Cosmic ray removal from Raman spectra for improved classification of biological specimens

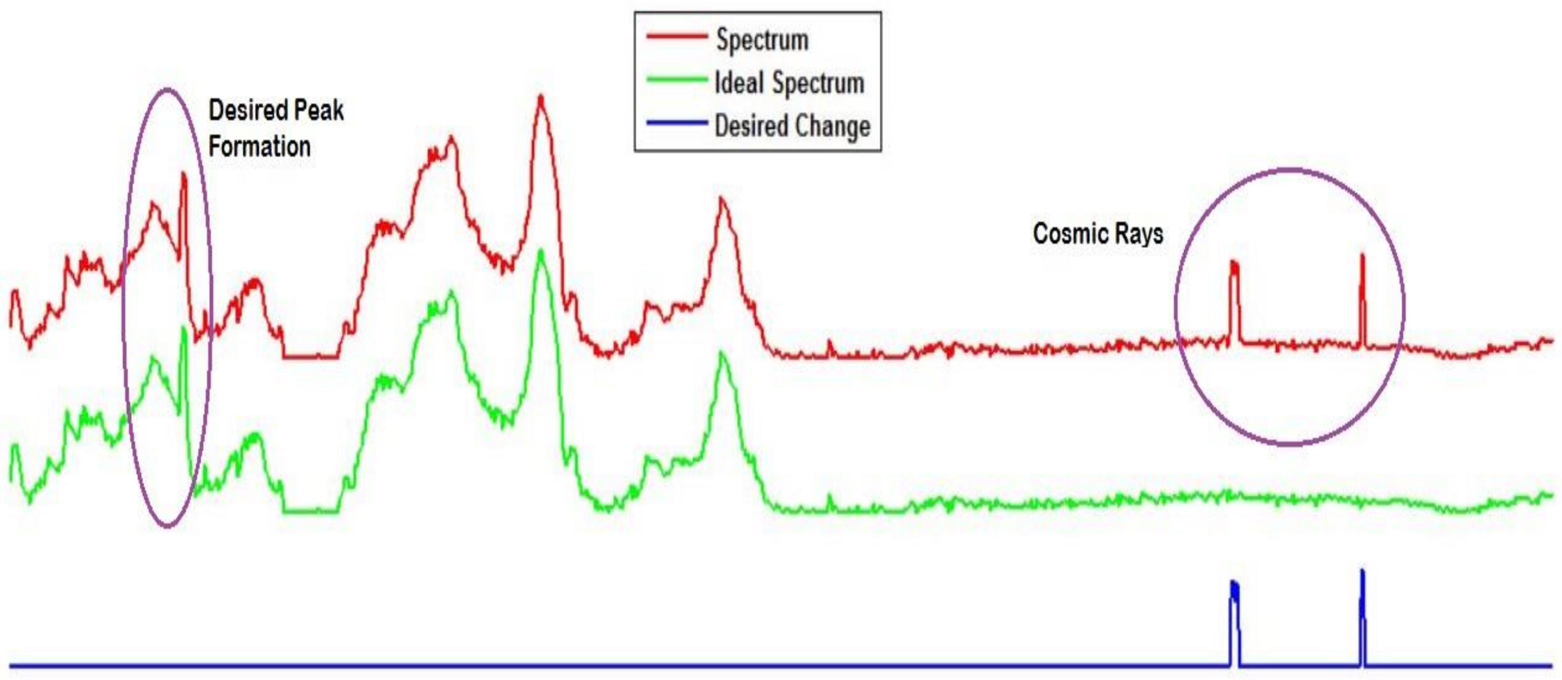
Maynooth University
National University of Ireland Maynooth

Sinead J. Barton[1] and Bryan M. Hennelly[1,2]

Biomedical Engineering Research Group

1. Abstract

Raman spectroscopy is an optical technique that can be used to evaluate the biomolecular composition samples in a non-invasive manner. This technique is based on inelastic light scattering as a result of the interaction between incident monochromatic light (laser) and a biological sample. Cosmic ray artefacts are random unidirectional sharp spikes that distort spectra collected from array detectors [1]. They not only reduce the signal to noise ratio of the system but may also have an impact on the post processing of spectra. The method proposed here is useful for applications where datasets of spectra are recorded. It uses normalised covariance to identify a similar spectrum in that dataset. A direct comparison reveals the presence of cosmic rays, which can then be replaced with the appropriate values from the most similar spectrum without interfering with any other spectral features.



2. Algorithm

a. A dataset of spectra is supplied to the algorithm, where each row of the matrix is a spectrum of the form: $S_o(n\Delta\lambda)$

Where $\Delta\lambda$ = 0.187 nm, n = {0,1, 2 ….. N -1} and $N\Delta\lambda$ = 1024

b. Spectra are first area normalised

Where $\overline{S_o}$ is the mean of $S_o(n\Delta\lambda)$

$$S_o(n\Delta\lambda) = \frac{S_o(n\Delta\lambda)}{\overline{S_o}}$$

c. The most similar spectrum in the dataset is identified by finding the spectrum with the highest normalised covariance [2]

$$N_{cov} = \frac{(S_o \cdot S_j)^2}{(S_o \cdot S_o)(S_j \cdot S_j)}$$

Where $S_o(n\Delta\lambda)$ is the spectrum being processed and S_j represents all other spectra in the dataset

d. A threshold is decided by calculating the standard deviation of the spectrum [3]

Where k is the kth element in the spectrum.

$$\sigma = \sqrt{\frac{1}{n}\sum_{k=1}^{n}\left(S_o(k\Delta\lambda) - \overline{S_o}\right)^2}$$

e. The threshold is applied to the residuals obtained by subtracting the most similar spectrum from the original such that:

1. $$S_o(k\Delta\lambda) = \begin{cases} S_p(k\Delta\lambda) \ldots\ldots if\ x_{res}(k) > \sigma \\ S_o(k\Delta\lambda) \ldots\ldots if\ x_{res}(k) < \sigma \end{cases}$$

2. $$S_o(k \pm 1\Delta\lambda) = \begin{cases} S_p(k \pm 1\Delta\lambda) \ldots\ldots if\ x_{res}(k \pm 1) > {}^{\sigma}/_{2} \\ S_o(k \pm 1\Delta\lambda \ldots\ldots if\ x_{res}(k \pm 1) < {}^{\sigma}/_{2} \end{cases}$$

Where $S_p(k\Delta\lambda)$ is the spectrum with the highest normalised covariance and x_{res} is the array of residuals. Step 2 is only performed on the condition that step 1 was already activated.

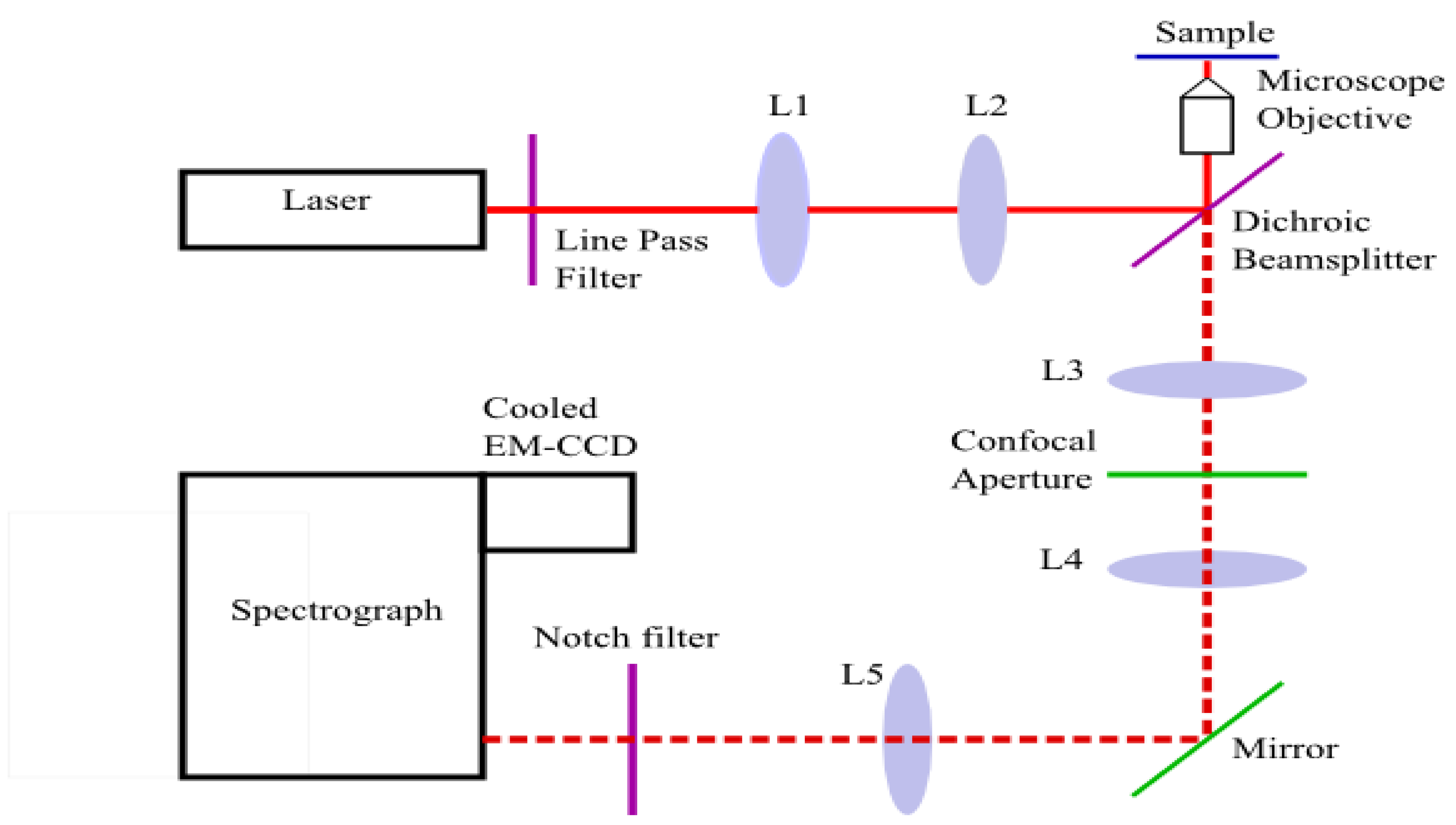


Fig 4: A diagram of the system used to collect the datasets used to test the algorithm. (L1 and L2 are collimating lenses)

3. Results

After background subtraction is applied [4, 5] spectra are processed as rows in a matrix which comprises the dataset.

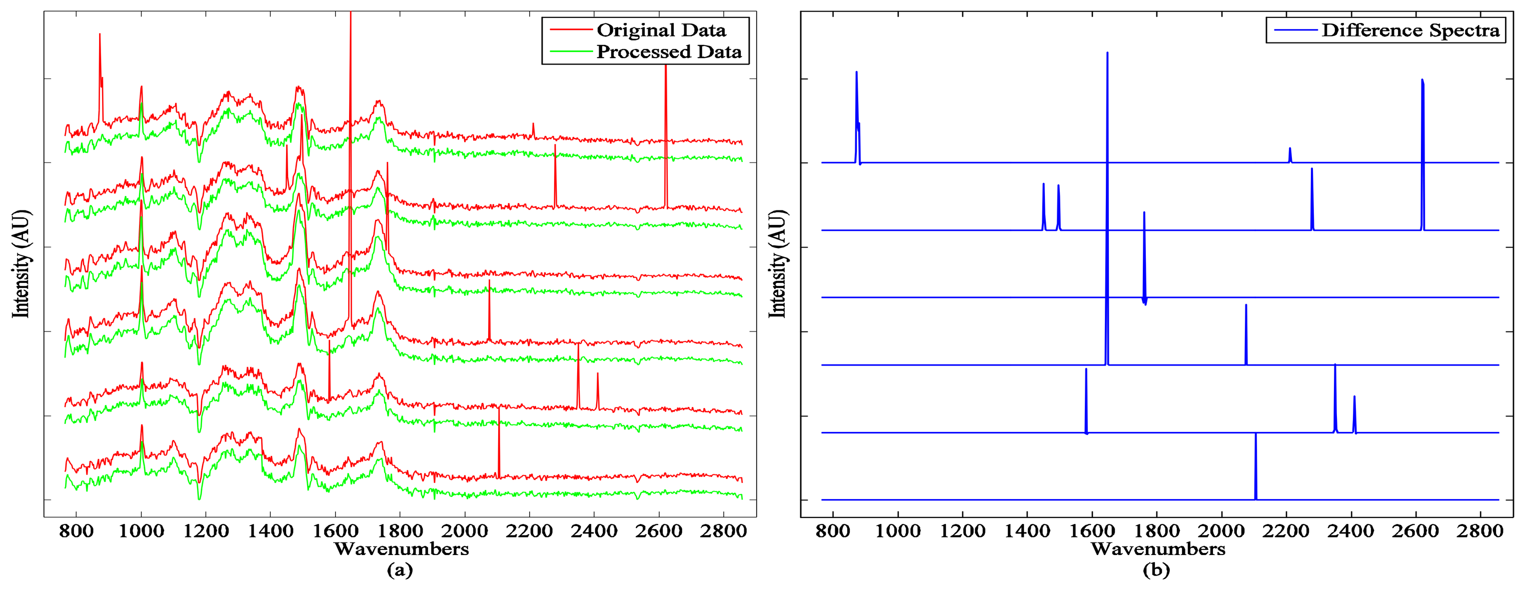


Fig 1: An image of 6 spectra, from an original dataset of 21, (a) before and after processing (b) the residuals when the green is subtracted from the red

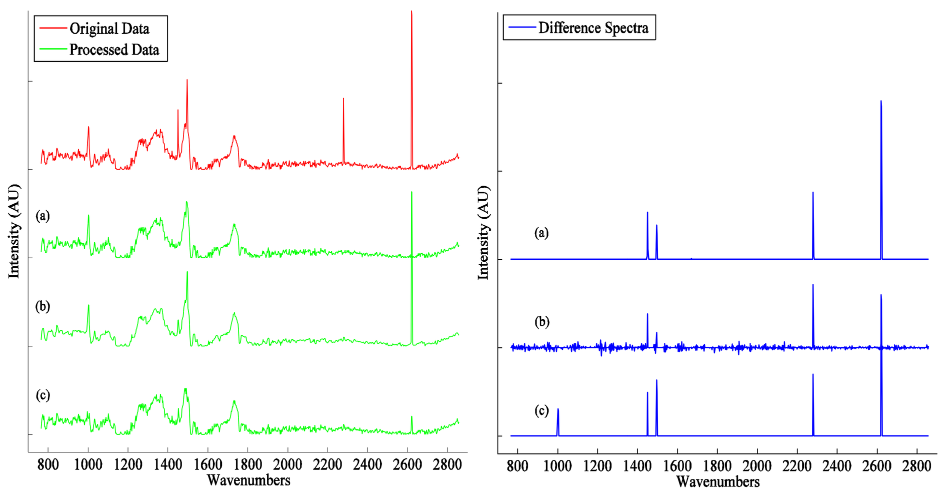


Fig 2: A comparison between the performance of (a) the algorithm developed here (b) median filtering and (c) where Savitsky-Golay smoothing is used to approximate a true spectrum and a threshold is applied to the residuals.

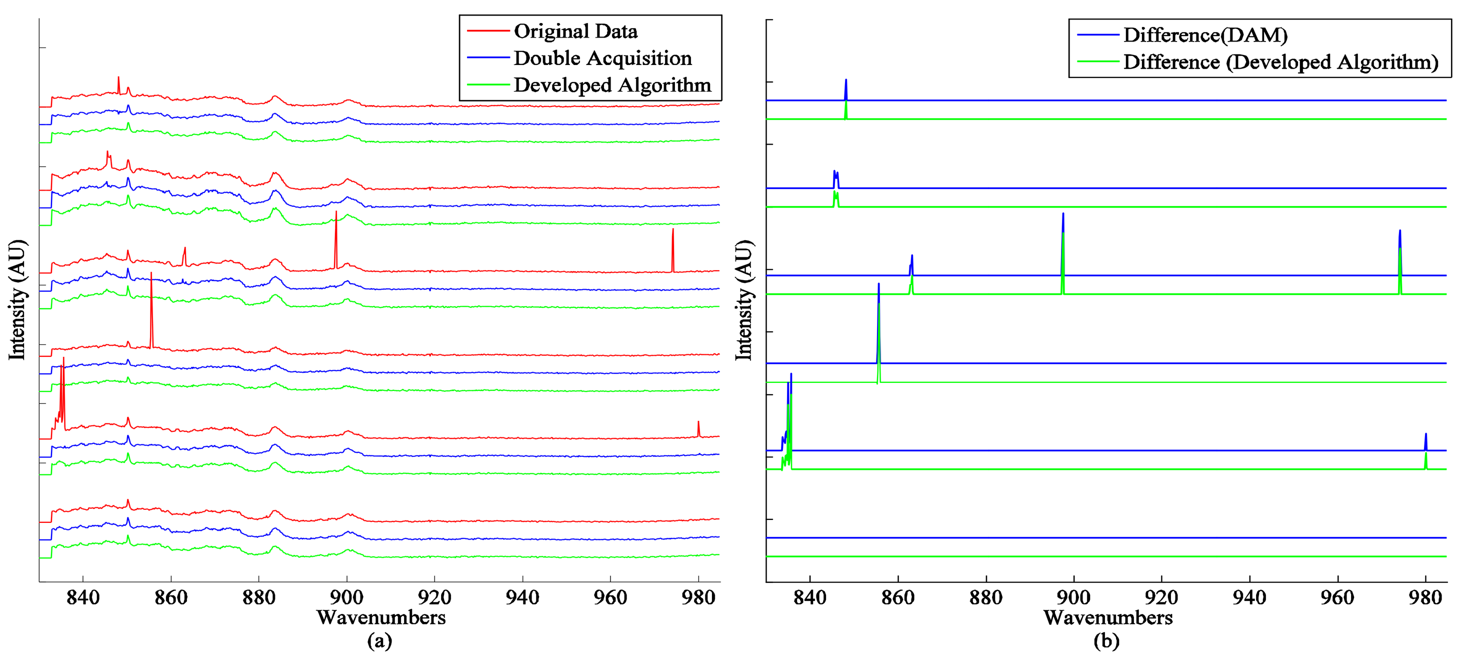


Fig3: A performance comparison between the developed algorithm and the double acquisition method [6].

4. Conclusion

The goal of removing cosmic rays while making no other change to the spectrum is challenging. The proposed algorithm is fully automated and can be applied to any dataset of similar spectra. When compared to median filtering and Savitsky-Golay processing, these two methods produced either false negatives or false positives which caused undesirable changes to the spectrum while the developed algorithm removed only the cosmic rays. The algorithm achieved a high sensitivity for datasets as small as 5 spectra as well as datasets which contained a mix of different cell lines. This method achieves similar results to a method proposed in 1993 [5], the double acquisition method, but it does not require the additional capture of spectra which could halve the time taken to record data.

Affiliations:
1. Electronic Engineering, Dept. Maynooth University, Co. Kildare, Ireland
2. Callan Institute, Maynooth University, Co. Kildare, Ireland

Acknowledgements:
This research was conducted with the financial support of the Irish Research Council (IRC) under project ID GOIPG/2013/1434 and Science Foundation Ireland (SFI) under Grant Number 11/SIRG/12140. I would like to thank the IRC and SFI for their support.

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

sfi
science foundation ireland
fondúireacht eolaíochta éireann

References:
[1] J.R. Janesick. *Scientific Charge-Coupled Devices.* SPIE (2001)
[2] U. Cappel, I. Bell and L. Pickard. *Applied Spectroscopy*, 64:195 – 200 (2010)
[3] Y. Katsumoto and Y. Ozaki. *Applied Spectroscopy*, 57:317:322 (2003)
[4] L.T. Kerr, H.J. Byrne and B.M. Henelly. *Analytical Methods*, Under Review (2014)
[5] B.D. Beier and A.J. Berger. *Analyst*, 134:1198 – 1202 (2009)
[6] H. Takeuchi and I. Harada. *Applied Spectroscopy*, 47:129 – 131 (1993)