Analyst

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Open Access Article. Published on 28 September 2022. Downloaded on 4/18/2026 7:40:22 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Check for updates

Cite this: *Analyst*, 2022, **147**, 5113

# Label-free fibre optic Raman spectroscopy with bounded simplex-structured matrix factorization for the serial study of serum in amyotrophic lateral sclerosis†

James J. P. Alix, *‡[a,b] Nick S. Verber,‡[a,b] Chlöe N. Schooling,[a,c] Visakan Kadirkamanathan,[c] Martin R. Turner,[d] Andrea Malaspina,[e] John C. C. Day[f] and Pamela J. Shaw*[a,b]

Amyotrophic lateral sclerosis (ALS) is an incurable neurodegenerative disease in urgent need of disease biomarkers for the assessment of promising therapeutic candidates in clinical trials. Raman spectroscopy is an attractive technique for identifying disease related molecular changes due to its simplicity. Here, we describe a fibre optic fluid cell for undertaking spontaneous Raman spectroscopy studies of human biofluids that is suitable for use away from a standard laboratory setting. Using this system, we examined serum obtained from patients with ALS at their first presentation to our centre ($n$ = 66) and 4 months later ($n$ = 27). We analysed Raman spectra using bounded simplex-structured matrix factorization (BSSMF), a generalisation of non-negative matrix factorisation which uses the distribution of the original data to limit the factorisation modes (spectral patterns). Biomarkers associated with ALS disease such as measures of symptom severity, respiratory function and inflammatory/immune pathways (C3/C-reactive protein) correlated with baseline Raman modes. Between visit spectral changes were highly significant ($p$ = 0.0002) and were related to protein structure. Comparison of Raman data with established ALS biomarkers as a trial outcome measure demonstrated a reduction in required sample size with BSSMF Raman. Our portable, simple to use fibre optic system allied to BSSMF shows promise in the quantification of disease-related changes in ALS over short timescales.

Received 6th June 2022,
Accepted 27th September 2022

DOI: 10.1039/d2an00936f

rsc.li/analyst

## Introduction

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative condition caused by the progressive loss of motor neurones in the brain and spinal cord. As a result, patients experience weakness of limb and respiratory muscles, as well as of the muscles controlling speech and swallowing. Despite significant progress in understanding the disease, ALS remains incurable with an average survival time of two years from the point of diagnosis.[1]

As a result, many clinical trials are underway which attempt to treat different aspects of the disease. Development of biomarkers of disease that can identify early changes are therefore a priority area of ALS research. Recent imaging,[2] electrophysiological[3] and serum[4–6] biomarker studies have demonstrated early changes in longitudinal measurements (within 3–6 months) with the potential to improve clinical trial design.[7]

Raman spectroscopy is a form of vibrational spectroscopy based upon the inelastic scattering of light. Interest in the application of Raman spectroscopy to neurological disorders is growing due to the simple, label-free nature of the technique.[8,9] Biofluids are particularly appealing in biomarker research due to the ease of sample acquisition. Thus far, biofluid based ALS studies have typically employed surface enhanced Raman spectroscopy,[10–12] in which inelastic scattering is potentiated by plasmon excitation in nanoparticles.[13] Spontaneous Raman spectroscopy, which is more straightforward to implement but generates a far weaker signal, has also been used, albeit less frequently, in studies on ALS[14] and other neurodegenerative diseases.[15] While most applications have employed standard laboratory-based microscope formats,

[a]Sheffield Institute for Translational Neuroscience, University of Sheffield, UK. E-mail: j.alix@sheffield.ac.uk, pamela.shaw@sheffield.ac.uk
[b]Neuroscience Institute, University of Sheffield, UK
[c]Department of Automatic Control and Systems Engineering, University of Sheffield, UK
[d]Nuffield Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK
[e]Blizard Institute, Queen Mary University of London, London, UK
[f]Interface Analysis Centre, School of Physics, University of Bristol, UK

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d2an00936f

‡ These authors contributed equally to this work.

Open Access Article. Published on 28 September 2022. Downloaded on 4/18/2026 7:40:22 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

fibre optic technologies are gaining momentum due to their potential for use in clinical environments.[16] Invasive, *in vivo* fibre optic based measurements have been a focus for development but bedside testing of easily obtainable samples, such as blood, would potentially avoid complex laboratory assays and provide rapid results upon which clinicians could act.

Parallel to technological advances, the development of data analysis algorithms is key for the incorporation of Raman spectroscopy into clinical research and, ultimately, clinical practice.[17,18] With the exception of deep learning methods, most statistical analyses begin with dimension reduction, a process which aims to improve data visualisation and remove redundant information. Non-negative matrix factorisation (NMF) is popular method, used for both dimension reduction and feature extraction within signal processing and biomedical fields.[19] The technique combines multivariate analysis and linear algebra to decompose the original data into two lower ranking matrices (*i.e.* with fewer dimensions). In the context of Raman spectroscopy, one of these matrices contains the spectral patterns, also termed 'modes', for which the non-negative constraint provides a physically realistic output. The other matrix contains the associated coefficients, typically termed 'weights', which represent the relative importance of a given pattern to a given sample. Recently, generalisations of NMF, termed simplex structured matrix factorisation, have been proposed which do not impose the non-negative constraint.[20] A newly developed variant of this approach, termed bounded simple structured matrix factorisation (BSSMF), uses the distribution of the original data to impose bounds on the approximation.[21] As for other forms of matrix factorisation, two lower rank matrices are produced, however, BSSMF imposes the interval found in the origin data upon the decomposition. Thus, when applied to spectral data, such restrictions should enhance the interpretability of the dominant spectral patterns that are identified.

In this proof-of-concept study, we constructed a portable fibre optic Raman system for the study of biofluids in clinical environments. We studied serum samples collected from ALS patients at two time points and analysed the Raman spectra using BSSMF. We compared Raman data to established clinical measures of disease severity, standard clinical analytes known as promising biomarkers of disease activity (C-reactive protein,[22] ferritin[23] and complement[24]) and the leading new serum biomarker for ALS (neurofilament-light, NfL[25,26]). Our data show promise in optimising the biomarker potential of serum Raman studies in ALS.

# Methods

### Participants and clinical assessments

Samples were collected as part of A Multicentre Biomarker Research Strategy in ALS (AMBRoSIA) study, a longitudinal, observational biomarker study. The study was approved by an NHS Research Ethics committee (reference: 16/LO/2136).

66 patients were recruited at their first presentation to the Royal Hallamshire Hospital, Sheffield, UK. After written consent was obtained blood samples were collected and clinical measures of disease completed. Forced vital capacity (FVC), assessment of respiratory function, was measured using a handheld spirometer and a percentage of the patient's predicted value calculated using subject age and height. The ALS Functional Rating Scale-Revised (ALSFRS-R), the established symptom severity score for ALS, was completed. A baseline disease progression rate (DPR) was calculated as (48-ALSFRS-R)/(months from symptom onset at the time of first sample collection). A second data collection visit, comprising repeat clinical assessments and venepuncture, was undertaken 4 months after the first visit in $n = 27$ patients.

### Serum assays

Following venepuncture, samples were separated out for further processing of the different biomarkers. For ferritin, CRP and complement (C3 and C4), samples were sent to the Clinical Chemistry and Clinical Immunology laboratories at Sheffield Teaching Hospitals NHS Foundation Trust. For NfL and Raman studies, serum was separated from blood (centrifuged at 3500 rpm at 4 °C for 10 minutes) and stored in liquid nitrogen. For NfL, these were thawed on ice and quantified using the Mesoscale Discovery (MSD) R-PLEX electrochemiluminescent (ECL)-ELISA platform, as per the manufacturer's instructions.

### Raman spectroscopy

Raman spectra were obtained from the samples of all patients using a custom made, fibre optic coupled, liquid measurement cell (Fig. 1; Clifton Photonics Ltd, Bristol, UK). This cell focusses laser excitation into a 40 μl disposable aluminium sample container with an objective lens of focal length 25 mm and collects scattered light in reflection mode. The internal optics provide filtering for the rejection of elastically scattered light and clean-up of the incident laser beam. The collection optics have focal length 25 mm and numerical aperture of 0.22. Optical fibres coupled the cell to an 830 nm diode laser (Process Instruments, Salt Lake City, USA) and a Raman Explorer spectrometer (Headwall Photonics Inc. Bolton, Massachusetts, USA). The spectrometer was used in conjunction with an Andor iDus 420 CCD and Andor Solis software (Andor Technology ltd, Belfast UK) for data acquisition.

Acquisition of Raman spectra was undertaken in a windowless hospital clinic room. Serum samples were thawed on ice and 25 μl pipetted into the aluminium sample holder. A laser power of 60 mW measured at the sample was used. PTFE spectra were collected at the start of each recording session for wavenumber calibration. In addition, a background signal was collected with an empty aluminium pan *in situ* within the recording chamber. For each serum sample, an 8 seconds exposure was used, and 20 spectra were recorded (total recording time therefore 160 seconds). The individual spectra were then averaged prior to analysis. Replicates were taken from 20 samples chosen at random;

 This journal is © The Royal Society of Chemistry 2022

Open Access Article. Published on 28 September 2022. Downloaded on 4/18/2026 7:40:22 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

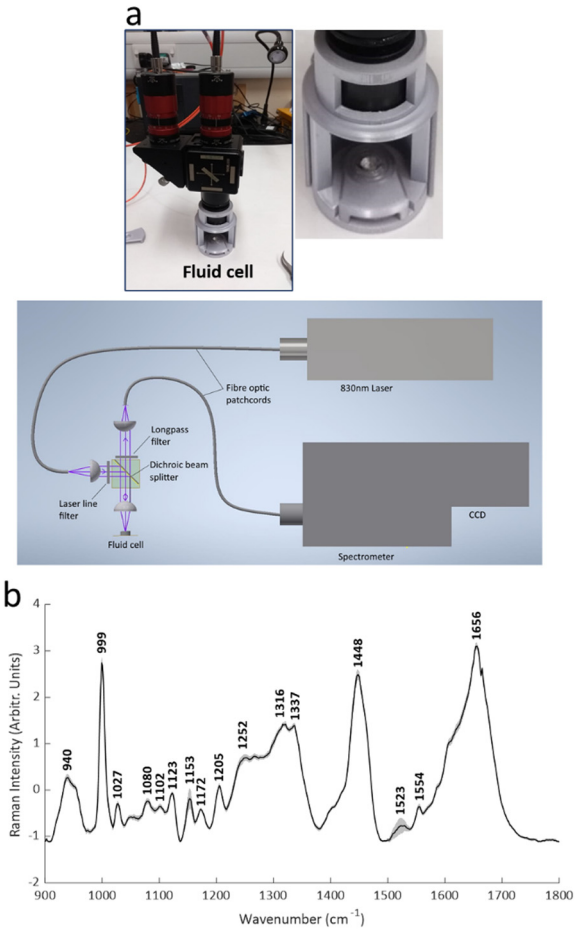


**Fig. 1** The portable fibre optic Raman system. (a) The 3D printed fluid cell and schematic of the system. (b) Mean (± standard deviation) spectra (all samples).

for these, a second 25 µl of serum was taken from the main aliquot and the spectral collection process noted above repeated.

### Analysis

Spectral analyses were undertaken using MATLAB (Version R2021b, The MathWorks, Inc., Natick, MA, USA). Analyses were undertaken across all spectra (*i.e.* visit 1 and visit 2), with separation of the two visits undertaken *post-hoc*. Spectra were first windowed between 900 and 1800 $cm^{-1}$. At <900 $cm^{-1}$ the spectra were dominated by silica-related artefact from the fibre optics. At >1800 $cm^{-1}$ the spectra consisted of non-biological noise. Windowed spectra were interpolated to integer wavenumber spacings, followed by background subtraction using the rubber band algorithm[27] and standard normal variate (SNV) normalisation. Peaks were identified using the findpeaks MATLAB function.

BSSMF[21] was also applied across all spectra *i.e.* visit 1 and visit 2 were analysed as one dataset. The data are built up as an $n \times m$ matrix, $\boldsymbol{X}$, where $n$ is the 900 wavenumbers and $m$ is the number of spectra (herein, 66 first visit samples plus 27 seconds visit samples equals a total of 93 spectra). For a rank $r$ factorisation, (where the rank, $r$, represents the number of spectral patterns output from the factorisation) BSSMF approximates $\boldsymbol{X}$ as the product of two low rank matrices:

$$\boldsymbol{X} \approx \boldsymbol{WH},$$

where $\boldsymbol{W}$ has $n$ rows and $r$ columns, and $\boldsymbol{H}$ has $r$ rows and $m$ columns. The unique characteristic of the BSSMF method is that entries in each column of $\boldsymbol{W}$ are bounded to belong in the interval of the observed dataset $\boldsymbol{X}$. Thus, the data in the matrix $\boldsymbol{W}$ represent the dominant spectral patterns in the dataset and are termed 'modes'. By bounding these modes into the interval of the observed data range we identify spectra which are physically realistic to observe. The columns of $\boldsymbol{H}$ belong to the unit simplex; these are the weightings and represent the relative importance of each spectral pattern (*i.e.* mode) to each sample. To select the rank of the decomposition, $r$, that is, the number of spectral patterns to be found, the root mean square residual ($\boldsymbol{R}$) between replicate data of 20 samples was calculated:

$$R = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left|\boldsymbol{X1}_{ij} - \boldsymbol{X2}_{ij}\right|^2}{n \times m}}$$

where $\boldsymbol{X1}$ is the first run of the replicates and $\boldsymbol{X2}$ is the repeat set (see ESI Fig. S1† for a subtraction spectrum of the technical replicates). For the matrix reconstruction, the root mean square residual ($\boldsymbol{D}$) between the dataset ($\boldsymbol{X}$) and the approximation ($\boldsymbol{WH}$) was determined:

$$D = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left|\boldsymbol{X}_{ij} - (WH)_{ij}\right|^2}{n \times m}}$$

The rank was chosen when the decomposition residual became smaller than the deviation between replicates (*i.e.* when $\boldsymbol{D} < \boldsymbol{R}$). This was satisfied for a rank of 5; thus, 5 dominant spectral modes were found (M1–M5). All subsequent analyses were applied to the weighting matrix, $\boldsymbol{H}$. These included Pearson correlations with the different markers of disease (NfL, CRP, ferritin and complement, FVC, ALSFRS-R, DPR) which were undertaken using visit 1 data. For assessment of visit 1 *vs.* visit 2 repeated measures one-way analysis of variance with a false discovery rate correction ($Q$ = 0.05) was performed using GraphPad Prism (Version 9).

In addition, principal component analysis (PCA) was applied to the 5 spectral weightings of the entire dataset, and the direction of maximum variance (PC1) was calculated (BSSMF-PCA). Considering the $j^{th}$ sample, the BSSMF-PCA value is given by $\sum_{r=1}^{5}(p^{[r]} \times \boldsymbol{H}_{rj})$, where $p^{[r]}$ is the PC1 coefficient for the $r^{th}$ dimension. For identification of patterns associated with more/less severe disease the five modes were split into two subgroups, depending upon whether the respective coefficient, $p^{[r]}$, was positive or negative. If a mode has a positive coefficient, then as its weighting increases, a relative increase in the BSSMF-PCA value is observed. By contrast, increasing weight for the modes with negative coefficients results in a relative decrease in BSSMF-PCA. The PC coefficients were then

Open Access Article. Published on 28 September 2022. Downloaded on 4/18/2026 7:40:22 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

used to derive a linear combination of the modes for the two subgroups (more/less severe disease). To do this, each BSSMF mode in the subgroup was multiplied by the absolute value of the respective PC1 coefficient, $|p^{[r]}|$. Hence, by assessing how the BSSMF-PCA value correlates with disease it is possible to assign labels of more severe and less severe disease to these two combined spectral patterns (as shown in Fig. 2b). For the visit 1 *vs.* visit 2 analysis, the paired data (mode weights and BSSMF-PCA scores) were identified for *post hoc* analysis and the same PC coefficient procedure applied.

The sample size for the number of patients required in a hypothetical clinical trial was calculated at 5% significance and 80% power:[28]

$$\text{Number of patients (sample size)} = 2 \times \left( \frac{Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}}{\text{ES} \times \frac{p}{100}} \right)^2$$

where $Z$ is the standard normal distribution for the respective significance level ($\alpha = 0.05$) and power ($1 - \beta = 80\%$), $p$ is the treatment effect percentage and ES is the effect size of the biomarker, calculated by assessing the mean change over time ($\mu$) and the standard deviation of the change ($\sigma$):

$$\text{ES} = \frac{\mu}{\sigma}$$

## Results and discussion

### Baseline data

Patient characteristics are shown in Table 1. Raman spectra were obtained and Raman spectra demonstrated features similar to those seen in other blood biofluid studies,[14,29] with prominent peaks relating to phenylalanine (999 $cm^{-1}$), the amide III region (between 1205 $cm^{-1}$ and 1340 $cm^{-1}$), the $CH_2$

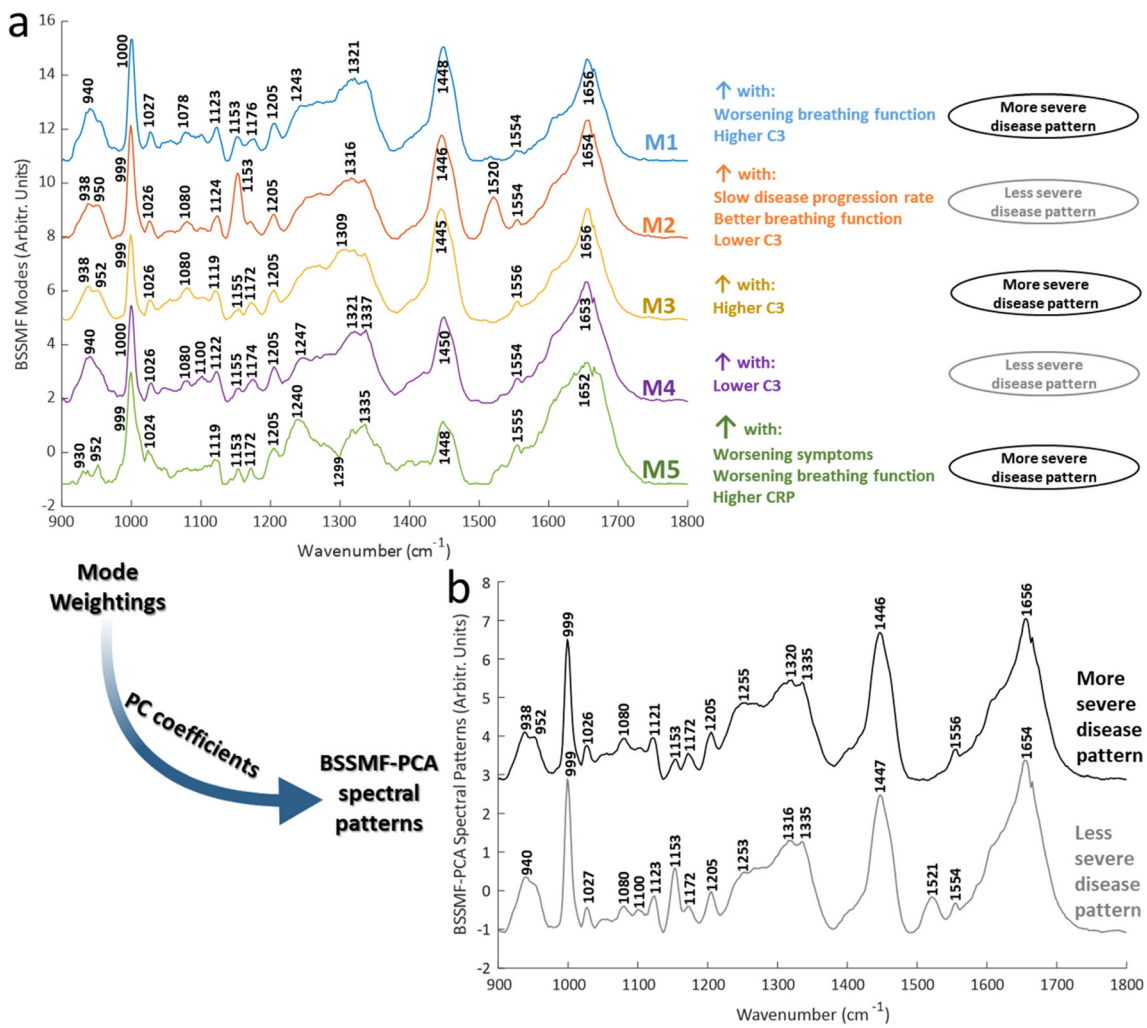


Fig. 2 Spectral patterns arising from the BSSMF and their linear combinations. (a) Modes from the BSSMF factorisation of all data. Visit 1 outputs were then correlated with visit 1 clinical/biomarker data to established associations with disease severity (n=66 samples from 66 patients). (b) Modes were combined with PCA and then the relevant PC1 coefficients used to derive spectral patterns associated with worsening disease (modes 1, 3, 5) and less severe disease (modes 2, 4).

 This journal is © The Royal Society of Chemistry 2022

Open Access Article. Published on 28 September 2022. Downloaded on 4/18/2026 7:40:22 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

**Table 1** Clinical details of the ALS patients

| | |
|---|---|
| Mean age (s.d.) | 62 years (12) |
| Gender (*n*, %) | Male: 37 (56%)<br>Female: 29 (44%) |
| Site of disease onset (*n*, %) | Limb: 55 (84%)<br>Bulbar: 8 (12%)<br>Respiratory: 1 (2%)<br>Cognitive: 1 (2%) |
| Mean disease duration (s.d.) | 22 months (21.7) |
| Mean ALSFRS-R score (range) | 37 (17–48) |
| Mean DPR (range) | 0.8 (0–6) |
| Mean % predicted FVC (range) | 86 (12–141) |

deformation of lipids/proteins (1448 $cm^{-1}$) and amide I region (1650–1660 $cm^{-1}$) (Fig. 1). Tentative assignments were taken from the literature (see ESI Table S1† for references).

BSSMF was undertaken and the five modes demonstrated similar prominent peaks to the raw spectra (Fig. 2a). Shifts were evident for certain peaks. These included phenylalanine peaks (999/1000 and 1024–1027 $cm^{-1}$), the $CH_2$ deformation of lipids/proteins (1446–1450 $cm^{-1}$) and α-helical protein content (938–942 $cm^{-1}$). There were also differences in the prominence, or even presence/absence of other peaks, including 1078–1080 $cm^{-1}$ (present in modes M1–M4), 1153 $cm^{-1}$ (present in M1, M2, M5), 1240 $cm^{-1}$ (M5), 1520 $cm^{-1}$ (prominent in M2). See ESI Table S1† for further mode peak assignments.

Correlations between spectra obtained from samples taken at the baseline visit (*i.e.* visit 1, $n = 66$) and clinical/biochemical characteristics were explored using mode weights. These demonstrated that some modes were associated with more severe disease (M1, M3, M5), and some with less severe disease (M2, M4; Fig. 2a; see ESI Table S2† for full correlation statistics). Particularly prominent correlations were seen for FVC and the inflammatory/immune proteins C-reactive protein (CRP) and C3. FVC is an established respiratory assessment used in ALS. CRP is an acute phase protein which acts as an activator of the complement system, a key component of the innate immune system. Both are increased in ALS and associated with a more severe form of disease.[22,30–34] In keeping with these prior reports, in our analyses modes associated with increasing CRP/C3 were also associated with worsening disease.

The combined BSSMF-PCA metric was also found to be associated with more severe disease (see ESI Table S2† for correlations and ESI Table S3† for PC coefficients). Thus, more/less severe disease spectral patterns were obtained (Fig. 2b). By using multiple measures of disease this approach provides a more comprehensive means of findings spectral features associated with worsening disease. These patterns demonstrated peak differences related to protein structure (938–952, 1253/55, 1316/1320, 1654/6 $cm^{-1}$) and lipids (1100 $cm^{-1}$). Peak shifts such as these have been described in other diseases, such as malignancy[35,36] and necrosis.[37] The understanding of the exact mechanisms driving such shifts is incomplete but may involve altered inter-molecular interactions.[38] Thus, changes in chemical bond length or symmetry, perhaps related to structural alterations of proteins, may drive such changes.

While environmental factors such as sample temperature could also contribute, we noted that other peaks demonstrated alterations in shape and intensity, or were absent in some modes (*e.g.* 1335, 1520 $cm^{-1}$). Instrument calibration was performed prior to each recording session and average spectra from each visit manifested several prominent common peaks (*e.g.* 999, 1080, 1123, 1153, 1205, 1554, 1656 $cm^{-1}$; Fig. 3). These observations would suggest that systematic differences in recording conditions were not the dominant reason for peak differences found in the matrix decomposition.

Interestingly, peak alterations indicative of changes in protein structure have also been identified in Raman serum studies of other neurodegenerative conditions.[15] These disorders, which include ALS, are increasingly recognised as proteinopathies, in which misfolded proteins play a key driving

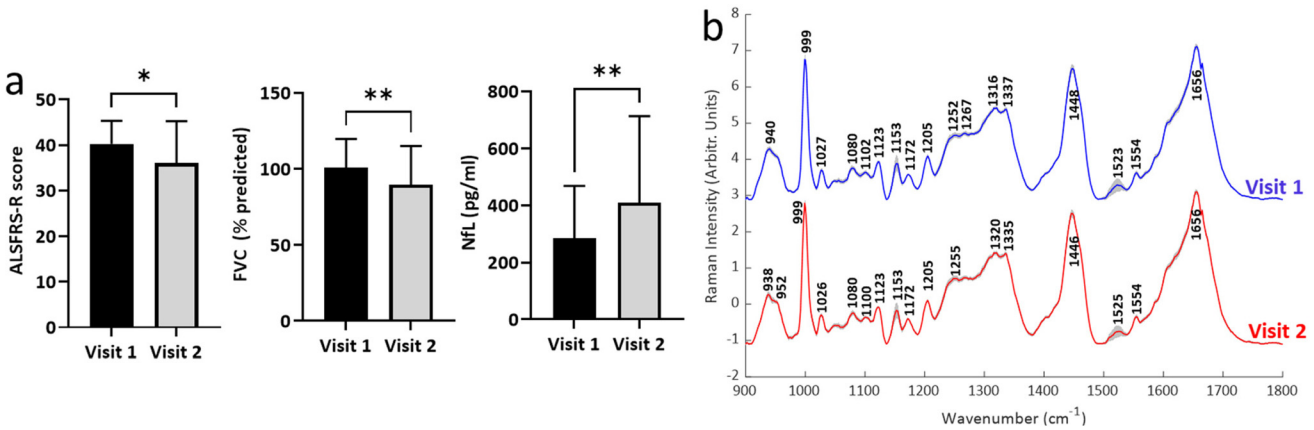


**Fig. 3** Change in clinical and biochemical measures of disease. (a) Changes in key clinical and biochemical measures of disease ($n = 27$; see ESI fig. S2† for data from other measures). (b) Mean (± standard deviation) for Raman spectra obtained from patients attending both visit 1 and visit 2 ($n = 27$). *$p < 0.05$, **$p < 0.01$.

Open Access Article. Published on 28 September 2022. Downloaded on 4/18/2026 7:40:22 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

role in disease initiation and spread.[39] For example, analysis of protein aggregates within blood from ALS patients demonstrated the presence of a large number of proteins involved in the proteasome, the clearance system for defective proteins, illustrating that the effects of abnormal proteins can spread across the blood–brain barrier.[40] Assessing whether Raman spectroscopy could provide a simple point of care assay on such aggregates would be a useful future work. In addition, peaks associated with carotenoids were more prominent in the less severe disease pattern (1153, 1521 $cm^{-1}$). Carotenoids exhibit free radial scavenging and interact with the Nrf2 signalling pathway, both of which have been implicated in ALS.[41,42] Interestingly, increased carotenoid intake has been shown to correlate with less severe disease,[43] although the relationship between intake and disease is likely to be complex.[44]

### Change over time: visit 1 *vs.* visit 2

Second visit assessments were performed 4 months after the first visit in $n$ = 27 patients. Established clinical (ALSFRS-R, FVC) and biochemical biomarkers (NfL) demonstrated significant change between the two visits, indicating disease progression in these patients (Fig. 3a; see ESI Fig. S2† for remaining biomarkers). The weightings of modes M3, M4 and M5 from the BSSMF also demonstrated significant changes (Fig. 4a). The BSSMF-PCA scores increased over time (moving from negative to positive; Fig. 4b), which represents a shift from the less severe to the more severe spectral pattern in the earlier analysis. Significant differences were observed in the visit 1/visit 2 BSSMF-PCA scores ($p$ = 0.0002; Fig. 4b).

Several reports utilising more traditional techniques for monitoring ALS (such electrophysiological and imaging studies) have demonstrated significant changes in disease state within 4 months.[2,3] The most promising serum biomarker at present is NfL, a structural protein found in axons which is released when axons degenerate. Recent multi-centre work from the AMBRoSIA study group found that levels increased over time and modelling its inclusion as a trial outcome measure reduced sample size requirements.[26]

To investigate the potential of our Raman paradigm in a longitudinal setting, a hypothetical clinical trial was constructed and sample sizes for the number of patients required (not the volume of serum) were calculated (Table 2). In comparison to established ALS biomarkers (ALSFRS-R, FVC, NfL), Raman data from modes M3-5 and the linear combination of all modes (BSSMF-PCA) required smaller sample sizes.

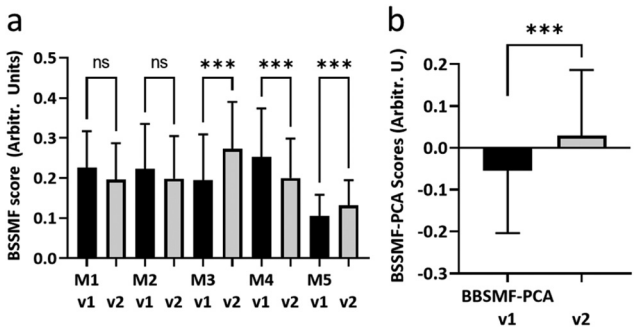


**Fig. 4** Longitudinal BSSMF mode and BSSMF-PCA changes. (a) Changes in individual BSSMF mode weights. (b) Change in BSSMF-PCA scores from visit 1 and visit 2. ns – not significant, ***$p$ < 0.001.

**Table 2** Hypothetical clinical trial sample sizes

| | Effect size | Sample size: 50% treatment effect | Sample size: 20% treatment effect |
|---|---|---|---|
| **Raman** | | | |
| M1 | −0.35 | 529 | 3310 |
| M2 | −0.26 | 941 | 5883 |
| M3 | 0.89 | 80 | 502 |
| M4 | −0.85 | 87 | 547 |
| M5 | 0.84 | 89 | 557 |
| BSSMF-PCA | 0.85 | 87 | 547 |
| ALSFRS-R | −0.44 | 323 | 2024 |
| FVC | −0.78 | 102 | 642 |
| Neurofilament-light | 0.7 | 128 | 798 |

Fibre optic Raman systems have been investigated for *in vivo* applications, for example, for deployment during surgery and endoscopy.[16] Serum studies have largely employed either standard microscope formats (*e.g.*[45]), but smaller portable systems are being investigated (*e.g.*[46]). More recently, optofluidic systems, which combine microfluidics to separate blood components with Raman, have been developed.[47,48] Whether using a portable system such as ours, or a miniature optofluidic design, being able to take Raman to the clinic offers the possibility for real-time molecular fingerprinting at the point of care. Such information could help guide clinical decision making in the precision medicine era. This might benefit a range of diseases, as well as a variety of care settings, for example, emergency care, surgery and outpatient clinics. In addition to immediate patient benefit, reduced decision-making times can also have potential knock-on effects in cost benefit analyses. Developing an evidence base for point of care Raman spectroscopy through high quality, multi-centre studies with parallel health economic assessments, will be a priority for the field as it continues to develop.

The complexity of Raman spectra and the subtle molecular differences found in disease necessitates multivariate analysis, often beginning with dimensionality reduction. The typical approach (PCA) can distort the spectral profile and present physically unrealistic results. Non-negative matrix factorisation can overcome some of these issues and provide a quantitative measure of the importance of different spectral patterns to disease.[49,50] The bounded simplex structured matrix factorisation used here does not impose the non-negative constraint; instead the factorisation matrix $\boldsymbol{W}$ is bound within an interval defined by the original data.[21] Thus, the data within $\boldsymbol{W}$ can be interpreted in the same way as the original data ($\boldsymbol{X}$). Bounded component analysis approaches are similar in their approach, although more difficult to interpret.[51] In our analyses we also combined modes to gain a more complete picture of the biochemical changes occurring over time. Analysing Raman data in this fashion could facilitate clinical trials, with reduced

 This journal is © The Royal Society of Chemistry 2022

Open Access Article. Published on 28 September 2022. Downloaded on 4/18/2026 7:40:22 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

sample size numbers reducing trial costs. This, in turn, can help utilise resources more efficiently, meaning more studies could be funded in the search for new treatments. Furthermore, the integration of advanced analysis algorithms with portable fibre optic technology, or indeed other platforms, can aid the development of biomedical Raman spectroscopy as a point of care technology.

## Conclusions

Herein, we have presented a novel, portable fibre optic system for the study of human biofluids. As the development of biomarkers for monitoring ALS is an area of intense investigation, we tested the system on serum from ALS patients collected at a 4-month interval. We analysed data using a matrix decomposition technique with enhanced physical interpretation constraints. Key spectral features appeared to relate to protein structure. Further assessment of serum samples over longer time periods could provide additional insights into the complex biochemical changes occurring the serum of patients with ALS and facilitate more efficient clinical trial design.

## Author contributions

J. J. P. A.: conceptualisation, formal analysis, investigation, methodology, visualisation, writing – original data. N. S. V.: conceptualisation, investigation, project administration, writing – review & editing; C. S.: formal analysis, writing – review & editing; V. K.: writing – review & editing; M. R. T. – funding acquisition, writing – review & editing; A. M. – funding acquisition, writing – review & editing; J. C. C. D,: conceptualisation, methodology, resources, writing – review & editing; P. J. S.: conceptualisation, funding acquisition, resources, project administration, writing – review & editing.

## Conflicts of interest

J. C. C. D. is the chief executive office of Clifton Photonics Ltd, a company that designs photonic instruments for industrial and research use.

## Acknowledgements

The work was supported by Motor Neurone Disease Association (AMBRoSIA 972-797) and the NIHR Sheffield Biomedical Research Centre (IS-BRC-1215-20017). P. J. S. is supported as a National Institute for Health Research (NIHR) Senior Investigator (NF-SI-0617–10077). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising. The authors gratefully acknowledge the patients who so enthusiastically participated in this study.

## References

1 O. Hardiman, A. Al-Chalabi, A. Chio, E. M. Corr, G. Logroscino, W. Robberecht, P. J. Shaw, Z. Simmons and L. H. van den Berg, *Nat. Rev. Dis. Primers*, 2017, **3**, 17071.
2 T. M. Jenkins, J. J. P. Alix, C. David, E. Pearson, D. G. Rao, N. Hoggard, E. O'Brien, K. Baster, M. Bradburn, J. Bigley, C. J. McDermott, I. D. Wilkinson and P. J. Shaw, *J. Neurol. Neurosurg. Psychiatry*, 2018, **89**, 248–255.
3 Y. Chan, J. J. P. Alix, C. Neuwirth, P. E. Barkhaus, J. Castro, T. M. Jenkins, C. J. McDermott, P. J. Shaw, M. de Carvalho, S. Nandedkar, E. Stalberg and M. Weber, *Muscle Nerve*, 2021, **65**, 203–210.
4 M. Benatar, L. Zhang, L. Wang, V. Granit, J. Statland, R. Barohn, A. Swenson, J. Ravits, C. Jackson, T. M. Burns, J. Trivedi, E. P. Pioro, J. Caress, J. Katz, J. L. McCauley, R. Rademakers, A. Malaspina, L. W. Ostrow, J. Wuu and C. R. Consortium, *Neurology*, 2020, **95**, e59–e69.
5 P. C. Chen, D. Wu, C. J. Hu, H. Y. Chen, Y. C. Hsieh and C. C. Huang, *J. Neurol. Sci.*, 2020, **418**, 117070.
6 R. Waller, E. F. Goodall, M. Milo, J. Cooper-Knock, M. Da Costa, E. Hobson, M. Kazoka, H. Wollff, P. R. Heath, P. J. Shaw and J. Kirby, *Neurobiol. Aging*, 2017, **55**, 123–131.
7 N. S. Verber, S. R. Shepheard, M. Sassani, H. E. McDonough, S. A. Moore, J. J. P. Alix, I. D. Wilkinson, T. M. Jenkins and P. J. Shaw, *Front. Neurol.*, 2019, **10**, 291.
8 G. Devitt, K. Howard, A. Mudher and S. Mahajan, *ACS Chem. Neurosci.*, 2018, **9**, 404–420.
9 M. Paraskevaidi, P. L. Martin-Hirsch and F. L. Martin, *Mol. Neurodegener.*, 2018, **13**, 20.
10 Q. J. Zhang, Y. Chen, X. H. Zou, W. Hu, X. L. Lin, S. Y. Feng, F. Chen, L. Q. Xu, W. J. Chen and N. Wang, *J. Biophotonics*, 2019, **12**, e201900012.
11 Q. J. Zhang, Y. Chen, X. H. Zou, W. Hu, M. L. Ye, Q. F. Guo, X. L. Lin, S. Y. Feng and N. Wang, *Ann. Clin. Transl. Neurol.*, 2020, **7**, 2010–2018.
12 C. Carlomagno, P. I. Banfi, A. Gualerzi, S. Picciolini, E. Volpato, M. Meloni, A. Lax, E. Colombo, N. Ticozzi, F. Verde, V. Silani and M. Bedoni, *Sci. Rep.*, 2020, **10**, 10175.
13 J. Langer, D. Jimenez de Aberasturi, J. Aizpurua, R. A. Alvarez-Puebla, B. Auguie, J. J. Baumberg, G. C. Bazan, S. E. J. Bell, A. Boisen, A. G. Brolo, J. Choo, D. Cialla-May, V. Deckert, L. Fabris, K. Faulds, F. J. Garcia de Abajo, R. Goodacre, D. Graham, A. J. Haes, C. L. Haynes, C. Huck, T. Itoh, M. Kall, J. Kneipp, N. A. Kotov, H. Kuang, E. C. Le Ru, H. K. Lee, J. F. Li, X. Y. Ling, S. A. Maier, T. Mayerhofer, M. Moskovits, K. Murakoshi, J. M. Nam, S. Nie, Y. Ozaki, I. Pastoriza-Santos, J. Perez-Juste, J. Popp, A. Pucci, S. Reich, B. Ren, G. C. Schatz, T. Shegai, S. Schlucker, L. L. Tay, K. G. Thomas, Z. Q. Tian, R. P. Van Duyne, T. Vo-Dinh, Y. Wang, K. A. Willets, C. Xu, H. Xu, Y. Xu, Y. S. Yamamoto, B. Zhao and L. M. Liz-Marzan, *ACS Nano*, 2020, **14**, 28–117.
14 C. F. Morasso, D. Sproviero, M. C. Mimmi, M. Giannini, S. Gagliardi, R. Vanna, L. Diamanti, S. Bernuzzi, F. Piccotti,

Open Access Article. Published on 28 September 2022. Downloaded on 4/18/2026 7:40:22 PM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

M. Truffi, O. Pansarasa, F. Corsi and C. Cereda, *Nanomedicine*, 2020, **29**, 102249.
15 A. Huefner, W. L. Kuan, S. L. Mason, S. Mahajan and R. A. Barker, *Chem. Sci.*, 2020, **11**, 525–533.
16 O. Stevens, I. E. Iping Petterson, J. C. Day and N. Stone, *Chem. Soc. Rev.*, 2016, **45**, 1919–1934.
17 S. Guo, J. Popp and T. Bocklitz, *Nat. Protoc.*, 2021, **16**, 5426–5459.
18 C. L. M. Morais, K. M. G. Lima, M. Singh and F. L. Martin, *Nat. Protoc.*, 2020, **15**, 2143–2162.
19 D. D. Lee and H. S. Seung, *Nature*, 1999, **401**, 788–791.
20 R. Wu, W.-K. Ma and X. Fu, A stochastic maximum-likelihood framework for simplex structured matrix factorization, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 2557–2561.
21 O. V. Thanh, N. Gillis and F. Lecron, Bounded Simplex-Structured Matrix Factorization, in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 9062–9066.
22 C. Lunetta, A. Lizio, E. Maestri, V. A. Sansone, G. Mora, R. G. Miller, S. H. Appel and A. Chio, *JAMA Neurol.*, 2017, **74**, 660–667.
23 E. F. Goodall, M. S. Haque and K. E. Morrison, *J. Neurol.*, 2008, **255**, 1652–1656.
24 J. Ganesalingam, J. An, C. E. Shaw, G. Shaw, D. Lacomis and R. Bowser, *J. Neurochem.*, 2011, **117**, 528–537.
25 E. Feneberg, P. Oeckl, P. Steinacker, F. Verde, C. Barro, P. Van Damme, E. Gray, J. Grosskreutz, C. Jardel, J. Kuhle, S. Koerner, F. Lamari, M. D. M. Amador, B. Mayer, C. Morelli, P. Muckova, S. Petri, K. Poesen, J. Raaphorst, F. Salachas, V. Silani, B. Stubendorff, M. R. Turner, M. M. Verbeek, J. H. Weishaupt, P. Weydt, A. C. Ludolph and M. Otto, *Neurology*, 2018, **90**, e22–e30.
26 A. G. Thompson, E. Gray, N. Verber, Y. Bobeva, V. Lombardi, S. R. Shepheard, O. Yildiz, E. Feneberg, L. Farrimond, T. Dharmadasa, P. Gray, E. C. Edmond, J. Scaber, D. Gagliardi, J. Kirby, T. M. Jenkins, P. Fratta, C. J. McDermott, S. G. Manohar, K. Talbot, A. Malaspina, P. J. Shaw and M. R. Turner, *Brain Commun.*, 2022, **4**, fcac029.
27 M. Paraskevaidi, C. L. M. Morais, K. M. G. Lima, J. S. Snowden, J. A. Saxon, A. M. T. Richardson, M. Jones, D. M. A. Mann, D. Allsop, P. L. Martin-Hirsch and F. L. Martin, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E7929–E7938.
28 T. P. Ryan, *Sample Size Determination and Power*, John Wiley & Sons, Inc., 2013.
29 C. A. Jenkins, R. A. Jenkins, M. M. Pryse, K. A. Welsby, M. Jitsumura, C. A. Thornton, P. R. Dunstan and D. A. Harris, *Analyst*, 2018, **143**, 6014–6024.
30 X. Chen, W. Feng, R. Huang, X. Guo, Y. Chen, Z. Zheng and H. Shang, *J. Neurol. Sci.*, 2014, **347**, 90–95.
31 I. L. Goldknopf, E. A. Sheta, J. Bryson, B. Folsom, C. Wilson, J. Duty, A. A. Yen and S. H. Appel, *Biochem. Biophys. Res. Commun.*, 2006, **342**, 1034–1039.
32 C. Cui, J. Sun, Y. Pawitan, F. Piehl, H. Chen, C. Ingre, K. Wirdefeldt, M. Evans, J. Andersson, J.-J. Carrero and F. Fang, *Brain Commun.*, 2020, **2**(2), fcaa152.
33 S. Kharel, R. Ojha, V. Preethish-Kumar and R. Bhagat, *Brain Behav.*, 2022, **12**, e2532.
34 S. E. Parker, A. M. Hanton, S. N. Stefanou, P. G. Noakes, T. M. Woodruff and J. D. Lee, *Neurobiol. Dis.*, 2019, **127**, 223–232.
35 C. J. Kirkby, J. Gala de Pablo, E. Tinkler-Hundal, H. M. Wood, S. D. Evans and N. P. West, *Analyst*, 2021, **146**, 581–589.
36 Y. Chen, J. Dai, X. Zhou, Y. Liu, W. Zhang and G. Peng, *PLoS One*, 2014, **9**, e93906.
37 E. Brauchle, S. Thude, S. Y. Brucker and K. Schenke-Layland, *Sci. Rep.*, 2014, **4**, 4698.
38 H. Ma, S. Liu, N. Zheng, Y. Liu, X. X. Han, C. He, H. Lu and B. Zhao, *Anal. Chem.*, 2019, **91**, 9376–9381.
39 D. Gagliardi, N. Bresolin, G. P. Comi and S. Corti, *Cell. Mol. Life Sci.*, 2021, **78**, 561–572.
40 R. Adiutori, F. Puentes, M. Bremang, V. Lombardi, I. Zubiri, E. Leoni, J. Aarum, D. Sheer, S. McArthur, I. Pike and A. Malaspina, *Brain Commun.*, 2021, **3**(3), fcab148.
41 E. D'Amico, P. Factor-Litvak, R. M. Santella and H. Mitsumoto, *Free Radical Biol. Med.*, 2013, **65**, 509–527.
42 R. J. Mead, A. Higginbottom, S. P. Allen, J. Kirby, E. Bennett, S. C. Barber, P. R. Heath, A. Coluccia, N. Patel, I. Gardner, A. Brancale, A. J. Grierson and P. J. Shaw, *Free Radical Biol. Med.*, 2013, **61**, 438–452.
43 J. W. Nieves, C. Gennings, P. Factor-Litvak, J. Hupf, J. Singleton, V. Sharf, B. Oskarsson, J. A. Fernandes Filho, E. J. Sorenson, E. D'Amico, R. Goetz and H. Mitsumoto, *JAMA Neurol.*, 2016, **73**, 1425–1432.
44 V. Böhm, G. Lietz, B. Olmedilla-Alonso, D. Phelan, E. Reboul, D. Bánati, P. Borel, J. Corte-Real, A. R. de Lera, C. Desmarchelier, J. Dulinska-Litewka, J.-F. Landrier, I. Milisav, J. Nolan, M. Porrini, P. Riso, J. M. Roob, E. Valanou, A. Wawrzyniak, B. M. Winklhofer-Roob, R. Rühl and T. Bohn, *Nutr. Rev.*, 2021, **79**, 544–573.
45 C. Chen, W. Wu, C. Chen, F. Chen, X. Dong, M. Ma, Z. Yan, X. Lv, Y. Ma and M. Zhu, *J. Raman Spectrosc.*, 2021, **52**, 1798–1809.
46 H. F. Nargis, H. Nawaz, H. N. Bhatti, K. Jilani and M. Saleem, *Spectrochim. Acta, Part A*, 2021, **246**, 119034.
47 J. J. S. Rickard, V. Di-Pietro, D. J. Smith, D. J. Davies, A. Belli and P. G. Oppenheimer, *Nat. Biomed. Eng.*, 2020, **4**, 610–623.
48 E. E. Storey, D. Wu and A. S. Helmy, *J. Lightwave Technol.*, 2021, **39**, 5634–5642.
49 A. L. Blee, J. C. C. Day, P. E. J. Flewitt, A. Jeketo and D. Megson-Smith, *J. Raman Spectrosc.*, 2021, **52**, 1135–1147.
50 C. N. Schooling, T. J. Healey, H. E. McDonough, S. J. French, C. J. McDermott, P. J. Shaw, V. Kadirkamanathan and J. J. P. Alix, *Physiol. Meas.*, 2021, **42**(10), DOI: 10.1088/1361-6579/ac2672.
51 S. Cruces, *IEEE Trans. Signal Process.*, 2010, **58**, 2141–2154.

 This journal is © The Royal Society of Chemistry 2022