# MSc/PhD research topics for advances in:

### ⊞ *XRF quantification methodology*

### ⊞ *MA-XRF imaging analysis*

### ⊞ *XRF spectra deconvolution*

*all topics requiring use/development of simulation programs and*

*programming skills*

X-ray Fluorescence (XRF) is an elemental analysis technique characterized by its exceptional versatility and simplicity in analyzing almost any kind of material (solid, liquid, bulk, thin filmσ, powder, particles, etc.), providing rapid qualitative and quantitative results. In XRF analysis an incident X-ray beam generates inner-shell (K-, L-, M) ionization of sample atoms and subsequently the emission of element and energy specific X-rays (the so-called characteristic X-rays).  With the use of energy dispersive X-ray spectrometers, the identification of the sample constituent elements (from atomic number Z=11 to Z=92) within a broad dynamic range of concentrations (from µg/g to wt.%) is simultaneous (qualitative analysis), whereas the recorded intensity of characteristic X-rays is used to deduce the respective analyte concentration (quantitative analysis). The ultimate goal in applying the XRF technique in various applications is to quantify the sample elemental composition and to generate element specific concentration maps (for 2D heterogenous samples) [1].

A key and distinctive feature of the XRF technique, amongst other analytical spectroscopies, is the existence of a mathematical formulation that converts the measured characteristic X-ray peak areas to element's weight percent. The relevant formalism developed back in 60's, was further refined and improved to account for various second order enhancement phenomena of the fluorescence intensity [2-7]. The methodology to describe the analyte fluorescence intensity as the result of all the contributing fundamental physics interactions between the exciting X-ray beam and analyte atoms, is generally referred to as the Fundamental Parameter (FP) approach in XRF analysis [8-10]. In FP-based methods, the relationship between XRF intensities and the unknown concentrations of a multi-element sample is expressed through a set of well-defined nonlinear equations which can be solved iteratively so as the calculated spectrum matches the experimental one. This methodology sets a robust framework for XRF quantitative analysis, ideally without the need of any reference sample. However, it pre-requires good knowledge of measurement and instrument geometrical parameters, appropriate modelling of all primary and multiple order physics interactions governing the generation of the primary X-ray tube spectrum, the transport of exciting and outgoing X-rays within the sample and in the semiconductor detector, but also generic information about the sample type.

The implementation of the FP methodology in XRF analytical applications includes several steps, starting from a reliable spectrum deconvolution procedure, communication with updated and evaluated databases of X-ray FPs and setting the required measurement, instrument and sample descriptive parameters using a friendly user-interface. So far, these basic requirements have been fulfilled to a great extent by the PyMca code developed by the software group of the European Synchrotron Radiation Facility (ESRF, http://pymca.sourceforge.net/). The PyMca Toolkit is a collection of Python tools for visualization and analysis of energy-dispersive XRF data. It builds its graphic interface and plotting routines on top of the C++ library Qt through its Python

binding PyQt or PySide. The data analysis routines can be used independently of any graphical interface and the code is platform independent (Linux, Windows, MacOS X, Solaris …) and freely available for non-commercial use.

The PyMca software initially developed to process XRF data acquired at synchrotron facilities (2007) with emphasis to imaging (qualitative) analysis. However, it further developed to encompass X-ray tube spectra, compositional analysis of multi-layered samples including multiple order fluorescence enhancement mechanisms, but also Monte-Carlo based modelling.

After almost of fifteen years of practical experience, certain shortcomings have been identified by the scientific community in view of the need to improve data analysis of complex spectra, to process more effectively large datasets generated by modern XRF scanners and to support compositional analysis of field portable XRF spectrometers. For example, specific issues such as the need to cope more effectively with the lack of freedom to define the fitting parameters of individual peaks, the lack of flexibility in the background subtraction, the non-linear channel-energy calibration effects observed sometimes in low energy X-ray detection and the necessity to include in some cases dominant multi-vacancy X-ray "shake" satellite lines have motivated the development of new software XRF analysis packages [1, 2].

The XRF laboratory of the Institute of Nuclear and Particle Physics at NCSR "Demokritos" has accumulated long lasting experience in processing quantitatively XRF spectra from a great variety of samples acquired by either laboratory or portable spectrometers equipped with different X-ray sources and energy dispersive detection systems, including data processed by XRF set-ups operated at synchrotron sources. Overall, based on the accumulated experience specific needs and projects have been identified to upgrade PyMca functionality by developing dedicated software tools aiming to improve the accuracy of the FP-based methodology, and to allow deeper and more comprehensive quantitative evaluation of XRF spectra:

1) **Comparative evaluation of MA-XRF imaging data using Machine learning and Multivariate methods**. In the case of macroscopic XRF (MA-XRF) analysis, tens of even hundreds of thousands of XRF spectra are generated by producing element specific distribution maps over large surfaces. In order to extract meaningful, interpretable and quantifiable information, the dimensionality of the raw data should be properly reduced. Multivariate techniques, such as Principle Component analysis followed by k-means clustering have been already applied in numerous of applications. Additional techniques, such as the t-Distributed Stochastic Neighbour Embedding technique (t-SNE) [3, 4] or more recently the Application of Uniform Manifold Approximation and Projection (UMAP) [5] seem very promising. Moreover, t-SNE has demonstrated that can generate high statistics spectra that would contain enough information for qualitative and/or quantitative analysis of major, minor and even trace elements even from low-statistics individual pixel spectra [3]. Apart the chemometric multivariate methods, unsupervised machine learning (ML) algorithms (SOM) based on neural networks, have also been proven very effective in reducing a large spectral image dataset to a handful of distinct clusters that share similar spectra [6]. Due to the broad range of MA-XRF applications and the subsequent necessity to extract from large datasets the hidden correlations, a systematic evaluation of ML and chemometric methods for MA-XRF imaging analysis must be performed to recommend optimum methodologies and reveal their limitations.

2) **XRF quantification using Machine Learning (ML) training approaches**. The proposed research aims to offer a novel contribution in the XRF quantitation problem based on

the development, application and evaluation of Machine Learning (ML) training approaches. For example, the GAs generally have been proven to be a useful tool in solving real problems in many scientific fields presenting good search accuracy, but are flawed by poor precision [7-8]. Being inspired by the evolutionary biology, start with a stochastic population of solutions and sequentially evolve next generations (through inheritance, mutation, selection and crossover), based only on the better set of solutions as indicated by the given fitness function until a convergence criterion is fulfilled. The first use and evaluation of Genetic Algorithms as an optimization procedure was proposed for the quantitative XRF analysis of single-layer and multiple-layer films with very encouraging results [9-10]. This application was motivated by the GAs inherent insensitivity in initial conditions and the robustness of the FP approach as theoretical XRF framework. There are also other applications which have reported the use Artificial Neural Network (ANN) for fast classification of samples based on their XRF spectroscopic pattern [11-13]. From the results of the proposed research the merits, requirements and limitations imposed by the adopted ML training algorithms will be evaluated and an assessment will be carried out to which extent they can contribute in the standardization and interoperability of different commercial or custom developed XRF equipment.

3) **Description and Evaluation of RRS spectra:** The Resonant Raman Scattering (RRS) is an inelastic X-ray scattering atomic process that occurs and significantly enhanced as the incident x-ray radiation approaches from lower energies the binding energy of an inner shell of the target atom [14-26]. This scattering process proceeds through an intermediate state with a virtual hole in the respective inner shell, whereas in the final state an emitted RRS photon and a photoelectron from a higher inner shell share together the available energy. The RRS emission energy dispersive spectrum includes distinct and predictable features for a specific element excited by an incident monoenergetic X-ray beam, similarly with the emission spectrum produced during the photo-ionization process. PyMca does not account for the presence of RRS peaks in XRF spectra, although there are cases when RRS features manifest and humper the analysis and quantification of important trace elements.

4) **Simultaneous quantification of multiple XRF datasets**: In field portable XRF analysis using for example handheld XRF spectrometers, the optimized analytical protocols recommend multiple measurements of the same sample by different operational and hardware conditions to attain the maximum possible analytical sensitivity. Thus, in such cases, multiple spectra need to be evaluated and processed together integrated under the same FP based software platform. Similar needs exist in tunable energy synchrotron excited XRF analysis, or in benchtop XRF spectrometers using Cartesian geometry with a set of secondary targets. PyMca can deal with multiple spectra only sequentially and not in simultaneous mode, thus requiring from the user to apply manually time consuming iterative runs. This problem becomes more pronounced and may even affect the consistency of quantitative analysis when the matrix elements are analyzed in more than one analysis condition.

5) **Study and parametrization of electron based secondary ionization processes in XRF analysis**: Secondary ionization processes following photon induced primary ionizations in matter may contribute significantly to the detected fluorescence radiation introducing significant errors in XRF quantitative analysis [27-30]. These contributions are totally ignored by quantitative analysis software packages such as the PyMca. in the case of XRF analysis of elements contained within a light matrix, energetic photoelectrons generated

by the primary ionization of light matrix atoms can next ionize locally (in short range from the point of their generation) the analyte low Z atoms, thus enhancing their fluorescence intensity. Also, in the case of low Z elements XRF analysis within a high Z matrix, apart the secondary fluorescence, the Auger electrons ionizations of low Z elements is also expected to provide significant contribution to their fluorescence intensity. Dedicated Monte Carlo estimations of the various contributions induced by electron based secondary ionization processes are needed to be carried out for bulk samples resembling cultural heritage materials and to develop appropriate mathematical framework to account for these phenomena in FP based calculations and if possible to incorporate these enhancement corrections within PyMca software

## References

1. Analysis of Synchrotron-Based X-Ray Fluorescence Measurements from Multiple Detector Modules, Catharina Ziska, Universitat Hamburg Deutsches Elektronen-Synchrotron, A thesis submitted for the degree of Bachelor of Science, 2020

2. An empirical derivation of the X-ray optic transmission profile used in calibrating the Planetary Instrument for X-ray Lithochemistry (PIXL) for Mars 2020, C. M. Heirwegh, W. T. Elam, D. T. Flannery, and A. C. Allwood, Powder Diffraction, 2018, doi:10.1017/S0885715618000416

3. Iva Božičević Mihalić, Stjepko Fazinić, Marko Barac, Andreas Germanos Karydas, Alessandro Migliori, Damir Doračić, Vladan Desnica, Domagoj Mudronja and Dragica Krstićg, Multivariate analysis of PIXE+XRF and PIXE spectral images, Accepted in JAAS

4. Matthias Alfeld, Silvia Pedetti, Philippe Martinez, Philippe Walter, Joint data treatment for Vis–NIR reflectance imaging spectroscopy and XRF imaging acquired in the Theban Necropolis in Egypt by data fusion and t-SNE, Comptes Rendus Physique, Volume 19, Issue 7, November 2018, Pages 625-635

5. Marc Vermeulen, Kate Smith, Katherine Eremin, Georgina Rayner, Marc Walton Application of Uniform Manifold Approximation and Projection (UMAP) in spectral imaging of artworks, Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy Volume 252, 5 May 2021, 119547

6. Sotiria Kogou, Lynn Lee, Golnaz Shahtahmassebi, Haida Liang, A new approach to the interpretation of XRF spectral imaging data using neural networks, XDOI: 10.1002/xrs.3188

7. D.E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning", Addison-Wesley Longman Publishing Co, 1989.
8. A. Hoecker et al., "TMVA - Toolkit for Multivariate Data Analysis", PoS ACAT, 2007.
9. A. D. Dane, P. A. M. Timmermans, H. A. Van Sprang, and L. M. C. Buydens, "A Genetic Algorithm for Model-Free X-ray Fluorescence Analysis of Thin Films", Anal. Chem., vol. 68, no. 14, pp. 2419–2425, 1996.
10. A. D. Dane, H. A. van Sprang, and L. M. C. Buydens, "A Two-Step Approach toward Model-Free X-Ray Fluorescence Analysis of Layered Materials", Anal. Chem., vol. 71, no. 20, pp. 4580–4586, 1999.
11. L. Luo, D. R. Chettle, H. Nie, F. E. Mcneill, and M. Popovic, "Curve fitting using a genetic algorithm for the X-ray fluorescence measurement of lead in bone", vol. 269, no. 2, pp. 325–329, 2006.

12. J. Breen, P. De Souza, G. P. Timms, and R. Ollington, "Onboard assessment of XRF spectra using genetic algorithms for decision making on an autonomous underwater vehicle", Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms, vol. 269, no. 12, pp. 1341–1345, 2011.

13. MI Kaniu, KH Angeyo, AK Mwala, FK Mwangi, "Energy dispersive X-ray fluorescence and scattering assessment of soil quality via partial least squares and artificial neural networks analytical modeling approaches", Talanta, vol. 98, 30, pp. 236-240, 2012.

14. J. Sparks Jr., *Phys. Rev. Let.* **33**, 262 (1974).

15. S. Manninen, *Radiat. Phys. Chem.* **50**, 77 (1997).

16. K. Hämäläinen, D.P. Siddons, J.B. Hastings and L.E. Berman, *Phys. Rev. Lett.* **67**, 2850 (1991).

17. A.G. Karydas, S. Galanopoulos, Ch.Zarkadas, T.Paradellis and N.Kallithrakas-Kontos. *J. Phys.: Condens. Matter* **14**, 12367 (2002).

18. J.J. Leani, H. Sánchez, D. Pérez and C. Pérez. *Anal. Chem.* **85**, 7069 (2013).

19. J.J. Leani, H. Sánchez, M. Valentinuzzi and C. Pérez. *J. Anal. At. Spectrom.* **26**, 378 (2011).

20. P. Eisenberger, P. M. Platzman and H. Winick, *Phys. Rev. Lett.* **36**, 623 (1976).

21. S. Kumar, V. Sharma, D. Mehta and N. Singh. *Nucl. Instr. and Meth. in Phys. Res. B* **264,** 1 (2007).

22. J. Szlachetko, J.-Cl. Dousse, M. Berset, K. Fennane, M. Szlachetko, J. Hoszowska, R. Barrett, M. Pajek and A. Kubala-Kukus, *Phys. Rev. A* **75**, 022512 (2007).

23. M. Müller, B. Beckhoff, G. Ulm and B. Kanngießer. *Phys. Rev. A* **74**, 012702 (2006).

24. H.J. Sánchez, M.C. Valentinuzzi and C. Pérez, *J. Phys. B: At. Mol. Opt. Phys.* **39**, 4317 (2006).

25. H. Czerwinski, F. Smend, D. Schaupp, M. Schumacher, A.H. Millhouse and H. Schenk-Strauss. *Z. Phys. A - Atoms and Nuclei* **322**, 183 (1985).

26. A. G. Karydas, M. Budnar, Z. Smit, Ch. Zarkadas and T. Paradellis, *Nucl. Instr. and Meth. in Phys. Res. B* **189**, 43 (2002).

27. B. Beckhoff et al., X-Ray Spectrom. 2008; 37: 462–465

28. L. Vincze, et al., European Mineralogical Union (EMU) Notes in Mineralogy, Vol. 8 (2010), Chapter 6, 169–237

29. J. E. Fernandez, X-Ray Spectrom. 2013, 42, 189–196

30. K. N. Stoev, "Theoretical calculation of secondary x-ray fluorescence due to Compton electrons", J. Phys. D: Appl. Phys. vol. 25, pp. 131–138, 1992.

**General Bibliography**

- "Handbook of Practical X-Ray Fluorescence Analysis", B. Beckhoff, B. Kanngießer, N. Langhoff, R. Wedell, H. Wolff (Eds), © 2006, First Edition, Published by Springer-Verlag Berlin Heidelberg,
- J. Sherman, "The theoretical derivation of fluorescent X-ray intensities from mixtures" Spectrochim. Acta, vol. 7, no. x, pp. 283–306, 1955.

- T. Shiraiwa and N. Fujino, "Theoretical Calculation of Fluorescent X-Ray Intensities in Fluorescent X-Ray Spectrochemical Analysis", Jpn. Appl. Phys., vol. 5, no. 10, p. 886, 1966.
- D. K. G. De Boer, "Calculation of x-ray fluorescence intensities from bulk and multilayer samples", X-Ray Spectrom., vol. 19, no. 3, pp. 145–154, 1990.
- M. Mantler and N. Kawahara, "How accurate are modern fundamental parameter methods?", *Rigaku J.*, vol. 21, no. 2, pp. 17–25, 2004.
- B. Beckhoff, "Reference-free X-ray spectrometry based on metrology using synchrotron radiation", *J. Anal. At. Spectrom.*, vol. 23, no. 6, p. 845, 2008.