

## Introduction

This note deals with an important deficiency in our TSG TIR unmixing algorithms. It explains the problem and suggests a research project to define scaling parameters which should substantially improve the accuracy of the unmixing weights these algorithms produce. When available they will be easy to include in existing TIR unmixing algorithms.

## Current Scaling in TSG.

Figure 1 shows Quartz spectrum and two Montmorillonite spectra selected from the USGS library. The quartz spectrum (in green) has a dynamic range of almost the full range. The montmorillonite in red has a dynamic range about half that of quartz and that of the fine-grained montmorillonite shown in black has one tenth the range. As these spectra were measured with a carefully controlled system and we can assume that the relative magnitude of the reflectance of these samples is a reasonable estimate of the actual reflectance.

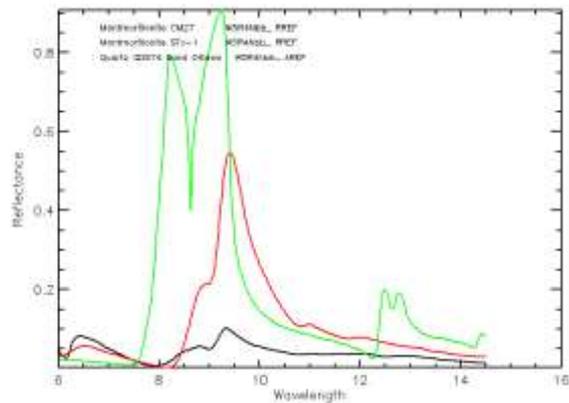


Figure 1 USGS Library spectra for quartz and two montmorillonites

Figure 2 shows the reflectance of a quartz sample and a montmorillonite sample from the TSG TIR Reference Set. This plot shows the spectra *as they are stored in TSG*, i.e. before they were put into TSG they were rescaled so they are on a 0 – 1 range. It is easy to see that the montmorillonite has been multiplied by a much larger factor than the quartz because we can see the noise much more clearly on the rescaled montmorillonite spectrum.

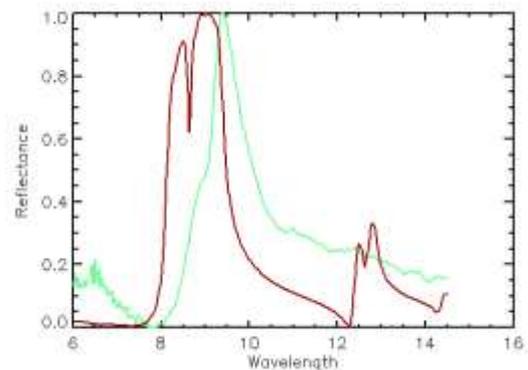


Figure 2. Rescaled quartz and montmorillonite spectra from the TSG Reference Set

The examples in Figure 1 show that while it is clear that the reflectance peak of a coarse-grained quartz sample will always be much greater than that of a fine-grained montmorillonite it is not easy to choose which montmorillonite sample to include in a reference set. This uncertainty about absolute reflectance permeates the selection of every reference spectrum in the TIR Reference Set and led to the decision to use rescaled spectra in the preliminary development of the TIR unmixing codes. It was always thought that a more sophisticated scaling would necessary in a subsequent stage.

The effect of this 0 – 1 rescaling on unmixing weights is seen in the somewhat extreme simplified example illustrated in Figure 3. Suppose we have a sample composed of quartz and montmorillonite each of which have true reflectance spectra shown in Figure 3a. Then,

assuming the rock is 80% montmorillonite and 20% quartz the resulting mixed spectrum is shown in Fig. 3b in black with the contributing effects of the quartz and montmorillonite shown in green and red.

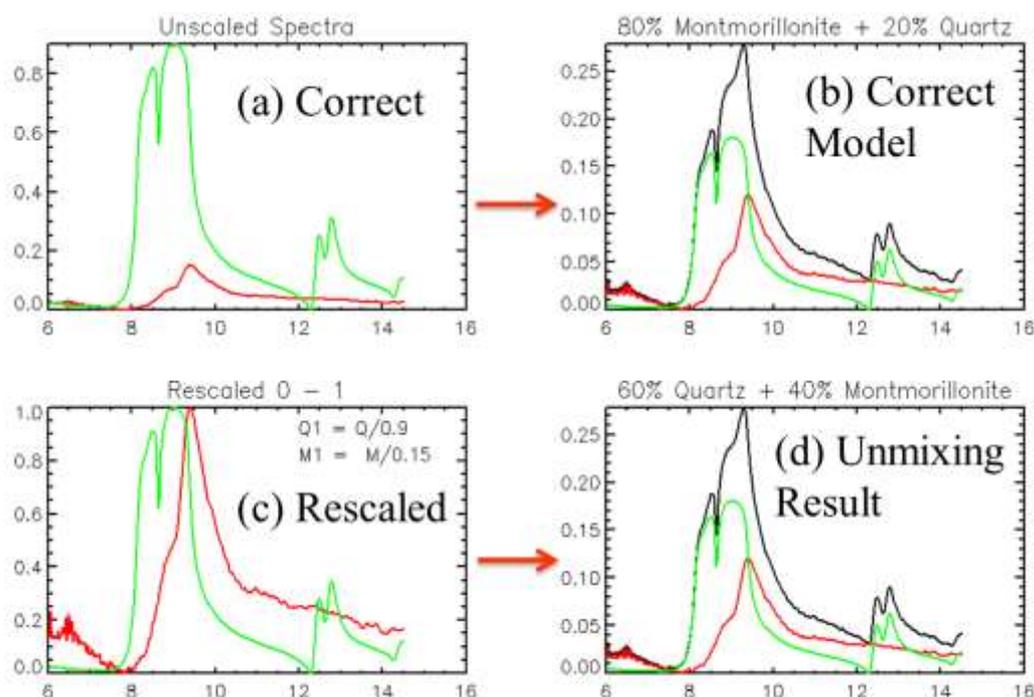


Figure 3 A simplified example to illustrate the effect rescaling on unmixing results.

When the rescaled spectra are used in the unmixing model the observed spectrum is reproduced exactly as are the contributions from the quartz and the montmorillonite (Fig.3d). However the proportions of the rescaled spectra that are used to create this mode are entirely different from the correct result. The quartz is three times the correct value and the montmorillonite is half the correct value.

Thus, scaling reference the spectra 0 - 1 causes:

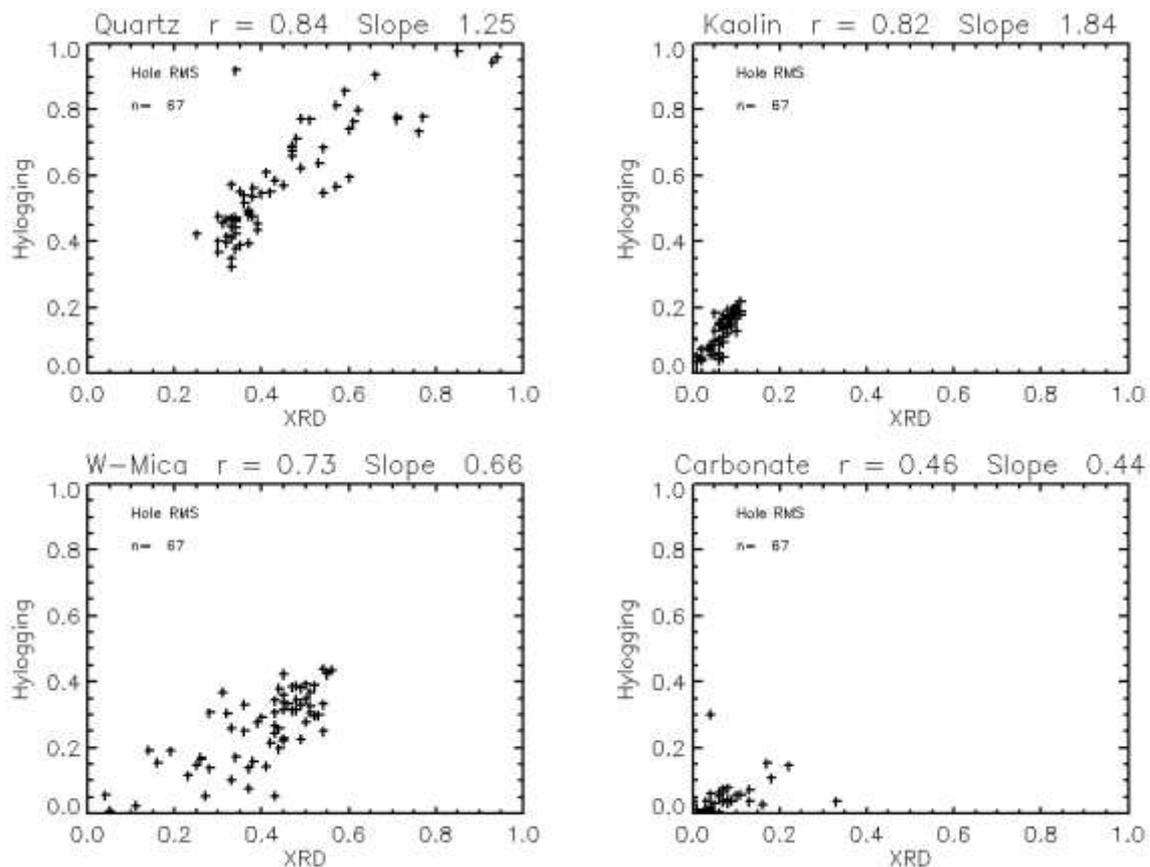
- the underestimation of weak reflectors and
- the overestimation of strong reflectors

It is important to note that it is only the relative unmixing weights that are affected by this problem. The quality of the fit and mineral selection process (e.g. TSA Subset Selection) will be exactly the same with either unscaled or scaled spectra.

## Performance of Current Algorithms

Given the simplistic assumptions explained above it is surprising that our current unmixing algorithms perform as well as they do. But we certainly have ample evidence of reasonable performance for minerals that are the major component of the rock. Given this we should expect that corrections for the current scaling will not make really gross changes to the existing results.

Alan Mauger's AJES article on Holdfast<sup>1</sup> showed that rescaling (through linear regression) established good relationships between the Hylogger results and XRD results<sup>2</sup>. In the figure below I have used Alan's XRD data and my own version of CLS and the results are consistent with Alan's except that there is no rescaling. As expected from our TIR mixing model, the relationships are linear passing through the origin but quartz and kaolin are overestimated and the white mica and carbonate are underestimated due to the 0 – 1 rescaling discussed above.



In other work on the Boddington hole WRD13075<sup>3</sup> found good correlation between XRD and the unmixing results for the more abundant minerals. But this work is marred by the fact that

<sup>1</sup> HyLogging unconventional petroleum core from the Cooper Basin, South Australia, A. J. Hill & A. J. Mauger AJES, 2016, Vol. 63, No. 8, 1087–1097

<sup>2</sup> However the results could be even better because the comparison has some problems because the samples used for XRD were collected before the HyLogger data was acquired and usually only adjacent HyLogger measurements could be used. Correct sample location has emerged a key issue in every HyLogger/XRD comparison study I have seen so far except the Boddington work. It is the reason I advocate the use of bulk sampling for XRD *after* HyLogging. When the XRD represents the average mineralogy over a reasonable depth range (say 0.5 m) then there are enough HyLogger samples to fully sample this range and allow the averaging of the HyLogger samples match that of the XRD

<sup>3</sup> Boddington Unmixing Study, Andy Green 5 December 2018

the VNIR/SWIR data and the TIR data were acquired at different times and on different machines. I have another dataset from the MacArthur Basin that has extensive XRD. The correlations here are not as good as one would like but this data is badly degraded because the core was heavily altered by post-drilling oxidation and it also suffers poor sample location.

### **The Next Step**

I think now we need a project to estimate the new scaling parameters. This should involve something like the following steps.

1. Collect a diverse set XRD/HyLogger-3 data that has our major active minerals represented over as wide a range of abundances as possible.
2. Check them all for sample location and oxidization/coating problems like those that produce surface sulphides and carbonates.
3. Model the HyLogger data with CLST using the known mineralogy
4. Invert the results for unknown scaling parameters as outlined in the Appendix to this note.

Appendix - Using XRD to resolve unknown scaling in unmixing  
 Andy Green 24 April 2015

Let  $\rho$  be a data set of  $N_i$  TIR reflectance spectra with  $N_\lambda$  channels. Neglecting  $\Delta T$  effects the model for the response of the  $i^{\text{th}}$  sample is

$$b_i \mathbf{R} \mathbf{p}_i = \rho_i \quad (1)$$

Where  $\mathbf{p}_i$  are the proportions of each reference spectrum in the unknown spectrum and  $b_i$  is a constant scale factor for each unknown spectrum to reflect the fact that geometry effects can mean that samples with the same proportions can have different “brightness”.  $\mathbf{R}$  is the  $(N_\lambda \times N_k)$  matrix of *unscaled* mineral reflectance spectra.

However, in practice, we unmix with a *scaled* Reference Set  $\mathbf{S}$  by inverting the model

$$\mathbf{S} \mathbf{w}_i = \rho_i.$$

We have  $\mathbf{R}\mathbf{A} = \mathbf{S}$  where  $\mathbf{A}$  is diagonal matrix of  $N_k$  scaling coefficients  $\alpha_k$  so that

$$\mathbf{R}\mathbf{A}\mathbf{w}_i = \rho_i$$

Thus using (1)

$$\mathbf{A}\mathbf{w}_i = b_i \mathbf{p}_i$$

Letting  $a_k = 1/\alpha_k$  and  $\beta_i = 1/b_i$  we have

$$w_{ki} = a_k b_i p_{ki} \quad \text{where} \quad \sum_k p_{ki} = 1$$

At this stage we would like to introduce the XRD results to estimate the scaling coefficients. However the XRD results are given as proportions among  $N_m$  minerals and the above proportions are for spectra. As our model can have more than one spectrum for a mineral we have  $N_k > N_m$  and thus a slightly more complicated problem. These considerations have led me to assume that all the  $\alpha_k$  associated with the  $m^{\text{th}}$  mineral are equal to  $\alpha_m$  so that there are now  $N_m$  unknowns.

Because each spectrum is associated with only one mineral we have

$$\sum_{k \in m} p_{ki} = p_{mi} = x_{mi} \quad \text{and thus} \quad \sum_{k \in m} w_{ki} \equiv w_{mi} = a_m b_i x_{mi} \quad (2)$$

Where  $x_i$  is the vector of  $N_m$  XRD results for the  $i^{\text{th}}$  sample with  $\sum_m x_{mi} = 1$ . Because of this

last constraint we also have  $\sum_m \frac{w_{mi}}{a_m b_i} = 1$  and thus

$$\sum_m \frac{w_{mi}}{a_m} = b_i \quad (3)$$

Also from (2) and (3)

$$\frac{\alpha_m w_{mi}}{\sum_m \alpha_m w_{mi}} = x_i$$

If the  $\beta$ s (and the  $b$ s) are initially set to 1 each  $a_m$  can be estimated by a regression ( $b_{mi} w_{mi}$  vs  $x_{mi}$ ) over all  $i$ . Then this process this can be iterated with (3) for the  $\beta$ s and new  $a$  after normalizing the  $\alpha$ s to have an average of one.

This algorithm has been implemented in the XRDRESALE routine and has been tested on the Boddington data set as illustrated by the following two figures. The first is the CLS uncorrected unmixing results (y-axis) vs the XRD results (x-axis) and the second shows the results after correction. The  $\alpha_m$  values for each mineral are shown of these plots.

