A spectral climatology for atmospheric compensation

John H. Powell* and Ronald G. Resmini

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. Most atmospheric compensation methods 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. This paper introduces a new climatological approach to atmospheric compensation in which empirically derived spectral information, rather than sensible atmospheric state variables, is the fundamental datum. An experimental archive of airborne HSI data is mined for representative atmospheric compensation coefficients, which are assembled in a scientific database of spectral and sensible atmospheric observations. We present the empirical techniques for extracting the coefficients, the modeling methods used to normalize the coefficients across varying collection and illumination geometries, and the resulting comparisons of adjusted coefficients. Preliminary results comparing normalized coefficients from representative scenes across several distinct environments are presented, along with a discussion of the potential benefits, shortfalls, and future work to fully develop the new technique.

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 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. 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 this work we introduce 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.

2. BACKGROUND

2.1 Empirical Line Method

The empirical line method 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 \( L_S(\lambda) \) can be written as:

\[
L_S(\lambda) = L_{dir}(\lambda) + L_{sky}(\lambda) + L_{path}(\lambda)
\]  

(1)

where \( L_{dir}(\lambda) \) is the direct path reflected radiance, \( L_{sky}(\lambda) \) is the indirect sky-illumination reflected-radiance and \( L_{path}(\lambda) \) is the path radiance. Adjacency and multiple surface scatter effects are neglected, as are thermal emissive radiance contributions. The direct path term is given by:

\[
L_{dir}(\lambda) = E_0 \tau_d(\lambda) \tau_u(\lambda) \frac{\rho(\lambda)}{\pi} \cos \sigma
\]

(2)

where \( E_0 \) is the solar irradiance at the top of the atmosphere, \( \tau_d(\lambda) \) is the downward path transmittance, \( \tau_u(\lambda) \) is the upward path transmittance, \( \rho(\lambda) \) is the surface reflectance factor, and \( \sigma \) is the incident angle to the surface. Implicit in the reflectance factor term is the assumption of a Lambertian surface. In the more general case, \( \rho(\lambda) \) would be replaced by the bi-directional reflectance distribution function (BRDF). The indirect reflected term is written as:

\[
L_{sky}(\lambda) = E_s \tau_u(\lambda) \frac{\rho(\lambda)}{\pi}
\]

(3)

where \( E_s \) is the skylight irradiance at the surface. Here it is assumed that the entire hemisphere of the sky is visible to the surface and again, the surface is Lambertian.

Equation 1 can then be written as a linear relationship (dropping the wavelength dependence notation for clarity):

\[
L_S = m\rho + b
\]  

(4)
where $m$ and $b$ are the gain and offset vectors given by:

$$m = \left( E_0 \tau_d \cos \sigma + E_\sigma \right) \frac{\tau_u}{\pi}$$  \hspace{1cm} (5)$$

$$b = L_{path}. \hspace{1cm} (6)$$

In ELM, the gain and offset coefficients $m$ and $b$ in equation 4 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. 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).

MODTRAN accepts as inputs a range of remote sensing geometries and most possible illumination conditions. It has a rich variety of options for specifying atmospheric constituents, state variables, and scattering models. 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. 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 is commonly used in HSI analysis because it shows results comparable to PB models under many conditions but without the requirement for exact atmospheric and illumination information. It involves less calculation than PB methods and is not as dependent on the accuracy of the data’s radiometric calibration. QUAC is used here as the method to which the ELM derived coefficients and reflectance retrievals are compared.

2.4 Hyperspectral data

The data used in this study were collected by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor. HYDICE is a pushbroom hyperspectral sensor with a spectral range from 0.4 to 2.5 microns. It uses a Schmidt prism dispersive spectrometer with a single indium antimonide (InSb) focal plane. HYDICE collects 210 spectral bands with a nominal bandwidth of 10 nm and 320 spatial samples.

HSI scenes were collected over a variety of environments representing distinct climate regimes. In each environment, data were collected at several altitudes ranging from 5-20 kft above ground level (AGL) 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.

3. METHODOLOGY

3.1 ELM analysis

The spectral climatology requires harvesting ELM-derived atmospheric compensation coefficients for a wide variety of environments. An extensive scientific archive of HYDICE imagery was reviewed to identify the scenes containing calibration panels for analysis. The QUAC algorithm was run on the full scene for maximum spectral diversity and the
resultant gain and offset coefficients saved for later analysis. Ground truth spectra were reconstructed from available records, including experiment reports, field logs, and metadata. Scenes with missing ground truth or ambiguous metadata were rejected, unless adequate substitute data were available, e.g., the same panel measured on the prior day.

Standard ELM techniques were used to derive gain and offset coefficients that describe the inversion from at aperture radiance to reflectance\(^2\). The in-scene pixels containing each calibration panel were manually identified and each pixel spectrum reviewed to ensure no noticeable background or edge contamination was present. The radiance spectra of the pixels for each panel were averaged. Low altitude (5 kft AGL) collects typically contained 25-30 pixels on each panel, ranging to only 1-2 pixels at the highest altitude (20 kft AGL). An ELM regression was then performed on the averaged panel radiance spectra against the corresponding ground truth reflectance. ELM produces vectors \(m\) and \(b\) containing the gain and offset coefficients for each channel and an RMSE vector \(r\).

In some cases, the regression performed using all calibration panels resulted in coefficients with large RMSE relative to the offset values. These results also produced unusual offset vectors, having strong negative values in the near- and short wave IR regions. In such cases, the ELM regression was repeated with one or more of the grayscale panels omitted to obtain coefficients with minimum RMSE. In all cases this resulted in offset coefficients that did not show strongly negative values. Figure 1 shows such a case. The regression using all four calibration panels resulted in large negative offset values and a large RMSE. Using only two panels (brightest and darkest) produced a more physical offset coefficient profile, albeit with smaller negative regions.

![Figure 1](image.png)

**Figure 1.** ELM gain (solid line) and offset (dashed) coefficients with RMSE (dot-dashed), using all calibration panels (left) and omitting the two mid-brightness (approximately 16% and 50%) reflectance panels (right). No RMSE is computed with only two regression points.

Figure 2 shows the resultant reflectance retrievals obtained using the two ELM cases plotted against the ground truth spectra. Both regressions produced very similar results. It is interesting, but perhaps not wholly unexpected, that there are numerous different sets of ELM gains and offsets that generate reasonable reflectance spectra. From a mathematical standpoint, any trade between gain and offset values in the regression that increases the goodness of fit is acceptable. To fit the physical model described in section 2.1, however, it is desirable to have offset coefficients that make physical sense for the path radiance term \(L_{\text{path}}(\lambda)\), i.e., not strongly negative. In theory, for i.i.d. observations, the regression should generally improve with more observation points, but that was not always the case here, whether because of non-representative ground truth measurements, panel contamination, sensor nonlinearity, or other factors. In any case the coefficients used for the climatology were those having minimum RMSE, so optimal goodness of fit is assured as long as the majority of the dynamic range of the scene is represented in the regression, i.e., very dark and bright panels are

©2014-The MITRE Corporation. All rights reserved.
included. Good ELM results have been shown with only two observation points, but in almost all cases in this work, three or more points were used so that quantifiable error terms (RMSE) were produced.

Figure 2. Reflectance retrievals from the 4-panel ELM regression (dashed line) and the 2-panel ELM regression (dot-dashed) plotted against the ground truth panel reflectance spectra (solid).

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 and FWHM of all collects.

3.2 Coefficient adjustment

The ELM-derived coefficients correct for illumination and geometric factors as well as atmospheric effects. To compare coefficients from different collects it is therefore necessary to adjust 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 models the atmosphere by treating it as a series of homogeneous layers characterized by their temperature, pressure, and molecular composition. MODTRAN models the absorption, scattering, and emission for each of the molecular constituents along a specified optical path, from the ultraviolet to far infrared range of the spectrum at up to 0.1 cm\(^{-1}\) resolution. MODTRAN can also provide solar illumination based on geographic position, date, and time. Thus the terms in equation 1 can be modeled explicitly.

To model the HYDICE collects, a MODTRAN spectral filter function (.FLT file) was created using the HYDICE mean spectral response function to provide output corresponding to the radiance measured by the instrument in MODTRAN’s channel output file (.CHN). Two MODTRAN runs are required to obtain the terms of equation 1. First, it was run with the surface reflectance set to albedo=1. The terms of interest are calculated from the MODTRAN outputs:

\[
L_{\text{dir}} = \text{DIRECT\_TRANS\_GRND\_REFLECT} \tag{7}
\]

\[
L_{\text{sky}} = \text{TOTAL\_TRANS\_GRND\_REFLECT} – \text{DIRECT\_TRANS\_GRND\_REFLECT}. \tag{8}
\]

The simulation was then repeated with surface reflectance set to albedo=0. This gives:
\[ L_{\text{path}} = \text{PATH TOTAL_SCAT_SOLAR}. \] (9)

To calculate the spectral radiance of the above terms, each must be divided by an integrating factor given by:

\[ IF = \frac{\text{CHANNEL_RADIANCE}}{\text{SPECTRAL_RADIANCE}} \] (10)

also from the .CHN file. This converts the radiance units to \( \text{W/cm}^2\cdot\text{sr}\cdot\mu\text{m} \). The terms are further multiplied by \( 10^{-6} \) to convert to \( \mu\text{W/cm}^2\cdot\text{sr}\cdot\mu\text{m} \) and by 75 to match the scale factor applied to the HYDICE spectral data (for numerical storage efficiency).

We can now calculate modeled gain and offset coefficients equivalent to those derived from ELM:

\[ m^* = (L_{\text{dir}} + L_{\text{sky}}) \] (11)

\[ b^* = L_{\text{path}} \] (12)

where the stars delineate the terms as MODTRAN modeled coefficients.

The modeled coefficients can then be used to estimate corrections to the ELM coefficients for differing illumination and sensor altitude. 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 the above procedure. 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 computed from the modeled coefficients:

\[ A_{21} = \frac{m_2^*}{m_1^*} \] (13)

\[ B_{21} = \frac{b_2^*}{b_1^*}. \] (14)

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

\[ m_2' = A_{21} \times m_1 \] (15)

\[ b_2' = B_{21} \times b_1 \] (16)

where the primes delineate estimated (adjusted) ELM coefficients.

Using this procedure it is possible to model the effects of differing illumination and geometry on ELM coefficients and therefore to compare coefficients derived under different imaging conditions. The accuracy of the results will decrease as the magnitude of the change increases, but within some bounds of variability, the method produces accurate estimates. The assumption implicit in this technique is that the fractional difference in the modeled coefficients accounts for the portion of the variability in the ELM coefficients that is due to the differing imaging conditions. The variability that is left is assumed to be due to differences in the atmosphere. By applying the ratios of modeled coefficients, it has the advantage of canceling out to first order any systematic errors in the modeled results while leveraging the model’s strength of computing varied illumination and geometric conditions.

4. RESULTS

4.1 Illumination adjustment

To produce a database of comparable atmospheric gain and offset coefficients, the methodology described above must be able to correct for variations in illumination from the differing times of day, seasons, and latitudes encompassed in the observation set. To demonstrate the effectiveness of the adjustment with respect to illumination changes, a sequential series of images was analyzed. The same panels were imaged six times over a 75-minute period from the same altitude (5 kft AGL). Figure 3 shows the ELM coefficients from the earliest, middle, and latest observation times, with the gain coefficients increasing ~25% over the period in the visible region due to the increased illumination approaching local noon (12:46 PM).
Figure 3. ELM gain (left) and offset (right) coefficients derived from three sequential observations approximately 2.5 to one hour before local noon.

The coefficient adjustment procedure was performed on the 10:14 AM coefficients to correct them for the illumination modeled at 11:31 AM. Figure 4 shows the adjusted coefficients plotted against the actual 11:31 AM image ELM coefficients (the $m^*$ plot is difficult to distinguish from the ELM gain in the visible region). Figure 5 shows the fractional error in the adjusted coefficients; with the exception of the low SNR absorption regions, the gain coefficients agree to within 2% and the offset to within 35% over most of the wavelengths.

It is noted from figure 4 that the directly modeled coefficients $m^*$ and $b^*$ contain large magnitude errors. The MODTRAN runs display a high sensitivity in the visible region to the input conditions and scattering models used. Figure 6 shows the variation of the modeled parameters for a number of MODTRAN input parameters – single scatter model, DISORT multiple scattering model with the default rural aerosol model, and DISORT with the desert aerosol scattering model. Despite the large variation in modeled coefficients, the adjustment procedure has the effect of grounding the model with initial empirically derived coefficients, and errors of 5% or less were seen in the adjusted gain coefficients in each of the cases shown in figure 6. The DISORT/desert aerosol model (used in the figure 4 case) had the lowest error; the environment was a summertime desert scene and the shape of the coefficient desert profile in figure 6 suggests that the scattering is well modeled.

Figure 4. Estimated gain (left) and offset (right) coefficients (dashed line) for a 77 minute change in illumination conditions, plotted against the actual derived coefficients (solid). The dot-dashed line shows the directly modeled coefficients ($m^*$ and $b^*$).

©2014-The MITRE Corporation. All rights reserved.
4.2 Altitude adjustment

The database also includes coefficients derived from various flight levels, so the adjustment procedure should correct for variations in observation level. The same procedure was applied as described in section 3.2, 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 7 shows the adjusted coefficients plotted against the actual 1:14 PM image ELM coefficients; Figure 8 shows the fractional errors in the coefficients. The errors in the estimated gain coefficients are larger than in the illumination only case in section 4.1, but estimates are still generally within 3-5% of the actual ELM values. The estimated offsets are similarly high, ranging from 5-10% in the visible range. The offset error is largest in the visible, suggesting that the aerosol scattering model may not be optimal in this case. The scene was a summertime mid-latitude, continental plains environment.
Figure 7. Estimated gain (left) and offset (right) coefficients (dashed line) for a 5.4 kft altitude change and 26 minute change in illumination conditions, plotted against the actual derived coefficients (solid). The dot-dashed line shows the directly modeled coefficients ($m^*$ and $b^*$).

Figure 8. Fractional errors in the adjusted gain (top) and offset (bottom) coefficients.

4.3 Comparison of environments
Since the goal of the database is to identify and separate distinct climatic regimes, coefficients must be compared across differing seasons and geographic regions. An example of a seasonal comparison is shown in figures 9 and 10. Coefficients from a representative summertime desert image were adjusted to a similar wintertime desert image. The two images were not over the same geographic area, but were in similar climatological environments. 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 atmospheric states. The fractional differences in the gain coefficients shown in figure 10 are significantly larger than the errors in the adjustment procedure shown in figure 8, by a factor of 5-
10 in most regions. Figure 8 shows a representative error for the adjustment procedure, not the error from the specific case in figure 10. However, the relatively large difference shown in figure 10 suggests that the seasonal atmospheric signal is substantially larger than the error incurred in the adjustment procedure. The offset coefficients are also larger over many regions than the comparable errors, but to a lesser degree than the gain coefficients.

Figure 9. Normalized gain (left) and offset (right) coefficients for desert environment in winter (solid line) and summer (dashed).

Figure 10. Fractional differences between summer and winter cases for the normalized gain (top) and offset (bottom) coefficients.

To compare many coefficients from multiple environments, we adjust each set of coefficients to a common set of illumination and geometric conditions. In this example the set of conditions is arbitrary but was selected to be near the center of the conditions of the observation set. With the coefficients adjusted to a common set of imaging conditions, direct intercomparison of the coefficients is possible. An example comparison is shown in figure 11. Representative adjusted coefficients from four different climatic environments are plotted against the mean of the four cases. The figure shows distinct differences in several of the environments, suggesting that separable climatological regimes are present in the ELM coefficient data.
5. CONCLUSION

This paper presents a method for extracting atmospheric variations from ELM coefficients using radiative transfer modeling to correct for the effects of varying illumination and imaging geometry. The method shows good results with small errors in test cases against actual ELM coefficients over a limited range of conditions. Comparisons of normalized coefficients across climatological environments suggest that the method can be used to separate climatic regimes. In order to validate these suggestions, the example cases presented must be expanded to a larger set of conditions to characterize the limits of the adjustment procedure. With a large sample of coefficients spanning numerous observations of the climatic regimes, the example shown in figure 11 can be expanded to a statistical analysis that can confirm the viability of the method.

Additional work is needed to fully understand the impacts of various MODTRAN modeling inputs. The method is resilient to absolute errors in the radiative transfer results, but the scattering models in particular can affect the coefficient normalization in the visible region. Sensible atmospheric observations are also available for the experimental collects, which can be used to aid the validation of atmospheric variability extracted from the spectral coefficients.

Once validated, the spectral database can be expanded to incorporate a growing number of observations, leading to a true climatological database of spectral coefficients. The method is potentially adaptable to other empirical atmospheric compensation algorithms such as QUAC or automated ELM methods. This would alleviate the requirement for high fidelity ground truth and greatly expand the number of observations available for the climatology.

REFERENCES


