Hi David and Matt!

This is the brief summary of what I have done. There are two main things:

1. Using VNIR data to predict the amount of bastnaesite.
2. Unmix different REE- bearing mineral spectrum in ‘DT’s custom REE- mineral library’, using multiple Gaussian modeling.

And the brief- illustrated results are:

1. The largest R squared of estimation is about 0.923 (small absorptions used in big absorption area, like 6 absorptions have been considered in 700nm- 775nm)
2. Multiple Gaussian modeling could fit well in some degree. The absorptions’ position could be extracted, including smaller ones.

Here come details of the algorithm I used and why I used them.

1. **REE-mineral amount estimation.**

The whole workflow is:

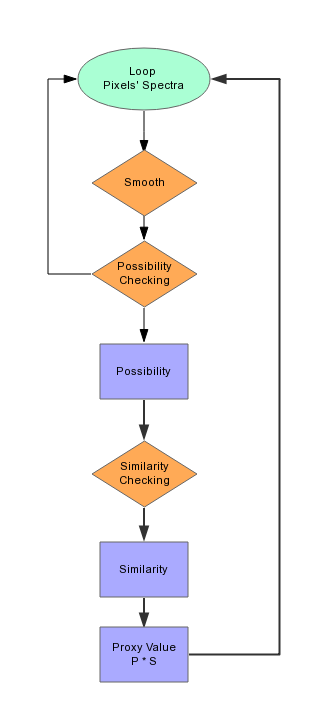


Fig1. Flow chart of whole process

1. **Loop all pixels**Loop the whole picture (totally, there are 6 rock samples) and get the spectrum from pixel.
2. **Smooth the spectrum**  
   there are too many noises that would impact the following result.  
   And these are smooth results. (Different Pixels’ spectrum)

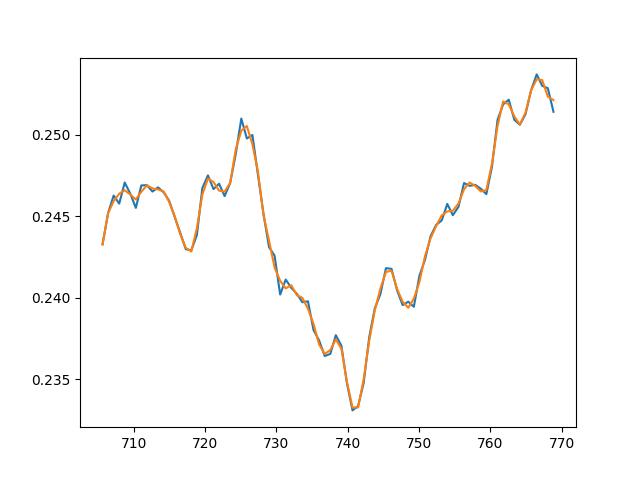
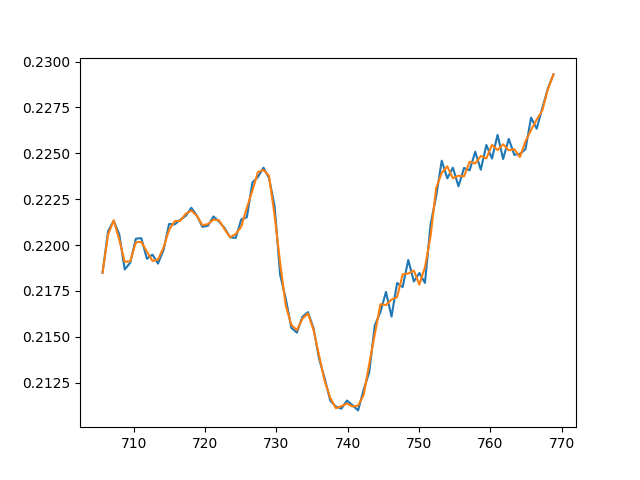
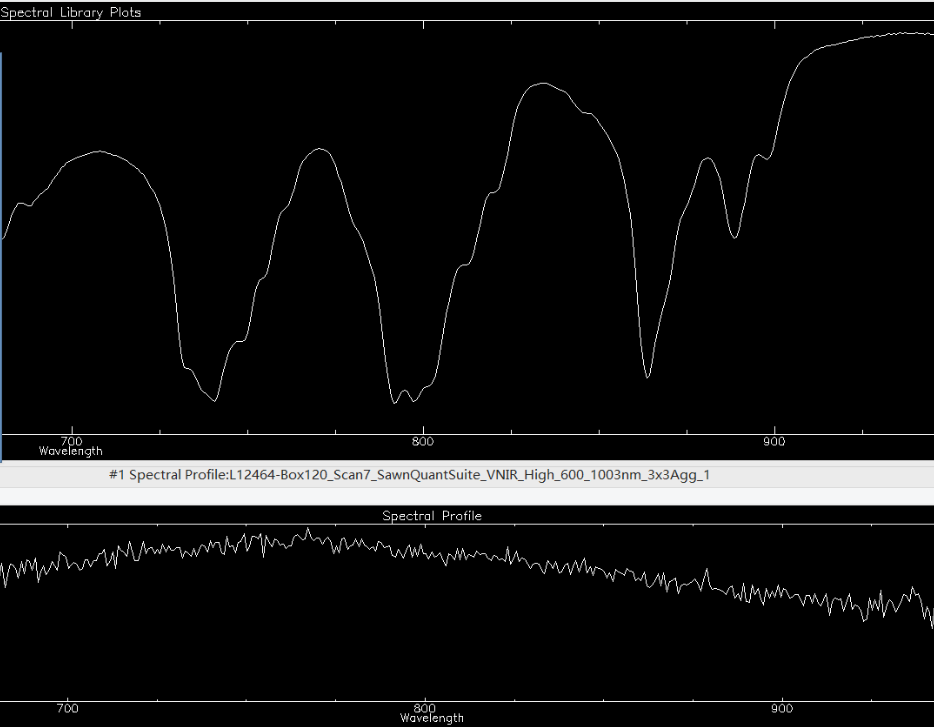


Fig 2. Smooth results

1. **Possibility Checking**

This procedure is to differentiate spectrums which is possible to be ‘bastnaesite’ and which is not (Lower one is possible, upper one is not.)



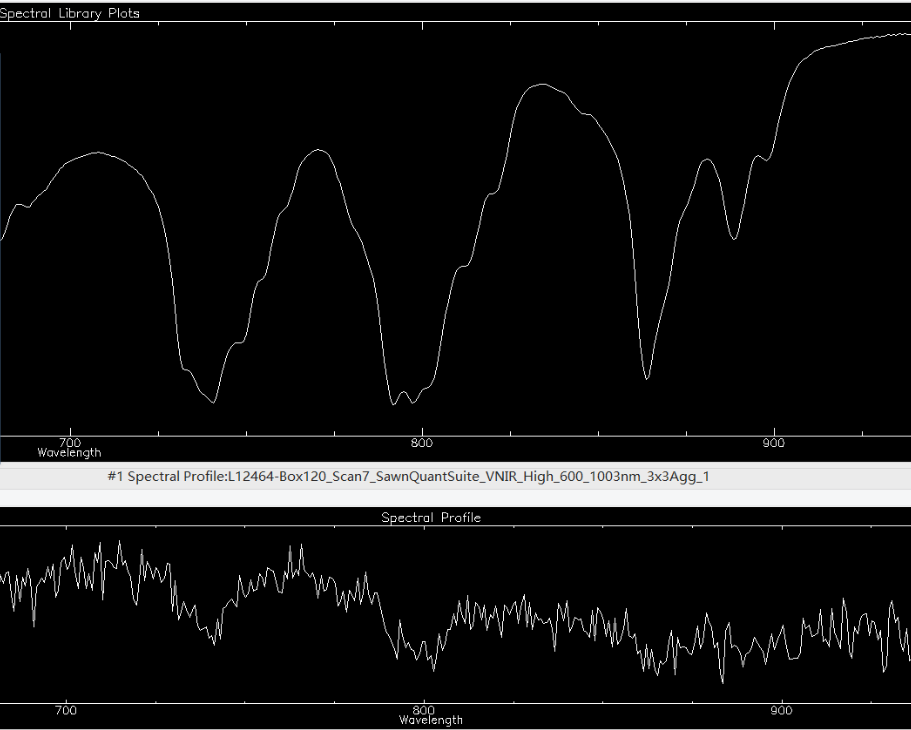


Fig 3. Lower one is possible while upper one is not.

The algorithm’s procedures I used to achieve this target (Possibility Checking) are:

1. **Depth checking**  
   Check the depth of big absorption area, if depth larger than **threshold** (Used DT’s test result, < Email to Benoit and Process>, 0.0075). Continue to do 2). Else, this area is not possible to be ‘Bastnaesite’ (or other minerals)

**Depth = (LeftShoulder + RightShoulder)/2 – Absorption Minimum  
Depth = ((b1+b2+b3)/3 + (b7+b8+b9)/3)/2 – (b4+b5+b6)/3**

* Three absorptions tested in VNIR (Unit: nm):  
  a. Absorption b1 b2 b3 b4 b5 b6 b7 b8 b9  
  b. 741(Nd3+) 705 706 707 740 741 742 767 768 769  
  c. 796 (Nd3+) 770 771 772 795 796 797 830 831 832  
  d. 864(Nd3+) 854 855 856 863 864 865 877 878 879  
  e. 889(Nd3+) 880 881 882 888 889 890 892 893 894

1. **Less inflection point checking:**   
   This procedure is to test whether multiple absorptions (smaller one included) exist in ‘Band’. If there is a absorption in that area, add weight of this absorption to Possibility.  
     
   Existence Checking: Get the extremum (inflection point), if its position is in that absorption area, this absorption exists.  
     
   Possibility =  if absorption ‘i’ exists.  
   1 =   
     
   **if the possibility > 0, we continue the algorithm (do similarity checking), else return 0 as proxy value for this pixel.**

For example, the weight and inflection points (minimum) in band1 (705-769nm):

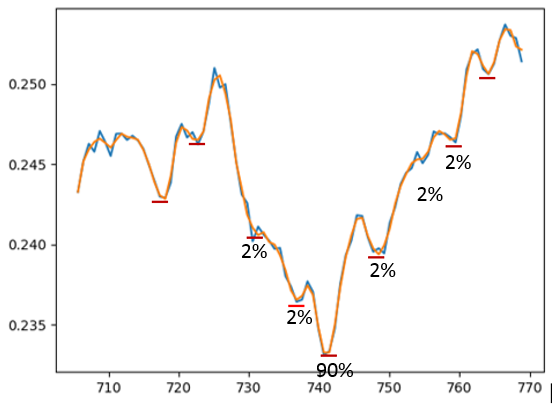


Fig4. Band1 spectrum of one pixel. Possibility Checking Result: 98%

* Band and weight list of Bastnaesite

a. Band range centers weight   
b. Band1 705 – 769 732, 736, 740, 750, 753, 759 2%, 2%, 90%, 2%, 2%, 2%

c. Band2 770 – 832 783, 791, 797, 803, 810, 820 2%, 90%, 2%, 2%, 2%, 2%

d. Band3 854 – 879 863, 868, 873 5%, 90%, 5%

e. Band4 880 – 894 889 100%

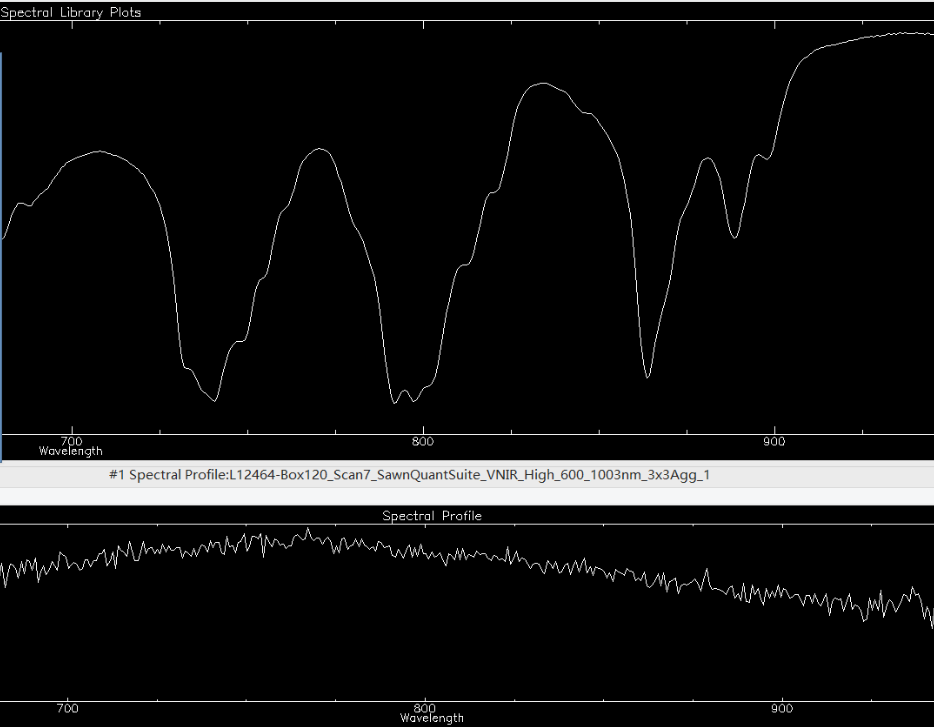
\*See more results using different weights, at ‘Result’ part

**3) Return Possibility**

After ‘Possibility Checking’, we got a possibility, which illustrates how much possible this pixel is to be Bastnaesite

**4) Similarity Checking**

This procedure is to differentiate pixels containing different amount of Bastnaesite.



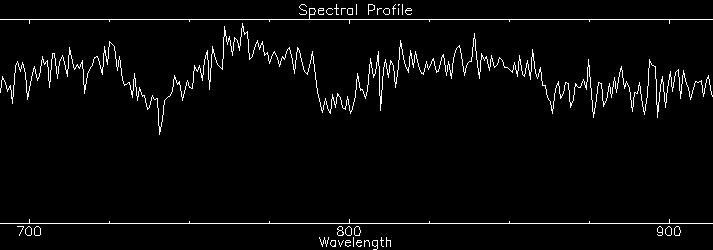
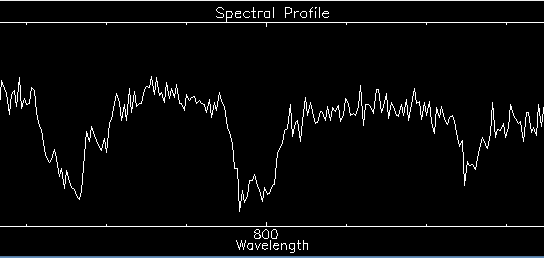
 

Fig5. Similarity Checking, left one should have lower similarity than the right one.

In this step, we mainly want to compare similarity between testing spectra (pixels) and reference spectra (library). Mostly, similarity comparing methods in math or signal processing were used.

In order to achieve this target, these methods were used:

1) Frechet distance

In mathematics, the Fréchet distance is a measure of similarity between curves that takes into account the location and ordering of the points along the curves.

2) Hausdorff distance

In mathematics, the Hausdorff distance, or Hausdorff metric, also called Pompeiu–Hausdorff distance, measures how far two subsets of a metric space are from each other. It turns the set of non-empty compact subsets of a metric space into a metric space in its own right. It is named after Felix Hausdorff.

3) Correlation Coefficients

A correlation coefficient is a number that quantifies a type of correlation and dependence, meaning statistical relationships between two or more values in fundamental statistics.

4) General method

Similarity \* Testing\_spectra = Reference\_spectra

And use least square to get the best similarity.

5) Ignore ‘Similarity Checking’.

Later, it is found that we cannot get the result what we expected, similarity values of all pixels are pretty similar, so ‘Possibility Checking’ plays the main role in this algorithm, still looking for solution.

**5) Proxy value**

At last, every pixel got a similarity and a possibility.

Proxy value (pixel) = Possibility \* Similarity

Proxy value (whole rock) = / total number of pixels in each scene

**6) Results**

* Prediction Result

Hausdorff Distance predicting results are showed here as an example. Others methods (mainly differentiate in ‘Similarity Checking’) are pretty similar, because the problem mentioned below and ‘Similarity Checking’ didn’t get what we expected.

Band1 is always the best predicting result in all methods.

Band2 cannot get good prediction result. Reason haven’t got yet.

Band3’s prediction R squared is around 0.85.

Band4 cannot be detected using this method, here is the reason I analyze: it’s kind of small and don’t clear in mixing spectrum.

Fig 6. Predicting Result using band1 (705-769nm)

1.9 is the amount in whole scene. 1.8129 is the predict result. R squared is 0.9226.

Fig7. Predicting Result using band2 (770- 832nm)

Fig8. Predicting Result using band3 (854- 879nm)

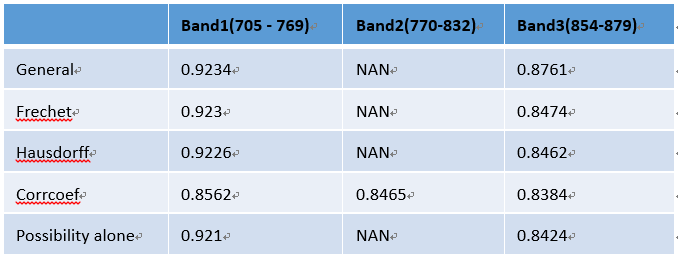
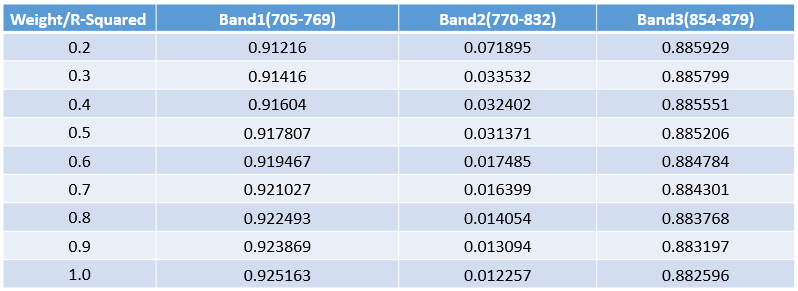
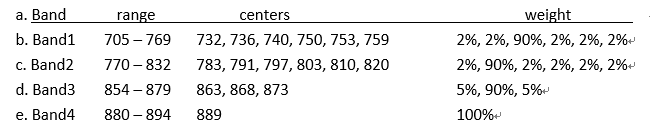


Fig9. Prediction R squared of all methods.

It should be noticed that using Correlation coefficient, every band could get a ‘kind of good’ R squared, including band2 which is always bad using other method.

* Sensitivity Detection – different weights of small absorptions in band.



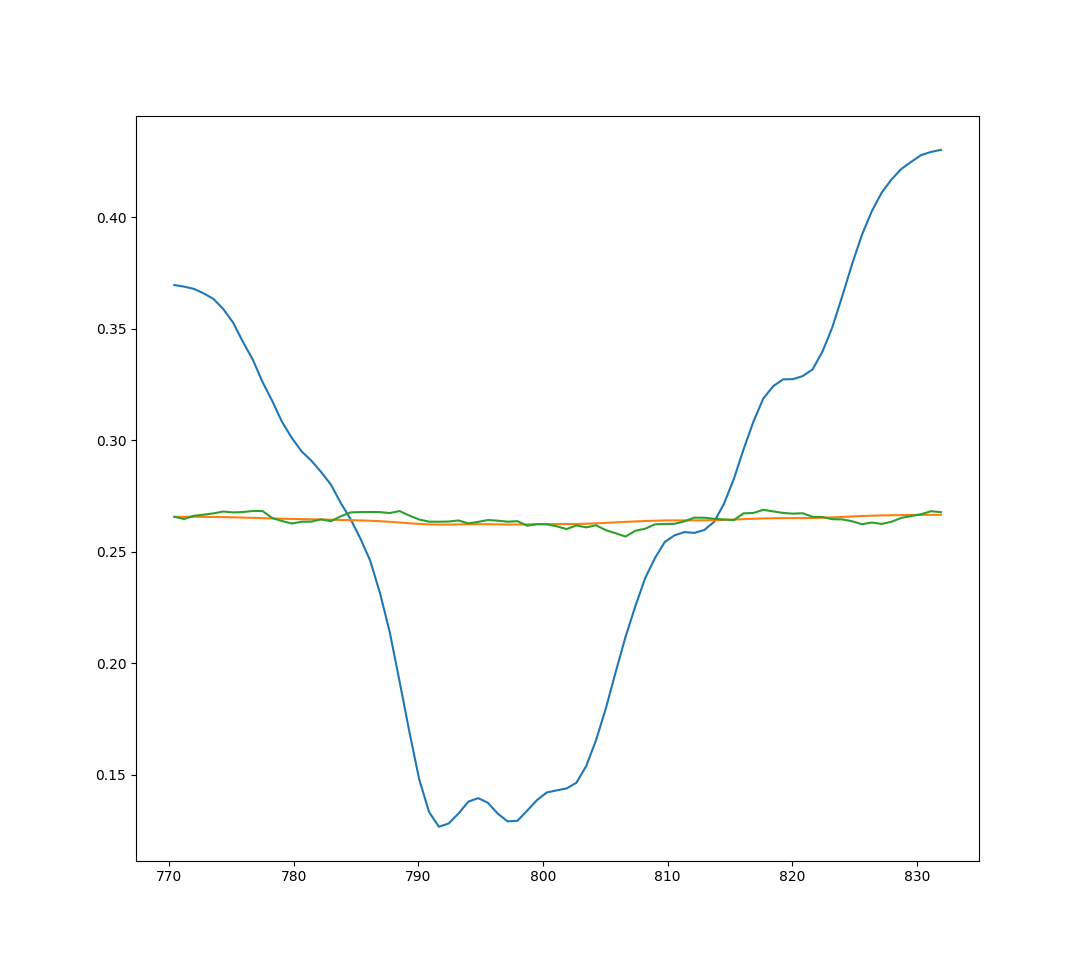


Weight is centers’ (741 796 864 889) weight, other small absorptions’ weight is

From result of different weight, we could say that this algorithm is kind of stable, not very sensitive to different weight. As the center’s weight increasing, band1 got a better prediction result whereas band3 has a contrary result.

* Problem in ‘Similarity Checking’

All ‘possible’ pixels has similar similarity, cannot be differentiated. This is because, the absorption is kind of small, comparing to reference spectrum. So we should enlarge it first and then do ‘Similarity Checking’, but scale it directly didn’t work. Solutions is still being found.

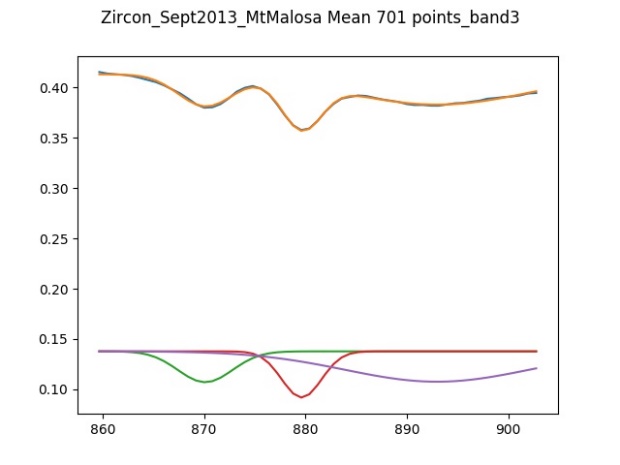
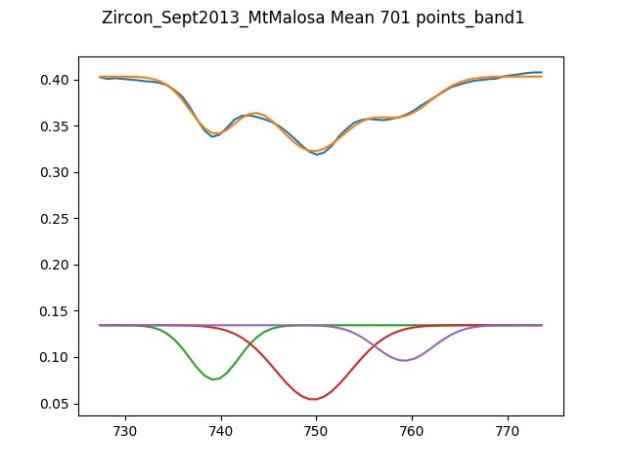


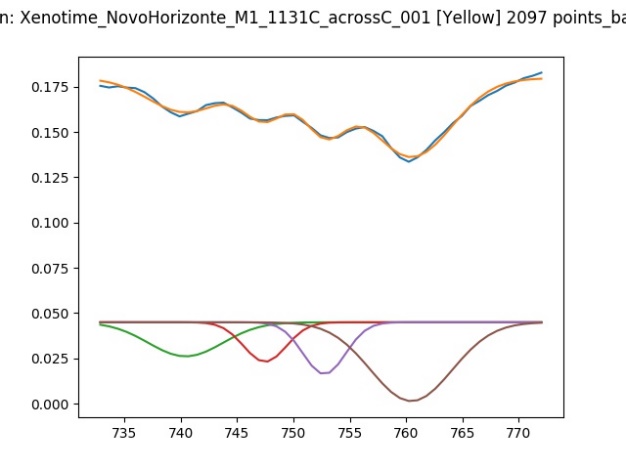
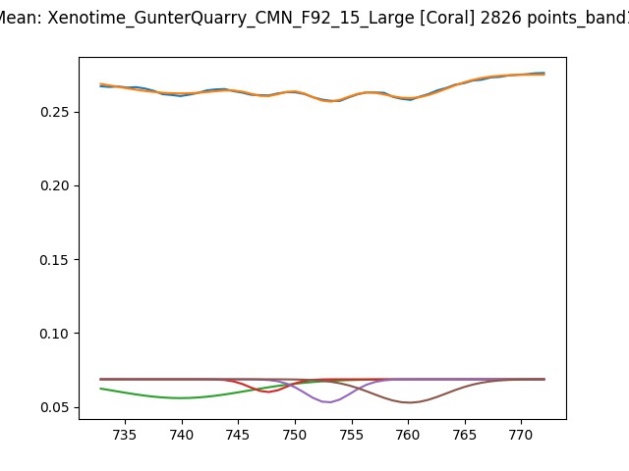
1. **Dominant mineral spectrum unmixing**

In this part, we mainly used multiple Gaussian model to fit the minerals’ spectrum in library. Modified Gaussian Model (MGM) was considered, but it seems Gaussian modeling is good enough so MGM wasn’t used.

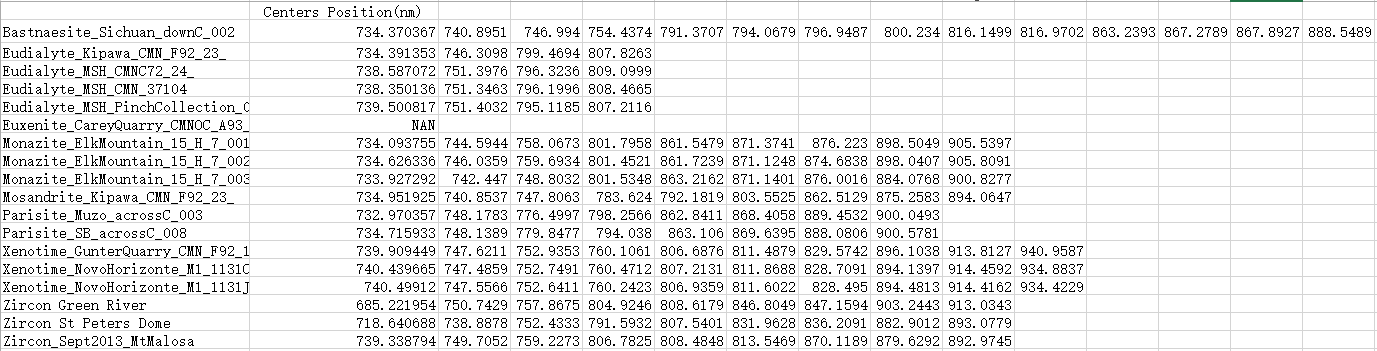
Least Square is used to get the best fitting parameters for multiple Gaussian models, initial parameters is read from a pre-written txt file.

These are fitting results (picture) of all minerals in ‘DT’s custom REE library’, all other results could be found in ‘REE fitting results.zip’.





Although we have extracted the multiple Gaussian parameters of all minerals, like height, width and center. It is still a problem to show how they are distributed, because there are too many parameters (which are dimensions in plot), the number of which is far more than two. In order to lower the dimension, PCA and LDA have been considered, but it didn’t work well (or match our problem exactly). And this is centers distribution of different minerals:



About results files:

Initial multiple Gaussian parameters could be found in ‘initialParams\_Minerals.txt’, the optimal fitting parameters could be found in ‘OptParams\_Minerals.txt’. True wavelength of minerals’ ‘band’ could also be found at ‘initialParams\_Minerals.txt’

All excel analysis could be found at ‘REE-patterns-vincent-internship.xlsx’