

AN AUTOMATED SPECTRAL DECONVOLUTION ALGORITHM: APPLICATION TO THERMAL INFRARED STUDIES OF EARTH AND MARS. N. V. Zorn¹ and M. S. Ramsey², ¹Department of Electrical Engineering, University of Pittsburgh, 348 Benedum Hall, Pittsburgh, PA 15261, nvzst@pitt.edu, ²Department of Geology and Planetary Science, University of Pittsburgh, 200 SRCC Building, Pittsburgh, PA 15260, ramsey@ivis.eps.pitt.edu

Introduction: The linear mixing of thermal infrared (TIR) emission spectra in multi-mineralic mixtures has been proven, and its limits and applicability have been quantitatively investigated [1,2]. Limiting factors in the accuracy of any linear retrieval (spectral deconvolution) algorithm include the spectral precision of the instrumentation as well as the fact that the number of end-members must be \leq the number of spectral bands in the TIR [1]. Because of this end-member constraint, there is no way to examine a multispectral TIR image using a large, spectral end-member library [3]. A possible solution is the implementation of an automated, blind end-member algorithm to analyze all possible subsets (of arbitrary size k) of minerals present in the mixture from within a mineral library (of size n). For example, in the only such study to employ this technique, the Kelso Dunes, California were examined using 375 unique combinations ($k4$) of the most likely ($n=$) 15 minerals present in the dunes [4]. The deconvolution model results were analyzed for their “goodness of fit” to laboratory spectra of collected sand samples. This investigation proved highly successful, further supporting the capabilities of linear retrieval in the TIR [4]. However, the process was computationally intensive because it was entirely manual. The development of an automated algorithm to accomplish this task would decrease the amount of time required for such an investigation by many orders of magnitude.

From a remote sensing perspective, an automated blind end-member spectral deconvolution algorithm could be useful to determine mineral abundance for any mixture (or pixel-by-pixel in a TIR image) using a large emissivity library of minerals as input. This approach is particularly timely as there multispectral TIR mapping instruments now orbiting Earth and Mars.

Algorithm: Combinatorial mathematics states that the number of unique k -subsets of n elements is given by the equation:

$$n C k = \frac{n!}{(n-k)!k!}$$

It is observed that as n increases for any given k , the result grows significantly. For example, $15 C 4 = 1,365$ and $20 C 4 = 4,845$. Algorithms have been written by combinatorial mathematicians to generate lists of these combinations with different ordering methods. In the

techniques developed here, a lexicographic (integer) ordering algorithm was implemented [5].

All routines were written in the MATLAB[®] environment (Version 6.1.0.450). This software was chosen due to its ability to efficiently manipulate large matrices. The algorithm accepts the following as input: (1) an “unknown” mixture spectrum, at high spectral resolution (518 data points); (2) a library composed of n individual mineral spectra, also at high resolution and possibly present in the unknown; (3) a user-selected maximum output subset size k ; and (4) a user-selected threshold abundance below which minerals should not appear in the modeled output (e.g. 0-5%). Concerning (4), the mathematical techniques utilized in the linear deconvolution (and explained in detail in [1]) involve a linear regression analysis that can result in negative values in the output matrix of end-member fractions. Because negative abundances are physically impossible, model results containing these are removed from further consideration. Linear deconvolution can also produce spurious results below the realistic accuracy of the model (~5%) [1].

A combinations list is then generated for $n C 2$, and each unique combination is used as indices for the modeled reconstruction of the unknown emissivity. For example, where $n=20$, the 190 resultant combinations {1,2}, {1,3}, {1,4}, ... {18,20} and {19,20} are each used to construct a narrowed end-member library, and linear deconvolution is performed on this narrowed library, resulting in a vector of abundance fractions for each combination. If any fraction is less than the low threshold parameter, the combination is removed. Then a candidate emissivity is constructed using each fraction vector, and the root-mean-squared (RMS) error (or “goodness of fit”) of the candidate is calculated [1,4]. The combination resulting in the lowest RMS error is considered the best fitting combination and stored. Next, the subset size is increased by one, and a new combination list is generated ($n C 3$ long). If a better-fitting combination of three elements is found within this list, that combination becomes the new best fit. This process is repeated until (k) is reached. Theoretically, the combination that results in the lowest RMS error will be the one that actually comprises the mixture.

Preliminary Results: Thermal emission spectra of several mixtures of known mineralic composition were used to initially test the capabilities of this algorithm.

The mixtures were composed of hornblende, microcline, oligoclase and quartz, and have been analyzed in prior investigations of spectral deconvolution [1].

Is RMS error sufficient? To conserve processor time and rigorously test the algorithm, a blind end-member library of size $n=20$ was initially used, in conjunction with the TES spectral library of 170 minerals [3]. An end-member library subset was constructed of 16 arbitrarily selected mineral spectra and the four known end-members. The k value was set at 4 for each execution, and the low threshold parameter was set at 0%. Therefore, the number of combinations examined was $(20 C 2) + (20 C 3) + (20 C 4) = 6,175$. The results for three mixtures are tabulated in Table 1.

Mixture	2		3		4	
	Real %	Mod. %	Real %	Mod. %	Real %	Mod. %
Hornblende	10	0.0	20	22.4	70	74.4
Microcline	10	12.9	20	20.5	5	9.1
Oligoclase	20	20.3	20	27.8	5	0.0
Quartz	60	53.0	40	30.2	20	15.6
Pyroxmangite	0	14.4	0	0.0	0	0.0
Biotite	0	0.0	0	0.0	0	1.6

Table 1. Algorithm output for three mixtures using high-resolution thermal emission spectra.

The modeled mixtures above exhibited the lowest RMS error, and were selected as the best-fitting combinations. The RMS errors for modeled mixtures 2, 3 and 4 were 0.0040, 0.0039 and 0.0025, respectively. The algorithm at times finds other minerals that compose a *mathematically* better fit to the unknown than the actual combination of minerals. This suggests that RMS error alone may not be capable of judging absolute “goodness of fit.” Also, it is observed that some minerals with abundance $\leq 15\%$ are not always detected, or other minerals are detected in their place. This was also noted in previous investigations and appears to depend not only on abundance, but on spectral features as well [1,2]. These and other characteristics of the algorithm are currently being studied in greater detail.

Algorithm runtime considerations. Performing thousands of spectral deconvolutions per unknown spectrum clearly requires a fast and capable computer. The trials performed in this study were done on a Windows® 2000 machine with a 1400 MHz Intel® Pentium® 4 CPU and 128 MB RAM. The total runtime of the algorithm was measured for end-member

libraries of different sizes, ranging from 4 to 50, and $k=4$. In the 20 end-member library trials (6,175 total combinations) the algorithm ran for ~8 seconds. In a 50 end-member library trial (251,125 total combinations) the algorithm ran for ~4,032 seconds (67 minutes). Further investigation into the effects of increasing the number of end-members in the spectral library is underway. Ideally, the algorithm should be able to choose from the entire library of recorded thermal emission spectra and quickly produce a resultant fit.

Effects of degraded spectral resolution. The effect of degrading the spectral resolution of the mineral and mixture spectra is also of interest, given that the Mars-orbiting Thermal Emission Imaging System (THEMIS) instrument only has 10 spectral bands in the TIR (6.5 and 14.5 μm). The capabilities of a linear retrieval algorithm at this resolution were tested by degrading the spectral resolution of the end-member and mixture spectra to the approximated THEMIS band centers. The resulting fractional abundances are tabulated in Table 2 (compare to the actual compositions of mixtures 2-4 in Table 1). The RMS errors for modeled mixtures 2, 3 and 4 were 0.0041, 0.0033 and 0.0016, respectively. Further analysis of the accuracy at low spectral resolution is also underway. If successful, this algorithm will be implemented for imaging instruments such as THEMIS in order to identify mineral percentages and track sediment transport over time on the surface of Mars and Earth [3].

Mixture	2	3	4
<i>Mineral</i>	<u>Mod. %</u>	<u>Mod. %</u>	<u>Mod. %</u>
Hornblende	5.4	21.5	70.0
Microcline	10.3	19.3	18.4
Oligoclase	32.3	29.2	0.0
Quartz	52.9	30.7	11.9

Table 2. Spectral deconvolution results using 10-band thermal emission spectra.

References: [1] Ramsey M. S. and Christensen P. R. (1998) *JGR*, 103, 577-596. [2] Feely, K.C. and Christensen P. R. (1999) *JGR*, 104, 24,195-24,210. [3] Christensen P. R., et al. (2000) *JGR*, 105, 9735-9739. [4] Ramsey M. S., et al. (1999) *GSA Bulletin*, 111, 646-662. [5] Kreher D. L. and Stinson D. R. (1999) *Combinatorial algorithms: generation, enumeration, and search*, Boca Raton, CRC Press, 43.