Adaptive HSI Data Processing for Near-Real-time Analysis and Spectral Recovery*

Su May Hsu,[1] Hsiao-hua Burke and Michael Griffin
MIT Lincoln Laboratory, Lexington, Massachusetts

1. INTRODUCTION

Hyperspectral imaging (HSI) sensors collect spatially resolved data in hundreds of spectral channels. While the technology matures and finds broad applications, data downlink from the collection platform and near real-time processing remain key challenges, especially for near-term spaceborne sensors. It is desirable to process the data on-board for near-real-time analysis and downlink compressed data allowing near full spectral recovery for post-mission analysis.

Principal component analysis (PCA) can be used to determine the reduced dimensionality and separate noise components in the data. While PCA is useful for image feature analysis such as smoke/cloud discrimination (Griffin et al., 2000), it can also be used as a data compression tool. With PCA, the majority of information in an HSI data cube is effectively compressed to a small number of principal components. The data volume is significantly reduced while the feature contrast is enhanced. Spectral information can be recovered from the compressed data with minimal loss. In this paper, the reconstructed data are compared to the original "truth" data with difference analysis using sample AVIRIS imagery.

This methodology also allows for the HSI data to be used adaptively for various multispectral band simulations without the constraint of data volume and processing burden. Based on AVIRIS data, emulation of MODIS sensor bands are carried out and compared with the PCA-reconstructed data. Two products are also derived and compared: Normalized Difference Vegetation Index (NDVI) and the integrated column water vapor (CWV) using the full set of AVIRIS data and the reconstructed spectral information.

2. PRINCIPAL COMPONENT ANALYSIS OF HSI DATA

Principal component analysis (PCA) is generally used to de-correlate data and maximize the information content in a reduced number of features [Richards, 1994; Geladi,1997]. It can be applied to HSI data to reduce the dimensionality, aid in anomaly detection and separate scene information from noise components. The covariance matrix is first estimated over the pixel spectra contained in the HSI data cube of interest. Eigenvalues and Eigenvectors are then obtained for the covariance matrix Σ as given below,

$$\Sigma = E\,\{(X - X_{\mathbf{m}})(X - X_{\mathbf{m}})^{\mathbf{T}}\} = \Phi\,\Lambda\,\Phi^{\mathbf{T}}\,,$$

where X represents the spectral vector data; $X_{\mathbf{m}}$, the mean spectral vector over the data cube and **E**, the average operator over the entire data cube. Φ is a matrix consisting of columns of Eigenvectors and Λ is a diagonal matrix of Eigenvalues.

Using the Eigenvector matrix Φ, the HSI data cube is then transformed into principal components, also called Eigenimages. The components at an image location with a spectral vector X is obtained as below,

$$\mathbf{PC\,(X)} = \Phi^{\mathbf{T}} X\,.$$

---

* This work was sponsored by the Department of Defense under Contract F19628-00-C-0002. Opinions, interpretations, conclusions and recommendations are those of the authors and not necessarily endorsed by the United States Air Force.
[1] Telephone (781) 981-2920; Fax (781) 981-7271; e-mail sumayhsu@ll.mit.edu, 244 Wood Street, Lexington, Massachusetts. 02420-9185

The components are ranked in descending order of the Eigenvalues (image variances). The Eigenimages associated with large Eigenvalues contain most of the image variations while the Eigenimages associated with small Eigenvalues are usually noise-dominated. The majority of variations in the scene are generally contained in the first few principal components although some anomalous image features may still be apparent in the 10th or later Eigenimage. Thus principal component transform allows the determination of the inherent dimensionality and segregation of noise components in the HSI data. With the information compression property in the leading principal components, the full spectral image data can be recovered from truncated sets of the principal components. The information content that was lost is indicated by the sum of the Eigenvalues below the level of component truncation.

The PCA processing is demonstrated on an AVIRIS scene taken over Brazil during the Smoke, Clouds, Aerosols and Radiation - Brazil (SCAR-B) experiment in August of 1995. Figure 1 shows RGB and pseudo color images of the cumulus cloud scene from an AVIRIS image on 20 August 1995. Eigenvalues and sample Eigenvectors obtained from the data cube are plotted in Figures 2a and b, respectively. In Figure 2a, there are approximately four orders of magnitude separation between the 1st and 10th Eigenvalues, which indicate the relative information content of the related Eigenimages. The Eigenimages can be considered as a weighted sum of the spectral images. The component weights over the spectral bands are represented by the Eigenvectors. Sample Eigenimages are shown in Figure 3. Scene features are well contrasted in the images. Clouds, water and shadows are the most contrasted features shown in the first Eigenimage. In the second Eigenimage, vegetation features are enhanced. Detailed cloud structures are seen in the 3rd Eigenimage. Burned areas are evident in the 4th Eigenimage.

The most computation-intensive operation in PCA is projecting the spectral data into Eigenimages. It requires 2*n* FLOPs (floating point operations) per pixel per Eigenimage where *n* is the number of spectral bands. To obtain 10 Eigenimages from an AVIRIS-like image of 500x500-pixel, 200 spectral-band HSI, it takes approximately 10^9 FLOPs. This is the same order of computation required for pixel-level atmospheric compensation. The basic computation performed in ATREM (Gao et al, 1996) requires about 15 FLOPs per pixel per band. This amounts to 7.5×10^8 FLOPs to compensate the same spectral image.

In a multispectral image, the number of data points can be expressed as,

$$S\ L\ n\ ,$$

where *S*, *L* and *n* are numbers of samples, lines and spectral bands, respectively in the image. The number of data points in the compressed image and the corresponding Eigenvectors are shown below,

$$(S\ L + n)\ p\ ,$$

where *p* is the number of principal components used. The Eigenvectors are necessary for spectral data recovery. Notice that number of spectral bands, *n* is typically much less than the number of pixels in the image, *S L*. The compression ratio is,

$$S\ L\ n\ /\ (S\ L + n)\ p \sim n/p\ .$$

Considering there are approximately 200 spectral bands in the AVIRIS data, the compression ratio using 10 PCs is about 20 for a typical image frame of 500x500 pixels.

3. SPECTRAL COMPARISON

The spectral image data are recovered with inverse PCA using 5, 10, and 15 principal components. The recovered data are compared to the original data for performance evaluation. Different spectra from various features in the image are plotted in Figure 4. Good agreements are seen for all features between the reconstructed and original data, water being the exception. The reflected signal from a water surface is very low in the NIR/SWIR. Therefore, even small errors in the reconstruction of the spectral information are exaggerated for low reflecting surfaces. The reconstruction error is largest for 5 PCs and reduces considerably for 10 PCs and higher.

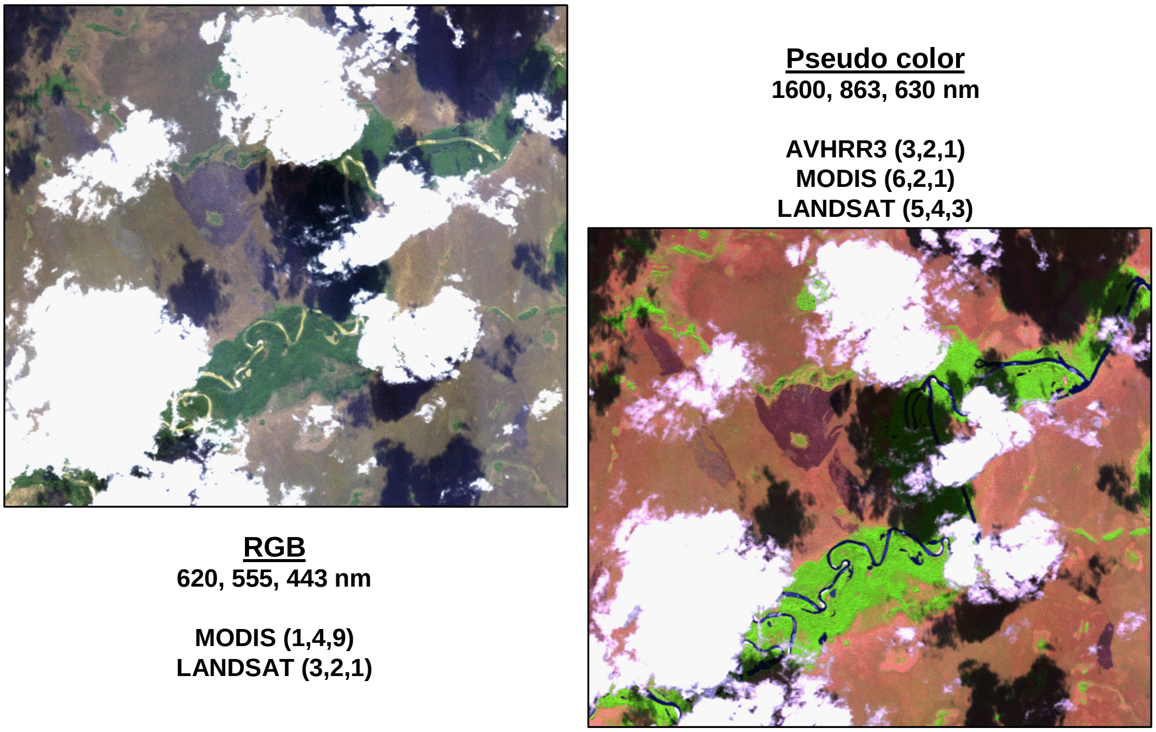

Figure 1. Color images of the AVIRIS SCAR-B data.

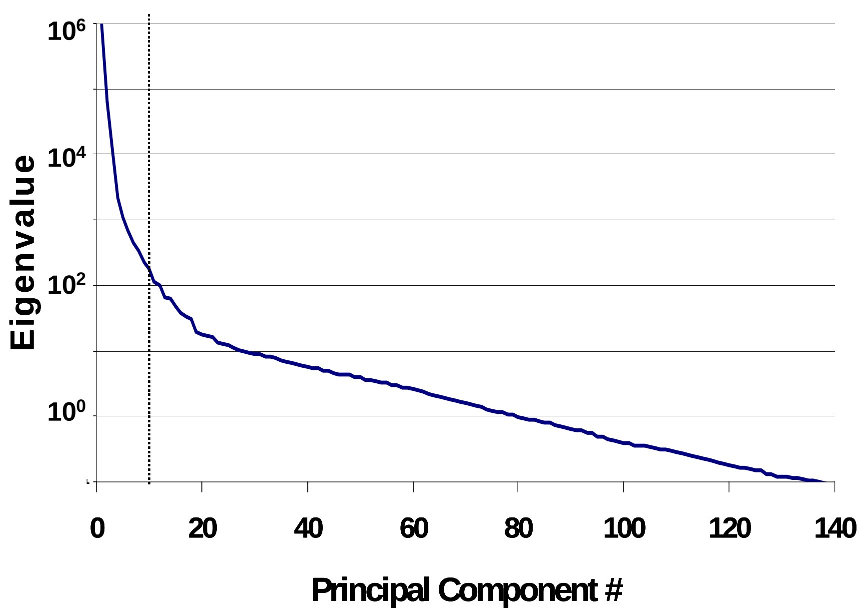

Figure 2(a). Eigenvalues of the SCAR-B HSI data.

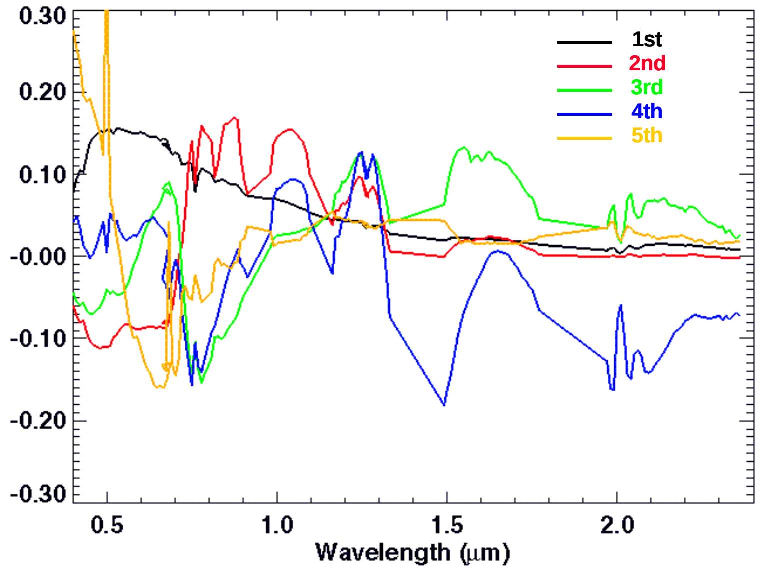

Figure 2(b). Eigenvectors of the SCAR-B HSI data.

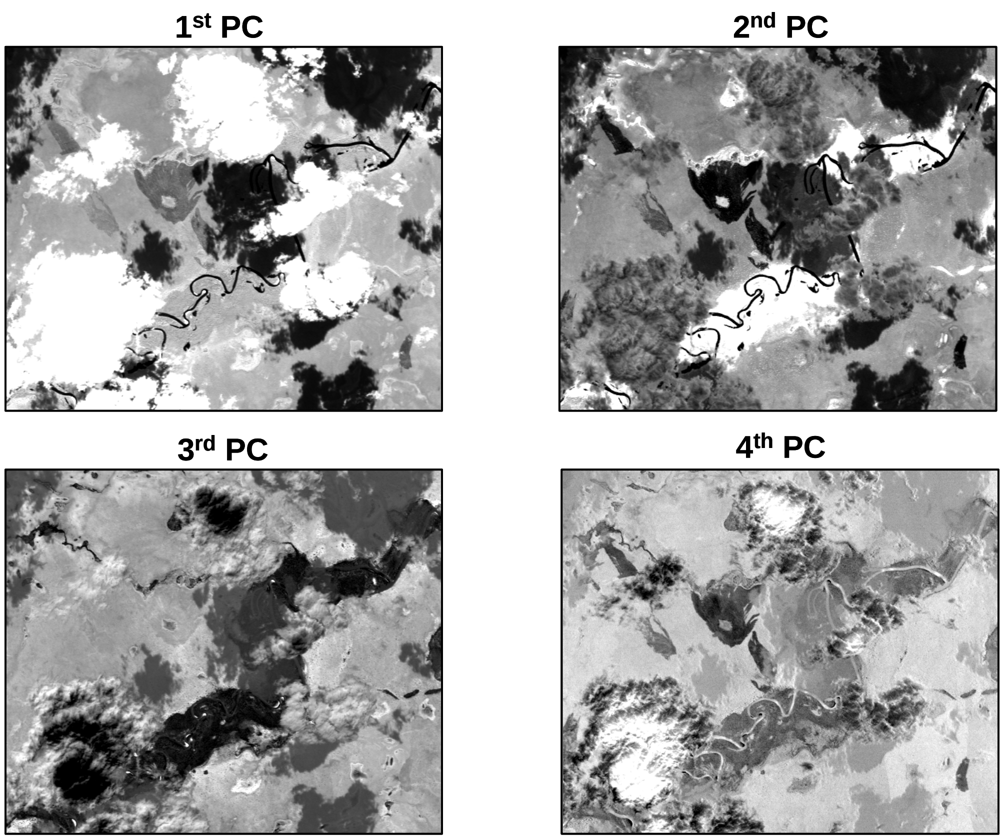

Figure 3. Sample principal components of SCAR-B HSI data.
Each component highlights various features in the scene,
1st PC: cloud, water, shadow; 2nd PC: vegetation;
3rd PC: cloud structure and 4th PC: burned area.

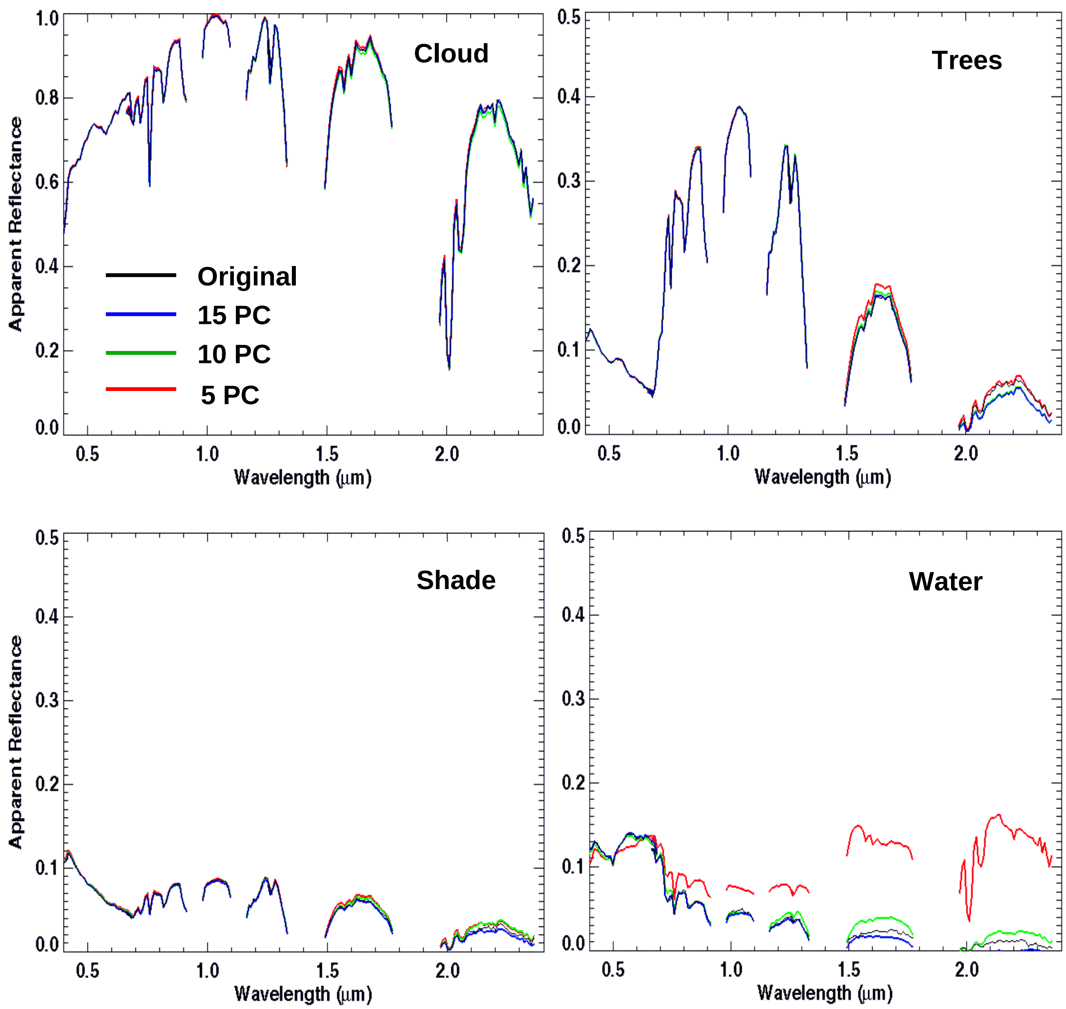

Figure 4. Comparison of reconstructed and original data for sample features in the SCAR-B scene.

Examples of reconstructed spectral images from 10 principal components are illustrated in Figure 5. The differences between the reconstructed and original images are also shown. The agreements are typically within a few percent. Quality of image reconstruction from overall spectral coverage can also be carried out. The original AVIRIS spectral data as well as the reconstructed data are band-integrated to simulate 16 MODIS spectral bands. Two measures, spectral angle and normalized spectral distance (Euclidean distance), are used to evaluate the reconstructed images. These are calculated from the formulas,

$$\textbf{Spectral Angle (X,M)} = \mathbf{COS^{-1}(X \cdot M / |X|\ |M|)}\ ,$$

$$\textbf{Normalized Spectral Distance (X,M)} = \mathbf{|X - M| / |M|}\ ,$$

where, M and X are the multispectral image spectra from the original and reconstructed AVIRIS data, respectively.

A zero angle and zero distance represent a perfect match of the two spectra. The images generated from the original AVIRIS data are considered the "truth" data here. The reconstructed images are compared to the truth data in terms of spectral angle and spectral distance. The spectral distance normalized by the pixel amplitude in the truth image is calculated for comparison. The frame-averaged differences are listed in Table 1. It shows that, as expected, the more principal components (PC) used, the closer is the reconstruction to the truth data. The 10-PC reconstruction appears to be a reasonable one in that it is much closer to the truth in spectral angle and distance compared to 5-PC, but only slightly worse than the 15-PC. As illustrated in the previous section, the 10-PC compression ratio in this case is about 20.

Table 1. Frame Average over the SCAR-B Scene.

p **(No. of PCs)**	**15**	**10**	**5**
(Variance Ratio)	**(3.8e-5)**	**(1.4e-4)**	**(8.7e-4)**
Compression Ratio	**~ 13**	**~ 20**	**~ 40**
Spectral Angle $\mathbf{COS^{-1}(X \cdot M / \lvert X \rvert \lvert M \rvert)}$	**0.3°**	**0.4°**	**0.7°**
Normalized Spectral Distance $\mathbf{\lvert X - M \rvert / \lvert M \rvert}$	**0.6%**	**1.0%**	**1.6%**

M: Pixel spectrum in reference image

X: Pixel spectrum in reconstructed image

4. IMAGE PRODUCT COMPARISON

Accuracy of the reconstructed image can also be evaluated based on spectral products such as vegetation index and column water vapor (CWV). An AVIRIS scene over Moffett Field, CA was employed for the image product comparison. The color images of the AVIRIS data are depicted in Figure 6 and the comparison on derived NDVI is illustrated in Figure 7. Except for the lake region, the 20-PC derived NDVI agreed with the truth to within 0.01. The range of NDVI is ~0.8, the agreement is of the order of a few percent. The Eigenvalue ratio of the 20^{th} and 1^{st} components of this data set is 1.9×10^{-4} which is comparable to the ratio of 10^{th} and 1^{st} components of the Brazil cloud scene.

The CWV field was also computed for both sets of AVIRIS data as a by-product of the ATREM model application for the Moffett Field scene shown in Figure 6. The scene could be characterized as relatively dry (from nearby radiosonde observations) with approximately 1.27 cm of CWV. ATREM was applied to both the original AVIRIS image and the reconstructed image. The primary product of the model being the spectral surface-leaving reflectance and a secondary product, the CWV, was used for this comparison. Figure 8 displays the retrieved CWV from the original AVIRIS data and from reconstructed data using 5, 10, 15, and 20 PCs. From the water vapor image and related RGB image, it is clear that topography plays a role in the distribution of water vapor across the scene. Water vapor is maximum over the urban area which is in close proximity to the San Francisco Bay, and decreases somewhat as the terrain elevation increases (upper part of the image). It is clear from the component difference images that accuracy of the water vapor retrieval (as compared with the original image) improves with increasing PC. Both of the product comparisons shown here suggest that to keep information loss errors below 1%, requires a minimum of 15 PCs to be saved.

5. SUMMARY

Future spaceborne HSI sensors will require onboard data volume reduction before transmission to surface receiving sites. Using principal component analysis (PCA) in a data compression mode, an end-to-end comparison was made on the accuracy of retrieving specific data products from compressed and reconstructed hyperspectral data. PCA was applied to HSI data to produce a set of component images. The original HSI data (at full spectral resolution) was reconstructed from selected sets of the component images. The reconstructed images were used in applications to derive the NDVI and CWV products as well as simulate multispectral channels on the MODIS sensor. Results of the product comparisons show that as few as 15 PCs could be used to obtain product values within 1% of those obtained using the full complement of HSI data.

Using adaptive channel processing, it was shown that multispectral channels could be simulated from reconstructed HSI data with minimal loss in image quality. The study does not deal with individual channel issues such as signal-to-noise which could be channel specific. In this case, the specifications of the HSI bands used in the simulation would determine the specific characteristics of the channel. The reconstruction of HSI data from a selected subset of PCs also has the effect of noise filtering, since much of the information lost due to the truncation of components was noise. On a similar note, the NDVI obtained from simulation of NOAA AVHRR channels using HSI could be improved by selectively excluding spectral bands which do not contribute to the vegetation signatures (i.e., water vapor absorption bands near 0.82 and 0.94 μm).

6. REFERENCES

Gao, B.C., K.B. Heiderbrecht, and A.F.H. Goetz, 1996, “Atmosphere Removal Program (ATREM) Version 2.0 Users Guide”, *Center for the Study of Earth from Space/CIRES*, University of Colorado, Boulder, Colorado.

Geladi, P. and H. Grahn, 1997, Multivariate Image Analysis, John Wiley & Sons.

Griffin, M.K., S.M. Hsu, H.K. Burke and J.W. Snow, 2000, “Characterization and Delineation of Plumes, Clouds and Fires in an AVIRIS Image,” *Proc. of the 2000 JPL Airborne Earth Science Workshop*, JPL, Pasadena, California.

Richards, J., 1994, Remote Sensing Digital Image Analysis, Springer-Verlag, p. 133.

ACKNOWLEDGEMENTS

The authors wish to thank CAPT Frank Garcia at DUSD (S&T), program manager for the Hyperspectral Technology Assessment Program (HTAP), and Dr. Edward Howard at NOAA, task manger for the GOES Program at Lincoln Laboratory, for their support and leadership.

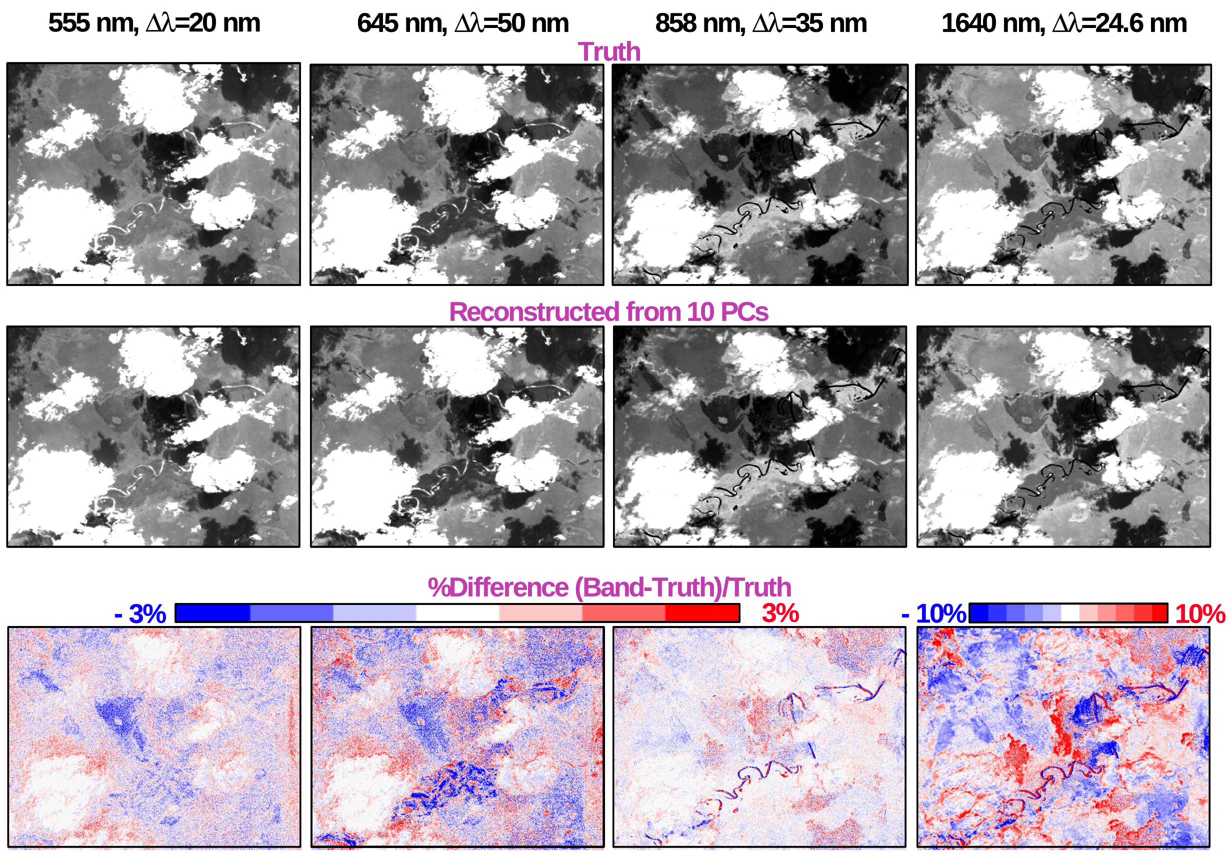

Figure 5. Comparison of reconstructed and original spectral data for simulated MODIS bands.

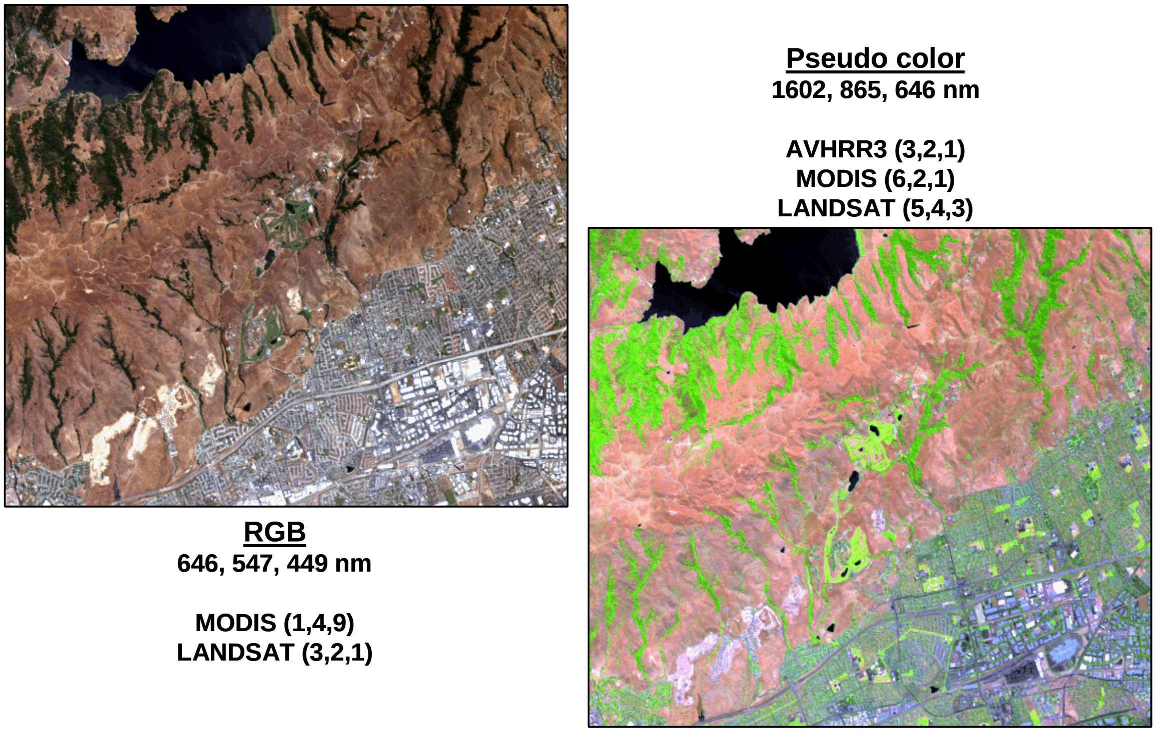

Figure 6. Color images of AVIRIS HSI data over Moffett Field.

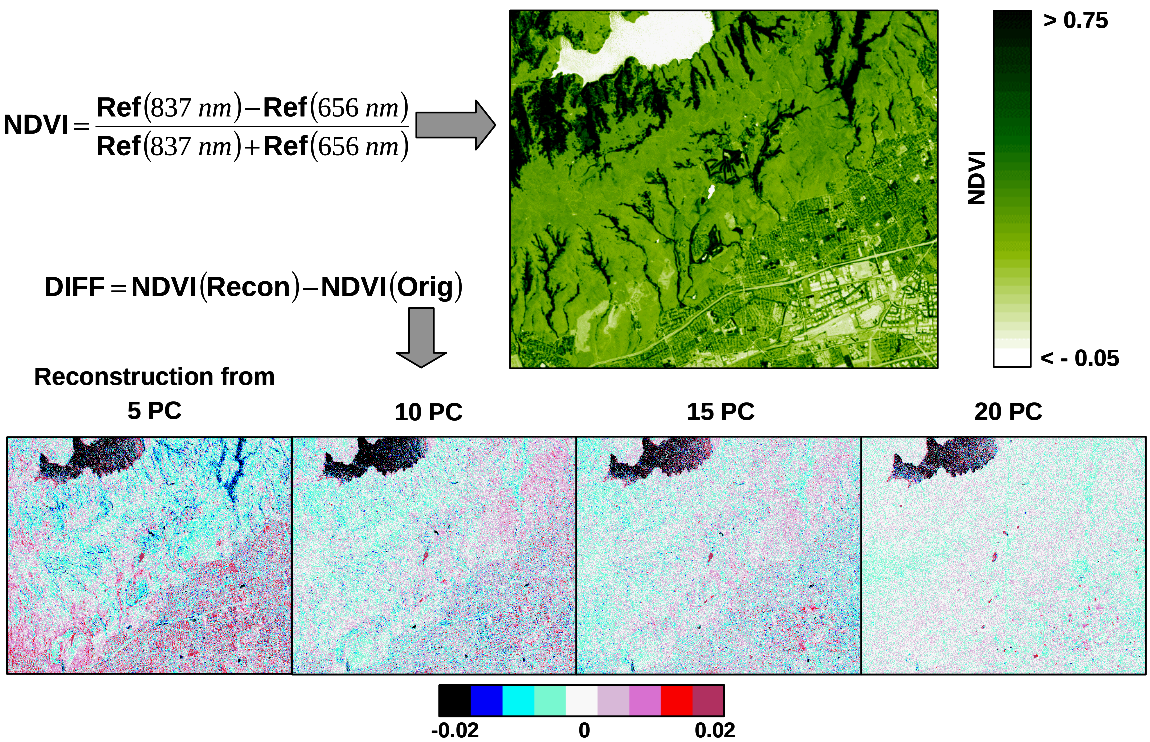

Figure 7. Comparison of NDVI derived from reconstructed and truth data over the Moffett Field scene.

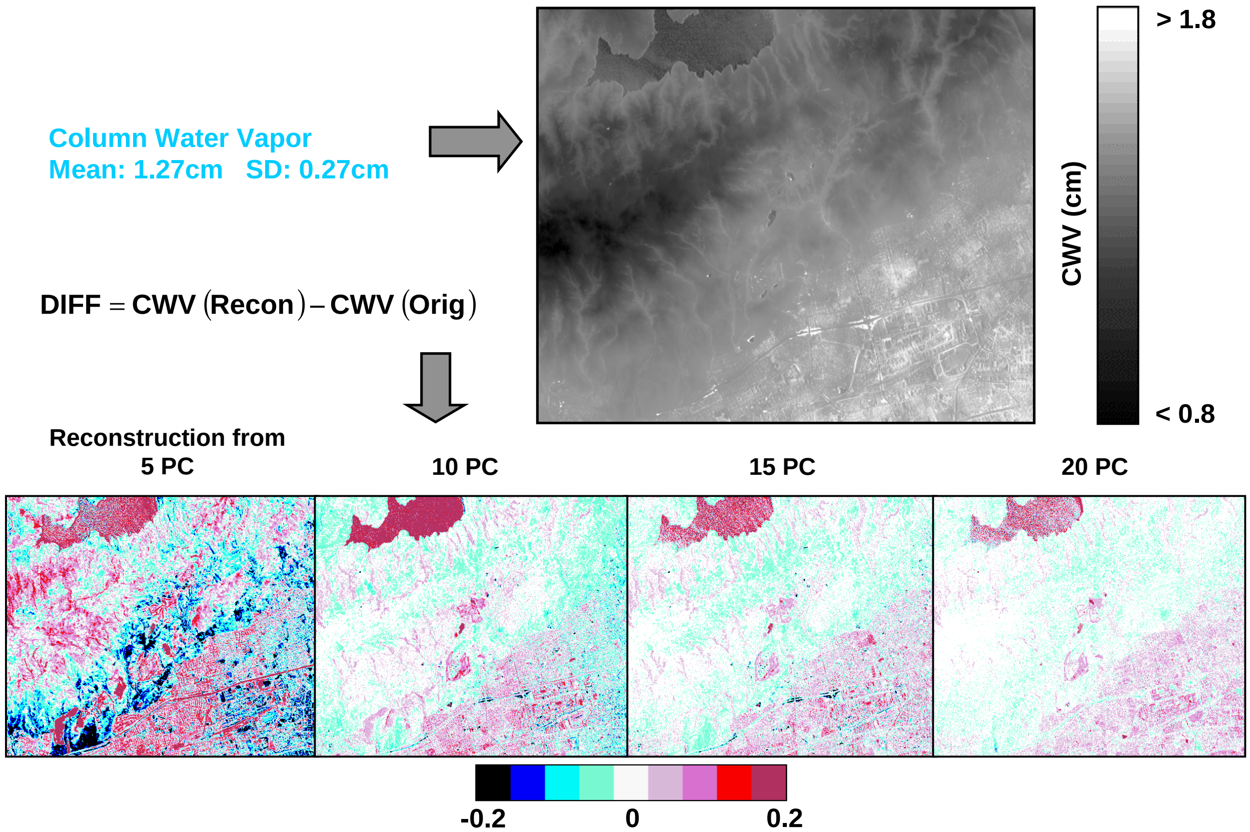

Figure 8. Comparison of CWV derived from reconstructed and truth data over the Moffett Field scene.