Earth Mineral dust source InvesTigation (EMIT)

EMIT L3 Algorithm: Aggregated Mineralogy

Theoretical Basis

Gregory S. Okin

Department of Geography, University of California, Los Angeles

David R. Thompson, Philip G. Brodrick

Jet Propulsion Laboratory, California Institute of Technology

Version 0.3

Feb 2020

Change Log

Version	Date	Comments
0.1	Oct, 2019	Initial Draft
0.2	Jan 2020	Post-PDR refinements
0.3	Feb 2020	Refinements from demo SDS

TABLE OF CONTENTS

TABLE OF CONTENTS

1. Key Teammembers

A large number of individuals contributed to the development of the algorithms, methods, and implementation of the L3 approach for EMIT. The primary contributors are the following:

- Gregory S. Okin (UCLA) – EMIT Co-I, L3 product lead
- David R. Thompson (Jet Propulsion Laboratory) – EMIT Co-I, Instrument Scientist
- Robert O. Green (Jet Propulsion Laboratory) – Mission PI, Radiometric modeling
- Bethany L. Ehlmann (Caltech) – EMIT Co-I, Mineral composition and abundance validation
- Philip G. Brodrick (Jet Propulsion Laboratory) – Algorithms Design and Implementation

2. Historical Context and Background on the EMIT Mission and its Instrumentation

Mineral dust aerosols originate as soil particles lifted into the atmosphere by wind erosion. Mineral dust created by human activity makes a large contribution to the uncertainty of direct radiative forcing (RF) by anthropogenic aerosols (USGCRP and IPCC). Mineral dust is a prominent aerosol constituent around the globe. However, we have poor understanding of its direct radiative effect, partly due to uncertainties in the dust mineral composition. Dust radiative forcing is highly dependent on its mineral-specific absorption properties. The current range of iron oxide abundance in dust source models translates into a large range of values, even changing the sign of the forcing (-0.15 to 0.21 W/m2) predicted by Earth System Models (ESMs) (Li et al., 2020). The National Aeronautics and Space Administration (NASA) recently selected the Earth Mineral Dust Source Investigation (EMIT) to close this knowledge gap. EMIT will launch an instrument to the International Space Station (ISS) to directly measure and map the soil mineral composition of critical dust-forming regions worldwide.

The EMIT Mission will use imaging spectroscopy across the visible shortwave (VSWIR) range to reveal distinctive mineral signatures, enabling rigorous mineral detection, quantification, and mapping. The overall investigation aims to achieve two objectives.

1. Constrain the sign and magnitude of dust-related RF at regional and global scales. EMIT achieves this objective by acquiring, validating and delivering updates of surface mineralogy used to initialize ESMs.

2. Predict the increase or decrease of available dust sources under future climate scenarios. EMIT achieves this objective by initializing ESM forecast models with the mineralogy of soils exposed within at-risk lands bordering arid dust source regions.

The EMIT instrument is a Dyson imaging spectrometer that will resolve the distinct absorption bands of iron oxides, clays, sulfates, carbonates, and other dust-forming minerals with contiguous spectroscopic measurements in the visible to short wavelength infrared region of the spectrum. EMIT will map mineralogy with a spatial sampling to detect minerals at the one hectare scale and coarser, ensuring accurate characterization the mineralogy at the grid scale required by ESMs. EMIT's fine spatial sampling will resolves the soil exposed within hectare-scale agricultural plots and open lands of bordering arid regions, critical to understanding feedbacks caused by mineral dust arising from future changes in land use, land cover, precipitation, and regional climate forcing.

The EMIT Project is part of the Earth Venture-Instrument (EV-I) Program directed by the Program Director of the NASA Earth Science Division (ESD). EMIT is comprised of a Visible/Shortwave

Infrared Dyson imaging spectrometer adapted for installation on the International Space Station (ISS).

Table 1 below describes the different data products to which the EMIT Mission will provide to data archives. This document describes the "Level 3" stage which relies on outputs from the Level 2A algorithms (cloud masking, standing water, vegetation cover) and the Level 2B mineral detection algorithms to produce aggregated mineral maps that can be ingested into Earth System models to evaluate Radiative Forcing (RF) impacts (Level 4).

Table 1. Emit Data Product Hierarchy

Data Product	Description	Initial Availability	Median Latency Post-delivery	NASA DAAC
Level 0	Raw collected telemetry	4 months after IOC	2 months	LP DAAC
Level 1a	Reconstructed, depacketized, uncompressed data, time referenced, annotated with ancillary information reassembled into scenes.	4 months after IOC	2 months	LP DAAC
Level 1b	Level 1a data processed to sensor units including geolocation and observation geometry information	4 months after IOC	2 months	LP DAAC
Level 2a	Surface reflectance derived by screening clouds and correction for atmospheric effects.	8 months after IOC	2 months	LP DAAC
Level 2b	Mineralogy derived from fitting reflectance spectra, screening for non-mineralogical components.	8 months after IOC	2 months	LP DAAC
Level 3	Gridded map of mineral composition aggregated from level 2b with uncertainties and quality flags	11 months after IOC	2 months	LP DAAC
Level 4	Earth System Model runs to address science objectives	16 months after IOC	2 months	LP DAAC

A high-level, yet complete workflow of the EMIT science data system is shown in Figure 1 for context.
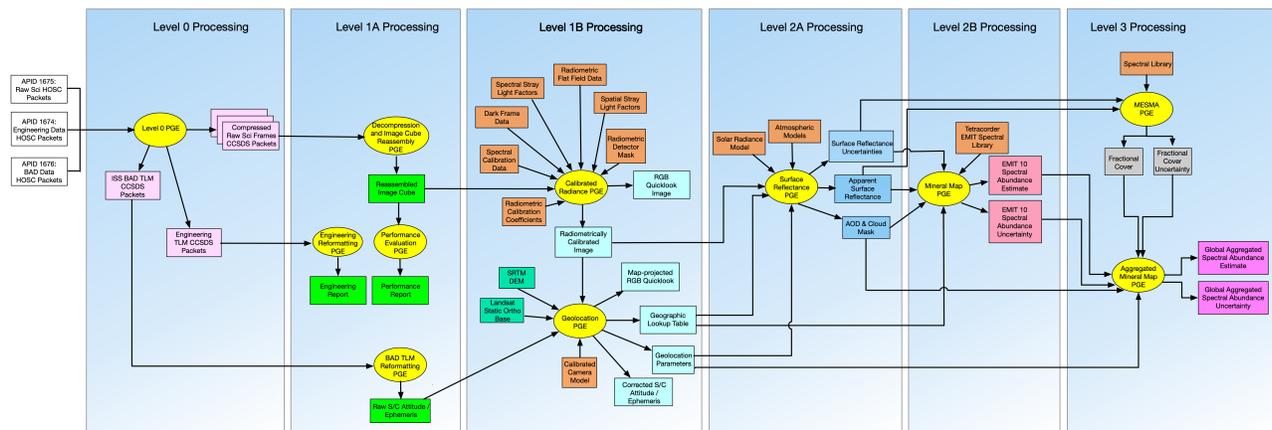

Figure 1. High-level workflow of the EMIT science data system.

3. Algorithm rationale

The EMIT L3 approach relies on the strength of the mineral detection algorithms described in L2B, which have a long development and verification history (e.g. Clark 2003, Swayze 1997). In the L3 step, these L2B products are adjusted for the sub-pixel vegetation and then aggregated to the half degree resolution required for L4 modeling. The aggregation step uses simple masked averaging, and as measurement noise propagated as inputs into L3 is expected to be random, uncertainties in this step should only decrease.

The main new algorithmic approach used in the L3 step is the sub-pixel vegetation fraction estimations, which are generated here using the Multiple Endmember Spectral Unmixing approach (MESMA; Roberts et al. 1998). Like the mineral identification algorithms used, MESMA also has a long history of development and validation. While linear spectral unmixing based on endmembers constructed from image acquisitions precedes MESMA (e.g. Roberts, 1993), the 1998 advancement described a more repeatable and generalizable approach by generating model endmember sets from large, *in situ* reference library endmember sets. Endmember selection techniques have undergone a series of advancements (Dennison & Roberts, 2003; Dennison et al., 2004; Schaaf et al., 2011; Roth et al., 2012), and the resulting products from MESMA have been field-validated and shown to be effective relative to other methods (Dennison et al., 2019).

4. Algorithm description

Below we detail the algorithms used to generate EMIT L3 products. The major processing steps, as well as input and output data, are outlined in Figure 1.

4.1 Input data

The EMIT input and output data products delivered to the DAAC use their formatting conventions, the system operates internally on data products stored as binary data cubes with detached human-readable ASCII header files. The precise formatting convention adheres to the ENVI standard, accessible (Jan 2020) at https://www.harrisgeospatial.com/docs/ENVIHeaderFiles.html. The header files all consist of data fields in equals-sign-separated pairs, and describe the layout of the file. In the file descriptions below, n denotes the number of lines particular to the given acquisition and c the number of columns.

The specific input files needed for the L3 stage are:

1. **Estimated mineral spectral abundance,** provided as n x c x 10 BIL interleave data cubes, where each band corresponds to the one of the 10 identified EMIT mineral classes. Each channel contains the estimated EMIT-10 mineral spectral abundance, as defined in L2B Section 4.2.2.
2. **Estimated mineral spectral abundance uncertainty,** provided as n x c x 10 BIL interleave data cubes, where each band corresponds to the one of the 10 identified EMIT mineral classes. Each channel contains the estimated EMIT-10 mineral spectral abundance uncertainty, as defined in L2B Section 5.
3. **Surface reflectance,** provided as n x c x b BIL interleave data cubes, where each of b bands corresponds to a different wavelength.
4. **Channelized surface reflectance uncertainty,** provided as n x c x b BIL interleave data cubes, where each of b bands corresponds to a different wavelength.
5. **Cloud, shade, and water mask,** provided as n x c x 1 binary files. Details on mask bit assignments are available in the L2A ATBD.

6. **Geospatial reference data,** provided as raw-space n x c x 3 BIL interleave data cubes. The three channels designate the x, y, z ground-coordinates of each pixel.

4.2 Theoretical description

To introduce the Level 3 aggregation, we first define the Spectral Abundance (SA) for an observed reflectance spectrum containing an absorption signature. The SA aims to estimate the effective areal fraction of the spectrum spatial footprint covered by the pure material. To estimate SA_i for mineral i, we calculate the distance from the deepest point of the mineral absorption to the local continuum and normalize it relative to the feature depth for a pure library spectrum of the material, and then scale by the proportion of mineral i in the library spectrum. It is a simple proxy for the areal coverage fraction of that mineral within the spectrum spatial footprint, relative to the library sample. Intimate mixtures count as aggregate materials which apportion their area to their constituent minerals. For a complete description of this process, we refer the reader to the L2B Algorithm Theoretical Basis Document (ATBD). These values are provided as inputs into the Level 3 aggregation. The L3 aggregation step uses the 60 m ground-level resolution SA output from L2B, in conjunction with various masks and adjustments for vegetation, in order to estimate the Aggregated Spectral Abundance (ASA) of each EMIT mineral for each ESM grid square. This L3 ASA product is defined as the expected normalized band depth that one would find upon measuring the surface reflectance at a random bare (not vegetated or within water) location within the grid square. Here we use "expectation" formally in its statistical sense to mean a numerical average.

Aggregation from the native-sensor resolution SA estimates to 0.5° x 0.5° model grid cells requires four basic steps: 1) the correction of mineral SA estimations to account for partial vegetation cover, 2) the conversion of raw-space input data products to map-space co-registered products, using provided geospatial information, 3) the aggregation of vegetation-corrected mineral SA to 0.5° x 0.5° model grid cell estimates, and 4) the propagation of uncertainty.

4.2.1 Bare Earth Percentage Adjustment

The goal of the L3 aggregation is to provide the Aggregated Spectral Abundance of the given mineral over only bare-ground regions of the half degree model grid cells. Consequently, prior to aggregation all areas that are not bare-ground need to be removed. This is done by first masking out areas where the bare-ground coverage f_b, does not exceed a particular threshold $f_{b,th}$, caused by vegetation, shade, water, or anthropogenic sources obscuring the bare-ground. We select a $f_{b,th}$ value of 0.5 based on published evaluations (Okin *et al.*, 2001). In areas where $1 > f_b > f_{b,th}$, the SA of 60 m ground-level resolution L2B outputs must also be adjusted so that they are only representative of the bare-ground component of the surface. This adjustment is proportional to the inverse of bare ground coverage, giving

$$SA_i^c = \frac{SA_i}{f_b} \tag{1}$$

where SA_i^c is the corrected spectral abundance.

To estimate fractional cover, we use the Multiple Endmember Spectral Unmixing approach (MESMA; Roberts et al. 1998, Dennison et al., 2019). MESMA is a spectral mixture analysis method that uses a linear combination of endmembers to estimate surface reflectance. Both the number and type of endmembers are allowed to vary throughout the image. Endmember selection is a critical for generalized applications of MESMA, and here we use endmembers derived from aggregations of a diverse set of field spectra, following Dennison et al. (2019).

4.2.2 Orthorectification and mosaicing

To aggregate into 0.5° x 0.5° model grid cells, all data (NBD_i^c and associated masks) must be transformed from raw to map space. Since pixel-specific georeferencing has already occurred, here we simply need to use the reference pixel coordinates as a look up table to convert all products to map space. This is done, however, on a line-by-line basis, and many acquisitions will overlap. Consequently, acquisitions will be mosaiced after masking (including input cloud, shade, and water masks as well as derived vegetation masks (Section 4.2.1)), by selecting for pixels with the minimum summed mineral measurement uncertainty.

4.2.3 Aggregation

Half-degree spatial aggregates of the corrected spectral abundances are not intended to account for the variable vegetation, quartz, and feldspar distributions within a grid cell, given that these will enter into L4 models from other sources (e.g., Scanza et al., 2015). This aggregated product, which we term the Aggregated Spectral Abundance (ASA) and is calculated for each of the EMIT 10 minerals (*i*), can be calculated as the simple average of relevant pixels within the grid, giving:

$$ASA_i = \frac{\sum_{j=1}^{N}(1 - m_j)SA_{i,j}^c}{\sum_{j=1}^{N}(1 - m_j)} \tag{2}$$

where *j* is an index over the two spatial dimensions within the given grid cell and m_j is a binary indication of whether each individual pixel is masked (either from an input mask or because $f_b <$ $f_{b,th}$).

The corresponding variability in ASA_i will be characterized as

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{N}(1 - m_j)\left(SA_{i,j}^c - ASA_i\right)^2}{-1 + \sum_{j=1}^{N}(1 - m_j)}} \tag{3}$$

where σ_i is the standard deviation of the estimates of ASA_i^c in each grid cell.

4.3 Practical Considerations
The various adjustment and aggregation code described above is implemented in Python 3.7, and can operate independently on different scenes. All computation is faster than the L1B, L2A, and L2B stages of analysis, limited mainly by the input / output throughput. Operations can be executed out-of-core to remove any memory limitations for long lines. Dependencies are all provided as input, making this code base easily operable and stand-alone.

4.4 Output Data
Level 3 output data include both *delivered* products, which are necessary for mission success, as well as *auxiliary* products, which are generated in the process of producing the delivered products, and preserved for transparency and issue tracking.

4.4.1 Delivered Products
1. **Mineral aggregated spectral abundance,** provided as a 10 band global image in GeoTiff format, EPSG:4326 at 0.5 x 0.5 degree resolution. Each channel contains the Aggregated Spectral Abundance (see section 4.2) of each EMIT mineral.
2. **Mineral aggregated spectral abundance uncertainty,** provided as a 10 band global image in GeoTiff format, in EPSG:4326 at 0.5 x 0.5 degree resolution. Each channel contains the EMIT mineral aggregated spectral abundance measurement uncertainty as defined in section 5.

4.4.2 Auxiliary Products
1. **Fractional cover,** provided as an *n* x *c* x 3 BIL interleave data cube, with *c* columns and *n* lines. Each channel contains the fractional cover as calculated by MESMA (see section 4.2.1).
2. **Fractional cover uncertainty,** provided as an *n* x *c* x 3 BIL interleave data cube, with *c* columns and *n* lines. Each channel contains the estimated uncertainty of the fraction cover, as defined in section 5.

5. Calibration, uncertainty characterization and propagation, and validation

Uncertainty characterization of the aggregated L3 product comes through a combination of uncertainty estimates from the L2B outputs and MESMA products, and is ultimately provided as a spectral abundance uncertainty for each EMIT mineral. To estimate the uncertainty of the MESMA results, we run 100 Monte Carlo experiments, randomly perturbing each reflectance spectrum by its estimated uncertainty. These results are used to estimate the standard deviation of f_b, or σ_b. Neglecting the uncertainty of the binary masking, and treating all uncertainty instances as random and independent, we can reduce relative ASA uncertainty (Ψ^i_{ASA}) to:

$$\Psi^i_{ASA} = \sqrt{\left(\frac{ASA_i}{\sum_{j=1}^{m}(1-m_j)}\right)^2 \left\{\sum_{j=1}^{N}(1-m_j)\left(\left(\frac{\Psi^i_{SA_j}}{SA_{i,j}}\right)^2 + \left(\frac{\sigma_b^j}{f_b^j}\right)^2\right)\right\}} \qquad (4)$$

where $\Psi^i_{SA_j}$ is the spectral abundance uncertainty for spatial index *j* and EMIT mineral *i,* as calculated in L2B. Notably, due to the random and independence assumptions, uncertainty decreases as a square root of the number of unmasked, observed pixels within the half degree grid cell. Following the 50% grid cell coverage requirement, this corresponds to a scaling factor of less than 4e[-5] at the equator.

While the independence and randomness assumptions are reasonable for the origins of the propagated uncertainty values (which stem from L2A), there are several other potential sources of uncertainty that are not considered here. These include:

- **Misidentification inside of the L2B spectral library.** If the mineral identification process in L2B mistakes a surface property for another mineral, this error would not be captured in the L2B outputs, and consequently is unaccounted for in the L3 output. This type of error is likely to have some form of spatial order, given the general spatial autocorrelation of land features, and consequently would likely not diminish by the ~4e[-5] factor shown above. However, there is no way of estimating this type of error without more complete surface knowledge, which is itself an objective of this mission.
- **Uncertainty from unmeasured areas.** Areas not observed by EMIT, or that were covered by cloud, shadow, or particularly high aerosol levels (at the time of observation) will not be included in the ASA calculations above. Consequently, up to 50% of the surface minerology could be unaccounted for. Due to the expected spatial autocorrelation of surface mineral composition, it is nevertheless likely that mapped areas within each half degree pixel are reasonable estimates of the ASA. This is particularly true at the global scale, where it is unlikely that unmeasured areas are systematically correlated with a specific mineral type.

6. Constraints and Limitations

No constraints or limitations are imposed on the L3 grid. All delivered data will have undergone quality control and should be considered valid calibrations up to the reported uncertainties in input parameters. Unanticipated data corruption due to factors outside the modeling, if discovered, will be reported in peer reviewed literature and/or addenda to this ATBD.

References

Ardila, D. R., Dennison, P., Green, R. O., Renzullo, J., Roberts, D. A., Thompson, D. R. Routine Production of a Terrestrial Ecosystem Product for Green Vegetation, Non-Photosynthetic Vegetation, and Substrate Fractions for AVIRIS, JPL Internal Report available on request, 2017.

Clark, R. N., and Roush, T. L., 1984, Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications: Journal of Geophysical Research, 89, p. 6329-6340.

Dennison, P. E., Halligan, K. Q., & Roberts, D. A. (2004). A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sensing of Environment*, *93*(3), 359-367.

Dennison, P. E., & Roberts, D. A. (2003). Endmember selection for multiple endmember spectral mixture analysis using endmember average RMSE. *Remote sensing of environment*, *87*(2-3), 123-135.

Dennison, P.E., Qi, Y., Meerdink, S.K., Kokaly, R.F., Thompson, D.R., Daughtry, C.S., Quemada, M., Roberts, D.A., Gader, P.D., Wetherley, E.B. and Numata, I., 2019. Comparison of Methods for Modeling Fractional Cover Using Simulated Satellite Hyperspectral Imager Spectra. *Remote Sensing*, *11*(18), p.2072.

Keshava and J. F. Mustard, "Spectral unmixing," *IEEE signal processing magazine*, vol. 19, no. 1, pp. 44–57, 2002.

Lawson, C. and Hanson, R. 1995, Solving Least Squares Problems, 1st edn. (Society for Industrial and Applied Mathematics).

Li, L., Mahowald, N., Balkanski, Y., Connelly, D., Ginoux, P., Ageitos, M. G., Hamilton, D., Kalashnikova, O., Klose, M., Miller, R. L., Obiso, V., Paynter, D. and Perez Garcia-Pando, C.: Large contribution of hematite and geothite to uncertainty in dust direct radiative forcing, Atmos. Chem. Phys., 2020.

Okin, G. S., Roberts, D. A., Murray, B., & Okin, W. J. (2001). Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. *Remote Sensing of Environment*, *77*(2), 212-225.

Roberts, D., M. Smith, and J. Adams. "Green vegetation, nonphotosynthetic vegetation, and soils in AVIRIS data." *Remote Sensing of Environment* 44.2-3 (1993): 255-269.

Roberts, D., Gardner, M., Church, R., Ustin, S., Scheer, G., & Green, R. 1998, Remote Sensing of Environment, 65, 267.

Roth, Keely L., Philip E. Dennison, and Dar A. Roberts. "Comparing endmember selection techniques for accurate mapping of plant species and land cover using imaging spectrometer data." *Remote Sensing of Environment* 127 (2012): 139-152.

Scanza, R. A., Mahowald, N., Ghan, S., Zender, C. S., Kok, J. F., Liu, X., et al. (2015). Modeling dust as component minerals in the Community Atmosphere Model: development of framework and impact on radiative forcing. *Atmos. Chem. Phys., 15*(1), 537-561. http://www.atmos-chem-phys.net/15/537/2015/

Schaaf, A. N., Dennison, P. E., Fryer, G. K., Roth, K. L., & Roberts, D. A. (2011). Mapping plant functional types at multiple spatial resolutions using imaging spectrometer data. *GIScience & Remote Sensing*, *48*(3), 324-344.