# A MULTIPLE ENDMEMBER UNMIXING APPROACH FOR MAPPING HEAVY METAL CONTAMINATION AFTER THE DOÑANA MINING ACCIDENT (SEVILLA, SPAIN)

Thomas Kemper, Javier Garcia Haro, Helen Preissler, Wolfgang Mehl and Stefan Sommer
European Commission, DG JRC, Space Applications Institute, EGEO Unit, Italy
email: thomas.kemper@jrc.it

**KEY WORDS:** Hyperspectral, Spectral Mixture Analysis, Mining Accident, Spain

## ABSTRACT

A Variable Multiple Endmember Spectral Mixture Analysis (VMESMA) Tool was used to extend the possibilities of multiple endmember unmixing. In order to make the program more flexible, a zonal partition of the image was introduced to allow the application of different submodels to the selected areas. Based on an iterative feedback process, the unmixing performance may be improved in each stage until an optimum level is reached.
This approach was applied to map residual contamination after a mining accident at the Aznalcóllar Mine (Southern Spain), where heavy metal contaminated sludge was distributed over large areas. Although the sludge and the contaminated topsoils had been removed mechanically in the whole area, still high sludge abundances were found.
The sludge abundances achieved by unmixing confirm the field observations and chemical measurements taken in the area.

## 1. Introduction

Spectral mixture analysis (SMA) is a widely used method to determine the sub-pixel abundance of vegetation, soils and other spectrally distinct materials that fundamentally contribute to the spectral signal of mixed pixels (e.g. Adams et al. 1989). This is of particular importance to obtain quantitative estimates of distinct materials, which is a typical application of hyperspectral data. SMA aims to decompose the measured reflectance spectrum of each pixel into the proportional spectral contribution of so-called endmembers (EM). These EMs are known reflectance spectra considered to represent the spectral characteristics of the relevant surface components constituting the pixel surface cover proportional to their spatial occurrence (i.e. the area covered) within the pixel. The strategy to select these EMs is one of the key issues in the successful application of SMA. It has to consider the changing spectral significance of EMs as a function of the variability of the occurring surface materials, the spatial and spectral resolution of the data and the thematic purpose of the study. Different strategies have been described in literature, the most widely used method consists in employing the same EMs to the whole image, and using all available EMs at the same time.
Although it is mathematically possible to use in the decomposition as many EMs as spectral bands available, which for imaging spectrometry data would allow to use dozens of EMs, usually a limited number of endmembers should be sufficient to explain in a physically meaningful way the mixed spectral signature of an individual pixel. In an absolutely noise free system it should be theoretically no problem to retrieve out of a large number of distinct spectra abundance values >0 only for those EMs that really contribute to the mixed signal, while for the other spectra the abundance value would be 0. In real data sets however, the natural variability of EM material reflectance, residual errors of calibration and atmospheric correction as well as detector system electronic noise prevent such ideal, absolutely unambiguous solutions of the spectral unmixing. Especially in overdetermined systems (i.e. higher number of EMs used than really represented in the pixel) there is a high risk to produce mathematically well fitting results, which are physically misleading and do not explain the real composition of the pixel.
In recent years, many authors have proposed and used a more complex model where both the number and the set of EMs vary dynamically on a per-pixel basis (e.g. Roberts et al. 1991, Lacaze et al. 1996), which has become known as multiple endmember spectral mixture analysis (MESMA). The idea consists in restricting the large set of possible EMs to a small set of appropriate EMs which can be different for each pixel, thereby allowing an accurate decomposition using a virtually unlimited number of EMs. These spectra are chosen to represent the spectral variability of a larger area of interest but would over determine the spectral signal of most of the single pixels. In this work we have combined two complementary unmixing approaches: (1) the use of standardised signatures to represent the scene materials and (2) an improved strategy called Variable MESMA (VMESMA), which allows a segmentation of the image to increase the flexibility and accuracy.

In this study the potential of VMESMA for the detection of contamination after a mining accident in Spain is investigated using hyperspectral image data obtained with the HyMap sensor. The work presented here is integrated in a bigger framework for the quantitative estimation of heavy metal contamination. The retrieved sludge abundances are used as semi-quantitative measure for the assessment of the heavy metal contamination level. Furthermore, the spectral mixture analysis allows an estimation of the vegetation cover and therefore the removal of the vegetation signal from pixels for recovery of pure soil spectra. This pure soil information will be used in the next step for the application of spectral models for the quantitative mapping of heavy metals.

## 2. Study Area

The study area is situated approximately 40 km west of Sevilla (Spain). After a collapse of a tailings dam at the Aznalcóllar Mine on 25th of April 1998, 4 million cubic meters of acidic water and 2 million cubic meters of toxic mud with highest heavy metal concentrations flooded the Agrio and Guadiamar rivers endangering the wetlands of the Doñana National Park, which is one of Europe's biggest wetlands areas and a very important refuge for migrating waterfowl. The toxic sludge had covered an area of 4286 ha, mainly agricultural and marshland. Sludge and contaminated topsoils had been removed within 6 months after the accident using heavy machinery. Monitoring activities have been started by Spanish institutions to control the possible mobility of residual contaminants due to oxidation processes (Grimalt et al. 1999) (Fig. 1).

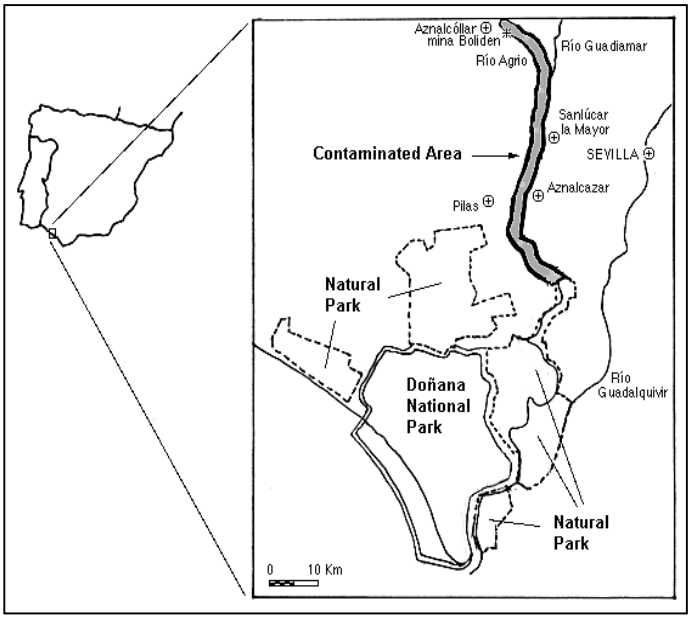



**Figure 1: Study Area**

## 3. Data Set

During an extensive field campaign representative sites along the affected river catchments were selected. Spectral measurements using GER S-IRIS resulted in a VIS to SWIR high-resolution library hierarchically organised of rocks, soils and vegetation, close related to their physiographic context. More than 200 spectra were measured to represent the spectral variability of the area. Furthermore a detailed soil sampling was carried out for soil chemical/physical laboratory analysis in order to characterise the contamination level of the soils. Finally, these have been integrated in the MedSpec database (Preissler et al. 1998).
The imaging spectrometry data were acquired with the DAIS & HyMap sensors in the framework of the HyEurope'99 airborne campaign in June 1999. The data take covered the entire contaminated area and was accompanied by a radiometric calibration field campaign. For this study only HyMap data were taken into consideration.
The atmospheric correction was done at DLR using the ATCOR software (RICHTER, 1996). Because of the good data quality no further preprocessing was necessary, merely 4 bands (487, 1346, 1401 and 1417 nm) were excluded from further analysis, because of strong atmospheric interference.

## 4. METHODS

We now introduce briefly the approaches used in this work. Instead of using reflectance spectra as input for the unmixing, an alternative SMA approach that overcomes the limitations of the shade EM, was applied. This approach consists in performing standardization on both the EMs and the spectral library, as a previous step before applying the SMA. This standardization transforms the data to a set of percentage variations about the mean value, which allows matching of the data in a manner independent of the reflectance scale (Mackin *et al.*, 1991):

$$\hat{\mathbf{r}} = \frac{\mathbf{r} - \mu_r}{\sigma_r} \tag{1}$$

where $\hat{\mathbf{r}}$ is the standardised vector associated to the pixel vector $\mathbf{r}$, with mean $\mu_r$ and standard deviation $\sigma_r$. This procedure steps the problems of shadowing and brightness variations, retaining and even enhancing the information due to the spectral shape, such as gradient (*i.e.* derivative features) and absorption bands. As the contribution of shade has

not to be modeled, it is possible either to use instead of the shade EM the spectral signature of a new different material or, alternatively, to apply the SMA with a smaller number of EMs increasing thus the reliability of estimated abundances. Figure 2 shows selected EMs as 'normal' reflectance and as standardised spectra. The use of 'normal' reflectance spectra for this special case would cause severe problems, because of the opacity and absence of spectral features in the sludge spectrum. If we used a shade EM this would not have a big spectral contrast in comparison to sludge and therefore cause confusion. If we did not use a shade EM, sludge would be used to model shade. The standardised sludge spectrum instead has enough spectral contrast to be distinguished from other EMs.

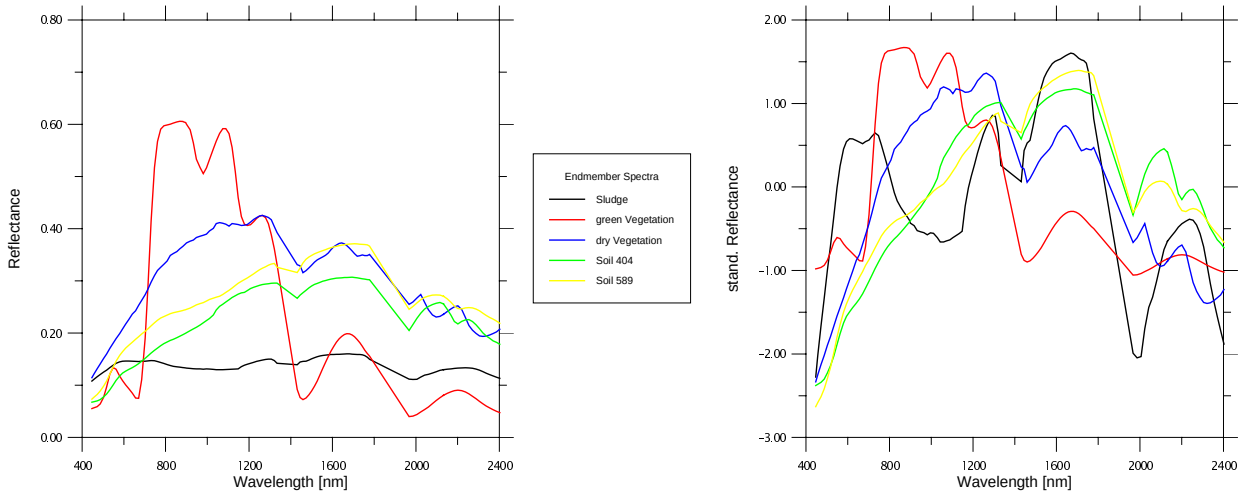



**Figure 2. Reflectance (left) and standardised (right) spectra of selected EM**

Though standardised SMA is based on the same principles as conventional SMA, it solves the problem using different standardised coordinates, which introduces some differences in the equations. We shall express the formulation of the standardised SMA as follows:

$$\hat{\mathbf{r}} = \sum_{i=1}^{c} \hat{\mathbf{E}}_i \hat{f}_i + \hat{\boldsymbol{\varepsilon}} \tag{2}$$

where $\hat{\mathbf{r}}$ is the standardised pixel vector, $\hat{\mathbf{E}}_i$ represents the *i*-th standardised endmember, $\hat{f}_i$ is the proportion of such endmember in the standardised coordinates, *c* is the number of components in the pixel and $\hat{\boldsymbol{\varepsilon}}$ is the residual vector (expressed in standardised units). The standardised SMA consists in unmixing the pixel vector $\hat{\mathbf{r}}$ using standardised endmembers $\hat{\mathbf{E}}_i$ (i =1,…c) in order to obtain the proportions $\hat{f}_i$. In this work, we have used a constrained analytical estimator (García-Haro et al., 1999). In standardised SMA, the sum-to-one condition expresses as follows:

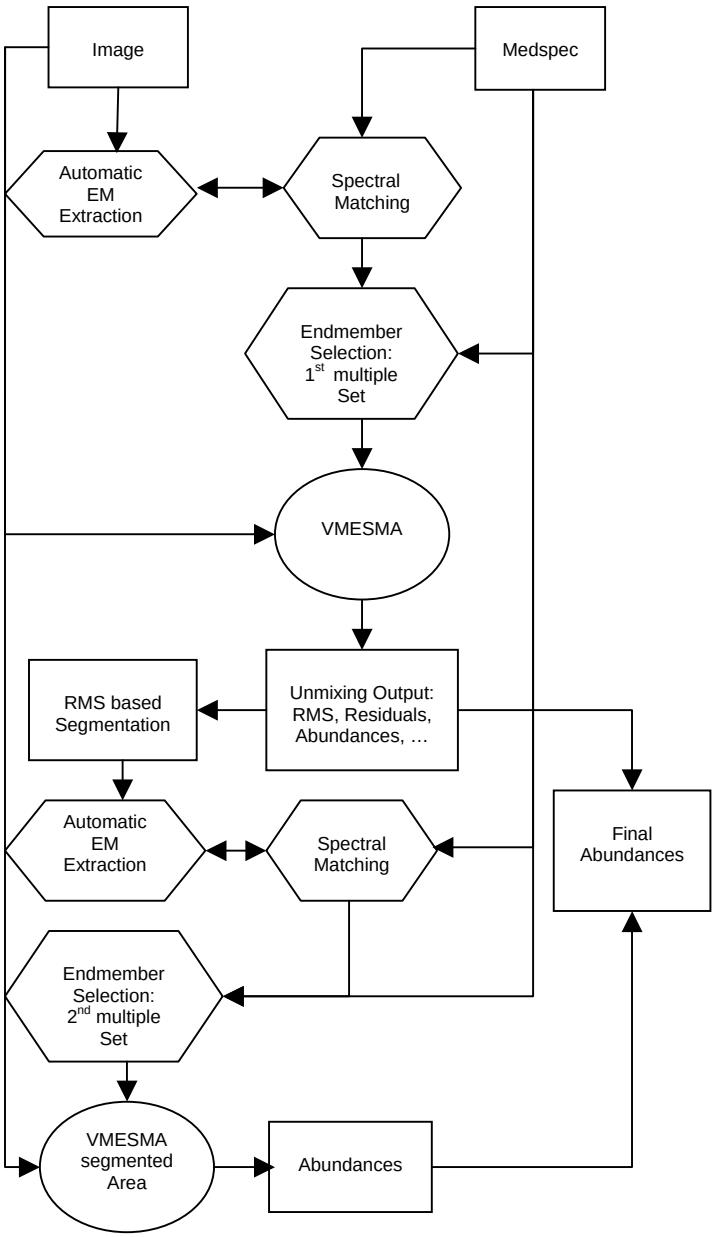
$$\sum_{i=1}^{c} \frac{\hat{f}_i}{\sigma_{E_i}} = \frac{1}{\sigma_r} \tag{3}$$

where $\sigma_{E_i}$ and $\sigma_r$ represent the standard deviation of the vectors $\hat{\mathbf{E}}_i$ and $\hat{\mathbf{r}}$, respectively. Finally, the proportions in the original system, $f_i$, are obtained using the expression:

$$f_i = \sum_{i=1}^{c} \frac{\sigma_r}{\sigma_{E_i}} \hat{f}_i \tag{4}$$

This transformation ensures the sum-to-one condition and preserves also the positive proportions.

VMESMA allows a zonal partition of the area and a zone-dependent choice of different EM submodels and methods. This method allows for the definition of standard and repeatable pathways to incorporate information dynamically derived from the unmixing results with other sources of data in order to optimize the algorithms and to increase the flexibility of the modeling approach. VMESMA is conceived as an iterative or feedback process, in which unmixing performance may be potentially improved in each stage until an optimum level is reached.




In this study VMESMA was applied in a hierarchical way (Fig. 3). For this application a two-step approach was sufficient, but the program allows in general an unlimited number of iterations. In the first step a simple EM set was selected (green and dry vegetation and the main soil types in the area) and applied to the whole image (see Fig. 2). The EM selection process was supported by an automated image EM selection procedure. This procedure visualises the distribution of the selected library EMs projected on the image cloud of the first primary PCs and extracts extreme EMs from the image. These image EMs are then spectrally matched against MedSpec, a spectral attribute data base (Preissler et al. 1998) using different standard methods (e.g. spectral angle mapping). After this visualization one extra soil was selected from the automated EM selection procedure and matched to a dark soil, which is rich in organic matter with little remaining dry vegetation. Only a maximum of 3 EMs were applied for every pixel, since the use of the standardised SMA accounts already for differences in shadow or shade. Moreover, it is not common to find four or more different materials in a pixel of about 6 m. Hence we have considered combinations of 2 and 3 EM submodels.

In the second step the modelling error obtained in the first unmixing was used for image segmentation. In this application the segmentation is based only on the RMS. The software allows incorporating also information from GIS or other classifications. In the areas with a higher modelling error more appropriate EM subsets were selected using the automated EM extraction procedure. To model the sludge signature a library spectrum was considered.

**Figure 3. Flowchart of VMESMA Application**

## 5. Results

The application of VMESMA is demonstrated at two subsets of the HyMap images with different soil backgrounds. The first site, Soberbina, is located in the North of the area, just in the South of the confluent of Agrio and Guadiamar rivers. This area is built up mainly by point bar deposits, formed by gravels and medium to coarse sand textures, and by finer grained overbank deposits. The gravel and sand deposits are mined in open pits (Gallart et al. 1999). The second site, Vado de Quema, is in the South of the sludge-affected area (South of Aznalcazar). In this area the main channel is limited by natural levees. The floodplain is built up by overbank deposits with finer sand and silt textures (Gallart et al. 1999). The two subsets are in areas in which soil samples have been analysed, and thus the contamination level is known. In both areas the sludge has been removed mechanically and ploughed so that remaining sludge was worked into the soil, except for three areas in the southern subset, which could be used as reference.

**Soberbina**

The analysis of the modelling performance combined with the use of RMS thresholds allowed a discrimination of sludge-affected areas. The result was a binary mask covering 29 % of the area, which corresponded to an RMS threshold of 0.35. This zone overlaps very well with the sludge-affected area as assessed by our observations and the intensive field works undertaken by different research groups after the mine accident (e.g. Lopez-Pamo et al.1999, Cabrera et. al. 1999). After the second step, including the sludge EM, the RMS for the whole area could be reduced from 0.38 to 0.21 (45 % reduction). This reduction was even significantly higher in the affected area. Mainly 3-EM sets were used for the modelling, only in 21 % of the pixels 2-EMs were sufficient.

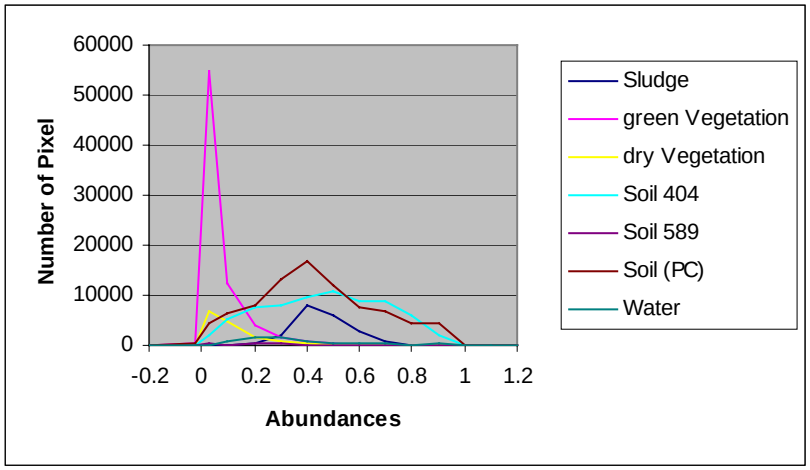



**Figure 4: Soberbina, Abundance Histograms**

The abundance histograms (see Fig. 4) show that even using a relatively small number of EMs, the model produces physical realistic abundances with only 0.1 % of values as negative as –0.05. The model is also able to detect low levels of green vegetation (mean value of 8 %), which are present in 76.4 % of the area.

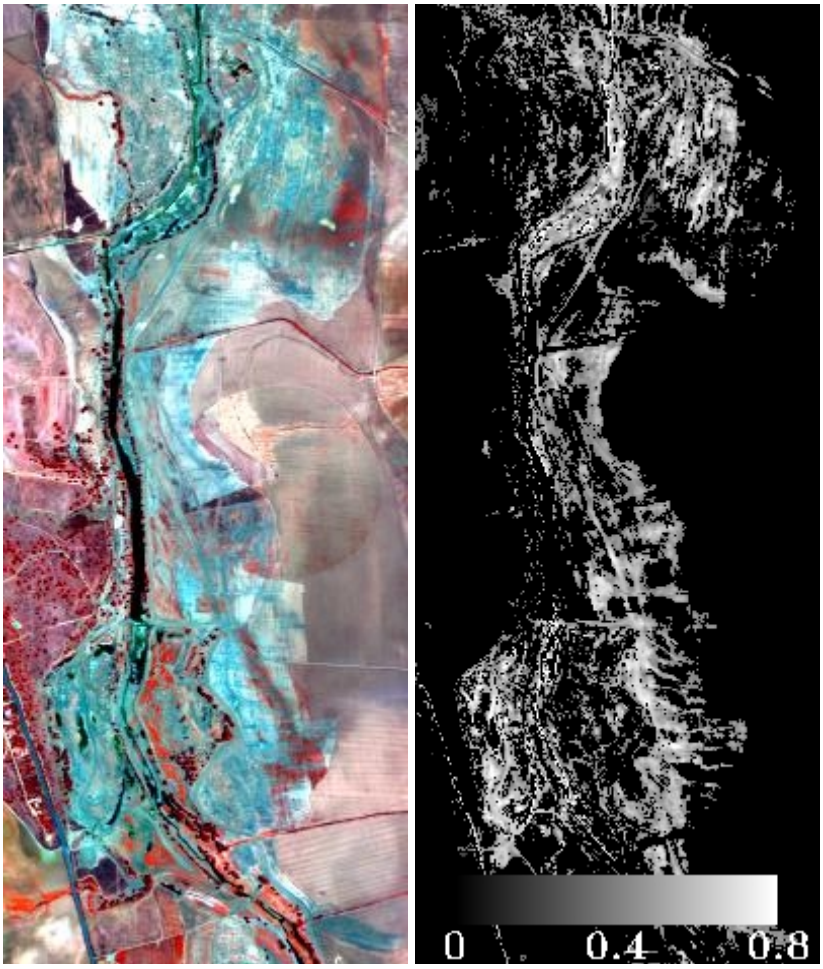



**Figure 5. Soberbina, False Color Composite (R=1019, G=587, B=487 [nm], left) & Sludge Abundance (right)**

The distribution of sludge is limited to an area of 21 percent with abundances between 30 and 60 percent. The highest values were found in gravel quarries in the South of the image, where the sludge could penetrate very deep into the soils due to the coarse soil composition (Fig. 5). Generally, the distribution is very irregular and no gradients were recognizable, because the distribution patterns after the accident have been changed drastically due to mechanical removal. Furthermore, dense annual herbaceous vegetation partly covers the area.

**Vado de Quema**

For the Vado de Quema site the same procedure was applied as described above for Soberbina. The RMS analysis resulted in a binary mask covering 59 percent of the area. This mask did not only cover the sludge-affected area, but also some other areas, where no appropriate modelling was possible with the initially selected EMs. Therefore new EMs were extracted from the image using the automated procedure. After spectral matching three image EM were selected. A water EM for the modelling of the ponds and the river, a soil for the modelling of the bright sandy areas and a sludge EM, which was more appropriate for the sludge in this area, were selected.

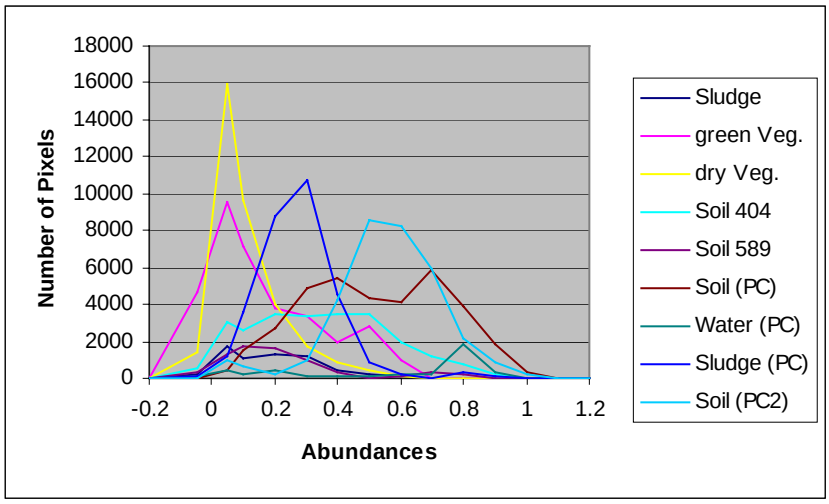



**Figure 6. Vado de Quema, Abundance Histograms**

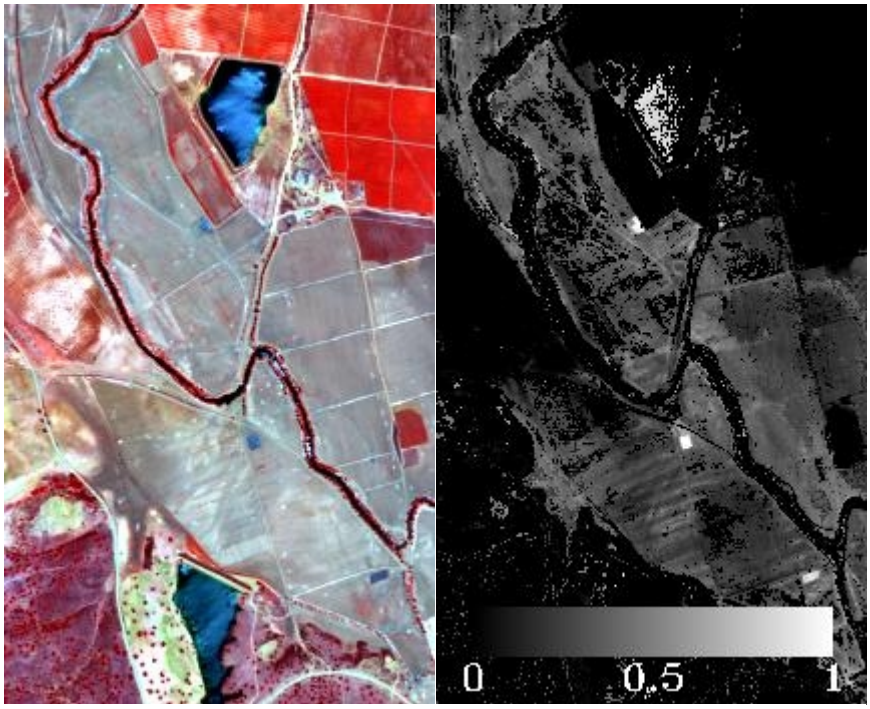



**Figure 7. Vado de Quema, False Color Composite (R=1019, G=587, B=487 [nm], left), Sludge Abundance (right)**

This submodel was applied to the segmented area. After the second unmixing step, the RMS for the whole area could be reduced from 0.50 to 0.22 (56 % reduction). The abundance histograms for Vado de Quema showed a slight increase in negative values compared to Soberbina with 0.36 percent of values as negative as –0.05. This increase can be explained due to the bigger number of 3-EM submodels (28 %). In Vado de Quema the sludge contamination ranges between 15 and 50 percent and covers an area of 49 %. In the abundance image the three areas in which pure sludge remained after the accident are clearly recognizable. Furthermore we found line structures in the southern part of the image, which were readily apparent in the field and were probably caused by the mechanical removal of the sludge. The riverbed is easily identifiable, because of the presence of shallow water and riparian vegetation that was not removed during the clean up. Also dark spots in the affected area correspond to residual vegetation (mainly annual herbaceous species).

Major sources of errors come from the incorrect modelling of all spectral components of the image, since the surfaces in the area show a high spectral variability. An example can found at one of the irrigation ponds in Fig. 7 because of the mismatch of the water EM. However, these problematic areas may be easily identified because of their high RMS values. These areas might eventually be either masked or modelled better using more suitable EMs. Other minor source of errors may be also attributed to the correction of the atmospheric effects.

The sludge abundances achieved with the VMESMA approach show a clear separation between areas known to be affected and the undisturbed surroundings. This demonstrates the ability of the approach to monitor the remaining contamination. Furthermore we have been able to distinguish different levels of contamination independent of the soil background. The differences in the level of contamination between Soberbina and Vado de Quema are confirmed by our own chemical analysis and the levels of contamination found by CABRERA et al. (1999). This approach will be applied in the near future to the whole area in order to retrieve sludge abundance maps of the whole area.

## 6. Conclusions

The VMESMA was applied successfully for the mapping of remaining contamination. This setup proved to be very useful for such complex tasks as the mapping of contamination, despite of the problem that sludge was worked into the soil so that is an intimate mixture rather than a distinct surface pattern.
It became clear again that in order to retrieve accurate abundances of sludge, it is essential to adequately address the spectral response of the rest of materials that make up the pixel. This underpins the importance of the fieldwork for characterisation of the study area.
VMESMA software also allows for the removal of the vegetation signal from each pixel in order to retrieve pure soil spectra. This pure soil spectral information will be used to quantify the heavy metal concentration with the help of spectral models, which are currently being derived and validated against the geo-chemical analyses of the field samples.
On the longer term repeated hyperspectral data takes could contribute to assess and monitor the oxidation of the remaining pyritic sludge. To start the monitoring new hyperspectral images were aquired during the HySens2000 flightcampaign in June this year. This flight campaign was again accompanied by spectral field measurements and intensive soil sampling campaign carried out by Soil & Waste Unit of the Environment Institute.

## Acknowledgements

We would like to express our gratitude to the team of the Soil & Waste Unit of the Environment Institute, JRC Ispra, for their support during the field campaign and the laboratory analysis of the soil samples. We also would like to thank the DLR team for assistance in ground calibration during the over flight and the radiometric correction of the images.

## References

Adams, J.B., Smith, M.O., Gillespie, A.R., 1989, Simple Models for complex natural surfaces: a strategy for the hyperspectral era of remote sensing, in: Proc. Int. Geoscience and Remote Sensing Symp. (IGARSS ‘89), July 10-14, Vancouver, pp. 16-21.
Cabrera, F., Clemente, L., Diaz Barrientos, E., López, R., Murillo, J.M., 1999, Heavy Metal Pollution of Soils affected by the Guadiamar toxic flood. Sci. Tot. Env., 242(1-3), pp. 117-130.
Gallart, F., Benito, G., Martín-Vide, J.P., Benito, A., Prió, J.M., Regüés, D., 1999. Fluvial Geomorphology and hydrology in the dispersal and fate of pyrite mud particles released by the Aznalcóllar mine tailings spill. Sci. Tot. Env., 242(1-3), pp. 13-26.
Grimalt, J.O., Ferrer, M., Macpherson, E., 1999. The mine tailing accident in Aznalcollar. Sci. Tot. Env., 242(1-3), pp. 3-11.
Preissler, H, Bohbot, H, Mehl, W and Sommer, S., 1998, MEDSPEC-A spectral database as a tool to support the use of imaging spectrometry data for environmental monitoring, 1st EARSeL Workshop on Imaging Spectrometry, Zurich, Switzerland, pp. 455-462.
Lacaze, B., Casselles, V., Coll, C., Hill, J., Hoff, C., De Jong, S., Mehl, W., Negendank, J.F.W., Riezebos, H., Rubio, E., Sommer, S., Teixeira Filho, J., Valor, E., 1996, Integrated approaches to desertification mapping and monitoring in the mediterranean basin. Final Report of the DeMon-1 Project.- EUR 16448, Bruxelles.
Lopez-Pamo, E., Barettino, D., Antón-Pacheco, C., Ortiz, G., Arránz, J.C., Gumiel, J.C., Martinez-Pledel, B., Aparicio, M., Montouto, O., 1999. The extent of the Aznalcóllar pyritic sludge spill and its effects on soil. Sci. Tot. Env., 242(1-3), pp. 57-88.
Mackin, S, Settle, J, Drake, N. and Briggs, S, 1991, Curve shape matching, end-member selection and mixture modelling of AVIRIS and GER data for mapping surface mineralogy and vegetation communities, proceedings of the Airborne VISIBLE/Infrared imaging spectrometer (AVIRIS). Airborne Geoscience workshop bibliographies.
Richter, R., 1996. Atmospheric Correction of DAIS Hyperspectral Image Data. SPIE Proc., Vol. 2758, AEROSENSE ’96 Conference, Orlando, 8-12 April.
Roberts, D. A., Smith, M. O., Adams, J. B. and Gillespie, A. R., 1991, Leaf Spectral Types, Residuals, and Canopy Shade in an AVIRIS Image, Proc. 3rd Airborne Sci. Workshop: AVIRIS, JPL, Pasadena, CA, 20-21, May, 43-50.