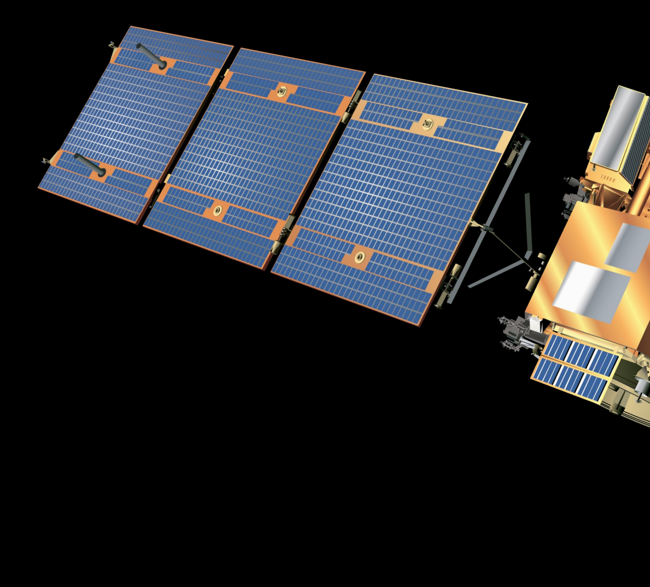
PROGRESS IN REPROCESSING THE HYPERION/EO-1 DATA SET WITH A PROTOTYPE PIPELINE FOR SBG

Jon M. Jenkins[1], Peter Tenenbaum[1,2], Yohei Shinozuka[1,3], Weile Wang[1], Bob Ciotti[1], Bill Wohler[1,2], Ian Brosnan[1], Jennifer Dungan[1], Vanessa Genovese[1,3], Andrew Michaelis[1], Michelle M. Gierach[5], Philip Townsend[6,5], Benjamin Poulter[7]

[1]NASA Ames Research Center, [2]SETI Institute, [3]Bay Area Environmental Research Institute, [4]California State University – Monterey Bay, [5]Jet Propulsion Laboratory, [6]University of Wisconsin, Madison, [7]NASA Goddard Space Flight Center

Abstract:

We describe the progress we've made in developing a pathfinder pipeline to reprocess the 55-TB Hyperion/EO-1 data set to mitigate risk in architecting the science data system for the Surface Biology and Geology (SBG) mission. This work is part of the Space-based Imaging Spectroscopy and Thermal pathfindER (SISTER) study, a collaboration between JPL, Ames, and Goddard. We integrate Python versions of the Hyperion science algorithms with a science pipeline control infrastructure, *Ziggy**, to reprocess the Hyperion data set on NASA's High-End Computing Capability (HECC) at NASA Ames Research Center. Ziggy was developed for NASA's Kepler and Transiting Exoplanet Survey Satellite (TESS) exoplanet survey missions and provides a scalable architecture for robust, repeatable, and replicable science and application products that can be run on a range of systems from a laptop to the cloud or a supercomputer. Ziggy is compliant with NASA Procedural Requirement (NPR) 7150.2C, is at a technical readiness level (TRL) of 7 and has been released to github.com/nasa/ziggy. We integrated Ziggy with EO-1/Hyperion workflows to build a prototype pipeline and ingested the 17-year mission archive that provides globally sampled visible through shortwave infrared spectra that are representative of SBG data types and volumes. We fully implemented the first stage and processed the entire 55 TB Hyperion data set from the raw data (Level 0) to top-of-the-atmosphere radiance (Level 1R). We are currently evaluating the ISOFIT atmospheric correction module to convert the L1R data to surface reflectance (Level 2) before reprocessing the full data set to L2. Crosschecks are being performed with RadCalNet with coincident observations by AVIRIS, and with forward simulations. As the next step, we plan to implement modern methods for georectifying the Hyperion scenes. Finally, we describe an analysis of the cost to conduct forward processing and reprocessing campaigns for SBG on HECC using the resurrected Hyperion pipeline as a proxy for full-volume SBG data, which demonstrate that SBG L0 data can be processed on HECC with full reprocessing campaigns every two years for ~$2.6M over a 7-year lifespan.



1. Reanimating the Hyperion Pipeline:

We've reconstructed the Hyperion pipeline in Python as shown in Figure 1. The processing is complete for the entire data (55 TB) to top-of-atmosphere radiance at full resolution (27 TB). This product, time-referenced and annotated with some ancillary information but not georeferenced, is referred to as level 1R (L1R) in the USG S archive and equivalent to NASA data processing level L1A. From this point, our pipeline allows the atmospheric correction with either the Imaging Spectrometer Optimal FITting (ISOFIT) open source package version 2.9.2 [4] or the non-open source Atmospheric Removal (ATREM) program version 5.0 [5]. Note that the SBG mission is fostering and adhering to NASA's Open Science Policy and thereby strives to implement open-source architecture, workflows, and software. We are now working to incorporate georeferencing and co-registration into the pipeline to complete the pipeline to an L1B georeferenced top-of-atmosphere radiance (approx. 27 TB), and georeferenced, orthocorrected level 2 (L2) reflectance product (approx. 124 TB). We also have begun conducting experiments to evaluate and verify the results and are investigating how the processing scales with data volume for both L0→L1A, L1A→L1B, and L1B→L2.

Note that in this context, ATREM is being used only for comparison purposes on selected scenes and we plan to use ISOFIT to process the global Hyperion data set to L2 for release to the community.

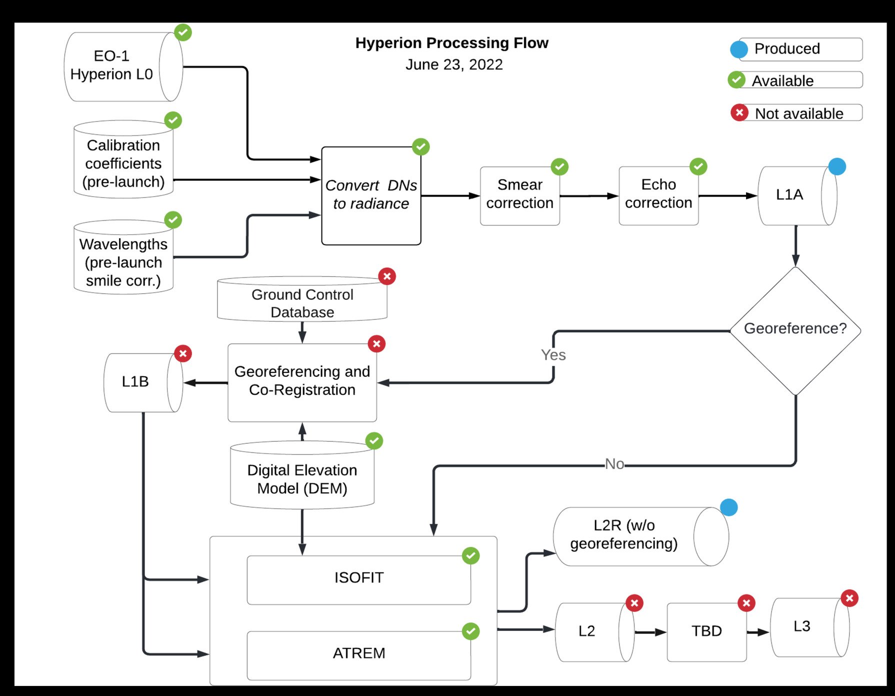


Figure 1. Hyperion Processing Flow.

2. Evaluating the Hyperion Pipeline:

We have been conducting exercises to evaluate and verify the results for specific scenes:

- Comparing the Hyperion L2 surface reflectance reconstructed with ISOFIT to the ground-based RadCalNet observations [6].
- Comparing the ISOFIT atmospheric products to the ground-based AERONET observations [7].
- Comparing surface reflectance spectra generated using ATREM with those using ISOFIT.
- Comparing surface reflectance spectra of Hyperion with airborne AVIRIS scenes.
- Comparing Hyperion results with forward modeling using MODIS data.

The Hyperion reflectance spectra show some unphysical patterns, which may indicate calibration corrections not available to this workflow [2]. Nonetheless, they show generally good agreement with near-coincident RadCalNet observations at the Railroad Valley site (Figure 2). The root-mean-square difference across the wavelength is 0.04-0.09 for the four pairs shown in Figure 2. Atmospheric changes between each pair of measurements may explain the reflectance differences such as those noticeable in the bottom panel.

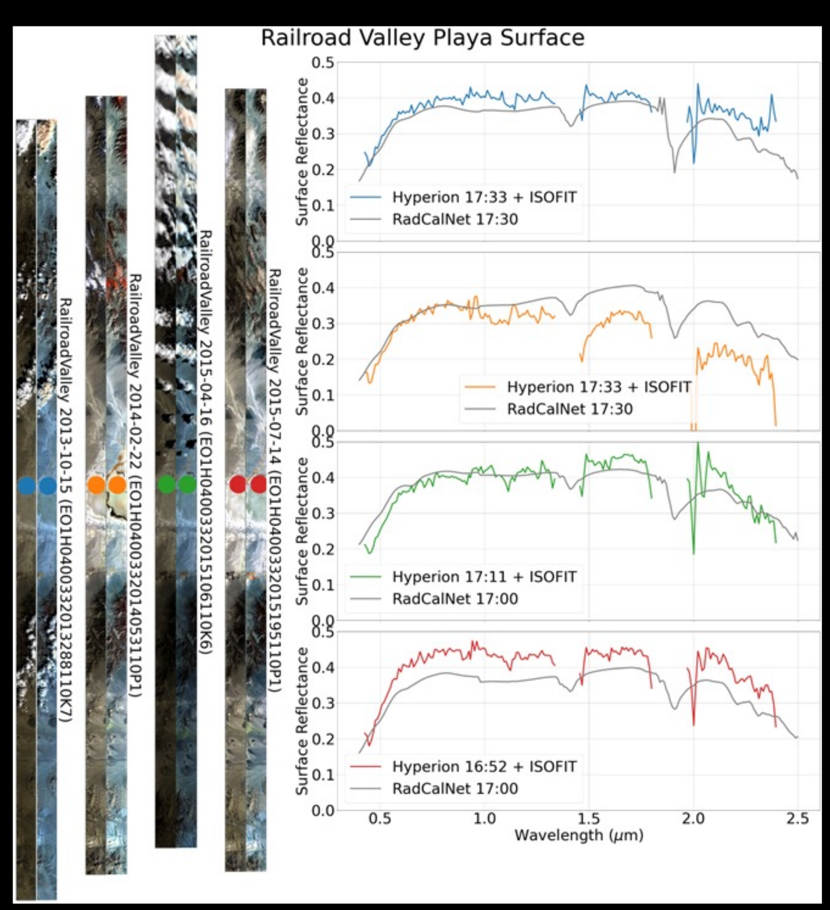


Figure 2. Comparison of RadCalNet measurements with Hyperion surface reflectance retrievals for scenes observed in Railroad Valley. Four VNIR/SWIR Hyperion image pairs appear to the left with markers indicating the pixel location for the surface reflectance plotted on the right-hand side vs. wavelength. The solid black curves are the results obtained by RadCalNet. The blue, orange, green and red curves show Hyperion results obtained using the ISOFIT package for atmospheric corrections.

3.Atmospheric Retrievals:

ISOFIT retrieves atmospheric products in addition to surface reflectance, enabling comparison of the Hyperion scenes that include Railroad Valley with the AERONET observations. The root-mean-square differences for near-coincident cases (< 1 km horizontal distance and <30-minute temporal gap, including scenes not shown in Figure 2) marked with closed circles in Figure 3 is 0.14 cm in column water vapor and 0.11 in 500-550 nm aerosol optical depth, respectively. The aerosol optical depth retrieved with Hyperion and ISOFIT tends to be greater than the ground-observed value. Systematic differences may be explained by calibration inaccuracy; their dependence on wavelength, location and year remain to be investigated.

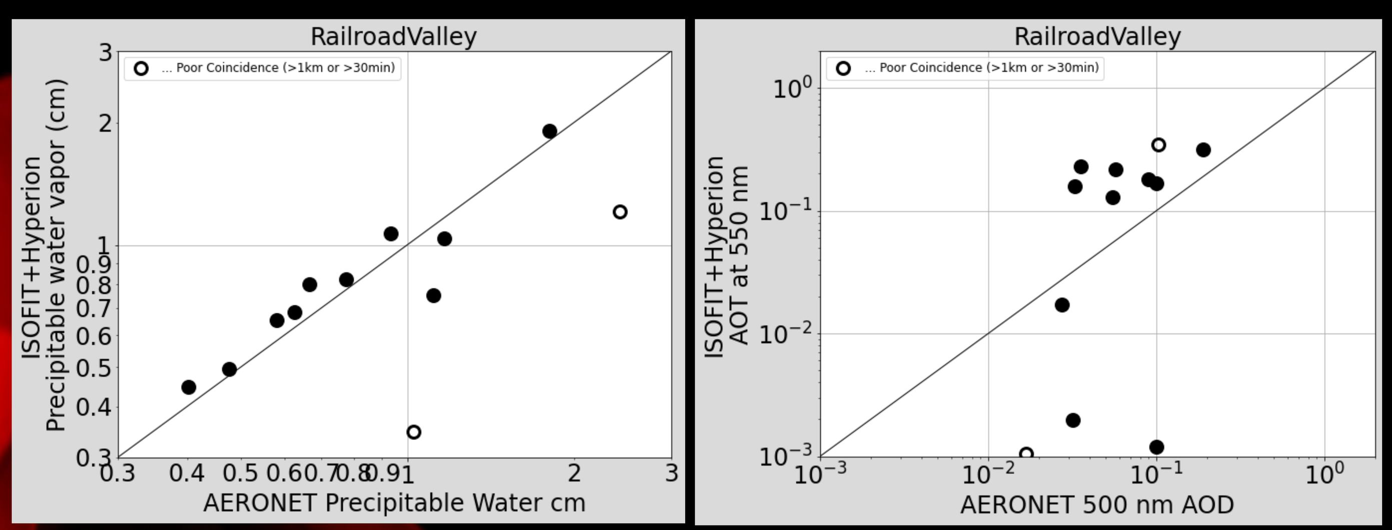
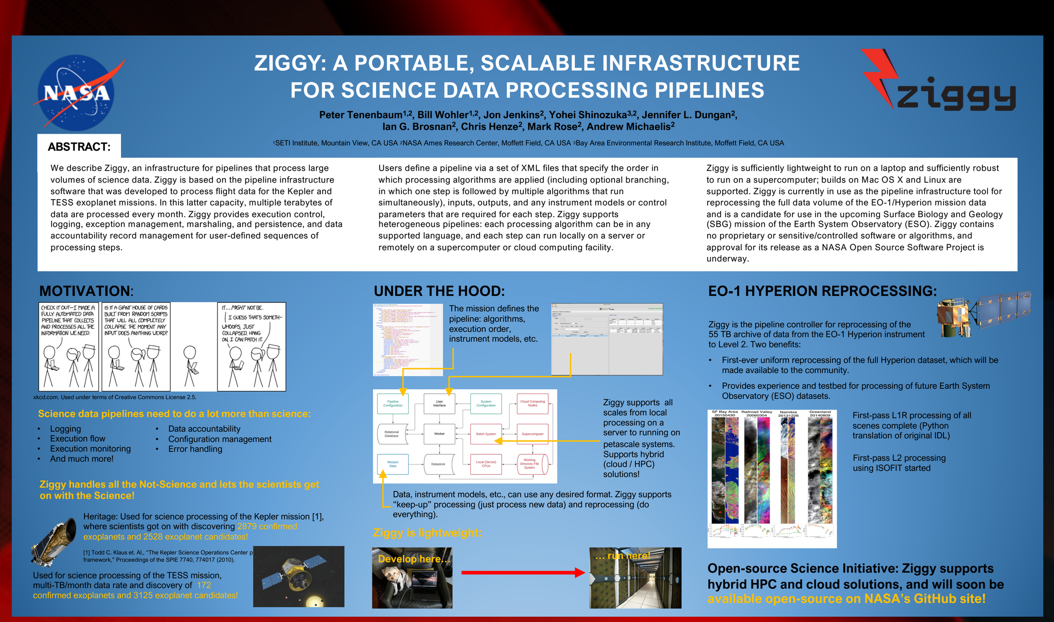


Figure 3. Comparison of Hyperion ISOFIT atmospheric products (left: water vapor, right: aerosol optical depth) with AERONET observations at Railroad Valley.



For more information on Ziggy, please consult this poster

4. Comparison of surface reflectance spectra retrieved from Hyperion and AVIRIS

Figure 4 shows the retrieval products for near-coincident airborne AVIRIS-Classic observations generated with the same two algorithms. Note that georeferencing errors are expected, perhaps by 5-10 pixels. The comparison between the two instruments highlights the unphysical spectral patterns of Hyperion, such as the noisiness across the entire wavelength range and the weak signal below 450 nm. While good agreement between the two instruments is evident near 600 nm and near 1600 nm, the Hyperion reflectance is smaller than the AVIRIS in other spectral regions.

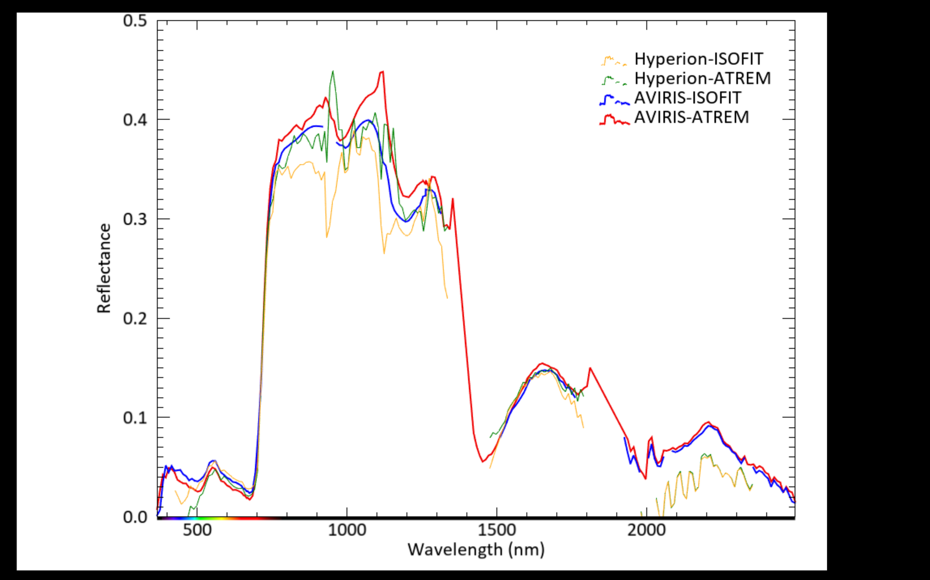
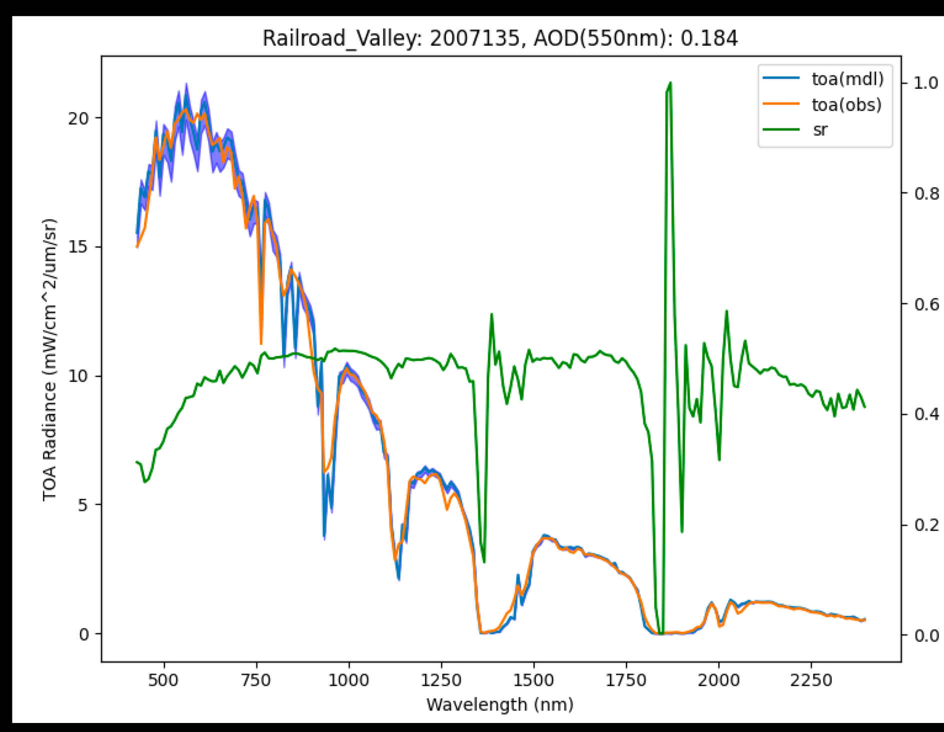


Figure 4. Comparison of surface reflectance spectra retrieved with the Hyperion and AVIRIS sensors, each with ISOFIT and ATREM retrieval algorithms. The results are shown for a vegetation site near Half Moon Bay, CA, observed on April 30, 2015.

5. Forward Modeling Results

Here we compare Hyperion TOA radiance spectra with forward models propogated through the atmosphere from the ISOFIT-derived surface reflectance spectra and arosol optical depth. The model TOA agrees well with the measured TOA from Hyperion

Figure 5. Comparison of Hyperion top of atmosphere (TOA) radiance (orange) for a Hyperion Railroad Valley scene with forward modeling. The orange line shows the TOA radiance measured by Hyperion. The green line is the surface reflectance (SR) retrieved by ISOFIT. The blue line indicates the TOA radiance simulated with ISOFIT SR and AOD retrievals. The purple shades indicate the variations of the TOA radiance by changing the AOD with a scaling factor from 0.5 to 2.



6. Processing SBG Data from L0 to L2 on NASA's High-End Computing Capability (HECC) Facility

Large data sets from NASA flight missions can be processed by NASA's Advanced Supercomputer (NAS) Division's supercomputer for as low as $0.115 per Standard Billing Unit (one hour of time on a Broadwell node). HECC enjoys economies of scale comparable to commercial cloud vendors. Projects can purchase their own compute and storage for their exclusive use on margin.

For this estimate we assume a 7-year mission and full reprocessing runs every two years. We take the 55-TB Hyperion data set as a proxy for SBG and scale the estimates for the SBUs required for reprocessing the Hyperion data set to L2. Here, 14,000 SBUs needed for processing Hyperion scale to 232,432 SBUs/year of SBG data. A 64-node Milan architecture can provide 2,130,522 SBUs/year assuming a typical 90% availability.

We include all costs to the project for these resources, including compute, storage, tape backups, maintenance, operations costs (staff), power, water, facility, software licensing, etc.

69% of the capacity is available for other activities: algorithm development, L3+ processing, more frequent re-processing, etc.

The bottom line: HECC can support forward and reprocessing for SBG from L0 to L2 for approximately $2.6M.

Costs Basis Est from next plan install to buy HPC on Margin

$0.115/SBU - 7 year

Scaling Hyperion up to SBG

55 TB L0 ⇔ ~22 days of SBG

14,000 SBUs for Hyperion is ~232,432 SBUs/year for SBG

7-year purchase, 64 nodes, 2 PB storage

2 PB of tape each year for L0 (2 copies)

Reprocessing every two years

2 hours compute/2 hours transfer @30gb/sec

	Forward (KiloSBU)	Reprocessing (KiloSBU)	Annual (KiloSBU)	Compute Tape	Data Storage (K $)	Projected HECC Cost (K $)
Year						
0				$ 1,719	$ 250	$ 1,969
1	232		232	$ 84		$ 84
2	232		232	$ 84		$ 84
3	232	464	696	$ 84		$ 84
4	232		232	$ 84		$ 84
5	232	928	1160	$ 84		$ 84
6	232		232	$ 84	$ 25	$ 109
7	232	1392	1856	$ 84	$ 25	$ 109
						$ 2,607
					percent utilization	31%

References:

[1] "USGS EROS Archive - Earth Observing One (EO-1) - Hyperion," Earth Resources Observation and Science (EROS) Center , 17 April 2019. [Online]. Available: https://www.usgs.gov/centers/eros/science/usgs-eros-archive-earth-observing-one-eo-1-hyperion. [Accessed 1 June 2022].

[2] X. Jing, L. Leigh, D. Helder, C. Teixeira Pinto and D. Aaron, "Lifetime Absolute Calibration of the EO-1 Hyperion Sensor and its Validation," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 57, pp. 9466-9475, 2019.

[3] P. Tenenbaum, J. M. Jenkins, Y. Shinozuka, B. Wohler, A. Michaelis, J. L. Dungan and I. G. Brosnan, *A Prototype Data Analysis Pipeline Infrastructure for the Surface Biology and Geology (SBG) Mission*, SBG Community White Papers, 2022.

[4] D. R. Thompson, V. Natraj, R. O. Green, M. C. Helmlinger, B.-C. Gao and M. L. Eastwood, "Optimal estimation for imaging spectrometer atmospheric correction," *Remote Sensing of Environment*, vol. 216, pp. 355-373, 2018.

[5] B.-C. Gao and A. F. H. Goetz, "Column atmospheric water vapor and vegetation liquid water retrievals from Airborne Imaging Spectrometer data," *J. Geophys. Res.*, vol. 95(D4), p. 3549–3564, 1990.

[6] M. Bouvet, K. Thome, B. Berthelot, A. Bialek, J. Czapla-Myers, N. Fox, P. Goryl, P. Henry, L. Ma, S. Marcq, A. Meygret, B. Wenny and E. Woolliams, "RadCalNet: A Radiometric Calibration Network for Earth Observing Imagers Operating in the Visible to Shortwave Infrared Spectral Range," *Remote Sensing*, vol. 11, p. 2401, 2019.

[7] B. N. Holben, T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak and A. Smirnov, "AERONET - A federated instrument network and data archive for aerosol characterization," *Remote Sensing of Environment*, vol. 66, pp. 1-16, 1998.

[8] W. Wang, J. Dungan, V. Genovese, Y. Shinozuka, B. Poulter and I. Brosnan, *Development of the Ames Global Hyperspectral Synthetic Dataset*, SBG Community White Papers, 2022.