

Multivariate Analysis in Infrared Spectroscopy

Paul DUMAS SOLEIL Synchrotron, France paul.dumas@synchrotron-soleil.fr

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Synchrotron IR data and « « basic »analysis





Synchrotron IR data and « « basic »analysis



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Microscopes are commercial and softwares are proprietary

Essentially used at SR-IR beamlines = OPUS (Bruker) OMNIC(Thermo Fischer)

Since data analysis is time consuming, users request the software for post analysis when back to their lab or institutes.

 Facing difficulties for providing the softwares to users

Users request beamtimes at different beamlines, and ask for merging data on equivalent samples, from different softwares

- Facing difficulties for exchanging format

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010





Some developers have made possible IR data analysis

PyMCA (A. Solé)



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Some developers have made possible IR data analysis

Axis2000 (A. Hitchcock, C. Jacobsen)



Not as friendly-user as the proprietary ones

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Statistical analysis on spectra and images





Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Typical example



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010





LEIL However, single cell spectra are often affected by physical phenomena

Analyst



The home of high impact research in analytical, bioanalytical and detection science.

Subscribers

O HTML article

Free access

Search for citing articles

Paper

Article citation: Paul Bassan, Analyst, 2009, DOI: 10.1039/b904808a

Resonant Mie scattering in infrared spectroscopy of biological materials – understanding the 'dispersion artefact'

Paul Bassan, Hugh J. Byrne, Franck Bonnier, Joe Lee, Paul Dumas and Peter Gardner

Infrared spectroscopic cytology is potentially a powerful clinical tool. However, in order for it to be successful, practitioners must be able to extract reliably a pure absorption spectrum from a measured spectrum that often contains many confounding factors. The most intractable problem to date is the, so called, dispersion artefact which most prominently manifests itself as a sharp decrease in absorbance on the high wavenumber side of the amide I band in the measured spectrum, exhibiting a derivative-like line shape. In this paper we use synchrotron radiation FTIR micro-spectroscopy to record spectra of mono-dispersed poly(methyl methacrylate) (PMMA) spheres of systematically varying size and demonstrate that the spectral distortions in the data can be understood in terms of *resonant* Mie scattering. A full understanding of this effect will enable us to develop strategies for deconvolving the scattering contribution and recovering the pure absorption spectrum, thus removing one of the last technological barriers to the development of clinical spectroscopic cytology Infrared radiation



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010







There are data treatment published (not yet satisfactory) to account for these effects

Estimating and Correcting Mie Scattering in Synchrotron-Based Microscopic Fourier Transform Infrared Spectra by Extended Multiplicative Signal Correction KOHLER A. ; SULE-SUSO J. ; SOCKALINGUM G. D. ; TOBIN M. ; BAHRAMI F. ; YANG Y. ; PUANKA J. ; DUMAS P. ; COTTE M. ; VAN PITTIUS D. G. ; PARKES G. ; MARTENS H.

Applied spectroscopy 2008, vol. 62, no3, pp. 259-266

Resonant Mie scattering in infrared spectroscopy of biological materials understanding the 'dispersion artefact' BASSAN Paul ; BYRNE Hugh J. ; BONNIER Franck ; LEE Joe ; DUMAS Paul ; GARDNER Peter Analyst 2009, vol. 134, no8, pp. 1586-1593 [



IR data preprocessing







Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



IR data preprocessing

3-Normalization



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Statistical softwares mostly used (all proprietary and expensives)

Unscrambler (<u>www.camo.com</u>) Individual spectra analysis

Cytospec (<u>www.cytospec.com</u>) Image analysis

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Importing data in Unscrambler

New New Design	Ctrl+N	<u>?</u>	?	+		X	1	<u> </u>	<u> 1</u>	÷ [<u>}</u>	0 0		$\leq \leq$	×.	- <u> </u>	A				
Open	Ctrl+O	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
Close		6	7	8	9	10	11	12	13	-14	15	16	17	18	19	20	21	22	23	24	25
Save	Ctrl+S	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
Save As		m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	n
		m		m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	n
Import	•	Unso	ramble	er Resi	ults		Ctrl+	Ū	m	m	m	m	m	m	m	m	m	m	m	m	n
Import 3-D	•	Unso	ramble	er Data	э				m	m	m	m	m	m	m	m	m	m	m	m	n
Export									m	m	m	m	m	m	m	m	m	m	m	m	n
Duplicate	•	050	/dld						m	m	m	m	m	m	m	m	m	m	m	m	n
Convert Vector to Data Table		Unse	ramble	er ASC	II				m	m	m	m	m	m	m	m	m	m	m	m	r
Print	Chrlad	ASC	II						m	m	m	m	m	m	m	m	m	m	m	m	r
Print Preview	Cuttr	Lotu	5						m	m	m	m	m	m	m	m	m	m	m	m	r
Print Lab Report		Exce	s						m	m	m	m	m	m	m	m	m	m	m	m	r
Print Setup									m	m	m	m	m	m	m	m	m	m	m	m	r
·····		. Mati	ab						m	m	m	m	m	m	m	m	m	m	m	m	r
Lock		NSA	s						m	m	m	m	m	m	m	m	m	m	m	m	r
Properties		Data	abase						m	m	m	m	m	m	m	m	m	m	m	m	r
System Setup									m	m	m	m	m	m	m	m	m	m	m	m	r
1 All spectra ND_CIRR	[d]	JCAI	MP-DX.	•••					m	m	m	m	m	m	m	m	m	m	m	m	r
2 D:\Paul\\CHC_PP_VCL1	[r]	Trac	ker						m	m	m	m	m	m	m	m	m	m	m	m	r
3 All spectra KO WT	[d]	APC							m	m	m	m	m	m	m	m	m	m	m	m	r
4 PCA-ND_CIRR_1475-1710	[r]	GRA	MD				1		m	m	m	m	m	m	m	m	m	m	m	m	r
Evit	ALLE4		CDF	9. Cnar	-trOn				m	m	m	m	m	m	m	m	m	m	m	m	r
	ALTE4	LLA. Todic		a oher	uron.				m	m	m	m	m	m	m	m	m	m	m	m	r
23 m m m	m	BEE	3				1		m	m	m	m	m	m	m	m	m	m	m	m	r
24 m m m	m	OPL	5				/		m	m	m	m	m	m	m	m	m	m	m	m	r
25 m m m	m r	Omr	ic						m	m	m	m	m	m	m	m	m	m	m	m	r
		VAR	IAN																		

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



PCA Output



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



PCA additional features



Selecting spectra inside clusters



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Average spectra of selected points



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010

SYNCHROTRON



What's missing in Unscrambler?

Extracting all spectra from a 2D recording

PCA of selected frequency (energy region) region and of two non-adjacent frequency (energy) regions

✓ Projecting the scores into the image of re-assembled spectra

Beta version of integrated imaging in June 2010

Weak points:

Expensive (between 15k€ to 35k€)

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010





Hyperspectral imaging with Cytospec (Mathlab based-)





Importing data in Cytospec



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Univariate imaging



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



Pre-processing options









SYNCHROTRON



HCA imaging options



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



HCA : cluster imaging

oft PowerPoint CytoSpec v. 1.4.02, bu	ld 00326b — Software for Hyperspectral Imaging,Paul Dumas	
	aging Multivariate Imaging Tools File Info Help	
S2KO-L2C5F1.map [## HCA imaging [memory setting: speed]		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
	dendrogram	12
distance method: D-Values	7000	11
use shortcut reduced HCA calc load save	6000	
Hierarchical clustering	5000	6
cluster method: Ward's algorithm	4000	4
calc load save	3000	2
Cluster imaging number of clusters:	2000	Hierarchical cluster map
imagespectradendro		
cancel	Two clusters	9
4000 3500	3000 2500 2000 1500 1000	7
coordinat	es, x: 10 y: 10 show processed spectrum:	6 5 4 3
help	disp options set color chem image quit	
O O O 4562 kB		S:0 Z: Mt 0

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



HCA : cluster imaging



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



HCA : cluster imaging

crosoft PowerPoint W CytoSpec v. 1.4.02,	build 00326b - Software for Hyperspectral Imaging, Paul Dumas	
		4 15 18
HCA imaging [memory setting: speed]		10 10
Distance matrix calculation	dendrogram 12	
distance method: D-Values	7000	
use shortcut reduced HCA calc load save	6000	
Hierarchical clustering		
cluster method: VVard's algorithm	4000	
calc load save	3000	
Cluster imaging number of clusters:	2000 - 14 Hierarchical cluster map	
image spectra dendro.		
	Four clustors	
4000 3500		
	CYTOSPEC 5	
coordi	nates, x: 10 y: 10 show processed spectrum: 3	
r → Bormes autor	disp options set color chem image quit	4 15 16
💿 💿 💿 4562 kB	S: 👩 Z: 🖱 M:	۳ 0 //



HCA : cluster imaging Average spectra of each cluster



Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010



IR

STXM

Where are the needs?

✓ Spectra treatment and imaging from nonproprietary software

Routine all exists, needs to reassemble them

Experiment type ?

« Specific» Menu

X-Fluo Diffraction UV fluorescence

Mathematical treatment package

✓ Statistical analysis of spectra and images , with a common approach for all micro-spectroscopic analytical tools

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010

P. Dumas

Others...



Where are the needs?

✓ Merging several spectral range on a same spectra



 Combining with other approaches on the same sample
resolution

- probing depth
- image alignment

Workshop « HDF5 as hyperspectral data analysis format » ESRF January 11-13, 2010





In summary

Data acquisition is usually achieved using a proprietary software:

- * they are all very good but...
- * difficulties in providing a copy to users
- * exchange data format difficult,

Statistical approaches do exist, most of the time proprietary.But niether one is fully satisfactory.. Mathlab routines, IDL... PyMCA are available...

Users demand for software practice after their experiment is high, and there is no easy way to fulfil their request.