Statistical clustering and Mineral Spectral Unmixing in Aviris Hyperspectral Image of Cuprite, NV

Mario Parente, Argyris Zymnis

I. Introduction

Hyperspectral Imaging is a technique for obtaining a spectrum in each position of a large array of spatial positions so that a recognizable image is obtained at each of a set of discrete wavelengths. The images might be of a rock in the laboratory, a field study site from an aircraft or a rover camera, or a whole planet from a spacecraft or Earth-based telescope. By analyzing the spectral features (generally neighborhoods of local minima in the spectra) one can map materials. A simplistic explanation for this being is that specific chemical bonds in different materials manifest themselves as absorption features at different wavelengths and by mapping where those bonds occur in the spectra one can uniquely identify what is called the unique *spectral signature* of the material . The factors affecting spectra of natural materials and the causes of absorption features are several and combine in complex ways. They are not the focus of this paper but a comprehensive tutorial can be found in [3]. Spectral unmixing is the procedure by which the measured spectrum of a pixel is decomposed into a collection of constituent spectra, or *endmembers*, and a set of corresponding fractions, or *abundances*, that indicate the proportion of each endmember present in the pixel. In the case of rocks or soils the endmembers can be consistent with the minerals present in the geologic surface observed. In this work we present a novel technique of endmember selection from a database of minerals based on simple convex optimization techniques.

Spectral unmixing can assume linear or nonlinear combination of the endmembers depending on the nature of the surface observed [9]. Unfortunately nonlinear schemes can be impractical for hyperspectral imaging because multiple views from different angles of the same scene are required [11].

Linear Spectral Unmixing is based on the assumption that the spectrum of each pixel of the scene is a convex combination of the spectra of its component minerals [13]. Deterministic modeling of the mixture lacks the ability to explain the statistical variability of the spectra within a class due for example to illumination differences, altimetry, grain size of the material and other causes. Several attempts have been made to correct this problem: these approaches allow the endmembers of the mixture to be random variables (mostly Gaussians) [6], [14].

The present work assumes a mixture of multidimensional pdf's for the statistical distribution of the spectra of the single pixels composing the scene. Each pdf (a multinomial Gaussian) represents the likelihood of a certain mineral mixture in the scene. We make use of the Gauss Mixture Vector Quantization algorithm [1], [7] as an alternative to the EM algorithm [5]to learn the mixture parameters. We also explore cluster analysis with correlation distance [8].

The clustering stage is useful to select only a few centroids each to be representative of the image and to perform only on them the unmixing as opposed to the whole image. This decreases the computational cost of the processing, especially when a large number of images is acquired.

Each gaussian mean produced by the clustering stage, thought to be representative of a specific mineral mixture, is unmixed in this work by a constrained least square algorithm which can be cast as a quadratic program.

II. Aviris Cuprite dataset

Spectral data collected over Cuprite, Nevada, USA, have been widely used to evaluate remote sensing technology and for spectral unmixing [4], [15] and references therein.

For the purpose of our study 50 out of the Aviris bands (172 - 221, or 1.99 to 2.48 μm) have been selected because of the better discrimination of mineral signatures in that range.

In figure 1(b) there is a plot of the spectra that represent the spectral variability in the image. Those were obtained by considering the image as a 50-dimensional data cloud and capturing its corners.

III. Clustering

Our goal in clustering is to pick up as much as possible of the spectral variability in the image. We tried some statistical measures of cluster validation such as the Gap statistics. By polling several experts in mineral identification we assessed that the statistical optimal number of clusters is too low to capture the variability in the data that those experts would consider if they scanned the pixel spectra.

Since this is a completely unsupervised classification and for the reasons just mentioned we consider capturing the most spectra of figure 1(b) as a reasonable goal for the clustering stage and to select the number of clusters.

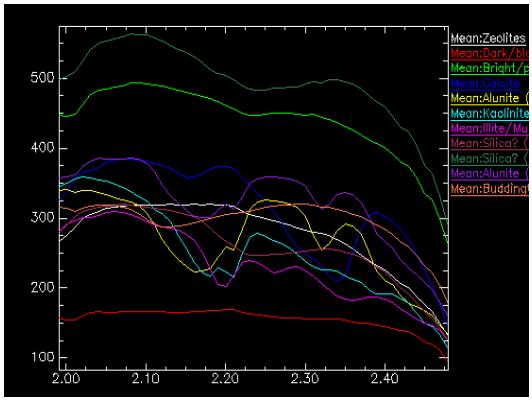
We experimented with 3 different setups: K-means, GMVQ and cluster analysis with the correlation measure.

We try for each k (number of clusters) 5 runs with random initial point and choose the run with minimum value for objective function (distortion or distance).

A. Lloyd Clustering Algorithm for Gauss Mixture design

The Gauss Mixture Vector Quantization algorithm can be seen as an alternative to the EM algorithm for fitting a finite

(a) RGB composite of bands 183, 193 and 207



(b) spectral variability

Fig. 1. Aviris hyperspectral image of Cuprite, NV

Gauss mixture $\{p_m, g_m\}$ to a training set $\{x_1, x_2, \ldots, x_N\}$ (see [1] section 4 and [10].

The design of a Gauss mixture implemented by GMVQ is as follows [1]:

Minimum Distortion (Nearest Neighbor) Step: For $i = 1, \ldots, N$ encode each training vector x_i into the index $\alpha^{c+1}(x_i)$ corresponding to the minimum distortion Gaussian model $f_m(x_i|\theta_m^c)$, that is [1],

$$\begin{aligned} \alpha^{c+1}(x_i) &= \arg\min_m \left(-\ln f_m(x_i|\theta_m^c) + \lambda \ln \frac{1}{p_m^c} \right) \quad (1) \\ &= \arg\min_m (\frac{1}{2} \ln |K_m^c| + \quad (2) \\ &\quad + \frac{1}{2}(x - \mu_m^c)^T (K_m^c)^{-1} (x - \mu_m^c) - \lambda \ln p_m^c). \end{aligned}$$

where μ_m^c and K_m^c are the current estimates of the mean and covariance of the distribution of the samples.

Centroid Step: Given all the vectors $\{x_1, x_2, \ldots, x_{N_m}\}$ belonging to the m-th partition determine a gaussian density (i.e. its parameters μ_m^c and K_m^c) $f_m(x_i \theta_m^c)$ so as to minimize

$$\sum_{i=1}^{N_m} \left[\frac{1}{2} \ln |K_m^c| + \frac{1}{2}(x_i - \mu_m^c)^T (K_m^c)^{-1} (x_i - \mu_m^c) \right] \quad (3)$$

The minimization can be performed by alternatively fixing μ_m^c or K_m^c and minimizing over the other one [1]. The solutions are the sample mean in the m-th partition

$$\mu_m^{c+1} = \frac{1}{N_m} \sum_{i=1}^{N_m} x_i \quad (4)$$

and the sample covariance in the m-th partition

$$K_m^{c+1} = \frac{1}{N_m} \sum_{i=1}^{N_m} (x_i - \mu_m^c)(x_i - \mu_m^c)^T \quad (5)$$

The algorithm includes one more step to calculate the optimal length function $\ln \frac{1}{p_m^c}$,given the partition and the centroids [1].

We proposed a variation on the GMVQ algorithm that discards the penalty term $\lambda \ln \frac{1}{p_m^c}$ in the distortion measure (by considering $\lambda = 0$) because we observed from simulations that algorithms that use measures of membership probability of each cluster (like the prior for each cluster in EM and optimal length function for GMVQ) penalize too much the clusters with fewer assigned sample.

1) Euclidean Distance: If we set the covariance term $K_m, \forall m$ equal to the identity, the distortion measure becomes the Euclidean distance and the GMVQ Algorithm reduces to the well known K-means (or vanilla Vector Quantization).

The Euclidean distance is invariant under orthogonal transformation of the data but it does not take the correlation of the variables into account.

The results in figure 2 (left) show that the euclidean distortion is able to pick up the relatively very big (in norm) and very small clusters in figure 1(b) which stand out as solitary but only a few of the spectra in the central range are distinguished. The reason being that those spectra differ mostly in shape. We also found that the segmentation map was a little bit too fragmented.

2) Mahalanobis Distance: If the only constraint is $\lambda = 0$ then the distance measure is similar to the Mahalanobis distance (with an additional term that considers the volume of the gaussian cluster).

The classifier becomes quadratic. The increased flexibility of the boundary is traded off by the decreasing of the between-cluster variance/within-cluster variance ratio.

From figure 3 (left) we see results similar to the previous case. The reason might rely on the fact that the covariances of the clusters are very similar and we know that in case the classification boundary is linear. We also notice that the GMVQ estimation for the mean is a simple average like in k-means. The fragmentation of the segmentation map is somewhat improved by the use of the second order statistics.

B. Correlation-based Cluster Analysis

The correlation-based distance between a pixel x_i and a centroid μ is:

$$d_c(x_i, \mu) = 1 - \rho(x_i, \mu) = 1 - \frac{(x_i - \bar{x}_i \mathbf{1})^T (\mu - \bar{\mu} \mathbf{1})}{\|x_i - \bar{x}_i \mathbf{1}\|_2 \|\mu - \bar{\mu} \mathbf{1}\|_2},$$

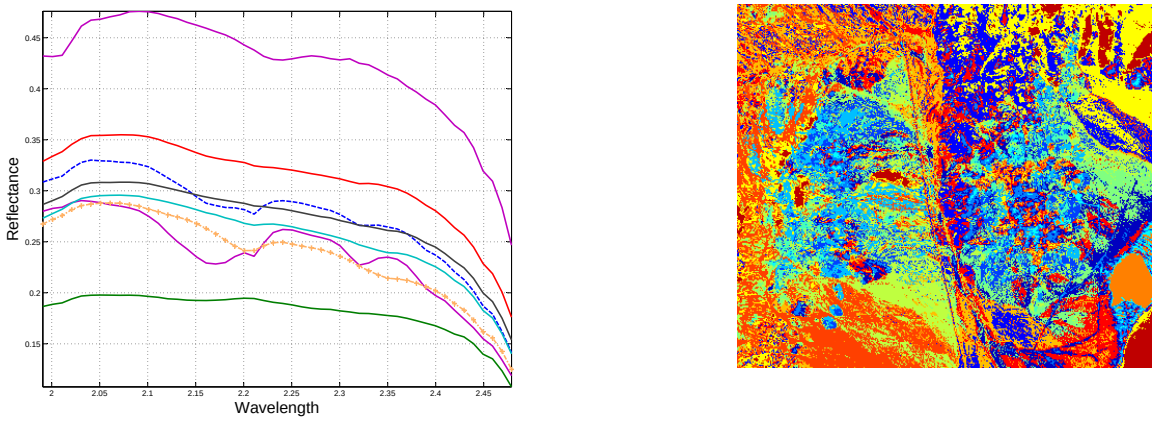


Fig. 2. Cluster centroids (left) and Cluster Map (right) for Euclidean Distance.

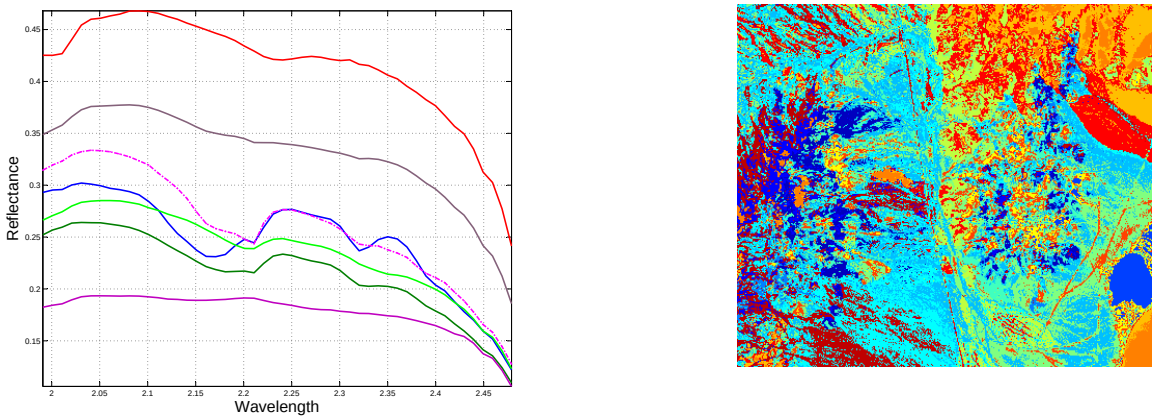


Fig. 3. Cluster centroids (left) and Cluster Map (right) for Mahalanobis Distance.

where $\bar{x}_i$, $\bar{\mu}$ are the vector means of x_i and μ respectively.

This is a measure that is invariant to scaling and shifting (vertically) of the expression values. We tried this setup to take into account the shape of the spectra considered as signals. The drawback is that the actual magnitudes of the spectra are ignored.

If the inputs are standardized, then the above distance is equivalent to the Euclidean distance. Another drawback of the distance is that $d_c(x_i, \mu) = 0$ only implies linear relationship between x_i and μ.Furthermore the centroids are not obvious to interpret.

We actually obtained the best results with the correlation-based distance, as we can see from figure 4 (left). The obvoius limitation of the measure is that the high norm cluster in figure 1(b) is misplaced because the measure is normalized. We on the other hand get almost the full variability in the data. In a development of this project we will explore a clustering algorithm based on Shape and Gain Vector Quantization that takes into account correlation (shape) and norm (gain) simultaneously. The segmentation map seems less fragmented.

IV. Mineral Identification and Unmixing

We assume that a dictionary of mineral spectra is available to us. For this particular dataset, we extracted the dictionary from [3] and [14]. Suppose that the mineral dictionary is given in D

$$D = \left[\begin{array}{cccc} v_1 & v_2 & \dots & v_n \end{array} \right],$$

where v_i for $i = 1, \dots, n$ are the individual mineral spectra. We want to find the abundances $a^{(j)}$, such that, for each cluster centroid (mean) μ_j, we have:

$$\mu_j \approx \sum_{i=1}^{n} a_i^{(j)} v_i = D a^{(j)} \tag{6}$$

in a least squares sense.

Since it is unreasonable to assume that a given spectrum is the linear combination of a large number of dictionary

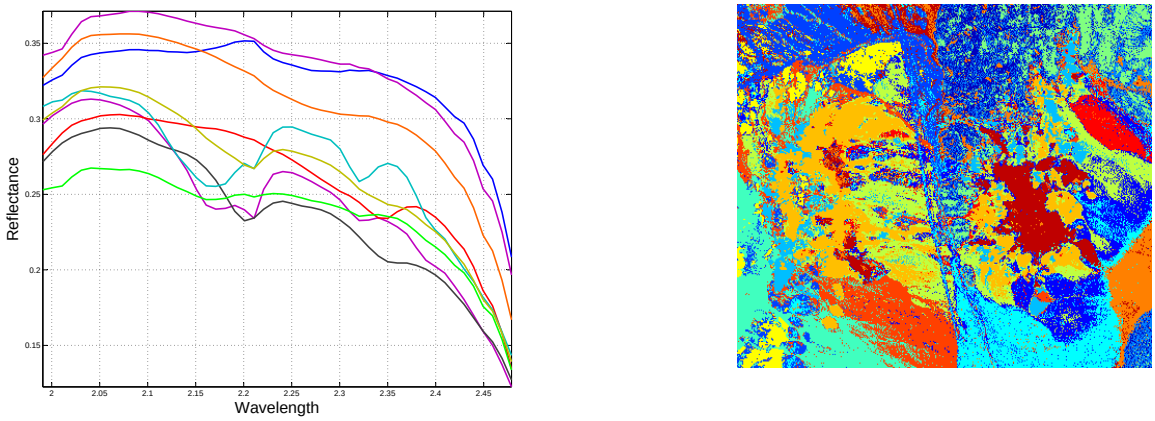


Fig. 4. Cluster Means (left) and Cluster Map (right) for Correlation Distance.

spectra, we want to impose a limit on the number of non-zero abundance coefficients. We can view this problem as selecting a small number of regressors out of a given set, in order to approximate (in a least-squares sense) a given vector. Specifically we would like to solve the following problem, for each cluster centroid:

$$\begin{array}{ll} \text{minimize} & \|Da^{(j)} - \mu_j\|_2 \\ \text{subject to} & \mathbf{Card}(a^{(j)}) \leq r \\ & a^{(j)} \geq 0 \end{array} \tag{7}$$

Here $\mathbf{Card}(a^{(j)})$ denotes the cardinality of $a^{(j)}$, i.e. the number of nonzero elements in $a^{(j)}$, or in other words the sparsity structure of the abundance vector.

For our particular problem, the dictionary contains $n = 117$ minerals. We would like to express each centroid as a linear combination of approximately $r = 5$ of those minerals.

Problem 7 reduces to a quadratic program (QP) if the cardinality constraint is removed. However, with this constraint present, it turns out that this problem is combinatorial and is thus very hard to solve. Specifically, if we wanted to find the global optimum of 7 we would have to solve $n!/r!(n-r)!$ quadratic programs. Each of these QPs would correspond to a different sparsity structure in the abundance vector. Obviously solving such a number of problems is intractable, even for a modest value of r.

There exist, however, efficient heuristics for finding approximate solutions to this problem. As explained in [2] section 6.3.2, one method that works satisfactorily is to first solve the following problem, for a range of values of λ:

$$\text{minimize} \quad \|Da^{(j)} - \mu_j\|_2 + \lambda\|a^{(j)}\|_1 \tag{8}$$

By increasing the value of λ, we are in essence putting more weight on minimizing the *l-1* norm of the abundance vector $a^{(j)}$. This causes the solution of 8 to be sparser. We can then use the sparsity pattern given by this problem to solve the original problem.

It turns out that problem 8 is equivalent to the following problem, for an appropriate choice of ϵ:

$$\begin{array}{ll} \text{minimize} & \|Da^{(j)} - \mu_j\|_2 \\ \text{subject to} & \|a^{(j)}\|_1 \leq \epsilon \end{array} \tag{9}$$

This problem can be expressed as a QP with a simple transformation in the variables.

The parameter ϵ puts a limit on the maximum allowable *l-1* norm of $a^{(j)}$. In particular, if we choose ϵ to be large, then the problem essentially becomes unconstrained. On the other hand, if ϵ is less than the *l-1* norm of the optimal solution of the unconstrained least-squares problem, then the constraint in 9 will be tight. In other words if we choose a small ϵ, then we can be certain that the solution $a_*^{(j)}$ of 9 will have $\|a_*^{(j)}\|_1 = \epsilon$. Thus, since an *l-1* norm constraint on $a^{(j)}$ will change its sparsity structure, we can change ϵ until we get the desired cardinality on the solution $a_*^{(j)}$.

Now suppose we obtain an acceptable solution $a_*^{(j)}$ to 9. We then construct the matrix $\tilde{D}$, which consists of the columns of D which correspond to non-zero entries in $a_*^{(j)}$. We then proceed to solve the following problem for each cluster centroid:

$$\begin{array}{ll} \text{minimize} & \|\tilde{D}\tilde{a}^{(j)} - \mu_j\|_2 \\ \text{subject to} & \tilde{a}^{(j)} \geq 0 \end{array} \tag{10}$$

Thus, the solution to 10, for a centroid μ_j will give us the abundances (weights) for that given cluster corresponding to equation 6. In order to express these in terms of percentages, we then have to normalize the vector $a^{(j)}$.

V. Unmixing Results

The results of figure 5 show the estimated abundance maps for three minerals, whose presence in this region is unanimously agreed on by experts (i.e. [4]). The maximum in the scale is 60% (dark red). We found reference for quantitative data for mineral abundances for this dataset, namely [15]. For the most common minerals our results our method produced abundance maps which are qualitatively similar to

those obtained used there. The values of the abundances are in broad accordance but we don't have a definitive answer of what are the most accurate for the lack of ground truth data on mineral abundances.

VI. Conclusion and Future Work

In this work we explored clustering techniques on a well-known hyperspectral image. We assessed that cluster analysis with use of the correlation distortion measure is a technique that picks up most of the variability in the dataset. Despite the lack of quantitative reference data for mineral abundances for this dataset, our results were qualitatively in accordance with other studies.

In future studies we will devise reliable performance measures for cluster validation and mineral unmixing. We will also explore the clustering performance of an algorithm that takes into account both shape and norm as an improvement of our clustering stage.

References

[1] A. Aiyer, K. Pyun, Y. Huang, D. O'Brien and R.M. Gray, *Lloyd Clustering of Gauss Mixture Models for Image Compression and Classification,* in Image Communication, Vol 20, pp. 459-485 (2005).

[2] S. Boyd and L. Vandenberghe, *Convex Optimization,* Cambridge University Press, (2004).

[3] R.N. Clark, *Spectroscopy of Rocks and Minerals, and Principles of Spectroscopy,* in Manual of Remote Sensing, John Wiley and Sons, A. Rencz Editor, New York, (1999).

[4] R. N. Clark, G. A. Swayze, K. E. Livo, R. F. Kokaly, S. J. Sutley, J. B. Dalton, R. R.McDougal, and C. A. Gent, *Imaging Spectroscopy: Earth and Planetary Remote Sensing with the USGS Tetracorder and Expert Systems,* Journal of Geophysical Research, Vol. 108, No. E12, , p. 5-1-44, (December 2003).

[5] A.P. Dempster, N.M. Laird and D.B. Rubin, *Maximum-likelihood from incomplete data via the EM algorithm,* Journal of the Royal Statistical Society, Ser. B, 39, (1977).

[6] M.T. Eissmann and R.C. Hardle, *Stochastic spectral unmixing with enhanced endmember class separation,* Applied Optics, Vol.43, No. 36, (December 2004).

[7] R.M. Gray, *Gauss Mixture Vector Quantization*, Proceedings of IEEE International Cnference on Acoustics, Speech and Signal Processing, (May 2001).

[8] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning,* Springer (2001)

[9] N. Keshava and J.F. Mustard, *Spectral unmixing,* IEEE Signal Processing Magazine, (January 2002).

[10] M. Parente, *An investigation of the Properties of Expectation-Maximization and Gauss Mixture Vector Quantization in Density Estimation and Clustering,* EE391 Report, Stanford University, (September 2004).

[11] M. Petrou, *Mixed pixel classification: an overview,* submitted to World Scientific, (1998).

[12] R. Redner and H. Walker, *Mixture densities, maximum likelihood and the EM algorithm,* SIAM Review, 26(2), pages 195-239, (April 1984).

[13] J.J Settle and N.A. Drake, *Linear mixing and the estimation of ground cover proportions,* International Journal of Remote Sensing, 14, pp.1159-1177, (1993).

[14] D. Stein, *Application of the Normal Compositional Model to the analysis of Hyperspectral Imagery,* IEEE, (2004).

[15] D.W. Stein, *The Normal Compositional Model with Applications to Hyperspectral Image Analysis,* MIT Lincoln Laboratory, Project Report NGA-8 , (March 2005).

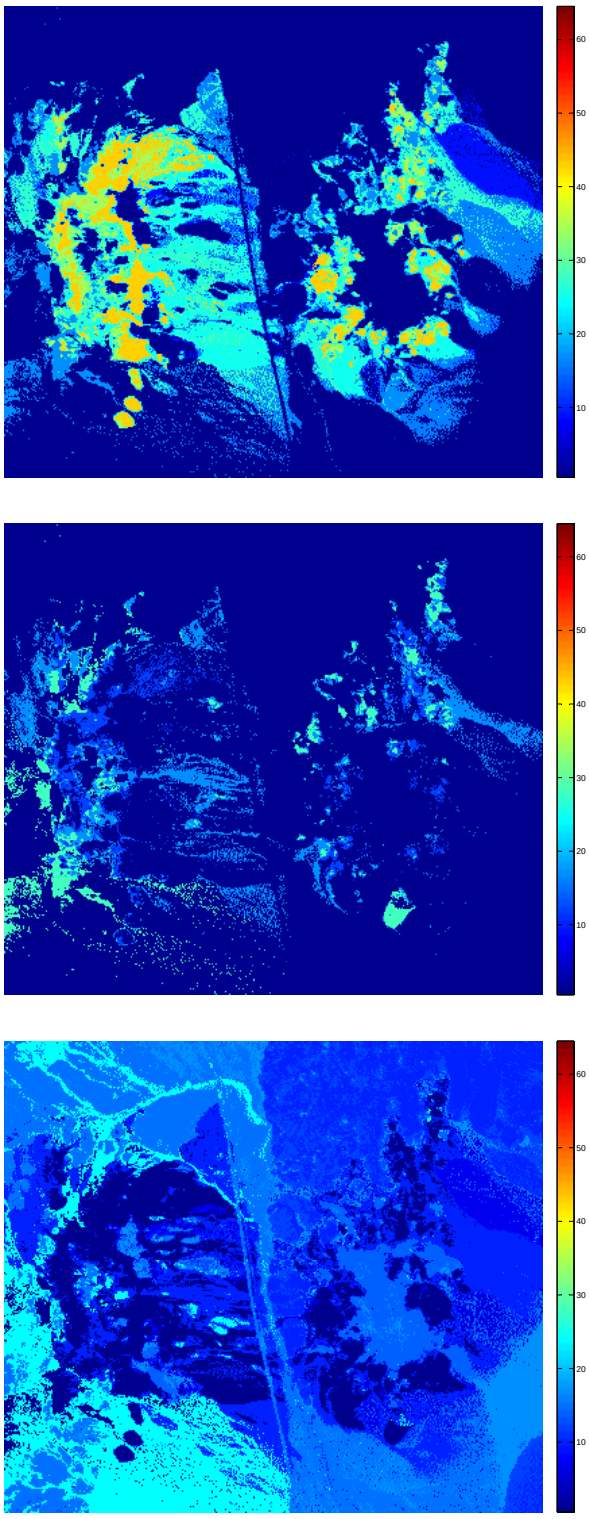


Fig. 5. Alunite HS295.3B (left), Kaolinite CM9 (middle) and Muscovite GDS113 (right) abundance maps.