


RESEARCH ARTICLE

Open Access



# The use of infrared spectroscopy and chemometrics to investigate deterioration in vegetable tanned leather: potential applications in heritage science

Elizabeth Dickinson<sup>1\*</sup> and Kirsty E. High<sup>2\*</sup> 

## Abstract

Vegetable tanned leather presents a unique challenge to conservators and curators of heritage collections, as little is known about how its physical and chemical properties change upon deterioration. Developing a better understanding of deterioration processes would be incredibly valuable in informing the conservation, storage, and restoration of leather objects. Fourier Transform infrared spectroscopy (FTIR) used with attenuated total reflectance (ATR) is increasingly applied in the heritage sector due to its relative ease of application and potential to be non-destructive. However, whilst FTIR has been applied successfully to the understanding of deterioration in other protein-based materials such as parchment, its application to the analysis of leather has been limited, largely due to the highly complex spectra obtained. Here, we have developed multivariate statistical methods for the analysis of FTIR data obtained from a time-series of leather samples artificially degraded at different pH values. Principal component analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA) and k-means clustering, when used together, are demonstrated as powerful tools in identifying early subtle differences in the FTIR spectra as leather degrades, identifying differences occurring over time and between different environmental conditions. We show that k-means clustering of time series data was able to highlight some areas of the spectrum that might be indicative of degradation, which more common chemometric techniques could not. The methods we describe here have the potential to widen the application of FTIR as a fast, non-destructive and reliable tool for assessing the condition of archaeological and historical leather objects, ultimately leading to better informed conservation, storage and restoration of these objects.

**Keywords:** FTIR, Acid hydrolysis, Collagen, Time series, Principal component analysis, *k*-means clustering

## Introduction

Archaeological artefacts made of leather are considered highly significant due to their comparative rarity and will often need to be conserved to ensure their long-term survival. One of the earliest existing examples of leather use dates from around 5000 BP in the clothing of Otzi

the Iceman, discovered in the Alps in the early 1990s [1], although the discovery of associated tools such as bone needles and scrapers indicate that large animals were exploited for their pelts as far back as the early Palaeolithic [2]. However, as leather is both structurally and chemically complex, artefacts can present a unique conservation challenge and determining the degree of deterioration in an artefact is a crucial step in identifying an appropriate conservation method [3, 4]. Leather present in historical artefacts such as books can also deteriorate in storage due to a variety of environmental factors (e.g. light, humidity and heat) [5]. Understanding degradation

\*Correspondence: elizabeth.dickinson@york.ac.uk; kirsty.high@gmail.com

<sup>1</sup> Department of Mathematics, University of York, Heslington YO10 5DD, York, UK

<sup>2</sup> Department of Chemistry, University of York, Heslington YO10 5DD, York, UK

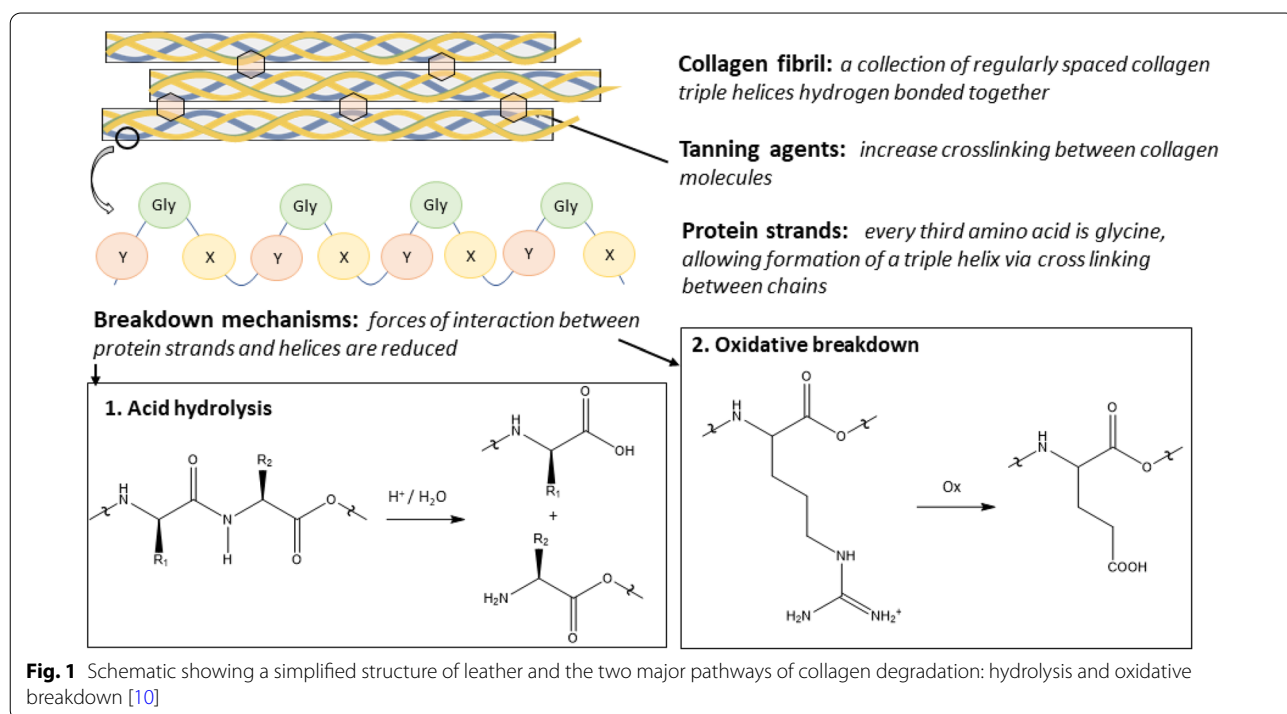
processes is therefore also an important part of determining the stability of museum or archive collections, for example in identifying increased degradation caused by storage or exhibition conditions. Identifying risks to an object due to accelerated degradation may be able to inform decisions made regarding the most appropriate storage or conservation of these objects. More sensitive analytical methods may provide the ability to recognise these risks at an earlier stage, hence leading to faster decision making.

Leather is a highly complex material formed from the processing of skin, which is composed primarily of layers of type I collagen: a hierarchical structure with a characteristic coil formed by bundles of right-handed protein triple helices hydrogen bonded together [6]. These triple helices self-assemble into a network of fibres through intra-molecular cross-linking to form a highly stable macro-structure (Fig. 1) [6]. Further molecular stability is gained by introducing cross-links between collagen molecules via the use of tanning agents; tanning also adds properties such as flexibility and softness to the leather, allowing it to be used for multiple purposes [7]. The exact nature and structure of the tanning agent present will vary depending on the date of origin of a sample, but in archaeological leather they will most commonly be plant-based compounds known as vegetable tannins: a diverse group of polyphenolic organic compounds [8, 9].

The exact route of leather degradation depends on factors such as the identity of the tanning agents and the

environmental conditions it has been subjected to [10, 11]. However, two key degradation pathways in vegetable tanned leather have been identified through experimental studies as acid hydrolysis and oxidative breakdown. [11, 12] Both processes are largely driven by environmental factors and cause breakdown of both the tanning agent and protein structure, reducing cross-linking [10]. Aside from protein cleavage, oxidative breakdown leads to deamination of the amino acid side chains, affecting hydrogen bonding between the protein strands. Therefore together, acid hydrolysis and oxidative breakdown reduce the stabilising forces of attraction afforded both by the natural composition of collagen and the tanning agents. Thus, the collagen eventually becomes a denatured random coil rather than an ordered triple helix [13]. These molecular changes manifest as changes in the mechanical physical properties of the leather (e.g. fraying, brittleness and loss of strength [14]). Both processes are affected by changes in pH, with oxidative breakdown being driven by alkaline conditions, and hydrolysis being acid catalysed [15].

Many analytical methods typically used to examine these changes (e.g. thermal analysis [14] or amino acid analysis [12]) result in loss of the sample, even if this is a very small amount, which has obvious drawbacks when applied to heritage objects. Analytical techniques also often require specialist instrumentation (e.g. scanning electron microscopy [3] or pyrolysis gas chromatography [16]) or involve lengthy preparation (e.g. amino acid



analysis [12]). Fourier Transform -Infrared spectroscopy (FTIR) can be used to detect alterations in protein composition and has been shown to reveal changes to the hierarchical structure of collagen. [17] When used with an attenuated total reflectance (ATR) attachment it is fast and requires little or no sample preparation [18]. Although FTIR-ATR can leave a small imprint on a malleable sample such as leather, the damage is minimal, and no sample is lost. These factors mean that multiple analyses can be carried out across an object, providing a better evaluation of the entire object on which to base conservation decisions and make it an ideal technique for the analysis of heritage objects. Advances in instrumentation, for example portable instruments or FTIR microscopes which do not require direct contact, increasingly provide the ability to analyse larger objects and remove risks of damage [19]. Despite this, whilst FTIR has been successfully applied to the analysis of many collagen-based materials including historic parchments [20] its application to leather is less well explored. This is largely because FTIR spectra obtained from leather is complicated by the presence of tanning agents, making it very difficult to accurately assign all peaks in the spectrum [21, 22]. However, peaks relating specifically to the collagen backbone (amide I, amide II and amide II) can be identified, and changes in these peaks have been shown to change when conformational changes within the collagen occur [21, 23]. The hydrolysis or oxidation breakdown pathways of leather will result in other changes to the functional groups present which will be represented by changes to the related peaks in the FTIR spectra, for example an increase in carboxylic acid and amine groups due to cleavage of the peptide chains (Fig. 1) [10].

The application of multivariate statistical methods to FTIR data has shown promise in recent years [17, 25, 26]. Principal component analysis (PCA) to explore data and methods such as partial least squares regression (PLS-R) or PCA-Linear Discriminant Analysis (PCA-LDA) for classification and prediction are customary techniques in metabolomic and analytical chemistry studies. These approaches provide a way of visualising change in a large series of samples, identifying spectral regions of interest, and allow the separation of samples displaying different chemical characteristics. However, although unsupervised techniques such as PCA describe the greatest variance in data this may not be of interest to a particular study; therefore alternative techniques need to be developed to identify subtle “of interest” differences between samples which may otherwise go undetected. We have recently demonstrated that *k*-means clustering is a powerful tool in identifying these subtle differences between sample groups or conditions using time-course data [25], when standard exploratory techniques were unable to

identify variance related to the problem in question. One other potential drawback of using PCA is that, when analytical samples are similar, and compositional variance between samples is low, small differences appear exaggerated by PCA scores plots. Points which would be expected to be grouped together on the plots (for example, technical replicates) tend to spread out and appear to be very different. One solution to this is to calculate the repeatability (the intraclass correlation or ICC) [26, 27], to gain confidence in data quality:

$$\text{repeatability} = \frac{\sigma_{\text{between}}^2}{\sigma_{\text{between}}^2 + \sigma_{\text{within}}^2}$$

where the  $\sigma_{\text{between}}^2$  is the variance between group means (e.g. groups of replicates, one sample is a group); the  $\sigma_{\text{within}}^2$  is the within-group variance (the pooled variance over all groups of replicates [26]) and gives an indication of technical measurement error. A score of 0–1 is obtained where the closer to 1, the more reliable and repeatable the data: 1 corresponds to no measurement error and zero means that all the variability in the data is due to experimental measurement error [27].

Clustering of time series is often used and developed in -omics technologies [28], and works by clustering similar trends or profiles over time. In this case, we have adapted the technique to cluster similar trends in intensity of FTIR wavenumbers, clustering and highlighting regions of the spectrum that change in a similar way over the time course. Like PCA, *k*-means clustering is an unsupervised technique, therefore it uses no classification variables to cluster, clustering only on the basis of similarity or closest trends, whichever classification group the data may be from. This also has the benefit that an ‘overfit’ model (one that fits the investigation data only) will not be produced. Despite not using class characteristics to cluster, it is possible to identify and isolate time series trends which are in separate clusters and belong to distinct classes, for example different pH conditions.

The aim of this investigation was to combine fast, relatively simple, and, most importantly, non-destructive FTIR analysis of leather with time-course multivariate statistics. This novel approach aimed to identify subtle compositional differences between leather samples artificially deteriorated at high temperature in the laboratory under different pH conditions (pH 3 and 5), replicating environmental conditions under which ancient leather may degrade, particularly via acid hydrolysis. By identifying regions of the FTIR spectra which subtly change, we aim to investigate whether applying chemometric approaches to FTIR data strengthens its applicability as a suitable tool for the analysis of structural and chemical changes in vegetable tanned leather as it degrades.

To the best of our knowledge, this is the first time that such statistical models have been successfully developed using FTIR time-course data from degrading leather. These models have the potential to facilitate the effective application of FTIR for the analysis of leather degradation more widely, broadening its use as a diagnostic tool within the heritage sector and furthering understanding of the breakdown mechanisms of leather.

## Materials and methods

### Controlled degradation experiments

In order to create a time-series of samples with increasing levels of degradation over a short time-scale, high temperature degradation experiments were carried out over 28 days. Experiments were carried out at both pH 3 and pH 5 to create two distinct groups of samples, since acid hydrolysis is known to be a major breakdown mechanism of leather [10, 23].

Modern vegetable tanned cow hide that had been drum-processed using mimosa as a tanning agent over a three-week period, was obtained from Thomas Ware and Sons LTD leather processors. The sample was cut into pieces approximately  $2 \times 2$  mm using a scalpel and one piece per experiment placed in a 2 mL glass ampoule before adding 1.5 mL of either pH 3 or pH 5 solution (created by adjusting ultrapure water with sulfuric acid; Fisher Scientific; tested using a calibrated glass pH probe; Denver instrument) [17]. Ampoules were heat sealed then placed in a furnace set to 80 °C for time periods of 1 h, 2 h, 3 h, 4 h, 7 h, 18 h, 1 day, 5 days, 14 days and 28 days. Each experiment was run in triplicate leading to a total of 63 experiments (10 time points  $\times$  3 replicates,  $\times$  2 different pH conditions, plus 3 control untreated replicates). At the end of the heating time, the liquid was immediately removed and samples left to air dry in laboratory conditions for at least 3 days.

### Analysis by FTIR

FTIR spectroscopy was carried out using an Agilent Cary 630 instrument with an ATR attachment and diamond window. Dry samples were placed directly onto the crystal window and pressure applied. All samples as well as the starting material were analysed at  $4 \text{ cm}^{-1}$  resolution between 400 and  $4000 \text{ cm}^{-1}$ , using an averaged 126 scans. Before chemometric analysis, all spectra were baseline corrected with no other data pre-processing.

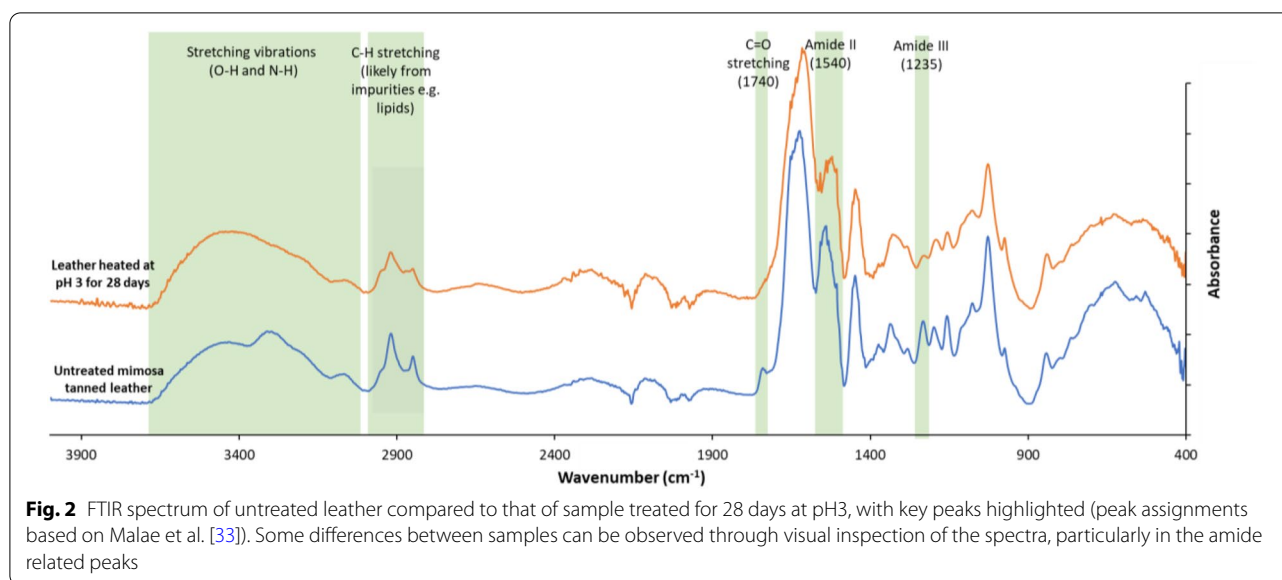
### Chemometric analysis of FTIR data

Analyses were conducted in R version 3.4.4 (R Core Team 2018, R Foundation for Statistical Computing, Vienna, Austria). Data was normalised to mean sum of the spectral integral, and initial data exploration was performed using PCA on all samples, then repeated with UV-scaled

data to prevent larger peaks dominating the analysis [29]. These analyses were repeated with a smaller spectral width of  $800\text{--}1800 \text{ cm}^{-1}$ , to remove the effect of possible varying humidity on the FTIR spectrum (as the later part of the spectra is dominated by peaks influenced by the presence of water). Regions of the spectrum with the greatest variance were identified from the PCA loadings, and a summary of results is shown in Table 2. PLS-DA was conducted (both full spectrum and wavenumbers  $800\text{--}1800 \text{ cm}^{-1}$ ), using the plsR package [30] with leave-one-out (LOO) cross validation, to identify regions specifically responsible for classifying data, and finding variance, according to pH i.e. pH3 or pH5. Repeatability of the data was calculated using the scores from the first two principal components to ensure that data was of high quality. To investigate differences between leather in pH 3 conditions and pH 5 conditions, scaled data was transposed to produce scaled time-series profiles of each FTIR peak for both pH 3 and pH 5, and the median time-series of each peak was calculated from the three replicates for each time point. Following this data preparation, the Elbow method [31] was used to determine the number of clusters required ( $k=8$ ) for  $k$ -means clustering, performed using the Hartigan–Wong algorithm. These observations were then reduced further by removing any for which the two profiles corresponding to the same peak (one from each pH group) clustered together. The remaining time-series, representing peaks that differ between the pH 3 and pH 5 data, were then ranked—for each peak represented, the Euclidean distance between the pH 3 and pH 5 time series was calculated from the original data and used to rank the peaks, to determine the profiles that changed most over time as a result of pH [26]. This  $k$ -means clustering procedure was also repeated over a smaller spectral width of  $800\text{--}1800 \text{ cm}^{-1}$ , which is expected to encompass the majority of collagen related peaks (Supplementary Information). To ensure robust results, the outcomes of some  $k$ -means and PLS-DA analyses were verified after removal of noise peaks between 400 and  $500 \text{ cm}^{-1}$  to ensure that this noise was not incorporated into any modelling.

## Results and discussion

Two replicates (5 days at pH 5 and 28 days at pH 5) failed due to drying out of the samples during the high temperature experiments and were discarded, leading to 61 successful experiments. Samples treated for the longest time points (14 and 28 days) at both pH values displayed obvious darkening and increase in brittleness, observed visually. These changes are reflected in FTIR analysis, where the spectra obtained from the most physically altered samples display substantial differences to the starting material (Fig. 2). These changes include a reduction in



intensity of peaks at between 2800 and 3000  $\text{cm}^{-1}$  which are likely to relate to non-structural components such as lipids and therefore not diagnostic of structural modifications to the leather [32]. A reduction in intensity of peaks at 1540  $\text{cm}^{-1}$  (amide II) and 1235  $\text{cm}^{-1}$  (amide III) are also very evident, and likely to relate to conformational changes within the collagen structure. A decrease in intensity of the C=O stretch at 1740  $\text{cm}^{-1}$  is also likely to relate to deterioration of the leather, either within the

collagen fibrils or tannin structure. Changes in the shape of spectral peaks can also be observed (e.g. the O–H and N–H stretch between 3000 and 3600  $\text{cm}^{-1}$ ). Differences in peaks which could be attributed to more than one component (including tannins and contaminants from the manufacturing or experimental processes) are also seen, although these are likely to be less diagnostic of structural changes due to the potential contribution from multiple components (Table 1). Changes in peak

**Table 1** Summary of characteristic peaks in the FTIR spectra of leather

Wavenumber (approx)	Functional group	Possible component
3600–3200	O–H stretch	Adsorbed water
3100–2800	C–H stretch	
3220 (broad)	N–H stretching	
2925 (broad)	C–H <sub>3</sub> stretching	Protein chain, lipids
2900	C–H bending	Protein chain, lipids
2830	O=C–H stretch	Aldehyde from degradation
2350 (broad)		Sulfur (possibly from contamination)
2360/2340 doublet	Carbon dioxide	
1740–1750	R–C=O	Esters in collagen, lipids
1660	C=O stretching	Amide I
1656–1500	C–N–H bending	Amide II
1380	C–O–C	Amino acid side chains
1490–1300	C–H bending	Deformation of protein and lipids
1258–1273	Protein backbone	Amide III
1150	C–O–C stretch	Cyclic ethers in tannins
1033	C–O–C=O	Esters in tannins
990	C–O stretch	
976	C=C	Amide III

Many of the peaks have possible contributions from more than one element of the leather [22, 24, 32]

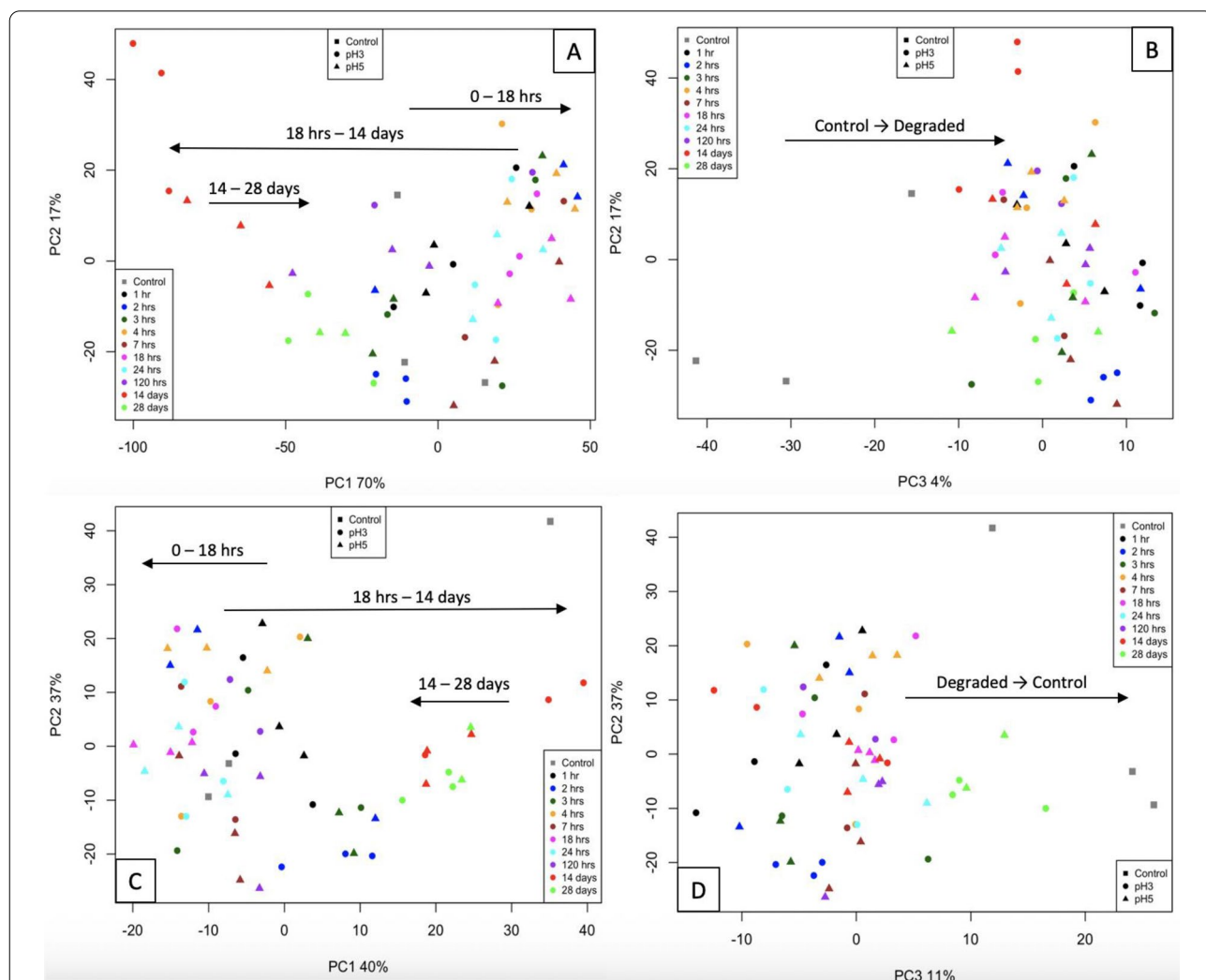


position are also observed, for example the large peak at  $1640\text{ cm}^{-1}$  (amide I) has shifted to a lower wavenumber in the degraded sample.

At shorter timepoints the differences between spectra were much more subtle, emphasising the possible benefits of using chemometric analysis to differentiate between less severely degraded samples. Further investigation using statistical methods aimed to evaluate these differences as well as highlight additional areas of the spectra that may be diagnostic of degradation.

PCA on the (unscaled) data resulted in identifying changes in intensity of the largest peaks of the spectrum some of which could be identified by eye, as expected—changes around  $2900\text{ cm}^{-1}$  (C-H bending),  $1520\text{ cm}^{-1}$  (Amide II) and  $1640\text{ cm}^{-1}$  (Amide II). However PCA of scaled data

(to prevent domination of larger peaks) showed that the greatest variance between samples was due to structural changes increasingly occurring over longer artificial degradation times; when looking at principal components 1–3 (accounting for total variance of 91%), little distinction was observed between samples degraded at pH3 and those at pH5 when the timepoint was the same (Fig. 3A, B), with the exception of samples at 14 days, where pH 3 samples appear somewhat apart from pH5 samples across PC2 (Fig. 3A). For PC3 the variance described is between untreated leather samples and those degraded (both at pH3 and pH5). When analysing this scaled data, the peaks changing were different to those visually identified, and were identified by the PCA loadings (summarised in Table 2). When focusing on the region of  $800\text{--}1800\text{ cm}^{-1}$



**Fig. 3** A, B PCA scores plots of the full spectrum analysis showing changes over time across PC1 and PC2 (A) and between control and degraded samples across PC3 (B). C, D PCA scores plots of the analysis of the  $800\text{--}1800\text{ cm}^{-1}$  region, showing changes over time across PC1 and PC2 (C) and between control and degraded samples across PC3 (D)

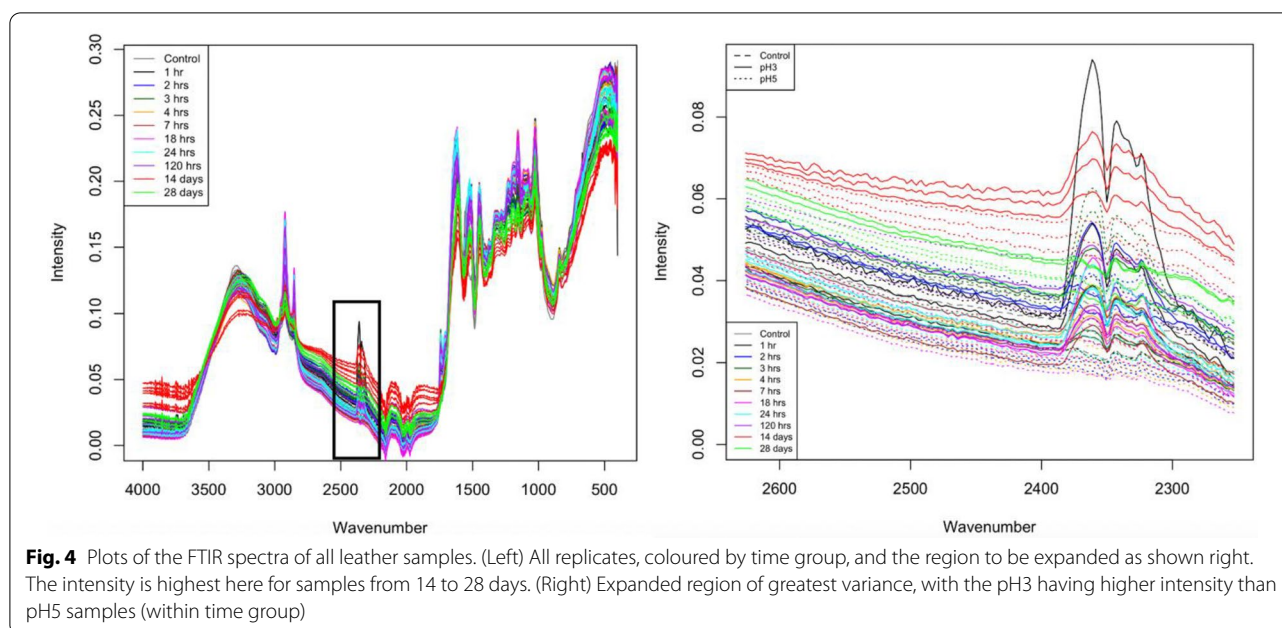
**Table 2** A summary of regions of the FTIR spectrum of leather, found through PCA/PLS loadings and k-means clustering, which change as a result of time, degradation or pH of surrounding conditions [22, 24, 32]

Analysis	Result observed	Influential regions (cm <sup>-1</sup> )	Possible peak assignment
PCA: full spectrum	PC1: changes over time	2441–2501	None: spectral feature
	PC2: changes over time	3209–3229	O–H stretch, N–H stretch
	PC3: Differences between control and degraded	1474–1496	C–N–H bend
PLS-DA: full spectrum	Differences between control and degraded Differences between pH 3 and pH 5	1490–1525	C–N–H bend
		1000–1080	C–O–C=O in tannin ester stretch
		2849–2933	C–H stretch
		2329–2370	<i>Sulfur (possible contamination)</i>
		1705–1753	C=O stretch
<i>k</i> -means: full spectrum	Differences between pH 3 and pH 5	2911–2925	C–H stretch
		1738–1746	C=O stretch
		1140–1160	C–O–C cyclic ether stretch
		1019–1034	C–O–C=O in tannin ester stretch
PCA: 800–1800 cm <sup>-1</sup>	PC1: changes over time PC2: changes over time	1151–1196	C–O–C cyclic ether stretch
		1526–1539	C–N–H bend
	PC3: Differences between control and degraded PC4: Differences between control and degraded	989–998	Amide III, C–O stretch
		1684–1686	C=O stretching
		1341–1343	C–N stretch, C–O stretch
		1030–1086	C–O stretch
		1258–1273	Amide III
PLS-DA: 800–1800 cm <sup>-1</sup>	Differences between control and degraded	1630–1648	C–N–H bend/Amide I
		1503–1531	C–N–H bend
		997–1061	C–O–C=O in tannin ester stretch
<i>k</i> -means: 800–1800 cm <sup>-1</sup>	Differences between pH 3 and pH 5	1023–1031	C–O–C=O in tannin ester stretch
		1490–1511	C–N–H bend

the same pattern was observed when plotting PCA scores (Fig. 3C, D), though no real distinction was observed between pH3 or pH5 samples at 14 days. This analysis gave much more informative peak assignments than the analysis of the full spectrum. Again, when looking at principal components 1–4 (accounting for total variance of 93%), no distinction was observed between samples degraded at pH 3 and those at pH 5 when the timepoint was the same. Both scores plots in Fig. 3A, C show that there is considerable overlap between samples from 0 to 18 h, and that structural change appears to occur from 18 h onwards, with the greatest change (across PC1) between 18 h and 14 days, with distinction between the groups of replicates, particularly between 24 h, 120 h, 14 days and 28 days.

The regions of the full FTIR spectrum which varied most, identified from the PCA loadings, are summarised in Table 2. The greatest variance in the data across PC1 was in the region of 2450 cm<sup>-1</sup> which appeared to change most between samples at 18 h and later time samples at 14 days (early time point samples had lower intensity than late time point samples; Additional file 1: Fig. S1). Although there is no specific peak assignment here, this area of the spectrum is within a much wider spectral feature for the late time point samples. Figure 4 shows that samples from 14 to 28 days have generally much higher

intensity in the general region of 2600–1800 cm<sup>-1</sup> compared to all other samples. By scaling the data to prevent dominance of larger peaks, the region around 2450 cm<sup>-1</sup> has been shown to have much higher intensity than other parts of the spectrum within this spectral feature. A second region which appeared to change most between samples at 18 h and later time samples at 14 days was 3220 cm<sup>-1</sup> (O–H/N–H stretch), increasing in intensity between 14 and 28 days (PC2). Although much of the variance in PC2 is from changes over time, some contribution is also from differentiation between pH 3 and pH 5 at 14 days, with pH3 having higher intensity than pH 5 (Fig. 4). For any other time group, no difference can be observed between pH of experimental conditions. Changes between degraded samples (both pH 3 and pH 5) and controls were in the region of 1480 cm<sup>-1</sup> (C–H bend), which had higher intensity for degraded samples. PCA of region 800–1800 cm<sup>-1</sup> identified peaks changing between 18 h to 14 days: C–O–C cyclic ether stretching (1150 cm<sup>-1</sup>) and C–N–H stretching (1530 cm<sup>-1</sup>), both of which reduced in intensity as time increased, whilst C–O stretching (990 cm<sup>-1</sup>) increased. Differences between controls and degraded leather samples (pH 3 and pH 5) were observed across PC3 (Fig. 3D) and PC4 (Additional file 1: Fig. 2), and peaks changing in intensity were C=O stretch (1685 cm<sup>-1</sup>; higher for degraded samples),



C–N and C–O stretch ( $1342$  and  $1050\text{ cm}^{-1}$ ; lower for degraded samples) and amide III ( $1260\text{ cm}^{-1}$ ; higher for degraded samples).

Despite the apparent “spread” of replicates on PCA plots, repeatability of replicates was calculated (using scores from PC1 and PC2) at 0.85, therefore demonstrating that the data was of high quality, but suggesting that FTIR spectral differences between time points are likely to be subtle.

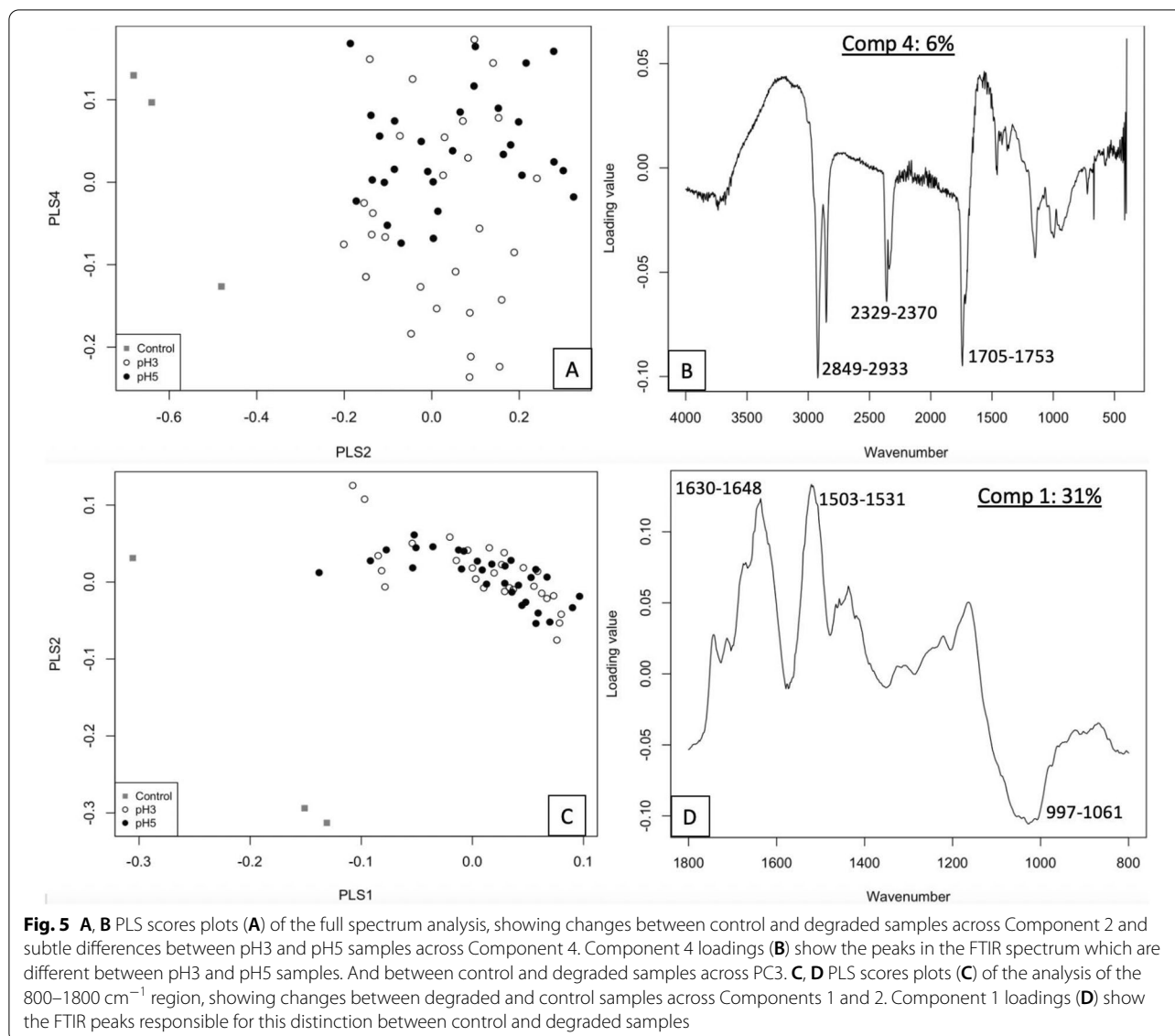
As with PCA, PLS-DA was conducted on the full spectrum followed by the more focused approach of analysing  $800\text{--}1800\text{ cm}^{-1}$ ; this technique was used to establish peaks changing most which related to the class of sample i.e. pH 3, pH 5 or control. Using the full spectrum, it is clear from the PLS scores plots that controls were simple to differentiate from degraded samples, whilst there is considerable overlap of pH3 and pH5 samples. By plotting Component 2 by Component 4 (Fig. 5A), the difference between control and degraded is observed (across Component 2) as well as some differentiation between pH3 and pH5 samples (across Component 4). Peaks changing between degraded and control samples (Component 2) were similar to those found in PCA i.e.  $1490\text{ cm}^{-1}$  (C–N–H bend), as well as  $1030\text{ cm}^{-1}$  (C–O–C=O stretch). Peaks which changed between samples degraded in pH 3 vs pH 5 were from C–H stretch, sulfur (assumed to be contamination from the sulfuric acid) and C=O stretch, at  $2900$ ,  $2350$  and  $1740\text{ cm}^{-1}$ , respectively (Fig. 5B). When PLS-DA was conducted between pH3 and pH5 samples alone (controls excluded), the same peaks as in earlier analysis Component 4 (Fig. 5B)

were found as most important in differentiating between the pH conditions. These peak intensity changes had not been identified by eye or unsupervised PCA. In contrast, when analysing the region  $800\text{--}1800\text{ cm}^{-1}$  only, this technique was much less successful in differentiating between pH 3 and pH 5 samples, with significant overlap of samples on the scores plots (Fig. 5C), even up to Component 4. Changes in FTIR peaks between degraded and control samples were identified by PLS loadings (Fig. 5D) as C–N–H bend/Amide I ( $1635\text{ cm}^{-1}$ ), C–N–H bend/Amide II ( $1530\text{ cm}^{-1}$ ) and C–O–C=O stretch ( $1030\text{ cm}^{-1}$ ), as discussed earlier.

PCA is therefore very useful in giving insight into structural changes over time, but unable to distinguish the more subtle structural differences occurring due to varying pH of experimental conditions. Only when analysing the full spectrum of day 14 samples were some differences highlighted between pH of experimental conditions. PLS-DA was only semi-successful, when the whole of the spectrum is considered.

*k*-means cluster analysis was successful in identifying the subtle differences in composition of the leather over-time as a result of pH of experimental conditions, which the more common techniques PCA and PLS-DA were unable to achieve adequately. Figure 6 shows the results of the *k*-means cluster analysis, where every line in the plots in the figure represents an FTIR data point, and how that data point or peak changes in intensity over time. There are twice as many lines in the figure as there are FTIR peaks in the spectrum, due to two different pH of experimental conditions;

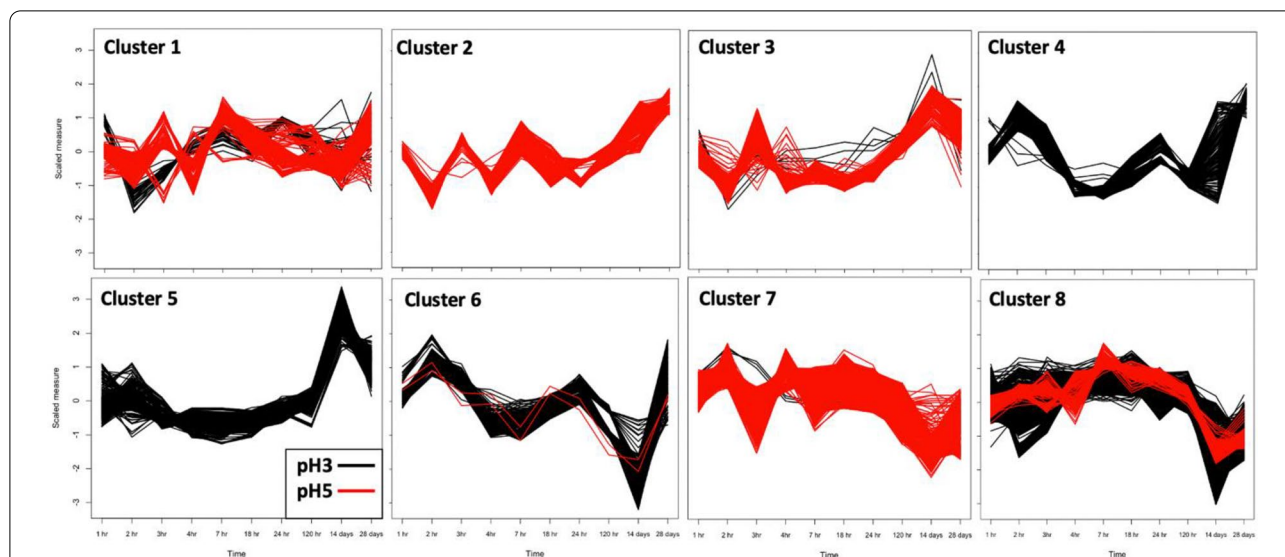




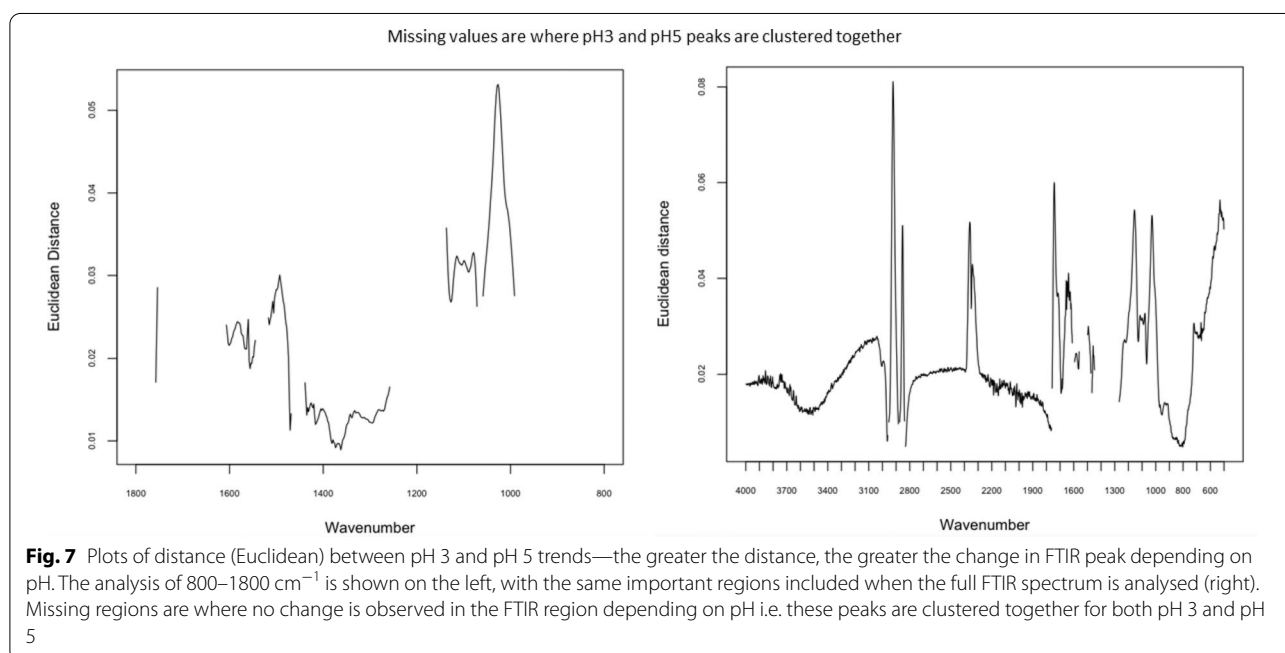
**Fig. 5** **A, B** PLS scores plots (**A**) of the full spectrum analysis, showing changes between control and degraded samples across Component 2 and subtle differences between pH3 and pH5 samples across Component 4. Component 4 loadings (**B**) show the peaks in the FTIR spectrum which are different between pH3 and pH5 samples. And between control and degraded samples across PC3. **C, D** PLS scores plots (**C**) of the analysis of the 800–1800  $\text{cm}^{-1}$  region, showing changes between degraded and control samples across Components 1 and 2. Component 1 loadings (**D**) show the FTIR peaks responsible for this distinction between control and degraded samples

the black time series represent every FTIR peak at pH 3, and the red time series represent each FTIR peak at pH 5. The y-axis represents the scaled peak intensity, so that larger peaks again do not dominate the analysis. The figure clearly shows that many time-series of FTIR peaks cluster separately depending on the pH of the conditions, with some clusters containing time series of peaks from only one pH. The peaks/regions of the FTIR spectrum which changed most can be observed in Fig. 7, where the distance between the pH 3 and pH 5 trends (for each peak) are plotted by wavenumber; the greater the distance, the more different the pH 3 time series is from its corresponding pH 5 time series. When analysing all wavenumbers in the full spectrum, the greatest differences in intensities

between pH conditions were exhibited in the regions of 2911–2925  $\text{cm}^{-1}$  and 1738–1746  $\text{cm}^{-1}$ , supporting the PLS-DA results obtained when analysing the full spectrum for changes between pH 3 and pH 5 samples. Other FTIR regions which changed depending on pH were also identified, including around 2350, 1150 and 1020  $\text{cm}^{-1}$ : these regions were also identified as changing most over time in the PCA results, and showed the greatest changes between control and degraded samples in the PLS analysis. The same spectral difference in the region 1023–1031  $\text{cm}^{-1}$  was found in the analysis using wavenumbers 800–1800  $\text{cm}^{-1}$  only, alongside a region around 1500  $\text{cm}^{-1}$ , most likely from C–N–H bend/Amide III peak. This subtle difference between pH 3 and pH 5 was not identified in the PLS-DA when



**Fig. 6** Clusters obtained from the k-means clustering of the time series trends for each peak in the FTIR spectrum of leather samples degraded at pH 3 (black) and pH 5 (red). Peaks following the same trends are clustered together, regardless of pH group



**Fig. 7** Plots of distance (Euclidean) between pH 3 and pH 5 trends—the greater the distance, the greater the change in FTIR peak depending on pH. The analysis of 800–1800  $\text{cm}^{-1}$  is shown on the left, with the same important regions included when the full FTIR spectrum is analysed (right). Missing regions are where no change is observed in the FTIR region depending on pH i.e. these peaks are clustered together for both pH 3 and pH 5

specifically modelling the data based on control vs pH 3 and pH 5, and was only identified as a region which may change between control and degraded samples (figures of the clustering of the 800–1800  $\text{cm}^{-1}$  subset is shown in the Additional file 1).

Importantly, as well as isolating pH-dependent trends over time, crucial information on specifically how individual peak intensities change over the duration of the

time series can be examined, and if peaks from functional groups are clustering together. K-means clustering has confirmed the results found using other chemometric techniques and found additional differences between pH3 and pH5 degraded leather, but we are able to use a targeted approach to compare how the FTIR regions of interest change over time from Table 2, specifically bonds in the collagen, depend on the pH of the surrounding

**Table 3** Regions of interest in the FTIR spectrum of degrading leather, and how differently the time series trends cluster, depending on the pH of the surroundings conditions

Peak region of interest (cm <sup>-1</sup> )	Possible peak assignment	Cluster for pH 3 (see Fig. 6)	Cluster for pH 5 (see Fig. 6)
3400–3600	O–H stretch	4	2
3200–3400	O–H stretch	6	1
2900–3100	C–H stretch (lipids)	4	2
1750	C=O stretch	1	7
1685	Amide I	8	7
1560–1590	C–N–H, Amide II	6	1
1490–1559	C–N–H, Amide II	7	8
1470–1490	Amide III	6	1
1257	Amide III	8	7
1030	Tannin ester C–O–C=O stretch	8	7
990	Amide III	8	7
Other regions	–	5	3

conditions: this is summarised in Table 3. It is clear from Fig. 6 that many of the pH3 trends (in black) are more ‘extreme’ with greater minima and maxima when compared to the corresponding pH 5 trend (in red). For example, cluster 6 (predominantly pH 3) has a general downward trend over time, with much greater minimum intensity at 14 days, when compared to cluster 1 (predominantly pH 5): these clusters contain O–H stretch, Amide III and C–N–H, Amide II peaks. Similarly, cluster 7 (predominantly pH 5) contains time series of the intensities of peaks from Amide I, Amide II and Amide III together, gradually lowering in intensity over time. The same trend is observed for cluster 8 with the same corresponding pH3 peaks, however, cluster 8 shows a later drop to much lower intensity for later time points at 14–28 days. The O–H stretch bonds at 3400 cm<sup>-1</sup> and in clusters 2 (pH 5) and cluster 4 (pH 3) are clearly also affected differently depending on pH, with cluster 4/pH 3 showing greater maxima and minima within the trend over time compared to the corresponding peaks for pH 5.

The application of chemometrics to FTIR data of leather has proven to be a powerful data analysis tool, in that subtle structural changes to collagen can be detected in the FTIR spectrum over a short time course of degradation. The application of ‘standard techniques’ such as PCA and PLS-DA have shown promise in detecting changes in the FTIR spectrum over time, and between control leather samples and degraded samples. Many of the peaks in the FTIR spectra for leather contain contributions from more than one element (e.g. collagen, tannins, or non-structural components such as lipids) and many overlap. This makes it difficult to assign peaks, and

more difficult yet to interpret differences in the peaks. We must therefore acknowledge that the changes we’ve demonstrated need to be investigated further before being used as indicative measures of deterioration. Certainly, some of the greatest changes in the FTIR spectra appear to be in the C–H stretch, which likely relates primarily to non-structural lipids and tannins. However, this highlights that these are the first changes to occur under such acidic conditions, and further analysis may reveal further information either about the degradation of tannins themselves or the tannin-collagen interface. We have nevertheless shown that subtle changes, depending on how acidic conditions are, can be detected using clustering of time series, and PLS-DA. The results of the *k*-means has supported some of the PCA and PLS-DA results, ensuring a more robust investigation. Undoubtedly, other statistical approaches or classification methods could be applied to or data, such as genetic algorithms, classification and regression trees (CART) [15] or neural networks, however, the *k*-means results of establishing differences between pH3 and pH5 spectra are more reliable than those of supervised techniques, including PLS-DA; *k*-means is an unsupervised technique, therefore there is no possibility of overfitting. The fact that the greatest changes were observed in the peaks assigned to C–H stretch of the lipids and C–O stretch of tannins (peaks outside the 800–1800 cm<sup>-1</sup> region) explains why PLS-DA was unable to differentiate between pH 3 and pH 5 degradation when analysing the 800–1800 cm<sup>-1</sup> region alone. The development of *k*-means clustering of time series to investigate differences based on pH has also circumvented the ‘problem’ or innate feature of PCA, that PCA finds the greatest variance within data, regardless of whether or not this may of interest: in this case, the greatest differences were based on time, rather than pH of conditions. These spectral differences over time were obviously of interest, but PCA was unsuitable to compare more subtle spectral changes. Although *k*-means clustering of time series would not be suitable to study archaeological leather discovered at one time point, it has been a crucial technique to show that specific functional groups change over time in the same way by clustering peak trends together, i.e. those assigned to amides and C–N–H bonds from the collagen, and those assigned to phenolic groups of tannins. It has identified specific FTIR regions of interest, and therefore highlighted areas of the spectra which may be indicative of important changes in chemistry involved in the degradation of leather in increasingly acidic conditions. It has also confirmed that the more acidic conditions show a much greater effect on the structure of collagen, as one might expect. Finally, a reduction in intensity of amide C–N–H bonds/amide over time confirms a breakdown

of collagen structure rather than a breakdown of tannins or non-structural components such as lipids. Although it is impossible to resolve peaks OH peaks (around  $3400\text{ cm}^{-1}$ ) of tannins from those within collagen, O–H bonds are clearly affected by changes in pH by the different clustering of these groups for pH 3 and pH 5—this spectral change could correspond to the breakdown of the tanning agent or the breakdown of hydrogen bonding within the collagen structure.

## Conclusions

FTIR is not currently widely used to determine changes in leather structure with degradation, largely due to the complexity of the spectra obtained making it difficult to confidently assign peaks and establish regions indicative of degradation. However, we have demonstrated that the development and combination of time-series chemometric techniques applied to FTIR have the potential to be powerful tools. We do not suggest that this work alone be used as a predictive model to discern the age of degrading leather; what we show is that our untargeted approach has undoubtedly identified families of peaks that are affected in the same way over time, together showing structural changes of the collagen in leather caused by acid hydrolysis, rather than just the breakdown of tannins. Moreover, we prove that these structural changes based on the pH of the surrounding conditions are very subtle, but clearly different and identifiable by combined FTIR and multivariate statistics. Our work supports that of Vyskočilová et al. [17], who also used chemometrics and FTIR to monitor the breakdown of collagen and degradation of leather. However, our study goes further, in that we are able to discern much more subtle changes in the leather structure earlier in the degradation process, and between different acidic conditions. Our work shows that there is significant promise in using FTIR to further investigate leather degradation, and this study could be used as a platform for targeted work aimed at understanding the deterioration of archaeological and historical leathers. Indeed, by acquiring FTIR data on leather samples of known age/environmental conditions and focusing on the most important peaks or spectral features reported in this work, it would be possible to build much simpler predictive models that could be used to predict the state of degradation of newly discovered archaeological leather samples, crucial to determining appropriate conservation treatment. These models could then potentially be applied effectively by heritage scientists who are not expert in the field of chemometrics. Furthermore, our *k*-means approach opens up the possibilities for FTIR to be used to assess and monitor degradation of other

heritage materials over shorter timescales, for example, whilst changing storage conditions or moving archaeological materials, allowing the evaluation of subtle differences. As such, potential risks to preservation may be noticed at an early stage, allowing appropriate action to be quickly taken and better safeguarding cultural heritage materials.

## Abbreviations

FTIR: Fourier transform infrared spectroscopy; ATR: Attenuated total reflectance; PCA: Principal component analysis; PLS-DA: Partial least squares discriminant analysis; PCA-LDA: PCA linear discriminant analysis; ICC: Intraclass correlation; UV: Unit variance; LOO: Leave one out; CART: Classification and regression trees.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40494-022-00690-w>.

**Additional file 1: Fig. S1.** Boxplots showing range of intensity values obtained for peaks at (left)  $2434\text{ cm}^{-1}$  and (right)  $3222\text{ cm}^{-1}$ , from all samples (regardless of pH). **Fig. S2.** PCA scores plots of analysis of FTIR data of leather, showing PC1 vs PC4. (Left) PCA of full spectrum, (Right) from PCA of  $800\text{--}1800\text{ cm}^{-1}$ . Later time points (14–28days) are clearly distinct from other time points in both analyses. **Fig. S3.** Clusters obtained from the *k*-means clustering of the time series trends for each peak in the FTIR spectrum region  $800\text{--}1800\text{ cm}^{-1}$  of leather samples degraded at pH3 (black) and pH5 (red). Peaks following the same trends are clustered together, regardless of pH group.

## Acknowledgements

The authors wish to thank Grace Clark and Edward Thirkell for helpful discussions regarding this research. Edward Thirkell also carried out part of the lab-based experiments as part of his undergraduate Master's project. Leather samples were supplied by Thomas Ware and Sons LTD courtesy of Barry Knight.

## Author contributions

ED: Conceptualization, methodology, project administration, formal analysis, software and validation, visualisation, writing of the original draft and reviewing/editing. KH: Conceptualization, methodology, project administration, laboratory investigation, data curation, funding acquisition, resources, supervision, visualisation, writing of the original draft and reviewing/editing. All authors read and approved the final manuscript.

## Funding

ED is supported by a Knowledge Transfer Partnership (KTP) between the University of York and Croda Europe Ltd, funded by Croda and Innovate UK. KH is supported by a Natural Environment Research Council Knowledge Exchange Fellowship (Grant Number NE/P005799/1).

## Availability of data and materials

The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

## Declarations

### Competing interests

None declared.

Received: 8 November 2021 Accepted: 11 April 2022  
Published online: 24 May 2022

## References

- Püntener AG, Moss S. Ötzi, the Iceman and his Leather clothes. *Chimia*. 2010;64:315–20.
- Charles R. The exploitation of carnivores and other fur-bearing mammals during the North-western European late upper Palaeolithic and Mesolithic. *Oxf J Archaeol*. 1997;16:253–77.
- Wallace A. Scanning electron microscopy and fibre shrinkage temperature analysis of archaeological waterlogged leather: Observations on Medieval leather from Swinegate, York. In: Hoffman P, Grant T, Spriggs JA, editors. Proceedings of the 6th ICOM group on Wet Organic Archaeological Materials Conference. ICOM Committee for Conservation Working Group on Wet Organic Archaeological Materials, Bremerhaven; 1997.
- Peacock EE. Conservation of severely deteriorated wet archaeological leather recovered from the Norwegian Arctic. Preliminary results. In: Proceedings of the 9th ICOM WOAM conference, Copenhagen, 2004.
- Calnan CN. Ageing of vegetable tanned leather in response to variations in climatic conditions. In: Calnan C, Haines B, editors. *Leather: its composition and changes with time*. Northampton: The Leather Conservation Center; 1991, pp. 41–50.
- Shoulders MD, Raines RT. Collagen structure and stability. *Annu Rev Biochem*. 2009;78:929–58.
- Covington AD. *Tanning chemistry: the science of leather*. The Royal Society of Chemistry, Cambridge; 2009.
- Khanbabaee K, Van Ree T. Tannins: classification and definition. *Nat Prod Rep*. 2001;18:641–9.
- Falcão L, Araújo MEM. Characterisations of vegetable tannins in historical leathers by ATR-FTIR spectroscopy and spot tests. *J Cult Herit*. 2013;14:499–508.
- Larsen R. The chemical degradation of leather. *CHIMIA Int J Chem*. 2008;62:899–902.
- Florian MLE. The Mechanisms of deterioration in leather. In: Conservation of leather and related materials. In: Kite M, Thomson R, editors. 2006; pp 36–57.
- Larsen R, Vest M, Nielsen K, Jensen AL. Amino acid analysis. In: STEP leather Project, Research Report No. 1. København: Bjørnholt Repro; 1994.
- Carsote C, Şendrea C, Micu M, Adams A, Badea E. Micro-DSC, FTIR-ATR and NMR MOUSE study of the dose-dependent effects of gamma irradiation on vegetable-tanned leather: the influence of leather thermal stability. *Radiat Phys Chem* 2021;189.
- Plavan V, Miu L, Gavryuk L. Evaluation of the amino acid composition, structure and properties of archaeological leather. *Proc Chem*. 2013;8:279–83.
- Chahine, C. Acid deterioration of vegetable tanned leather. In: *Leather, its composition and changes with time*. In: Calnan C, Haines B, editors. Leather Conservation Centre: Northampton, UK, pp. 75–87; 1991.
- Sebestyén Z, Czégény Z, Badea E, Carsote C, Şendrea C, Barta-Rajnai E, Bozi J, Miu L, Jakab E. Thermal characterization of new, artificially aged and historical leather and parchment. *J Anal Appl Pyrol*. 2015;115.
- Vyskočilová G, Ebersbach M, Kopecká R, et al. Model study of the leather degradation by oxidation and hydrolysis. *Herit Sci*. 2019;7:26.
- Analytical Methods Committee Technical Brief. Fourier Transform infrared spectroscopic analysis of organic archaeological materials: background paper. *Anal Methods*. 2021;26.
- Margariti C. The application of FTIR microspectroscopy in a non-invasive and non-destructive way to the study and conservation of mineralised excavated textiles. *Herit Sci*. 2019;7:63.
- Badea E, Miu L, Budrugaec P, Giurginca M, Masic A, Badea N, Della Gatta. Study of deterioration of historical parchments by various thermal analysis techniques complemented by SEM, FTIR, UV-Vis-NIR and unilateral NMR investigations. *J Therm Anal Calorimetry*. 2008;91.
- Bicchieri M, Monti M, Piantanida G, Pinzari F, Sodo A. Non-destructive spectroscopic characterization of parchment documents. *Vib Spectrosc*. 2011;55(2):267–72.
- dos Santos Grasel F, Flôres Ferrão M, Rodolfo Wolf C. Development of methodology for identification the nature of the polyphenolic extracts by FTIR associated with multivariate analysis. *Spectrochim Acta Part A Mol Biomol Spectrosc*. 2016;153:94–101.
- Carsote C, Badea A, Miu L, Gatta GD. Study of the effect of tannins and animal species on the thermal stability of vegetable leather by differential scanning calorimetry. *J Therm Anal Calorim*. 2016;124:1255–66.
- Mehta M, Naffa R, Maidment C, Maidment C, Holmes G, Waterland M. Raman and ATR-FTIR spectroscopy towards classification of wet blue bovine leather using ratiometric and chemometric analysis. *J Leather Sci Eng*. 2020;2:3.
- Dickinson E, Rusilowicz MJ, Dickinson M, Charlton AJ, Bechtold U, Mullineaux PM, Wilson J. Integrating transcriptomics techniques and *k*-means clustering in metabolomics to identify markers of abiotic and biotic stress in *Medicago truncatula*. *Metabolomics*. 2018;14:126–37.
- Wehrens R, Hageman JA, van Eeuwijk F, Kooke R, Flood PJ, Wijmker E, Keurentjes JJB, Lommen A, van Eekelen HDLM, Hall RD, Mumm R, de Vos RCH. Improved batch correction in untargeted MS-based metabolomics. *Metabolomics*. 2016;12:88.
- Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol*. 2008;31:466–75.
- Kiddle SJ, Windram OPF, McHattie S, Mead A, Beynon J, Buchanan-Wollaston V, Denby KJ, Mukherjee S. Temporal clustering by affinity propagation reveals transcriptional modules in *Arabidopsis thaliana*. *Bioinformatics*. 2010;26(3):355–62.
- Craig A, Cloarec O, Holmes E, Nicholson JK, Lindon JC. Scaling and normalization effects in NMR spectroscopic metabolomic data sets. *Anal Chem*. 2006;78:2262–7.
- Mevik B-H, Wehrens R. The pls package: principal component and partial least squares regression in R. *J Stat Softw*. 2007;18:1–23.
- Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: an R package for determining the relevant number of clusters in a data set. *J Stat Softw*. 2014;61(6):1–36.
- Hassan RRA. A preliminary study on using linseed oil emulsion in dressing archaeological leather. *J Cult Herit*. 2016;21:786–95.
- Malea E, Boyatzis SC, Kehagia M. Cleaning of tanned leather: testing with infra red spectroscopy and SEM-EDAX. Joint Interim Meeting of Five ICOM-CC Working Groups: Leather And Related Materials, Rome. 2010.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)