



北京同步辐射装置
Beijing Synchrotron Radiation Facility

XANES/EXAFS谱的 PCA分析和分峰方法

安鹏飞

1W1B吸收谱实验站

anpf@ihep.ac.cn



中国科学院高能物理研究所

内容提要

- PCA基本原理
- PCA适用体系
- PCA软件的使用及应用
- 分峰拟合的基本原理
- 分峰拟合适用体系
- 分峰拟合软件的使用及应用举例

北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分析讲习班

武汉·2014



中国科学院高能物理研究所



PCA的基本原理

主元分析（Principal Component Analysis）

- ◆ 本质上是一种数学中的线性代数变换方法
- ◆ 是一种掌握事物主要矛盾的统计分析方法
- ◆ 在图像处理、生物、化学以及控制等各个领域有着广泛的应用
- ◆ 可以分析复杂体系中的吸收谱





北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA的基本原理

形象地理解PCA：这个美女是谁？



究所



PCA的基本原理

基本原理

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ A_{31} & A_{32} & \cdots & A_{3n} \\ \vdots & \vdots & \cdots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} = \begin{bmatrix} E_{11} & E_{12} & \cdots & E_{1n} \\ E_{21} & E_{22} & \cdots & E_{2n} \\ E_{31} & E_{32} & \cdots & E_{3n} \\ \vdots & \vdots & \cdots & \vdots \\ E_{m1} & E_{m2} & \cdots & E_{mn} \end{bmatrix} \cdot \begin{bmatrix} v_{11} & 0 & \cdots & 0 \\ 0 & v_{12} & \cdots & 0 \\ \vdots & \vdots & v_{ij} & \vdots \\ 0 & 0 & \cdots & v_{nm} \end{bmatrix} \cdot \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix}$$

$[A]$

$=$

$[E]$

\cdot

$[V]$

\cdot

$[w]^t$

$M \times N$

$M \times N$

$N \times N$

$N \times N$

吸收谱数据组

组元数组

本征值

权重因子

(在整个数据组的权重) (单个谱的权重重要性)

拆分主元





PCA的基本原理

基本原理

$$\begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \\ \vdots \\ T_m \end{bmatrix} = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \\ E_{31} & E_{32} \\ E_{41} & E_{42} \\ \vdots & \vdots \\ E_{m1} & E_{m2} \end{bmatrix} \cdot \begin{bmatrix} E_{11} & E_{21} & E_{31} & E_{41} & \cdots & E_{m1} \\ E_{12} & E_{22} & E_{32} & E_{42} & \cdots & E_{m2} \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \\ \vdots \\ S_m \end{bmatrix}$$

$$[T] = [E] \cdot [E]^t \cdot [S]$$

变换谱

组元数组

组元数组的转置矩阵

标样谱

目标变换——寻找合适的标样



PCA适用体系

- ✓ 化学反应中间过程的分析
- ✓ 物质相变过程的分析
- ✓ 环境科学中的组分分析

北京同步辐射装置
Beijing Synchrotron Radiation Facility

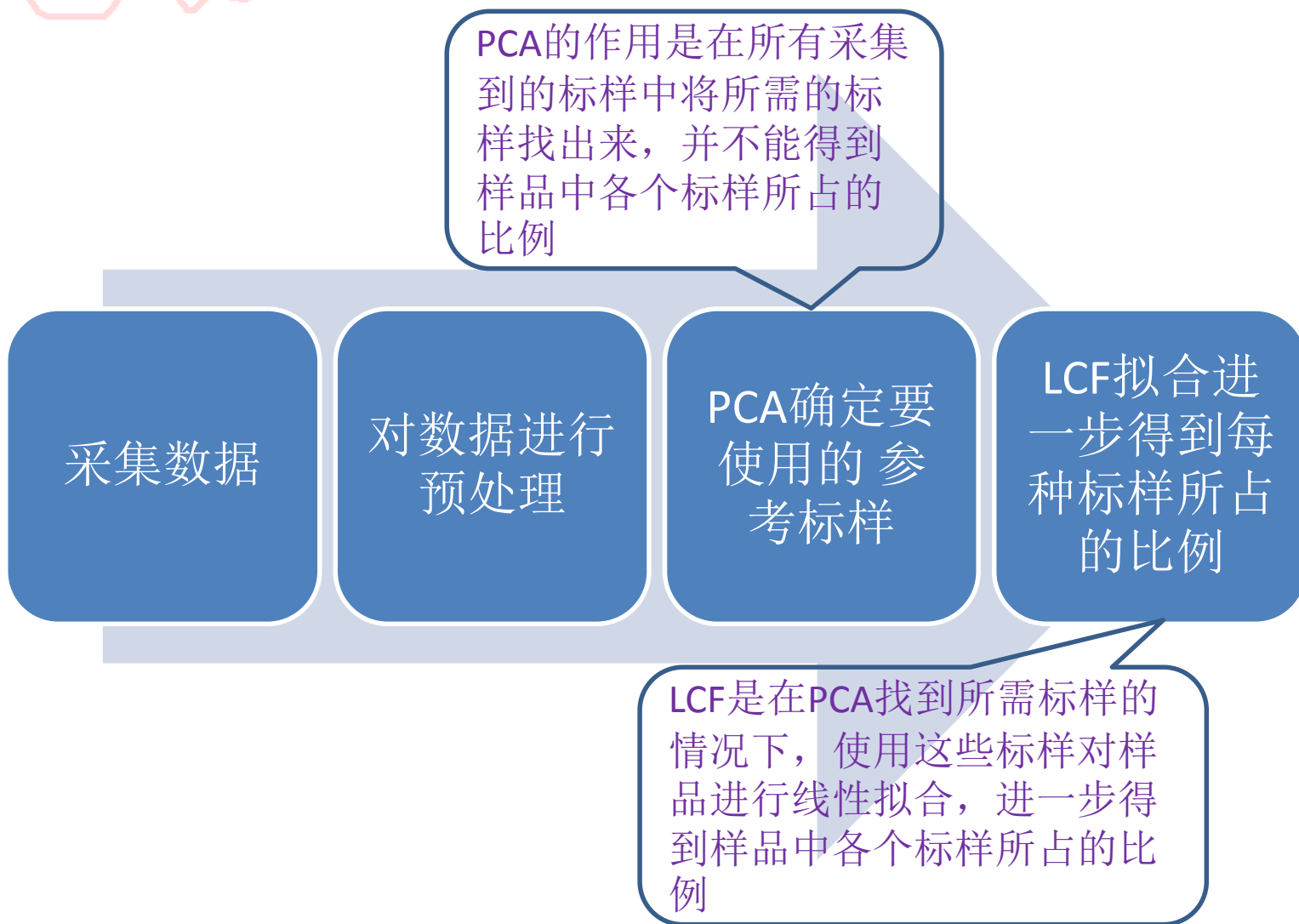
×射线吸收谱学实验
和数据分折讲习班

武汉·2014

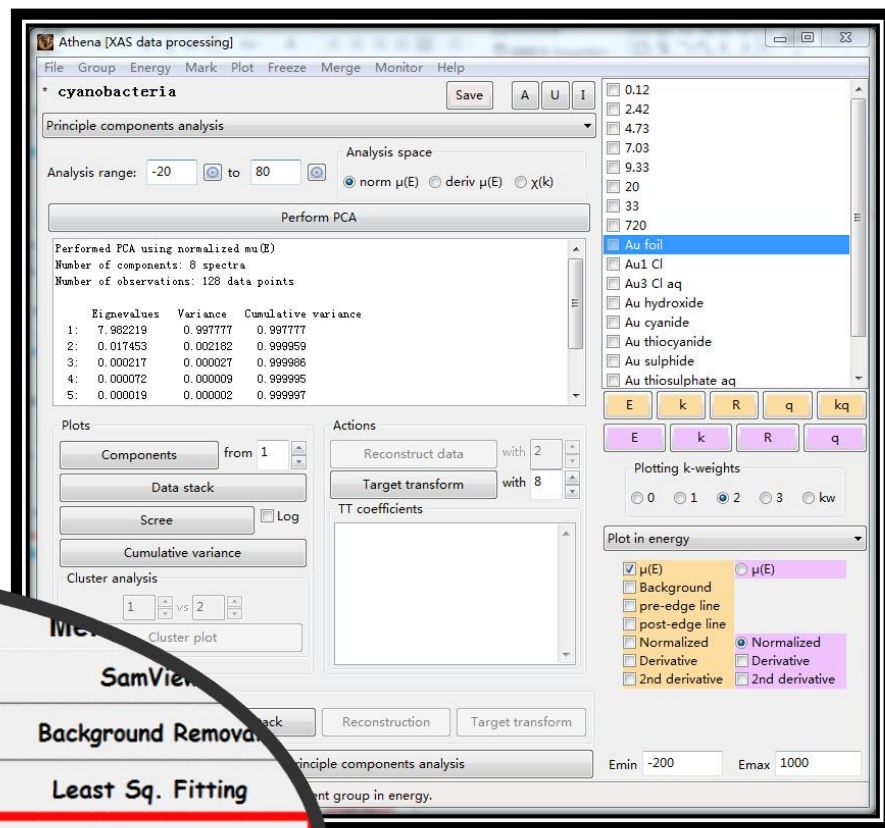
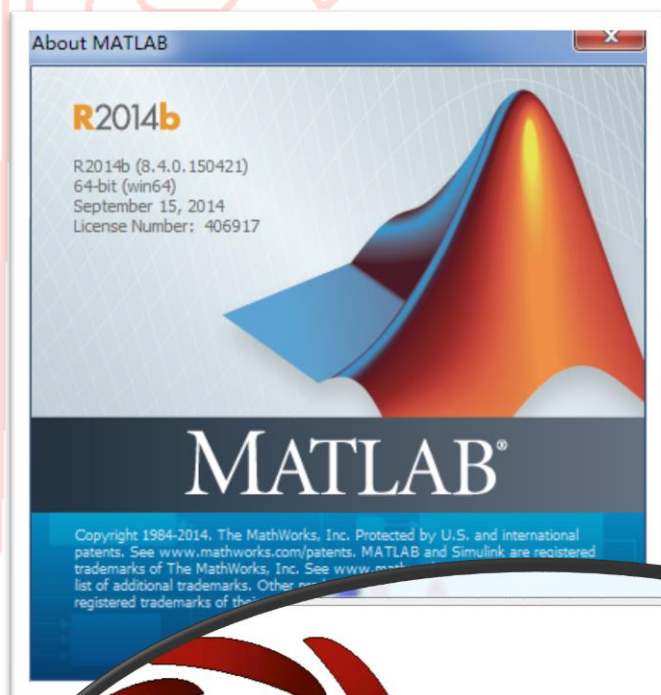


中国科学院高能物理研究所

进行主元分析的一般步骤



PCA软件的使用



SIXpack

Background Removal

Least Sq. Fitting

PC Analysis

FEFF EXAFS Fitting

Make FEFF SS Paths

About SixPACK/SamX

Quit

北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

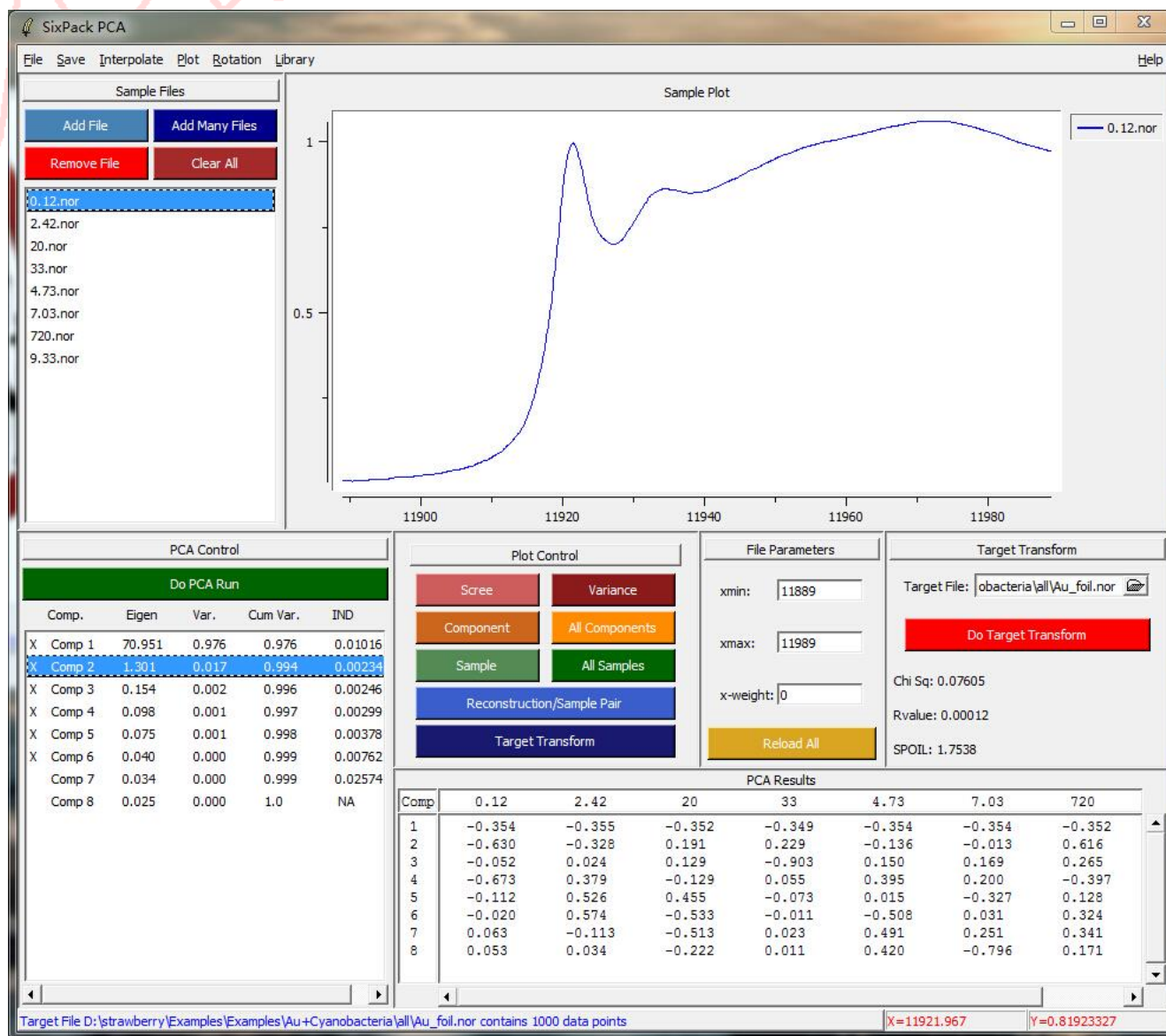
武汉·2014



中国科学院高能物理研究所

PCA软件的使用

旧版本ifefit软件包SixPack里面的PCA界面



北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

中国科学院高能物理研究所

PCA软件的使用

新版集成在Athena软件中的PCA分析组件

Athena [XAS data processing]

File Group Energy Mark Plot Freeze Merge Monitor Help

* cyanobacteria Save A U I

Principle components analysis

Analysis range: -20 to 80 Analysis space: ☒ norm $\mu(E)$ ☐ deriv $\mu(E)$ ☐ $\chi(k)$

Perform PCA

Performed PCA using normalized $\mu(E)$
 Number of components: 8 spectra
 Number of observations: 128 data points

	Eigenvalues	Variance	Cumulative variance
1:	7.982219	0.997777	0.997777
2:	0.017453	0.002182	0.999959
3:	0.000217	0.000027	0.999986
4:	0.000072	0.000009	0.999995
5:	0.000019	0.000002	0.999997

Plots: Components from 1 Data stack Scree ☐ Log Cumulative variance Cluster analysis 1 vs 2 Cluster plot

Actions: Reconstruct data with 2 Target transform with 8 TT coefficients

Save things to files: Components Data stack Reconstruction Target transform

Document section: principle components analysis

Plot first derivative data when plotting the current group in energy.

0.12 2.42 4.73 7.03 9.33 20 33 720

☒ Au foil ☐ Au1 Cl ☐ Au3 Cl aq ☐ Au hydroxide ☐ Au cyanide ☐ Au thiocyanide ☐ Au sulphide ☐ Au thiosulphate aq

E k R q kq

E k R q

Plotting k-weights: ☐ 0 ☐ 1 ☒ 2 ☐ 3 ☐ kw

Plot in energy: ☒ $\mu(E)$ ☐ Background ☐ pre-edge line ☐ post-edge line ☐ Normalized ☒ Normalized ☐ Derivative ☐ 2nd derivative ☐ $\mu(E)$ ☐ Derivative ☐ 2nd derivative

Emin -200 Emax 1000

北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014



中国科学院高能物理研究所

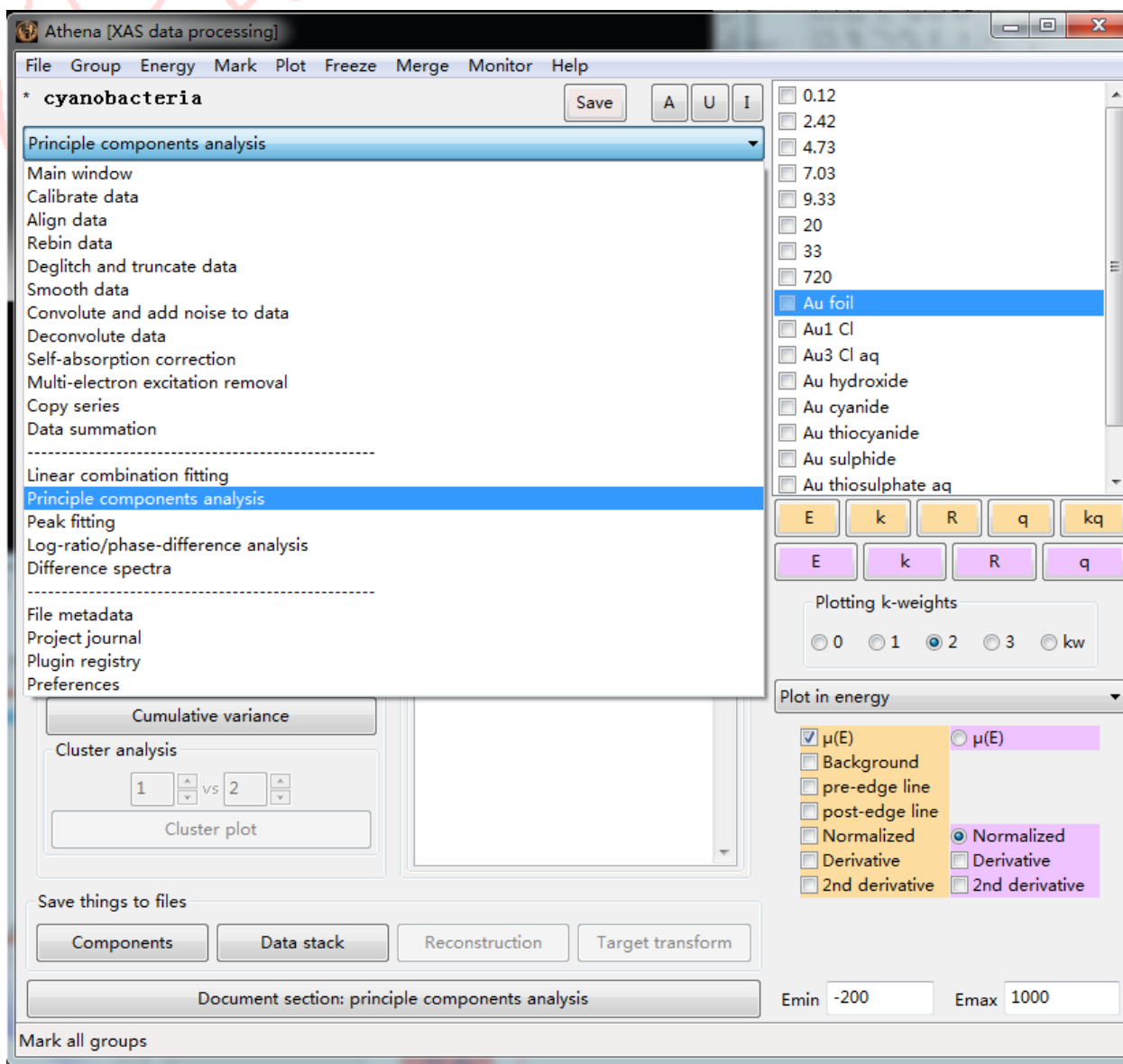


北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分析讲习班

武汉·2014

PCA软件的使用



中国科学院高能物理研究所



PCA软件的使用

相比于SixPack PCA，Athena PCA的优势：

数据调整灵活

更方便、多样的
保存数据方法

非常方便的选择做
PCA分析的数据以
及对标样进行对比

可以很方便的在不
同的空间进行PCA
分析





北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用

数据预处理

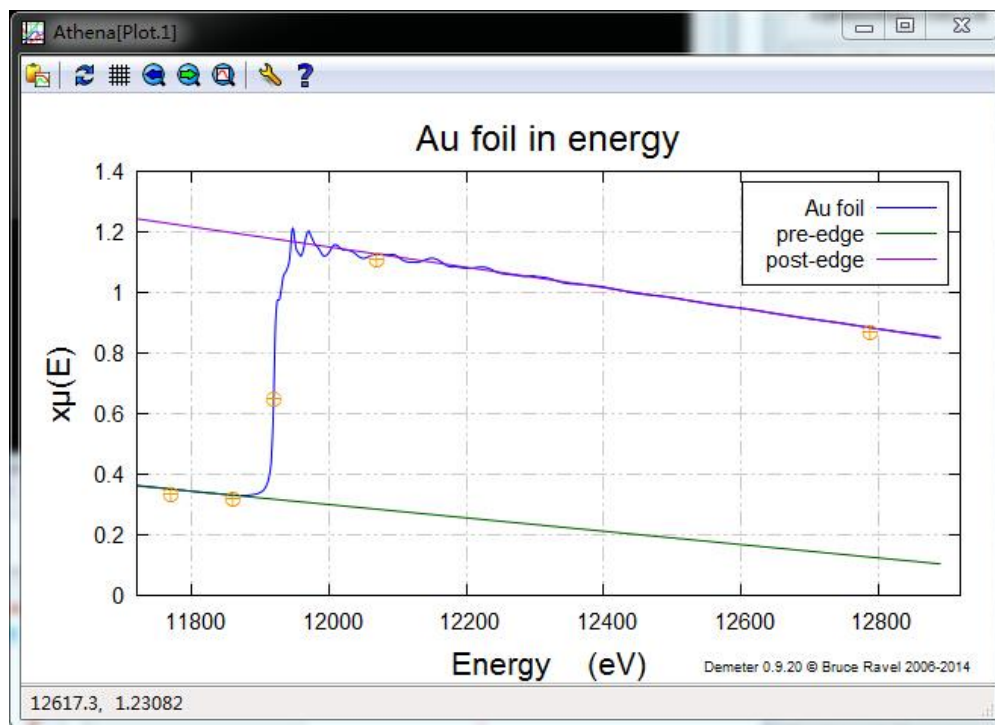


中国科学院高能物理研究所



PCA软件的使用

数据预处理——归一化



- 同一组样品归一化区间尽量相同
- 归一化之前需要对样品数据进行一些处理





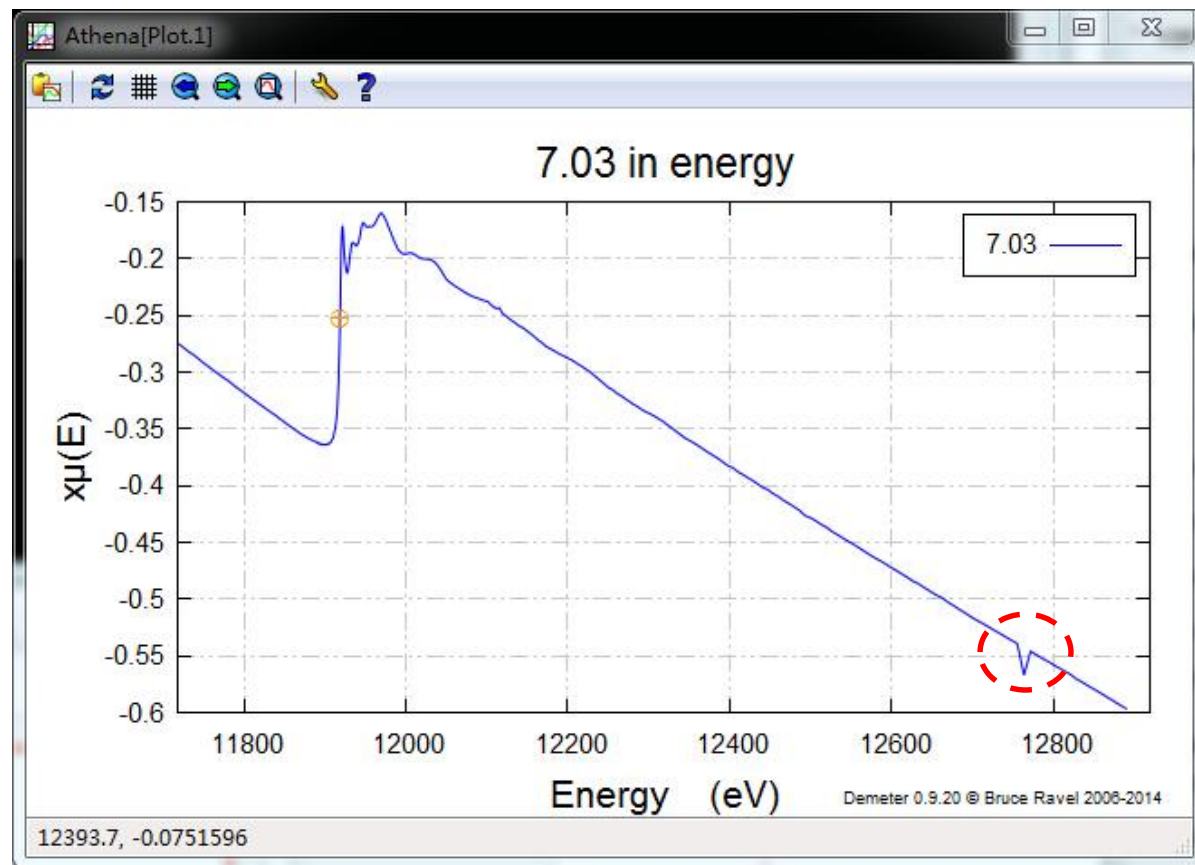
北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分析讲习班

武汉·2014

PCA软件的使用

数据预处理——去Glitch



中国科学院高能物理研究所



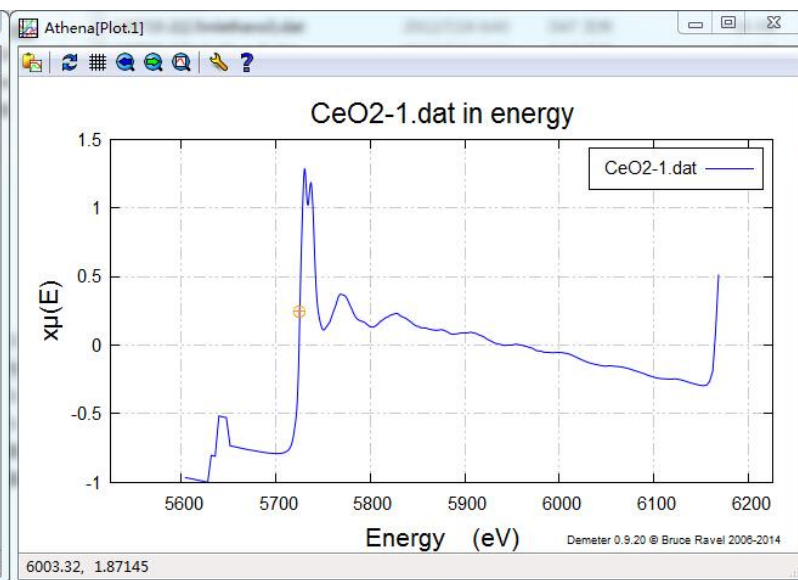
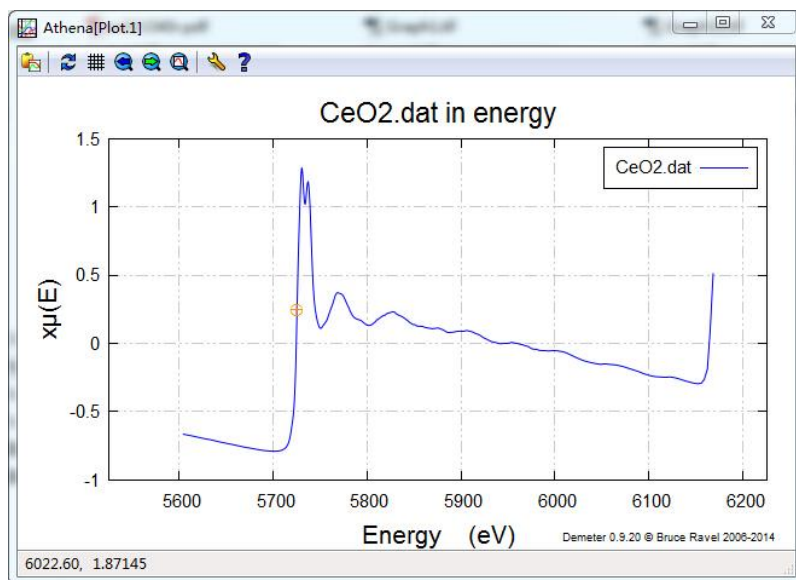
北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用

数据预处理——截取数据



Truncate data

Drop points

☐ before ☒ after

12856.0

Replot Truncate data Truncate marked

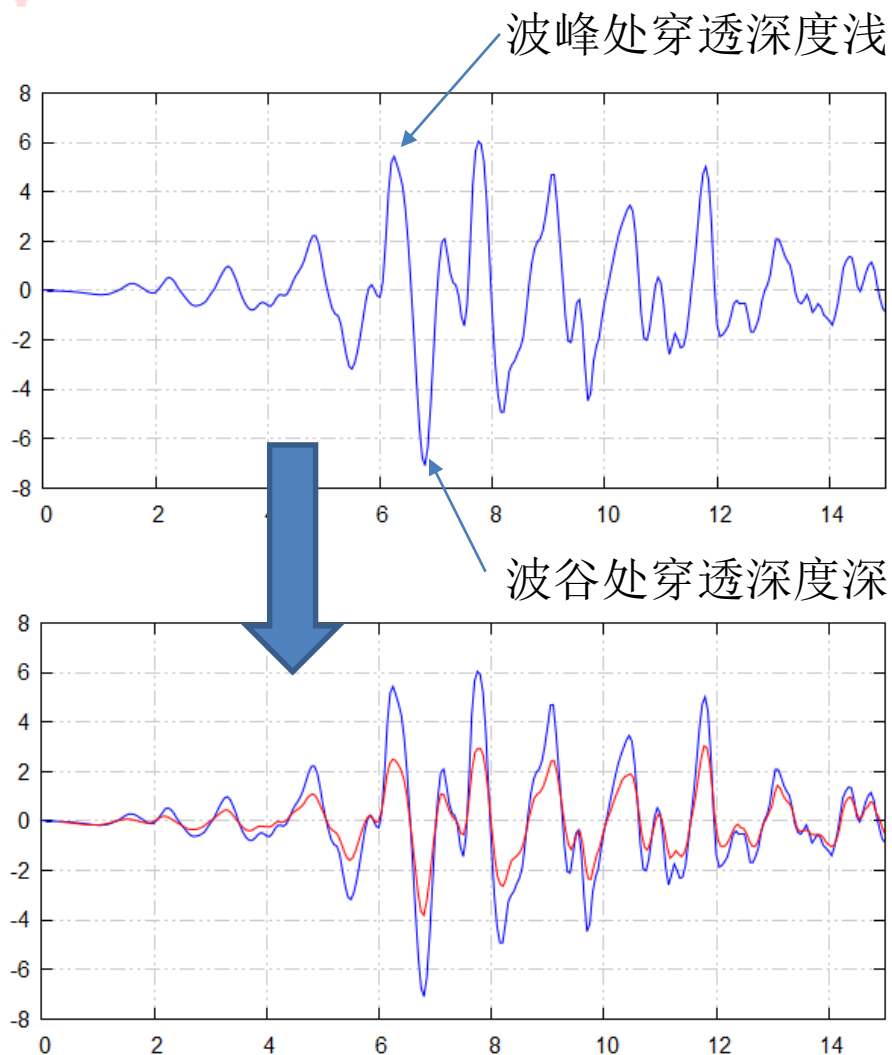


中国科学院高能物理研究所



PCA软件的使用

数据预处理——自吸收校正



在荧光测量模式下，探测器测量的是入射光激发的二次荧光信号。

在震荡信号的波峰位置，X光穿透深度浅，产生的荧光信号相对减弱；同理，在波谷的位置，产生的荧光信号增强，这样就将震荡信号的强度进行了衰减。

自吸收效应的产生





PCA软件的使用

数据预处理——自吸收校正

需要注意自吸收效应的情况：

- ◆ 样品中所测元素的含量太高
- ◆ 样品太厚
- ◆ 在荧光模式下

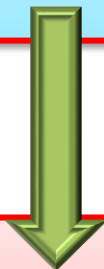




PCA软件的使用

数据预处理——自吸收校正

- ◆ 尽量在实验上规避自吸收效应
- ◆ 对于样品含量较高的，能做透射不做荧光
- ◆ 薄膜样品可采用掠入射方法
- ◆



数据处理上考虑

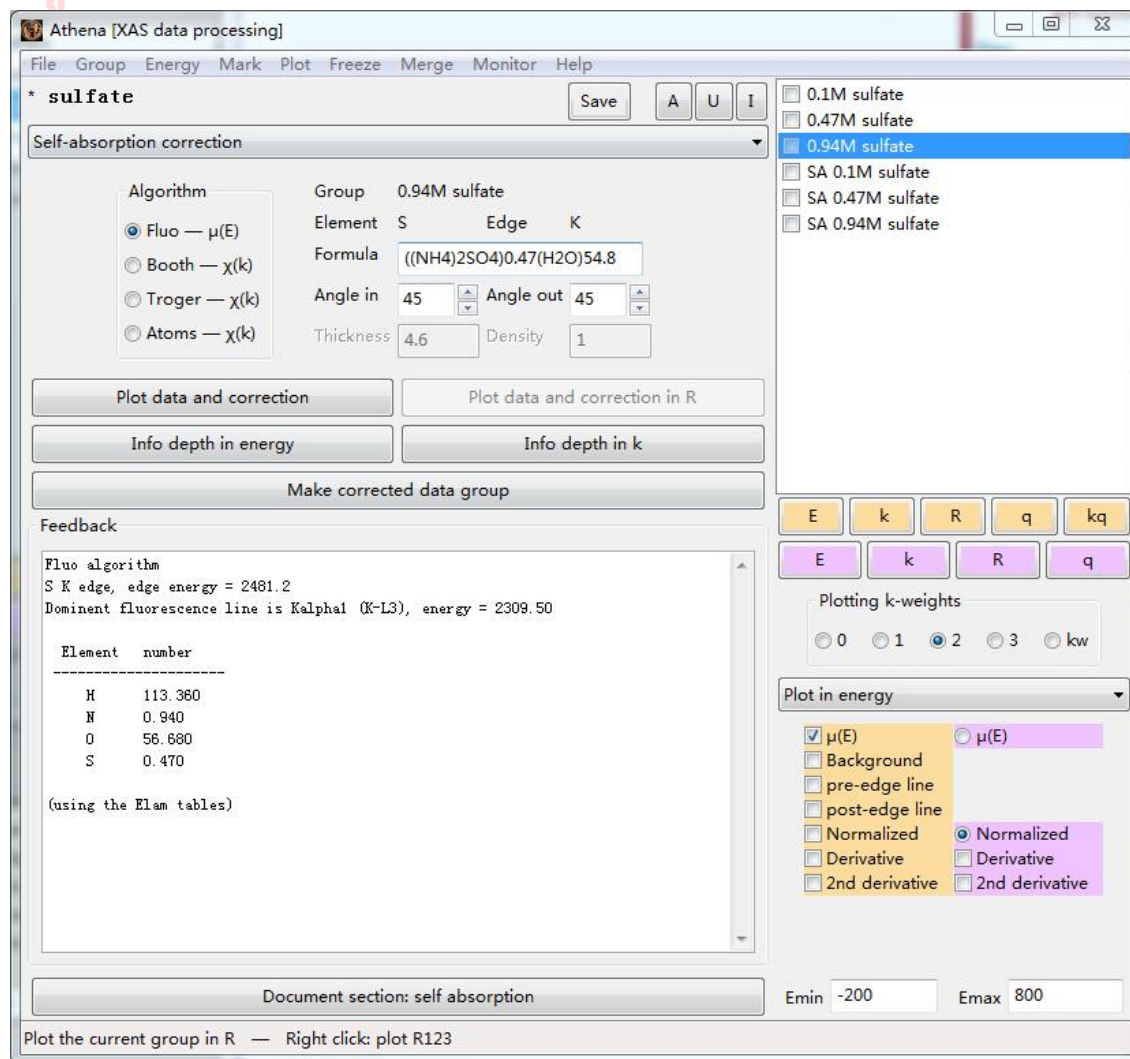
自吸收效应的校正





PCA软件的使用

数据预处理——自吸收校正



● 针对荧光模式

● 有四种算法



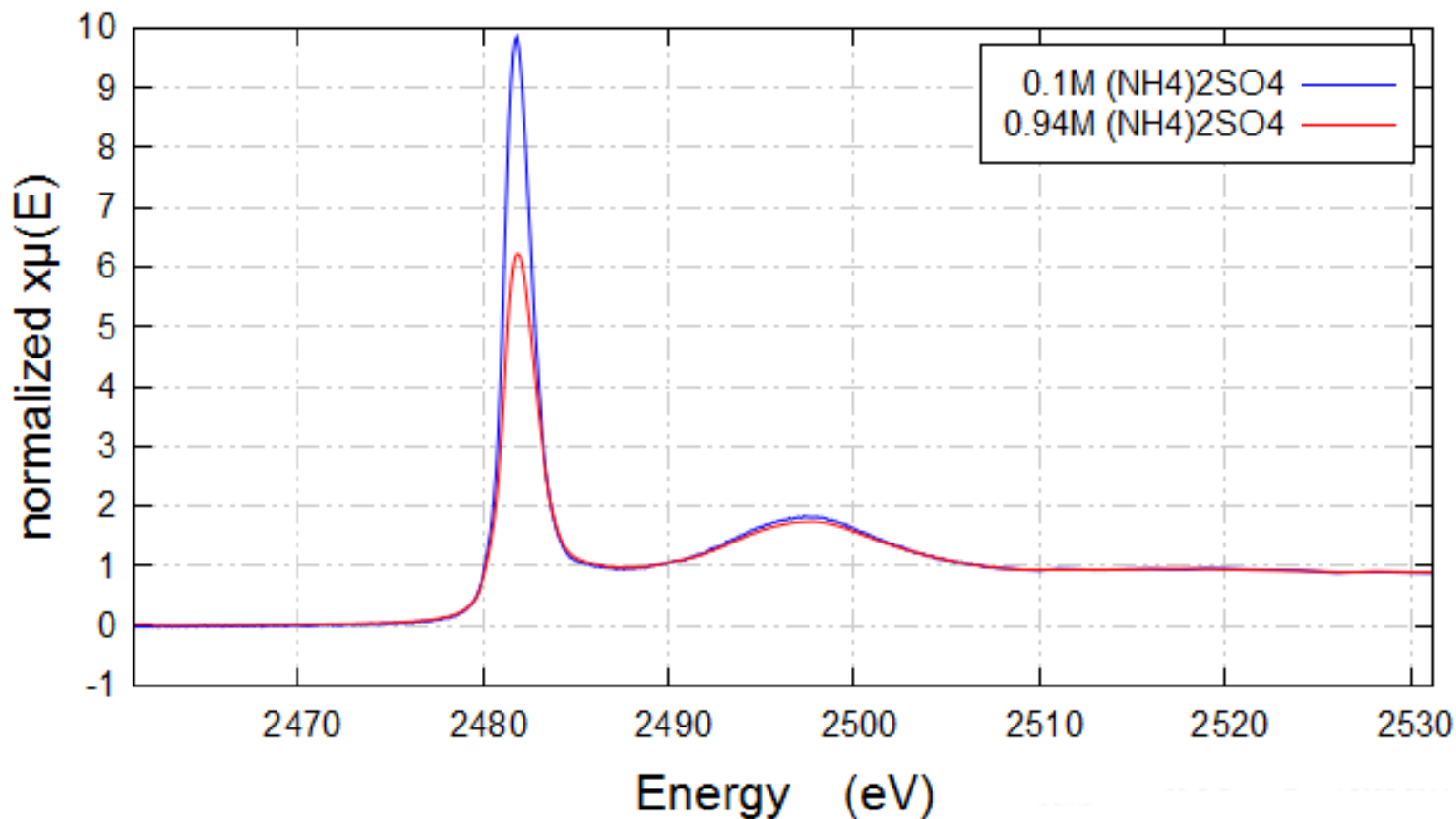
北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用

数据预处理——自吸收校正



中国科学院高能物理研究所



PCA软件的使用

数据预处理——自吸收校正

Self-absorption correction

选择算法

Algorithm

- ☒ Fluo — $\mu(E)$
- ☐ Booth — $\chi(k)$
- ☐ Troger — $\chi(k)$
- ☐ Atoms — $\chi(k)$

Group 0.94M (NH₄)₂SO₄

Element S

Edge K

填写相关参数

Formula ((NH₄)₂SO₄)_{0.94}(H₂O)_{54.8}

Angle in 45

Angle out 45

Thickness 4.6

Density 1

Plot data and correction

作图和修正

Info depth in energy

Info depth in k

Make corrected data group 建立新的数据

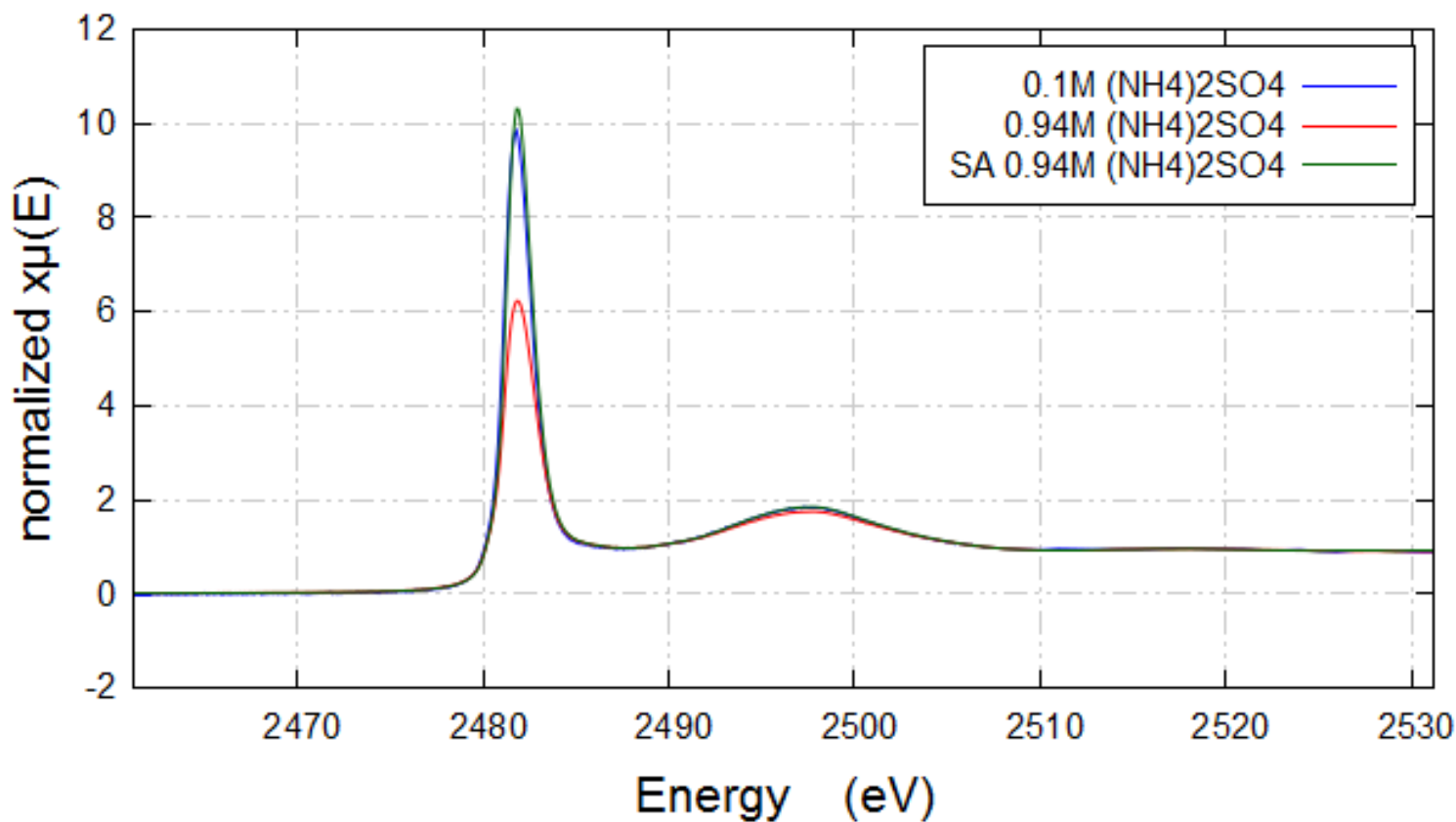
Athena软件进行自吸收校正的操作





PCA软件的使用

数据预处理——自吸收校正



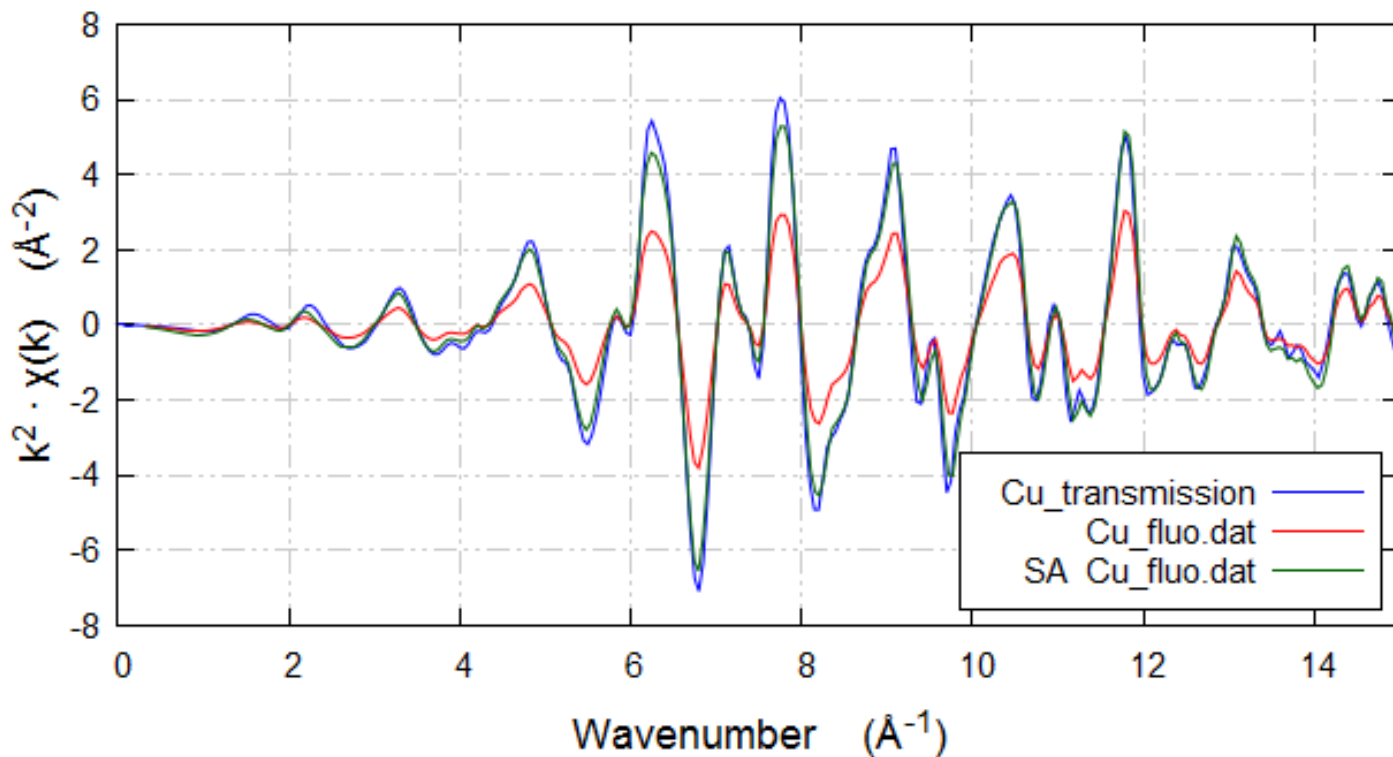
*Athena*软件进行自吸收校正的操作





PCA软件的使用

数据预处理——自吸收校正



*Athena*软件进行自吸收校正的操作





北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据解析讲习班

武汉·2014

PCA软件的使用

数据预处理——自吸收校正

！ 再次提醒

**不要过分迷信于软件对自吸收的修正
尽量在实验上进行考虑**



中国科学院高能物理研究所



Beijing Synchrotron Radiation Facility

北京同步辐射装置

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用

PCA软件的具体操作

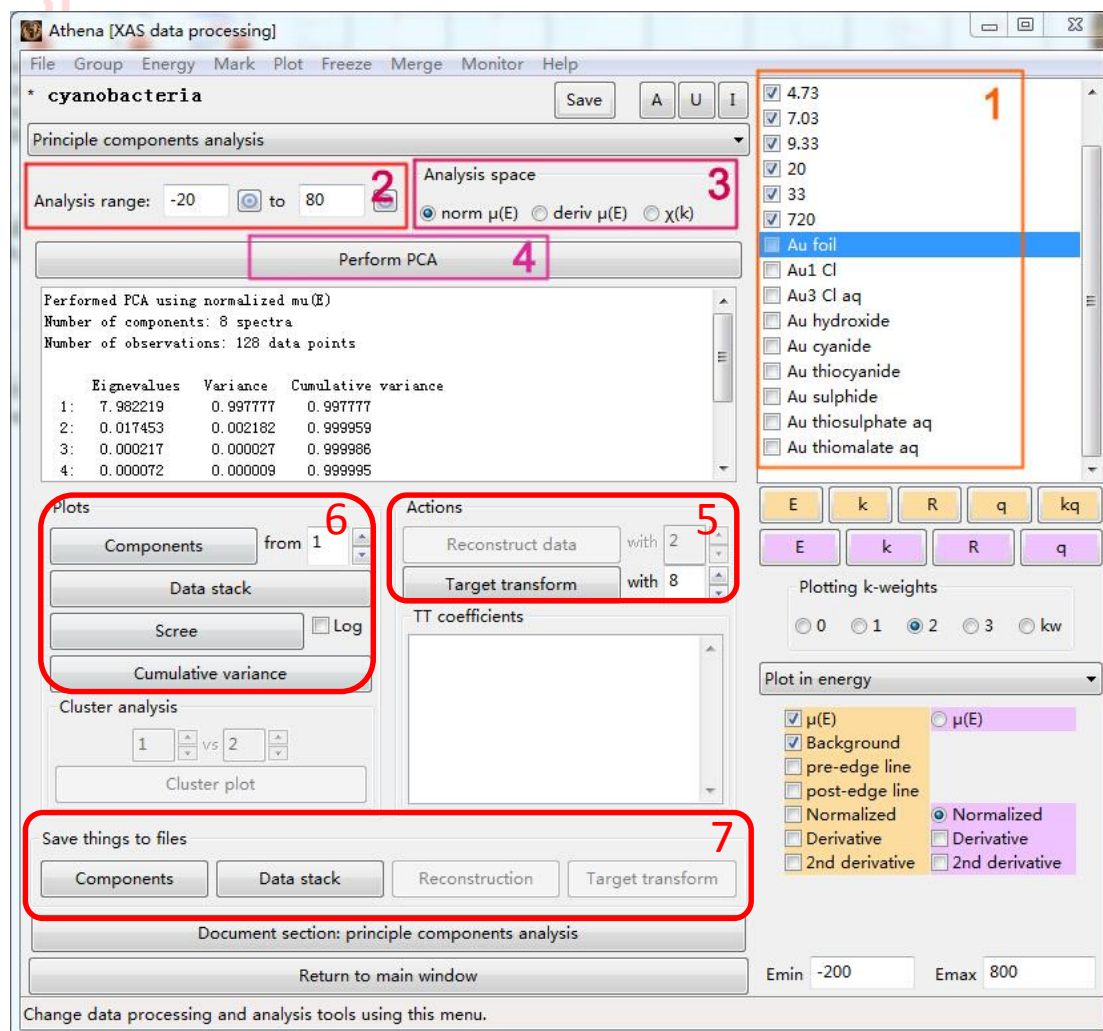


中国科学院高能物理研究所



PCA软件的使用

软件主要操作步骤



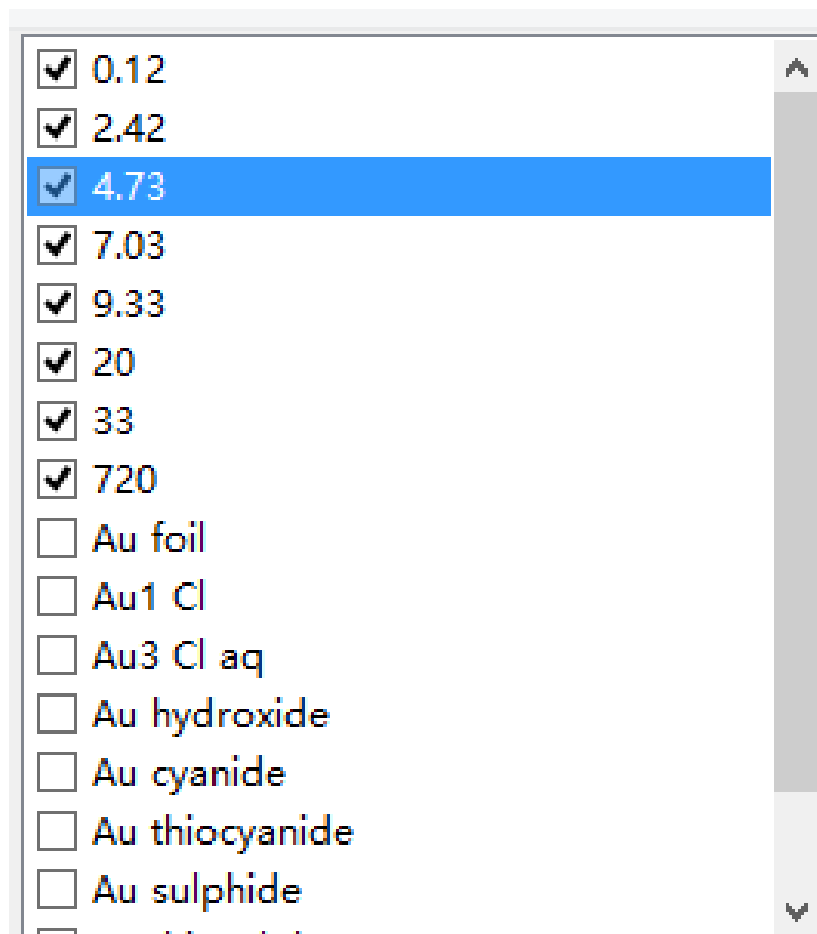
1. 数据
2. 选择PCA范围
3. 选择模式
4. 进行主元拆分
5. 重建数据和目标变换
6. 作图选项
7. 保存选项





PCA软件的使用

软件主要操作步骤



- ✓ 不要选择标样
- ✓ 数据组的个数至少要3个

选择数据组





PCA软件的使用

软件主要操作步骤

选择进行PCA分析的范围

选择模式

Principle components analysis

Analysis range: -20 to 80

Analysis space
☒ norm $\mu(E)$ ☐ deriv $\mu(E)$ ☐ $\chi(k)$

Perform PCA 进行主元拆分

Performed PCA using normalized $\mu(E)$
Number of components: 8 spectra
Number of observations: 128 data points

	Eignevalues	Variance	Cumulative variance
1:	7.982219	0.997777	0.997777
2:	0.017453	0.002182	0.999959
3:	0.000217	0.000027	0.999986
.	-----	-----	-----

➤ 对于同一组数据，可以分别进行近边及扩展边的分析，以相互印证





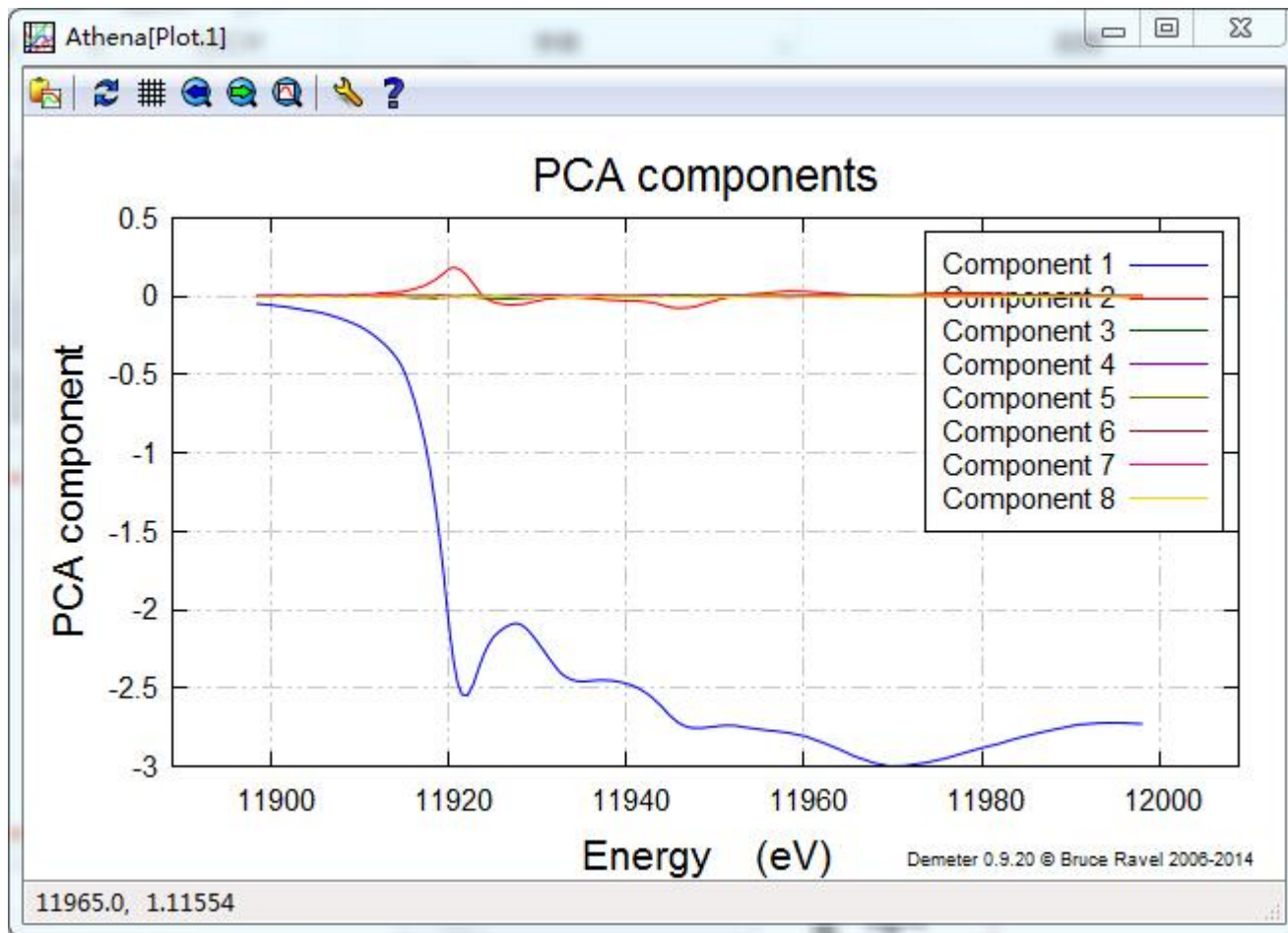
北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用

主元



中国科学院高能物理研究所



PCA软件的使用

软件主要操作步骤

Plots

Components	from	1	▲ ▼
Data stack			
Scree		<input type="checkbox"/> Log	
Cumulative variance			

- 作图界面
 - 主元
 - 数据组
 -

Actions

Reconstruct data	with	4	▲ ▼
Target transform	with	4	▲ ▼

- 寻找标样
 - 重建数据
 - 目标变换





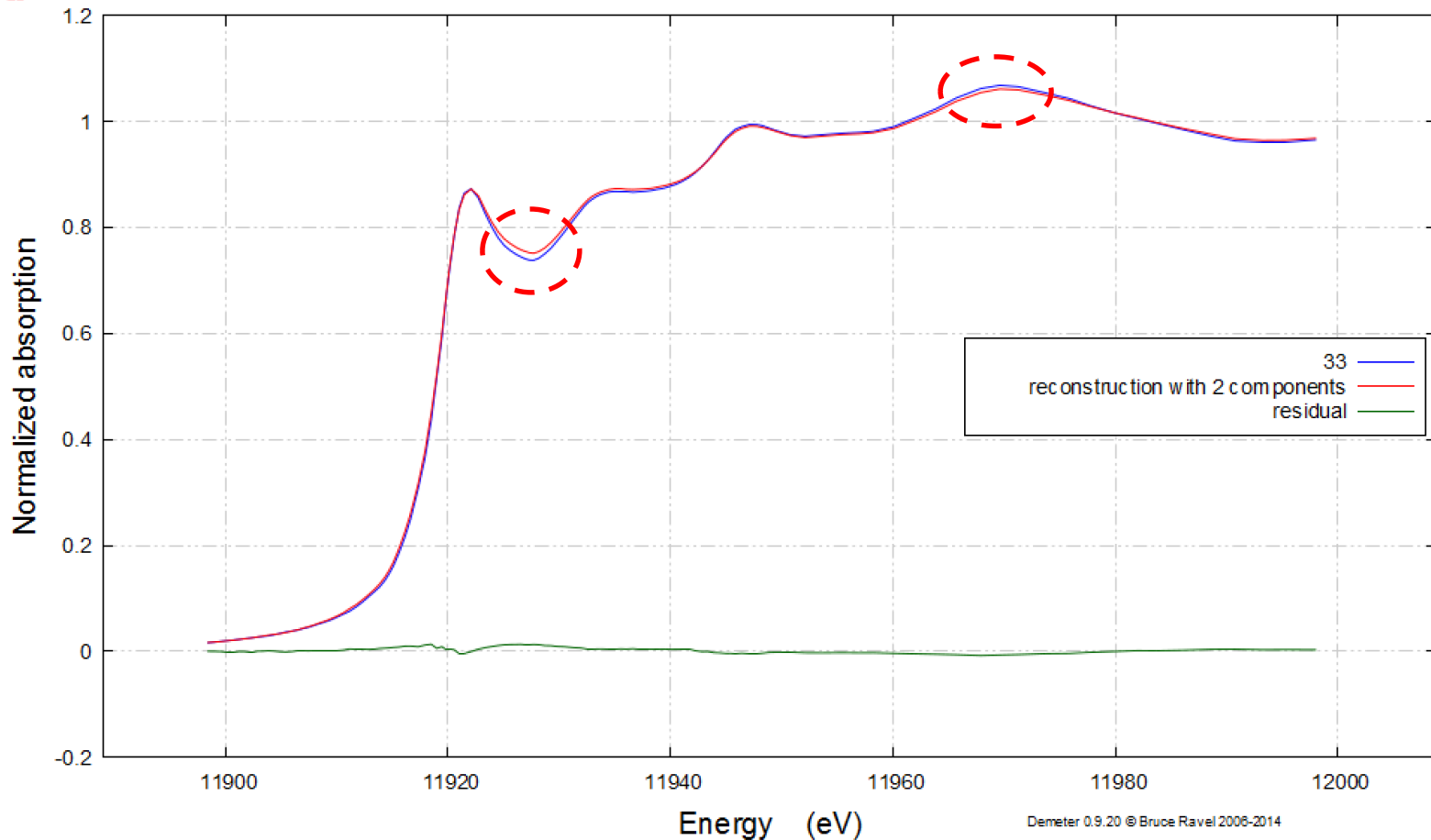
北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用

□ 2主元 ×



利用主元重建数据



中国科学院高能物理研究所



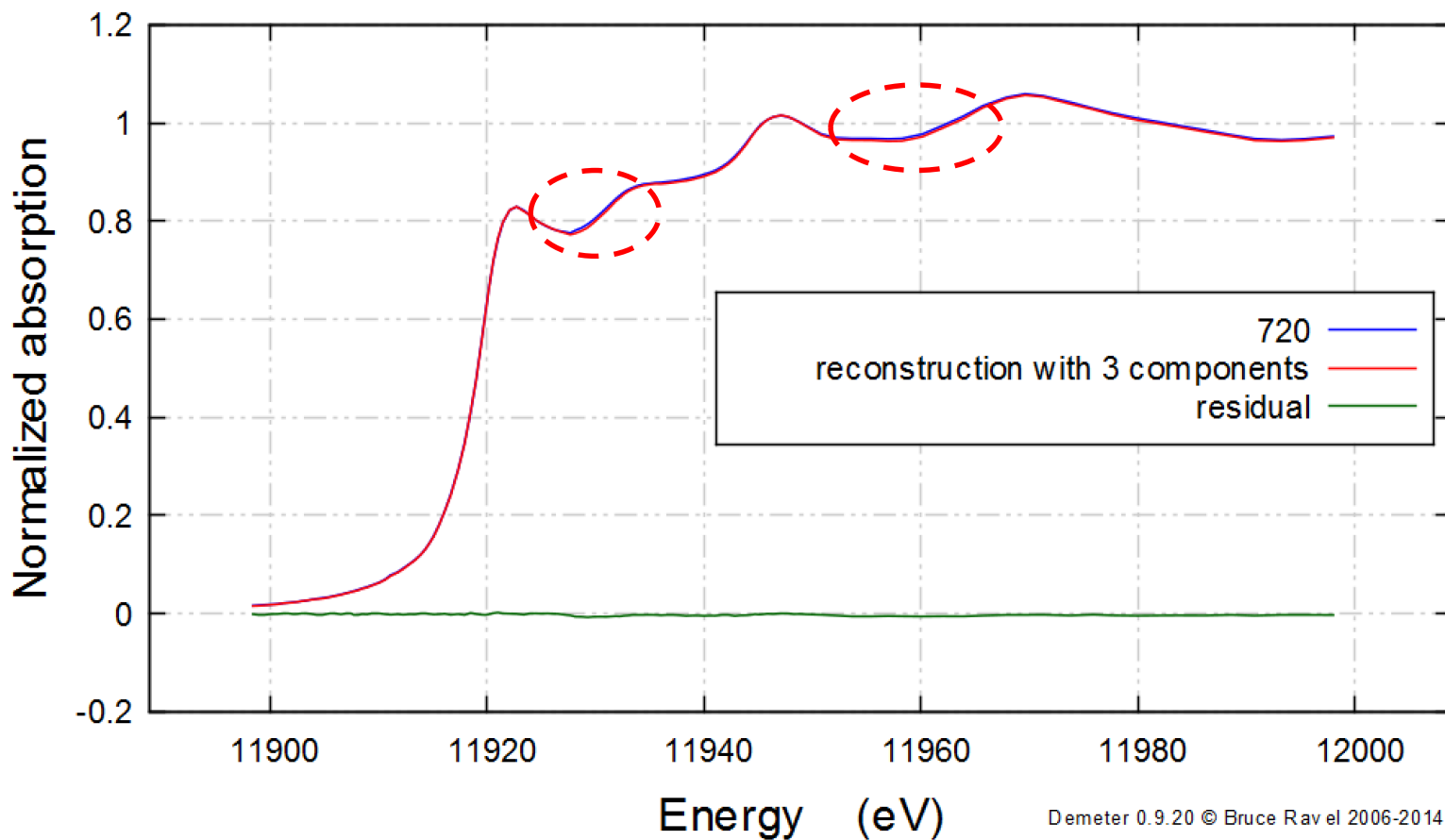
北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用

□ 3主元 ×



利用主元重建数据
中国科学院高能物理研究所



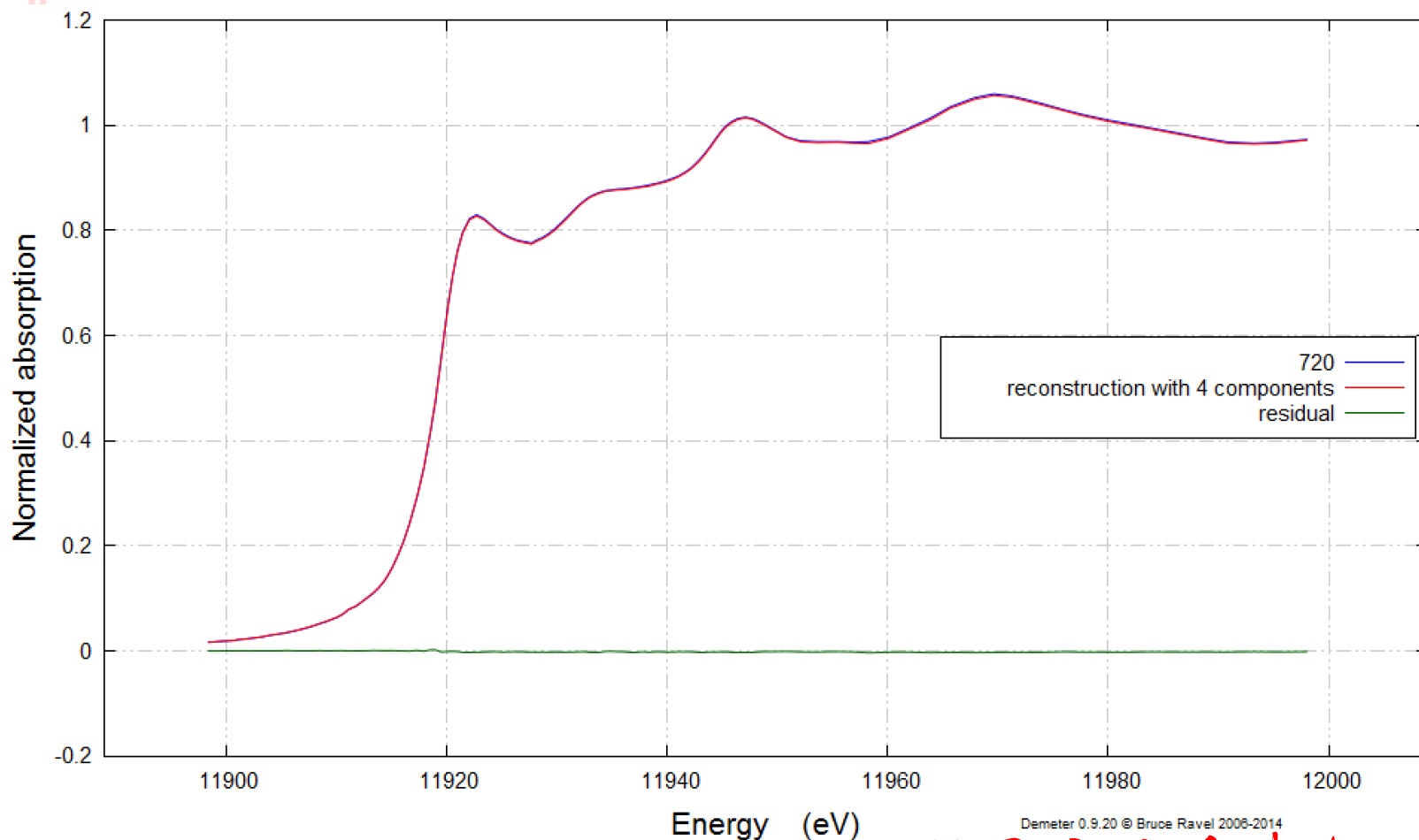
北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分析讲习班

武汉·2014

PCA软件的使用

□ 4主元 ✓



利用主元重建数据
中国科学院高能物理研究所

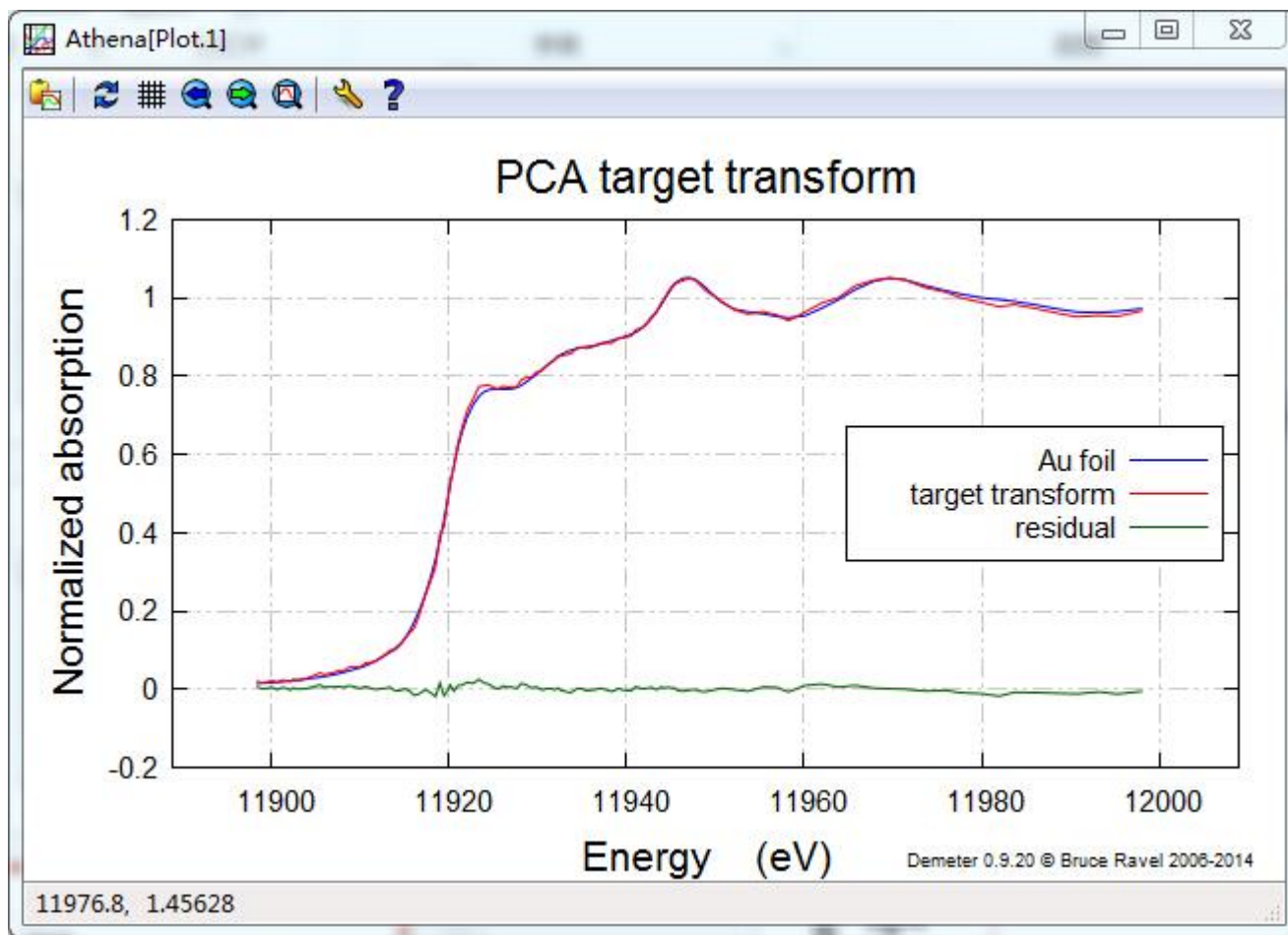


北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用



进行目标变换寻找标样



中国科学院高能物理研究所

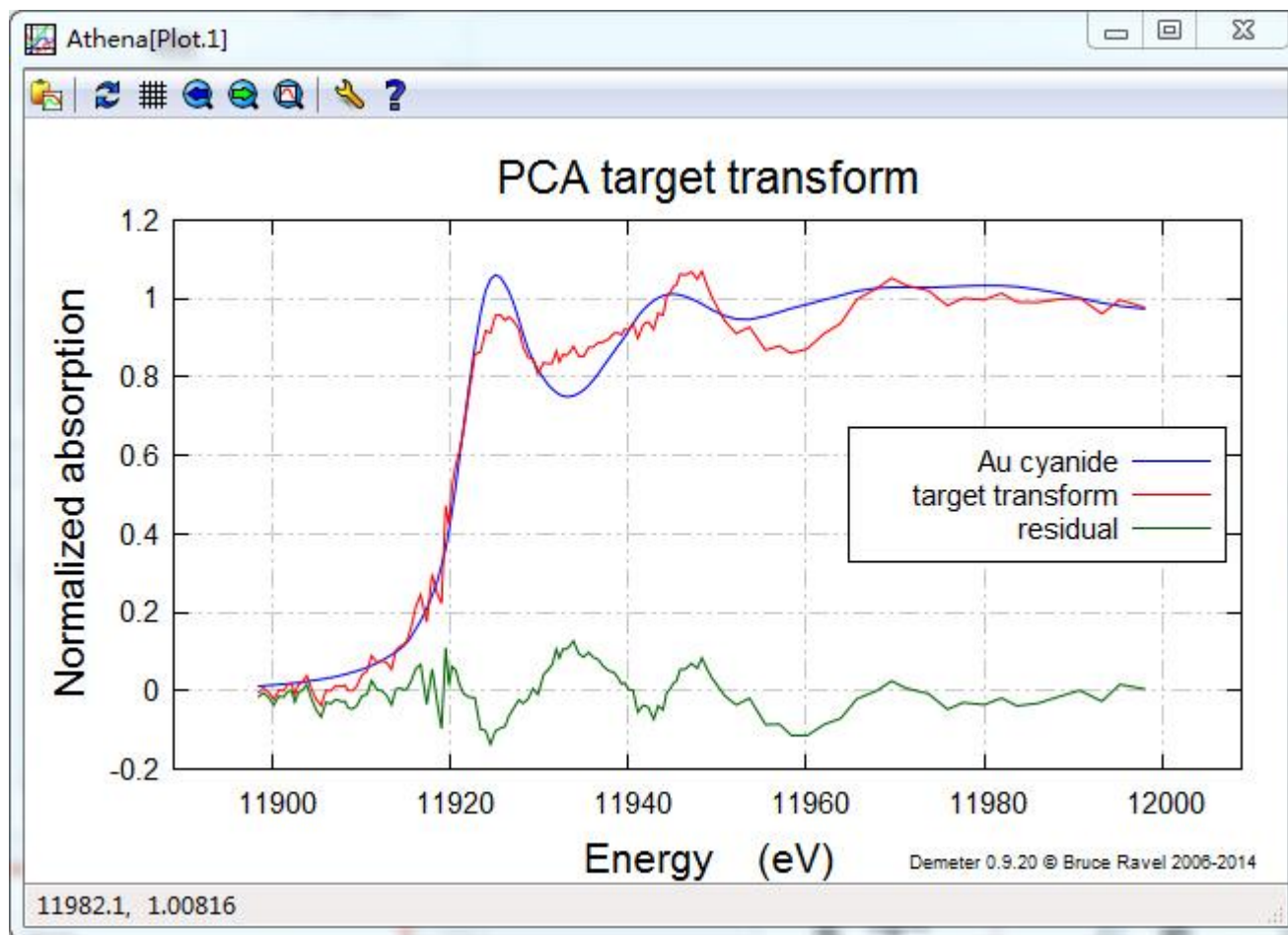


北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分折讲习班

武汉·2014

PCA软件的使用



进行目标变换寻找标样

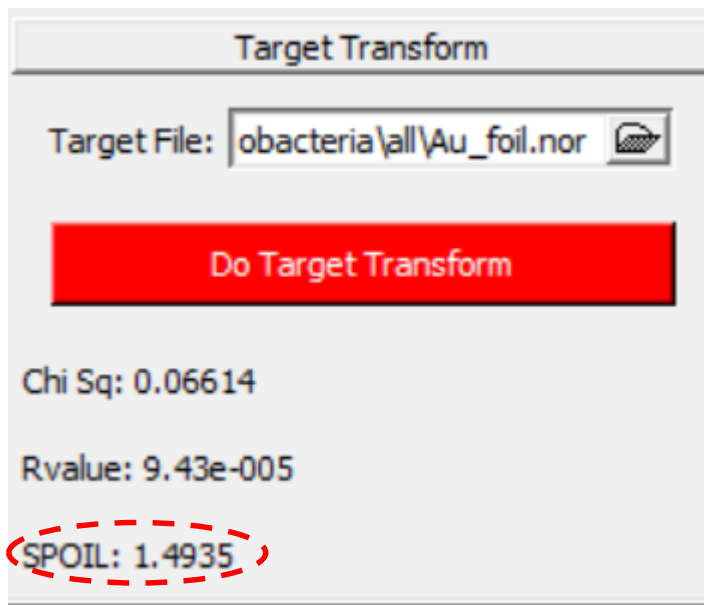


中国科学院高能物理研究所

PCA软件的使用

寻找标样——准则

$$SPOIL = \sqrt{\frac{N(M-C) \sum_i (\tilde{\chi}_i^{ref} - \chi_i^{ref})^2}{(N-C) \sum_{\alpha=C+1}^M \lambda_{\alpha}^2 \sum_{a=1}^C \left[\lambda_a^{-1} \sum_i E_i^a \chi_i^{ref} \right]^2} - 1}$$



参考SPOIL的值可以来取舍标样：

- SPOIL<1.5 相当好
- 1.5<SPOIL<3 不错
- 3<SPOIL<4.5 一般
- 4.5<SPOIL<6 差
- SPOIL>6 不可接受

(旧版本可参考高能所徐伟老师的课件)



中国科学院高能物理研究所



北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分析讲习班

武汉·2014

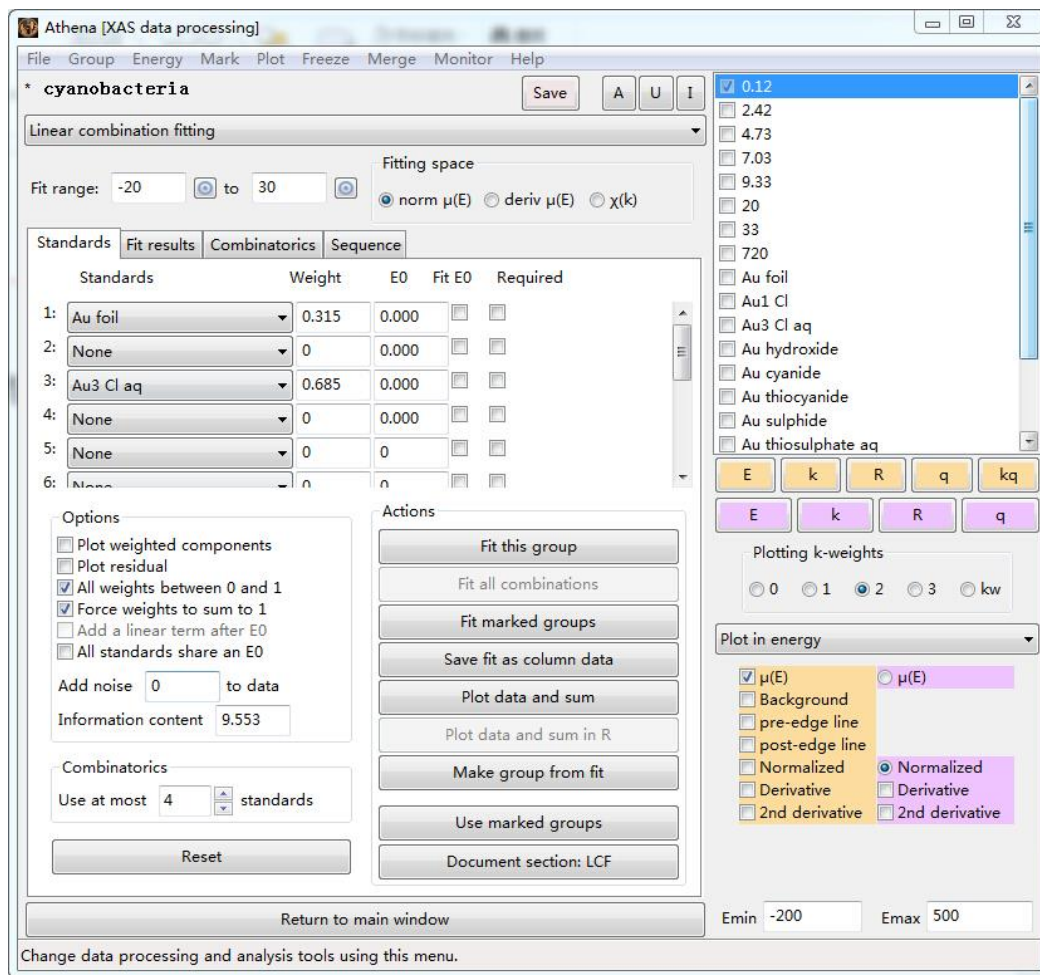
LCF软件的使用

LCF软件的使用



中国科学院高能物理研究所

LCF软件的使用



◆ 必须建立在PCA分析的基础上
◆ 能够进一步得到未知样品中各组分含量

北京同步辐射装置
Beijing Synchrotron Radiation Facility

×射线吸收谱学实验
和数据分析讲习班

武汉·2014



中国科学院高能物理研究所