distances between rank-transformed spectra, and as already discussed, RT is not an appropriate transformation for phase corrected data with large amounts of empty baseline.

Fig. 4(b) also shows a heatmap of TPRs and TNRs, but now for the outcomes from application of Lab 1 models to the Lab 2 test set. For PLSDA and RF, models trained on raw Lab 1 data are unable to transfer to raw Lab 2 spectra, with both methods yielding TPRs or TNRs closer to 0 than 1, or highly unstable outcomes with respect to the training/test partitioning (see the top row of blocks in the heatmap). This is clear evidence that the inter-instrument differences disrupt the transferability of these models to Lab 2. Notably, the NN-OCC method remains able to function reasonably well on magnitude data, yielding similar success rates to the Lab 1 test set. This hints at a greater robustness of this

approach in the face of instrument effects, likely due to the nature of the distance metrics utilised. Pre-treatment improves the outcomes for the two-group methods. The effects are somewhat variable, but overall, RF outperforms PLSDA, reaching TPRs and TNRs > 0.95 for the combination of SNV or RT pre-treatment and magnitude data.

The analogous heatmaps for the argan oil model are given in Supplementary Fig. 5. Panel (a) again represents outcomes under the ideal circumstance in which test partition spectra were acquired on the same instrument (Lab 1) as those used in model development. A generally similar pattern of outcomes is obtained as for olive, although with slightly more differentiation between the efficacies of the various spectral form/pre-treatment/method combinations. This may indicate that olive oil is more distinct from the non-authentic oils under consideration than argan, making it an easier classification problem. It could also indicate that additional systematic effects are making the argan class harder to model: recall that Figs. 1 and 2 revealed an appreciable batch difference within the Lab 1 argan oil data. In contrast, no such differences affected the olive oil spectra, since the sample delivery schedule meant that these were represented in Batch 2 only. Supplementary Fig. 5(b) shows the inter-instrument outcomes from application of the Lab 1 argan models to Lab 2 data. Likewise, these are somewhat less successful than their analogues for olive shown in Fig. 4 (b). Most importantly, only the use of magnitude data can yield reasonably stable and acceptable outcomes, with TPRs and TNRs > 0.9. A summary of the mis-classified items is given in Supplementary Fig. 6.

The full complement of olive authentication models was also applied to the two Archive test sets. The PDS pre-treatment requires a small set of transfer spectra obtained on each instrument from which to calculate the mapping matrix. Since both archives dated back several years and the spectrometers concerned have been respectively refurbished (Archive 1) and decommissioned (Archive 2), it was impossible to collect any further spectra. The workaround used was to randomly select a number of spectra covering the same oil types present in the training set (with the exception of argan oil, which was not represented in either of the archived collections).

The outcomes from these test sets are summarised in Fig. 5, together

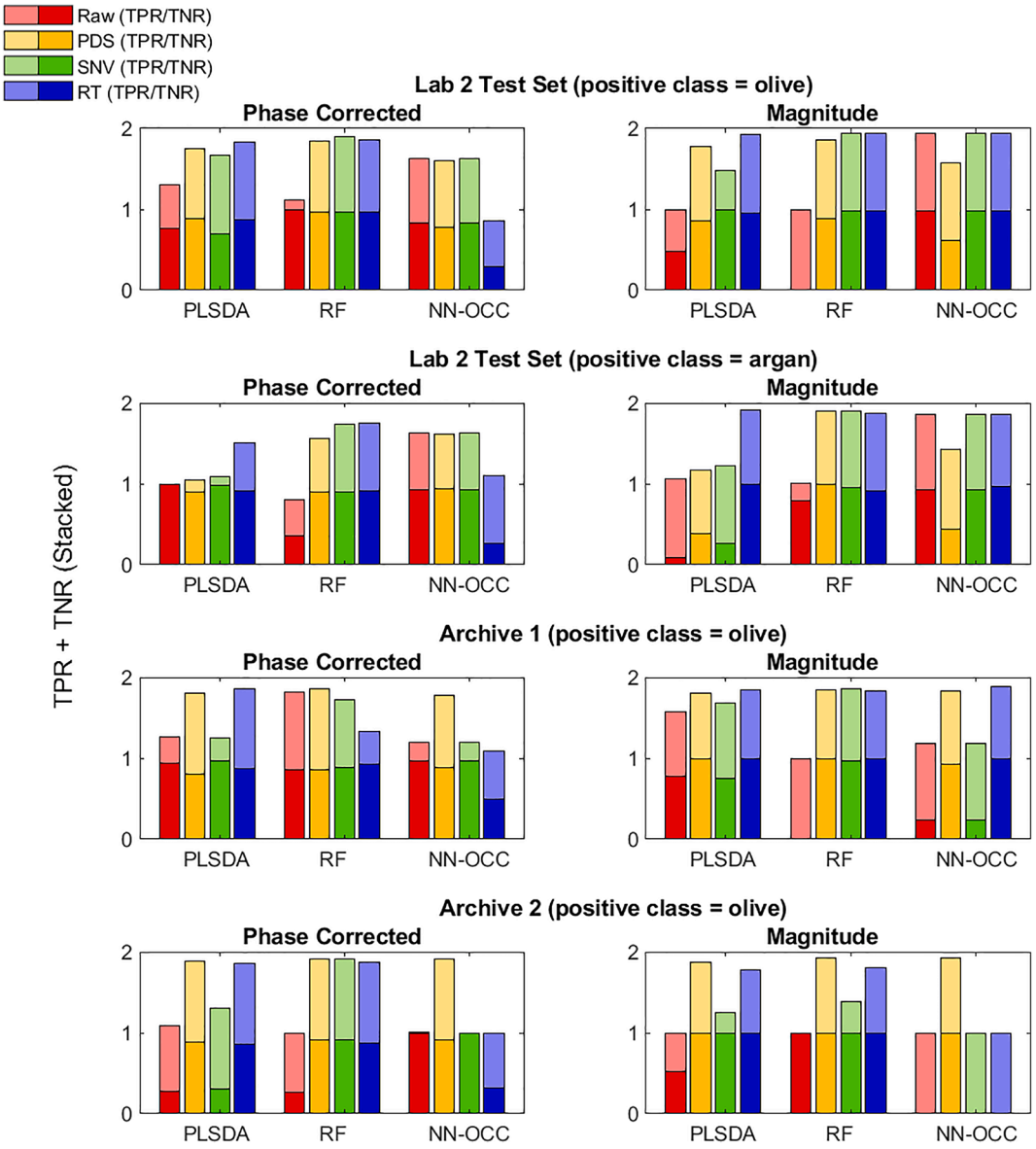


Fig. 5. Stacked bar charts showing the TPR and TNR for each of the method/pre-treatment/spectral form combinations. Each row of panels corresponds to one of the four inter-instrument test sets (respectively, Lab 2 (olive model); Lab 2 (argan model); Archive 1 and Archive 2 (both olive model). The two columns of panels correspond respectively to the phase corrected and magnitude spectral forms. Within each panel, the stacked bars are grouped by method as indicated on the horizontal axis, and coloured by pre-treatment as shown in the legend.

with those from the other inter-instrument test sets discussed above (Lab 2, olive and argan) for ease of comparison. From this figure, it can be concluded that using raw data when transferring between instruments will likely fail. The most reliable combination of pre-treatment and spectral form is RT applied to magnitude data, which gives acceptable outcomes for all modelling methods and test sets (with the sole exception of NN-OCC and Archive 2). SNV and PDS are less consistently effective, although good outcomes can be obtained, most often in combination with the RF method. An alternative presentation of this summary of results is given in Supplementary Table 4.

4. Conclusions

Benchtop NMR is well-suited to analysis by chemometric methods that draw on information from across whole spectral profiles. This is contrast to high field NMR, where the focus of data analysis is generally on integrals of well-resolved peaks. As in other high-throughput spectroscopy methods, unwanted systematic variances can occur in benchtop data from a variety of sources. Within spectra collected even from a single spectrometer, there may be incremental or step changes over time that cause batching effects. Inter-instrument differences may be greater still and impact on the transfer of databases and accompanying chemometric models to different spectrometers. Some of these systematic effects may only come to light with hindsight.

For chemometric models developed on one machine to reliably work on another, some form of spectral pre-treatment is needed. In this work, three conceptually different multivariate approaches were employed to model authentication scenarios for olive and argan oils. These were used in conjunction with three different data pre-treatments: PDS, SNV and RT. The first two are established spectral normalisation methods; the latter is not well-known in conjunction with spectral data. Without pre-treatment, models developed using data from one instrument have unpredictable, and likely poor, generalisation ability to new spectrometers. All the pre-treatment types considered improved outcomes, although not always sufficiently to allow effective transfer of the models between instruments.

The use of magnitude as well as phase corrected data was also examined. The magnitude spectral form is a direct representation of the NMR measurement which avoids any artefacts arising from phase correction. Magnitude data do not have the conventional appearance of a baseline with peaks; however, trading a visually familiar spectral form for greater spectral reproducibility offers clear benefits for chemometric modelling, producing effective multivariate models that are comparatively more robust to instrument differences. If mitigating batch or instrumental effects is important, then serious consideration should be given to working not only with phase corrected spectra but also with magnitude data.

It was found that the combination of RT applied to magnitude data offers particular advantage. Rank transformed spectra are visually very different from their raw counterparts, but nevertheless retain and indeed augment spectral features. The large dynamic range in peak intensities usual in NMR is thus much compressed. This property makes RT unsuitable for data with large amounts of baseline, but ideally suited to magnitude spectra where the signal rather than just noise is present across a wide frequency range. In particular, small peaks from contaminants, which are easily overlooked in typical phase corrected views, are amplified to become much more noticeable, making it a useful approach for data inspection. Finally, multivariate models developed from rank transformed magnitude data appear the most likely to remain effective when transferred to different spectrometers, particularly in conjunction with the RF or NN-OCC modelling methods.

CRediT authorship contribution statement

Yvonne Gunning: Data Curation, Investigation, Methodology, Writing – review & editing. **Fouad Taous:** Conceptualization, Methodology, Resources. **Tibari El Ghali:** Methodology, Resources. **James D. Gibbon:** Investigation. **E. Wilson:** Formal Analysis, Software, Writing – review & editing. **Rachel M. Brignall:** Investigation, Writing – review & editing. **E. Kate Kemsley:** Data Curation, Formal Analysis, Software, 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.

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 BB/CCG1860/1. We also thank our collaborators from the FAO/IAEA Coordinated Research Project 'Field-deployable Analytical Methods to Assess the Authenticity, Safety and Quality of Food' (project code D52040) for providing authentic argan oils.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2021.131333.

References

Alam, T. M., Alam, M. K., McIntyre, S. K., Volk, D. E., Neerathilingam, M., & Luxon, B. A. (2009). Investigation of Chemometric Instrumental Transfer Methods for High-Resolution NMR. *Analytical Chemistry, 81*(11), 4433–4443. https://doi.org/10.1021/ac900262g

Ancora, D., Milavec, J., Gradišek, A., Cifelli, M., Sepe, A., Apih, T., … Domenici, V. (2021). Sensitivity of Proton NMR Relaxation and Proton NMR Diffusion Measurements to Olive Oil Adulterations with Vegetable Oils. *Journal of Agricultural and Food Chemistry*. https://doi.org/10.1021/acs.jafc.1c00914

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

Boknaes, N., Jensen, K. N., Andersen, C. M., & Martens, H. (2002). Freshness assessment of thawed and chilled cod fillets packed in modified atmosphere using near-infrared spectroscopy. *Lebensmittel-Wissenschaft Und-Technologie-Food Science and Technology, 35*(7), 628–634. https://doi.org/10.1006/fstl.923

Bouveresse, E., & Massart, D. L. (1996). Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration. *Chemometrics and Intelligent Laboratory Systems, 32*(2), 201–213. https://doi.org/10.1016/0169-7439(95)00074-7

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Comins, J. A., & Hussey, T. W. (2015). Compressing multiple scales of impact detection by Reference Publication Year Spectroscopy. *Journal of Informetrics, 9*(3), 449–454. https://doi.org/10.1016/j.joi.2015.03.003

Feundale, R. N., Woody, N. A., Tan, H. W., Myles, A. J., Brown, S. D., & Ferre, J. (2002). Transfer of multivariate calibration models: A review. *Chemometrics and Intelligent Laboratory Systems, 64*(2), 181–192. <Go to ISI>://WOS:000179135600006.

Fonollosa, J., Fernandez, L., Gutierrez-Galvez, A., Huerta, R., & Marco, S. (2016). Calibration transfer and drift counteraction in chemical sensor arrays using Direct Standardization. *Sensors and Actuators B-Chemical, 236*, 1044–1053. https://doi.org/10.1016/j.snb.2016.05.089

Galvan, D., Bona, E., Borsato, D., Danieli, E., & Montazzolli Killner, M. H. (2020). Calibration Transfer of Partial Least Squares Regression Models between Desktop Nuclear Magnetic Resonance Spectrometers. *Analytical Chemistry, 92*(19), 12809–12816. https://doi.org/10.1021/acs.analchem.0c0090210.1021/acs.analchem.0c00902.s001

Gerdova, A., Defernez, M., Jakes, W., Limer, E., McCallum, C., Nott, K., … Kemsley, E. K. (2015). 60 MHz 1H NMR spectroscopy of triglyceride mixtures. In F. Capozzi, L. Laghi, & P. S. Belton (Eds.), *Magnetic Resonance in Food Science: Defining Food by Magnetic Resonance* (pp. 19–30). Cambridge: Royal Soc Chemistry.

Gunning, Y., Jackson, A. J., Colmer, J., Taous, F., Philo, M., Brignall, R. M., … Kemsley, E. K. (2020). High-throughput screening of argan oil composition and authenticity using benchtop(1)H NMR. *Magnetic Resonance in Chemistry, 58*(12), 1177–1186. https://doi.org/10.1002/mrc.5023

Harrington, P. D., & Wang, X. Y. (2017). Spectral Representation of Proton NMR Spectroscopy for the Pattern Recognition of Complex Materials. *Journal of Analysis and Testing, 1*(2), 11. https://doi.org/10.1007/s41664-017-0003-y

Indahl, U. G., Sahni, N. S., Kirkhus, B., & Naes, T. (1999). Multivariate strategies for classification based on NIR-spectra - with application to mayonnaise. *Chemometrics and Intelligent Laboratory Systems, 49*(1), 19–31. https://doi.org/10.1016/s0169-7439(99)00023-4

Jiang, M. M., Wang, C. H., Zhang, Y., Feng, Y. F., Wang, Y. F., & Zhu, Y. (2014). Sparse Partial-least-squares Discriminant Analysis for Different Geographical Origins of Salvia miltiorrhiza by H-1-NMR-based Metabolomics. *Phytochemical Analysis, 25*(1), 50–58. https://doi.org/10.1002/pca.2461

Jimenez-Carvelo, A. M., Perez-Castano, E., Gonzalez-Casado, A., & Cuadros-Rodriguez, L. (2017). One input-class and two input-class classifications for differentiating olive oil from other edible vegetable oils by use of the normal-phase liquid chromatography fingerprint of the methyl-transesterified fraction. *Food Chemistry, 221*, 1784–1791. https://doi.org/10.1016/j.foodchem.2016.10.103

Kim, J. H., Lee, H. J., Kwon, K., Chun, H. S., Ahn, S., & Kim, B. H. (2018). A 43 MHz Low-Field Benchtop H-1 Nuclear Magnetic Resonance Method to Discriminate Perilla Oil Authenticity. *Journal of Oleo Science, 67*(5), 507–513. https://doi.org/10.5650/jos.ess17243

Köppel, R., van Velsen, F., Ganeshan, A., Pietsch, K., Weber, S., Graf, C., … Licina, A. (2020). Multiplex real-time PCR for the detection and quantification of DNA from chamois, roe, deer, pork and beef. *European Food Research and Technology, 246*(5), 1007–1015. https://doi.org/10.1007/s00217-020-03468-1

Lindberg, W., Persson, J. A., & Wold, S. (1983). Partial least-squares method for spectrofluorimetric analysis of mixtures of humic-acid and ligninsulfonate. *Analytical Chemistry, 55*(4), 643–648. https://doi.org/10.1021/ac00255a014

McVey, C., McGrath, T. F., Haughey, S. A., & Elliott, C. T. (2021). A rapid food chain approach for authenticity screening: The development, validation and transferability of a chemometric model using two handheld near infrared spectroscopy (NIRS) devices. *Talanta, 222*, 8. https://doi.org/10.1016/j.talanta.2020.121533

Parker, T., Limer, E., Watson, A. D., Defernez, M., Williamson, D., & Kemsley, E. K. (2014). 60 MHz H-1 NMR spectroscopy for the analysis of edible oils. *Trac-Trends in Analytical Chemistry, 57*, 147–158. https://doi.org/10.1016/j.trac.2014.02.006

Quintanilla-Casas, B., Strocchi, G., Bustamante, J., Torres-Cobos, B., Guardiola, F., Moreda, W., … Vichi, S. (2021). Large-scale evaluation of shotgun triacylglycerol profiling for the fast detection of olive oil adulteration. *Food Control, 123*, 107851. https://doi.org/10.1016/j.foodcont.2020.107851

Riegel, S. D. (2015). Determination of Olive Oil Adulteration With 60-MHz Benchtop NMR Spectrometry. *American Laboratory, 47*(2), 16–19. <Go to ISI>://WOS:000351333500005.

Sobolev, A. P., Thomas, F., Donarski, J., Ingallina, C., Circi, S., Cesare Marincola, F., … Mannina, L. (2019). Use of NMR applications to tackle future food fraud issues. *Trends in Food Science & Technology, 91*, 347–353. https://doi.org/10.1016/j.tifs.2019.07.035

Sota-Uba, I., Bamidele, M., Moulton, J., Booksh, K., & Lavine, B. K. (2021). Authentication of edible oils using Fourier transform infrared spectroscopy and pattern recognition methods. *Chemometrics and Intelligent Laboratory Systems, 210*, 9. https://doi.org/10.1016/j.chemolab.2021.104251

Tang, F. F., Green, H. S., Wang, S. C., & Hatzakis, E. (2021). Analysis and Authentication of Avocado Oil Using High Resolution NMR Spectroscopy. *Molecules, 26*(2), 11. https://doi.org/10.3390/molecules26020310

van Beek, T. A. (2021). Low-field benchtop NMR spectroscopy: Status and prospects in natural product analysis. *Phytochemical Analysis, 32*(1), 24–37. https://doi.org/10.1002/pca.v32.110.1002/pca.2921

Velasco, A., Ramilo-Fernandez, G., Denis, F., Oliveira, L., Shum, P., Silva, H., & Sotelo, C. G. (2021). A New Rapid Method for the Authentication of Common Octopus (Octopus vulgaris) in Seafood Products Using Recombinase Polymerase Amplification (RPA) and Lateral Flow Assay (LFA). *Foods, 10*(8), 14. https://doi.org/10.3390/foods10081825

Wang, X., Wang, G. L., Hou, X. W., & Nie, S. D. (2020). A Rapid Screening Approach for Authentication of Olive Oil and Classification of Binary Blends of Olive Oils Using Low-Field Nuclear Magnetic Resonance Spectra and Support Vector Machine. *Food Analytical Methods, 13*(10), 1894–1905. https://doi.org/10.1007/s12161-020-01799-z

Williamson, D., Kemsley, E. K., Sutcliffe, O., & Mewis, R. (2019). A method for screening psychoactive substances. GB2571817A·2019-09-11.

Zhu, W. R., Wang, X., & Chen, L. H. (2017). Rapid detection of peanut oil adulteration using low-field nuclear magnetic resonance and chemometrics. *Food Chemistry, 216*, 268–274. https://doi.org/10.1016/j.foodchem.2016.08.051