Applying Noise Adjusted Principal Component Analysis for Noise Reduction of Raman Micro Spectroscopic Image Datasets

Frederik J. Bjerring[1] and Martin A. B. Hedegaard[2]

[1]Department of Mathematics and Computer science (IMADA), University of Southern Denmark, Campusvej 55, 5230, Odense M, Denmark
[2]Department of Chemical Engineering, Biotechnology and Environmental Technology, University of Southern Denmark, Campusvej 55, 5230, Odense M, Denmark

One of the ever-occurring problems when doing Raman spectroscopic mapping is background noise. This noise originated from CCD noise or cosmic rays to faults in the equipment and other experimental factors and is especially significant when using very short acquisition times <1s. These short acquisition times are often required to examine biological samples such as cells and tissue to prevent damage. This however increases the background noise levels that will then need correction in pre processing.

Previous research in the field of noise reduction has resulted in a number of ways to reduce the noise after measurements. One way of reducing noise in image datasets is using the factor analysis such as Principal Component Analysis (PCA). PCA however does have the problem that it might result in loss of signal due to PCA sorting the bands according to variance. A proposed solution to noise reduction is the maximum noise fraction (MNF) introduced by Green et al.[1] which sorts the components according to S/N-ratio. A similar technique is the noise adjusted principal component analysis (NAPC)[2] which in essence operate in the same way. The sorting of the components in both cases is based on prior knowledge of the noise in the dataset. This means that we have to devise a method to effectively estimate the noise based on the dataset itself. To investigate the use of NAPC could be used we compared several approaches, firstly PCA noise reduced data retaining 5 components and then comparing to NAPC using shifted image approach, NAPC using low signal spectra around the cell to estimate the noise and NAPC using low signal spectra around the cell to and mean centering to estimate the noise.

The best method for reducing the noise using the same number of components for all methods was found to be NAPC using low signal spectra around the cell to estimate the noise, which retained most information while reducing the noise. It should also be noted that incorrect noise estimation can result in significant artifacts being introduced into the dataset.

References

[1] A.A. Green, M. Berman, P. Switzer, M.D. Craig, *IEEE Transactions on Geoscience and Remote Sensing* 26, 65–74 (1988).
[2] J.B. Lee, A.S. Woodyatt, M. Berman, *IEEE Transactions on Geoscience and Remote Sensing* 28, 295-304 (1990).
[3] R.K. Reddy, R. Bhargava, *Analyst* 135(11), 2818-25 (2010).
[4] B. Bird, M. Miljković, S. Remiszewski, A. Akalin, M. Kon, M. Diem, *Laboratory Investigation* 92, 1358–1373 (2012).