

# A glimpse at the Rising Sun Envelope Method, with some examples

R. Monteiro, I. Miyasato, K. Takahashi

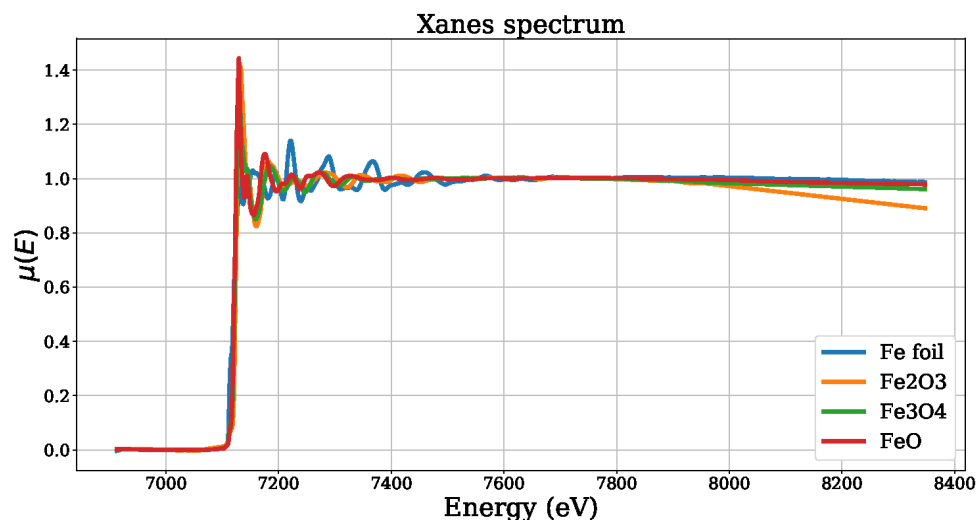
9/9/2019

## Overview

This is a quick overview to the algorithm presented in the paper “The Rising Sun Envelope Method: an automatic and accurate peak location technique for XANES measurements”, by R. Monteiro (MathAM-OIL/AIST, Sendai, Japan), I. Miyazato, and K. Takahashi (Hokkaido University, Japan).

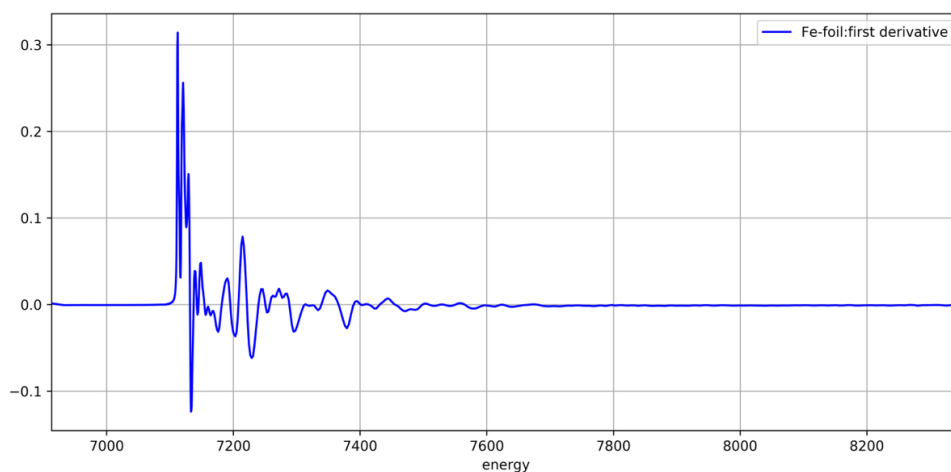
## Introduction

The problem we were concerned with is this: how do we find peaks and crests in XANES measurements.

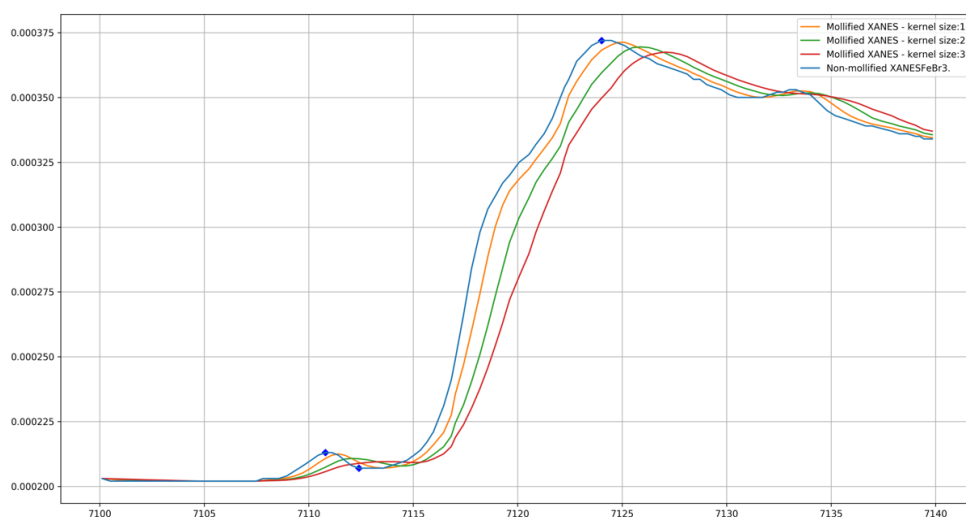


XANES (X-ray Absorption Fine Structure) is a common tool in Material Sciences that concerns to information about microscopic structure of materials. The idea, founded on the photo-electric effect, is that X-ray bombards the material and is scattered, reflected, and absorbed by it, giving information about its local geometrization and further properties of atoms and their neighboring structure.

It turns out that many techniques to find the peaks in these measurements fail brutally because they are non-differentiable.



Several standard approaches to locate these peaks have been tried, but all of them smear out peak location, as shown below:



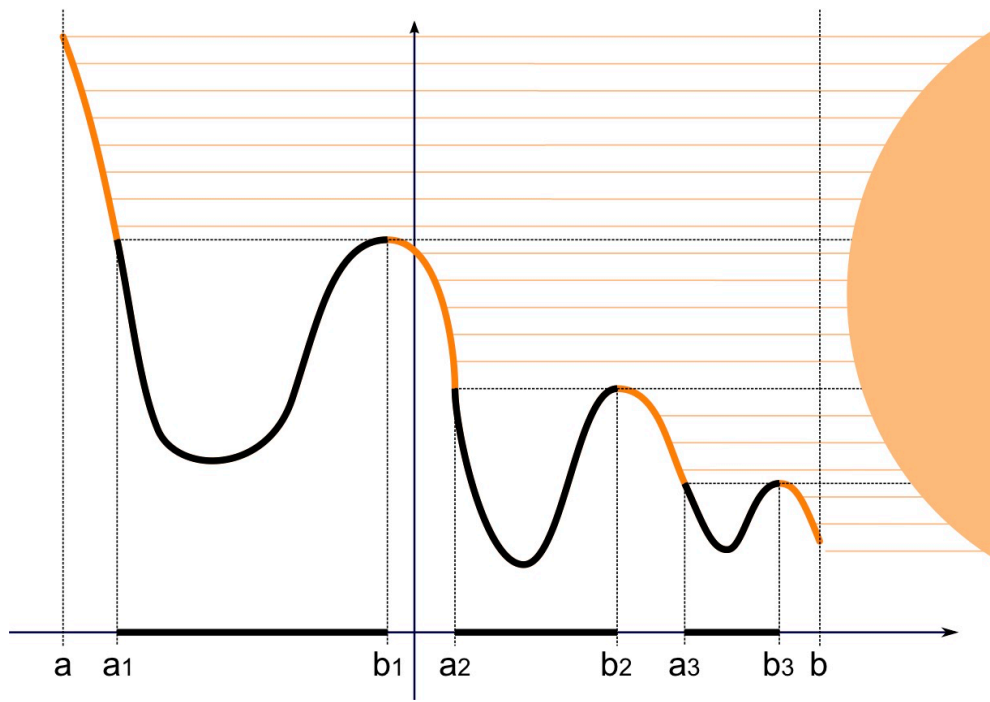
## Why do people care about it?

There are no equations for XAFS in the XANES region, hence a lot of our understanding on it comes from experiments. Experimentalists use the horizontal shift of the peaks in XANES as a measure of change in its valence, or say, oxidation state.

The initial project was motivated by the work (Miyazato, Takahashi, and Takahashi 2019), where a measure for the aforementioned peak shift was designed. In this project we decided to tackle the problem of locating these peaks. In passing we found other applications for the method, which we call here as the *Rising Sun Envelope Method*.

## The Rising Sun Envelope Method

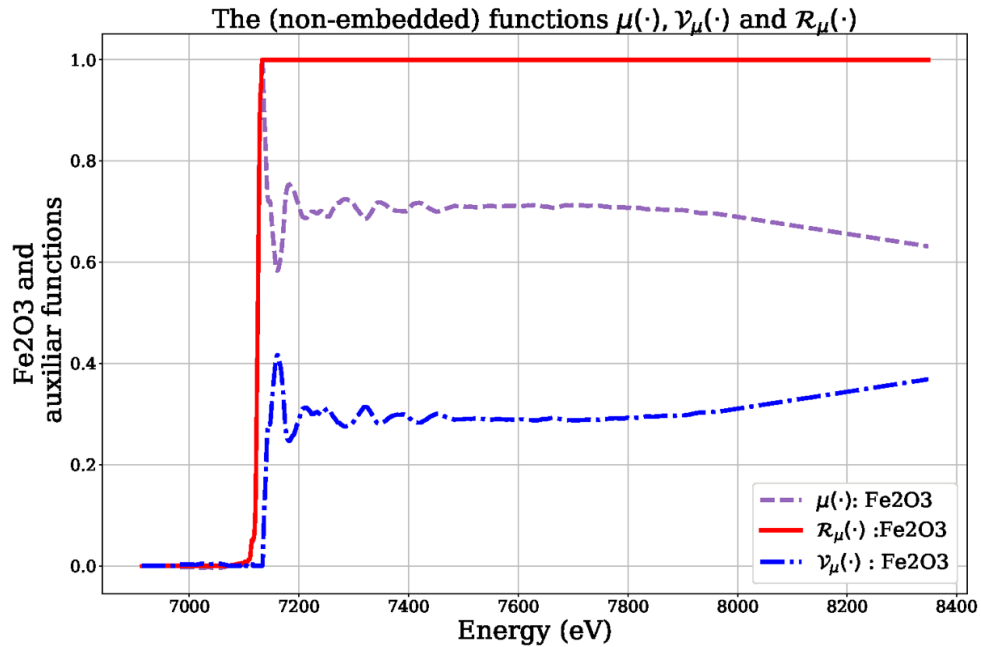
The Rising Sun Lemma is a common tool from Harmonic Analysis (Stein and Shakarchi 2005), created with the purpose of helping analysts to better understand differentiability properties of continuous functions. Our construction is pretty similar, but created with a different purpose: that of finding peaks of a non-differentiable function.



The method is based on iterations over the spectrum  $\mu(E)$  defined in energy ranges  $I_n$  that are decreasing in a chained fashion:

$$I_0 \supset I_1 \supset I_2 \supset \dots$$

In even index intervals, we define the Rising Sun function  $\mathcal{R}_\mu(E)$  of the measurement  $\mu(E)$ .



The Rising Sun function of a measurement  $\mu(E)$  in the energy range  $e_{-\infty} \leq E \leq e_{+\infty}$  is defined as

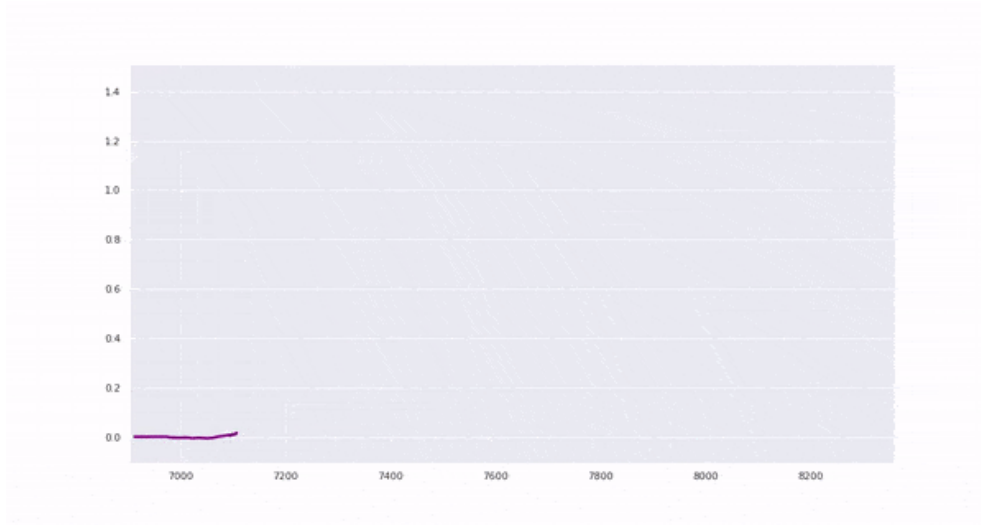
$$\mathcal{R}_\mu(E) := \max_{e_{-\infty} \leq F \leq E} \mu(F).$$

The main property that we exploit here is that both the Rising Sun function  $\mathcal{R}_\mu(E)$  and  $\mu(E)$  have the same location for their highest peak in the corresponding energy range interval. In that sense, the Rising Sun function gives the silhouette of the XANES curve, thus offering an information about the topology of its peaks

The Valley of Shadows in constrast gives an information about the “valleys”, about local minima. It is defined as

$$\mathcal{V}_\mu(E) = \mathcal{R}_\mu(E) - \mu(E),$$

which is always a non-neagative function. Note that, given the highest peak  $e_0$  that  $\mathcal{R}_\mu(E)$  and  $\mu(E)$  in an interval  $e_{-\infty} \leq E \leq e_{+\infty}$  , the highest peak of  $\mathcal{V}_\mu(E)$  in the range  $e_0 \leq E \leq e_{+\infty}$  gives the location of the first local minimum (i.e., “valley”) after the peak  $e_0$  . The idea can be seen in practice in the following gif:



## Discerning true peaks from random noise spikes.

The main difficulty of functions with noise is how to distinguish peaks from random spikes. Two quantities are introduced with this goal: in the vertical direction,  $h_*$  has the purpose of discerning peaks and random noise fluctuations; the other measurement is horizontal, a quantity  $d_*$  that takes into account how far appart peaks are.

**DEFINITION (peak criterion with thresholds [Mo,M,T-2019]):** given a measurement in energy range  $e_{-\infty} \leq E \leq e_{+\infty}$  and fixed positive constants  $(h_*, d_*)$ . Let  $e_{-\infty} \leq e_* < \bar{E} \leq e_{+\infty}$ . We say that  $\bar{E}$  is a local maximum relative to  $e_*$  with thresholds  $(h_*, d_*)$  whenever

$$\left| \mu(\bar{E}) - \mu(e_*) \right| > h_*, \quad \text{and} \quad \mu(\bar{E}) \geq \mu(E)$$

holds for all  $e_{-\infty} \leq E \leq e_{+\infty}$  with  $|E - \bar{E}| \leq d_*$  (see Figure fig:threshold\_slecting\_peaks). Similarly, one says that point  $e_* < \underline{E} \leq e_{+\infty}$  is a local minimum relative to  $e_*$  with thresholds  $(h_*, d_*)$  whenever

$$\left| \mu(e_*) - \mu(\underline{E}) \right| > h_*, \quad \text{and} \quad \mu(\underline{E}) \leq \mu(E)$$

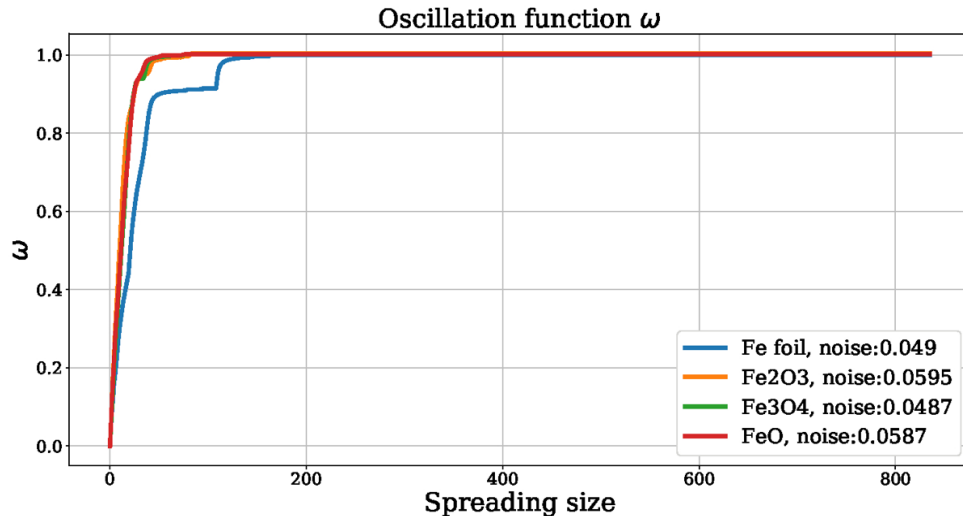
holds for all  $e_{-\infty} \leq E \leq e_{+\infty}$  with  $|E - \underline{E}| \leq d_*$ .

It turns out that the best way to evaluate this property is by using a Rising Sun function.

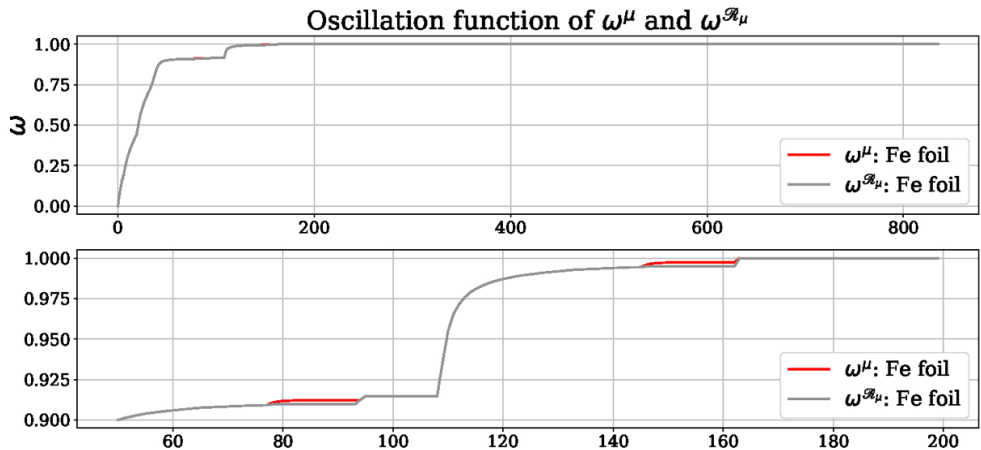
However, how to choose these thresholds? The oscillation threshold  $h_*^\mu$  is much simpler to motivate: we shall use

$$h_* = \omega_*^\mu(\delta), \quad \text{for} \quad \omega_*^\mu(j\delta) = \max_{|E-E'| \leq j\delta} |\mu(E) - \mu(E')|.$$

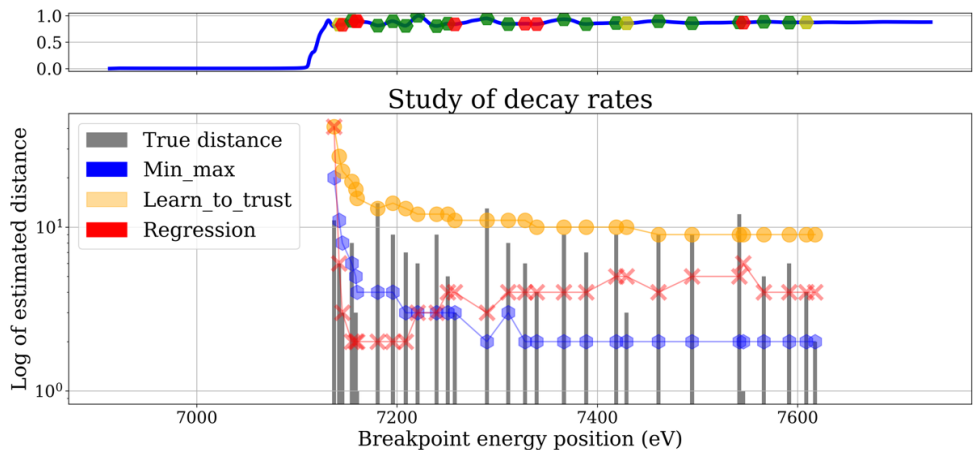
where  $\omega_*^\mu(E)$  is called *oscillation function* of  $\mu(E)$ . The oscillation function  $\omega(n)$  (where  $n$  denotes the width of the meshgrid, or the spreading size) is a non-decreasing function that is 0 only in the case of constant functions. An example is given below:



In the paper we show that , that is, the Rising Sun function has less noise than the measurement Curiously, this difference can be very small in some cases: for example, in the measurement of Fe foil their oscillation differ in 31 points in a 836 grid.



Quantification of how peaks spread in space is more complicated and several different techniques are used, ranging from simple averaging or simple regression to weighted regressions; see the paper for further details.



## Why does it work?

The main benefit of our construction is that it does not rely on differentiability properties. Furthermore, it does not destroy the information we are interested at, namely, peak intensity and peak location (i.e., they are invariants)

These initial results leave open the possibility for hyperparameter tuning using Machine Learning, besides the implementation of a package designed for peak location and decomposition of spectral measurements

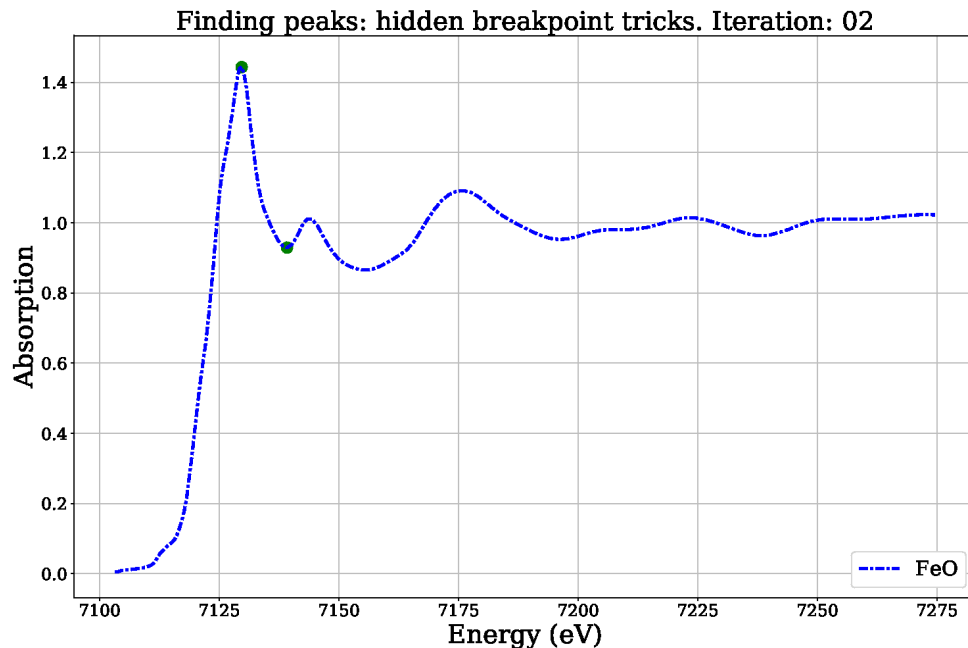
**Proposition [ Mo,I, T- 2019]:** The following properties are an immediate consequence of the construction of the Rising Sun  $\mathcal{R}_\mu(E)$ , Valley of Shadow functions  $\mathcal{V}_\mu(E)$  of a continuous measurement  $\mu(E)$ : 1. Fix the positive thresholds  $h_*$  and  $d_*$ . Let the point  $\bar{e}_0$  be the first local maximum of  $\mu$  in the range  $e_{-\infty} \leq E \leq e_{+\infty}$  with threshold  $(h_*, d_*)$ . Then,  $\bar{e}_0$  is the a local maximum of  $\mathcal{R}_\mu$  in the range  $e_{-\infty} \leq E \leq e_{+\infty}$  with threshold  $(h_*, d_*)$ , that is,

$$\mu(\bar{e}_0) = \mathcal{R}_\mu(\bar{e}_0) = \mathcal{R}_\mu(E), \quad \bar{e}_0 \leq E < \bar{e}_0 + d_*.$$

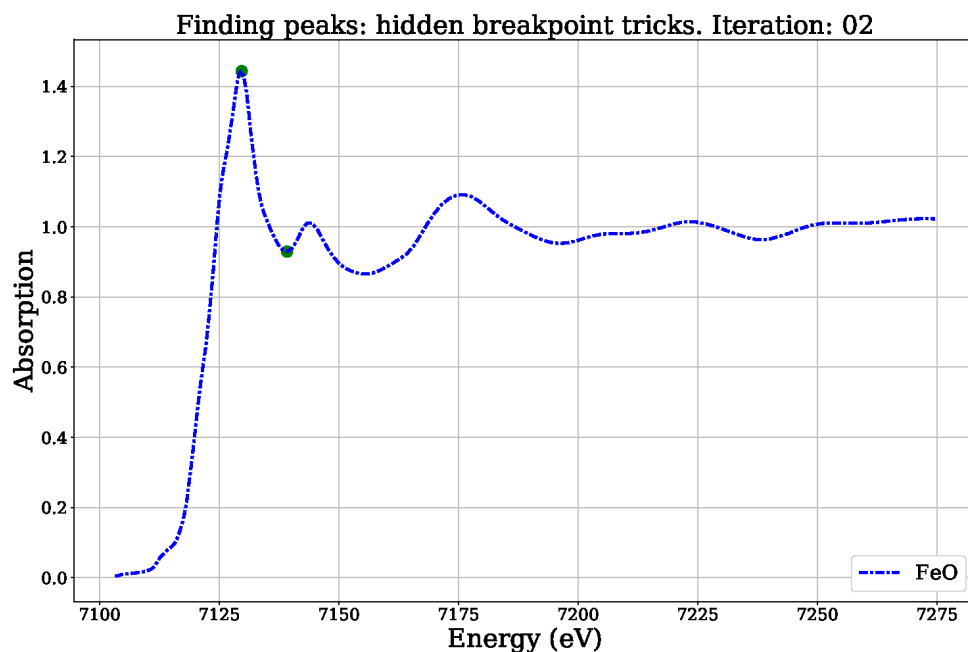
In other words, the functions  $\mu(E)$  and its associated Rising Sun function  $\mathcal{R}_\mu$  have the same first maximum in the range  $e_{-\infty} \leq E \leq e_{+\infty}$ ; Consider the first peak  $\bar{e}_0$  with threshold  $(h_*, d_*)$ . Let  $\bar{e}_1$  be the first local maximum with threshold  $(h_*, d_*)$  of  $\mathcal{V}_\mu(E)$  in the range  $\bar{e}_0 \leq E \leq e_{+\infty}$ . Then  $(\bar{e}_1, \mu(\bar{e}_1))$  is a local minimum of  $\mu|_{[e_0, e_{+\infty}]}(E)$  with threshold  $(h_*, d_*)$ . Furthermore,  $e_{-\infty} \leq e_0 < e_1 \leq e_{+\infty}$  is a sequence of 2-breakpoints with threshold  $(h_*, d_*)$ .

## Skipping peaks?

The algorithm could naturally skip peaks, regardless of how well we choose the resolution thresholds! we designed two methods, *hidden peak tricks*, to circumvent this issue: the *middle step*,

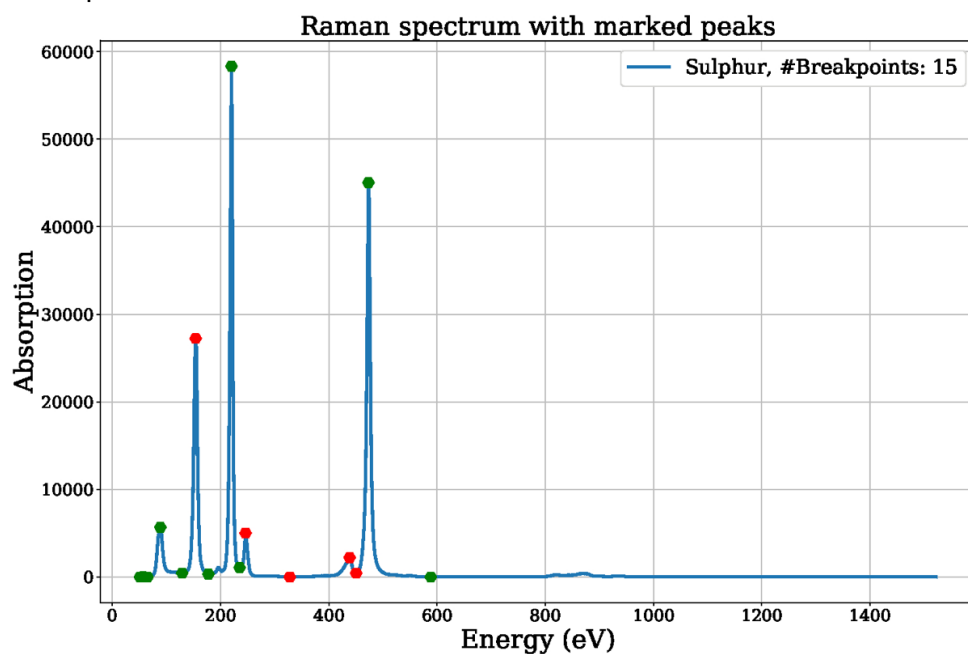


and the *back step*:



## Some applications

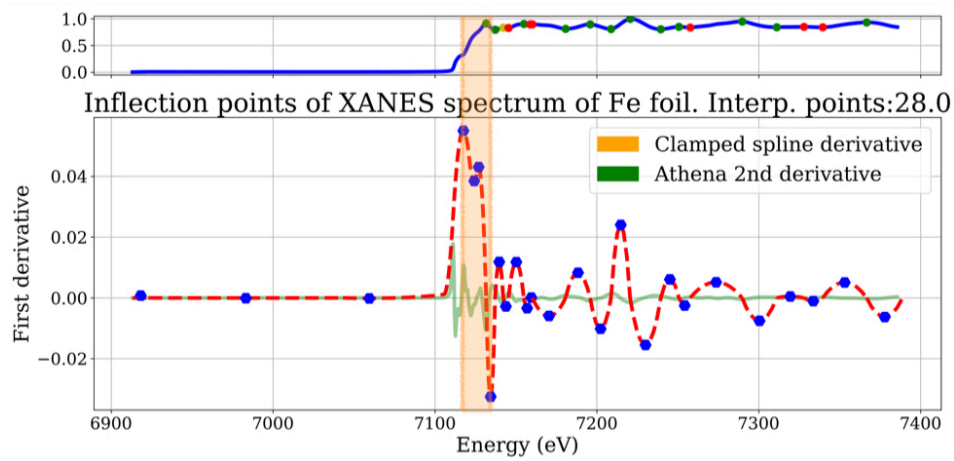
1. The first direction application concerns peak detection. We show two examples here. The first in XANES  
The second, to Raman spectrum.



2. Inflection points

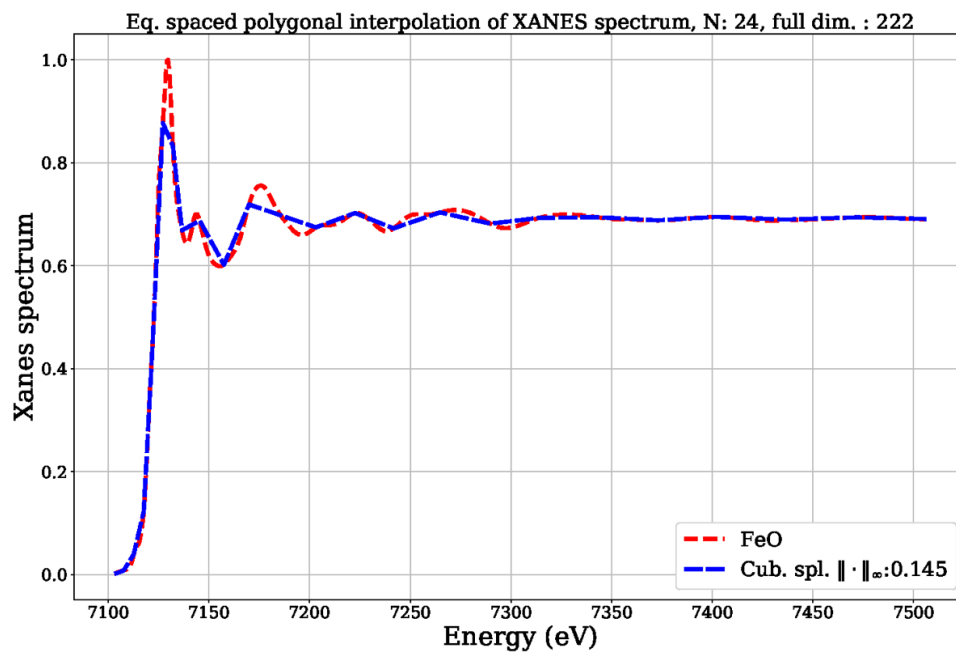
No other code can do this automatically and with so much quality (even in noisy cases)



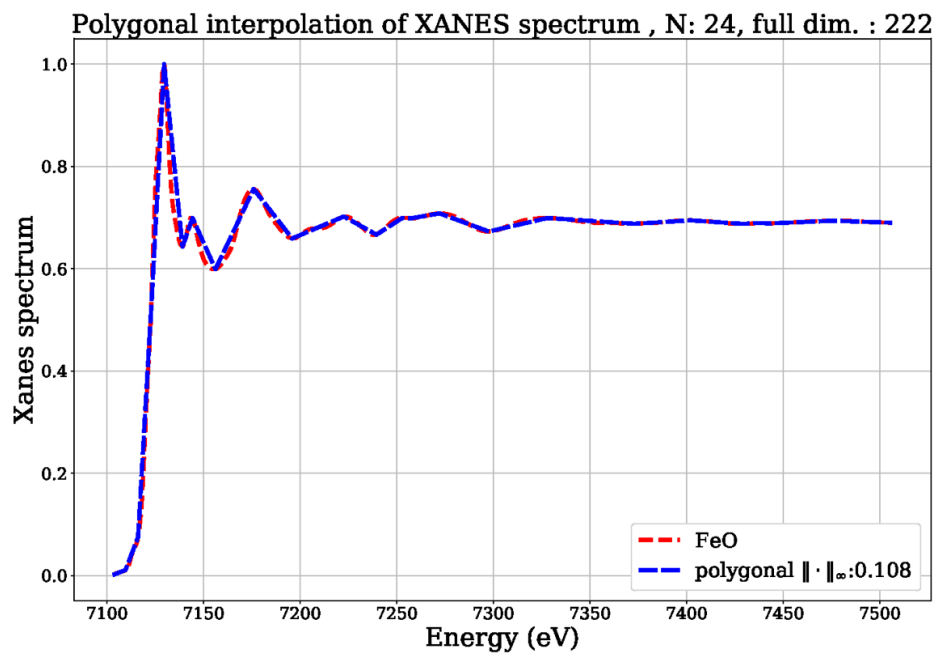


### 3. Dimension reduction

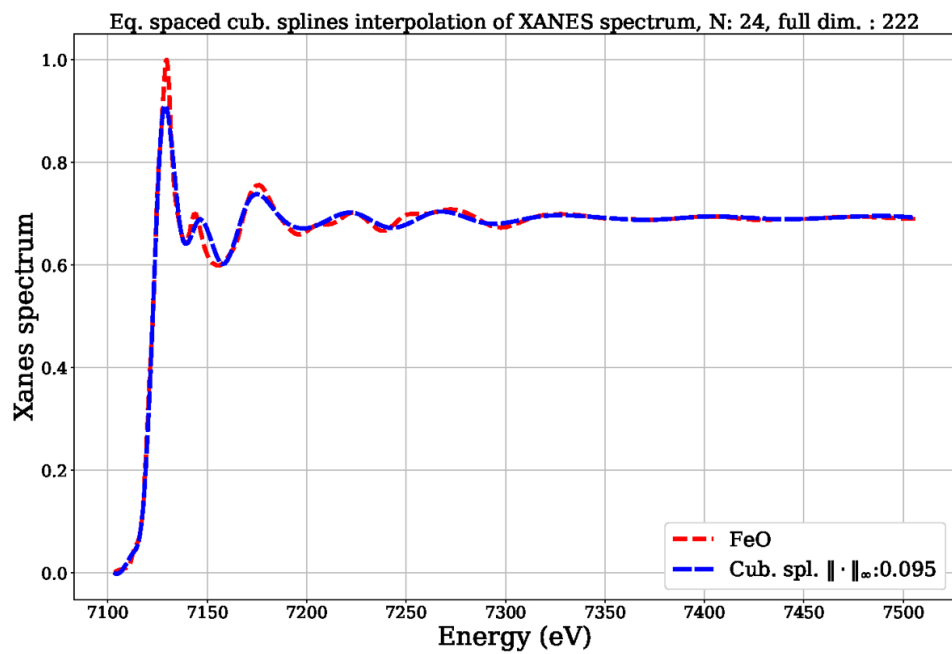
In passing, this technique gives a way to do a dimension reduction. For instance, normal polygonal interpolation with equally spaced mesh grid gives this



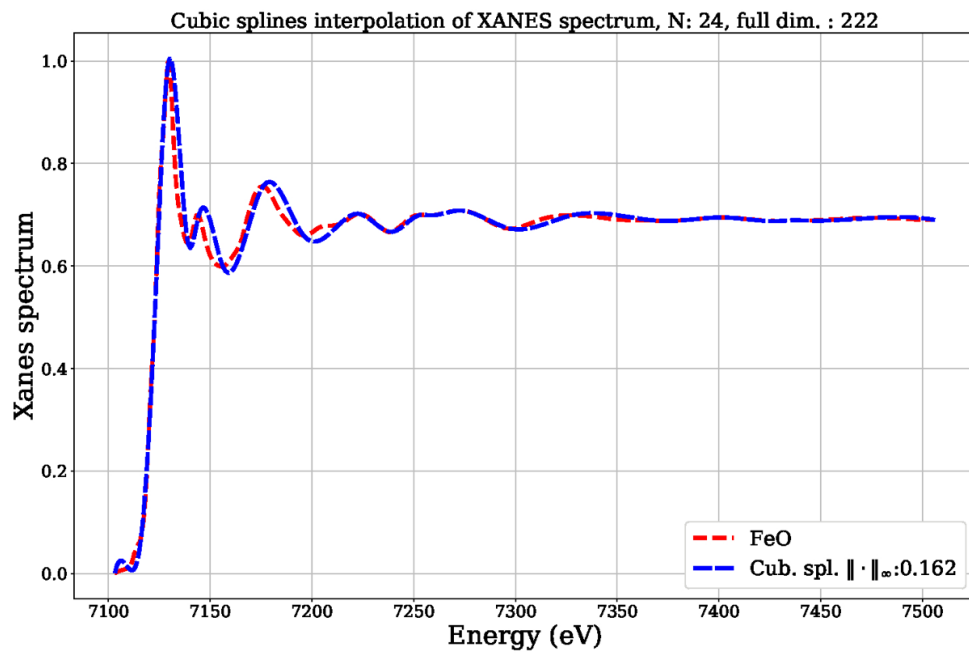
Choosing points judiciously with the Rising Sun envelope method gives a better result



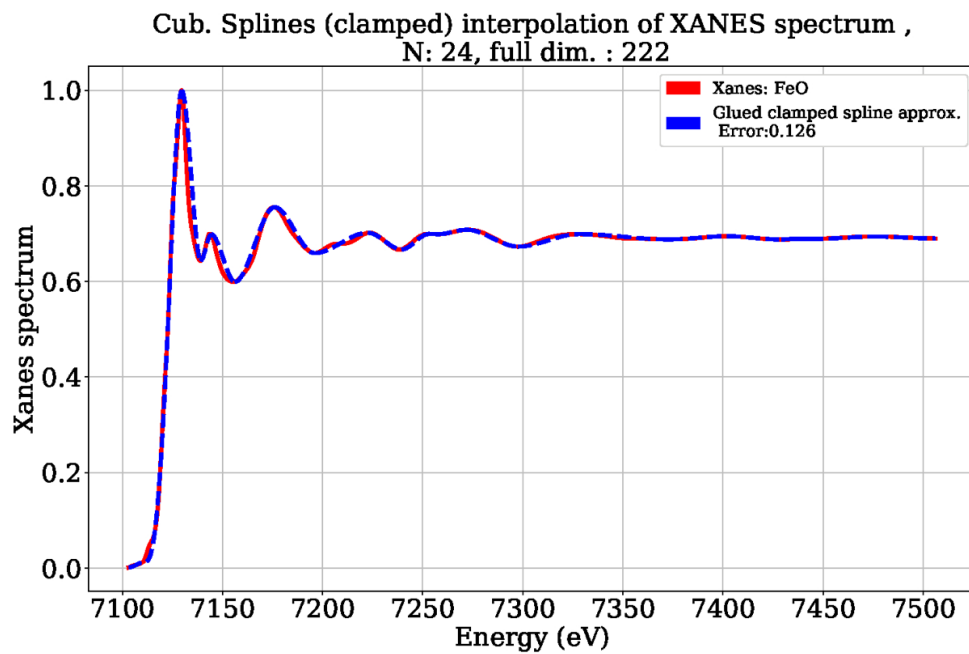
Normal cubic spline interpolation with equally spaced mesh grid gives this



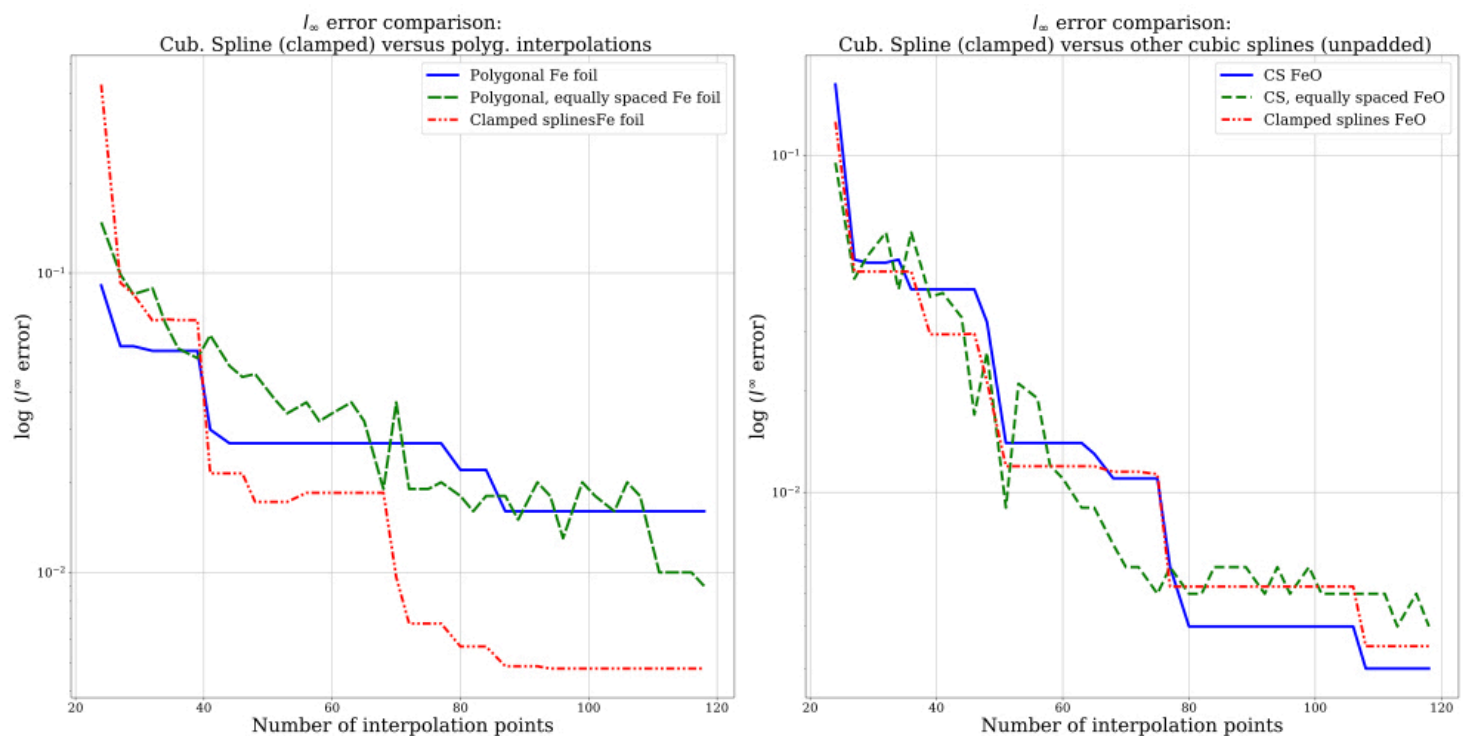
Using the points chosen by the Rising Sun envelope method gives gives this (still not satisfactory)...



...but it can be fixed using clamped splines



The error can be estimated and shown to be much lower than equally spaced interpolation



## References

Miyazato, Itsuki, Lauren Takahashi, and Keisuke Takahashi. 2019. "Automatic oxidation threshold recognition of XAFS data using supervised machine learning." *Molecular Systems Design & Engineering*.

Stein, Elias M., and Rami Shakarchi. 2005. *Real analysis*. Vol. 3. Princeton Lectures in Analysis. Princeton University Press, Princeton, NJ.