Computational Fluorescence Suppression in Shifted Excitation Raman Spectroscopy

Nia C. Jenkins, Katjana Ehrlich, András Kufcsák, Stephanos Yerolatsitis, Susan Fernandes, Irene Young, Katie Hamilton, Harry A. C. Wood, Tom Quinn, Vikki Young, Ahsan R. Akram, James M. Stone, Robert R. Thomson, Keith Finlayson, Kevin Dhaliwal, Sohan Seth

***Abstract*— Fiber-based Raman spectroscopy in the context of *in vivo* biomedical application suffers from the presence of background fluorescence from the surrounding tissue that might mask the crucial but inherently weak Raman signatures. One method that has shown potential for suppressing the background to reveal the Raman spectra is shifted excitation Raman spectroscopy (SER). SER collects multiple emission spectra by shifting the excitation by small amounts and uses these spectra to computationally suppress the fluorescence background based on the principle that Raman spectrum shifts with excitation while fluorescence spectrum does not. We introduce a method that utilizes the spectral characteristics of the Raman and fluorescence spectra to estimate them more effectively, and compare this approach against existing methods on real world datasets.**

***Index Terms*— Biomedical, Fluorescence, Lung Tissue, Machine Learning, Optical Fiber, Raman Spectroscopy, Regularization, Shifted Excitation, Sparsity, Smoothness,**

I. Introduction

INTERVENTIONAL pulmonology is a sub-specialty of pulmonary medicine that involves diagnostic or therapeutic interventions to be undertaken by bronchoscopy involving endoscopy of the trachea to access the endobronchial tree. Recent developments in bespoke optical fibres now allows access to the distal lung through the working channel of a bronchoscope to perform precise microscopic optical imaging and spectroscopy for assessing the physiological and molecular signatures of tissue [1]. This is a promising approach for intra-procedural tumor tissue delineation with a translational drive that is motivated by relatively low-cost systems, minimally invasive instrumentation, use of non-ionizing radiation, and real-time continuous data acquisition leading to improved tissue characterisation and diagnostic performance [2] [3].

While fiber-based fluorescence spectroscopy techniques are currently making the transition from bench-to-bedside, Raman spectroscopy (RS) is heralded as a promising candidate for translation into clinics. Raman spectroscopy is a molecular spectroscopy technique offering high chemical sensitivity and specificity as well as the advantage of rapid and label-free (i.e., without tagging the molecule of interest) analysis. It investigates the inelastic scattering of light from a sample when excited with monochromatic light and allows unique structural fingerprinting of molecules to be explored. Current research has explored RS in several organs such as the colon [5], lung [6], cervix [7] and Barrett's oesophagus [8] and it is observed that there are characteristic differences between the Raman spectra of normal and abnormal tissue [9].

Raman scattering is, however, inherently weak and, especially for *in vivo* applications, the Raman signal is masked by the Raman background from the optical fiber used to reach remote areas of the human body and the autofluorescence of the surrounding tissue. Fiber-based spectroscopy relies on carrying the excitation light through the fiber, and as the excitation light propagates within the fiber, it generates a Raman scattering that is directly proportional to the fiber length [10]. Recent advancements in the use of hollow-core negative-curvature optical fibers (NCFs) where light is guided in air, instead of e.g. fused-silica optical fibers, minimizes the interactions of light with the silica of the fiber core significantly allowing strong Raman signatures to be identified [4]. Yet the intensity of the fluorescence of the surrounding tissue is often several orders of magnitude larger than the Raman signal [11], and may still mask the relatively weak characteristic peaks of the Raman spectrum that are crucial for downstream analysis (see Fig. 1).

Several methods exist to reduce background fluorescence in Raman measurements, and they can be broadly categorized into two groups, i.e., technological and computational. Tech-

The authors thank the BioResource for access to tissue (NHS Lothian BioResource, Scotland Research Ethics Service, reference 15/ES/0094). NCJ was supported through the University of Edinburgh funding award. JMS, SY and HACW are supported through an EPSRC fellowship (EP/S001123/1). SF is supported through an MRC fellowship (MR/R017794/1). ARA is supported by a CRUK Clinician Scientist Fellowship (A24867). NCJ and SS are with the School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh, EH8 9AB, UK. (e-mails: {njenkin3,sohan.seth}@ed.ac.uk). KE, AK, SF, IY, KH, TQ, VY, ARA, RRT, KF, KD, and SS are with the Translational Healthcare Technology Group, Centre for Inflammation Research, Queen's Medical Research Institute, University of Edinburgh, 47 Little France Crescent, Edinburgh, EH16 4TJ, UK. (e-mails: {Katjana.Ehrlich,A.Kufcsak,Susan.Fernandes, iyoung3 K.Hamilton,tquinn,Vikki.Young,Ahsan.Akram,Keith.Finlayson, Kev.Dhaliwal}@ed.ac.uk). SY, HACW. and JMS are with the Centre for Photonics and Photonic Materials, University of Bath, Claverton Down, Bath, BA2 7AY, UK. (emails: {S.Yerolatsitis,H.Wood,J.M.Stone}@bath.ac.uk). RRT is with Scottish Universities Physics Alliance (SUPA), Institute of Photonics and Quantum Science, Heriot-Watt University, Edinburgh, EH144AS, UK. (email: R.R.Thomson@hw.ac.uk).

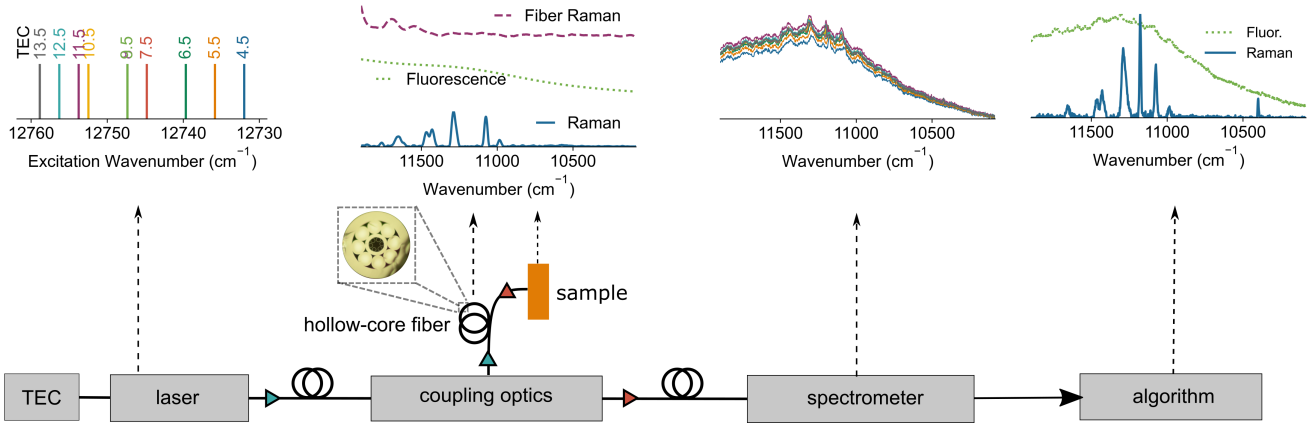

Fig. 1: The figure shows the experimental set-up for SER. From left to right, the plots show the change in excitation due to change in TEC setting; *representative* Raman, fluorescence and fiber Raman background spectra; the observed spectra due to change in excitation; and inferred Raman and fluorescence spectra after fluorescence suppression. The experimental set up is similar to that of Yerolatsitis *et al.* [4]. The arrows along hollow-core fiber show the direction of light through the system.

nological methods involve instruments designed to remove the fluorescence from measurements, for example time-gating [12][13][14] and frequency-domain methods [11] but they are often costly and require highly specified instrumentation. On the other hand, computational techniques involve statistical models to suppress fluorescence from measurements, for example polynomial fitting procedure [15] and least squares methods [16], but they often do not perform well in low signal-to-noise ratio environments and may require careful parameter selection.

A promising compromise between the two approaches is shifted excitation Raman spectroscopy (SER). The principles of SER are founded on Kasha's rule and Stokes shift. Kasha's rule states that the fluorescence emission is unaffected by a small change in excitation wavelength [17], while Stokes shift dictates that the Raman spectrum shifts by the same amount in wavenumber as the shift in excitation. SER collects multiple emission spectra by shifting the excitation by small amounts and uses statistical tools to separate these spectra into fluorescence and Raman components. Existing approaches for separating these two spectra, however, are sub-optimal since they do not explicitly utilize their spectral characteristics.

Our objectives are to, 1) perform a literature survey to review existing methods of removing fluorescence based on statistical and machine learning tools (section II), 2) present an algorithm that utilizes the underlying structure of fluorescence and Raman signals as prior information to effectively estimate them from observed spectra (section III), 3) compare existing methods and the proposed approach on real reference datasets as well as human lung tissue data over several diverse evaluation metrics (section IV), 4) present a Python package implementing these tools to underpin reproducibility.

II. Existing Approaches

We denote a function over wavenumber $\tilde{\nu}$ by lower case letters, e.g., $x(\tilde{\nu})$, where *wavenumber* in cm^{-1} is the number of wavelengths (λ in nm) in unit length or $\tilde{\nu} = 10^7/\lambda$. We denote the column vector holding the values of the function at N wavenumbers with lower case bold letters, e.g., $\mathbf{x} = [x(\tilde{\nu}_1), \ldots, x(\tilde{\nu}_N)]^\top$ where $\cdot^\top$ denotes *transpose*. We denote the n-th entry of the vector $\mathbf{x}$ using lower case letters, i.e., x_n which is equivalent to $x(\tilde{\nu}_n)$. We denote matrices using upper case bold letters, e.g., $\mathbf{X}$, and the (i,j)-th entry of the matrix as x_{ij}. We denote the n-th column of the matrix $\mathbf{X}$ as $\mathbf{x}_n$, i.e., $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$.

Let $r(\tilde{\nu})$ and $f(\tilde{\nu})$ denote the Raman and fluorescence (also referred to as *background* in this context) spectrum respectively, and let $y^{(k)}$ denote the k-th observed spectrum in SER where each observed spectrum is acquired by shifting the excitation by a small amount $\Delta\tilde{\nu}_k$ ($\Delta\tilde{\nu}_1 = 0$ for the first excitation). Each observed spectrum can be described (in a noise-free situation) as a combination of the underlying Raman and fluorescence spectra, i.e.,

$$y^{(k)}(\tilde{\nu}) = r(\tilde{\nu} - \Delta\tilde{\nu}_k) + f(\tilde{\nu}). \tag{1}$$

We assume that each observed spectrum is acquired at N equispaced wavenumbers $\tilde{\nu}_n$ with $n \in \{1, \ldots, N\}$, and that shifting the excitation by $\Delta\tilde{\nu}_k$ in wavenumbers is equivalent to shifting it by Δn_k indices, i.e,

$$r(\tilde{\nu}_n - \Delta\tilde{\nu}_k) = r(\tilde{\nu}_{n-\Delta n_k}). \tag{2}$$

However, Δn_k for $k \in \{1, \ldots, K\}$ where K is the total number of excitations or shifts (counting $\Delta n_1 = 0$ as a shift) do not have to be contiguous. Collecting the values of the spectra in vectors, we can express Eq. (1) as

$$\mathbf{y}_k = \mathbf{L}^{(\Delta n_k)}\mathbf{r} + \mathbf{f} \tag{3}$$

where $\mathbf{y}_k = [y^{(k)}(\tilde{\nu}_1), \ldots, y^{(k)}(\tilde{\nu}_N)]^\top \in \mathbb{R}^N$ is the column vector of the k-th observed spectrum, $\mathbf{L}^{(\Delta n_k)} \in \mathbb{R}^{N\times N}$ is a lower shift matrix that shifts the vector $\mathbf{r}$ by Δn_k indices, i.e.,

$$l_{ij}^{(\Delta n_k)} = \begin{cases} 1 & \text{if} \quad i = j + \Delta n_k \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

and $\mathbf{r} \in \mathbb{R}^N$ and $\mathbf{f} \in \mathbb{R}^N$ are the vectors of the Raman and fluorescence spectra respectively. *Fluorescence suppression* in

TABLE I: Summary of existing fluorescence suppression methods and their applications.

Paper	Method	Parameter	Initialization	K	Experimental Conditions	Samples
Zhang *et al.* [15]	AIRPLS	Background smoothing		1	BWTEK i-Raman-785 Excitation: 785 nm Acquisition time: 5-7.5 s	Prednisone acetate and glibenclamide tablets
Gebrekidan *et al.* [18]	RSERDS	Center line estimate smoothing, ALS smoothing, ALS asymmetry		2	Dual fiber Raman probe design with distal end optics Excitation: 784-785 nm Power: 100 mW Acquisition time: 10 s	*Ex vivo* pig tissue and human oral tissue
Mazilu *et al.* [19]	PCA				Free space Raman microscope Excitation: 784.96-785.04 nm Power: 5 mW Acquisition time: 10 s	Polystyrene
McCain *et al.* [20]	PIP	Max. iterations	$f_n = \min_k y_{nk}$ $r_n =$ estimate given $\mathbf{f}$	8	Multicore fiber Raman probe Excitation: 782.6-794.3 nm Power: 5-50 mW Acquisition time: 128 s	Dye dissolved in ethanol
Cooper *et al.* [21]	PIP	Max. iterations	$f_n = \min_k y_{nk}$ $r_n = \mathrm{std}_k\, y_{nk}$	4	Free space Raman microscope Excitation: 784.6-785.3 nm Power: 50 mW Acquisition time: 3-141 s	Dimethyl gloxime, acenaphthylene, 4-bromo-N, N-dimethylaniline
Marshall and Cooper [22]	PIP	Max. iterations	$f_n = \min_k y_{nk}$ $r_n = \max_k y_{nk}$	32	Free space Raman microscope Excitation: 849.68-852 nm Acquisition time: 320 s	Laser dye
Mørup *et al.* [23]	SNMF		Random	9		Simulation
Mørup *et al.* [24]	SICA		Random	10		Simulation
Proposed	MSERS	Raman sparsity, fluorescence smoothing	$f_n = \min_k y_{nk}$ $r_n = \mathrm{std}_k\, y_{nk}$	10	Multicore fiber Raman probe Excitation: 783.8-785.4 nm Power: 20 mW Acquisition time: 100-500 s	Cyclohexane, sesame oil, and lung tissue

SER addresses the problem of estimating the Raman spectrum $\mathbf{r}$ given K *noisy* spectra collected at different shifted excitations. The available methods for fluorescence suppression (in Raman spectroscopy) can be broadly divided into three categories based on how many shifted measurements are considered by the method. Table I summarizes these methods.

A. Single spectrum

Adaptive iteratively reweighted penalized least squares (AIRPLS) estimates the fluorescence recursively from a single spectrum $\mathbf{y}$ and subtracts it from the observed spectrum to estimate the Raman spectrum [15]. At each iteration, i, AIRPLS updates the estimated fluorescence $\mathbf{f}^{[i]}$ by minimizing $\min_{\mathbf{f}} \sum_{n=1}^{N} w_n^{[i]} (y_n - f_n)^2 + \lambda \sum_{n=2}^{N} (f_n - f_{n-1})^2$ where the first term accounts for the fidelity to the observed spectrum while the second term accounts for the smoothness of the estimated spectrum, and a large λ increases the smoothness. Starting with $\mathbf{w}^{[0]} = \mathbf{1}$, in the subsequent iterations, the weights $w_n^{[i]} = 0$ for $y_n > f_n^{[i]}$ are chosen to gradually ignore the peaks, and $w_n^{[i]} = \exp\left(i(y_n - f_n^{[i-1]})/\left|\mathbf{d}^{[i]}\right|\right)$ for $y_n < f_n^{[i-1]}$ are chosen to learn the baseline or fluorescence, where $d_n^{[i]} = (y_n - f_n^{[i]})_-$. The estimated fluorescence $\mathbf{f}$ is then subtracted from $\mathbf{y}$ to obtain the Raman spectrum $\mathbf{r}$. AIRPLS is able to adequately eliminate fluorescence to recover Raman spectra with large peaks, i.e., high signal-to-noise ratio. However, it is unable to differentiate between noise and the Raman spectrum under low signal-to-noise ratio, and the estimated Raman spectrum is usually noisy due to not modeling the Raman spectrum explicitly but only implicitly through fluorescence subtraction.

B. Two spectra

Shifted-excitation Raman difference spectroscopy (SERDS) uses two emission spectra $\mathbf{y}_1$ and $\mathbf{y}_2$ and estimates the Raman spectrum by first constructing the difference spectrum $\mathbf{e} = \mathbf{y}_1 - \mathbf{y}_2 = \mathbf{r} - \mathbf{L}^{(\Delta n_2)}\mathbf{r}$ and then constructing the Raman spectrum recursively as $r_{i+\Delta n_2} = r_i + e_{i+\Delta n_2}$ with $r_1 = \ldots = r_{i+\Delta n_2 - 1} = 0$. In the presence of noise and differences in relative intensities due to photobleaching, however, further corrections are required. For example, Gebrekidan *et al.* [18] propose an algorithm with the following steps to address these situations: first, perform z-score normalization to ensure that the intensities of the two spectra are equivalent before subtracting them, second, zero-center the difference spectrum using asymmetric least squares (ALS) to minimize the effect of left-over background before estimating the Raman spectrum from the difference spectrum, and third, correct baseline of the reconstructed Raman spectrum further using piecewise ALS. We refer to this algorithm as RSERDS (Reconstruction of SERDS). Subtracting two noisy signals, however, increases the noise in the estimated difference spectrum. We observe that the estimated Raman peaks are usually broader using RSERDS, and the estimated spectrum may contain artefacts potentially due to its heuristic parameter selection.

C. Multiple spectra

A commonly used method for obtaining the Raman difference spectrum from SER is principal component analysis (PCA). Here the intuition is that the direction of maximum variance, i.e., the largest eigenvector of the covariance matrix $\mathbf{C} = \sum_{k=1}^{K}(\mathbf{y}_k - \bar{\mathbf{y}})(\mathbf{y}_k - \bar{\mathbf{y}})^\top$ where $\bar{\mathbf{y}}$ is the average spectrum, corresponds to the difference spectrum, and the peak location can be determined through zero-crossings of this spectrum [25, p. 13687]. PCA is fast and does not require any parameter selection or the shifts, i.e., Δn_ks to be known. However, it is difficult to reconstruct the Raman signal from the difference spectrum [21]. Although this is applied in biological applications [25]–[27], we observed that it is often difficult to determine the peak locations from the estimated spectrum. We refer to this approach as PCA. A more principled approach for estimating the Raman spectrum from SER is treating it as a latent variable and using probabilistic modeling to infer this latent spectrum from observed spectra as an inverse problem.

1) Poisson inverse problem: McCain *et al.* [20] observed that Eq. (3) can be expressed as

$$\mathrm{vec}(\mathbf{Y}) \sim \mathrm{Poisson}(\mathbf{Hh}) \text{ with } \mathbf{H} = \begin{bmatrix} \mathbf{I}_N & \mathbf{L}^{(\Delta n_1)} \\ \vdots & \vdots \\ \mathbf{I}_N & \mathbf{L}^{(\Delta n_K)} \end{bmatrix}$$

under the assumption of Poisson noise where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_K] \in \mathbb{R}^{N\times K}$ is the pooled data matrix, $\mathbf{h} = [\mathbf{f}^\top, \mathbf{r}^\top]^\top \in \mathbb{R}^{2N}$, and $\mathbf{H} \in \mathbb{R}^{KN\times 2N}$ is a known operator. The authors proposed an expectation-maximization (EM) algorithm to estimate $\mathbf{h}$ recursively using the Lucy-Richardson formula as $\mathbf{h}^{[i+1]} = \mathbf{h}^{[i]} \circ \left(\mathbf{H}^\top(\mathrm{vec}(\mathbf{Y}) \oslash (\mathbf{H}\mathbf{h}^{[i]}))\right)/K$ where i denotes iteration and $\circ$ and $\oslash$ represent element-wise multiplication and division respectively. We refer to this approach as PIP. The authors found this approach to be more robust for noisy observations when compared to other methods including polynomial fitting and RSERDS [20, p. 10990]. Cooper *et al.* [21] addressed the high-computational cost of multiplying the $\mathbf{H}$ operator, and showed that this can be achieved more efficiently due to the specific structure of the matrix. Although PIP works well in practice, it does not take into account the relative variation in intensities due to photobleaching and laser output power variations explicitly, and we observe that the estimated Raman spectrum may not be sparse.

2) Shifted matrix factorization: Mørup *et al.* [23], [24] addressed a generic problem of separating an arbitrary number (M) of shifted latent spectra with unknown shifts and weights,

$$\mathbf{y}_k = \sum_{m=1}^{M} a_{km}\mathbf{L}^{(\Delta n_{km})}\mathbf{r}_m. \tag{5}$$

Mørup *et al.* proposed shifted nonnegative matrix factorization (SNMF) to solve Eq. (5) under the normal noise model and the assumption that $\mathbf{r}_m > \mathbf{0}$ [23] while Mørup *et al.* proposed shifted independent component analysis SICA to solve Eq. (5) under normal noise model and the assumption that $\mathbf{r}_m$s are independent [24]. The flexibility of these methods lie in estimating unknown shifts which can also be non-integer but this comes with additional computational complexity, and although both methods have shown to correctly identify latent spectra on synthetic data with unknown shifts, they have not been applied on real datasets. This approach can be adapted to fluorescence suppression by letting $M = 2$ with $\mathbf{r}_1 = \mathbf{r}$ and $\mathbf{r}_2 = \mathbf{f}$, and by letting $\Delta n_{2\cdot} = 0$ to ensure fluorescence does not vary over shifted excitations while $\Delta n_{1\cdot}$s are known. To differentiate the generic formulation from this specific one, we refer to the latter as SNMF* and SICA* respectively. We observe that although SNMF* can recover latent spectra adequately the resulting Raman spectrum is not sparse, and SICA* often fails to adequately recover the Raman spectrum since it can take negative values even though independence between the two spectra is a desired assumption.

III. Proposed Approach

While a number of principled approaches exist to suppress fluorescence, they either do not exactly match our requirements or they do not utilize the spectral characteristics of the spectra. For example: 1. AIRPLS and RSERDS do not use multiple spectra, 2. PCA only finds the Raman difference spectrum, 3. PIP assumes that the relative intensities of fluorescence and Raman do not change over shifted excitations, 4. while SICA* models the relative intensities, it does not assume non-negativity of the spectrum, 5. while SNMF* does both these, it may not produce a sparse Raman spectrum, and 6. none of the methods use the difference in the spectral characteristics of the two spectra, i.e., while Raman spectrum comprises sharp peaks, often intermittent but possibly overlapping, over a zero baseline, the fluorescence spectrum varies smoothly, and it is usually present over the entire region of interest.

To address these, we develop a method to 1. estimate the spectra from more than 2 measurements, i.e., $K \geq 2$, 2. estimate the Raman spectrum, i.e., $\mathbf{r}$, explicitly rather than the difference spectrum 3. allow relative intensities of the fluorescence, i.e., β, and Raman, i.e., α, to vary to accommodate the effect of photobleaching and variations in the laser output power, and include a bias term, i.e., b, to account for additive noise from the sensor (e.g. 'dark current') 4. use regularization for both Raman and fluorescence spectra, i.e., λ_r and λ_f respectively, to encode their spectral characteristics, i.e., Raman spectrum is 'sparse' and fluorescence spectrum is 'smooth', and finally, 5. use shifts reliably estimated from data due to the presence of characteristic oxygen and nitrogen peaks from NCF.

A common feature of subsequent Raman measurements is *photobleaching* where fluorescence intensity reduces over time due to sustained laser exposure [28]. Additionally, the spectrometer can produce a small output signal even in the absence of incident light, also known as *dark current*. To accommodate these features, we extend Eq. (3) as,

$$\mathbf{y}_k = \alpha_k\mathbf{L}^{(\Delta n_k)}\mathbf{r} + \beta_k\mathbf{f} + b_k \tag{6}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ denote the intensities of the Raman and fluorescence spectrum respectively, and $\mathbf{b}$ denotes the bias term.

The sparsity of the Raman spectrum can be encoded in several ways, and a typical solution is to minimize the ℓ_1

norm of the spectrum. This approach does not capture the complete spectral characteristics of the Raman spectrum, e.g., the local 'smoothness' in regions where the peaks appear, but we observe that the ℓ_1 regularization distinguishes the Raman spectrum from fluorescence well.

The smoothness of the fluorescence spectrum can also be modeled in several ways, and a typical solution is to use Tikhonov regularization as in AIRPLS. This can be described as $\|\mathbf{Df}\| = \mathbf{f}^\top\mathbf{D}^\top\mathbf{Df}$, where $\mathbf{D} = \mathbf{I} - \mathbf{L}^{(1)}$ is the difference matrix of size $N \times N$. However, we observe that this regularization might not sufficiently suppress the fluorescence spectrum (see supp. I). We suggest using a more generalized regularization of the form $\mathbf{f}^\top\mathbf{K}^{-1}\mathbf{f}$ where $\mathbf{K}$ is a suitable covariance matrix, e.g., $k_{nn'} = \exp(-(n-n')^2/\ell^2)$ where ℓ controls the smoothness of the covariance function. The parameters of the covariance function can be optimized as part of the optimization. However, for simplicity, we set it to $l = N/4$. We also use a small regularization parameter $\epsilon = 10^{-6}$ to avoid instability of the matrix during inversion, i.e., the resulting regularization term is $\mathbf{f}^\top(\mathbf{K}+\epsilon\mathbf{I})^{-1}\mathbf{f}$.

We assume a normal noise model. We can use more flexible distributions such as Poisson or negative binomial. However, choosing a standard noise model simplifies the resulting optimization problem. Thus, we minimize the following loss function to find $(\mathbf{r}, \mathbf{f}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{b})$,

$$c = \sum_{k=1}^{K} \left\| \mathbf{y}_k - \left(\alpha_k \mathbf{L}^{(\Delta n_k)}\mathbf{r} + \beta_k \mathbf{f} + b_k\right)\right\|^2 + \lambda_f \mathbf{f}^\top\tilde{\mathbf{K}}\mathbf{f} + \lambda_r\|\mathbf{r}\|_1 \tag{C}$$

where λ_f and λ_r are regularization parameters for the fluorescence and Raman spectrum respectively, and $\tilde{\mathbf{K}} = (\mathbf{K}+\epsilon\mathbf{I})^{-1}$ can be pre-computed. $\|\cdot\|$ and $\|\cdot\|_1$ denote ℓ_2 and ℓ_1 norm.

We use Block Coordinate Descent (BCD) to minimize the loss [29], i.e., we split the problem into $K+2$ blocks, such that at each iteration, the loss function for a specific block is minimized while fixing the values of the other blocks, and we repeat this over each block until convergence. The $K+2$ blocks are the Raman spectrum $\mathbf{r}$, the fluorescence spectrum, $\mathbf{f}$, and K triplets (α_k, β_k, b_k) respectively. We update these blocks using either non-negative least squares (NNLS) or non-negative least absolute shrinkage and selection operator (NNLASSO). NNLS solves $\min_{\mathbf{x}\geq\mathbf{0}}\|\mathbf{d}-\mathbf{Cx}\|^2$ [30] and NNLASSO solves $\min_{\mathbf{x}\geq\mathbf{0}}\|\mathbf{d}-\mathbf{Cx}\|^2 + \lambda\|\mathbf{x}\|_1$ [31]. We express the quadratic cost as $\mathbf{x}^\top\mathbf{Ax} - 2\mathbf{b}^\top\mathbf{x}$, however, the two forms are equivalent with $\mathbf{C} = \textsc{Chol}(\mathbf{A})$ and $\mathbf{d} = (\mathbf{C})^{\top^{-1}}\mathbf{b}$ where CHOL denotes Cholesky decomposition. We refer to this algorithm as *multi-spectral estimation of regularized spectra* or MSERS (see Algorithm 1 and supp. II for concise and detailed pseudocode).

a) Intensities and bias update: The (α_k, β_k, b_k) values for each $k = 1, \ldots, K$ shifts are updated by solving the following optimization,

$$\min_{\alpha_k,\beta_k,b_k\geq 0}\left\|\mathbf{y}_k - (\alpha_k\mathbf{L}^{(\Delta n_k)}\mathbf{r} + \beta_k\mathbf{f} + b_k)\right\|^2. \tag{7}$$

This is a NNLS optimization problem, i.e.,

$$\min_{\mathbf{x}\geq\mathbf{0}}\|\mathbf{y}_k - \mathbf{Cx}\|^2 \tag{P1}$$

Algorithm 1 MSERS

Input: $\mathbf{Y}$, $\{\Delta n_k\}_{k=2}^K$, $\tilde{\mathbf{K}}$, λ_r, λ_f, $i_{\max}$, tol
Require: $\mathbf{Y} \geq 0$
  Initialize Raman $\mathbf{r}^{[0]}$ and fluorescence $\mathbf{f}^{[0]}$ spectra
  $i = 0$, $c^{[-1]} = 0$, $c^{[0]} = 0$
  while $i \leq i_{\max}$ & $|c^{[i]} - c^{[i-1]}| \geq \text{tol} \times c^{[i-1]}$ **do**
    $i = i + 1$
    for $k \in \{1, \ldots, K\}$ **do**
      update $(\alpha_k, \beta_k, b_k)^{[i]}$ using Eq. (P1)
    end for
    update $\mathbf{f}^{[i]}$ using Eq. (P2)
    update $\mathbf{r}^{[i]}$ using Eq. (P3)
    update $c^{[i]}$ using Eq. (C)
  end while
  return $(\mathbf{r}, \mathbf{f}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{b})$

with $\mathbf{C} = [\mathbf{L}^{(\Delta n_k)}\mathbf{r}, \mathbf{f}, \mathbf{1}]$ and $\mathbf{x} = [\alpha_k, \beta_k, b_k]^\top$. $\mathbf{L}^{(\Delta n_k)}\mathbf{r}$ does not need to be computed explicitly through matrix multiplication but the vector $\mathbf{r}$ can be shifted (supp. III).

b) Fluorescence update: The fluorescence $\mathbf{f}$ is updated by solving the following optimization,

$$\min_{\mathbf{f}\geq\mathbf{0}}\left\|\mathbf{y}_k - (\alpha_k\mathbf{L}^{(\Delta n_k)}\mathbf{r} + \beta_k\mathbf{f} + b_k)\right\|^2 + \lambda_f\mathbf{f}^\top\tilde{\mathbf{K}}\mathbf{f} \tag{8}$$

This can again be expressed as a NNLS optimization as

$$\min_{\mathbf{f}\geq\mathbf{0}}\sum_{k=1}^{K}\left\|\mathbf{y}_k^f - \beta_k\mathbf{f}\right\|^2 + \lambda_f\mathbf{f}^\top\tilde{\mathbf{K}}\mathbf{f} \equiv \min_{\mathbf{f}\geq\mathbf{0}}\mathbf{f}^\top\left[\left(\sum_{k=1}^{K}\beta_k^2\right)\mathbf{I} + \lambda_f\tilde{\mathbf{K}}\right]\mathbf{f} - 2\left[\sum_{k=1}^{K}\beta_k\mathbf{y}_k^f\right]^\top\mathbf{f} \tag{P2}$$

with $\mathbf{y}_k^f = \mathbf{y}_k - \alpha_k\mathbf{L}^{(\Delta n_k)}\mathbf{r} - b_k$. $\mathbf{L}^{(\Delta n_k)}\mathbf{r}$ can be evaluated efficiently by shifting $\mathbf{r}$.

c) Raman update: The Raman $\mathbf{r}$ is updated by solving the following optimization

$$\min_{\mathbf{r}\geq\mathbf{0}}\sum_{k=1}^{K}\left\|\mathbf{y}_k - (\alpha_k\mathbf{L}^{(\Delta n_k)}\mathbf{r} + \beta_k\mathbf{f} + b_k)\right\|^2 + \lambda_r\|\mathbf{r}\|_1 \tag{9}$$

This can be expressed as a NNLASSO optimization as

$$\min_{\mathbf{r}\geq\mathbf{0}}\sum_{k=1}^{K}\left\|\mathbf{y}_k^r - \alpha_k\mathbf{L}^{(\Delta n_k)}\mathbf{r}\right\|^2 + \lambda_r\|\mathbf{r}\|_1 \equiv \min_{\mathbf{r}\geq\mathbf{0}}\mathbf{r}^\top\left[\left(\sum_{k=1}^{K}\alpha_k^2\right)\mathbf{I}\right]\mathbf{r} - 2\left[\sum_{k=1}^{K}\alpha_k\mathbf{y}_k^{r\top}\mathbf{L}^{(\Delta n_k)}\right]\mathbf{r} + \lambda_r\|\mathbf{r}\|_1 \tag{P3}$$

where $\mathbf{y}_k^r = \mathbf{y}_k - \beta_k\mathbf{f} - b_k$. $\mathbf{y}_k^{r\top}\mathbf{L}^{(\Delta n_k)}$ can be evaluated efficiently by shifting $\mathbf{y}_k^r$.

d) Fiber Raman background: Although hollow-core NCFs reduce the Raman background from the optical fibers significantly, it is nonetheless present in the observed spectra (see CYCLOHEXANE). Raman background shifts with shifts in excitation similar to the Raman spectrum of interest, and therefore it cannot be explicitly modeled as either $\mathbf{f}$ or $\mathbf{r}$ in MSERS. However, we observe that MSERS can still remove this

background effectively to reveal a Raman spectrum with zero baseline since although this background (referred as $\mathbf{g}$) shifts with Δn_k, its relative smoothness allows it to be estimated as fluorescence, i.e., $\mathbf{f}+\mathbf{L}^{(\Delta n_k)}(\mathbf{g}+\mathbf{r}) \approx (\mathbf{f}+\mathbf{g})+\mathbf{L}^{(\Delta n_k)}\mathbf{r}$ over sufficiently small Δn_k, and $\mathbf{f}+\mathbf{g}$ is smooth.

e) Initialization and convergence: Following Marshall and Cooper [22], we initialize the Raman spectrum with the standard deviation across the raw spectra for each wavenumber whilst the fluorescence spectrum is initialized as the minimum value across the raw spectra for each wavenumber (see supp. IV for other initializations). The algorithm is said to converge if the relative change in loss between subsequent iterations is at most $\text{tol} = 10^{-3}$ and a maximum number of iterations $i_{\max} = 100$ is set to manage the run-time.

f) Parameter selection: We express the regularization parameters as $\lambda_{(\cdot)} = KN10^{-\lambda^e_{(\cdot)}}$ and select λ^e_f and λ^e_r using internal validation. We run the algorithm for several parameter values, i.e., $\lambda^e_r \in \{11, 10, 9, 8\}$ and $\lambda^e_f \in \{12, 11, 10, 9, 8, 7\}$ and choose the solution with adequate sparsity (see IV-C.4) and distinguishability. The distinguishability is calculated as the inverse correlation (see IV-C.3) between the Raman and fluorescence spectra. We choose the Raman spectrum with moderate sparsity (between 0.3 and 0.7) that is most distinguishable from the fluorescence spectrum (see supp. V).

g) Estimating shifts: Although the shift in excitation can be estimated explicitly through a second spectrometer, we instead choose to estimate it through the emission spectra. The air in the hollow-core NCF results in two distinct Raman peaks at 1555 cm^{-1} and 2331 cm^{-1} (in Raman shift), (see Fig. 4) due to the presence of oxygen and nitrogen in the air. These characteristic peaks can be used to compute the excitation wavelength and the shift in excitation. Although both peaks provide similar estimation (see supp. VI), we used the oxygen peak as after converting to wavenumber, the resolution of the observed spectra around the oxygen peak is higher than around the nitrogen peak. This inherent calibration step simplifies the experimental set-up by eliminating the need to introduce a second spectrometer and does not add further uncertainties such as calibration and synchronization.

IV. Experiments

A. Data Collection

The Raman spectra were taken using a custom-built optical set-up which can be seen in Fig. 1. The system is designed for shifted excitation wavelengths around 785 nm, utilizing the lower autofluorescence presence in the near-infrared window in biological tissue, and efficient coupling and collection into the bespoke Raman fiber. A 785 nm laser diode (DBR785S, Thorlabs; line width < 0.1 nm) was coupled with a thermo-electric cooler (TEC) control. A change in temperature tunes the wavelength of the laser diode to match different laser cavity modes and hence allows access to a range of excitation wavelengths. The average laser output power at the distal end of the fiber is 20 mW but this varies slightly between wavelengths. This results in different measured intensities over shifts. The hollow-core NCF is used to transport light to the sample whilst the surrounding multimode fibers collect the Raman scattering from the sample. The collected Raman signal is coupled via the dichroic mirror into a multimode patch cable and fiber-fed to the spectrometer (QePro Raman, Ocean Insight) which records the Raman spectra via the commercial software OceanView (Ocean Insight) within the spectral range of 840 nm to 992 nm (834 cm^{-1} to 2658 cm^{-1} in Raman shift for excitation 785 nm). We investigated two reference samples, CYCLOHEXANE and SESAME OIL, and *ex vivo* LUNG TISSUE from human lung. For liquid compounds, the distal end of the optical fiber was immersed in the compound but for tissue it had direct contact with the surface.

1) Cyclohexane: Cyclohexane (1.02822, cyclohexane for spectroscopy Uvasol®, Supelco, Merck KGaA) is a chemical compound popular in Raman spectrometer calibration due to its well known Raman spectrum and the fact that it has no fluorescence background. An example observed spectrum of cyclohexane is shown in Fig. 4 (left) (the background observed in the figure is the Raman background from the fiber). It is characterized by three large peaks of similar intensities occurring around 1029 cm^{-1}, 1267 cm^{-1} and 1445 cm^{-1}. Additionally there is a less intense peak at 1158 cm^{-1} and a weaker, broader peak at 1347 cm^{-1} [32, Table I].

2) Sesame oil: The Raman spectrum of sesame oil (Toasted Sesame Oil 250mL, Tesco) has large, well known peaks but also a high level of fluorescence background. An example observed spectrum of sesame oil is shown in Fig. 4 (middle). The Raman spectrum is characterized by two large peaks at 1441 cm^{-1} and 1657 cm^{-1} with smaller peaks at 1267 cm^{-1} and 1304 cm^{-1} creating a double peak and further small peaks at 1083 cm^{-1} and 1747 cm^{-1} [33, Table 1].

3) Lung tissue: An *ex vivo* human lung tissue sample was obtained from a patient who was recruited from Royal Infirmary of Edinburgh (NHS Lothian BioResource, reference 15/ES/0094), diagnosed with suspected or confirmed lung cancer, undergoing thoracic resection surgery. Peripheral lung tissue (>5 cm away from tumor margin) was obtained from the resection sample. An example observed spectrum of tissue is shown in Fig. 4 (right). Raman peaks from tissue are weak and complex to interpret as the peaks can overlap. They are almost completely masked by the strong autofluorescence of tissue. Following Huang *et al.* [34, Fig. 5], we consider the following eight peaks in our study at 1078 cm^{-1}, 1265 cm^{-1}, 1302 cm^{-1}, 1445 cm^{-1}, 1602 cm^{-1}, 1618 cm^{-1}, 1665 cm^{-1}, and 1745 cm^{-1} where the first four and the last one are distinct sharp peaks while the rest are overlapping peaks. We do not consider peaks with intensity below that of the 1745 cm^{-1} peak as they may be noisy and thus hard to detect, and we do not consider peaks 1518 cm^{-1}, 1552 cm^{-1} and 1582 cm^{-1} due to their proximity to the oxygen peak. While there is an additional sharp peak at 855 cm^{-1}, this peak is only partially observed since the respective wavelength (between 839.79 nm and 841.96 nm for different excitations) is close to our lower detection threshold at 840 nm to 992 nm, and therefore, we exclude this peak from the analysis.

B. Data Processing

Each measurement, for a shift k, consists of 1044 data points, spanning a spectral range of 840 nm to 992 nm. The

spectra were taken on a wavelength axis with unequal resolution, decreasing from 0.167 nm at lower wavelengths to 0.126 nm at longer wavelengths (see supp. VII). We first convert the wavelengths to wavenumbers $\tilde{\nu}$. The multi-spectra algorithms feature a 'shift', i.e., $\Delta\tilde{\nu}_k$, and applying a shift to the uneven wavelengths will result in misalignment. Therefore, we use linear interpolation to project the intensity values onto an equispaced grid. The resolution of this grid is equal to the highest resolution (smallest spacing) of the original uneven grid. This changes the length of each spectrum to $N = 1418$. The spectra were converted to Raman shifts $\delta\tilde{\nu} = \tilde{\nu} - \tilde{\nu}_0$ for visualization and comparison with known characteristic peak locations where $\tilde{\nu}_0$ is the first excitation wavelength corresponding to $\Delta\tilde{\nu}_1 = 0$.

For each sample, 14 temperature steps were set on the TEC control. Due to modelocking conditions, not all steps resulted in an excitation wavelength shift, and due to environmental conditions, such as laboratory temperature changes, the exact excitation wavelength positions were difficult to repeat. We considered the first 10 spectra with lower TEC setting as the higher TEC settings (with lower excitation wavelength) do not necessarily enter a new modelock, i.e., $K = 10$. For each k, 10 repeated measurements were taken to avoid saturation of the sensors with exposure time t_{exp}. We summed these repeated measurements, i.e., the integration time is $t_{\text{int}} = 10t_{\text{exp}}$ and the acquisition time is $t_{\text{acq}} = Kt_{\text{int}}$. For CYCLOHEXANE and SESAME OIL, $t_{\text{acq}} = 10\,\text{s}$, and for LUNG TISSUE, $t_{\text{acq}} = 50\,\text{s}$.

We implement the algorithms in Python3. For MSERS we use existing solvers for NNLS and NNLASSO, and existing peak detection tools for estimating shifts. We set any small negative values (i.e., $y_{kn} < 0$) to 0 and normalize the pooled spectra $\mathbf{Y}$ to be between 0 and 1. We apply the same initialization as MSERS to PIP, SNMF* and SICA*, and randomly initialize a_{km}s in SNMF* and SICA*. PCA, SNMF* and SICA* do not require any parameter selection. For AIRPLS, the smoothing parameter, λ, is incrementally increased from 1 and the resulting spectra are plotted until the fluorescence does not feature any Raman peaks. RSERDS involves tuning 3 parameters such that the difference spectrum is sufficiently smooth and the Raman spectrum is smooth with a zero baseline. This is done by plotting the results with small values and increasing them until the results are satisfactory. For PIP, the maximum number of iterations is chosen to be 2000, as is used by Cooper *et al.* [21, p. 979]. The peaks of the spectrum reconstructed with RSERDS and the zero-crossings of the difference spectrum of PCA may not align with the first observed spectrum as is the case with the other algorithms. Therefore, these spectra have been shifted by half of the maximum shift, to be in line with the other spectra for comparison.

C. Evaluation metrics

1) Peak evaluation: We compare the peaks detected from the estimated Raman spectrum with their respective true or *suggested* locations (see section IV-A) in terms of precision, i.e., number of true peaks detected over the total number of peaks detected, and recall, i.e., number of true peaks detected over the total number of true peaks. We count a true positive if the true peak location falls within the peak width of the detected peak. Detected peaks are those peaks whose height from the baseline is at least 5% of that of the oxygen peak for CYCLOHEXANE and SESAME OIL, and nitrogen peak for LUNG TISSUE, and the peak width is less than 200.

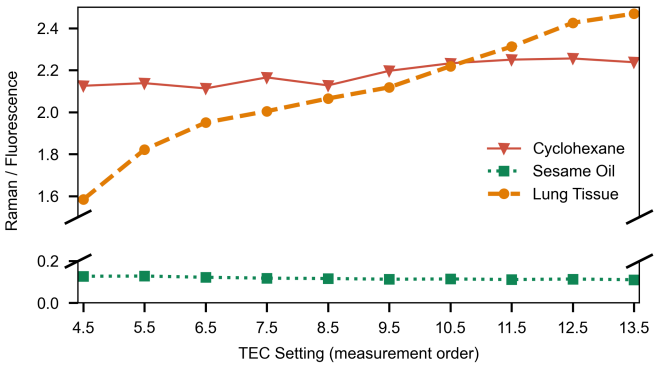

Fig. 2: The figure shows the relative intensity of Raman spectrum with respect to fluorescence spectrum intensity for MSERS over progressive measurements. See section IV-D.1.

2) Signal-to-noise ratio: We quantify SNR in terms of the ratio of the peak intensity (oxygen peak for CYCLOHEXANE and SESAME OIL and nitrogen peak for LUNG TISSUE, both from the baseline) and the standard deviation of a Raman free, fluorescence only area. This region is taken as the spectrum from 889 cm^{-1} to 942 cm^{-1} (in Raman shift) for CYCLOHEXANE and 1150 cm^{-1} to 1200 cm^{-1} for SESAME OIL and LUNG TISSUE.

3) Correlation: The fluorescence and Raman spectra should be independent of each other since they follow different generative mechanisms. Therefore, if the two spectra have been separated adequately then we should expect a small correlation between them. We quantify this using Pearson's correlation coefficient.

4) Sparsity: We expect the Raman spectra to be moderately sparse, i.e., if the fluorescence has been suppressed adequately then the resulting Raman spectrum should have intermittent sharp peaks. We quantify this as the proportion of wavenumbers with intensity value less than 0.1% of the maximum intensity value.

5) Run-time: We report the run-time of the algorithm, measured in seconds. We define run-time as the time it takes the algorithms to run once for a given parameter setting. However, this does not include the time for the user to adjust parameters which would impact the total implementation time.

D. Results

1) Effect of photobleaching: MSERS explicitly captures the effect of photobleaching where the relative intensity of the Raman spectrum compared to the fluorescence background vary over progressive measurements. Fig. 2 shows the ratio of Raman and fluorescence intensities, i.e., $(\alpha_k \max_n r_n)/(\beta_k \max_n f_n)$, in the order the measurements were taken for each dataset. We observe an upward trend for LUNG TISSUE indicating photobleaching while CYCLOHEXANE and SESAME OIL do not vary. This is expected

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2023.3243866

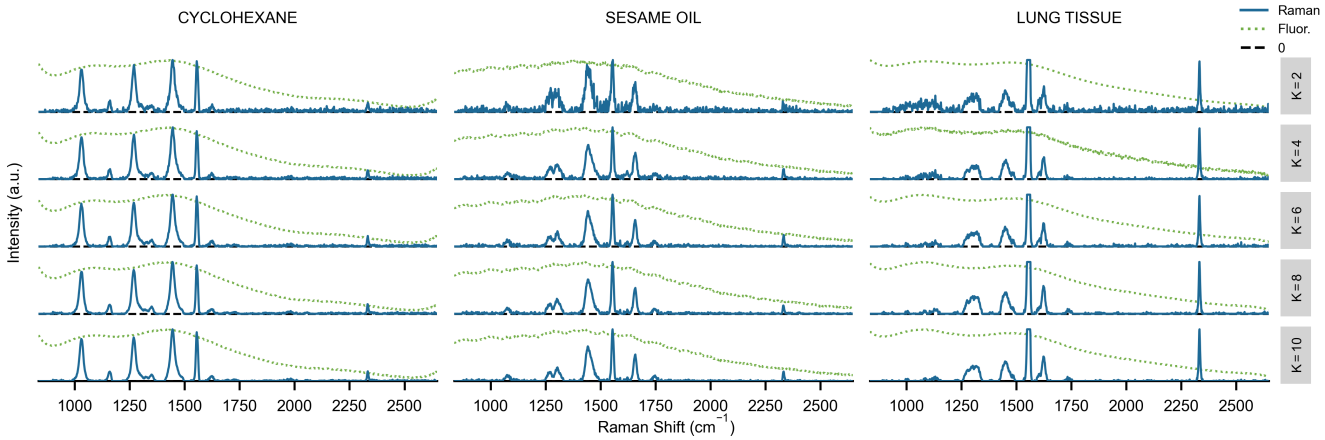

Fig. 3: The figure shows the inferred Raman and fluorescence spectra from MSERS for different number of excitations K. The inferred spectra progressively improve with increasing K. See section IV-D.2. a.u. is arbitrary unit.

TABLE II: The table summarizes the characteristics of spectra inferred using MSERS for different number of excitations. ↓ (↑) implies that lower (higher) value is better. See section IV-D.2. Run-time is shown in seconds. # P. is number of detected peaks.

	CYCLOHEXANE							SESAME OIL							LUNG TISSUE						
K	SNR$^{\uparrow}$	Corr.$^{\downarrow}$	Spar.	# P.	Prec.$^{\uparrow}$	Rec.$^{\uparrow}$	Time$^{\downarrow}$ [Itr]	SNR$^{\uparrow}$	Corr.$^{\downarrow}$	Spar.	# P.	Prec.$^{\uparrow}$	Rec.$^{\uparrow}$	Time$^{\downarrow}$ [Itr]	SNR$^{\uparrow}$	Corr.$^{\downarrow}$	Spar.	# P.	Prec.$^{\uparrow}$	Rec.$^{\uparrow}$	Time$^{\downarrow}$ [Itr]
2	72.1	0.291	0.301	57	0.088	1.0	378 [100]	34.0	0.388	0.371	113	0.053	1.0	361 [100]	23.6	0.344	0.353	128	0.055	0.875	37 [8]
4	80.3	0.312	0.309	15	0.333	1.0	51 [12]	74.9	0.480	0.321	31	0.194	1.0	233 [63]	55.4	0.354	0.646	23	0.217	0.625	39 [9]
6	107.9	0.306	0.341	7	0.714	1.0	46 [11]	103.8	0.467	0.408	18	0.333	1.0	250 [69]	88.2	0.237	0.587	22	0.273	0.750	109 [29]
8	99.8	0.319	0.348	7	0.714	1.0	48 [11]	92.7	0.448	0.453	8	0.750	1.0	193 [50]	149.7	0.257	0.662	9	0.556	0.625	63 [16]
10	120.9	0.338	0.382	6	0.833	1.0	47 [11]	135.9	0.452	0.497	8	0.750	1.0	269 [73]	198.9	0.264	0.677	8	0.625	0.625	92 [23]

for CYCLOHEXANE since it does not have any fluorescence. SESAME OIL shows the lowest value indicating relatively high presence of fluorescence compared to LUNG TISSUE.

2) Changing number of excitations: Although a higher K is expected to infer better spectra, a lower K may be preferred for *in vivo* applications to reduce data collection time and motion artefacts, e.g., due to breathing. We assess if fewer measurements can provide adequate accuracy. We compare the performance of MSERS when changing the number of excitations, K. We use $K \in \{2, 4, 6, 8, 10\}$ where the spectra are chosen to maximize the separation of the excitations (see supp. VIII). Table II summarizes the key metrics for all K, and Fig. 3 shows the estimated spectra. We observe that, both qualitatively and quantitatively, the estimated spectra progressively improve in signal-to-noise ratio as well as in precision and recall for increasing K, however, 1) even $K = 2$ provides effective fluorescence suppression, e.g., in CYCLOHEXANE, and 2) $K = 8$ and $K = 10$ provide similar performance in precision and recall for all datasets.

3) Comparison to existing methods: We use $K = 10$ for comparison with other methods. Fig. 4 shows the inferred spectra for different algorithms, and Table III summarizes their characteristics. We observe that qualitatively MSERS provides a 'peaky'-er Raman spectrum over zero baseline and a smoother background spectrum. This is supported quantitatively by a relatively higher sparsity and SNR compared to other approaches and a relatively more accurate location of the peaks as well as narrow peak widths (see supp. IX)). We observe that 1) AIRPLS provides a relatively noisy Raman spectrum in terms of SNR while RSERDS provides a broader Raman spectrum (see e.g., peaks in LUNG TISSUE), and both methods result in low precision and recall. 2) PCA works well for CYCLOHEXANE but performs poorly on the other datasets. 3) SICA* provides uncorrelated spectra that can take negative values (see e.g., LUNG TISSUE). The rest of the methods work well in terms of precision and recall, however, 4) MSERS provides better SNR and sparsity than PIP and SNMF* (see e.g., SESAME OIL), and Raman spectrum that is less correlated with the background. (see e.g., SESAME OIL).

V. Conclusions

SER is an emerging technology with promising applications in intra-operative real-time tumor delineation with the potential to improve surgical resection accuracy and patients outcome in the long term. Biomedical *in vivo* applications of SER suffers from the presence of tissue background fluorescence and background from Raman fiber that masks the weak Raman peaks of interest. Existing computational tools for suppressing fluorescence are inadequate for such applications due to the low signal-to-noise ratio and photobleaching. In this paper, we proposed a computational framework MSERS that is more suitable for such applications. We demonstrated that MSERS suppresses fluorescence and recovers Raman spectra more effectively than existing approaches, both qualitatively and quantitatively, by capturing the effect of photobleaching, modeling the fluorescence as a smooth spectrum, and modeling the Raman spectrum to be sparse. The computational efficacy of MSERS along with recent technological development in sources such as laser diodes could pave the way for clinical deployment of this technology in the near future.

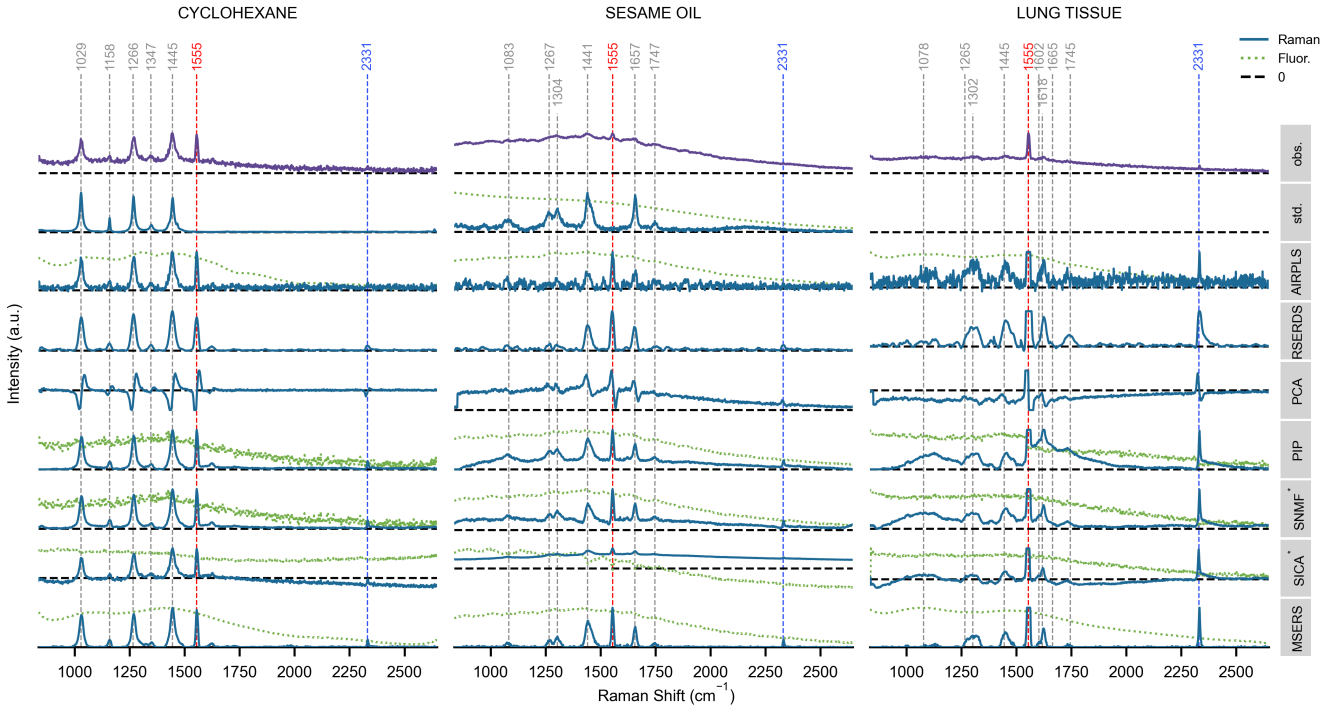

Fig. 4: The figures shows the observed, standard and inferred spectra for different fluorescence suppression methods. One observed spectrum (obs.) ($k = 1$) is shown for better visibility. Tissue spectra have been scaled to diminish the oxygen peak to focus on tissue peaks. The known and suggested peak locations are shown with gray vertical dotted lines while the oxygen and nitrogen peaks are shown with red and blue vertical dotted lines respectively. Standard spectra (std.) (except LUNG TISSUE) are collected with an inVia™ confocal Raman microscope (Renishaw plc.); spectra are taken without a fiber probe element and Raman spectra extracted using Wire™ 4.4 software (Renishaw plc.). See section IV-D.3.

TABLE III: The table summarizes the performance of fluorescence suppression methods. ↓ (↑) implies that lower (higher) value is better. unk is unknown and i/d is indistinguishable. # P. is number of detected peaks. See section IV-D.3.

		Peak Location ± Width/2 (cm^{-1})								SNR↑	Corr.↓	Spar.	# P.	Prec.↑	Rec.↑	Time↓
CYCLOHEXANE	True	1029±unk	1158±unk	1267±unk	1347±unk	1445±unk										
	AIRPLS	1028±36	1159±33	1270±34	1343±21	1442±42				24.8	0.153	0.009	156	0.032	1.000	0.091
	RSERDS	1030±33	1157±53	1267±30	1345±28	1445±40				38.1		0.072	7	0.714	1.000	0.347
	PCA	1031±75	1159±28	1270±55	1350±26	1446±60						0.025	282	0.018	1.000	0.001
	PIP	1030±44	1159±18	1268±36	1349±20	1445±40				192.3	0.356	0.059	6	0.833	1.000	1.899
	SNMF	1030±36	1159±19	1268±40	1349±20	1445±40				152.3	0.556	0.010	6	0.833	1.000	15.693
	SICA	1029±32	1159±17	1268±29	1350±20	1445±43				38.3	0.300	0.006	45	0.111	1.000	42.964
	MSERS	1030±28	1159±17	1268±35	1350±11	1445±37				120.9	0.338	0.382	6	0.833	1.000	47.390
SESAME OIL	True	1083±unk	1267±unk	1304±unk	1441±unk	1657±unk	1747±unk									
	AIRPLS	1078±15	1270±15	1300±47	1435±24	1660±18	1750±3			12.6	0.239	0.011	183	0.033	1.000	0.012
	RSERDS	1076±45	1271±19	1302±29	1441±36	1655±34	1755±36			42.6		0.035	17	0.353	1.000	0.233
	PCA	i/d	i/d	i/d	i/d	i/d	i/d					0.000	0		0.000	0.003
	PIP	1075±79	1273±15	1304±58	1442±47	1659±34	1748±28			51.8	0.854	0.017	9	0.667	1.000	1.925
	SNMF	1075±54	1270±10	1302±72	1442±33	1659±12	1748±20			59.0	0.855	0.000	14	0.429	1.000	18.479
	SICA	1075±26	i/d	1304±41	1442±36	1659±17	1748±23			27.0	0.316	0.000	6	0.833	0.833	47.090
	MSERS	1075±22	1270±15	1304±42	1442±38	1659±19	1748±20			135.9	0.452	0.497	8	0.750	1.000	269.300
LUNG TISSUE	True	1078±unk	1265±unk	1302±unk	1445±unk	1602±unk	1618±unk	1665±unk	1745±unk							
	AIRPLS	1079±4	1263±2	1304±10	1447±38	1600±3	1618±3	1668±14	1750±8	9.7	0.150	0.020	242	0.033	1.000	0.014
	RSERDS	1069±26	i/d	1296±46	1451±39	i/d	1624±36	i/d	1741±52	49.1		0.094	21	0.238	0.625	0.271
	PCA	i/d	i/d	i/d	i/d	i/d	i/d	i/d	i/d			0.000	0		0.000	0.002
	PIP	i/d	i/d	1316±47	1455±36	i/d	1624±36	i/d	1733±28	33.0	0.255	0.045	10	0.400	0.500	1.899
	SNMF	i/d	i/d	1295±45	1454±39	i/d	1624±33	i/d	1733±36	38.8	0.520	0.006	5	0.800	0.500	16.440
	SICA	i/d	i/d	1291±45	1446±37	i/d	1620±34	i/d	1729±51	45.9	-0.048	0.017	7	0.571	0.500	48.495
	MSERS	1087±10	i/d	1302±44	1449±36	i/d	1624±25	i/d	1733±12	198.9	0.264	0.677	8	0.625	0.625	91.562

REFERENCES

[1] J. A. Kim *et al.*, "Optical spectroscopy for *in vivo* medical diagnosis—a review of the state of the art and future perspectives," *Prog. Biomed. Eng*, vol. 2, no. 4, p. 042 001, Aug. 2020.

[2] M. J. Baker *et al.*, "Clinical applications of infrared and raman spectroscopy: State of play and future challenges," *Analyst*, vol. 143, no. 8, pp. 1735–1757, 2018.

[3] R. X. Xu *et al.*, "Diffuse optical imaging and spectroscopy for cancer," *Expert Rev. Med. Devices*, vol. 4, no. 1, pp. 83–95, 2007.

[4] S. Yerolatsitis *et al.*, "Ultra-low background raman sensing using a negative-curvature fibre and no distal optics," *J. Biophotonics*, vol. 12, no. 3, e201800239, 2019.

[5] B. Brozek-Pluska *et al.*, "Analysis of human colon by raman spectroscopy and imaging-elucidation of biochemical changes in carcinogenesis," *Int. J. Mol. Sci.*, vol. 20, no. 14, p. 3398, 2019.

[6] H. Wang *et al.*, "Screening and staging for non-small cell lung cancer by serum laser raman spectroscopy," *Spectrochim. Acta A Mol. Biomol. Spectrosc.*, vol. 201, pp. 34–38, 2018.

[7] A. Mahadevan-Jansen *et al.*, "Near-infrared raman spectroscopy for in vitro detection of cervical precancers," *Photochem. Photobiol.*, vol. 68, no. 1, pp. 123–132, 1998.

[8] C. Kendall *et al.*, "Raman spectroscopy, a potential tool for the objective identification and classification of neoplasia in barrett's oesophagus," *J. Pathol.*, vol. 200, no. 5, pp. 602–609, 2003.

[9] H. C. McGregor *et al.*, "Real-time endoscopic Raman spectroscopy for *in vivo* early lung cancer detection," en, *J. Biophotonics*, vol. 10, no. 1, pp. 98–110, 2017.

[10] J. May *et al.*, "Fiber raman background study and its application in setting up optical fiber raman probes," *Appl. Opt.*, vol. 35, no. 15, pp. 2527–2533, 1996.

[11] D. Wei *et al.*, "Review of fluorescence suppression techniques in raman spectroscopy," *Appl. Spectrosc. Rev.*, vol. 50, no. 5, pp. 387–406, 2015.

[12] F. Knorr *et al.*, "Development of a time-gated system for raman spectroscopy of biological samples," *Opt. express*, vol. 18, no. 19, pp. 20 049–20 058, 2010.

[13] J. Kostamovaara *et al.*, "Fluorescence suppression in raman spectroscopy using a time-gated CMOS SPAD," *Opt. express*, vol. 21, no. 25, pp. 31 632–31 645, 2013.

[14] K. Ehrlich *et al.*, "pH sensing through a single optical fibre using SERS and CMOS SPAD line arrays," *Opt. Express*, vol. 25, no. 25, pp. 30 976–30 986, 2017.

[15] Z.-M. Zhang *et al.*, "Baseline correction using adaptive iteratively reweighted penalized least squares," *Analyst*, vol. 5, pp. 1138–1146, 2010.

[16] P. Cadusch *et al.*, "Improved methods for fluorescence background subtraction from raman spectra," *J. Raman Spectrosc.*, vol. 44, no. 11, pp. 1587–1595, 2013.

[17] M. Kasha, "Characterization of electronic transitions in complex molecules," *Discussions of the Faraday Society*, vol. 9, pp. 14–19, 1950.

[18] M. T. Gebrekidan *et al.*, "A shifted-excitation raman difference spectroscopy (SERDS) evaluation strategy for the efficient isolation of raman spectra from extreme fluorescence interference," *J. Raman Spectrosc.*, vol. 47, no. 2, pp. 198–209, 2016.

[19] M. Mazilu *et al.*, "Optimal algorithm for fluorescence suppression of modulated raman spectroscopy," *Opt. express*, vol. 18, no. 11, pp. 11 382–11 395, 2010.

[20] S. T. McCain *et al.*, "Multi-excitation raman spectroscopy technique for fluorescence rejection," *Opt. Express*, vol. 16, no. 15, pp. 10 975–10 991, 2008.

[21] J. B. Cooper *et al.*, "Sequentially shifted excitation raman spectroscopy: Novel algorithm and instrumentation for fluorescence-free raman spectroscopy in spectral space," *Appl. Spectrosc.*, vol. 67, no. 8, pp. 973–984, 2013.

[22] S. Marshall *et al.*, "Quantitative raman spectroscopy when the signal-to-noise is below the limit of quantitation due to fluorescence interference: Advantages of a moving window sequentially shifted excitation approach," *Appl. Spectrosc.*, vol. 70, no. 9, pp. 1489–1501, 2016.

[23] M. Mørup *et al.*, "Shifted non-negative matrix factorization," in *IEEE Mach. Learn. Signal Process.*, 2007, pp. 139–144.

[24] M. Mørup *et al.*, "Shifted independent component analysis," English, in *Indep. Component Anal. and Signal Sep.*, 2007, pp. 89–96.

[25] A. C. De Luca *et al.*, "Modulated raman spectroscopy for enhanced cancer diagnosis at the cellular level," *Sensors*, vol. 15, no. 6, pp. 13 680–13 704, 2015.

[26] L. Woolford *et al.*, "Towards automated cancer screening: Label-free classification of fixed cell samples using wavelength modulated raman spectroscopy," *J. Biophotonics*, vol. 11, no. 4, e201700244, 2018.

[27] V. O. Baron *et al.*, "Label-free optical vibrational spectroscopy to detect the metabolic state of M. tuberculosis cells at the site of disease," *Sci. Rep.*, vol. 7, no. 1, pp. 1–9, 2017.

[28] I. Barman *et al.*, "Effect of photobleaching on calibration model development in biological raman spectroscopy," *J. Biomed. Opt.*, vol. 16, no. 1, p. 011 004, 2011.

[29] L. Taslaman *et al.*, "A framework for regularized non-negative matrix factorization, with application to the analysis of gene expression data," *PloS One*, vol. 7, no. 11, e46331, 2012.

[30] C. L. Lawson *et al.*, *Solving least squares problems*. SIAM, 1995.

[31] S. Kim *et al.*, "An interior-point method for large-scale ℓ_1-regularized least squares," *IEEE J-STSP*, vol. 1, no. 4, pp. 606–617, 2007.

[32] M. Pelletier, "Effects of temperature on cyclohexane raman bands," *Appl. Spectrosc.*, vol. 53, no. 9, pp. 1087–1096, 1999.

[33] C. Berghian-Grosan *et al.*, "Raman spectroscopy and machine-learning for edible oils evaluation," *Talanta*, p. 121 176, 2020.

[34] Z. Huang *et al.*, "Near-infrared raman spectroscopy for optical diagnosis of lung cancer," *Int. J. Cancer*, vol. 107, no. 6, pp. 1047–1052, 2003.

Supplement

I. SMOOTHING

Tikhonov and generalized smoothing have been used to smooth the fluorescence spectrum (see section III for details), and are compared in Fig. 5. Parameter values are chosen in the same way for Tikhonov as for generalized regalarization, with the same values for λ_r but larger ones for λ_f, i.e. $\lambda_r^e \in \{11, 10, 9, 8\}$ and $\lambda_f^e \in \{6, 5, 4, 3, 2, 1\}$. We observe, both methods produce fluorescence with a similar overall shape but generalized smoothing is able to reduce the noise further than Tikhonov. Additionally, generalized regularization is able to diminish the fluorescence better between Raman shifts 1000 and 1250.

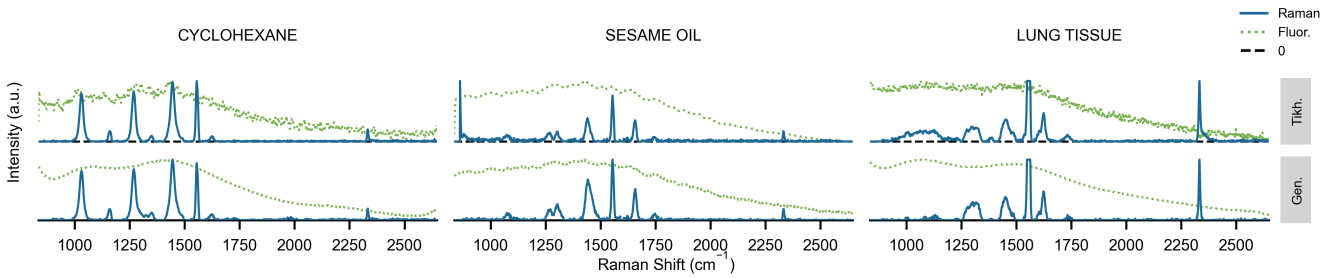

Fig. 5: The figure shows the inferred fluorescence and Raman spectra using Tikhonov (Tikh.) and generalized (gen.) smoothing for MSERS.

II. MSERS DETAILED ALGORITHM

Algorithm 2 MSERS

Input: $\mathbf{Y}$, $\{\Delta n_k\}_{k=2}^{K}$, $\tilde{\mathbf{K}}$ λ_r, λ_f, $i_{\max}$, tol
Require: $\mathbf{Y} \geq 0$
$r_n^{[0]} = \sqrt{\frac{\sum_k (y_{kn} - \bar{y}_n)^2}{K}}$, $n = 1, \ldots, N$
$f_n^{[0]} = \min_k y_{kn}$, $n = 1, \ldots, N$
Set $\mathbf{L}$
$i = 0$, $c^{[0]} = 0$ $c^{[-1]} = 0$
while $i \leq i_{\max}$ & $|c^{[i]} - c^{[i-1]}| \geq \text{tol} \times c^{[i-1]}$ **do**
  $i = i + 1$
  for $k \in \{1, \ldots, K\}$ **do**
    $\mathbf{C} = [\mathbf{L}^{(\Delta n_k)}\mathbf{r}^{[i-1]}, \mathbf{f}^{[i-1]}, \mathbf{1}]$
    $\mathbf{x} = \text{NNLS}(\mathbf{C}, \mathbf{y}_k)$
    $\alpha_k^{[i]} = x_1$, $\beta_k^{[i]} = x_2$, $b_k^{[i]} = x_3$
  end for
  for $k \in \{1, \ldots, K\}$ **do**
    $\mathbf{y}_k^f = \mathbf{y}_k - \alpha_k^{[i]}\mathbf{L}^{(\Delta n_k)}\mathbf{r}^{[i-1]} - b_k^{[i]}$
  end for
  $\mathbf{C}^f = \text{CHOL}\left(\left(\sum_{k=1}^{K} \beta_k^{[i]^2}\right)\mathbf{I} + \lambda_f \tilde{\mathbf{K}}\right)$
  $\mathbf{b}^f = \sum_{k=1}^{K} \beta_k^{[i]} \mathbf{y}_k^f$
  $\mathbf{d}^f = {\mathbf{C}^f}^{\top^{-1}} \mathbf{b}^f$
  $\mathbf{f}^{[i]} = \text{NNLS}(\mathbf{C}^f, \mathbf{d}^f)$
  for $k \in \{1, \ldots, K\}$ **do**
    $\mathbf{y}_k^r = \mathbf{y}_k - \beta_k^{[i]}\mathbf{f}^{[i]} - b_k^{[i]}$
  end for
  $\mathbf{C}^r = \left(\sqrt{\sum_{k=1}^{K} \alpha_k^{[i]^2}}\right)\mathbf{I}$
  $\mathbf{b}^r = \sum_{k=1}^{K} \alpha_k^{[i]} {\mathbf{y}_k^r}^{\top} \mathbf{L}^{(\Delta n_k)}$
  $\mathbf{d}^r = {\mathbf{C}^r}^{\top^{-1}} \mathbf{b}^r$
  $\mathbf{r}^{[i]} = \text{NNLASSO}(\lambda_r, \mathbf{C}^r, \mathbf{d}^r)$
  $c^{[i]} = \sum_{k=1}^{K} \left\| \mathbf{y}_k - \left(\alpha_k^{[i]}\mathbf{L}^{(\Delta n_k)}\mathbf{r}^{[i]} + \beta_k^{[i]}\mathbf{f}^{[i]} + b_k^{[i]}\right)\right\|_2^2$
    $+\lambda_f {\mathbf{f}^{[i]}}^{\top} \tilde{\mathbf{K}} \mathbf{f}^{[i]} + \lambda_r \|\mathbf{r}^{[i]}\|_1$
end while
return $(\mathbf{r}, \mathbf{f}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{b})$

III. OPERATOR MATRIX

For example, for a shift of size 2, $\Delta n = 2$, for $N = 5$, i.e. if the Raman intensity values are shifted by 2 places to the right, the following would apply,

$$\mathbf{L}^2\mathbf{r} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \\ r_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ r_1 \\ r_2 \\ r_3 \end{bmatrix}.$$

IV. Initialization

McCain *et al.* [20], Cooper *et al.* [22] and Marshall and Cooper [22] agree that the initial background should be calculated as the minimum value across the recorded spectra, for each $n = 1, \ldots, N$. However they suggest different approaches to the Raman spectrum. Cooper *et al.* use the standard deviation across the recorded spectra, for each $n = 1, \ldots, N$ whilst Marshall and Cooper use the maximum across the spectra. McCain *et al.* subtract the initial background from each recorded raw spectrum, un-shift them, i.e. apply $\mathbf{L}^{-(\Delta n_k)}$, and take the average. Fig. 6 shows the initial spectra as well as the inferred spectra using MSERS with SESAME OIL. It is evident that all three initializations produce very similar inferred Raman spectra. However, we observe that the maximum provides the least 'Raman-like' initial estimate and average un-shifting has a large increase on the far right of the spectrum, due to the shifting. The standard deviation provides an initial estimate closest to that of the inferred. Therefore, although all provide very similar Raman spectra, standard deviation is used (see section III).

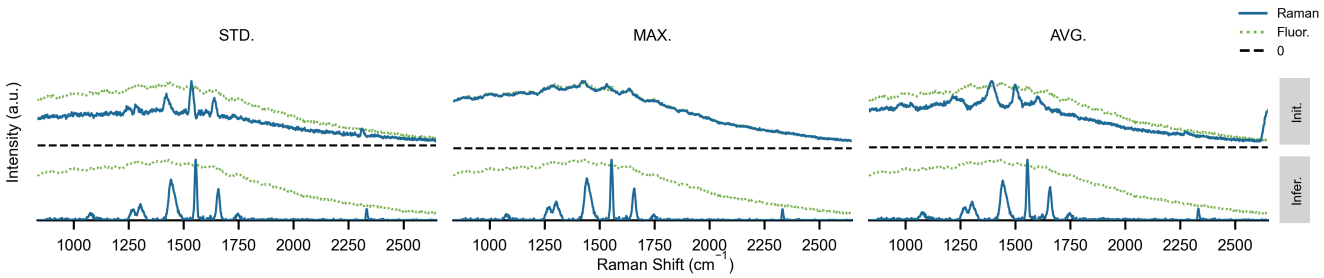

Fig. 6: The figure shows the initial (init.) Raman and fluorescence estimates along with the respective inferred (infer.) spectra for SESAME OIL using MSERS.

V. Regularization

To choose the optimum regularization parameter pair (see section III), Fig. 7 is used. The combinations are first filtered to those which produce inferred Raman and fluorescence with a derived sparsity between 0.3 and 0.7 (shown in blue). Of those, the pairing with the lowest correlation is chosen at the optimum (shown in red). In cases where there are multiple pairs with the same correlation, as in LUNG TISSUE, the higher sparsity value is chosen. For SESAME OIL $\lambda_r = NK \times 10^{-8}$ gives a zero Raman spectrum.

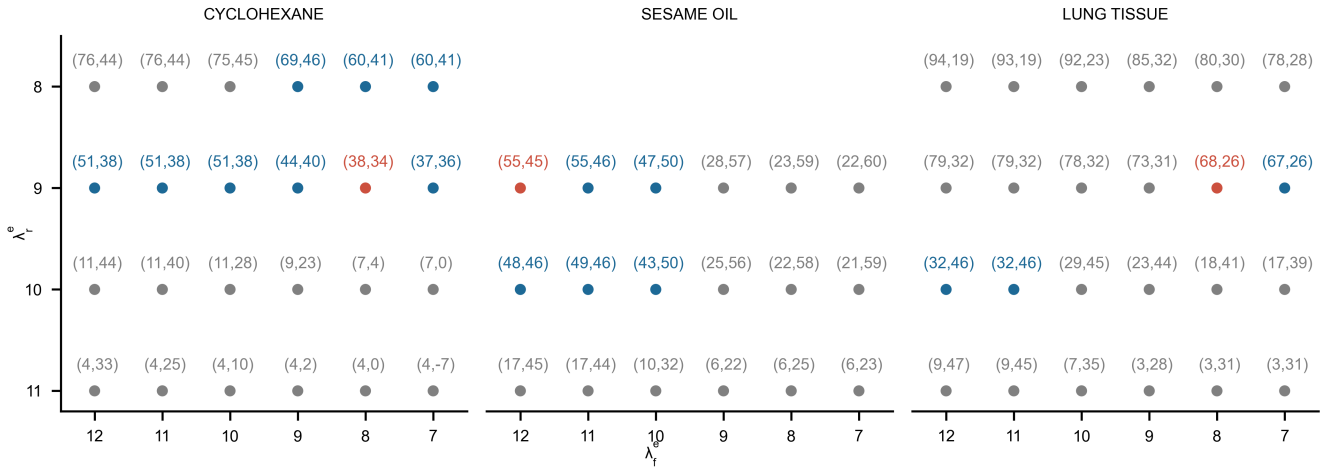

Fig. 7: The figure shows the derived sparsity of the Raman spectrum and correlation between the Raman and fluorescence spectrum, for each regularization parameter combination. The labels on each point give (sparsity$\times 10^2$, correlation$\times 10^2$). The blue points are those whose sparsity lies within the $[0.3, 0.7]$ threshold and the red points are those which are chosen to be the 'best' i.e. those with sparsity within the threshold and with lowest correlation.

Fig. 8 demonstrates the effect of varying regularization parameters on resulting inferred spectra. This is clearest in CYCLOHEXANE where the background spectrum's overall shape remains the same but its noise is reduced with increasing λ_f (decreasing λ_f^e). Additionally, the Raman spectrum does not lie on $y = 0$ for low λ_r (high λ_r^e) values whereas increased values maintain the key Raman peaks and reduce the surrounding spectra to 0.

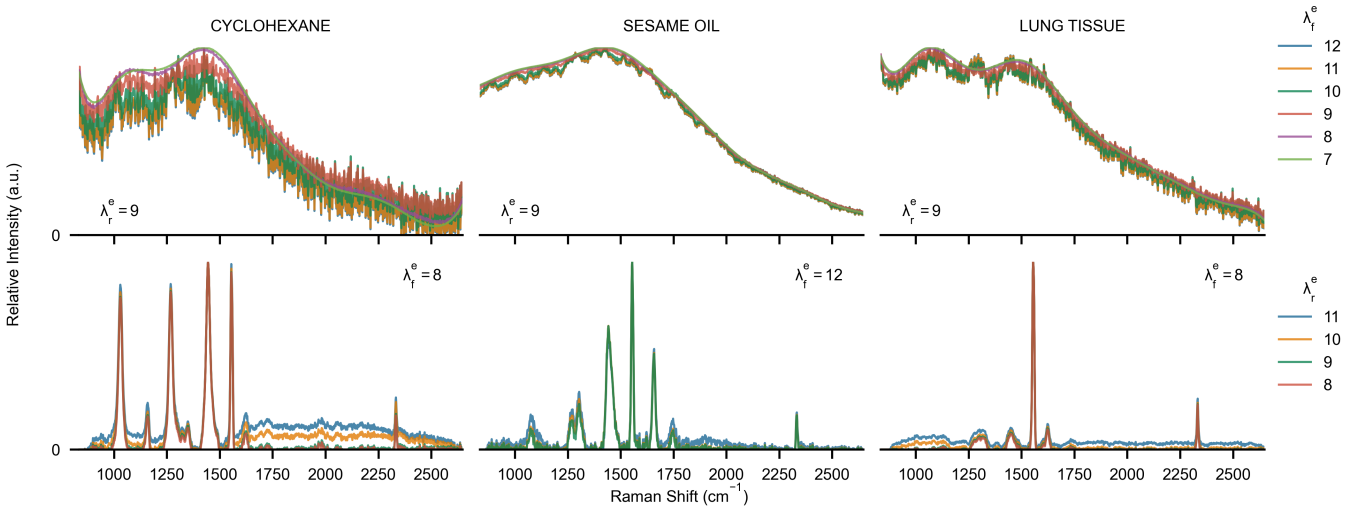

Fig. 8: The figure shows **(top)** the fluorescence spectra for constant sparsity parameter value λ_r and varying smoothness parameters λ_f, and **(bottom)** the Raman spectra for constant smoothness parameter λ_f and varying sparsity parameter values λ_r, for the chosen optimum parameter values for CYCLOHEXANE ($\lambda_r = KN10^{-9}, \lambda_f = KN10^{-8}$), SESAME OIL ($\lambda_r = KN10^{-9}, \lambda_f = KN10^{-12}$) and LUNG TISSUE ($\lambda_r = KN10^{-9}, \lambda_f = KN10^{-8}$).

Table IV provides the optimum parameter value pairs for the spectra given over varying number excitation wavelengths, k (see section IV-D.2).

TABLE IV: Chosen parameter values for varying number of excitation wavelengths, given as $KN10^{-\lambda^e_{(\cdot)}}$.

	CYCLOHEXANE		SESAME OIL		LUNG TISSUE	
Exc. (K)	λ^e_r	λ^e_f	λ^e_r	λ^e_f	λ^e_r	λ^e_f
2	7	8	9	12	9	7
4	9	8	10	12	9	9
6	9	8	9	12	9	8
8	9	8	10	12	9	7
10	9	8	9	12	9	8

VI. SPECTRUM SHIFT

The oxygen and nitrogen peaks have been detected by finding the peak position within the ranges 1500-1620 cm^{-1} and 2300-2350 cm^{-1} (in Raman Shift) respectively. There is little difference in the value of shift estimated from oxygen and nitrogen peaks (see section III) for CYCLOHEXANE and LUNG TISSUE, seen in Fig. 9. The differences seen for SESAME OIL are due to the large amount of background fluorescence around the nitrogen peak, resulting in difficulty in finding its exact location.

For CYCLOHEXANE, $K = 10$ measurements were obtained, resulting in the shift indices (Δn_ks) $\{0, 4, 7, 10, 12, 14, 16, 18, 19, 21\}$. For SESAME OIL, $K = 10$ measurements were obtained, resulting in the shift indices (Δn_ks) $\{0, 3, 6, 10, 12, 12, 16, 17, 19, 21\}$. For LUNG TISSUE, $K = 10$ measurements were obtained, resulting in the shift indices (Δn_ks)$\{0, 4, 7, 10, 12, 15, 16, 18, 19, 20\}$.

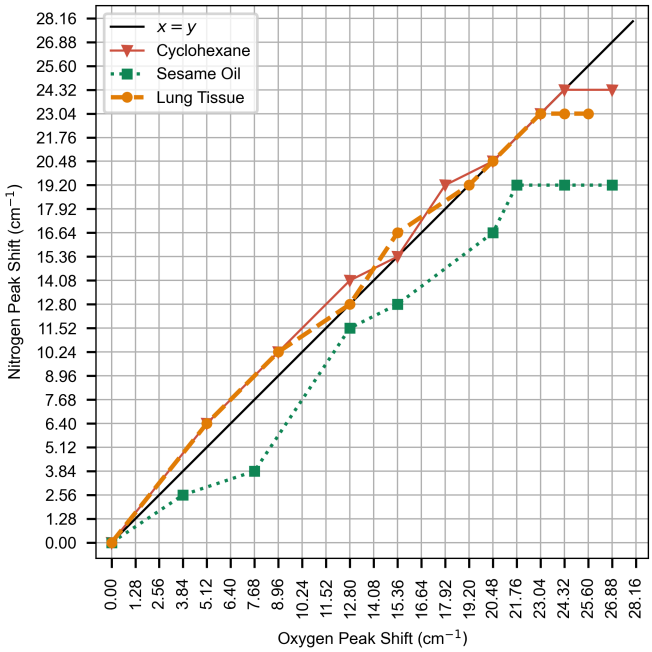

Fig. 9: The figure shows the shift in excitations estimated using oxygen and nitrogen peaks with respect to the first excitation. The grid shows the resolution of the spectra.

VII. Interpolation

The collected data is processed by converting from wavelength to wavenumber and then interpolated onto an equispaced grid (see section IV-B for details). Fig. 10 shows how the spectra's resolution changes with interpolation.

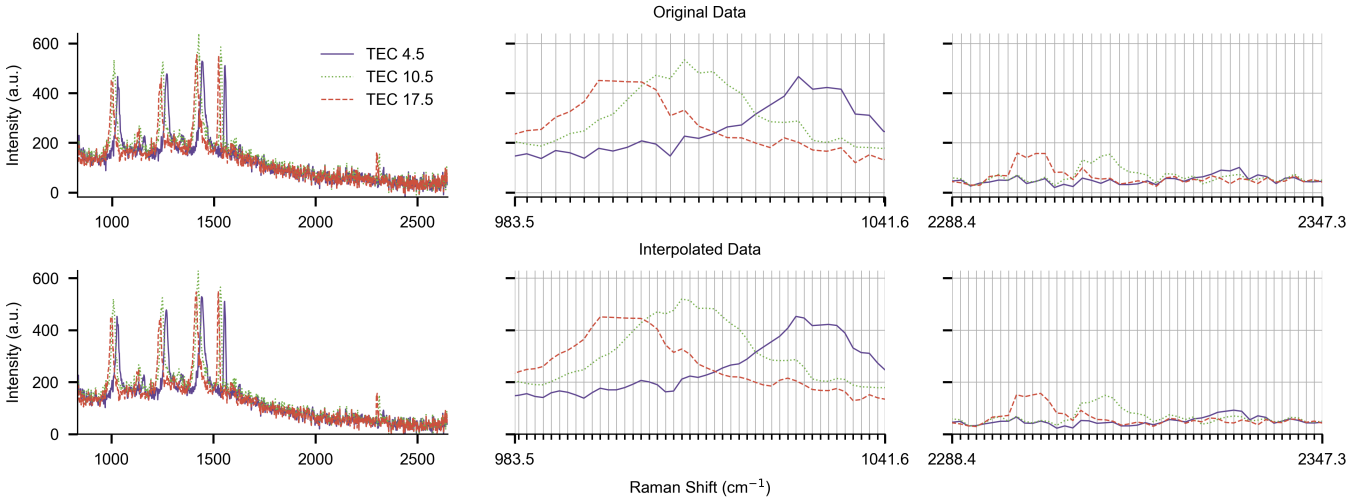

Fig. 10: The figure shows 3 of the original and interpolated spectra for CYCLOHEXANE, with a focused view on the left most peak ($\sim$1028 cm^{-1}) and the nitrogen peak ($\sim$2331 cm^{-1}). The grid lines show the positions of the data points.

VIII. Changing Number of Excitations

When changing the number of excitations (see section IV-D.2), the spectra are chosen such that the difference in the excitation wavelength between each spectrum is maximized. For example, when $K = 2$, we choose the 1st and 10th spectrum; when $K = 4$, we choose the 1st, 4th, 7th and 10th spectrm; when $K = 6$, we choose the 1st, 3rd, 5th, 7th, 9th and 10th and when $K = 8$, we choose the 1st to the 8th.

IX. Additional Inferred Spectra Evaluation

Table V presents additional characteristics of the resulting spectra in Fig. 4 (see section IV-D.3). We expect all the information in the observed spectra to be contained in the inferred Raman and fluorescence spectra, resulting in residuals that are 'white noise', i.e. residuals with no autocorrelation. The Durbin-Watson test statistic provides a measure of this autocorrelation in residuals, where values around 2 suggest no autocorrelation and values below (above) suggest positive (negative) autocorrelation. The test statistic was calculated for each observed spectrum and the value furthest from 2 was reported. Although all values in Fig. 4 are below 1, MSERS shows the highest across all datasets.

The table also gives peak start and end positions, without the assumption of symmetry as well the relative intensities of the peaks that are not presented in Table III.

This article has been accepted for publication in IEEE Transactions on Biomedical Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TBME.2023.3243866

TABLE V: The table summarizes the performance of fluorescence suppression methods (see section IV-D.3). a.u. is arbitrary unit. i/d is indistinguishable. DW is Durbin-Watson test statistic.

		Peak Location (Peak Start , Peak End) [Relative Intensity (a.u.)]									DW
CYCLOHEXANE	AIRPLS	1028 (994,1066) [0.86]	1159 (1116,1181) [0.27]	1270 (1239,1307) [0.87]	1343 (1327,1369) [0.22]	1442 (1410,1495) [1.00]					
	RSERDS	1030 (1002,1069) [1.00]	1157 (1088,1194) [0.22]	1267 (1239,1299) [1.09]	1345 (1314,1369) [0.17]	1445 (1413,1494) [1.19]					
	PCA	1031 (942,1092) []	1159 (1125,1180) []	1270 (1202,1312) []	1350 (1321,1372) []	1446 (1392,1511) []					
	PIP	1030 (998,1087) [0.82]	1159 (1135,1171) [0.19]	1268 (1238,1311) [0.85]	1349 (1319,1359) [0.13]	1445 (1415,1495) [0.99]					0.167
	SNMF	1030 (1001,1074) [0.82]	1159 (1135,1172) [0.22]	1268 (1231,1311) [0.86]	1349 (1319,1359) [0.14]	1445 (1415,1495) [0.99]					0.603
	SICA	1029 (1001,1065) [0.77]	1159 (1138,1171) [0.36]	1268 (1248,1307) [0.87]	1350 (1321,1360) [0.37]	1445 (1410,1496) [1.01]					0.595
	MSERS	1030 (1001,1056) [0.86]	1159 (1136,1170) [0.19]	1268 (1241,1312) [0.89]	1350 (1337,1359) [0.13]	1445 (1415,1490) [1.06]					0.845
SESAME OIL	AIRPLS	1078 (1057,1088) [0.38]	1270 (1253,1282) [0.23]	1300 (1240,1334) [0.34]	1435 (1422,1469) [0.50]	1660 (1632,1668) [0.54]	1750 (1746,1752) [0.28]				
	RSERDS	1076 (1053,1143) [0.14]	1271 (1244,1281) [0.18]	1302 (1285,1344) [0.16]	1441 (1415,1488) [0.65]	1655 (1607,1675) [0.60]	1755 (1710,1782) [0.11]				
	PCA	i/d	i/d	i/d	i/d	i/d	i/d				
	PIP	1075 (999,1157) [0.38]	1273 (1257,1286) [0.47]	1304 (1220,1336) [0.51]	1442 (1400,1495) [0.79]	1659 (1611,1679) [0.64]	1748 (1709,1765) [0.31]				0.003
	SNMF	1075 (1055,1162) [0.34]	1270 (1259,1280) [0.37]	1302 (1244,1387) [0.45]	1442 (1426,1492) [0.63]	1659 (1646,1670) [0.63]	1748 (1724,1765) [0.31]				0.317
	SICA	1075 (1057,1110) [0.27]	i/d	1304 (1253,1335) [0.53]	1442 (1419,1491) [0.82]	1659 (1637,1671) [0.76]	1748 (1712,1759) [0.51]				0.049
	MSERS	1075 (1056,1099) [0.12]	1270 (1255,1286) [0.18]	1304 (1249,1332) [0.26]	1442 (1418,1495) [0.67]	1659 (1638,1677) [0.51]	1748 (1724,1765) [0.11]				0.631
LUNG TISSUE	AIRPLS	1079 (1071,1080) [0.08]	1263 (1262,1266) [0.10]	1304 (1287,1307) [0.15]	1447 (1422,1499) [0.14]	1600 (1596,1602) [0.08]	1618 (1613,1619) [0.12]	1668 (1654,1682) [0.07]	1750 (1744,1760) [0.08]		
	RSERDS	1069 (1046,1098) [0.04]	i/d	1296 (1257,1349) [0.11]	1451 (1423,1501) [0.14]	i/d	1624 (1595,1666) [0.16]	i/d	1741 (1691,1794) [0.07]		
	PCA	i/d	i/d	i/d	i/d	i/d	i/d	i/d	i/d		
	PIP	i/d	i/d	1316 (1254,1349) [0.10]	1455 (1428,1500) [0.10]	i/d	1624 (1590,1663) [0.26]	i/d	1733 (1691,1746) [0.13]		0.008
	SNMF	i/d	i/d	1295 (1253,1343) [0.11]	1454 (1422,1500) [0.14]	i/d	1624 (1578,1645) [0.16]	i/d	1733 (1691,1764) [0.05]		0.464
	SICA	i/d	i/d	1291 (1248,1339) [0.11]	1446 (1422,1496) [0.14]	i/d	1620 (1574,1642) [0.17]	i/d	1729 (1682,1783) [0.07]		0.294
	MSERS	1087 (1076,1097) [0.01]	i/d	1302 (1257,1344) [0.07]	1449 (1422,1494) [0.09]	i/d	1624 (1592,1642) [0.11]	i/d	1733 (1725,1750) [0.02]		0.658