A spectral climatology for atmospheric compensation of hyperspectral imagery

John H. Powell*[a] and Ronald G. Resmini[a]
[a]College of Science, George Mason University, 4400 University Drive, Fairfax, VA 22030

ABSTRACT

Most Earth observation hyperspectral imagery (HSI) detection and identification algorithms depend critically upon a robust atmospheric compensation capability to correct for the effects of the atmosphere on the radiance signal. Atmospheric compensation methods typically perform optimally when ancillary ground truth data are available, e.g., high fidelity in situ radiometric observations or atmospheric profile measurements. When ground truth is incomplete or not available, additional assumptions must be made to perform the compensation. Meteorological climatologies are available to provide climatological norms for input into the radiative transfer models; however no such climatologies exist for empirical methods.

The success of atmospheric compensation methods such as the empirical line method suggests that remotely sensed HSI scenes contain comprehensive sets of atmospheric state information within the spectral data itself. It is argued that large collections of empirically-derived atmospheric coefficients collected over a range of climatic and atmospheric conditions comprise a resource that can be applied to prospective atmospheric compensation problems. A previous study introduced a new climatological approach to atmospheric compensation in which empirically derived spectral information, rather than sensible atmospheric state variables, is the fundamental datum. The current work expands the approach across an experimental archive of 127 airborne HSI datasets spanning nine physical sites to represent varying climatological conditions. The representative atmospheric compensation coefficients are assembled in a scientific database of spectral observations and modeled data.

Improvements to the modeling methods used to standardize the coefficients across varying collection and illumination geometries and the resulting comparisons of adjusted coefficients are presented. The climatological database is analyzed to show that common spectral similarity metrics can be used to separate the climatological classes to a degree of detail commensurate with the modest size and range of the imaging conditions comprising the study.

Keywords: Hyperspectral, HSI, atmospheric compensation, empirical line method, remote sensing, climatology

1. INTRODUCTION

The fundamental problem addressed in this research is atmospheric compensation in an Earth remote sensing context. For an imaging sensor at some altitude above the Earth's surface, atmospheric compensation is the process of deriving the surface reflectance values from the at-aperture radiance images recorded by the sensor. The magnitude of the atmospheric effects on measured electromagnetic energy can be strongly wavelength dependent, varying across the absorption regions of water vapor and trace gas constituents in the atmosphere. Scattering by molecules and suspended aerosol particles is also wavelength dependent. Successful analysis of remotely sensed hyperspectral imagery (HSI) is particularly dependent upon a robust atmospheric compensation capability. Most HSI applications rely on precise relationships between spectral bands and virtually any quantitative HSI analysis must therefore begin with an inversion problem to derive the surface reflectance or emittance from the measured at-aperture radiance. Many methods have been developed to accomplish this inversion. Most can be categorized as either empirical or physics-based methods.

HSI datasets contain complete sets of spectral measurements of light passing through the atmosphere at each pixel; therefore, information about the atmospheric transmission is present in the measured radiance signal. Empirical atmospheric compensation methods use this information along with some additional information about the scene to

*jpowell8@gmu.edu

statistically derive the relationship between radiance and reflectance. Alternatively, the physics of radiative transfer is well understood and can be accurately modeled using radiative transfer algorithms. Physics-based (PB) methods use radiative transfer codes to estimate the atmospheric effects on transmission and determine the surface reflectivity from the model. Both approaches perform best when ancillary ground truth data are available – high fidelity in situ radiometric observations and/or atmospheric profile measurements.

In practice, outside of controlled experiments, detailed ground truth data are rarely available. PB methods are routinely applied using broad climatological atmospheric parameters as input to the models. PB methods can use the atmospheric transmission information contained within the radiance data to refine the parameters; for example by using band ratios to extract the information and convert it to conventional meteorological parameters (water vapor mixing ratios, aerosol concentrations, etc.). The PB models then use radiative transfer algorithms to translate the meteorological information back into spectral effects during the reflectance inversion. Undesirable artifacts are inevitably introduced into the data with each translation between domains. Empirical methods must rely on indirect methods to supply reference reflectance signatures, either through supervised or statistical means; no climatologies currently exist for empirical methods.

Where well-calibrated ground truth measurements are available, the empirical line method (ELM) has been shown to produce reflectance retrievals that are overall equivalent or superior to those produced by other methods.[1] However, the native spectral information about the atmosphere captured in the ELM coefficients is not compiled for use outside of the scene from which they were derived. These coefficients comprise an untapped resource for climatological information relevant to HSI atmospheric compensation. In a previous work[2] we introduced a new type of atmospheric climatology, using a statistical approach in which empirically derived spectral information is the fundamental datum rather than sensible atmospheric state variables. This paper expands the empirical techniques for extracting the coefficients and correcting for small nonlinear features, and updates the modeling methods used to standardize the coefficients across varying collection and illumination geometries. The resulting climatological database is analyzed to determine the separability of the climatological classes.

2. BACKGROUND

2.1 Empirical Line Method

The empirical line method[3] is a proven empirical method of atmospheric compensation in which the conversion factors to retrieve reflectance values are derived by performing a linear regression of observed at aperture radiance values against corresponding measured ground truth reflectance factors. The context of this model is hyperspectral Earth remote sensing in the visible through short wave infrared part of the spectrum. The radiance reaching the sensor L_S can be written[2] as:

$$L_S = m\rho + b \tag{1}$$

where m and b are the gain and offset vectors given by:

$$m = \left(E_0 \tau_d \cos\sigma + E_s\right)\frac{\tau_u}{\pi} \tag{2}$$

$$b = L_{path}\,. \tag{3}$$

E_0 is the solar irradiance at the top of the atmosphere, τ_d is the downward path transmittance, E_s is the skylight irradiance at the surface, τ_u is the upward path transmittance, ρ is the surface reflectance factor, and σ is the incident angle to the surface. Each quantity is a function of wavelength. Here it is assumed that the entire hemisphere of the sky is visible to the surface and the surface is Lambertian. Adjacency and multiple surface scatter effects are neglected, as are thermal emissive radiance contributions.

In ELM, the gain and offset coefficients m and b in equation 1 are assumed constant across the image, and are therefore one-dimensional vectors in wavelength space. They are determined by selecting two or more groups of pixels for which the reflectance values are known (or assumed known) and performing a linear regression of the measured radiance against the ground truth reflectance. These vectors are then applied against each pixel to estimate the reflectance.

2.2 MODTRAN

The Moderate Resolution Transmittance (MODTRAN) radiative transfer code serves as the U.S. Air Force standard moderate spectral resolution radiative transport model for wavelengths extending from the visible through the thermal infrared region.[4,5] MODTRAN solves the radiative transfer equations from a fundamental physics approach, using a narrow band model of molecular and particulate absorption, emission, and scattering, as well as surface reflection and emission. The software models the solar and lunar illumination based on geographic location, date, and time. It can simulate a number of different remote sensing geometries, including the airborne Earth surface sensing application treated in this study. The atmosphere is modeled as stratified layers that can be user-defined or defaulted to one of several standard climatological profiles. MODTRAN is also used as the radiative transfer engine for many atmospheric compensation programs, including the commercial standard Fast Line-of-sight Atmospheric Analysis of Spectral Hypercubes (FLAASH).[6]

MODTRAN provides great flexibility for the user to specify input parameters defining the model operating modes and features, as well as the environmental conditions. The use of MODTRAN in this study was not intended to provide the absolute most realistic simulation for a given set of conditions, but rather to generate a reasonable estimate of the effects of changing imaging geometry and illumination. In this research, the parameters that were exercised consist mainly of geometric, spatial and temporal inputs (for illumination definition), and of those describing the modeled atmospheres. MODTRAN is used here to adjust ELM coefficients derived from different HSI scenes to account for differences in geometry and illumination conditions.

2.3 QUAC

The Quick Atmospheric Correction (QUAC) model of Bernstein, et al. is an unsupervised empirical atmospheric correction algorithm.[7] QUAC assumes a linear radiative transport equation like ELM, but uses a ratio of scene-derived statistics to those of a reference scene to calculate the gain coefficients. The reference scene is a spectrally diverse collection of laboratory reflectance measurements.

QUAC has become a prevalent method for visible to short wave infrared (VNIR/SWIR) HSI atmospheric compensation among practicing HSI analysts. It is fast, needs no external ground truth or atmospheric information, is tolerant of radiometric uncertainty, and highly robust. Even when other methods are ultimately used, QUAC often serves as a baseline for comparison. QUAC's main deficiency is in absolute accuracy of the reflectance returns, shown to be accurate within approximately 15% compared to the best FLAASH results (best meaning highly accurate radiometric input and characterization of the atmosphere).[7] QUAC is included in this research because it is a widely used, well-documented approach that yields consistent relative accuracy. QUAC gain and offset coefficients are analogous to ELM coefficients; however, QUAC defines the gain to be $\rho/(L-b)$ rather than equation 1, so the QUAC coefficients were inverted for direct comparison to ELM coefficients.

2.4 Hyperspectral data

The hyperspectal data used in this research was collected by the HYDICE sensor between the years 1995 and 2000 over a range of climatic regions, backgrounds and seasons. Each collection was accompanied by ground truth information to characterize the scene. HYDICE was a pushbroom hyperspectral sensor with a spectral range of 0.4 to 2.5 micrometers (VNIR/SWIR). It used a Schmidt prism spectrometer with a single indium antimonide (InSb) focal plane. HYDICE collected 210 spectral bands with a nominal bandwidth of 10 nm and 320 spatial samples. A 0.5 mrad instantaneous field of view (IFOV) produced ground sample distances (GSD) ranging from approximately 1 m to 4 m at typical operating altitudes (5000 to 20,000 ft above ground level (AGL), respectively).[8] The HYDICE sensor employed an onboard tungsten-halogen calibration source for in-situ calibration measurements. Error sources have been well studied and absolute radiometric uncertainty is approximately 5%.[9,10]

HSI scenes were collected over a variety of environments representing distinct climate regimes. In each environment, data were collected at several altitudes and with varied illumination conditions. Each environment was typically collected over a 2-4 day period. Several of the environments were imaged in different seasons. Each scene analyzed contained four to six gray scale calibration panels ranging from 2% to 64% reflectance. The calibration panels were measured with a field spectrometer, averaging 5-15 spot measurements across the panel to provide ground truth reflectance spectra for the scenes. In most cases the panels were measured on the day of the collects. The panels were sized relative to the HYDICE instantaneous field of view to ensure that at least one image pixel would fall entirely on the calibration panel at maximum flight altitude, providing a pure pixel corresponding to the ground truth reference spectra. A total of 181 HSI images were

used in the research, 127 of which produced atmospheric compensation coefficients. Most of the remaining 54 images lacked acceptable ground truth information or metadata to permit full analysis.

2.5 Imaged sites

Sites included in the study were selected for diversity of climate, ground cover and season. The images included in the research encompass 14 collections over nine geographic sites. The sites include continental plain and mountainous terrain as well as littoral regions. Environments range from tropical to mid-latitude temperate to arid conditions. Ground elevations range from sea level to nearly 10,000 ft. Background land cover includes bare earth, open shrub, agricultural vegetation, forest and urban environments. Each collection event occurred over a period of two to five days, and three to five images from each day are included in this research. Imaging times are from 9 am to 3 pm local time, with the majority of the images nearer to local noon. Imaging conditions were mostly clear skies but occasionally included broken clouds or thin high clouds.

Table 1 summarizes the characteristics of the sites and images used in the research; Table 2 and Table 3 define the associated climate and land cover classes. Land cover classes are based on USGS National Land Cover Database (NLCD) 2001 categorization[11]. The classification scheme was developed for remote sensing-based land classification, specifically the LANDSAT Thematic Mapper. In this research, categories are assigned by manual review of the HSI imagery in true color and infrared false color composites compared to the descriptions in the classification system. Two categories are assigned for each site. The primary category refers to the immediate environment surrounding the calibration panels (within approximately 10 pixels). The secondary category describes the dominant land cover of the region (within approximately 500 m).

Table 1 Summary of site characteristics. Climate and land cover categories are described in the following tables.

Site	Climate Category	Elevation (ft)	Land Cover Category (pri, sec)	Seasons Imaged (MMM)	Number of Images
1	H	6810	31, 51	SON	19
2	H	8498	31, 42	SON	12
3	H	9754	31, 51	SON	12
4	BW	5267	31, 51	JJA, DJF	8
5	BW	786	31	JJA, SON, DJF	24
6	Cs	62	23, 51	JJA	4
7	Cf	1025	71, 82	JJA	11
8	Dc	18	71, 41	JJA	22
9	Ar	193	71, 41	MAM	15

Table 2 USGS Land Cover Institute (LCI) land cover class definitions.

Land Cover Category	Level I Class	Level II Class
23	Developed	Commercial/Industrial/Transportation
31	Barren	Bare Rock/Sand/Clay
41	Forest Upland	Deciduous Forest
42	Forest Upland	Evergreen Forest
51	Shrubland	Shrubland
71	Herbaceous Upland	Grasslands/Herbaceous
82	Herbaceous Planted/Cultivated	Row Crops

Table 3 Köppen–Trewartha climate classification categories.

Climate Category	Climate Class	Type
Ar	Tropical	Rainy
BW	Dry	Arid or desert
Cf	Subtropical	Humid
Cs	Subtropical	Dry-summer maritime subalpine
Dc	Temperate	Continental
H	Highland	N/A

The modified Köppen climatic classification system as described by Trewartha[12,13] is used to characterize the climatic zones of the sites. The Trewartha-Köppen system considers prevailing atmospheric factors (e.g., temperature, humidity, rainfall) as well as geographic and geological aspects in defining the classes. This produces climatic regions that are better related to dominant flora and soil types, and thus better related to the typical atmospheric characteristics than are strictly geographic/morphological categorization systems.

3. METHODOLOGY

The central argument of this study is that empirically derived reflectance inversion coefficients can be used to characterize the atmosphere. The mathematical model used to represent the observed radiance is described in section 2.1, reducing the reflectance retrieval to the form of the empirical gain and offset coefficients. The methodology to derive the coefficients is presented in this section, as well as the analysis employed to evaluate and correct the coefficients.

ELM coefficients are specifically tied to the illumination and imaging geometry present at the time of the image from which they are derived, as all the factors affecting the radiative transfer are encapsulated within the coefficients. In order to isolate the atmospheric effects and to compare coefficients amongst different images, a method to account for varying illumination and geometry is required. This methodology is described in the latter part of this section.

3.1 ELM analysis

Standard ELM techniques were used to derive gain and offset coefficients that describe the inversion from at aperture radiance to reflectance.[3] The method is described in detail in a preliminary study.[2] ELM produces vectors *m* and *b* containing the gain and offset coefficients for each channel and an RMSE vector *r*. An example of the coefficients is shown in Figure 1. HYDICE radiance cubes were provided in scaled spectral radiance units of (1/75) * W / m2 sr μm. All radiance quantities presented in this study are converted to "microflick" units (μW / cm2 sr μm), and all gain coefficients are scaled such that they produce standard radiance units when applied to normalized reflectance values (ranging from 0 to 1). The band center wavelengths of the HYDICE sensor shift slightly from collect to collect due to prism distortion, so the sensor is calibrated for each flight path. To facilitate comparison, all ELM coefficients were resampled to the mean wavelengths averaged over all collects. The gain and offset coefficients are plotted on different axes in order to overlay them on the same plot. These coefficients comprise the basic observational quantity used in the study.

Figure 1 (left) depicts a classic, physically reasonable shape for the ELM coefficients – rising to a peak in the visible region and exponentially decaying toward the longer wavelengths – with relatively low residuals. The majority of the images produced ELM coefficients with similar spectra, as expected. A sizable number of images, however, produced coefficients with unexpected offset spectra. Normally when ELM is used to obtain a reflectance image, the shape and values of the coefficients are of little consequence as long as they produce an acceptable reflectance inversion. For that reason the coefficients themselves have been rarely studied in the literature. In this application, however, we attribute the gain and offset to a physical partitioning of radiative transfer effects, so non-physical anomalies are important to address.

The artifacts of greatest concern are similar to those depicted in the right hand side of Figure 1. The areas of negative offset values are completely nonphysical from the standpoint of our imaging model, i.e., the offset representing the path radiance. These artifacts appear episodically; for instance, several images might produce normal coefficients, and subsequent images on the same flight using the same ground truth produce coefficients with the negative gain artifacts. For this reason, the artifacts are attributed to the ELM linear regression not fully partitioning the radiative transfer in the way described by the

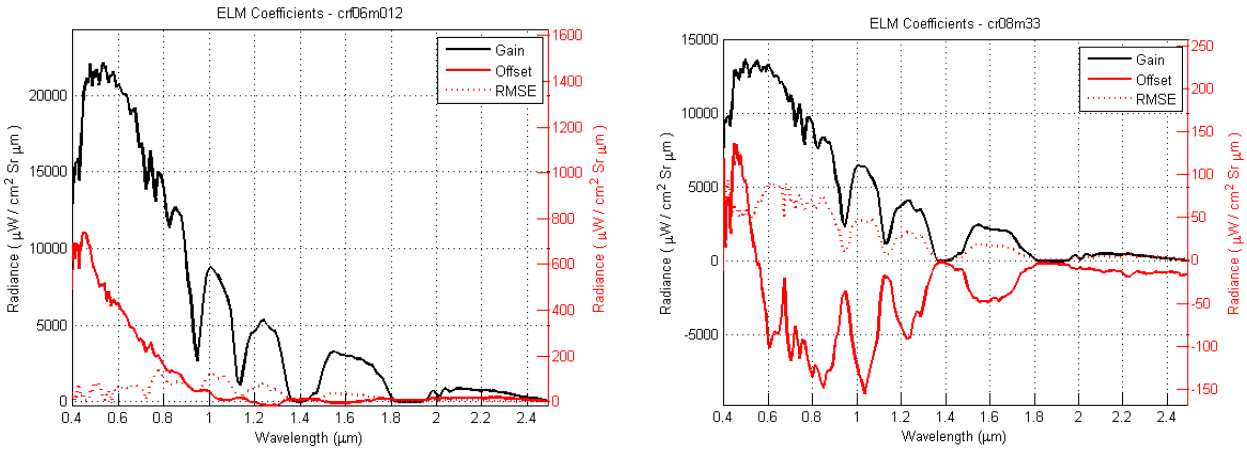

image model, rather than some systematic error in the analysis.

Figure 1 Example of typical ELM coefficients. The left y-axis scales the gain (*m*, in black) and the right y-axis scales the offset and RMSE residual (*b* and *err*, in red). Left plot (Site 4, 1998-08-25, 17:14z) shows typical behavior; right plot (Site 5, 1995-10-21, 20:20z) exhibits strongly negative offset values from 0.55 to 1.8 µm.

3.2 ELM coefficient correction

It should perhaps not be surprising that this sort of variability is observed in the offset coefficients, particularly when the RMSE residuals are large. ELM is fitting a line to a set of 4-6 radiance-reflectance pairs (one set per wavelength), and scatter in the data can easily shift the intercept when the values are so small relative to the gain (commonly over an order of magnitude smaller). In a preliminary paper on this work, the issue was minimized by leaving one or more reflectance panel observations out of the ELM analysis, trying various combinations to reduce RMSE and improve the offset coefficients[2]. This was largely effective at producing more physical offset coefficients, but there was no objective measure to indicate whether the regression was improved or harmed by the reduction of observation points because the goodness of fit is often improved simply by reducing the number of data points from four to three, for example.

To understand the causes of the negative offset coefficients, it is necessary to scrutinize the wavelength-by-wavelength linear regressions in the ELM process. This was done by performing the linear regression at each wavelength in sequence, animating the line fit against the data points, and summing up the residuals at each point. Every scene has a very dark panel, approximately 4% reflectance, that was useful to show how well the linear fit is "anchored" at low reflectance value near the y-axis. It was observed that while there were regions of scatter-related negative offsets (i.e., the linear fit passing below the dark panel observation point due to scatter in the data), that was not a consistent pattern in the large regions of the spectrum where the negative values occur. This was born out by the arithmetic sums of the dark panel residuals, which did not in general show a negative bias in the problem regions. The pattern that emerges is a slight apparent nonlinearity in the dark tail of the distribution, such that the slope of the line is reduced near the y-axis.

Second and third order fits were compared, but over the full range of wavelengths either worsened the overall fit or over-fit the data in images containing only four panels, in agreement with the body of work indicating a linear relationship. Instead, a small nonlinear term was added to equation 1 of the form:

$$\frac{h}{g\rho+1}. \tag{4}$$

This term has value h at $\rho = 0$, decaying toward zero with an effective width defined by the parameter g (Figure 2). To constrain the parameters g and h, we require that the slope be positive at $\rho = 0$, requiring:

$$g \le \frac{m}{h}. \tag{5}$$

To express the constraints in terms of departure from linearity as shown in figure, we introduce a new parameter α such that $g=m/\alpha h$. Then the slope at ρ = 0 is given by:

$$slope_0 = m\left(1-\frac{1}{\alpha}\right), \tag{6}$$

so α can be used to constrain the departure of the slope from m and h gives the departure of the intercept from b.

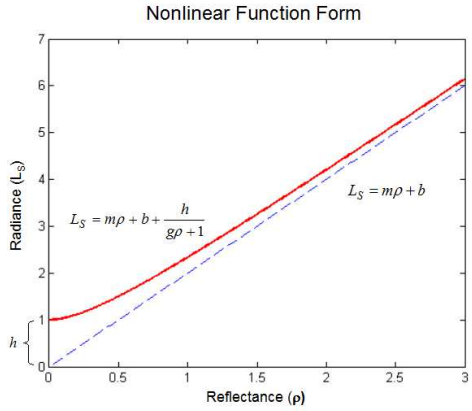

Figure 2 Generalized nonlinear correction function (solid red) and base linear function (dashed blue).

The correction can be incorporated in the ELM regression as follows. The ELM equation with the nonlinear correction is linearized as:

$$L_S = m\rho + b + h\delta, \tag{7}$$

where δ is an estimated nonlinear term given by:

$$\delta = \left(\frac{m\rho}{\alpha h_{est}}+1\right)^{-1}. \tag{8}$$

Here h_{est} is an initial estimate of h. The linearized form (equation 7) can then be fitted to the data using conventional linear regression. The problem remains of how to determine h_{est} from the data. An algorithm was developed to calculate the estimated intercept in four ways: 1) linear regression (conventional ELM); 2) second order polynomial fit; 3) third order polynomial fit; and 4) using the linear regression value of the slope but forcing the line to exactly intersect the lowest reflectance/radiance pair. These four vectors (across all wavelengths) were evaluated by a set of objective criteria to reject sets that deviated too greatly from the ELM result (in peak magnitude), and those that had significant negative values. The estimated intercepts from the remaining methods were averaged to determine h_{est}. The α parameter controls how quickly the nonlinear term falls to zero and was set empirically to a constant value that matched the bulk of the data well.

Equation 7 was then fitted to the data using linear regression and a set of nonlinear ELM coefficients was generated. Because the corrections are small in all cases and negligible in the majority of cases, it is desirable to keep the improved

offset values but still form a linear fit to the data. The linear regression was therefore repeated by forcing the offsets to the nonlinear values and optimizing the gain to minimize the RMSE residuals. These adjusted linear coefficients were used in the end analysis for all coefficients for consistency. The difference in the goodness of fit due to the adjustment procedure averaged across the entire collection is shown in Table 4. Over all images used in the study, the nonlinear function provided a slightly better fit than the linear regression. The adjusted linear regression produced larger residuals as expected, but only 5.6% greater than the linear case on average.

Table 4 RMSE residuals for linear, nonlinear, and adjusted linear regressions, averaged across all wavelengths and all image coefficients in the study.

	Linear Regression	Nonlinear Regression	Adjusted Linear Regression
RMSE (μW / cm^2 sr μm)	65.50	62.18	71.13

The improved fit of the nonlinear regression suggests that the functional form shown in Figure 2 more accurately matches the data than the linear form. Care was taken to minimize the departure from linearity, and to confine the nonlinearity to the low reflectance region of the relationship, so it is unlikely to be a result of over-fitting with a higher order polynomial. Additional tests were conducted on the reflectances returned from the adjusted linear coefficients, comparing to ground truth measurements, to validate the fact that the adjustment procedure did not adversely affect the end reflectance retrievals. The specific form of the nonlinear term in fact has a physical basis. Conel develops the form as an effect of the background spherical albedo, or the portion of incident radiance that is scattered isotropically.[3]

3.3 Coefficient adjustment

ELM derived coefficients correct for illumination and geometric factors as well as atmospheric effects. In order to compare coefficients derived from images collected under differing imaging conditions, it is necessary to modify the coefficients for variations in illumination and altitude of the observation. To diagnose these effects, the scenes were modeled using the MODTRAN radiative transfer software. MODTRAN can also provide solar illumination based on geographic position, date, and time. Thus the terms in equation 1 can be modeled explicitly.

In preliminary work, a “two stream method” was used to model the radiative transfer terms in equation 1, in which two MODTRAN simulations were run. Input parameters were identical except for the ground surface albedo, which was set to zero in one run and one in the other. This isolated the path radiance term and allowed the other terms to be calculated using ratios of the MODTRAN outputs.[2] The method was not optimal for two reasons. The method tries to force the PB model to provide radiative partitioning to the terms in equation 1, but the MODTRAN output fields are not consistent with that model and the results are therefore inexact. Second, setting the ground albedo to one across the entire field of view is an artificial extreme that amplifies any errors or artificialities present in the scattering model.

Fortunately, the updated MODTRAN 5.3 software provides access to radiometric fields that directly align with the model described in section 2.1. The derivation is fully developed in Appendix G of the MODTRAN 5.3 User’s Manual;[14] the salient relationship is equation 11 from the reference:

$$L_S \approx L_0 + \frac{A\rho + B\bar{\rho}}{1 - \bar{\sigma}\bar{\rho}}, \tag{9}$$

where ρ is target (pixel) reflectance, $\bar{\rho}$ is the average reflectance of the surrounding area, and $\bar{\sigma}$ is the spherical albedo. L_S is the sensor radiance and L_0 is the sensor radiance for the zero reflectance case. A and B are numerically derived coefficients described below. (The variable names have been changed from the reference to be consistent with the model developed in section 2.1, and the wavelength notation has been dropped. All variables except $\bar{\sigma}$ are wavelength dependent, and radiances L_S and L_0 are integrated over the instrument spectral channel.)

Each of the variables on the right hand side of equation 9 is either an input to the simulation or can be calculated from MODTRAN output data. Coefficient A is defined as the product of the total transmitted solar irradiance and the sensor-to-ground direct transmittance, and B is the product of the total transmitted solar irradiance and the sensor-to-ground diffuse transmittance, each convolved with the channel spectral response function (SRF)[14]. Spherical albedo is the fraction

of the incident irradiance that is reflected by the surface in all directions, summed over all wavelengths. This radiative transfer construct can be related to the ELM model by writing equation 8 as:

$$L_S = m^* \rho + b^* \tag{10}$$

where:

$$m^* = \frac{A}{1 - \bar{\sigma}\bar{\rho}} \tag{11}$$

$$b^* = L_0 + \frac{B\bar{\rho}}{1 - \bar{\sigma}\bar{\rho}}. \tag{12}$$

The star superscript is introduced to specify modeled rather than empirical coefficients. Equation 11 can be thought of as modifying the gain by accounting for the contribution of the spherical albedo term, and equation 12 as separating the offset into direct and diffuse parts.

Using equations 11 and 12, PB modeled gain and offset coefficients can be computed from the output of a single MODTRAN run with ρ = 0. This approach has been shown to be superior to two- or three-stream methods[14]. The computation of the coefficients is simple as the variables are readily available in the MODTRAN output. The ability to model the gain and offset coefficients is the key enabler for this research. The PB model allows us to simulate coefficients under a variety of observational conditions, which can be used to standardize ELM coefficients so they can be compared and studied together as a whole, despite being derived under varying observational conditions.

3.4 Coefficient standardization

One of the strengths of ELM is that the method implicitly accounts for illumination conditions, imaging geometry in addition to correcting for the atmospheric propagation and secondary surface interactions. However, the goal of this research is to use the empirical coefficients to compile a broad climatology relevant to HSI remote sensing. Therefore the illumination and geometric factors implicit in the coefficients must be factored out before the coefficients can be compared.

As described in the preceding section, modeled gain and offset coefficients can be computed for the imaging parameters relevant to each set of empirical coefficients derived from the imagery. The modeled coefficients can then be used to estimate corrections to the ELM coefficients for differing illumination and sensor altitude above ground. All images are nadir-looking, so geometry is completely described by the altitude above ground level. For a set of ELM coefficients m_1 and b_1 derived from one image, modeled coefficients m_1^* and b_1^* are computed for the imaging conditions using equations 11 and 12 as described above. To estimate the ELM coefficients under differing conditions, at a later time, for instance, modeled coefficients m_2^* and b_2^* are computed for the new conditions. Scale corrections are then computed from the modeled coefficients:

$$M_{21} = m_2^* / m_1^* \tag{13}$$

$$B_{21} = b_2^* / b_1^*. \tag{14}$$

Then the ELM coefficients at the new time are estimated as:

$$m_2' = M_{21} * m_1 \tag{15}$$

$$b_2' = B_{21} * b_1 \tag{16}$$

where the primes delineate estimated (adjusted) ELM coefficients.

Using this procedure it is possible to model the effects of differing illumination and geometry and to scale the ELM coefficients accordingly. This provides an estimate of what the ELM coefficients would be under differing imaging conditions and therefore allows comparison of coefficients across varied times and geometries. The atmospheric inputs to the PB model are identical in the two runs, so the scale correction will only account for changes in the illumination and imaging geometry. The variability that remains is assumed to be due to differences in the atmosphere. This method leverages strengths of the PB approach, e.g., calculating precise illumination and high fidelity atmospheric propagation, but by applying the *ratios* of modeled coefficients, it has the advantage of offsetting any systematic errors in the modeled results. Any artifacts caused by artificialities in the scattering models are present in both model runs and, to first order,

cancel each other out. The accuracy of the estimated coefficients will decrease as the magnitude of the change in imaging conditions increases, but within some bounds of variability, the method produces accurate estimates.

4. ANALYSIS

The standardized ELM coefficients constitute the basis for the climatological analysis. When properly standardized to remove the effects of illumination and scene geometry variability, they are directly comparable and are analyzed as proxy atmospheric variables (transformed through atmospheric transmission and the ELM process into spectral space). A series of optimizations and validations are performed to produce the most consistent standardization of the coefficients. Lastly, the standardized coefficients are clustered and analyzed to develop separable climatological classes.

4.1 Modeled coefficient optimization

In preliminary work, the coefficient standardization method was shown to be resilient to magnitude errors in the modeled coefficients, for the reasons cited in section earlier[2]. However, unusual scattering behavior was noted in some environments. MODTRAN provides tremendous flexibility to tailor the environmental inputs and it is desirable to use the most realistic modeling results possible in the analysis. To that end, an optimization study was completed for each site to determine the optimal set of input parameters to use.

The list of major radiative transport driver parameters used in the all trials is given in Table 5. Complete descriptions of all input parameters and their use are contained in the MODTRAN Users Manual[14]. Multiple scattering and the first-principle, plane parallel atmosphere discrete ordinate multiple scattering algorithm (DISORT) are required to produce the atmospheric compensation outputs used to compute the modeled coefficients, as are the related DSAZM and DSALB parameters. Surface reflectance is specified as Lambertian, and the surface skin temperature is set to 1 K to eliminate thermal emission in the computation of the reflected/scattered fluxes. The full 1 cm-1 resolution band model is used in the final MODTRAN runs.

Table 5 List of major radiative transfer driver parameters (common to all MODTRAN runs).

Input card	Field name	Value	Meaning
1	MODTRN	M	Use MODTRAN band model
1	ITYPE2	2	Slant path between two altitudes
1	IEMSCT	2	Spectral thermal plus solar/lunar radiance
1	IMULT	1	Multiple scattering mode
1	SURREF	LAMBER	Lambertian surface approximation
1	TPTEMP	1	Boundary temperature of image pixel (K)
1A	DISORT	t	DISORT scattering algorithm is used
1A	DISAZM	t	Azimutal depedence enabled for DISORT
1A	DISALB	t	Sperical albedo calculated
1A	LLFLTNM	t	Apply instrument SRF filter
1A	H2OAER	t	Aerosol properties modified per water vapor specification
1A	LBMNAM	f	Band model data file name (defines model resolution)
1A	CO2MX	390	CO2 mixing ratio

The MODTRAN input parameters defining the sensor geometry and illumination conditions are populated from the image metadata, specifically geographic location, date, time, sensor altitude, and ground elevation. Those related to atmospheric profiles (aerosol and water vapor models), seasonal models, scattering models and surface reflectance are also specified based on the characteristics of the site and environment. The multiple scattering mode includes radiance that scatters off of the background surface (not in pixel IFOV) and into the sensor. Experimentation with the generic MODTRAN backgrounds showed the modeled offset coefficients to be sensitive to these values, and best values did not always correspond to the type of land cover in the images. Representative HSI scenes were analyzed to extract mean reflectance signatures from each site; these were written to the MODTRAN library data and used as the background reflectance spectra. The spatial extent of the background that will contribute to the image will vary depending on altitude and atmospheric conditions, but an average effective radius of ~25 pixels was used. This is consistent with the background smoothing kernel size used in FLAASH[15].

Much more optimization was required to set the scattering related input parameters. An initial run was executed for image at each site, using the most appropriate settings for the known environments. With that run as a baseline, parameters controlling atmospheric profile, aerosol extinction model, visibility, and wind speed were modified to find the best set of

input parameters. At each iteration, the differences between the modeled and corresponding actual ELM coefficients, averaged over all images at the site, served as the metric for suitability for that site.

4.2 Modeled coefficient optimization

To validate that the coefficient standardization can adjust for illumination variations, the method is tested against a time series of images at a single site. The series of seven images were collected from similar altitudes (~10,000 ft AGL) over a 90-minute period. The ELM coefficients from the first image (10:14 Local time) ELM coefficients were adjusted to the time of the last image (11:42L) using the method described in section 3.4. Figure 3 shows the ELM coefficients before the illumination adjustment, and Figure 4 shows the coefficients standardized to 11:42L. The figures show that the gain variability from the change in illumination is almost completely corrected. The offset adjustment is not as complete, but the difference between the coefficients is reduced by the standardization by approximately 50%.

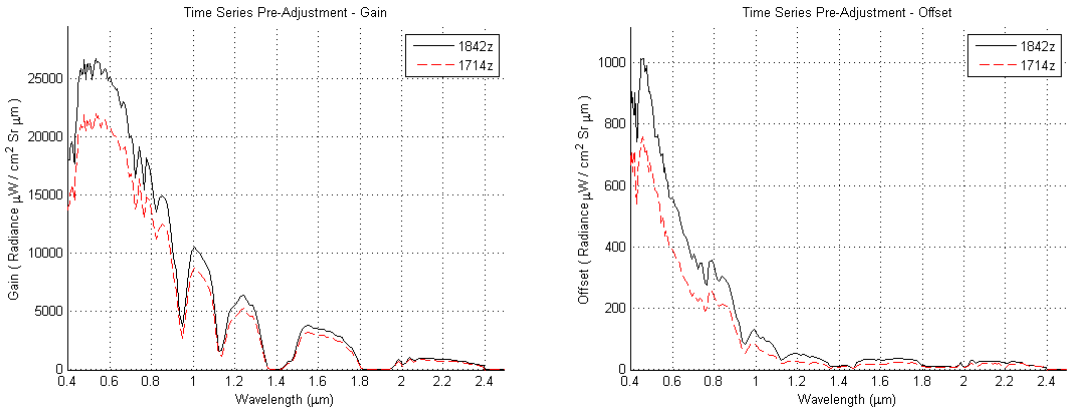

Figure 3 Gain (left) and offset (right) ELM coefficients for beginning and ending times in image sequence.

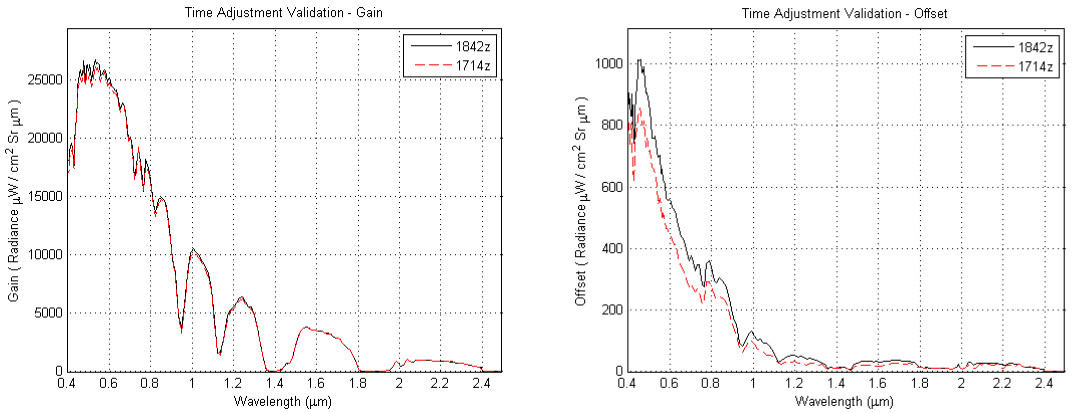

Figure 4 As in Figure 3 but after 1714z coefficients are adjusted to 1842z.

4.3 Geometric adjustment validation

Since all images in the study are nadir looking, the only major geometric variables are the sensor altitude and site elevation, i.e., the sensor altitude referenced to ground level (AGL). The validation method used above was repeated, but where the two observations differ in altitude in addition to time. In this example, the first scene was imaged at 12:48 PM local time from an altitude of 5.0 kft AGL. The second scene was imaged 26 minutes later from an altitude of 10.4 kft AGL. Figure 5 shows the ELM coefficients before the geometric adjustment, and Figure 6 shows the coefficients standardized to 10,407 ft AGL.

The pre-adjustment gain coefficients changed little between the two observation times. This is because the altitude change (higher altitude means greater transmission loss) and the time change (higher sun angle means greater illumination) have opposing influences on the gain coefficient. The standardization routine decreases the gain for altitude change but increases gain for the time difference. With both geometry and illumination changing, the error is expected to be higher than in the illumination only case, but the size of the difference suggests that the geometric correction has higher uncertainty. This makes physical sense, because in the altitude correction, the PB model is simulating transmission path that is not present in the ELM coefficients that are being standardized. The results are therefore sensitive to variability in the atmospheric properties with altitude. The gain is corrected to within 10% RMS and the offset to within 60% RMS, averaged across the spectrum (excluding low SNR wavelengths in the broad water absorption bands and where the offset values are near zero).

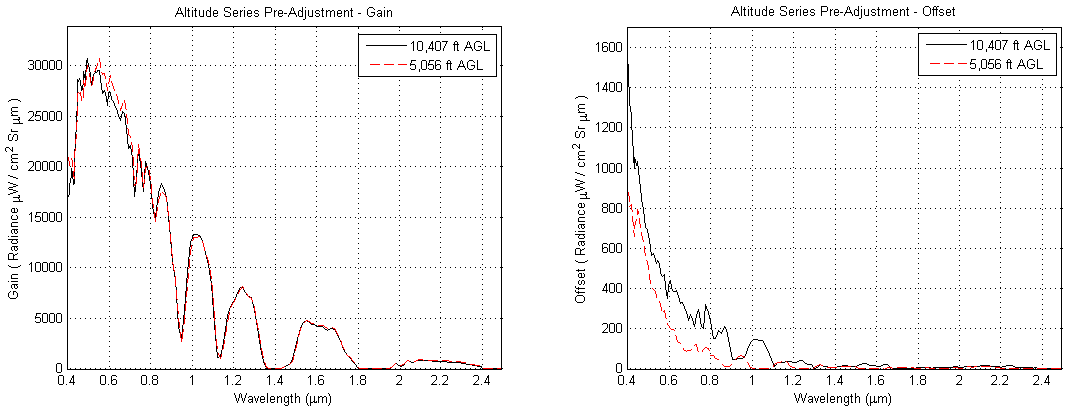

Figure 5 Gain (left) and offset (right) ELM coefficients for 5 and 10 kft images.

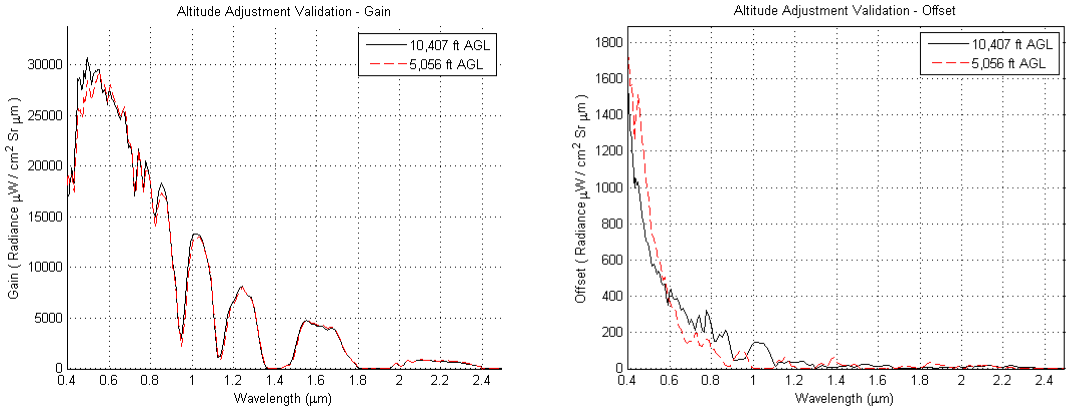

Figure 6 As in Figure 5 but after 5 kft coefficients are adjusted to 10 kft altitude AGL.

4.4 Seasonal adjustment validation

The goal of the study is to identify and separate distinct climatic regimes, so the coefficients must be compared across differing seasons and geographic regions. An example of a seasonal comparison is shown in Figure 7. Coefficients from a summertime collection of seven images were standardized to an image collected at the same site in the wintertime. To the extent that the coefficient adjustment procedure adequately corrected for the differing illumination and scene geometry conditions, the differences in the coefficients are due to differing environmental states. In Figure 7 the wintertime ELM

coefficients are shown (thick black line) against the collection of standardized summertime ELM coefficients (colored lines).

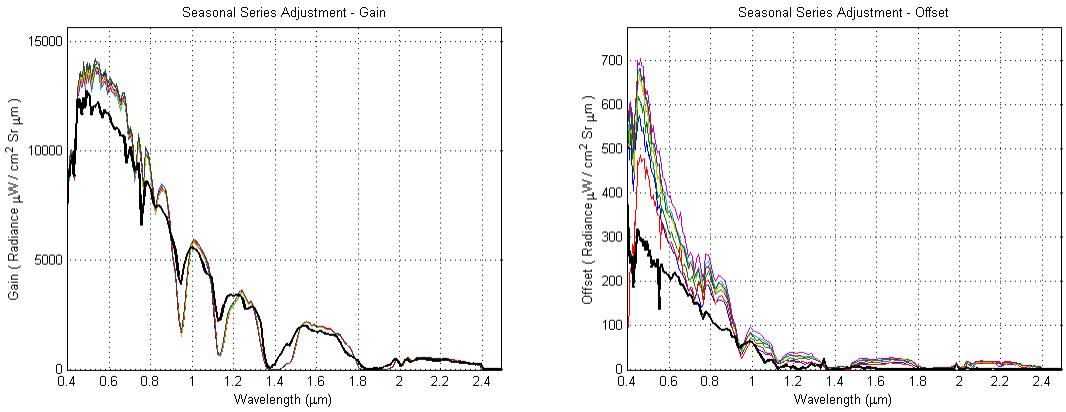

Figure 7 Standardized gain (left) and offset (right) coefficients for the seasonal adjustment validation. The heavy black lines are the Dec reference ELM coefficients; the colored lines are the standardized August coefficients.

The ensemble of standardized summertime coefficients is clearly separate from the wintertime ELM coefficient, suggesting that the environmental variability is larger than the variability remaining after the illumination and geometric adjustments. This is confirmed by the fractional seasonal differences (not shown), which are significantly larger than the variability among the ensemble members. RMSE for the gain is 40% in the inter-seasonal case versus 3% for the summer ensemble. RMSE for the offset is 42% versus 27%. The images in the summertime ensemble were all collected from the same altitude as the wintertime image, so although the time of day varied by up to 3.5 hours, the consistent altitude gives this case less variability than one would expect on average. Still, the study suggests that the seasonal atmospheric signal can be large enough to detect through any residual error from the standardization procedure.

4.5 Global coefficient standardization

All examples shown to this point have coefficients from some number of images standardized to correspond to the imaging conditions of another image to facilitate direct comparison. In order to form classes from the global collection of coefficients, it is necessary to standardize them all to a single reference point. The exact geospatial and temporal coordinates of the reference point can be somewhat arbitrary, but to minimize error, the reference point was chosen to lie near the median of the various imaging observational conditions comprising the study. Thus all coefficients are standardized relative to the coordinates listed in Table 8.

Table 6 List of standard reference coordinates for coefficient comparison

Coordinate	Reference Value
Geographic location	35° N, 95° W
Altitude	10,000 ft AGL
Date	8/15/1997
Time	17:00z

Now standardized to a set of common geometry and illumination conditions, the coefficients can be compared to determine variation in the atmospheric conditions under which they were collected. The variation of the coefficients within each site/collection event varies from site to site depending on the range of imaging conditions encompassed in the collection. Figure 8 shows a typical distribution of the standardized gain coefficients (Site 3, a high mountain location). Figure 9 shows the distribution of the mean standard gain and offset coefficients from all sites and imaging events. Separate means are included in cases where the same site imaged in different seasons or years. The tropical site (Site 9) showed a distinct bimodal distribution, with separation in the visible wavelength region greater than two standard deviations. There are no apparent indicators in the ground truth or image metadata to suspect an observational reason for the bimodal pattern, so it is likely caused by an environmental change during the five-day imaging timeframe. Site 9 is treated in the figure and in the remaining analysis as two separate distributions.

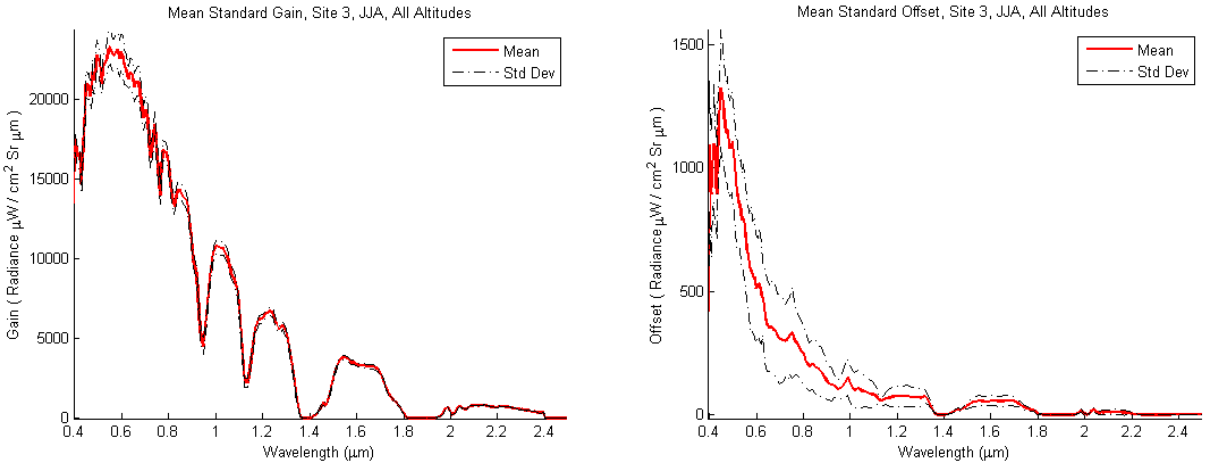

Figure 8 Site 4 mean standard gain (upper) and offset (lower) coefficients (JJA). Standard deviation is plotted in black.

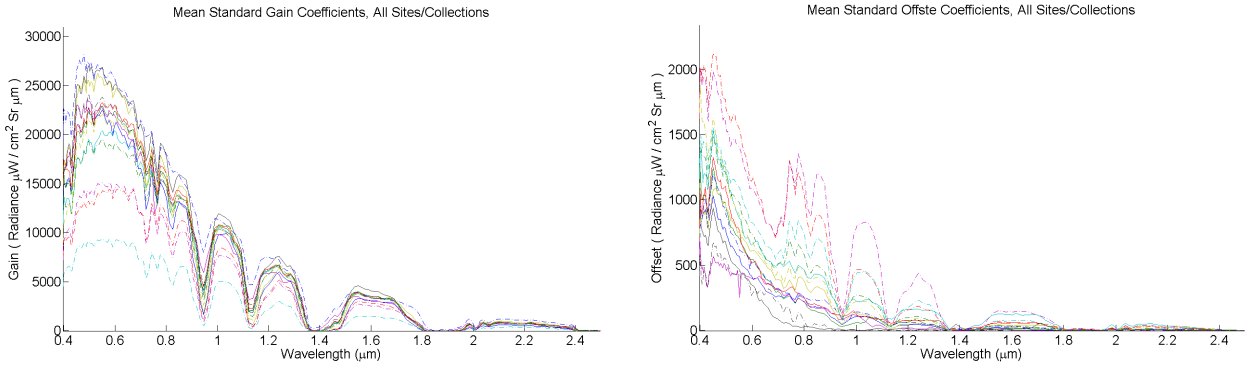

Figure 9 Means of the standard gain (left) and offset (right) coefficients for all sites and imaging events.

4.6 Coefficient classes

The mean standard coefficient spectra in Figure 9 visually appear to have some groupings, and it is desirable to determine classes of coefficients from the data to compare to climatological characteristics of the corresponding sites and images. The groupings of coefficients are explored by clustering the coefficients based on band-to-band spectral magnitude, and then by clustering based on the common spectral similarity metrics.

4.6.1 K-means clustering

A k-means clustering[16] was performed on the set of all coefficients to compare to the site means. The clusters were formed by concatenating each standard gain coefficient with its corresponding offset, the computing the k-means cluster analysis for k=7. The number of clusters (7) was chosen as an approximate number of climate classes represented in the data. Figure 10 shows the means of the gain and offset coefficient clusters superposed against the site means from Figure 9. The plot shows correspondence between some of the cluster means found from the full set of gain coefficients and the means of the site-specific gains.

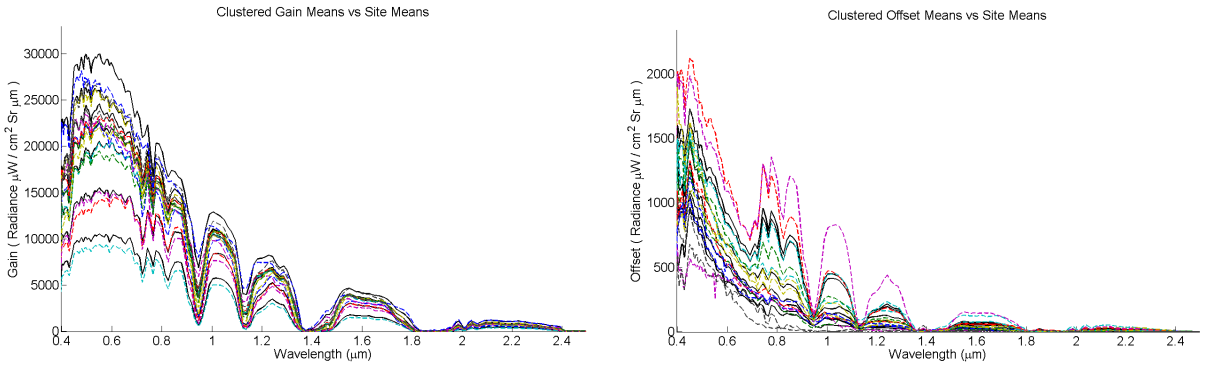

Figure 10 K-means clustered gain means (black) with site means in color (left), clustered offset means (right).

Figure 11 maps the cluster members to their sites' respective Köppen –Trewartha climate categories; e.g., 66% of the coefficients in cluster 1 are from the highland category, 7% from dry/desert, etc. Overall 67% of the coefficients clustered together with coefficients from the same category to form the majority, and 77% clustered with the same or a similar category.

Figure 11 Aggregate Köppen –Trewartha climate category membership of the k-means coefficient clusters. Green indicates the primary category, orange indicates a category with similar characteristics, and other colors indicate dissimilar categories.

4.6.2 Spectral similarity clustering

Because k-means uses a distance metric to define the clusters, the clustering is heavily dominated by the contribution of the gain coefficients, which are typically an order of magnitude larger than the offset coefficients. The effects of aerosol

scattering and background adjacency will have the greatest effect on the empirical offset coefficients, so it is preferable to have a contributing input signal to the clustering algorithm.

To better explore the contributions from and relationships between the gain and offset coefficients and the empirical clustering of the data, the two spectral similarity metrics are introduced into the cluster analysis - Spectral Angle Mapping (SAM)[17] and Euclidean Distance (ED)[18]. These metrics are complimentary because SAM is more sensitive to spectral features such as depth of absorption features and relative shape of the spectra, while ED is heavily weighted by the magnitudes of the spectra. The overall mean of all gain and offset coefficient site means was used as the reference spectra against which to apply the similarity metrics. SAM and ED were applied to each gain and offset coefficient, denoted SAM_m, SAM_b, ED_m, and ED_b. All binary combinations of these metrics are plotted in Figure 12. The colors and symbols represent differing sites and collection seasons. The scatter plots suggest which combinations of similarity metrics may be used to separate environmental classes. The ED_m vs. SAM_b plot is selected here for cluster analysis.

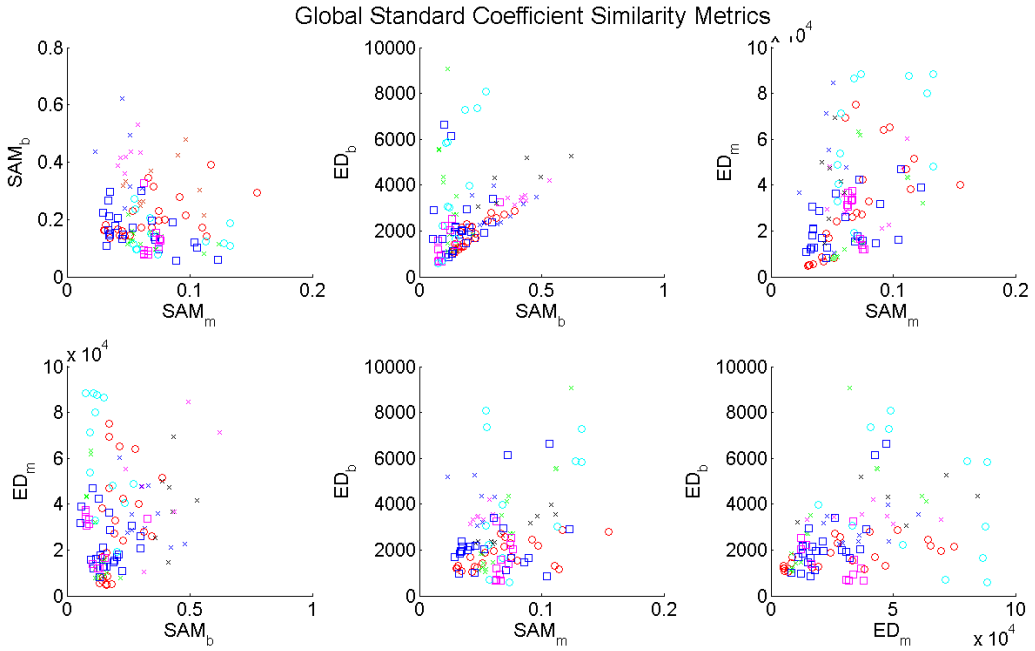

Figure 12 Similarity metrics SAM and ED for all standard gain and offset coefficients plotted against each other to show relationships. Colors/symbols represent differing sites and collections.

Because of the vastly different scales for SAM and ED, the metrics are normalized prior to clustering. An agglomerative hierarchical clustering scheme[16] is used in which a dendrogram is created from closest proximity pairs, which are then consolidated pair-wise in successive iterations until the tree is reduced to only two classes. A clustering criterion can then be applied to any level of the dendrogram desired to produce the desired number of clusters. The distance metric used is Euclidean Distance. Many differing objective functions can be used to define the clusters. The best results were obtained using Ward's minimum variance criterion[19], which minimizes the variance of the points within the cluster at each step.

Figure 13 shows the full set of standard coefficients mapped as SAM_b vs. ED_m. The upper left plot colors/symbols represent the sites and collection groups of the source coefficients. The other plots show representative levels from the hierarchical clustering from 6 to 16 clusters. In these the colors/symbols represent cluster membership. Note that the specific color and symbol combination assigned to a coefficient is arbitrarily assigned and does not map from one plot to another. Rather, the groupings of color/symbol combinations are the points of comparison. As the number of clusters increases, the cluster membership becomes more aligned to the site membership in the upper left panel. The figures show that the similarity metrics are, to a significant degree, able to separate the coefficients in accordance with their source image collections. Single clusters generally represent cohesive groupings coefficients from the same or similar sites (i.e., groups of same

colored symbols in the upper left plot), particularly as the number of clusters increases. Those site member coefficients that are not grouped together in the site grouping scatter plots are of course not separable in this analysis.

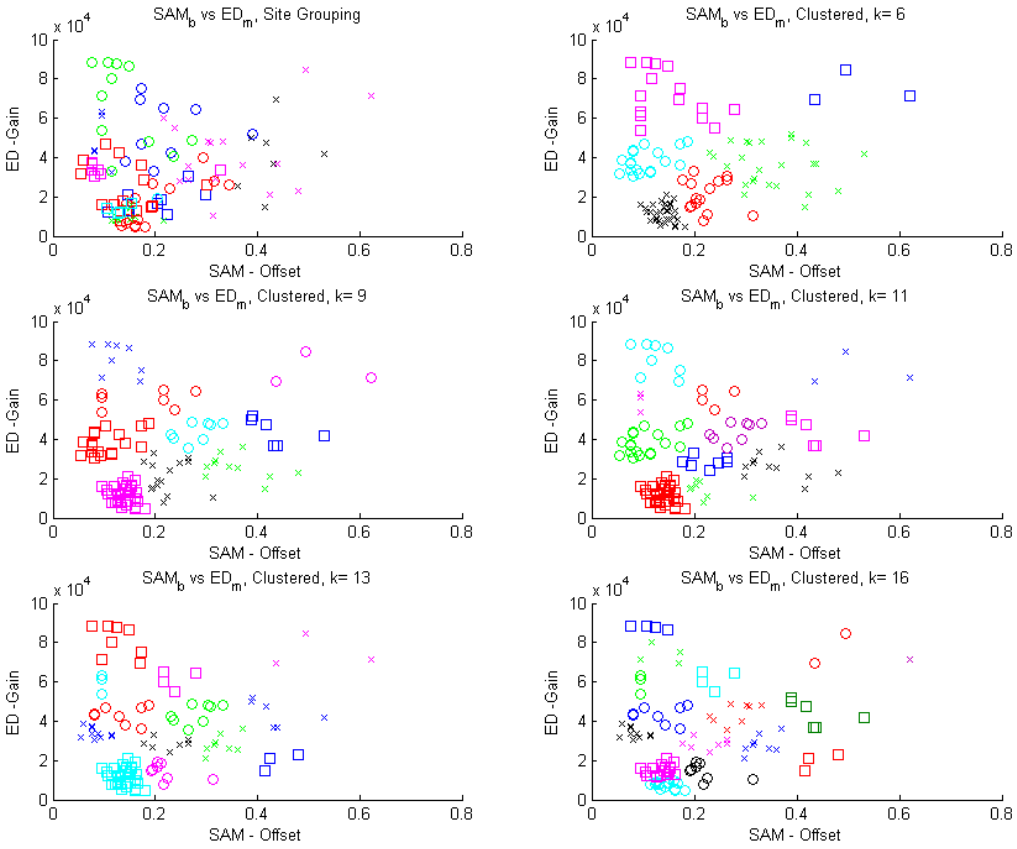

Figure 13 Full set of standard coefficients plotted as SAM_b vs. ED_m. Upper left colors/symbols represent sites and collection groups of source coefficients. Other plots show clustering with increasing numbers of clusters; colors/symbols represent cluster membership.

These analyses were repeated for other pairs of similarity metrics, by mapping clusters to climate category and land cover in addition to site number, and with varying numbers of clusters. Table 7 summarizes some of these results. The overall performance with respect to numbers of coefficients clustered with same-site coefficients is not markedly improved by using the spectral similarity metrics; the percentage is ~40-50% in most cases. In fact, the greatest correlation to site occurs in an 11-cluster k-means trial (60% of the primary members of the clusters were from the same sites). However, the use of the expanded similarity metrics does provide greater separability of environments in some cases as noted in the Table 7, particularly in distinguishing between certain land cover types. The results suggest that the gain coefficients are the primary drivers for classifying the remote sensing environment. The offset coefficients add discrimination power in cases where the background scattering is a large component, such as those sites containing dense vegetation. Despite the pains taken to ensure the offset coefficients were as consistent as possible, the relatively high noise in the offset coefficient signal, likely caused by transient, synoptic scale variations in aerosol content, makes it difficult to characterize in all but the strongest cases.

Table 7 Summary of cluster statistics.

Clustering Scheme	# Clusters	Identity Mapping	Members Mapped to			Clusters w/Pri Majority	Clusters w/Pri/Sim Majority	Notes
			Primary Category	Primary or Similar Category	Dissimilar Category			
k-means	7	site	53%	74%	26%	57%	100%	Highland sites separated; vegetative sites are not
		climate cat.	67%	77%	23%	86%	100%	Vegetative sites not separated
ED_m vs. SAM_m	7	site	37%	60%	40%	43%	71%	Highland sites in single cluster
		climate cat.	57%	69%	31%	57%	100%	Desert sites mixed among 3 clusters
ED_m vs. SAM_b	7	site	40%	62%	38%	43%	86%	Vegetative sites separated
		climate cat.	56%	67%	33%	43%	100%	Vegetative sites separated
ED_m vs. SAM_b	11	site	47%	68%	32%	45%	82%	High and low deserts separated; summer and winter deserts separated

4.6.3 Cross validation

In the preceding sections, the dataset used to form the clusters was also used to "score" the results in terms of homogeneity of clusters and ability to identify component climate or site membership. Without an independent test dataset, the results are likely positive biased, because the training data are not independent of the validation dataset. To reduce this bias, an "n-1" cross validation[16] was performed on the data. In this technique, one coefficient is held out as the validation sample, and the other n-1 coefficients are used to form the clusters or means, where n is the total number of samples. Then the next coefficient is held out and the other n-1 coefficients (including the first validation sample) are used to form the clusters or means. The process is repeated n times such that each sample is used once as the validation dataset and each validation trial is independent of the data used to form the means. The results are averaged across all n trials. The cross validation was performed on the clusters resulting from k-means clustering for k=3 to k=20 clusters. Site membership is used as the test criterion, expanded such that multiple-season sites are treated as multiple categories (i.e., one for each season).

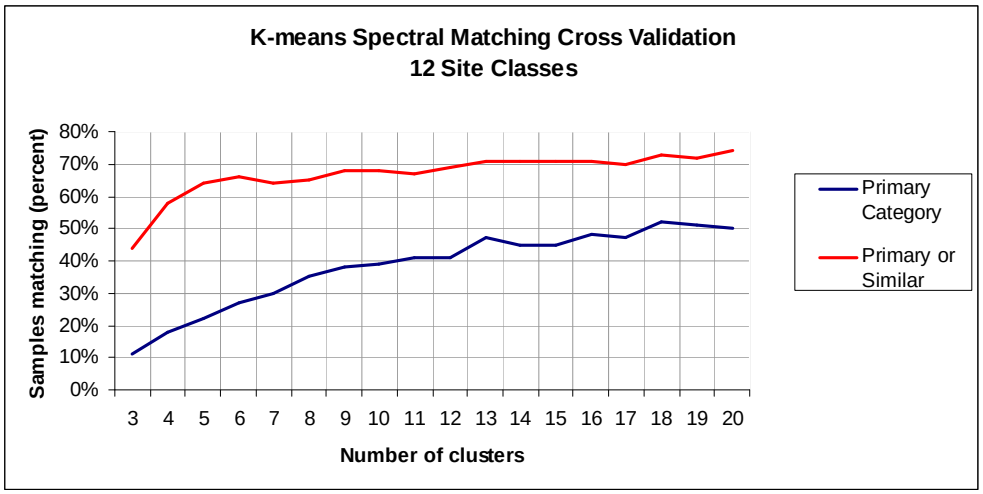

Figure 14 Result of the cross validation of the k-means clusters, from n=3 to n=20 clusters, showing the percentage of samples correctly matched to site classes using the SAM similarity metric.

Figure 14 shows the result of the cross validation. The figure confirms that seven is too few clusters, at least in the expanded twelve-class case. The maximum score of with respect to site identification occurs between 13 and 18 clusters, suggesting that the number of separable classes is in that range. The values for primary identification and primary or similar identification for k=11 compare well with the corresponding values in Table 7.

5. CONCLUSION

The research shows that environmental information embedded within the ELM coefficients can be standardized, cataloged, and used to form a climatological compendium in spectral space. Common spectral similarity metrics are used to show that the climatological classes are separable to a degree of detail commensurate with the relatively modest size and range of the imaging conditions comprising the study.

Largely unstudied aspects of ELM process and coefficients are revealed in the work. The small but highly leveraged trades between gain and offset coefficients that are not significant to ordinary ELM retrievals become problematic in this context because the imaging model equates the offset coefficient with path radiance, which should physically never be negative. In addition to random scatter, a small nonlinear effect of the spherical albedo was discovered in the data and a model developed to correct the coefficients to account for the effect.

Some of the imaged sites and climatological conditions were very conducive to the model and aligned nearly exactly with the standardization process. Others showed considerable variability that was not explained by the model. This is not surprising; although the body of conditions represented in the study data is relative large for an HSI collection, it is tiny compared to the universe of environmental conditions that can be present on the Earth. Much more data will be required to more fully characterize climatological means and variability. Even so, the data in the study is over broad enough conditions to demonstrate separable climatological classes.

REFERENCES

[1] Stewart, A., Bauer R. and Kaiser, R., “Performance Assessment of Atmospheric Correction Algorithms on Material Identification for VIS-SWIR Hyperspectral Data II,” Proc. SPIE, Imaging Spectrometry VI 4132, 206-217 (2000) [doi: 10.1117/12.406589].

[2] Powell, J. H. and Resmini, R. G., “A spectral climatology for atmospheric compensation,” Proc. SPIE 9088, Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX (2014) [doi: 10.1117/12.2050596].

[3] Conel, J. E., “Determination of surface reflectance and estimates of atmospheric optical depth and single scattering albedo from Landsat Thematic Mapper data,” Int. J. Remote Sensing 11(5), 783-828 (1990) [doi: 10.1080/01431169008955057].

[4] Berk, A., Anderson, G. P., Acharya, P. K., Bernstein, L. S., Muratov, L., Lee, J., Fox, M., Adler-Golden, S. M., Chetwynd, J. H., Hoke, M. L., Lockwood, R. B, Gardner, J. A., Cooley, T. W., Borel, C. C., Lewis, P. E., and Shettle, E. P., "MODTRAN5: 2006 Update," Proc. SPIE, Vol. 6233, 62331F, 2006.

[5] Berk, A., and Anderson, G.P. “Impact of MODTRAN 5.1 on Atmospheric Compensation,” IGARSS 2008, 127-129 (2008) [doi: 10.1109/IGARSS .2008.4779299].

[6] Perkins, T., Adler-Golden, S., Matthew, M., Berk, A., Anderson, G., Gardner, J. and Felde, G., “Retrieval of atmospheric properties from hyper- and multi-spectral imagery with the FLAASH atmospheric correction algorithm,” Proc. SPIE, Remote Sensing of Clouds and the Atmosphere X 5979, 59790E (2005) [doi: 10.1117/12.626526].

[7] Bernstein, L. S., Adler-Golden, S. M., Sundberg, R. L. and Ratkowski, A. J., “In-scene-based atmospheric correction of uncalibrated VISible-SWIR (VIS-SWIR) hyper- and multispectral imagery,” Proc. SPIE, Remote Sensing of Clouds and the Atmosphere XIII 7107, 710706 (2008) [doi: 10.1117/12.808193].

[8] Kappus, M.E., Aldrich, W.S., Resmini, R.G. and Mitchell, P., “The flexible HYDICE sensor's first year of operation,” Proc. of the 11th Thematic Conference on Geologic Remote Sensing 1, 433-441 (1996).

[9] Nischan, M. L., Kerekes, J. P., and Baum, J. E., “Analysis of HYDICE noise characteristics and their impact on subpixel object detection,” Proc. SPIE 3753, Imaging Spectrometry V (1999) [doi: 10.1117/12.366274].
[10] Slater, P. N., Basedow, R. W., Aldrich, W. S., and Coiwell, J. E., "In-flight radiometric stability of HYDICE for large and small uniform reflectance targets under various conditions", SPIE Proc., 2821, 300-310 (1996) [doi:10.1117/12.257178].
[11] Homer, C. C., Huang, L., Yang, B., Wylie, B., and Coan, M., “Development of a 2001 National Landcover Database for the United States,” Photogrammetric Engineering and Remote Sensing, 70 (7), 829-840 (2004).
[12] Trewartha, G. T. and Horn, L. H., An introduction to climate, McGraw-Hill, New York (1980).
[13] Baker, B., Diaz, H., Hargrove, W., and Hoffman, F., “Use of the Köppen–Trewartha climate classification to evaluate climatic refugia in statistically derived ecoregions for the People’s Republic of China,” Climatic Change 98, 113–131 (2010) [doi: 10.1007/s10584-009-9622-2].
[14] Berk, A, Anderson, G. P., and Acharya, P. K., "MODTRAN 5.3.2 User’s Manual," Air Force Research Laboratory (2013).
[15] Adler-Golden, S. M., Matthew, M. W., Berk, A., Fox, M. J., Lee, J., and Ratkowski, A. J., “Improvements in aerosol retrieval for atmospheric correction,” Geoscience and Remote Sensing Symposium, IGARSS (2008).
[16] Wilks, D. S., [Statistical Methods in the Atmospheric Sciences], Academic Press, San Diego (1995).
[17] van der Meer, F. D. and de Jong, S. M., [Imaging Spectrometry Basic Principles and Prospective Applications], 31-41, Kluwer Academic Publishers, Dordrecht, Netherlands (2003).
[18] Chang, C.-I., [Hyperspectral Imaging, Techniques for Spectral Detection and Classification], 20-21, Kluwer Academic Publishers, New York (2003).
[19] Ward, J. H., "Hierarchical Grouping to Optimize an Objective Function", J of the American Stat Assoc, 58, 236–244 (1963).