IUVSTA 59

8th July, 2010

Final Report for 59th IUVSTA Workshop – Surface Chemical Analysis – Improving Data Interpretation by Multivariate and Informatics Techniques

Chair: Prof Bonnie Tyler, University of West Indies, Trinidad
Co-chair: Joanna Lee, National Physical Laboratory, UK
Dates: 11-16th April, 2010
Location: Salybia Nature Resort, Trinidad and Tobago
Website: http://www.iuvsta59.org

Introduction

The 59th IUVSTA workshop to advance the analysis and interpretation of surface chemical data was held in Trinidad on 11-16th April 2010. Data interpretation has become perhaps the most important barrier limiting the application and impact of modern surface analysis techniques. State of the art, high-throughput instruments for surface chemical analysis produce overwhelming amounts of data. Improvements in instrument speed, mass or energy resolution, spatial resolution and depth profiling have added not only to the amount of data which can be collected but also to the complexity of the data. Unfortunately, advances in instrumentation have often not been matched with advances in data analysis capability. Consequently, data analysis, rather than the instrument time, has become the major bottleneck for analysts. Furthermore, the complexity of these data is a barrier for the wider uptake of surface analysis techniques particularly in the life sciences.

The workshop focused on improvements in data interpretation for surface chemical analysis using multivariate and information theory based techniques. Advances in these methods have happened rather independently between SIMS and XPS, two of the principal surface chemical analytical techniques. Multivariate methods are also widely used in Analytical Chemistry. The workshop provided a unique forum to bring together international experts from these communities. There is a plethora of different methods for analysts to choose from ranging from the well-known Principal Component Analysis and Multivariate Curve Resolution to newer methods such as Maximum Autocorrelation Factors and Mutual Information Theory. Each of these techniques have advantages and disadvantages, are affected by data pre-processing methods, and are suited to different applications such as image analysis, spectrum and profile interpretation, classification, prediction and correlation with other analytical techniques. The workshop greatly facilitated the understanding and validity of both the well-known and emerging methods. The workshop was particularly timely since modern PCs have the high performance to process data on a minutes timescale (this is now being challenged by 3D/4D datasets!) and developing algorithms are beginning to provide the firm conclusions that practical analysts require.

Participants

The workshop brought together 32 scientists from 8 countries including representatives from industry, academia and national laboratories. The attendees list is included in Appendix 1.

Invited talks were provided by Dave Castner, Joanna Lee, Vincent Smentkowski, John Walton, Tae Geol Lee, Bonnie Tyler, Michael Keenan, Willem Windig, Alan Spool, Ian Gilmore, Buddy Ratner, Roy Goodacre, Patrick Bertrand, Satoka Aoyagi, Ron Heeren and Alex Henderson. The participants included a broad mixture of expertise, ranging from experts in surface science, chemometricians, industrial analysts, instrument manufacturers, software vendors and analysts. This facilitated a holistic view to data analysis as well as open and friendly discussions which will significantly help to advance the science.

Workshop Program

The workshop opened with a plenary talk by Professor David Castner, University of Washington, entitled "Multivariate Processing of Surface Analysis Data: Past Successes and Future Opportunities". In total 16 invited talks and 10 contributed talks were given. The workshop was organised around 14 themed sessions, covering time-of-flight secondary ion mass spectrometry (TOF-SIMS) and X-ray photoelectron spectroscopy (XPS) raw data, analysis of images, extracting chemical information from spectra, analysis of biological samples, considerations for practical analysis, and the use of expert systems and libraries. Each session began with an invited lecture and one or two contributed talks, followed by an hour or more of extended discussion time.

The in depth discussions, rather than the talks, were the main focus of the meeting. The participants contributed to a list of key issues and questions that was circulated by e-mail prior to the meeting to spark off the debate. All the participants were very candid and open during the intense discussions, and there was a strong agreement that the community must work collaboratively to advance the field. An informal data sharing session was held on the first day of the workshop. This allowed participants the opportunity to try new data analysis algorithms using their data sets and test their algorithms on new data. A summary session was held on the last day of the meeting to summarise the key points raised during the workshop and identify key action items that are needed to move the field forward.

In addition to the ample time allocated for discussion, extended interactions took place throughout the day as the participants stayed and shared meals together at the intimate venue. Several social events, including a night-time excursion to watch leatherback turtles nesting and an evening banquet with a steelpan band, helped to promote an open and friendly atmosphere throughout the workshop.

The full programme is included in Appendix 2.

Summary of Discussion

The stimulating discussions at the workshop covered many important issues, including subtleties in the choice of data preprocessing and multivariate methods, the effects of biological sample preparation, methods for extracting relevant chemical information from the data dominated by noise and low-information fragments, improving spectral or image contrast between features, and presenting, reporting, interpreting and validating multivariate

analysis (MVA) results. Some of the issues discussed are often encountered by analysts but rarely examined in detail, including peak selection, effects of mass resolution and mass accuracy, computational limitations and the nature and definition of 'noise' in surface chemical data. Several key overarching issues, which were highlighted throughout the week, are listed below. A number of action items were proposed based on these issues.

1. *Guidance for MVA:* There is no one method ideally suited to solving every problem and the best multivariate method depends both on the nature of the data (e.g. Is it sparse? Is it simple spatially or spectrally? Is it noisy?) and the aim of the analyses. Appropriate preprocessing, including but not limited to dead time correction, calibration, peak selection, binning, normalization, and scaling, are crucial for extracting useful information and avoiding potential bias or artefacts. Simple guidelines are needed to help analysts, especially new MVA users, select the methods best suited to their research questions. In addition, the best data analysis method cannot compensate for bad data or misleading interpretation. The community needs to establish and implement best practices for procedures ranging from instrument operation to color schemes for published images. It is recognized that MVA should not be treated as a black box approach, and a good understanding of the chemistry of the samples, proper design of experiment (e.g. suitable replicates) and validation of MVA results are important before drawing conclusions.

   Proposed action: Completion of an ISO technical report providing guidelines and, when possible, best practices for MVA in the surface science community. This would assist researchers in selecting the most appropriate multivariate methods, preprocessing steps, validation tools and interpretation procedures for their work. Creation of a workshop blog for continuation of discussions on best practices.

2. *Software requirements:* There is a critical need for a simple interface between MVA programs and instrument manufacturer software. Researchers need to know the details of how data is processed in the instrument software and they need straightforward access to the raw, unmanipulated data for MVA purposes. Analysis routines embedded in instrument software are often treated as a black box. Proprietary formats and proprietary data reduction methods make it very difficult to do the accurate inter-laboratory and inter-technique comparisons that are essential to good science. A fast, easy way is needed to export data to an open format suitable for use in 3rd party software or for developing custom analysis routines. This is currently the largest barrier to wider use of advanced data analysis methods. Many participants expressed an interest in importing results like peak lists, spectra and regions of interest generated in MVA routines back into the instrument software for further use. In addition, computation problems need to be overcome with extremely large data sets, such as 3D depth profiles – the use of a sparse matrix is one approach explored at the workshop but new software/hardware developments are also needed.

   Proposed action: A formal request to instrument manufacturers to support a single, complete, non-proprietary format for data exchange along the lines of the standard being developed by the MALDI community (imzML) or the existing VAMAS standard data transfer format (ISO 14976).

3. *Tutorial:* There is a need in the surface science community for better training and mentoring in the use of advanced data analysis methods, which have a steep learning curve. For example, nearly everyone at the workshop was introduced to the use of MVA in surface science by a colleague who guided them through their first analyses. Practical hands-on workshops at conferences are essential for increasing uptake especially for industrial analyses.

Proposed action: Creation of a MVA website in cooperation with NESAC/Bio, which will serve as a community resource to facilitate data and software exchange and provide a repository for user guides, tutorials, software code, worked examples, FAQs and bibliographical resources. The website is now available at http://mvsa.nb.uw.edu/.

4. _Spectra library / database for identification:_ The participants identified the need for an updated spectral library/database with advanced spectral searching, for example fingerprint matching, autocorrelation of peaks or fragmentation search and better link-in between MVA methods and spectral libraries. Existing spectral library needs to be updated with new primary ions and energies. The range of materials in the library needs to be expanded especially for biological analysis. There is a need for the community to learn from other mass spectrometry databases. It is recognised that better mass resolving power and accuracies will be essential for proper identification especially for biological materials.

Proposed action_:_ Further study and review on fragmentation of biological molecules in SIMS is needed alongside the development of a community database. A number of options were discussed at the workshop and will be explored.

Conclusion

The workshop was a great success and many new links were forged with participants sharing ideas of future collaborative work. The community worked together to identify a number of action items which are now being taken forward. The workshop was very successful in moving the field forward and will be remembered by all for a long time.

Acknowledgements

Financial Statement

Appendix 1 – Workshop Attendees List

Last name	First name	Affiliation	Country
Abel	Marie Laure	Surrey University	UK
Alexander	Morgan	Nottingham University	UK
Aoyagi	Satoka	Shimane University	Japan
Artyushkova	Kateryna	University of New Mexico	USA
Bertrand	Patrick	UCL Catholic University of Louvain	Belgium
Castner	Dave	Washington University	USA
Conard	Thierry	IMEC	Belgium
Fischer	Anne	Strategic Analysis, Inc.	USA
Fletcher	John	Manchester University	UK
Gilmore	Ian	NPL	UK
Goodacre	Roy	Manchester University	UK
Graham	Dan	University of Washington	USA
Heeren	Ron	FOM Institute for Atomic and Molecular Physics	Netherlands
Heichler	Winfried	SPECS GmbH	Germany
Henderson	Alex	Manchester University	UK
Keenan	Michael	Independent Consultant	USA
Lea	Scott	Pacific Northwest National Laboratory	USA
Lee	Joanna	NPL	UK
Lee	Tae Geol	KRISS	Korea
Moffitt	Chris	Kratos Analytical	USA
Ohlhausen	Tony	Sandia National Laboratories	USA
Peterson	Richard	University of West Indies	Trinidad
Ratner	Buddy	University of Washington	USA
Sano	Naoko	Surrey University	UK
Smentkowski	Vincent	General Electric Global Research	USA
Spool	Alan	Hitachi Global Storage Technologies Inc.	USA
Tyler	Bonnie	University of West Indies	Trinidad
Walker	Amy	University of Texas at Dallas	USA
Walton	John	Manchester University	UK
Windig	Willem	Eigenvector Research Inc.	USA
Wolstenholme	John	Thermo Fisher Scientific	UK
Yang	Jing	Nottingham University	UK

Appendix 2 – Workshop Programme

Monday 12 April

Opening Session (Discussion leader: Patrick Bertrand)
8:30 am	**Welcome:** Bonnie Tyler
8:45 am	**Invited:** Dave Castner
	"Multivariate processing of surface analysis data: past successes and future opportunities"
9:15 am	**Discussion**
10:10 am	**Tea / Coffee Break**

TOF-SIMS (Discussion leader: Tae Geol Lee)
10:30 am	**Invited: Joanna Lee**
	"Chemometrics in TOF-SIMS: benefits and potential pitfalls"
11:00 am	**Dan Graham**
	"The raw facts about raw data"
11:15 am	**Richard Peterson**
	"Particle Analysis: The importance of data preprocessing"
11:30 am	**Discussion**
12:30 pm	**Lunch**

Practical Considerations (Discussion leader: Alan Spool)
1:30 pm	**Invited: Vincent Smentkowski**
	"Complete analysis of raw data files in a research and development analytical laboratory"
2:00 pm	**Winfried Heichler**
	"In situ characterisation of thin film solar cells and lithium ion batteries with PES and the related quantification issues"
2:15 pm	**Kateryna Artyushkova**
	"Evaluation of structure-to-property relationships in fuel cell durability performance by multivariate analysis"
2:30 pm	**Discussion**
3:30 pm	**Tea / Coffee Break**
3:50 pm	**Informal Discussion and Data Sharing**
7:00 pm	**Dinner**

Tuesday 13 April

XPS Imaging (Discussion leader: Scott Lea)
8:30 am	**Invited: John Walton**
	"Chemical and nano-structural microscopy using XPS"
9:00 am	**Chris Moffitt**
	"Information implications of statistics in imaging parameter choices"
9:15 am	**John Wolstenholme**
	"Practical implementation of multi-variant analysis techniques for XPS"
9:30 am	**Discussion**
10:30 am	**Tea / Coffee Break**

Imaging I (Discussion leader: Amy Walker)
10:50 am	**Invited: Tae Geol Lee**
	"Multivariate Analysis of TOF-SIMS Data Obtained from Biological Samples"
11:20 am	**Tony Ohlhausen**
	"The application of multivariate analysis to imaging NEXAFS"
11:35 am	**Discussion**
12:30 pm	**Lunch**

Imaging II (Discussion leader: Kateryna Artyushkova)
1:30 pm	**Invited: Bonnie Tyler**
	"Analysis of Image Series and 3D Image Stacks"
2:00 pm	**Naoko Sano**
	"The Transfer of Organics onto Glass Studied by TOF-SIMS"
2:15 pm	**Discussion**
3:15 pm	**Tea / Coffee Break**

Imaging III (Discussion leader: Dan Graham)
3:35 pm	**Invited: Michael Keenan**
	"Wavelets in multivariate image analysis: characterizing noise and improving performance"
4:05 pm	**Amy Walker**
	"Optimized analysis of imaging TOF SIMS data"
4:20 pm	**Discussion**
7:00 pm	**Dinner**
8:00 pm	**Excursion: Leatherback Turtle Watching**

Wednesday 14 April

Chemical Information I (Discussion leader: Michael Keenan)
8:30 am	**Invited: Willem Windig**
	"A comparison of different approaches for self-modelling mixture analysis"
9:00 am	**Thierry Conard**
	"On the influence of the fitting algorithm on the depth profiles reconstructed from AR-XPS data"
9:15 am	**Discussion**
10:10 am	**Tea / Coffee Break**

Chemical Information II (Discussion leader: Dave Castner)
10:30 am	**Invited: Alan Spool**
	"Information content in TOF-SIMS"
11:00 am	**Invited: Ian Gilmore**
	"Complementarity of MVA and G-SIMS: building molecules using intrinsic information from the fragmentation process"
11:30 am	**Discussion**
12:30 pm	**Lunch**
1:30 pm	**Excursion: Several Options**
7:00 pm	**Dinner**

Thursday 15 April

Biological Analysis (Discussion leader: Ron Heeren)
8:30 am	**Invited: Buddy Ratner**
	"The Tissue Engineering-SIMS Interface: Analysis of Really Complex Data Sets"
9:00 am	**Jing Yang**
	"TOF-SIMS analysis of fibronectin pre-adsorbed to a polymer microarray revealing the effect of protein and surface chemistry on cellular adhesion"
9:15 am	**Discussion**
10:20 am	**Tea / Coffee Break**

Biological Analysis II (Discussion leader: Buddy Ratner)
10:40 am	**Invited: Roy Goodacre**
	"Knowledge discovery using metabolomics: challenges and opportunities for the analysis of complex biological systems"
11:10 am	**Morgan Alexander**
	"Characterisation of Saccharide Microarrays using Time of Flight Secondary Ion Mass Spectrometry with Multivariate Analysis"
11:25 am	**Discussion**
12:30 pm	**Lunch**

Biological Analysis III (Discussion leader: Roy Goodacre)
1:30 pm	**Invited: Patrick Bertrand**
	"Principal Component Analysis for Interpreting Complex TOF-SIMS Data from Organic Materials"
2:00 pm	**Invited: Satoka Aoyagi**
	"TOF-SIMS spectrum analysis methods for protein evaluation"
2:30 pm	**Discussion**
3:30 pm	**Tea / Coffee Break**

Biological Analysis IV (Discussion leader: Morgan Alexander)
3:50 pm	**Invited: Ron Heeren**
	"Multimodal molecular imaging of tissue surfaces: MALDI and SIMS combined"
4:20 pm	**John Fletcher**
	"Adventures in multivariate analysis of TOF-SIMS data from biological samples"
4:35 pm	**Discussion**
7:00 pm	**Workshop Banquet**

Friday 16 April

Expert Systems (Discussion leader: Vincent Smentkowski)
8:30 am	**Invited: Alex Henderson**
	"Have I seen this before? Spectral Comparison in Static SIMS"
9:00 am	**Scott Lea**
	"A near real-time data analysis package to enhance the information extracted from XPS spectra"
9:15 am	**Discussion**
10:20 am	**Tea / Coffee Break**

<u>Summary Session</u> (Discussion leader: Ian Gilmore)
10:40 am   **Discussion**
                   "Towards best practice for data analysis and interpretation"
12:15 pm   **Close: Bonnie Tyler**
12:30 pm   **Lunch**