Project Report
Summer Student Program 2009
Hasylab at Desy

Improving of KEMP2 program

Dorota Szepietowska

AGH University of Science and Technology in Krakow

Supervisors: Dr. Gerd Wellenreuther
Dr. Wolfram Meyer-Klaucke

September 2009

1. Introduction

X-ray absorption spectroscopy is an experimental method, which allows to probe the physical and chemical structure of matter in atomic scale. In biological samples XAS is used to find electronic and geometric properties of metal atom. However, in biological XAFS the concentration of metal in the sample is low. It causes that date are usually collected in fluorescence mode by multielement solid-state detector. One scan provides *m* individual spectra, where m is the number of detector elements. If *n* is the number of scans the hole data pool is the *n x m* matrix. The subsets cannot be simply summed, because spectra from each detector element depends on the number of channel. Moreover, X-ray can cause some damages in sample so shape of the spectra can depend on the *n* number. However, to achieve good signal-to-noise ratio scans should be summed. Firstly all scans for one detector element are summed. Then after checking quality of each detector channel spectra are averaged. Using averaged spectrum one can extract XANES and EXAFS for measured sample.

KEMP2 was written to process fluorescence XAS data. It contains two main parts. In the beamline-dependent part program creates file, which contains only energy, I0 and counts for all fluorescence channels. Now program can process files created at three beamlines: the XAFS-beamlines A1, C, E4 and X at DORIS / HASYLAB, Hamburg, Germany, the beamline "Super XAS" at the SLS, Villigen, Switzerland, the DUBBLE-beamline at the ESRF, Grenoble, France. In the generic part all steps of analysis data are included. Program stats with averaging scans for detector elements and automatic selection of good detector channels. Then the time-development of XANES is displayed. Next step is subtracting the pre-edge background and fitting a spline to the data to extract the XAS. Then user has opportunity to remove glitches. In the end the non-phaseshift corrected Fourier-Transform is calculated.

One of my task during Summer Students Program was improvement of KEMP2 source code written in Python and adding part of program, which allows to select detector channels automatically. For first few weeks I worked with my partner Kamila Biernacka. At the same time we worked together on Microscope Images Stitching program.

2. Improving of KEMP2 program

In the beginning I started with writing new functions which allows to get energy setting for processed spectrum. In the function `get_energy_setting` user is asked for edge position, start of pre-edge background region, end of pre-edge background region, start point above edge, start point for XANES extrapolation and end point for XANES extrapolation. If user does not write any values all energy setting are taken from remove.par as a default values. When all energy setting are known, the function finds the indexes of this values in the list which contains energy values. All queries about energy settings use the function `query_float`. All indexes are found by the function `get_index`. This values and indexes are necessary to evaluate data and extract XANES and EXAFS.

Afterward I designed the menu, which occurs when program is started. In menu user can write the path to files, the sample name and choose one of three available data formats: FIO, SLS, SRS. After selection of data format user is asked about energy, I0, and fluorescence channels columns indexes. For each data format default indexes are defined in source code and program can take them if user does not select new indexes.

Next I implemented algorithm which was used in previous version of KEMP and was written in C++. This algorithm can identify the worst spectrum by comparing each subset C_j with its references R_j. The references is the average over all other subsets. First all subsets are normalized to one. Then for each spectrum program creates the difference spectrum D_j:

$$D_j(i) = C_j(i) - R_j(i), \quad i = 1,2,\dots,np, \quad j = 1,2,\dots,ns, \tag{2.1.}$$

where np is the number of data points in each subsets and ns is the number of subsets. This difference spectrum should be flat line if all subsets are the same. But if only outlier appears the deviation score can be calculated as mean square deviation of the difference spectrum points from their average $\bar{D_j}$:

$$\delta_j = (1/np)\sum_{i=1}^{np}[D_j(i) - \bar{D_j}], \tag{2.2.}$$

where:

$$\bar{D_j} = (1/np)\sum_{i=1}^{np} D_j(k), \tag{2.3.}$$

After evaluation of all deviations score, the subset with the highest score is temporary excluded and new value of R_j, D_j, and δ_j are calculated. The iterative process is continued until one subset is left. The threshold is defined as the score calculated in last iteration $\delta(ns-1)$. If in subset there is no outliers the maximum deviation score δ_{max} is near to $\delta(ns-1)$. The criterion for subsets which are included:

$$[\delta_{max}(r) - \delta(ns-1)]/\delta(ns-1) < t_d, \quad r = 1,2,\ldots \qquad (2.4.)$$

where $\delta_{max}(r)$ is the maximum value of the deviation in the *r*th iteration and parameter $t_d < 1$.

After calculation user can see plot of $\delta_{max}(r)/\delta(ns-1)$ for each detector element.

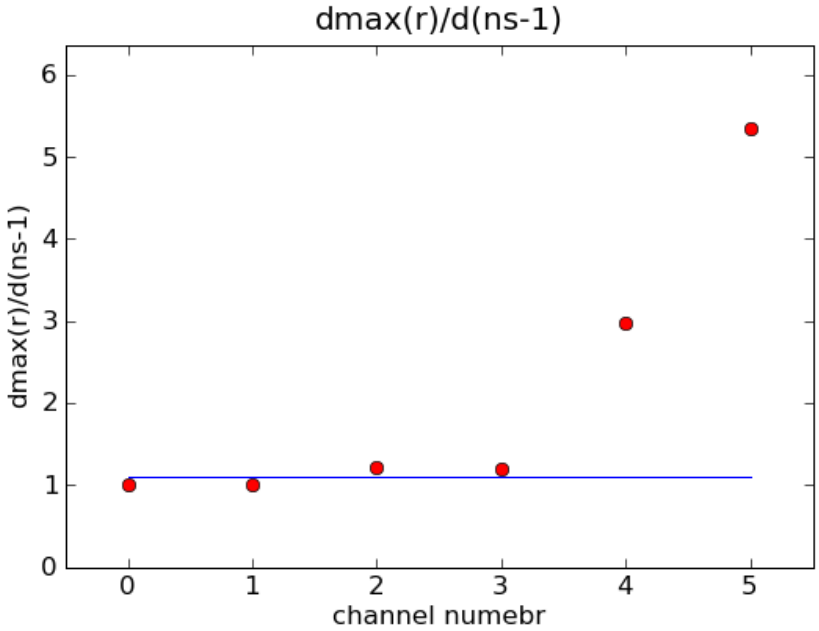

Fig. 1 Value of $\delta_{max}(r)/\delta(ns-1)$ for each detector. Points, which are below the blue line, correspond to the best detector elements.

Then KEMP2 plots overview detector channels. By using different colors program suggests the quality of each detector elements. Next user is asked which detectors elements should be taken to the average spectrum. There are four possibilities: program can take only green spectra, green and yellow spectra, all spectra or user can make a choice of detector elements manually.

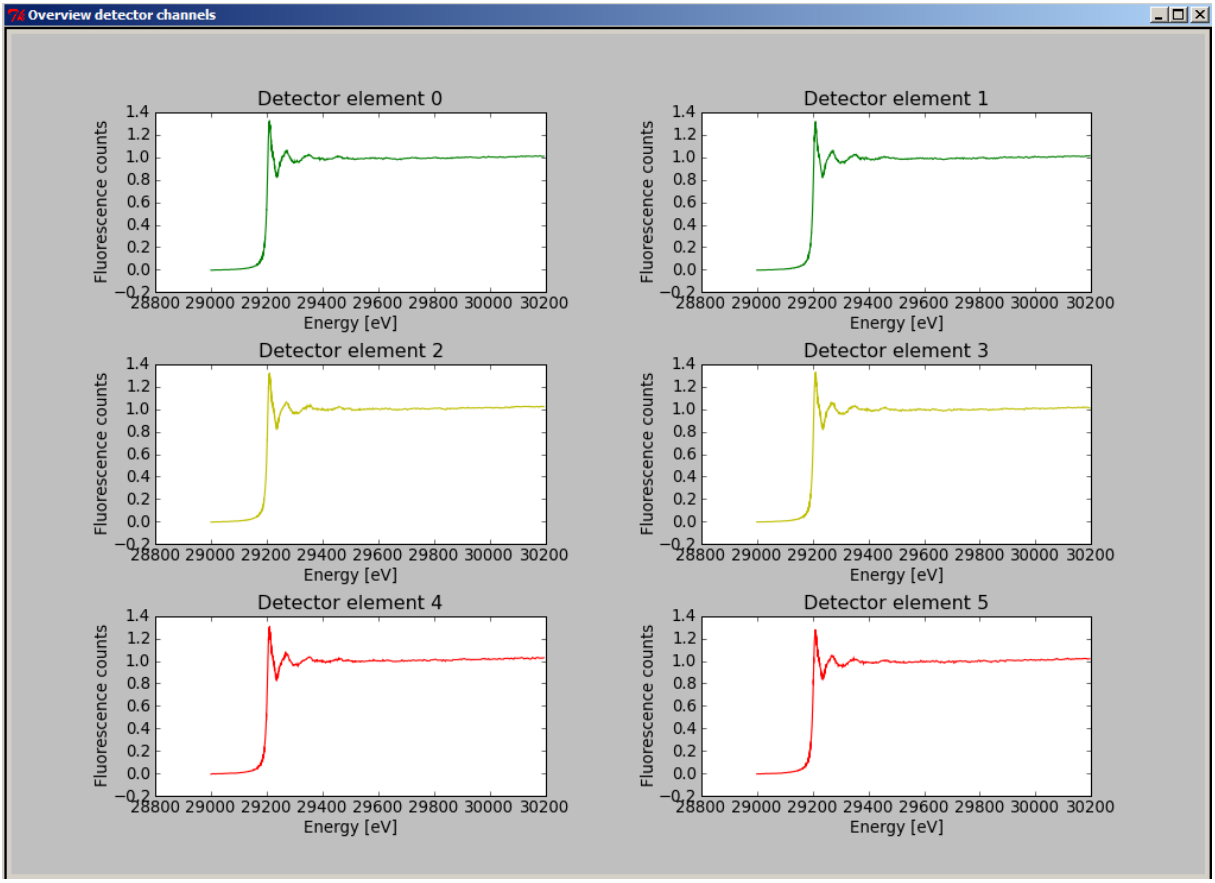

Fig.2 Overview detector channels: green spectra - parameter $t_d < 0.1$, yellow spectra – parameter $t_d < 0.3$, red spectra $t_d > 0.3$.

After detector selection program calculates the averaged spectrum, which is used to extract XANES and EXAFS and to evaluate the non-phaseshift corrected Fourier-Transform.

References:

[1] http://www.embl-hamburg.de/~gwellenr/KEMP2/KEMP2.htm

[2] Malgorzata Korbas, Daniel Fulla Marsa, and Wolfram Meyer-Klaucke, *KEMP: A program script for automated biological x-ray absorption spectroscopy data reduction*, REVIEW OF SCIENTIFIC INSTRUMENTS **77**, 063105 (2006)