Progress in Developing a Prototype Science Pipeline and Full-Volume, Global Hyperspectral Synthetic Data Sets for NASA's Earth System Observatory's Upcoming Surface, Biology and Geology Mission

Jon Jenkins[1], Peter Tenenbaum[2], Yohei Shinozuka[3], Bill Wohler[2], Andrew Michaelis[1], Jennifer Dungan[1], Ian Brosnan[1], Vanessa Genovese[4], Weile Wang[1], Michelle Gierach[5], Philip Townsend[6], and Ben Poulter[7]

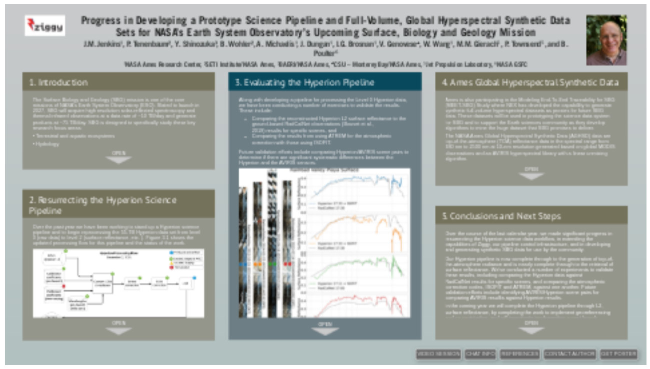
[1]NASA Ames Research Center
[2]SETI Institute/NASA Ames
[3]BAERI/NASA Ames
[4]CSU – Monterey Bay/NASA Ames
[5]Jet Propulsion Laboratory
[6]University of Wisconson/Jet Propulsion Laboratory
[7]NASA Goddard Space Flight Center

November 24, 2022

Abstract

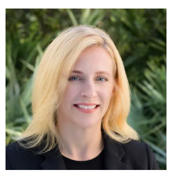
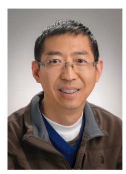
The Surface Biology and Geology (SBG) mission is one of the core missions of NASA's Earth System Observatory (ESO). SBG will acquire high resolution solar-reflected spectroscopy and thermal infrared observations at a data rate of ˜10 TB/day and generate products at ˜75 TB/day. As the per-day volume is greater than NASA's total extant airborne hyperspectral data collection, collecting, processing/re- processing, disseminating, and exploiting the SBG data presents new challenges. To address these challenges, we are developing a prototype science pipeline and a full-volume global hyperspectral synthetic data set to help prepare for SBG's flight. Our science pipeline is based on the science processing operations technology developed for the Kepler and TESS planet-hunting missions. The pipeline infrastructure, Ziggy, 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 an on-site supercomputer. Our effort began by ingesting data and applying workflows from the EO- 1/Hyperion 17-year mission archive that provides globally sampled visible through shortwave infrared spectra that are representative of SBG data types and volumes. We have fully implemented the first stage of processing, from the raw data (Level 0) to top-of-the-atmosphere radiance (Level 1R). We plan to begin reprocessing the entire 55 TB Hyperion data set by the end of 2021. Work to implement an atmospheric correction module to convert the L1R data to surface reflectance (Level 2) is also underway. Additionally, an effort to develop a hybrid High Performance Computing (HPC)/cloud processing framework has been started to help optimize the cost, processing throughput and overall system resiliency for SBG's science data system (SDS). Separately, we have developed a method for generating full-volume synthetic data sets for SBG based on MODIS data and have made the first version of this data set available to the community on the data portal of NASA's Advanced Supercomputing Division at NASA Ames Research Center. The synthetic data will make it possible to test parts of the pipeline infrastructure and other software to be applied for product generation.

Progress in Developing a Prototype Science Pipeline and Full-Volume, Global Hyperspectral Synthetic Data Sets for NASA's Earth System Observatory's Upcoming Surface, Biology and Geology Mission



J.M. Jenkins[1], P. Tenenbaum[2], Y. Shinozuka[3], B. Wohler[2], A. Michaelis[1], J. Dungan[1], I.G. Brosnan[1], V. Genovese[4], W. Wang[1], M.M. Gierach[5], P. Townsend[5], and B. Poulter[6]

[1]NASA Ames Research Center, [2]SETI Institute/NASA Ames, [3]BAERI/NASA Ames, [4]CSU – Monterey Bay/NASA Ames, [5]Jet Propulsion Laboratory, [6]NASA GSFC




PRESENTED AT:

1. INTRODUCTION

The Surface Biology and Geology (SBG) mission is one of the core missions of NASA's Earth System Observatory (ESO). Slated to launch in 2027, SBG will acquire high resolution solar-reflected spectroscopy and thermal infrared observations at a data rate of ~10 TB/day and generate products at ~75 TB/day. SBG is designed to specifically study these key research focus areas:

• Terrestrial and aquatic ecosystems

• Hydrology

• Weather

• Climate

• Solid Earth

Since the per-day volume is greater than NASA's total extant airborne hyperspectral data collection, collecting, processing/re-processing, disseminating, and exploiting the SBG data presents new challenges.

As part of SBG's Space-based Imaging Spectroscopy and Thermal pathfindER (SISTER) study, we are working to develop hybrid, high-performance computing (HPC)-cloud computing frameworks as a potential architecture for SBG. The pipeline control infrastructure used in this effort, "Ziggy (https://agu2021fallmeeting-agu.ipostersessions.com/?s=17-C2-C6-82-8B-51-EF-16-20-A3-8C-AF-09-9F-12-41)", is based on that developed for NASA's Kepler and Transiting Exoplanet Survey Satellite missions (see IN45G-0517 (https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/966160)). Using Ziggy, we've reprocessed the entire 17-year Hyperion data set as a proxy for SBG data from L0 to L1R, top-of-the-atmosphere radiance. We are currently processing all the L1R data to L2, surface reflectance.

In addition, Ames is also participating in the Modeling End-To-End Traceability for SBG (MEET-SBG) Study and has developed the capability to generate synthetic full-volume hyperspectral datasets similar to those expected from SBG using MODIS data and an Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) hyperspectral library with a linear unmixing algorithm. The first data set, corresponding to a full year of SBG data, has been released on NASA's Advanced Supercomputing Division's Data Portal (https://data.nas.nasa.gov/aghsd/). These datasets will be used in prototyping the science data system for SBG and to support the Earth sciences community as they develop algorithms to mine the huge dataset that SBG promises to deliver.

[1]SISTER is led by NASA's Jet Propulsion Laboratory in collaboration with Ames Research Center and Goddard Space Flight Center.

2. RESURRECTING THE HYPERION SCIENCE PIPELINE

Over the past year we have been working to stand up a Hyperion science pipeline and to begin reprocessing the 55-TB Hyperion data set from level 0 (raw data) to level 2 (surface reflectance, etc.[2]). Figure 3.1 shows the updated processing flow for this pipeline and the status of the work.

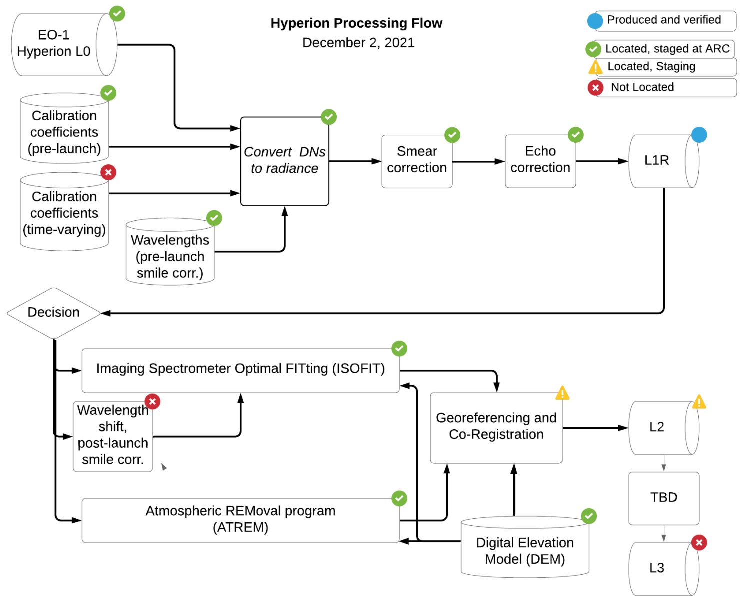


Figure 2.1 Hyperion Processing Flow.

Currently the pipeline is complete to level 1R, top-of-the-atmosphere radiance and we've completed reprocessing the entire 55-TB Hyperion data set to this stage. We are working to complete the pipeline to level 2, surface reflectance, and are working to incorporate georeferencing and co-registration into the pipeline. The atmospheric correction can be made using either the Imaging Spectrometer Optimal FITting (ISOFIT – Thompson et al., 2018) algorithm, or the Atmospheric Removal (ATREM – Gao & Geotz 1990 (https://www.sciencedirect.com/science/article/abs/pii/003442579390014O)) algorithm. We also have begun conducting experiments to evaluate and validate the results and are investigating how the processing scales with data volume for both L0->L1R and L1R->L2.

With regard to the "not located" items in Figure 2.1:

- Time-varying (post-launch) calibration coefficients were not applied for the Level 1 products as the relative trends derived from the lunar observations did not show significant long term changes for most of the spectral range (Lawrence Ong, personal communication),
- We are currently evaluating the need for a wavelength shift/smile correction, and
- We intend to select and implement L3 science algorithms in the near future.

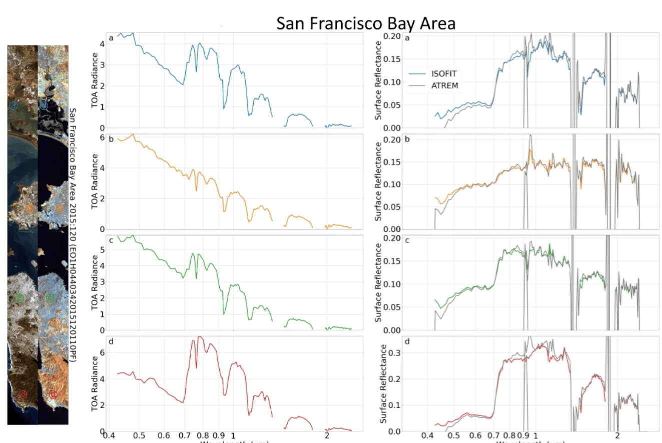


Figure 2.2 Top-of-the-atmosphere (TOA) radiance and surface reflectance spectra (SR) derived from Hyperion observations of the San Francisco Bay Area taken on 2015-04-30 at 17:28. The image pair on the left are false color composites made from the visible and near infrared (VNIR) and short-wave infrared (SWIR) observations of the Bay Area (from left to right). The labels, a, b, c and d, on the composite images indicate the locations for which the TOA and SR are shown vs wavelength in the two sets of graphs to the right of the images. TOA is shown in the left column of line plots, and both ATREM and ISOFIT results are shown in the right column of line plots of surface reflectance.

[2]Note that ISOFIT also retrieves additional products in addition to surface reflectance, including atmospheric optical depth and water vapor.

3. EVALUATING THE HYPERION PIPELINE

Along with developing a pipeline for processing the Level 0 Hyperion data, we have been conducting a number of exercises to validate the results. These include:

- Comparing the reconstructed Hyperion L2 surface reflectance to the ground-based RadCalNet observations (Bouvet et al., 2019) results for specific scenes, and
- Comparing the results from using ATREM for the atmospheric correction with those using ISOFIT.

Future validation efforts include comparing Hyperion/AVIRIS scene pairs to determine if there are significant systematic differences between the Hyperion and the AVIRIS sensors.

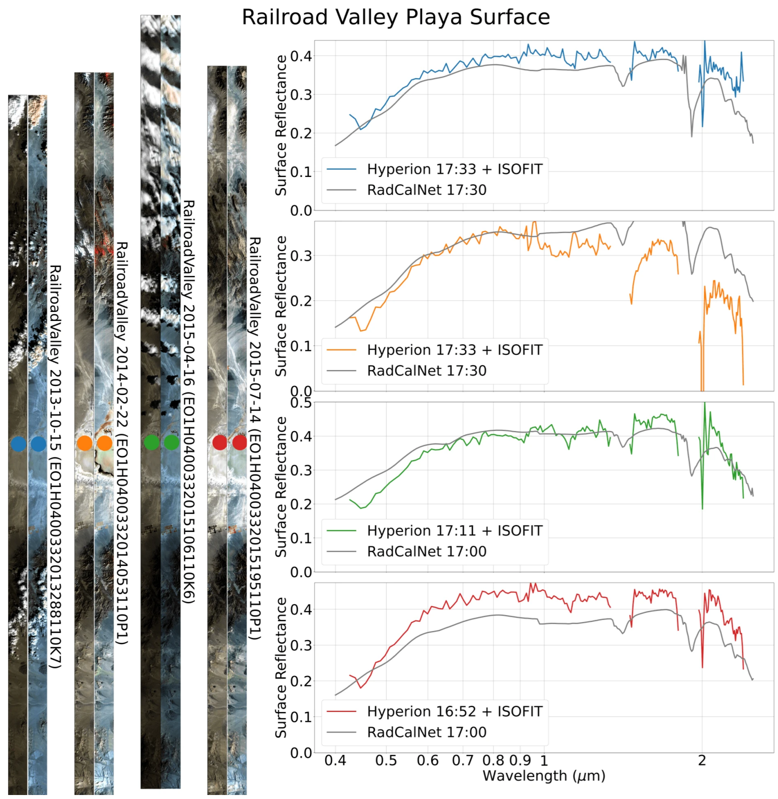


Figure 3.1 Comparison of RadCalNet measurements with Hyperion surface reflectance retrievals for scenes observed in Railroad Valley. Four VNIR/SWIR Hyperion false color composite image pairs appear to the left with markers indicating the pixel location for the surface reflectance plotted on the right hand side vs. wavelength. Each of the data sets were collected at 2013-10-15 17:33, 2014-02-22 17:33, 2015-04-16 17:11, and 2015-07-14 16:52, from left to right, respectively. The solid dark gray 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. Atmospheric changes between each pair of measurements may explain the reflectance differences such as those noticeable in the bottom panel.

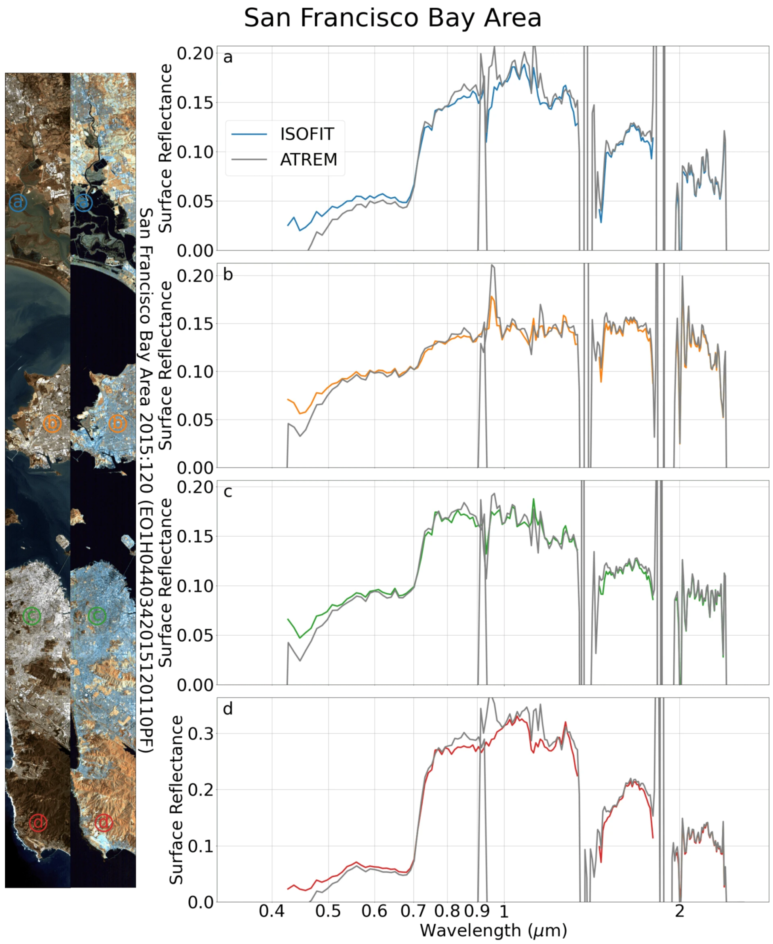


Figure 3.2 Comparison of ATREM with ISOFIT for the atmospheric correction used to convert top-of-the-atmosphere radiance to surface reflectance from observations made at 2015-04-30 17:28. A Hyperion VNIR/SWIR false color composit image pair obtained in the San Francisco Bay Area appears on the left with labels a, b, c and d indicating the pixel locations for which ISOFIT- and ATREM-derived surface reflectance is plotted vs wavelength on the right.

4. AMES GLOBAL HYPERSPECTRAL SYNTHETIC DATA

Ames is also participating in the Modeling End-To-End Traceability for SBG (MEET-SBG) Study where NEX has developed the capabililty to generate synthetic full-volume hyperspectral datasets as proxies for future SBG data. These datasets will be used in prototyping the science data system for SBG and to support the Earth sciences community as they develop algorithms to mine the huge dataset that SBG promises to deliver.

The NASA Ames Global Hyperspectral Synthetic Data (AGHSD) data are top-of-the-atmosphere (TOA) reflectance data in the spectral range from 380 nm to 2500 nm at 10-nm resolution generated based on global MODIS observations and an AVIRIS hyperspectral library with a linear unmixing algorithm.

Currently the first full year of these synthetic SBG data are available for download on the NAS Data Portal: https://data.nas.nasa.gov/aghsd/ (https://data.nas.nasa.gov/aghsd/)

[VIDEO] https://res.cloudinary.com/amuze-interactive/video/upload/vc_auto/v1638405234/agu-fm2021/BA-FC-6F-80-E8-23-64-5A-4C-90-30-09-CD-B7-CA-DA/Video/AGHSD_global_v2_qw40ao.mp4

Figure 4.1 An Ames Global Hyperspectral Synthetic Data (AGHSD) animation depicting the top-of-the-atmosphere reflectance in the spectral range from 380 nanometers (nm) to 2,500 nm with 10 nm bandwidths and ~1-kilometer spatial resolution at a global scale. The animation cycles through the 213 available bands for July 1, 2019.

[VIDEO] https://res.cloudinary.com/amuze-interactive/video/upload/vc_auto/v1638405163/agu-fm2021/BA-FC-6F-80-E8-23-64-5A-4C-90-30-09-CD-B7-CA-DA/Video/hypercube-BayArea_ihogio.mp4

Figure 4.2 An Ames Global Hyperspectral Synthetic Data (AGHSD) animation of a hypercube for a section (tile h09v03) located in the San Francisco Bay Area, for data collected on June 15, 2019.

5. CONCLUSIONS AND NEXT STEPS

Over the course of the last calendar year, we made significant progress in resurrecting the Hyperion science data workflow, in extending the capabilities of Ziggy, our pipeline control infrastructure, and in developing and generating synthetic SBG data for use by the community.

Our Hyperion pipeline is now complete through to the generation of top-of-the-atmosphere radiance and is nearly complete through to the retrieval of surface reflectance. We've conducted a number of experiments to validate these results, including comparing the Hyperion data against RadCalNet results for specific scenes, and comparing the atmospheric correction codes, ISOFIT and ATREM, against one another. Future validation efforts include identifying AVIRIS/Hyperion scene pairs for comparing AVIRIS resullts against Hyperion results.

In the coming year we will complete the Hyperion pipeline through L2, surface reflectance, by completing the work to implement georeferencing and co-registration, and will investigate implementing a post-launch wavelength shift and smile correction code. We will then reprocess the entire Hyperion data set to L2, which will represent the first time that the entire global 55-TB data set has been uniformly processed to L2.

Further, we plan to implement a few level 3 and higher algorithms in the pathfinder Hyperion pipeline with candidate algorithms to be drawn from a selection of algorithms related to terrestrial vegetation traits, aquatic scenes, snow and ice properties, and/or geology.

We also made excellent progress in streamlining Ziggy and signficantly reducing the amount of code required to integrate new science modules into a pipeline, in extending its capabilities for the Hyperion use case, and in investigating hybrid cloud-HPC frameworks. In the coming year we plan to continue to develop the hybrid Ziggy concept and plan to conduct experiments to learn the relative cost/compute efficiency between the NAS Pleiades supercomputer and Amazon Web Services.

SISTER, MEET and NEX are funded by NASA's Earth Science Directorate.

REFERENCES

Bouvet, M., K. Thome, B. Berthelot, A. Bialek, J. Czapla-Myers, N. P. Fox, P. Goryl, P. Henry, L. Ma, S. Marcq, A. Meygret, B.N. Wenny, E.R. Woolliams (2019). RadCalNet: A Radiometric Calibration Network for Earth Observing Imagers Operating in the Visible to Shortwave Infrared Spectral Range. (https://gcc02.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdoi.org%2F10.3390%2Frs11202401&data=04%7C01%7Cjon.jenkins%40nasa.gov%7Cd25b28eabd0c4468f4ea08d9b987fd56%7C7005d45845be48ae8140d4 *Sens.* **11**, 2401. DOI: 10.3390/rs11202401

Gao, B.C., & A.F. Goetz (1990). Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. (https://doi.org/10.1029/JD095iD04p03549) *Journal of Geophysical Research – Atmospheres*, **95**(D4), 3549–3564. DOI:10.1029/JD095iD04p03549

Thompson, D. R., V. Natraj, R. O. Green, M. Helmlinger, B.-C. Gao, and M. L. Eastwood (2018). Optimal Estimation for Imaging Spectrometer Atmospheric Correction. (https://doi.org/10.1016/j.rse.2018.07.003) *Remote Sensing of Environment* **216**, 355-373. DOI: 10.1016/j.rse.2018.07.003.