Supplementary Information

Utilization of Spectral Vector Properties in Multivariate Chemometrics Analysis of Hyperspectral Infrared Imaging Data for Cellular Studies

Suat-Teng Tan, Kejia Chen, Suyun Ong, Wee Chew[*]

Institute of Chemical and Engineering Sciences (ICES), Agency for Science, Technology and Research (A*STAR). 1 Pesek Road, Jurong Island, Singapore 627833. Singapore.

Vector Representation of Infrared Spectrum, Spectral Vector Properties and Multivariate Chemometrics

The underpinning concept that is common in the multivariate chemometrics approach of this paper is the utilization of various *spectral vector properties* (SVP) that characterizes certain spectral qualities.

$$A_{ij}(\bar{\nu}) = -\log_{10}\left(\frac{I_{ij}(\bar{\nu})}{I_{ij}^{o}(\bar{\nu})}\right) = \log_{10}\left(\frac{1}{T_{ij}(\bar{\nu})}\right) = c_i \cdot \varepsilon_{ij}(\bar{\nu}) + \phi_{ij}(\bar{\nu}) \tag{1}$$

$$\vec{\mathbf{a}}_{1\times\nu} = c_i \cdot \vec{\boldsymbol{\varepsilon}}_{1\times\nu} + \boldsymbol{\varphi}_{1\times\nu} \tag{2}$$

If the linear Lambert-Beer-Bouguer law is valid, according to Eq. (1) the decadic infrared absorbance $A_{ij}(\bar{\nu})$ of each *i*-th chemical species relates to its molar concentration c_i and conditional molar absorptivity $\varepsilon_{ij}(\bar{\nu})$ at a specific pathlength l for every *j*-th wavenumber $\bar{\nu}$. $\phi_{ij}(\bar{\nu})$ is the associated experimental error, $I_{ij}(\bar{\nu})$ and $I_{ij}^{o}(\bar{\nu})$ are the

---

[*] Corresponding author (W. Chew). Tel.: (65) 6796 3961; Fax: (65) 6316 6185. Email: chew_wee@ices.a-star.edu.sg.

attenuated and source intensities respectively, and $T_{ij}\left(\overline{\nu}\right)$ is the transmittance in decimals. Therefore, for a range of wavenumber $[\overline{\nu}_{\min},\overline{\nu}_{\max}]$ with intervals $d\overline{\nu}$, the entire FTIR spectrum is treated as a wavelength-dependent multivariable one-dimensional (1D) *vector quantity* $\vec{\mathbf{a}}_{1\times v}$ with v spectroscopic channels according to Eq. (2), which contains both *magnitude* and *direction*. The concept of *vector magnitude* for FTIR spectrum signifies the amount of attenuation as infrared radiation interacts with the sample, and is indicated by the absorbance units (a.u.) at each *j*-th wavenumber $\overline{\nu}$. The *vector direction* in FTIR spectrum encapsulates the functional group fingerprint of each *i*-th chemical species,[1, 2] which is represented by its unique infrared absorption band profile and quantified as an *abstract spectral direction/ coordinates* by its molar absorptivity spectrum $\vec{\boldsymbol{\varepsilon}}_{1\times v}$.

By the Lambert-Beer-Bouguer law, the infrared absorbance $A_{ij}^{\text{mix}}(\overline{\nu})$ of a mixture of pure components is simply the sum of absorptivities $\varepsilon_{ij}(\overline{\nu})$ of its constituents weighted (or magnified) by the respective concentrations c_i at every *j*-th wavenumber $\overline{\nu}$, as given by Eq. (3). Thus, the abstract spectral direction of a multi-component mixture infrared spectrum is the weighted sum of the individual component molar absorptivity spectra. For a set of *m* number of mixture infrared spectra with *s* number of pure components and v spectroscopic channels, its matrix form is written as Eq. (4).[3]

$$A_{ij}^{\text{mix}}(\overline{\nu}) = \sum_{i=1}^{s} c_i \cdot \varepsilon_{ij}(\overline{\nu}) + \phi_{ij}(\overline{\nu}) \tag{3}$$

$$\mathbf{A}_{m\times v}^{\text{mix}} = \mathbf{C}_{m\times s}\mathbf{E}_{s\times v} + \mathbf{\Phi}_{m\times v} \tag{4}$$

The above vector representation and matrix treatment is not foreign in the quantitative analysis of infrared spectra.[3, 4, 5, 6] Certain mathematical operations common in vector/ matrix algebra can be performed on a FTIR spectrum $\vec{\mathbf{a}}_{1\times\nu}$ to yield different numerical (scalar) values that characterizes various spectral vector properties (SVP) of the spectrum. The following SVP were employed for the chemometrics analyses herein.

1. Sum of spectral derivatives (absolute values) of different orders applied to entire spectrum: first (D_1), second (D_2) or fourth (D_4) orders [7, 9, †]

$$D_\partial = \sum_j d_j^\partial\left(\vec{\mathbf{a}}_{1\times\nu}\right), \text{ where } j = 1, 2, ..., \nu \text{ and}$$

$$\partial = \begin{cases} 1 \;\Rightarrow\; d_j^1\left(\vec{\mathbf{a}}_{1\times\nu}\right) = \left|\dfrac{\mathrm{d}A_j(\bar{\nu})}{\mathrm{d}\bar{\nu}}\right| & \forall\, j = [2, \nu-1] \\ 2 \;\Rightarrow\; d_j^2\left(\vec{\mathbf{a}}_{1\times\nu}\right) = \left|\dfrac{\mathrm{d}^2A_j(\bar{\nu})}{\mathrm{d}\bar{\nu}^2}\right| & \forall\, j = [2, \nu-1] \\ 4 \;\Rightarrow\; d_j^4\left(\vec{\mathbf{a}}_{1\times\nu}\right) = \left|\dfrac{\mathrm{d}^4A_j(\bar{\nu})}{\mathrm{d}\bar{\nu}^4}\right| & \forall\, j = [3, \nu-2] \end{cases} \tag{5}$$

2. Integrated area (α) [7, 9]

$$\alpha = \int_1^\nu \vec{\mathbf{a}}_{1\times\nu}\mathrm{d}\bar{\nu} \approx \frac{\mathrm{d}\bar{\nu}}{2}[\vec{\mathbf{a}}(1) + \vec{\mathbf{a}}(\nu)] + \mathrm{d}\bar{\nu}\cdot\sum_{j=2}^{\nu-1}\vec{\mathbf{a}}(j) \tag{6}$$

3. Maximum value (L$^\infty$-norm) of the reciprocal of transmittance ($\|\vec{\boldsymbol{\tau}}\|_\infty$) [10]

$$\|\vec{\boldsymbol{\tau}}\|_\infty = \max_j\left\{|\tau_j| : j = 1, 2, ..., \nu\right\}, \text{ where } \tau_j = \frac{1}{T_j(\bar{\nu})} \tag{7}$$

[†] All numerical derivatives were computed using the central difference approximation.

4. Abstract spectral (statistical) distance (δ) based on L$^{\infty}$-norm spectra [8]

$$\delta\left(\vec{\mathbf{a}}_x^{\infty}, \vec{\mathbf{a}}_y^{\infty}\right) = \sqrt{\sum_{j=1}^{\nu}\left[A_{jx}^{\infty}(\bar{\nu}) - A_{jy}^{\infty}(\bar{\nu})\right]^2} \tag{8}$$

5. Generalized spectral (information) entropy (H) [9, 11, 15]

$$H\left(\vec{\mathbf{a}}_{1\times\nu}\right) = -\beta_h \sum_{j}^{\nu} h_j^{\partial} \ln h_j^{\partial} + \beta_{\partial} D_{\partial} + \beta_{\alpha}\alpha\,,$$
$$\text{where } \beta_h, \beta_{\partial}, \beta_{\alpha} = 1 \text{ or } 0\,,\ h_j^{\partial} = \frac{d_j^{\partial}\left(\vec{\mathbf{a}}_{1\times\nu}\right)}{D_{\partial}}\,,\ \partial = 1, 2 \text{ or } 4 \tag{9}$$

Mathematically, the sums of spectral derivatives (absolute values) is equivalent to the L^1-norms of respective spectral derivative vectors of the FTIR spectrum $\vec{\mathbf{a}}_{1\times\nu}$, that is $\left\|d^{\partial}\left(\vec{\mathbf{a}}_{1\times\nu}\right)\right\|_1$. D_1 characterizes the total gradient/ slope of the infrared spectrum, whilst D_2 and D_4 quantifies its total curvature and degree of noise level respectively. [9] The integrated area α measures the area under the FTIR spectrum and also reflects the amount of translational baseline differences between different spectra. The maximum value of the reciprocal of transmittance $\left\|\vec{\boldsymbol{\tau}}\right\|_{\infty}$ is derived from the mathematical definition of vector norm,[10] in particular, the L$^{\infty}$-norm of reciprocal transmittance spectrum $\vec{\boldsymbol{\tau}}_{1\times\nu}$ calculated from its absorbance spectrum $\vec{\mathbf{a}}_{1\times\nu}$. The maximum degree of attenuation, or concomitantly the minimum transmittance, in the FTIR spectrum $\vec{\mathbf{a}}_{1\times\nu}$ is indicated by the dimensionless value of $\left\|\vec{\boldsymbol{\tau}}\right\|_{\infty}$, which provides an indirect bearing on the degree of spectral linearity[‡] and

---

[‡] The linear Lambert-Beer-Bouguer law usually holds if maximum absorbance is less than ca. 2.0 a.u. (or equivalently, a minimum transmittance of more than 0.01). Beyond which, non-linear response of observed absorbance with concentration changes may set in.

aids outlier detection. The abstract Euclidean spectral distance (δ) between two normalized (L^{∞}-norm) infrared spectra $\vec{\mathbf{a}}_x^{\infty}$ and $\vec{\mathbf{a}}_y^{\infty}$ reveals the degree of their similarity/ dissimilarity, which is extremely useful for the unsupervised leader-follower cluster analysis (LFCA) described in this paper. The generalized concept of information entropy H for infrared spectroscopy as defined previously by Garland and co-workers was utilized for the self-modeling curve resolution of pure component spectra.[9] The coefficients β_h, β_{∂} and β_{α} in Eq. (9) are set to unity to activate its corresponding generalized entropy term (sum of ∂-order spectral derivative based L^1-normalized Shannon entropy h_j^{∂}, sum of ∂-order spectral derivative D_{∂}, and integrate area α respectively), or otherwise set to zero to deactivate the particular term. The band-target entropy minimization (BTEM) curve resolution algorithm was applied with Eq. (9), which has been proven successful in resolving multiple spectral components with highly-overlapping infrared bands and containing some degree of spectral non-linearity,[9, 11, 12] including elucidating new (previously unknown) infrared spectra in complex organometallic chemistry.[13, 14, 15]

Additional Figures

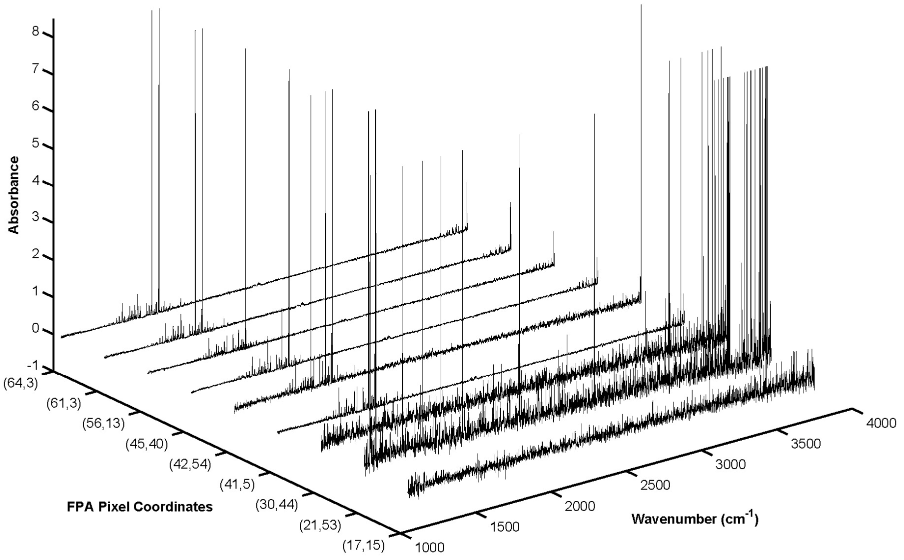

Figure S1 Outliers detected via constrained hierarchical cluster analysis of combined five-SVP values

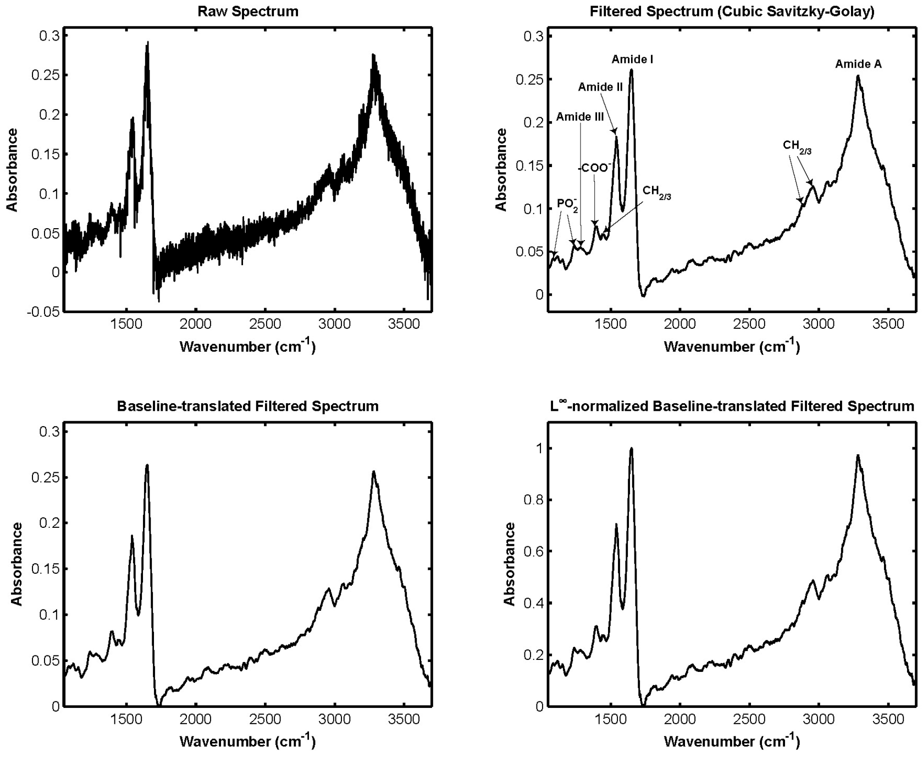

Figure S2 Cubic Savitzky-Golay filtering, baseline translation and L$^{\infty}$-normalization of infrared spectra

REFERENCES

---

[1] B. H. Stuart, in *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, ed. R. A. Meyers, John Wiley & Sons, West Sussex, **2000**, vol. 1, pp. 529-559.

[2] D. Naumann, in *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*, ed. R. A. Meyers, John Wiley & Sons, West Sussex, **2000**, vol. 1, pp. 102-131.

[3] P. R. Griffiths and J. A. de Haseth, *Fourier Transform Infrared Spectrometry*, Wiley-Interscience, 2nd edn., New York, **2007**.

[4] E. R. Malinowski, *Factor Analysis in Chemistry*, Wiley-Interscience, New York, 3rd edn., **2002**.

[5] R. G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, Wiley, West Sussex, **2003**.

[6] H. Mark and J. Workman, Jr., *Chemometrics in Spectroscopy*, Academic Press, London, **2007**.

[7] J. D. Faires and R. Burden, *Numerical Methods*, Brooks/Cole Publishing Co., Pacific Grove, 2nd edn., **1998**.

[8] B. K. Lavine and C. E. Davidson, in *Practical Guide to Chemometrics*, ed. P. Gemperline, CRC Press, Boca Raton, 2nd edn., **2006**, ch. 9, pp. 347-351.

[9] E. Widjaja, C. Z. Li, W. Chew, M. Garland, *Anal. Chem.*, **2003**, *75*, pp. 4499-4507.

[10] J. T. Scheick, *Linear Algebra with Applications*, McGraw-Hill, Singapore, **1997**.

---

[11] W. Chew, E. Widjaja, M. Garland, *Organom.*, **2002**, *21,* pp. 1982-1990.

[12] B. H. Kee, W.-S. Sim, W. Chew, *Anal. Chim. Acta*, **2006**, *571*, pp. 113-120.

[13] C. Z. Li, E. Widjaja, W. Chew, M. Garland, *Angew. Chem. Int. Ed.*, **2002**, *41,* pp. 3786-3789.

[14] C. Z. Li, E. Widjaja, M. Garland, *J. Am. Chem. Soc.*, **2003**, *125,* pp. 5540-5548.

[15] W. Chew, Ph.D. Thesis, National University of Singapore, **2003**.