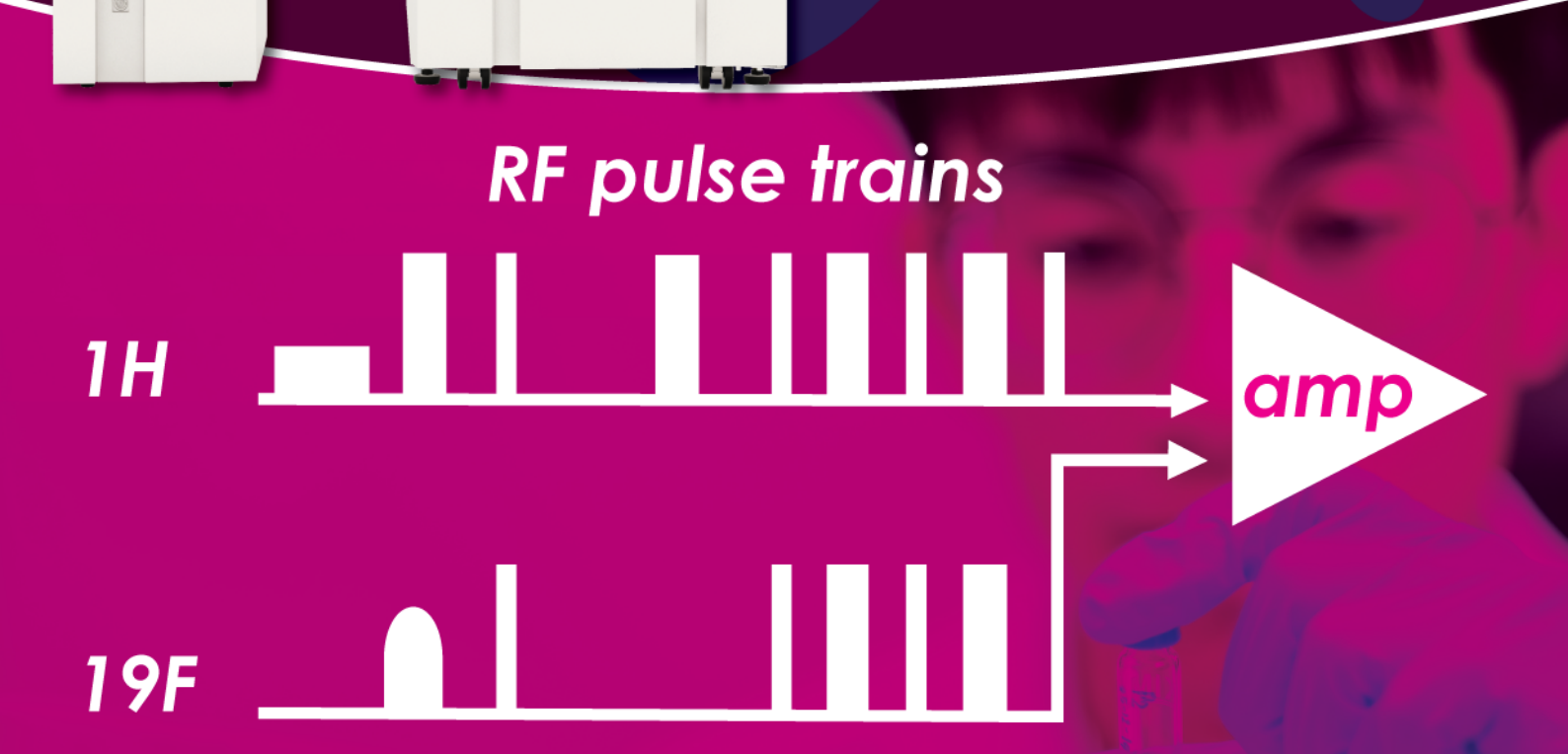
JEOL
Solutions for Innovation

Unlock the Potential of Multi-Resonance Experiments

Download the Guide



Received: 19 March 2023 | Revised: 24 May 2023 | Accepted: 30 May 2023

DOI: 10.1002/mrc.5373

RESEARCH ARTICLE

WILEY

Graphical exploration of 600- and 60-MHz proton NMR spectral datasets from ground roast coffee extracts

E. Kate Kemsley

Core Science Resources Group, Quadram Institute Bioscience, Norwich, UK

Correspondence
E. Kate Kemsley, Core Science Resources Group, Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, UK.
Email: kate.kemsley@quadram.ac.uk

Funding information
Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/CCG1860/1

Abstract

This article uses a variety of graphical and mathematical approaches to analyse 600- and 60-MHz ('benchtop') proton NMR spectra acquired from lipophilic and hydrophilic extracts of roasted coffee beans. The collection of 40 authenticated samples comprised various coffee species, cultivars and hybrids. The spectral datasets were analysed by a combination of metabolomics approaches, cross-correlation and whole spectrum methods, assisted by visualisation and mathematical techniques not conventionally employed to treat NMR data. A large amount of information content was shared between the 600-MHz and benchtop datasets, including in its magnitude spectral form, suggesting the potential for a lower cost, lower tech route to conducting informative metabolomics studies.

KEYWORDS

benchtop NMR, coffee, compact NMR, graphical exploration, high-resolution NMR, inter-correlation, magnitude spectra, metabolomics, spectral form, visualisation

1 | INTRODUCTION

High-resolution NMR spectroscopy is utilised extensively in many scientific disciplines. In the life sciences, major uses include the elucidation of structures of biochemical compounds, metabolite quantitation and metabolomics experiments. The current work concerns a 600-MHz ^{1}H NMR metabolomics investigation of hydrophilic (polar) and lipophilic (non-polar) extracts of ground roast coffee, combined with a parallel study of lipophilic extracts obtained from the same sample collection using 60-MHz ('benchtop' or 'compact') ^{1}H NMR.

The genesis of modern benchtop NMR spectroscopy dates back to the early 2000s when significant advances were made in the design and manufacture of rare-earth ferromagnets. These permanent magnets produce fields of ~1.5 T. Corresponding to a proton Larmor frequency of 60 MHz, this was a typical field strength in state-of-the-art NMR spectroscopy of the 1970s. However, developments in electronics and computing power make today's 60-MHz spectrometer an entirely different beast: Fourier transform rather than continuous wave, with high resolution on the frequency scale and much-improved signal to noise. Their size and weight are suitable for truly 'benchtop' operation, and there is a strong emphasis on ease of use: The only part of the system requiring user interaction is the probe, located between the poles of the magnet and accommodating the sample tube.

Taking a more commercial view, a key attraction of a benchtop NMR spectrometer is its much lower capital, maintenance and running costs,[1] which are an order of magnitude less expensive than a research grade instrument. Most reported studies that involve benchtop NMR

have a focus on industry-oriented or high-throughput applications. Examples include a variety of quality control and screening tasks,[2] and benchtop versus high-field comparisons, often in terms of the performance of predictive models.[3–7]

Coffee has been examined by both high-field and benchtop ^{1}H NMR for a variety of purposes: structural determination of individual compounds[8] and compositional analysis of green[9,10] and roasted coffee beans[11,12] and of coffee oil and spent coffee grains.[13] Many other agricultural commodities have likewise been studied by NMR including beverages (tea[14,15] and wine[16]) and food too, such as vegetable oils[17] and meat,[18] with appropriate extraction techniques for handling the different sample matrixes. Another recorded area of application is product authentication. There is a long history of fraud in the global food supply chain, and analytical tests can help by detecting instances of adulteration of high-value products with cheaper alternatives. With an authentication focus, NMR has been used to study pulp-wash from citrus fruits[19] and the spices saffron[20,21] and vanilla.[22]

In the coffee sector, cases of fraud usually involve the substitution of *Coffea arabica* L. ('Arabica') with the other commercially important species, *C. canephora* Pierre ex A. Froehner (also known as *C. canephora v. robusta*). The former generally trades at twice the price of the latter. The presence of canephora in a blend is betrayed by its aftertaste, described as 'peanutty',[23] although the limit of detection even by trained sensory experts is around 15% w/w. Analytical tests developed using NMR can perform substantially better than this,[24–27] down to detection limits commensurate with 'adventitious contamination'[28,29]: very low amounts of adulteration arising from some kind of mistake during transport or processing rather than deliberate fraud.

Spectra of complex biological samples such as coffee invariably contain large numbers of peaks, typically in the hundreds, and the dynamic range of the spectral intensities is such that the largest resonances dwarf the smallest by orders of magnitude. This makes holistic inspection of the spectra quite a challenge. In the present work, various strategies for visual exploration and comparison of such datasets are explored, using a range of different graphical methods and resources.[30–34] These are inspired by approaches used in other disciplines but unconventional in NMR data analysis, perhaps because researchers tend to utilise only software provided with the instrument console. It is shown that these contemporary, data-agnostic software tools and graphics resources offer many possibilities for the compelling visualisation and effective exploration of complex NMR datasets.

The datasets used in the work comprise 60- and 600-MHz ^{1}H NMR spectra of lipophilic extracts and 600-MHz ^{1}H NMR spectra of hydrophilic extracts from roast coffee beans of assured origin. Also considered are 60-MHz magnitude spectra from the same samples, as well as two additional collections: known blends and hospitality/retail-purchased surveillance samples. The spectra were obtained during the development of a method for coffee species authentication via a selected peak integral.[28] None have been previously treated as whole spectral profiles, and the magnitude data have not been previously disclosed. All datasets have been placed in the public domain[35] to encourage further analysis.

2 | MATERIALS AND METHODS

2.1 | Datasets

A public domain data repository[35] of NMR spectra from extracts of roast coffee beans was used in this work. It contains 60- and 600-MHz ^{1}H NMR spectra of lipophilic extracts and 600-MHz ^{1}H NMR spectra of hydrophilic extracts, giving three matched datasets of 40 spectra from samples of the same original coffees. Details of the samples are given in Table 1. The sample preparation methods and spectral acquisition have been previously described in Gunning et al.[28] but for convenience are summarised in Table S1.

Also used in this work are 60-MHz spectra of lipophilic extracts from two further sample collections: 27 serial blends of assured Arabica and canephora beans and 63 hospitality/retail-purchased ground roast coffees. Sample preparation and spectral acquisition were as above; sample listings are given in Table S2. These spectra in their magnitude form together with those of the 40 assured origin coffees detailed above are made available to accompany this manuscript.

2.2 | Data processing

Following routine FID post-processing (see Table S1), all spectra were exported via MestreNova (Mestrelab Research, Santiago de Compostela, Spain) from their native binary format into .csv files for further data analysis.

All subsequent data processing was carried out in the Matlab technical computing language installed with the Statistics & Machine Learning and Signal Processing toolboxes (The Mathworks, Cambridge, UK). In addition, use has been made of various third-party, public domain tools for the Matlab environment, as follows: the Crameri collection of scientific colour maps,[31] the 'circularGraph' toolbox[32] and the sparse canonical correlation analysis (CCA) toolbox.[34]

TABLE 1 Details of the assured origin coffee beans used to obtain the spectra utilised in the present work.

#	Sample code	Species/hybrid	Country of origin	Category	# Gunning et al.[28]
1	WB_01	Arabica	Ethiopia	Wild type	16
2	WB_02	Arabica	Ethiopia	Wild type	8
3	WB_03	Arabica	Ethiopia	Wild type	15
4	WB_04	Arabica	Uganda	Cultivar	25
5	WB_05	Arabica	Ethiopia	Wild type	6
6	WB_06	Arabica	Brazil	Cultivar	24
7	WB_07	Arabica	La Reunion	Cultivar	23
8	WB_08	Arabica	Colombia	Cultivar	22
9	WB_09	Arabica	Ethiopia	Wild type	14
10	WB_10	Arabica	Ethiopia	Wild type	13
11	WB_11	Arabica	Ethiopia	Wild type	1
12	WB_12	Arabica	Brazil	Cultivar	21
13	WB_13	Arabica	Ethiopia	Wild type	2
14	WB_14	Arabica	Ethiopia	Wild type	3
15	WB_15	Arabica	Ethiopia	Wild type	4
16	WB_16	Arabica	Ethiopia	Wild type	5
17	WB_17	Arabica	Ethiopia	Wild type	7
18	WB_18	Arabica	Ethiopia	Wild type	9
19	WB_19	Arabica	Ethiopia	Wild type	10
20	WB_20	Arabica	Ethiopia	Wild type	11
21	WB_21	Arabica	Ethiopia	Wild type	12
22	WB_22	Arabica	Ethiopia	Wild type	17
23	WB_23	Arabica	Colombia	Wild type/cultivar	18
24	WB_24	Arabica	Colombia	Cultivar	19
25	WB_25	Arabica	Colombia	Cultivar	20
26	WB_26	Arabica	Colombia	Cultivar	26
27	WB_27	Arabica	Costa Rica	Cultivar	27
28	WB_28	Arabica	Kenya	Cultivar	28
29	WB_29	Arabica × robusta < arabica (backcross)	Kenya	Cultivar (hybrid)	29
30	WB_30	Arabica × robusta < arabica (backcross)	Kenya	Cultivar (hybrid)	30
31	WB_31	Arabica/canephora (nominal 70:30 blend)	Cameroon		n/a[a]
32	WB_32	Arabica × robusta (‘Arabusta’ hybrid)	Kenya	Hybrid	31
33	WB_33	Canephora	Vietnam		34
34	WB_34	Canephora	Rwanda		36
35	WB_35	Canephora	India		32
36	WB_36	Canephora	India		37
37	WB_37	Canephora	Brazil		35
38	WB_38	Canephora	Indonesia		33
39	WB_39	Congensis × robusta	India	Hybrid	38
40	WB_40	Liberica	Uganda		40

Note: All samples were supplied as roasted whole beans by the Royal Gardens at Kew to the original research project reported in Gunning et al.[28] in which additional details (precise geographical origin, cultivar, etc) can be found. Code numbers as used in that work are also supplied here.

[a]Sample not used in Gunning et al.[28]

3 | RESULTS AND DISCUSSION

3.1 | Exploring the 600-MHz lipophilic spectra

Hundreds of compounds are present in the lipophilic and hydrophilic extracts of ground roast coffee. These include primary and secondary metabolites of coffee beans and breakdown products from the roasting process. The ^{1}H 600-MHz spectra reflect this complexity. An example of a lipophilic extract spectrum is shown in Figure 1a. Coffee beans typically contain 10%–15% w/w coffee oil, composed mostly (>80% w/w) of triglycerides,[9] so the lipophilic spectrum is dominated by resonances attributable to fatty acids (FA) and the glycerol backbone (Gly), as indicated on the figure. Other peaks include caffeine (Caf) and the various diterpenes (Dit), although these smaller and more interesting peaks from a compositional perspective are poorly visible at this scale, thus hard to examine in the context of the spectrum as a whole. This is because the full complement of resonances accessible by modern high-field NMR comprises intensities across a very large dynamic range. Dealing with data that contains information at vastly different scales is a familiar problem in other disciplines; consider satellite imagery, for example, where it is a common experience to magnify aerial views of the ground only to lose sight of the bigger picture.

3.1.1 | Log compression

Another analogue of the dynamic range problem is found in audio processing, where the solution is to apply compression to the signal intensities. In the simplest

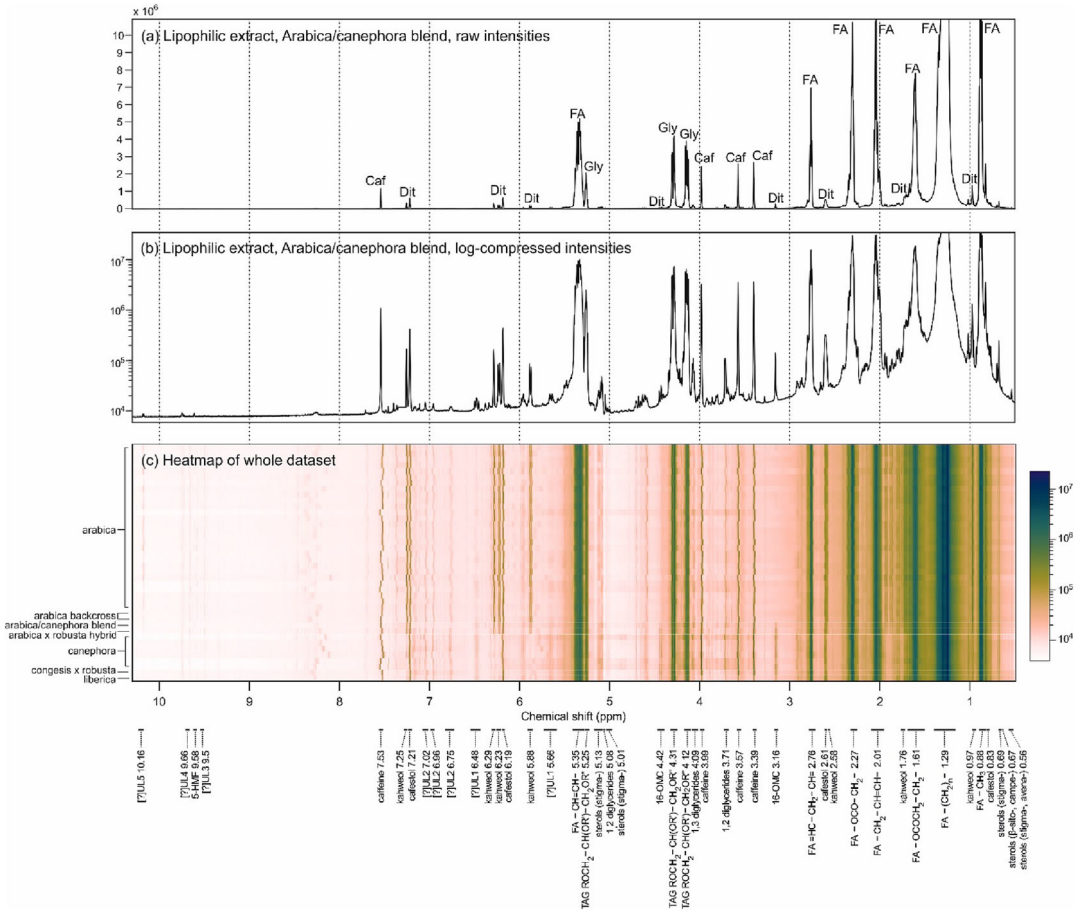


FIGURE 1 (a) The 600-MHz ^{1}H NMR spectrum of a lipophilic extract from an Arabica/canephora blend (sample WB_31). Peaks attributed to compound classes are marked as follows: FA, fatty acids; Gly, glyceride backbone; Caf, caffeine; Dit, diterpenes (cafestol, kahweol, 16-OMC). The 7.26 ppm solvent peak (d-chloroform) has been suppressed here and in the panels to follow. (b) The same spectrum after the application of a logarithmic compression function. (c) A heatmap view of the complete set of log-compressed spectra from all 40 lipophilic extracts. To aid clarity, the pixels are elongated vertically so that each compressed intensity value is represented by a vertical bar. Each row represents one spectrum; rows have been arranged by sample type, as indicated on the left-hand side. Colours are mapped to the log-compressed intensities using 'Batlow-W' (the 'scientific rainbow' from the Crameri colour map collection). Indicated below the heatmap are the integration regions used in the bucketing procedure, with nominal peak centres as marked.

approach, each intensity value I_ν is transformed by a logarithmic function, and this approach is equally applicable to NMR spectra. The processed spectrum is given by $f(I_\nu) = a \log_{10}(1 + b I_\nu)$, where the parameter b is chosen to produce the required amount of compression and a controls the output amplitude. By choosing suitable values for these, the spectral profile can be substantially altered, relatively amplifying smaller features and suppressing larger ones. Notice that this transformation is applied to individual data values, in contrast to scaling methods that operate on the data matrix row-wise (e.g. standard normal variate [SNV] or total area normalisations) or column-wise (e.g. standardisation or Pareto scaling), where each row contains a spectrum.

Log transformations of NMR spectra have been previously reported as beneficial pre-treatments in multivariate analysis.[36,37] In the present work, the primary interest is in improving visualisation. The log-compressed spectrum is illustrated in Figure 1b. Many more peaks are revealed; some of which were three orders of magnitude smaller on the original scale than the largest. The compression function has allowed all to be examined simultaneously.

3.1.2 | Heatmap representations of spectral datasets

Log-compressed spectra are also more amenable to visualisation as a heatmap than raw spectral intensities. Figure 1c shows the complete collection of lipophilic extract spectra, arranged row-wise into a matrix and presented as an image. This form of data representation is encountered in other scientific fields but has only occasionally been reported for NMR spectra.[21,38] It provides a clear way of presenting a large, complex dataset in a single illustration. This can be helpful in revealing systematic effects that are not apparent from overlaid or stacked spectra. For example, by organising the rows of the data matrix by sample type, differences between Arabica and other coffee species are made clear. This is especially obvious amongst the diterpenes that are known to be strongly associated with coffee species.[39,40] In contrast, other regions (e.g. >8 ppm, aromatics and aldehydes) are seen to vary individually by sample.

The heatmap also makes it easy to examine the alignment of peaks across the dataset: Some resonances are very well aligned (FA, for instance), others less so (Caf). Rather than using the solvent peak as a reference, the dataset shown here has been registered using correlation-optimised alignment (COA).[41] This produces a better alignment of peaks across the whole spectral range. The chemical shift scale was attached post hoc by setting the average position of the largest FA peak, present in all lipophilic extracts, to 1.25 ppm. For comparison, Figure S3(a) shows the heatmap of the dataset aligned to the solvent peak (d-chloroform, 7.26 ppm).

The colour map used in Figures 1c and S3(a) is 'Batlow-W'[42] from the Crameri collection of scientific colour maps, which provide perceptually uniform colour gradients and colour combinations readable by people with colour vision deficiencies. How these look in human eyes is self-evidently a subjective experience, but we can experiment further still with alternative maps to best communicate the information at hand. This is illustrated in Figure S3(b)–(c), which compares heatmaps of the chloroform-aligned, log-compressed dataset constructed using a variety of different colour mappings.

3.1.3 | Using heatmaps to assist with bucketing and annotation

A well-established treatment of NMR spectra is 'bucketing', which involves integrating peak area(s) within selected regions along the chemical shift scale. Ideally, each bucket is associated with a single functional group, and buckets attributed to the same molecule are expected to be highly correlated. In studies of natural products, which contain numerous compounds of varying concentrations, relative sideways shifting of peaks from sample to sample is unavoidable. This means that stated peak chemical shifts are approximate and may differ with experimental conditions. As such, bucketing has become a routine approach in metabolomics work. It aims to mitigate this imprecision by effectively de-resolving the spectrum from tens of thousands of data points to no more than hundreds of buckets.

However, misaligned peaks can present challenges during the bucketing process. This has long been recognised as an issue in NMR.[43] There have been a variety of attempts to automate the procedure,[44,45] but currently, human oversight is still advised. For large studies, it is important to consider the whole sample collection in setting appropriate integration limits, rather than just a single spectrum. With this in mind, the lipophilic extract dataset was bucketed using the integration regions as indicated at the bottom of Figure 1c; these were informed by inspection of heatmaps local to each peak. The integration regions and nominal peak centres are also tabulated in Table S4. For many of the smaller peaks, local COA was necessary to fine-tune the integration region. The benefit of doing this is illustrated in Figure S5 for the cafestol (at 6.19 ppm) and kahweol peaks (at 6.23 and 6.29 ppm).

Annotations have been drawn from established literature assignments,[46–48] with some revisions that were

made possible as a direct result of the careful bucketing process. One such example is the peak at 2.60 ppm, which has been assigned in the literature to cafestol.[46] However, close inspection of the spectral profiles in this region strongly suggests contributions from more than one compound (Figure S6). A likely candidate is the other major diterpene found in coffee, kahweol, which has a broadly similar molecular structure to cafestol and is mostly absent in non-Arabica species.[40] This hypothesis is well-supported by correlations of the region's integral with those of other annotated cafestol and kahweol peaks elsewhere in the spectrum. This leads us to conclude that the complex peak profile around 2.60 ppm contains resonances from both cafestol and kahweol. Although not fully resolved at 600-MHz field strength, it is possible to access information on their relative abundances through the calculation of partial integrals, as illustrated in the figure.

Peak intercorrelations can help make decisions on which buckets to retain for further analysis. Where reliable annotation was absent, it was elected to keep only those that are highly correlated with other buckets in the dataset or that contain a clearly resolved singlet or multiplet. Figure 2 shows all bucket intercorrelations, with an associated dendrogram giving the hierarchical cluster tree constructed from the correlation matrix. Serving as a useful check on the integrity of bucketed data, the lowest level clusters are seen to arise from integrals assigned to the same compound.

Higher levels of clustering show associations between groups of compounds. These can broadly be divided into three, as indicated: cluster A comprises diglycerides, caffeine, sterols and 16-OMC. The latter two are recognised as markers for distinguishing coffee species[39,49]; the bitter alkaloid caffeine also tends to be present in larger amounts in canephora beans compared with Arabica. Also in this cluster are five unassigned buckets which, from their correlation patterns, likely comprise two distinct compounds: UL1 with two peaks at 5.66 and 6.48 ppm and UL2 with three at 6.75, 6.96 and 7.02 ppm.

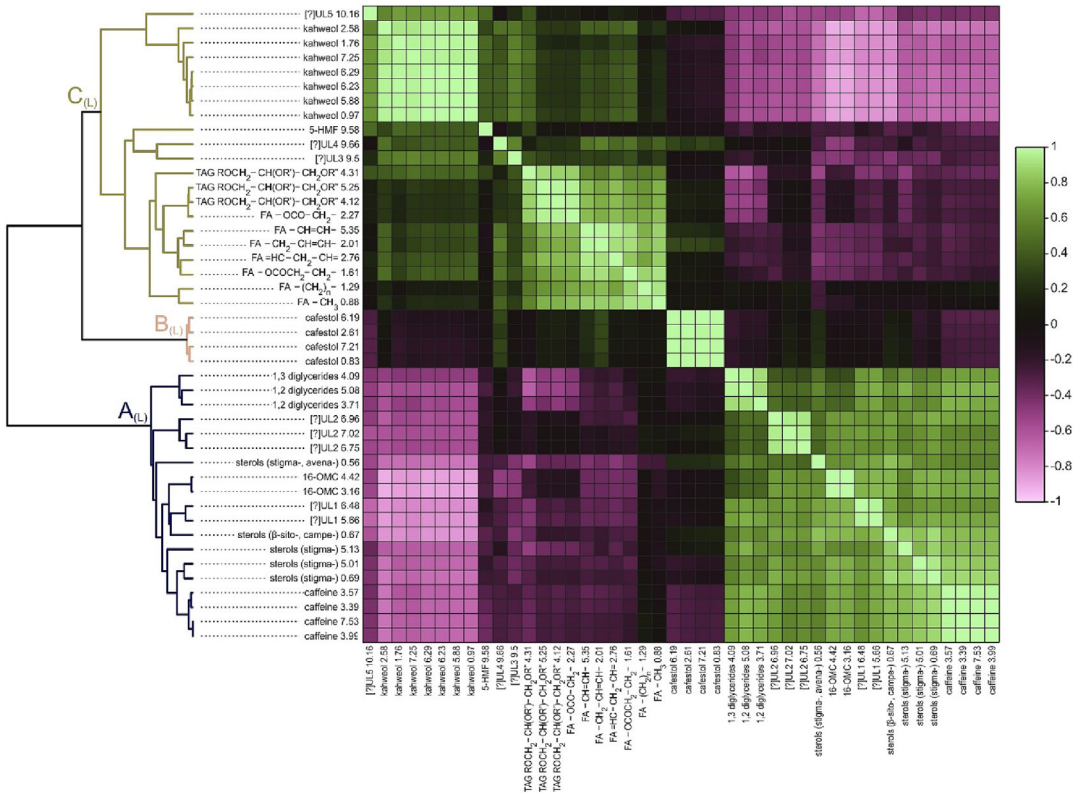


FIGURE 2 Heatmap showing the correlation matrix from the 600-MHz bucketed lipophilic dataset, with an associated dendrogram showing the agglomerative hierarchical cluster tree constructed from the correlation matrix using unweighted average distances (UPGMA) linkage. The blocks of high correlations along the leading diagonal correspond to the lowest level clusters: These arise from integrals assigned to the same compound. Three higher level clusters ($A_{(L)}$, $B_{(L)}$, $C_{(L)}$) indicate groups of compounds that vary similarly across the dataset. Cluster A comprises diglycerides, caffeine, sterols, 16-OMC and two unassigned compounds, UL1 and UL2. Cluster B comprises the buckets from cafestol only and is uncorrelated with all other compounds. Cluster C, especially kahweol and UL5, is generally anti-correlated with cluster A. The remaining compounds in cluster C include the triglycerides, fatty acids and 5-HMF.

From their chemical shifts, it is possible that these arise from as-yet unassigned compounds from the extensive sterolic fraction.[50,51]

Cluster A is uncorrelated with cluster B, which comprises cafestol only and is uncorrelated with all other compounds. Cluster C is generally anti-correlated with cluster A, especially kahweol and another unannotated peak (UL5) at 10.16 ppm. The remaining compounds in cluster C include the triglycerides, fatty acids and 5-HMF, an aldehyde-furan formed from reduction of sugars upon roasting.[52,53] The literature contains annotations of other features in the peroxide and aldehyde region (>7 ppm). However, in the present dataset, the profile across samples is highly inconsistent here, with peaks varying in number, size and precise location. This can be seen from the heatmaps in Figures 1 and S3.

3.2 | Exploring the 600-MHz hydrophilic extract spectra

The spectra of the hydrophilic extracts were processed similarly. The raw and log-compressed spectra of the Arabica/canephora blend and a heatmap of the whole dataset are shown in Figure S7. Compared with those of the lipophilic extracts, the raw signal intensities are lower and the spectra are more crowded. The main phenolic compounds in coffee beans are chlorogenic acids (CGAs, mainly 3-, 4- and 5-caffeoylquinic acid), which, along with their breakdown products such as N-methylpyridine, contribute significant peaks in multiple regions of the spectrum.[8,53,54] Other resonances arise from the alkaloid trigonelline, various polysaccharides and organic acids.[55] The latter are particularly susceptible to variation on the chemical shift scale. Caffeine also appears in the hydrophilic extracts, although its peaks in the middle-field region are highly mobile on the chemical shift scale and sometimes entangled with other resonances making them difficult to integrate; a partial integral approach was adopted in these cases, analogous to Figure S6 for cafestol/kahweol.

The hydrophilic spectra were likewise bucketed with the assistance of visualisation and registration tools, again retaining only those from which a reliable integral (or partial integral) could be obtained (Table S8). Of these, 46 were confidently annotated by cross-referencing with the literature, although published annotations for the hydrophilic fraction are notably less consistent than those for coffee oil. This may be because the compositional profile is more strongly influenced by factors such as the species, cultivar, or grade of bean, as well as variability arising from the roasting process.[10–12,56] A further six buckets are tentatively ascribed to three different unknown compounds (UA1–3). These could not be identified from reviewing the literature but are considered of interest because of their intercorrelation with other compounds. Experimental investigations (e.g. 2-D NMR) to establish their identity are precluded, however, because the sample collection is no longer available for further study.

The correlation heatmap and dendrogram for the hydrophilic extracts are given in Figure S9. The lowest level clusters are again from buckets arising from the same compounds. This is reassuring, as the greater peak movement and overlap in the hydrophilic spectra make the bucketing process more prone to errors. The next level of clustering highlights some interesting associations between compound classes. Cluster A comprising sucrose and CGAs is inversely correlated with cluster E containing N-methylpyridine along with an unassigned compound UA2. N-methylpyridine is formed by the breakdown of trigonelline during roasting and has been found to be inversely related to CGA concentration[57]; so by association, it is probable that UA2 is also a breakdown product. This could potentially be gamma-quinide,[58] although the presence of this compound and its chemical shift(s) are inconsistently reported in the literature.

Cluster B comprises a disparate mix of compounds present in green coffee beans (citric and malic acids, trigonelline) and breakdown products of sucrose and other polysaccharides (lactic and formic acids[59] and 5-HMF). The latter is notably anti-correlated with N-methylpyridine, even though both are associated with the roasting process; this can be explained by their different precursors. Clusters C and D contain further organic acids (acetic and quinic) along with caffeine.

3.3 | Relationships between the lipophilic and hydrophilic 600-MHz bucket datasets

3.3.1 | Using circular graphs

Because the lipophilic and hydrophilic spectra were obtained from the same collection of coffees, the bucketed datasets can be concatenated and explored collectively, beginning with an examination of their intercorrelations. Thus far, correlation values have been depicted as heatmaps, but other graphical forms are available that can elegantly illustrate these relationships. Undirected graphs are commonly used to represent connections in networks but are also well suited to the task of illustrating intercorrelations. In the present work, use has been made of a publicly available package for the

Matlab environment, circularGraph.[32] Screenshots taken directly from the tool are illustrated in Figure S10. A key attraction of this tool is its interactivity: Different networks can be easily activated, greatly facilitating exploration of the correlation matrix.

Figure 3 contains circular graphs showing positive intercorrelations in the concatenated dataset. Each of the 96 nodes corresponds to a bucket (with peak centres as labelled), and each of the edges has a correlation; only values significant at $p < 0.01$ (with Bonferroni multiple test adjustment) have been included, to avoid overcrowding the graph. Highlighted in Figure 3a is a single closed network that includes many of the between-fraction correlations. Caffeine is extracted in both the lipophilic and hydrophilic extracts, and all its buckets are found in this network. The other participating buckets from the hydrophilic extracts are from the FA (methylene) and UA3, which along with caffeine are highly correlated with sterols, aromatics and 16-OMC from the lipophilic extracts.

A second network (Figure 3b) links only the buckets from 5-HMF, which is present at very low concentrations in both fractions (see Figures 1 and S7; its resonances are barely visible even in the log-compressed profiles). Nevertheless, this breakdown product of the roasting process is regarded as potentially harmful and worthy of monitoring.[53,60] Finding that its buckets are well-correlated

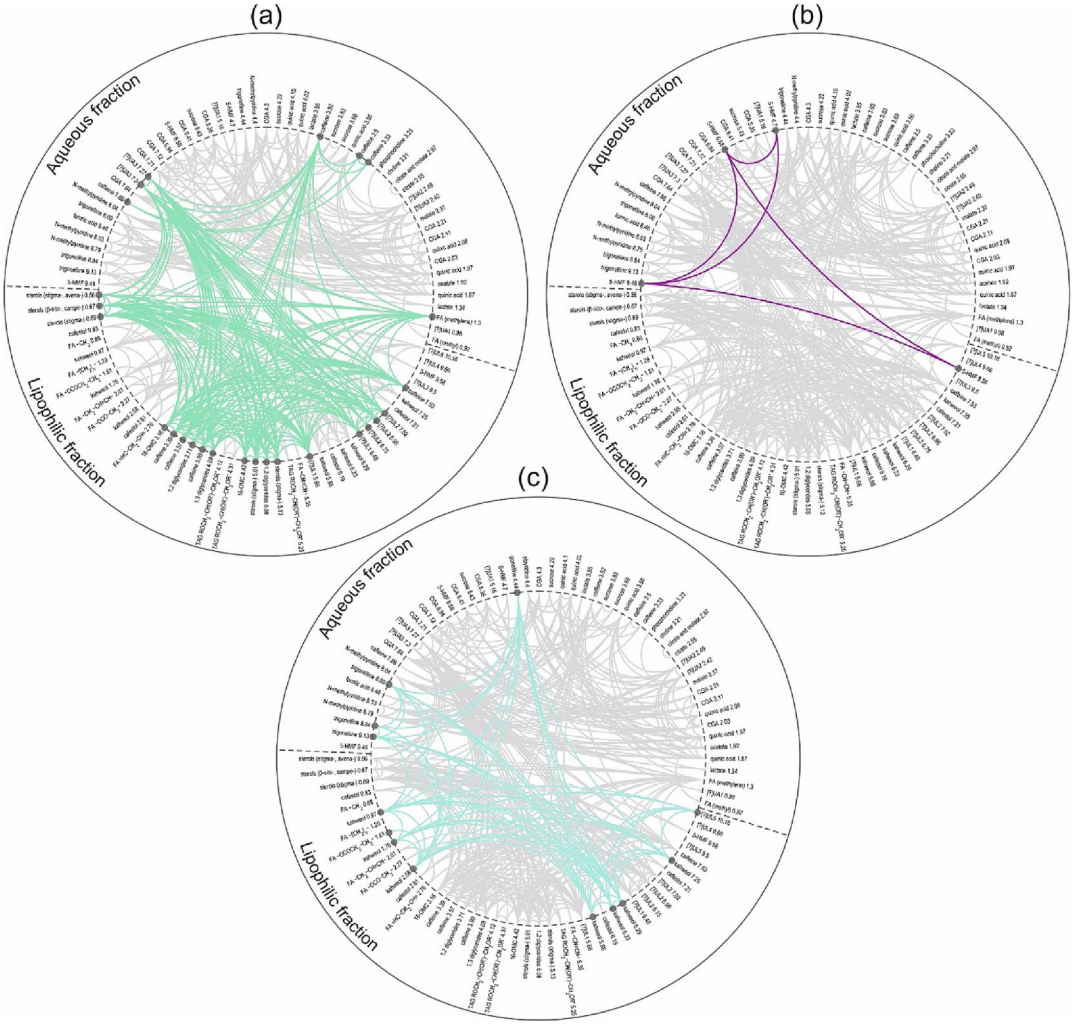


FIGURE 3 Circular graphs showing positive intercorrelations in the 600-MHz lipophilic and aqueous concatenated dataset. Each of the 96 nodes corresponds to a bucket, with peak centres as labelled, and each of the edges a correlation. Only values significant at $p < 0.01$ (Bonferroni multiple test adjustment) have been included, to avoid overcrowding the graph. Panel (a) highlights a closed network that includes many of the between-fraction correlations at this significance level, such as caffeine from both the lipophilic and hydrophilic extracts. Panel (b) shows a second network which contains only the buckets from 5-HMF, present at only very low concentrations in both fractions. The final closed network that links the two fractions is shown in panel (c) and includes all the buckets from trigonelline and kahweol in respectively the hydrophilic and lipophilic extracts. Both these compounds are known to be significantly higher in Arabica coffee.

across fractions reassures on the reliability of both the sample preparation and data integration procedures. The final closed network that links the two fractions (Figure 3c) includes all the buckets from trigonelline and kahweol in respectively the hydrophilic and lipophilic extracts. Both of these compounds are known to be significantly higher in Arabica coffee.[61]

3.3.2 | Biplots of the principal components

To examine these inter-species differences in the sample collection, the concatenated peak integrals were analysed using principal component analysis (PCA). As there are orders of magnitude differences in the scale of the integrals, both within and between the two datasets, the correlation matrix form of PCA was employed. This variant of PCA is routinely used in such circumstances; it is equivalent to standardising the buckets (mean-centring and scaling to unit variance) before analysis.

A biplot showing loadings and scores for the first two PC dimensions is given in Figure 4, which together account for ~57% of the dataset variance. The scores scatterplot highlights the contribution to this variance from the coffee species. In the first PC dimension, the Arabica and Arabica back-crosses cluster together and are clearly

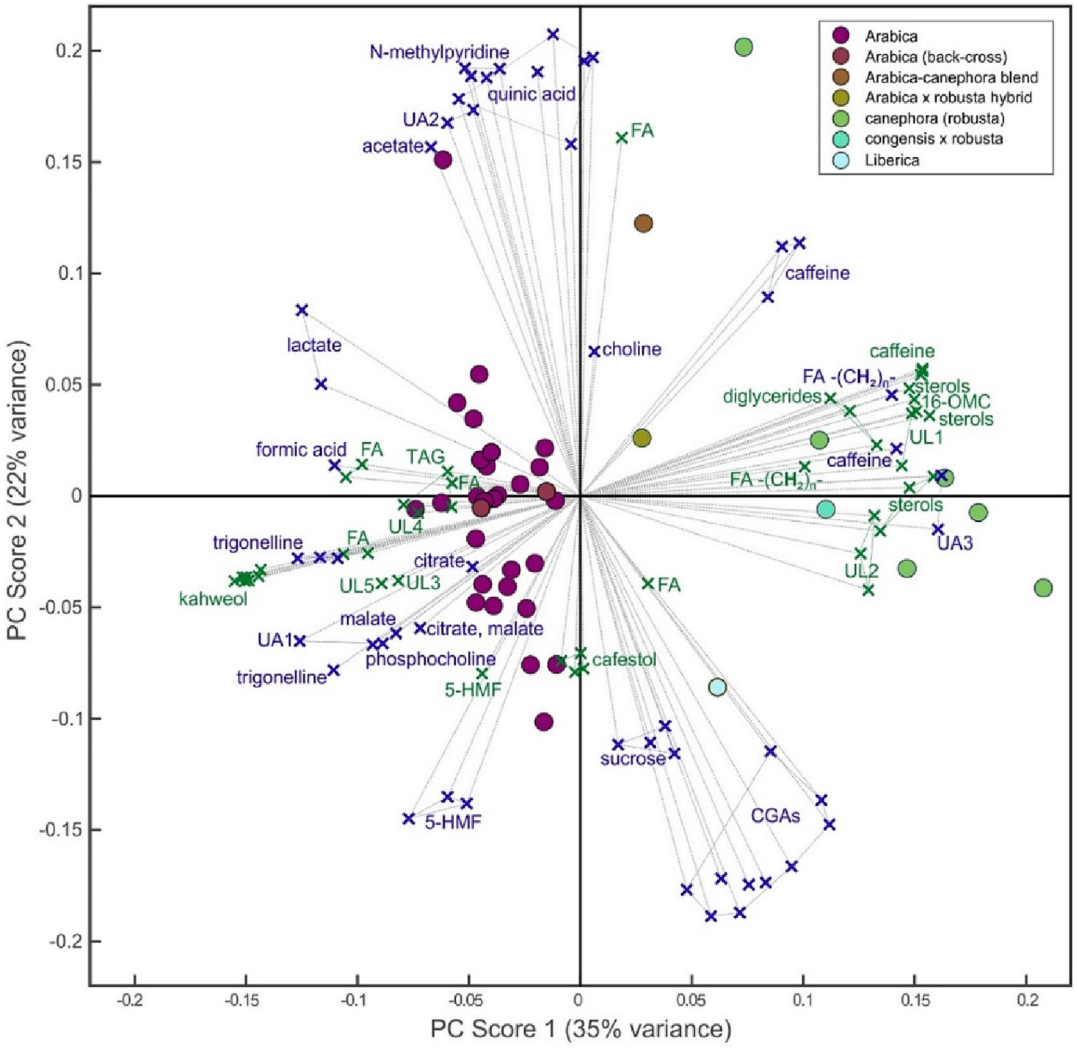


FIGURE 4 Two-dimensional PCA biplot illustrating the first and second scores and loadings obtained from the 600-MHz concatenated dataset. The scores scatterplot highlights the role of coffee species in determining the variance in the dataset. There is clear clustering of the Arabica (and back-cross) samples; the first PC axis is sufficient to distinguish these from the canephora and congensis x robusta samples. Note also the intermediate placement of the Arabica/canephora blend, hybrid and Liberica samples with respect to this axis. The association of the various metabolites with species is depicted by the loading weights, which are shown for each bucket by the vectors extending from the origin. For buckets in the lipophilic dataset, the terminal crosses are green, and crosses are blue for the hydrophilic metabolites; convex hulls are used to indicate loadings from the same compound within each fraction.

separated from the canephora and congensis samples; the Liberica, hybrid and blend are found between the two groups. The association of the various metabolites with this species distinction is depicted by the loading weights, shown for each bucket by vectors from the origin. The buckets most associated with Arabica are from kahweol, trigonelline, 5-HMF (in both fractions), most of the organic acids potentially including UA1, and other unidentified compounds (likely aldehydes) UL3, UL4 and UL5. Diametrically opposite are caffeine and methylene (in both fractions), 16-OMC, diglycerides, sterols and the unidentified (possible sterols) UL1, UL2 and potentially UA3. The latter group of compounds maps directly onto the closed network in the discussion of Figure 3a. There is little association with species in the second PC dimension; instead, it may relate to the roasting process, with CGAs and sucrose weighted in opposition to the breakdown products N-methylpyridine and quinic acid. The PC biplot presented here is qualitatively similar to the biplot shown in a study of 38 coffees of different species by de Souza and Benassi,[61] although fewer compounds were studied in that work. This suggests that the combined lipophilic and hydrophilic bucketed datasets have provided a good, general representation of the key characteristics of ground roast coffee composition.

3.4 | Benchtop NMR spectra

3.4.1 | Correlation with 600-MHz NMR bucket dataset

This section examines whether similarly detailed information can be obtained from 60-MHz NMR spectra of the same collection of samples. Lipophilic extracts only are considered, as the issue of water suppression is more difficult to address at benchtop field strengths because of the greater intrinsic width of spectral bands. For the same reason, 60-MHz spectra of complex biological samples are not generally amenable to bucketing either, apart from a few serendipitous cases. One such exception concerns the 16-OMC peak in coffee, which can be quantified and used as a marker of non-Arabica species. The spectra presented here were used for that purpose previously[28] but have never been explored as whole spectral profiles.

The collection of raw 60-MHz spectra is shown in Figure 5a. The wide range of peak intensities limits what can be conveyed by this representation of the dataset, in contrast to Figure 5b where the spectra have been log-compressed. In both panels, the spectra are colour coded by coffee species. In the log-compressed data, some of the small diterpene peaks have been amplified sufficiently for inter-species differences to be seen. These are annotated in the figure; an assignment of other features in the 60-MHz spectrum is given in Figure S11. In panel (c), the log-compressed dataset is presented as a heatmap, analogous to that in Figure 1c. This shows the highlighted species distinctions in the kahweol and 16-OMC regions very clearly. The spectra have been likewise registered using COA, and the chemical shift scale attached using a signal from the triglycerides common to all samples. Sideways movement of peaks is as much of an issue at low field strengths, although the greater peak widths mask the effect to an extent. Panel (d) presents a heatmap of the correlations between the 60-MHz raw spectral intensities and the 600-MHz bucketed data, also from the lipophilic extracts. To assist with the examination of the heatmap, it is aligned with the chemical shift scale of panel (c). The colour mapping indicates that there are very high correlations (the brightest green pixels in the heatmap) for many of the 600-MHz peak integrals with the 60-MHz intensities.

Circular graphs can again assist with detailed inspection of these values. For example, in Figure 6a, a graph has been constructed from the bucketed data and the subset of 60-MHz spectral data points that have a correlation of >0.9 with any bucket. The network highlighted is for kahweol, and it indicates two main regions (5.94–5.96 ppm and 6.26–6.35 ppm) in the 60-MHz spectrum that are highly correlated with the 600-MHz kahweol buckets. Reasonably distinct peaks are found here in the 60-MHz spectra (see Figure 5b), and we can conclude that these correspond to 600-MHz kahweol buckets with peak centres of 5.88, 6.23 and 6.29 ppm. The third region indicated by the correlation analysis is 1.63–1.66 ppm, which is heavily overlapped in the 60-MHz spectra. It is plausible that this pinpoints information from kahweol, because there is a 600-MHz bucket centred at 1.76 ppm. Note, however, that there is not always direct correspondence in chemical shift values at different field strengths.[62]

3.4.2 | CCA

Multivariate statistical methods are well-established for dealing with spectral data. One such method is CCA, which seeks out shared information in two multivariate datasets and expresses it as successive, maximally correlated linear combinations of variables. It is a powerful method for analysis and visualisation of two collections of observations made on the same samples. However, CCA cannot be applied directly to high-dimensional datasets (i.e. with more variables than observations, such as in the present study). Further, CCA easily tends

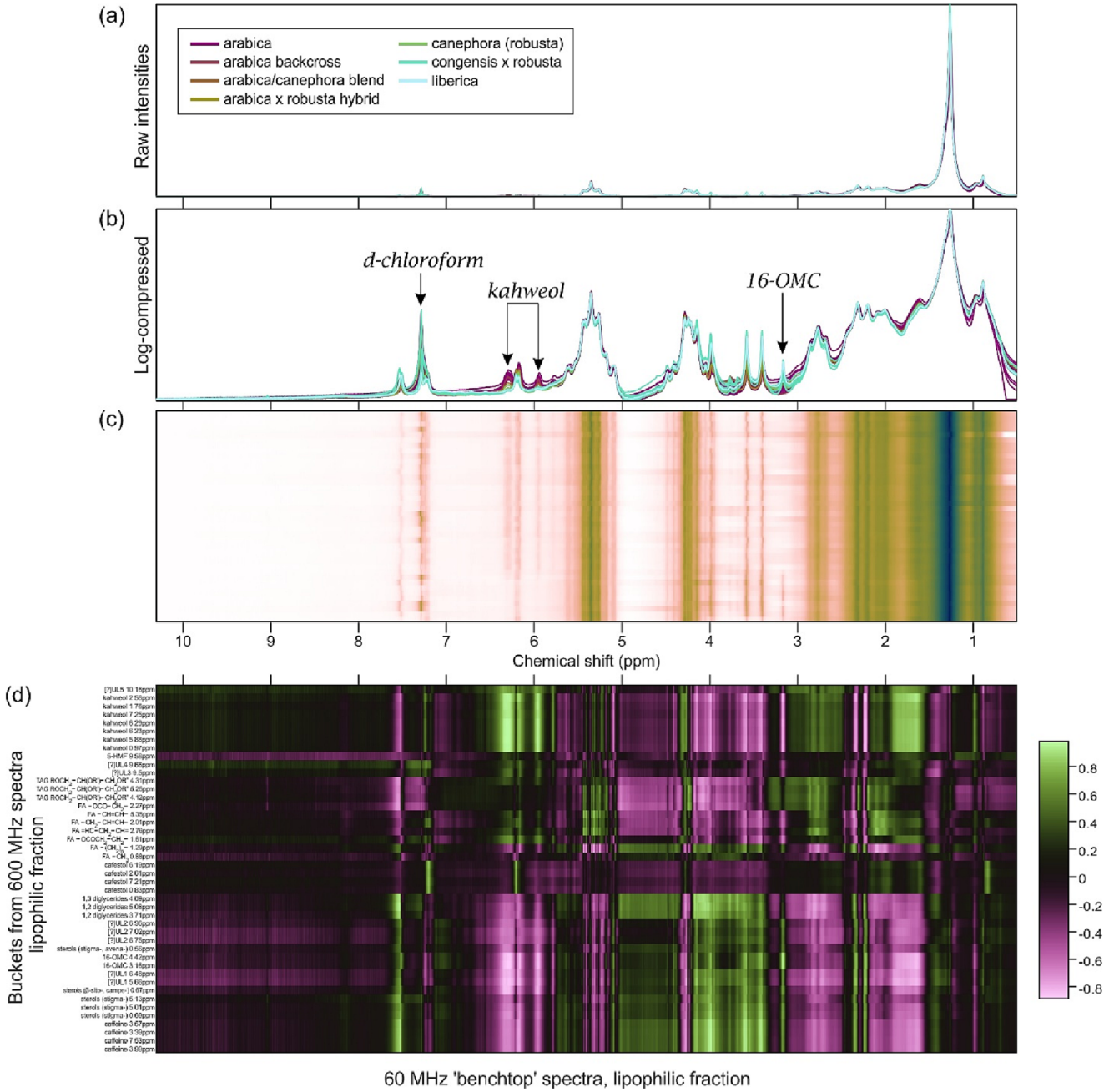


FIGURE 5 Panel (a) shows the 60-MHz ^{1}H NMR spectra of the lipophilic extracts from the collection of coffee sample with raw intensities as acquired; in panel (b), the same spectra have been log-compressed, revealing the information present in the spectral profile in a more accessible form. The spectra have been colour coded here according to coffee species, and the clearest peaks for distinguishing between them are annotated. In panel (c), the log-compressed dataset is presented as a heatmap. This plainly shows the species distinctions in the kahweol and 16-OMC peaks. Panel (d) presents a heatmap of the correlations between the 60-MHz spectra and the 600-MHz lipophilic buckets, aligned with the chemical shift scale of panel (c). There are many high correlations, indicated by the brightest green pixels; however, not all 600-MHz buckets are expected to point to distinct features in the 60-MHz spectra, in which the information content of a typical peak is heavily entangled with that of its equally broad neighbours.

towards 'overfitting', conceptualised as looking too hard for something in the data that may not actually be there. Overfitting increases with the number of variables used, even for data that are not high-dimensional. One workaround is to use PCA as a dimension-reducing pre-treatment and pass just a few PC scores to CCA, although this likely means that we are discarding at least some pertinent information and, in the worst case, could even be throwing it all away; the subset of PCs may simply not contain the information common to both datasets.

These issues can be circumvented by modifying the CCA calculations. One such variant, large-scale sparse kernel CCA[33] uses regularization to deal with the rank limitation and can be applied directly to high-dimensional data. Implemented in cross-validation form to mitigate overfitting, it is used in the present work to obtain an estimate of the canonical correlation between the 60- and 600-MHz data. Outputs from the procedure are presented graphically in Figure 6b. Both the loadings and scores plots show high correspondence between the two datasets. This is an important finding, because whereas the original 600 MHz were condensed into compositionally meaningful buckets, the 60-MHz spectra were treated as full spectral profiles. The analysis shows that these contain almost all of the same information as the bucketed data, albeit in overlapped form. Further, the scores plot shows that this shared information clearly distinguishes between coffee species. Collectively, the results presented in Figure 6 make a convincing case for the use of benchtop NMR for high-throughput profiling of the lipophilic fraction for the purpose of coffee species differentiation and authentication; this will be discussed further below.

3.4.3 | Magnitude spectra: an alternative spectral form

To avoid operator subjectivity in high throughput NMR studies, it is desirable to automate the phase correction. Nevertheless, some unwanted spectral variation is introduced by this step, particularly in benchtop spectra. Even very small changes in the magnet temperature affect the shimming and field homogeneity at the probe. Downstream this causes variation in the peak shape, symmetry, width and baseline in the phase-corrected spectra. An alternative spectral form, the magnitude spectrum, is calculated from the real and imaginary parts of the Fourier-transformed FIDs and offers a direct measure of induced magnetization with no phase correction step involved. The raw phase-corrected and magnitude spectra of the 60-MHz lipophilic extracts are compared in Figure S12. The qualitative difference between the two is obvious. Log compression can again be applied to magnitude spectra (panel c) to reduce the dynamic range and aid heatmap visualisation (panel d), helped by the fact that magnitude spectra contain only non-negative values. It is recognised that using this spectral form is advantageous

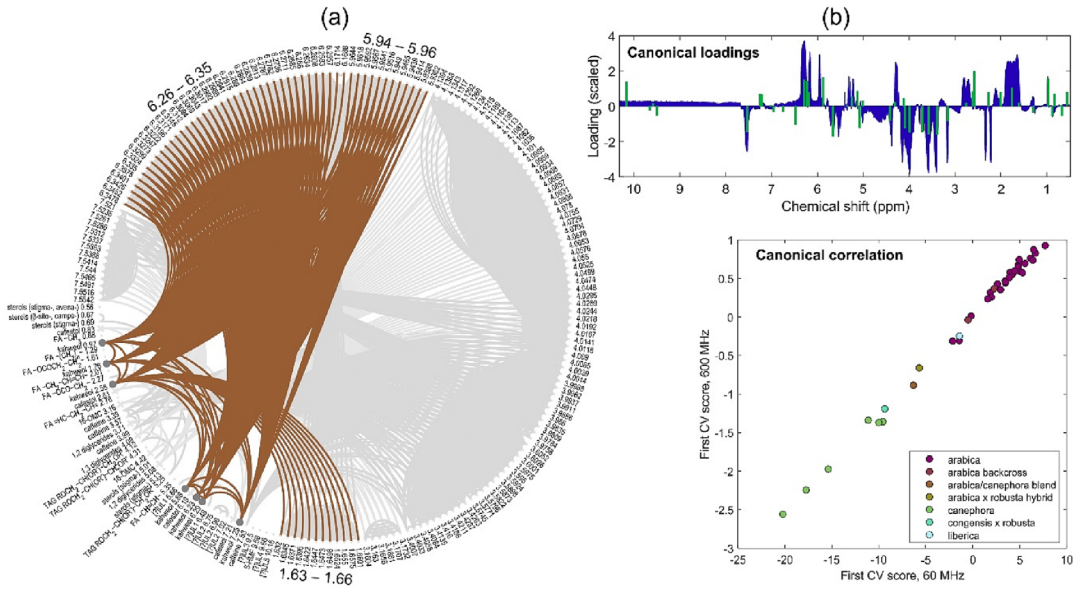


FIGURE 6 Panel (a) shows a circular graph constructed from all correlations greater than 0.9 between the 60-MHz spectral data points and the 600-MHz lipophilic extract buckets. The highlighted network is for kahweol. It identifies two main regions (5.94–5.96 ppm and 6.26–6.35 ppm) where distinct peaks are found in the 60-MHz spectra that map onto 600-MHz kahweol buckets. A third highly correlated region (1.63–1.66 ppm) is found to be heavily overlapped in the benchtop spectra. Panel (b) shows the first loadings and scores obtained by sparse CCA. In both, there is high correspondence between the two datasets, with a canonical correlation of ~0.99 (cross-validated). This is important because whereas the original 600 MHz were condensed buckets, the 60-MHz spectra were treated as full spectral profiles; this analysis shows that the latter nevertheless contain almost all of the same information as the bucketed data, just in overlapped form. Further, the scores plot shows that this shared information clearly distinguishes between coffee species. Note that the grouping information is not utilised in the CCA transformation at all; rather it is added post hoc as colour coding of the data points.

for pattern recognition,[63] including in its application to benchtop spectra.[64]

Another interesting form of post-processing is to take the first derivative of the magnitude spectra (panel e). This functions as a high-pass filter, directly extracting information on spectral peaks which in the magnitude form appear as 'ripples' on a broader underlying curve. Maxima in the derivative spectrum identify the locations of maximum ripple gradients and also correspond to peak centres in the phase-corrected spectrum. The use of derivative spectra has been reported for treating high-field NMR spectra,[65] although it is much more routine in other forms of spectroscopy that produce broad overlapped peaks, such as FTIR or UV–Vis[66] where whole collections of spectra are processed in their derivative form. Given their qualitative similarity, there is likely similar potential for handing benchtop NMR data in this way, although proceeding further down this route is beyond the scope of the present work.

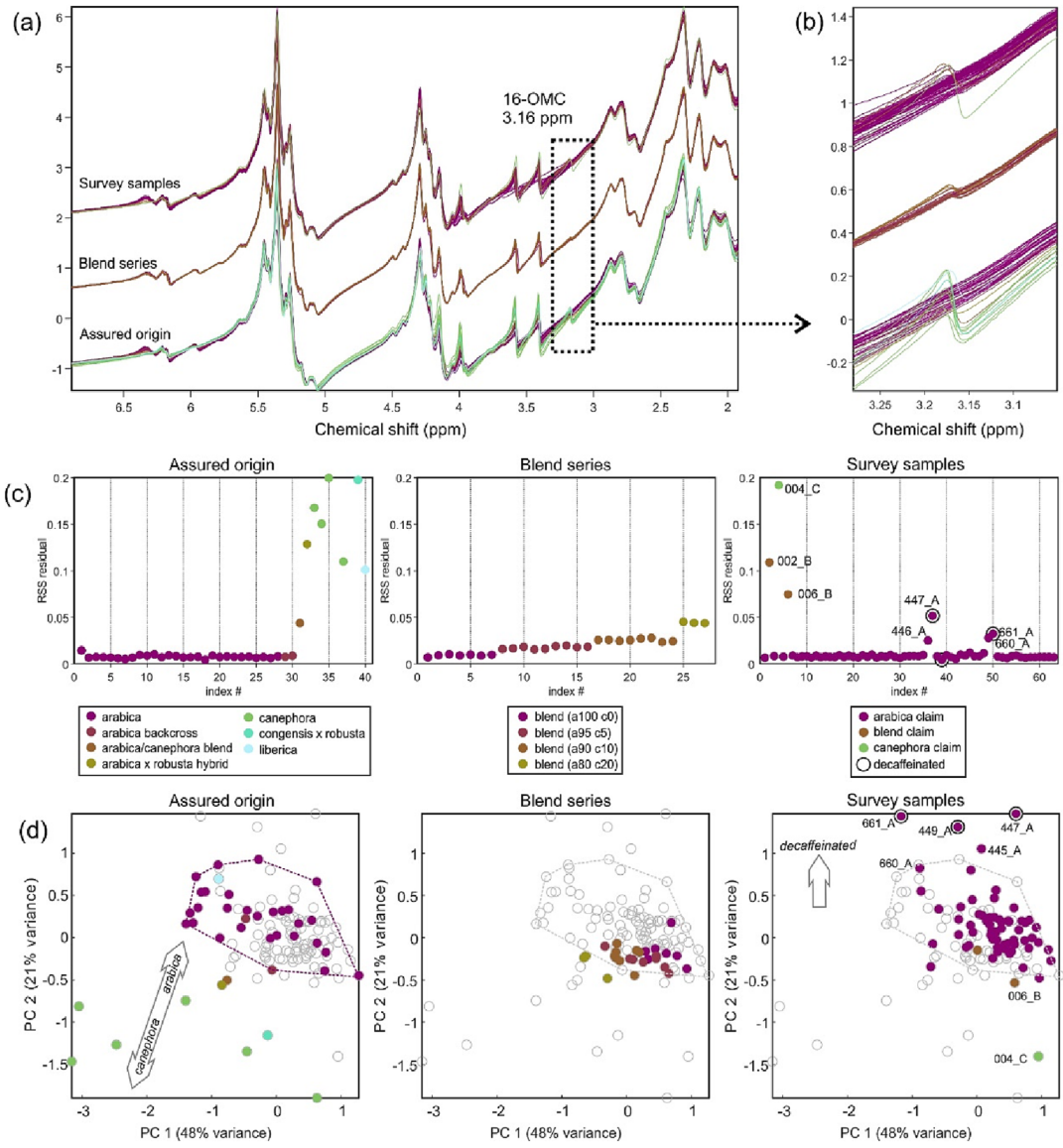


FIGURE 7 Panel (a) shows 60-MHz magnitude spectra of lipophilic extracts from the three coffee collections (assured origin, blend series and survey samples), offset for clarity, SNV-treated and focussing on the chemical shift region containing resonances from compounds that are known to distinguish coffee species. Panel (b) shows a magnification of the 3.16 ppm feature from 16-OMC which is strongly associated with non-Arabica coffee species. The RSS from regressing this region onto a local linear baseline gives a measure of the size of the feature, shown for each of the data collections in panel (c). PCA was applied to concatenated dataset from all three collections. The first two scores are shown in panel (d). Two directions in the PCA subspace are associated with aspects of coffee composition, as marked.

3.4.4 | Benchtop NMR for coffee species authentication

The 60-MHz phase-corrected spectra have been previously used in an authenticity test based on quantifying the 3.16 ppm 16-OMC peak.[28] Establishing a normal range for this peak area in authentic Arabica coffees gave a straightforward test for the likely presence of non-Arabica species. We have found that the same information can readily be accessed in the magnitude spectra. Figure 7a shows the lipophilic 60-MHz magnitude spectra in the region between 2 and 6.8 ppm, which contains resonances from the majority of compounds that distinguish coffee species. Also shown (offset in the figure) are the two additional data collections, from serial blends of assured Arabica and canephora beans, and from 63 retail and hospitality coffee samples surveyed at the time of the method development (sample listings are given in Table S2).

The inset panel (Figure 7b) shows a magnification of the 3.16 ppm region. There is a clear distinction in the spectral profile between the Arabica and other species. A simple estimate of the feature size is given by the sum of squared residuals (RSS) between this profile and a linear baseline. This is shown for each of the data collections in panel (c). For the assured origin samples, the most obvious is the vast difference in the RSS between the Arabicas and other species; further, the RSS is shown to be linearly correlated with the 16-OMC peak area calculated from both 60- and 600-MHz phase-corrected spectra (Figure S13). The outcomes for the blend series confirm that the size of the feature measured in this way monotonically increases with canephora concentration. The outcomes for the survey samples are also in agreement with previously reported work,[28] in which several clearly suspicious '100% Arabicas' had 16-OMC contents more consistent with Arabica/canephora blends. This clearly demonstrates the utility of the magnitude spectral form should phase correction be sub-optimal.

Finally, the magnitude spectra are amenable to whole spectral processing. The region as shown in Figure 7a in all three 60-MHz collections was passed to PCA. An amount of log compression was applied to maximise the information content in the first two dimensions, which together account for ~70% variance. The scores are shown in the three panels of Figure 7d, for clarity each highlighting a different sample collection. A convex hull is marked in all panels around the Arabicas of assured origin. Although not intended to function as a classifier, this outlines the extent of the authentic Arabicas in the two dimensions under consideration. The non-Arabica species and blends with higher canephora contents are seen to fall outside it, as do three of the survey samples confirmed as decaffeinated. Of these, two were deemed suspicious from their 16-OMC content as measured by the 3.16 ppm peak[28]; the two other highly anomalous samples in that analysis also appear as edge cases with respect to the convex hull.

This analysis demonstrates that the magnitude spectra are capable of conveying detailed information on differences between the samples. The raw and SNV magnitude spectra from the assured origin samples are made available to accompany this manuscript, as are the SNV magnitude spectra from the blend and survey collection. In addition, Table S14 provides a selection of Matlab code snippets used in the analysis as presented here.

4 | CONCLUSIONS

This paper demonstrates the usefulness of graphical tools and chemometric methods in the untargeted analysis of NMR spectra from complex biological samples. The combination of log compression and heatmapping enabled effective representation of whole datasets, aiding in a 600-MHz metabolomics study of the lipo- and hydrophilic extracts from roasted coffees. The approaches were also useful in fine-tuning the integration window for each bucket by local peak registration using COA.

Establishing a universal annotation of coffee spectra is challenging, especially for the hydrophilic fraction, because of natural product variability and the breakdown of compounds during the roasting process. Tools for visualizing bucket and dataset intercorrelations helped to obtain, for the collection of samples under consideration, reliable cross-validated assignments for around 100 buckets across the two fractions. These were grouped into clusters that were found to correspond to known aspects of coffee biochemistry, suggesting that the concatenated 600-MHz bucketed datasets yielded a meaningful coffee composition profile.

The 60-MHz lipophilic spectra contain much broader, overlapped peaks. Although visualisation tools to deal with the intensity dynamic range remain useful, the peak integration approach used in the 600-MHz analysis is of limited utility. However, focusing exclusively on bucketing is arguably a cul-de-sac, making too many problems inaccessible to benchtop NMR. Indeed, whole-spectrum analysis of the 60-MHz data showed that several spectral regions were highly correlated with the 600-MHz lipophilic buckets, including those containing resonances from minor constituents such as diterpenes. This finding was supported by CCA, which gave a canonical correlation between the two datasets of ~0.99, indicating a very high proportion of common information. Because of extensive overlapping, distinct peaks were not visible in

all of these highly correlated regions, although post-processing by taking the first derivative of the magnitude spectral form showed that the peak information is nevertheless embedded in the 60-MHz data.

The magnitude spectral form was also used to examine the 16-OMC resonance at 3.16 ppm, which appears as a ripple on a near-linear baseline. This may be quantified by its squared residual difference from a straight line and is found to correlate highly with conventional peak integrals calculated in phase-corrected spectra. Finally, the collection of 60-MHz magnitude spectra was treated with PCA. The representation of inter-sample differences obtained corresponded closely to that from the concatenated 600-MHz bucketed datasets. This suggests that 60-MHz NMR spectra, including their magnitude form, can be treated with appropriate data handling tools to offer a lower cost, lower tech route to informative metabolomics studies.

ACKNOWLEDGEMENTS

The author gratefully acknowledges the support of the Biotechnology and Biological Sciences Research Council (BBSRC). This work was funded by the BBSRC Core Capability Grant number BB/CCG1860/1. Thanks also to my colleagues at QIB for inspiring me with their varied attempts to visualise the microbiome and Sarah Jane Wilson for her helpful advice on colour mapping.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/mrc.5373.

ORCID

E. Kate Kemsley https://orcid.org/0000-0003-0669-3883

REFERENCES

[1] T. A. van Beek, *Phytochem Anal* **2021**, *32*(1), 24.
[2] Y. Gunning, A. J. Jackson, J. Colmer, F. Taous, M. Philo, R. M. Brignall, T. el Ghali, M. Defernez, E. K. Kemsley, *Magn Reson Chem* **2020**, *58*(12), 1177.
[3] S. Danoun, S. Balayssac, V. Gilard, R. Martino, M. Malet-Martino, *J Pharm Biomed Anal* **2023**, *223*, 10.
[4] J. F. Araneda, I. W. Burton, M. Paleologou, S. D. Riegel, M. C. Leclerc, *Can J Chem* **2022**, *100*(11), 799.
[5] R. Burger, S. Lindner, J. Rumpf, X. T. Do, B. W. Diehl, M. Rehahn, Y. B. Monakhova, M. Schulze, *J Pharm Biomed Anal* **2022**, *212*, 8.
[6] J. Gracia-Vitoria, M. Rubens, E. Feghali, P. Adriaensens, K. Vanbroekhoven, R. Vendamme, *Ind Crop Prod* **2022**, *176*, 7.
[7] D. Galvan, A. A. Tanamati, F. Casanova, E. Danieli, E. Bona, M. H. Killner, *Food Chem* **2021**, *365*, 10.
[8] F. F. Wei, K. Furihata, M. Zhang, T. Miyakawa, M. Tanokura, *J Agric Food Chem* **2011**, *59*(17), 9065.
[9] N. D'Amelio, E. de Angelis, L. Navarini, E. Schievano, S. Mammi, *Talanta* **2013**, *110*, 118.
[10] D. J. Kwon, H. J. Jeong, H. Moon, H. N. Kim, J. H. Cho, J. E. Lee, K. S. Hong, Y. S. Hong, *Food Res Int* **2015**, *67*, 175.
[11] F. M. Mehaya, A. A. Mohammad, *Heliyon* **2020**, *6*(11), e05508.
[12] F. F. Wei, K. Furihata, M. Koda, F. Hu, T. Miyakawa, M. Tanokura, *J Agric Food Chem* **2012**, *60*(4), 1005.
[13] K. Williamson, T. Banker, X. Zhao, J. Ortega-Anaya, R. Jimenez-Flores, Y. Vodovotz, E. Hatzakis, *LWT–Food Sci Technol* **2022**, *165*, 9.
[14] A. L. Davis, Y. Cai, A. P. Davies, J. R. Lewis, *Magn Reson Chem* **1996**, *34*(11), 887.
[15] G. Le Gall, I. J. Colquhoun, M. Defernez, *J Agric Food Chem* **2004**, *52*(4), 692.
[16] F. H. Larsen, F. van den Berg, S. B. Engelsen, *J Chemometr* **2006**, *20*(5), 198.
[17] M. D. Guillen, A. Ruiz, *Trends Food Sci Technol* **2001**, *12*(9), 328.
[18] O. Al-Jowder, F. Casuscelli, M. Defernez, E. K. Kemsley, R. H. Wilson, I. J. Colquhoun, High resolution NMR studies of meat composition and authenticity, in *International conference on applications of magnetic resonance in food science*, Averio, Portugal **2000**.
[19] G. Le Gall, M. Puaud, I. J. Colquhoun, *J Agric Food Chem* **2001**, *49*(2), 580.
[20] A. Yilmaz, N. T. Nyberg, P. Mølgaard, J. Asili, J. W. Jaroszewski, *Metabolomics* **2010**, *6*(4), 511.
[21] Y. Gunning, K. S. Davies, K. Kemsley, *Food Chem* **2023**, *404*, 9.
[22] G. S. Remaud, Y. L. Martin, G. G. Martin, G. J. Martin, *J Agric Food Chem* **1997**, *45*(3), 859.
[23] S. Rattan, A. K. Parande, V. D. Nagaraju, G. K. Ghiwari, *Environ Sci Pollut Res* **2015**, *22*, 6461.
[24] M. V. de Moura Ribeiro, N. Boralle, H. R. Pezza, L. Pezza, A. T. Toci, *J Food Compos Anal* **2017**, *57*, 24.
[25] E. Belmonte-Sanchez, R. Romero-Gonzalez, A. G. Frenich, *J Sci Food Agric* **2021**, *101*(9), 3541.
[26] R. Consonni, L. R. Cagliani, *Magn Reson Chem* **2019**, *57*(9), 558.
[27] M. Defernez, E. Wren, A. D. Watson, Y. Gunning, I. J. Colquhoun, G. le Gall, D. Williamson, E. K. Kemsley, *Food Chem* **2017**, *216*, 106.
[28] Y. Gunning, M. Defernez, A. D. Watson, N. Beadman, I. J. Colquhoun, G. le Gall, M. Philo, H. Garwood, D. Williamson, A. P. Davis, E. K. Kemsley, *Food Chem* **2018**, *248*, 52.
[29] Y. Gunning, M. Defernez, E. K. Kemsley, *MethodsX* **2023**, *10*, 102132.
[30] F. Kaspar, F. Crameri, *Angew Chem Int Ed* **2022**, *61*(16), e202114910.
[31] F. Crameri, *Scientific colour maps (7.0.1)*. Zenodo **2021**. https://doi.org/10.5281/zenodo.5501399
[32] *circularGraph*. **2023**, Paul Kassebaum: GitHub.
[33] V. Uurtio, S. Bhadra, J. Rousu. *Large-scale sparse kernel canonical correlation analysis*. in *International Conference on Machine Learning*. 2019. PMLR.
[34] *Sparse canonical correlation analysis*. **2023**.
[35] QIBChemometrics, *Coffee_NMR_Spectra*. **2023**, Zenodo.
[36] H. M. Parsons, C. Ludwig, U. L. Günther, M. R. Viant, *Bmc Bioinf* **2007**, *8*, 16.

[37] P. V. Purohit, D. M. Rocke, M. R. Viant, D. L. Woodruff, *Omics-a J Integr Biol* **2004**, *8*(2), 118.
[38] Y. M. Yusoff, G. Abbott, L. Young, R. A. Edrada-Ebel, *Metabolites* **2022**, *12*(1), 25.
[39] W. Kamm, F. Dionisi, L. B. Fay, C. Hischenhuber, H. G. Schmarr, K. H. Engel, *J am Oil Chem Soc* **2002**, *79*(11), 1109.
[40] C. Finotello, C. Forzato, A. Gasparini, S. Mammi, L. Navarini, E. Schievano, *Food Control* **2017**, *75*, 62.
[41] T. N. Vu, K. Laukens, *Metabolites* **2013**, *3*(2), 259.
[42] F. Crameri, G. Shephard, P. Heron, *batlow–a Scientific colour map*.
[43] S. A. A. Sousa, A. Magalhaes, M. M. C. Ferreira, *Chemom Intel Lab Syst* **2013**, *122*, 93.
[44] G. Tomasi, F. Savorani, S. B. Engelsen, *J Chromatogr a* **2011**, *1218*(43), 7832.
[45] J. W. H. Wong, C. Durante, H. M. Cartwright, *Anal Chem* **2005**, *77*(17), 5655.
[46] K. Williamson, E. Hatzakis, *Food Res Int* **2019**, *119*, 683.
[47] D. W. Lachenmeier, J. Teipel, A. Scharinger, T. Kuballa, S. G. Walch, F. Grosch, M. Bunzel, A. O. Okaru, S. Schwarz, *J AOAC Int* **2020**, *103*(2), 306.
[48] Y. B. Monakhova, W. Ruge, T. Kuballa, M. Ilse, O. Winkelmann, B. Diehl, F. Thomas, D. W. Lachenmeier, *Food Chem* **2015**, *182*, 178.
[49] F. Pablos, A. G. Gonzalez, M. J. Martin, M. S. Valdenebro, M. Leon-Camacho, *Analyst* **1999**, *124*(7), 999.
[50] H. Cizkova, V. Soukupova, I. Voldrich, R. Sevcik, *J Food Nutr Res* **2007**, *46*(1), 28.
[51] F. Carrera, M. León-Camacho, F. Pablos, A. G. González, *Anal Chim Acta* **1998**, *370*(2–3), 131.
[52] M. Murkovic, N. Pichler, *Mol Nutr Food Res* **2006**, *50*(9), 842.
[53] G. del Campo, I. Berregi, R. Caracena, J. Zuriarrain, *Talanta* **2010**, *81*(1–2), 367.
[54] F. F. Wei, K. Furihata, F. Hu, T. Miyakawa, M. Tanokura, *Magn Reson Chem* **2010**, *48*(11), 857.
[55] R. Consonni, L. R. Cagliani, C. Cogliati, *Talanta* **2012**, *88*, 420.
[56] M. A. Farag, A. Zayed, I. E. Sallam, A. Abdelwareth, L. A. Wessjohann, *Foods* **2022**, *11*(6), 18.
[57] U. Boettler, N. Volz, G. Pahlke, N. Teller, C. Kotyczka, V. Somoza, H. Stiebitz, G. Bytof, I. Lantz, R. Lang, T. Hofmann, D. Marko, *Mol Nutr Food Res* **2011**, *55*(5), 798.
[58] N. Happyana, E. Hermawati, Y. M. Syah, E. H. H. Hakim, *Curr Res Nutr Food Sci* **2020**, *8*(2), 479.
[59] S. E. Yeager, M. E. Batali, J. X. Guinard, W. D. Ristenpart, *Crit Rev Food Sci Nutr* **2023**, *63*(8), 1010.
[60] N. Villalon-Lopez, J. I. Serrano-Contreras, D. I. Téllez-Medina, L. G. Zepeda, *Food Res Int* **2018**, *106*, 263.
[61] R. Souza, M. T. Benassi, *J Braz Chem Soc* **2012**, *23*, 1347.
[62] W. Jakes, A. Gerdova, M. Defernez, A. D. Watson, C. McCallum, E. Limer, I. J. Colquhoun, D. C. Williamson, E. K. Kemsley, *Food Chem* **2015**, *175*, 1.
[63] P. D. Harrington, X. Y. Wang, *J Anal Test* **2017**, *1*(2), 11.
[64] Y. Gunning, F. Taous, T. El Ghali, J. D. Gibbon, E. Wilson, R. M. Brignall, E. K. Kemsley, *Food Chem* **2022**, *370*, 10.
[65] A. D. Maher, D. Crockford, H. Toft, D. Malmodin, J. H. Faber, M. I. McCarthy, A. Barrett, M. Allen, M. Walker, E. Holmes, J. C. Lindon, J. K. Nicholson, *Anal Chem* **2008**, *80*(19), 7354.
[66] R. M. P. da Silva, J. F. Mano, R. L. Reis, *Macromol Chem Phys* **2008**, *209*(14), 1463.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: E. K. Kemsley, *Magn Reson Chem* **2023**, 1. https://doi.org/10.1002/mrc.5373