Use Case Document: Large-Scale Spectral Mapping (NAU)

Note: the red font indicates missing/partial/unclear information.

# Descriptions

* **Collaborator:** Dr. Mark Salvatore, future NAU graduate student
* **Field**: Terrestrial geology
* **Measure of success:** Spatial coverage of surveyed area, reasonable estimates on atmospheric contributions, comparisons to a spectral library of known geologic materials.
* **Primary challenges:** Need a pipeline for automated processing of satellite imagery, automated detection and removal of snow, ice, water, and shadows from the scene, automated atmospheric characterization and removal, and automated “stretching” of the scenes. Lastly, we also need to save all of these steps and archive them for future use.
* **Platforms used:**
  + **Scripting languages/library:** None at the moment, as these steps are done through human-used image processing software packages (e.g., ENVI)
  + **System:** Currently Windows 7 workstation, 16 GB RAM, 3.4 GHz processor
* **Physical Systems:** 
  + **System:** Multispectral (8-band) data acquired from the WorldView-2 and WorldView-3 satellite systems. Eventually will work the 16-band WorldView-3 VNIR and SWIR data.
  + **Parameters:**
    - * + Parameters required from image metadata files, including solar elevation, radiometric correction factors, and effective bandwidths.
        + Table of Earth-sun distances (derived from Julian calendar).
        + Spectral library of laboratory-derived spectral signatures of different geologic units. Includes latitude and longitude of the sampling locations, which will be used to “unmix” specific scenes.
* **Description:** A series of algorithms to calibrate multispectral data from raw digital number through to spectral parameters derived from calibrated surface reflectance data. Intermediate steps include the derivation of top-of-atmosphere radiance, the estimation of atmospheric contributions and their removal from the scene, the calibration to surface reflectance, the removal of “non-geological” surfaces (e.g., ice, snow, water, shadow), the parameterization of reflectance data, and the “spectral unmixing” of orbital data using a library of spectral endmembers.
* **Components:**
  + **Stage 1:** Input multispectral data, calibrate to top-of-atmosphere radiance using parameters drawn from image metadata file.
  + **Stage 2:** Automatically perform an atmospheric correction algorithm
    - Identify spectrally homogeneous regions with varying surface illumination and shadowing;
    - Collect ~200 spectra from these locations, ensuring that you are only cataloging variations in brightness (due to illumination) and not variations due to variable surface composition or other chemical/physical properties;
    - Regress a line through these data at each band relative to Band 8, deriving a Y-intercept for each regression, which predicts the radiance value for each band when completely in shadow (attributed solely to atmospheric scattering);
    - Subtract these Y-intercept values from each image band to calculate surface radiance (atmosphere removed) from top-of-atmosphere radiance;
  + **Stage 3:** Use data from uploaded tables (Earth-Sun distance, ESUN values) to calibrate from surface radiance to surface reflectance;
  + **Stage 4:** Automatically remove non-geological surfaces using automatically derived spectral values;
    - Linearly “unmix” the image with a series of spectra from a given library that will include ice, snow, and shadowed and illuminated rock surfaces
    - Pixels meeting some threshold of non-illuminated geologic rock surfaces will be removed from the scene, leaving behind only illuminated rocky surfaces to be characterized geologically
  + **Stage 5a:** Parameterize spectral data into known parameter mapping products
    - Generate simple spectral ratios and parameter products that can be tied to known geologic and geochemical phenomena
    - Relative values of these parameters can be used to understand the distribution of mafic compositions, oxidative weathering, and other surface properties
  + **Stage 5b:** Linearly “unmix” these illuminated geological surfaces using a different endmember library, derived automatically from a library of hundreds of samples based on which library data were derived from a latitude/longitude contained within the scene
    - Will hopefully generate unmixing models where each pixel can be assigned a given percentage of specific rock types
* **Stage 1: Calibrate to TOA Radiance**
  + **Description:** Data in each band will undergo a relatively simple band-specific mathematical function to convert the data from raw digital number to observed radiance. The equation for each band is as follows:  
    where RadλTOA is the band-specific top-of-atmosphere radiance, RadCalFactλ is the band-specific Radiometric Calibration Factor found in the image metadata, DNλ is the band-specific raw digital number data from the image, and EffBandλ is the band-specific effective wavelength value found in the image metadata.
  + **Python files:**
* None (at the moment)
  + **Cores:** ??
* **Input files:**
* Raw image data (~1.3 GB each, GeoTIFF file)
* Associated image metadata (~9 kB each, XML file)
* **Output files:**
* Top-of-atmosphere radiance file (~1.5 – 2.5 GB each, GeoTIFF file)
  + **Time for completion:** When run by hand, the calculation typically takes 3-6 minutes per image.
  + **Storage Space:** For the entirety of geology in the Antarctic, this will probably require several TB of storage. Will likely want to limit the number of images to reduce overlap.
* **Stage 2: Atmospheric Correction**
  + **Description**: This is likely the hardest task to automate, but there may be a way to get “close” by simply using a library of previously derived atmospheric spectra. The first step will be to identify geologic regions that are (1) spectrally homogeneous and (2) exhibit variations in illumination (e.g., shadowed and unshadowed terrain). Part 1 might be accomplished by comparing individual spectra to their neighbors to try to find large regions where spectral variability is minimal. Once Part 1 is identified, finding locations near these regions that are dark and exhibit spectral signatures consistent with shadows will be required. When these regions are identified, spectra across the homogeneous region under varying illumination scenarios must be collected, stored, and output. The next step requires the assumption that the last spectral band (Band 8) does not exhibit any influences from atmospheric scattering, which is a “good enough” sort of assumption. Once made, you can plot each band against Band 8 to find a (hopefully) linear series of data points. Because we’ve assumed that Band 8 exhibits no influence from atmospheric scattering, a Band 8 value equal to zero will indicate a completely shadowed pixel. So, we must now regress a trend line through the data to predict the Y-intercept of each band relative to Band 8. For example, if the predicted value for Band 1 is 40 when Band 8 is equal to zero, that means that the atmosphere is contributing 40 W m-2 sr-1 to Band 1 even when the pixel is in complete shadow. Subtracting each of these derived band-specific atmospheric contributions from the top-of-atmosphere radiance scene is the final step, which will convert the data from top-of-atmosphere radiance (surface + atmosphere components) to surface radiance (minus the atmosphere).
  + **Input:** Top-of-atmosphere radiance image.
  + **Output**:
* Text file with the collected spectra from the homogeneous area under differing illumination conditions;
* Surface radiance image (~1.5 – 2.5 GB each, GeoTIFF file)
* **Stage 3: Calibration to Surface Reflectance**
  + **Description**: Use a relatively simple algorithm to calculate surface reflectance at each band from the band-specific radiance file. This can be performed using the following equation:

where ReflλSurf is the derived band-specific surface reflectance, π is the term pi, equal to 3.14159, RadλSurf is the band-specific surface radiance derived in Stage 3, d is the Earth-Sun distance, derived from the look-up table uploaded earlier, ESUNλ is the band-specific solar irradiance, derived from a look-up table uploaded earlier, and θ is the solar elevation for that given image, derived from the original image metadata file.

* + **Input:** Surface radiance image, Earth-Sun distance look-up table, ESUNλ look-up table, and original image metadata file.
  + **Output**: Surface reflectance image (~1.5 – 2.5 GB each, GeoTIFF file)
  + **Storage Space:** ~200GB (images), ~3GB (database rasters)
* **Stage 4: Removing “Non-Geological” Surfaces**
  + **Description**: Perform a linear unmixing of the surface reflectance data using a pre-defined spectral library that contains snow, ice, and shadowed and illuminated rock surfaces. Any surface that is modeled at less than ~80% illuminated rock surfaces will be removed from the scene, as to eliminate all non-rocky and non-illuminated pixels from the scene.
  + **Input:** Surface reflectance image, endmember library to unmix the data.
  + **Output**:
* Unmixed image file, where each band corresponds to a modeled abundance of each endmember used in the unmixing process (~1.5 – 2.5 GB each, GeoTIFF file)
* Surface reflectance image, where all “non-geological” surfaces have been removed (~1.5 GB each, GeoTIFF file)
* **Stage 5a: Parameterization**
  + **Description**: Employ several “standard” equations to this newly derived surface reflectance dataset to highlight relative spectral variability (RSV). These equations may include (but are not limited to) the equations defined in Salvatore (2015).
  + **Input:** Surface reflectance image, where all “non-geological” surfaces have been removed.
  + **Output**: Several one-band parameter images of the initially input reflectance image (~1 GB each, GeoTIFF file)
* **Stage 5b: Unmixing Images to Derive Surface Geology**
  + **Description**: Similar to Stage 4, we will linear unmix the surface reflectance image using an endmember library with only geological materials. This endmember library has already been created by Salvatore, and includes samples from throughout the Transantarctic Mountains. Salvatore will provide a “standard” endmember library that can be used to unmix any scene. However, it would be nice to automatically search through the spectral library for the corresponding sample latitude/longitude to see if any samples were acquired from within the imaged area. If so, replacing one of the “standard” endmembers with this true surface endmember would be ideal. In the most ideal instances, there will be multiple geologic targets acquired within a given image, and so the endmember library used to unmix the data will be completed derived from within the scene.
  + **Input:** Surface reflectance image, where all “non-geological” surfaces have been removed, as well as a geological endmember library.
  + **Output**:
* Unmixed image file, where each band corresponds to a modeled abundance of each endmember used in the unmixing process (~1.5 – 2.5 GB each, GeoTIFF file)