Imaging Spectroscopy Processing Environment on the Cloud (ImgSPEC)

E. Natasha Stavros[1], Philip Townsend[2], George Chang[3], Hook Hua[3], Thomas Huang[3], Justin Merz[4], Winston Olson-Duvall[5], William Phyo[1], Sujen Shah[3], and David Thompson[5]

[1]NASA Jet Propulsion Laboratory
[2]University of Wisconsin
[3]Jet Propulsion Laboratory
[4]University of California Davis
[5]Jet Propulsion Laboratory, California Institute of Technology

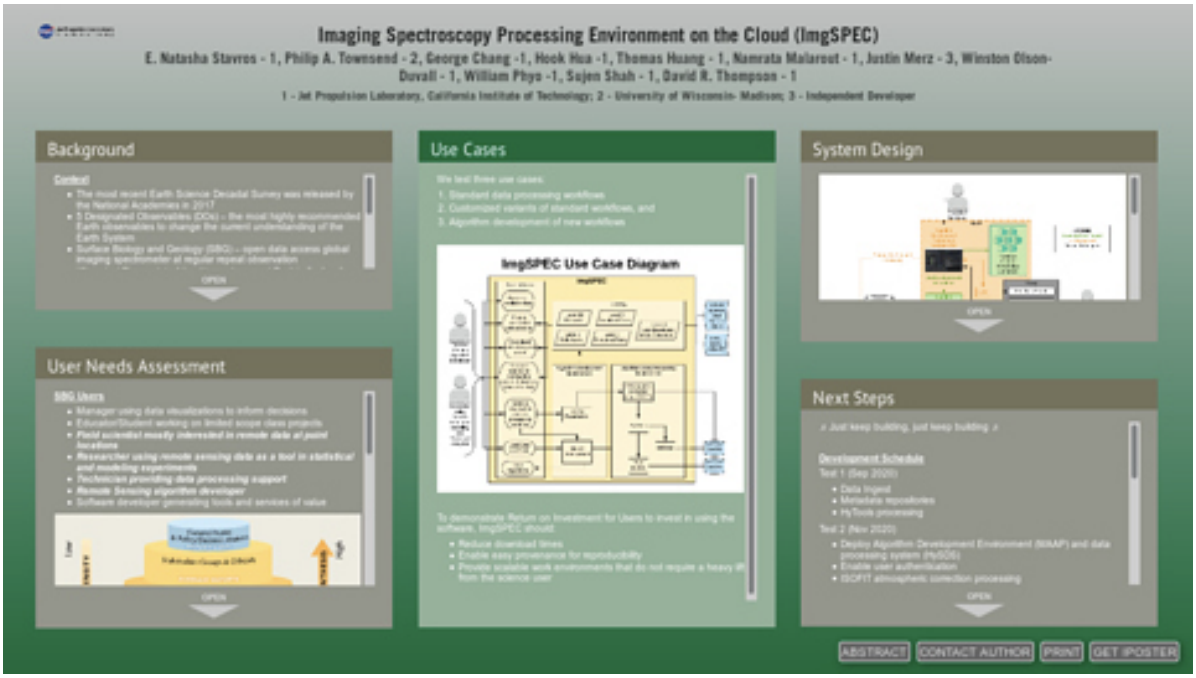
November 24, 2022

Posted on 24 Nov 2022 — CC-BY-NC-ND 4 — https://doi.org/10.1002/essoar.10504790.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Abstract

The geospatial Imaging Spectroscopy Processing Environment on the Cloud (ImgSPEC; formerly GeoSPEC) pioneers an on-demand science data processing system (SDPS) producing user-customized Level 1 calibrated radiance to Level 3+ data products in anticipation for the 2017-2027 Earth Decadal Survey prioritized spaceborne global imaging spectrometer to advance the study of Surface Biology and Geology (SBG). SBG data volumes (˜20 TB/day) of high dimensionality (>224 bands) would be infeasible to download and the breadth of applications of the data across dozens of disciplines presents a need to evolve the traditional NASA SDPS. ImgSPEC streamlines processing data into key SBG observables that have demonstrated algorithms at local-to-regional scales and may vary locally. As such, a traditional, monolithic SDPS could not fully exploit the information in SBG measurements. To remove this barrier to use, ImgSPEC demonstrates an on-demand SDPS prototype that improves imaging spectroscopy data discovery, access, and utility enabling shared knowledge transfer from advanced imaging spectroscopy users to less experienced users such as decision makers and the general public. We test three use cases: 1) standard data processing workflows, 2) customized variants of standard workflows, and 3) algorithm development of new workflows. We create collaborative algorithm development environments that offer services typically restricted to NASA SDPSs such as data product provenance and bulk processing. We leverage existing NASA-funded information technologies such as the hybrid on-premise/ cloud science data system (HySDS), the Multi-mission Algorithm and Analysis Platform (MAAP), ECOSIS – a crowd-sourced spectral database, and ECOSML – a crowd-sourced model database. We demonstrate ImgSPEC on the Terrestrial Ecosystem use case processing through to foliar traits and fractional cover, thus aligning with driving thrusts for the SBG Science and Applications Communities.

Imaging Spectroscopy Processing Environment on the Cloud (ImgSPEC)



E. Natasha Stavros - 1, Philip A. Townsend - 2, George Chang -1, Hook Hua -1, Thomas Huang - 1, Namrata Malarout - 1, Justin Merz - 3, Winston Olson-Duvall - 1, William Phyo -1, Sujen Shah - 1, David R. Thompson - 1

1 - Jet Propulsion Laboratory, California Institute of Technology; 2 - University of Wisconsin- Madison; 3 - Independent Developer

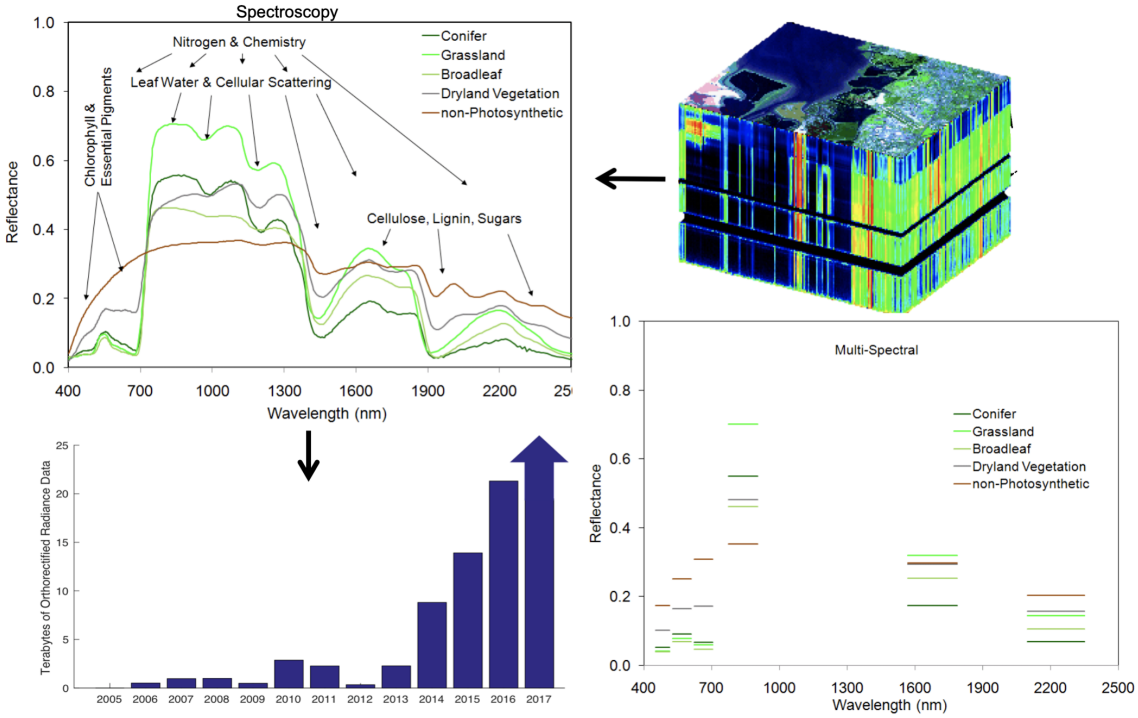
PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1–17 December 2020

BACKGROUND

Context

- The most recent Earth Science Decadal Survey was released by the National Academies in 2017
- 5 Designated Observables (DOs) – the most highly recommended Earth observables to change the current understanding of the Earth System
- Surface Biology and Geology (SBG) – open data access global imaging spectrometer at regular repeat observation
- "Chemical Fingerprint of the Atmosphere and Earth's Surface" = dozens of information value-added products



- Backwards compatible with a 30 year record = huge existing user base
- Barriers to use: Big data – volumes and processing

Objective

For the terrestrial ecology use case, demonstrate an end-to-end, on-demand, processing platform on the cloud for imaging spectroscopy Level 1 calibrated radiance data through Level 3+ information products

USER NEEDS ASSESSMENT

SBG Users

- Manager using data visualizations to inform decisions
- Educator/Student working on limited scope class projects
- ***Field scientist mostly interested in remote data at point locations***
- ***Researcher using remote sensing data as a tool in statistical and modeling experiments***
- ***Technician providing data processing support***
- ***Remote Sensing algorithm developer***
- Software developer generating tools and services of value

Disciplines

- Hydrological Cycles, Water Resources and Aquatic Ecosystems
- Weather and Air Quality
- Terrestrial Ecosystems and Natural Resource Management
- Climate variability and Change
- Earth Surface/ Geology

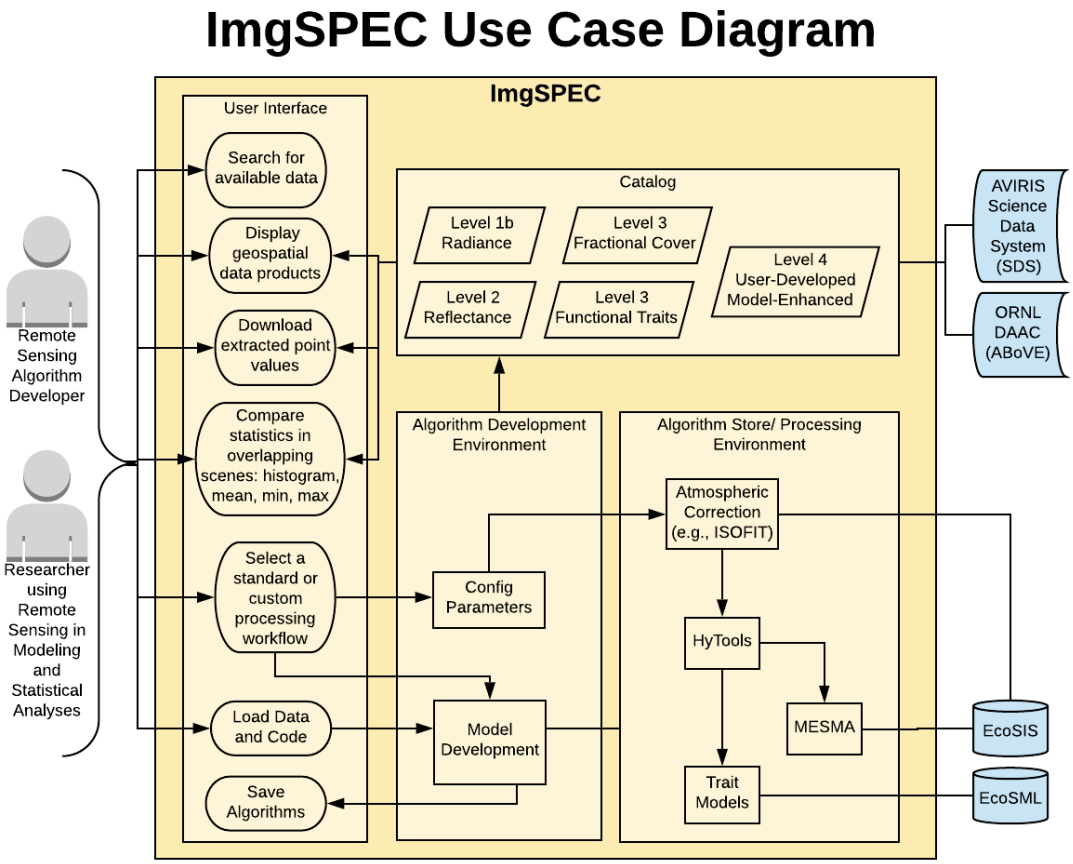
Method

1. Interviewed a Field Scientists, Researchers, Tecnicians and Remote Sensing Algorithm Developers in the Terrestrial Ecology Biodiversity discipline to understand user workflows and define the functional requirements of our system to meet their needs
2. Used a quantitative framework to help prioritize developments and functionality needs.

USE CASES

We test three use cases:

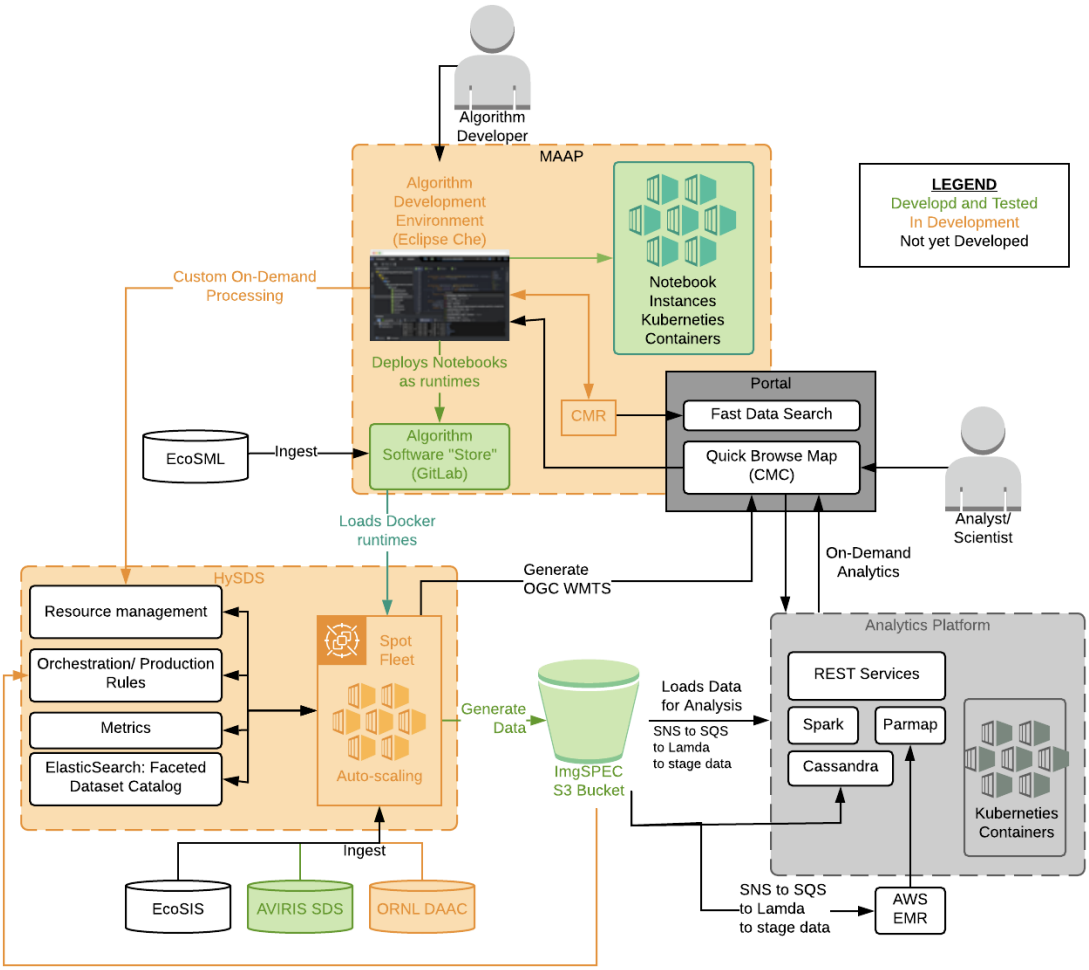
1. Standard data processing workflows
2. Customized variants of standard workflows, and
3. Algorithm development of new workflows



To demonstrate Return on Investment for Users to invest in using the software, ImgSPEC should:

- Reduce download times
- Enable easy provenance for reproducibility
- Provide scalable work environments that do not require a heavy lift from the science user

SYSTEM DESIGN



We create:

- Enable a collaborative algorithm development environment (Eclipse Che - Jupyter Hub, R Studio, etc)
- Offer services typically restricted to NASA Science Data Systems such as data product provenance and bulk processing

We leverage existing NASA-funded information technologies:

- Hybrid on-premise/ cloud Science Data System (HySDS)
- Multi-mission Algorithm and Analysis Platform (MAAP)
- ECOSIS – a crowd-sourced spectral database
- ECOSML – a crowd-sourced model database

NEXT STEPS

♬ Just keep building, just keep building ♬

Development Schedule

Test 1 (Sep 2020)

- Data Ingest
- Metadata repositories
- HyTools processing

Test 2 (Nov 2020)

- Deploy Algorithm Development Environment (MAAP) and data processing system (HySDS)
- Enable user authentication
- ISOFIT atmospheric correction processing

Test 3 (Jan 2021)

- Interface with EcoSIS

Test 4 (Feb 2021)

- Variant processing pathways
- Metadata for provenance

Test 5 (April 2021)

- Interface with EcoSML
- Trait and fractional cover processing

Test 6 (Oct 2021)

- Test custom, standard, and hybrid custom/stadard processing pathways

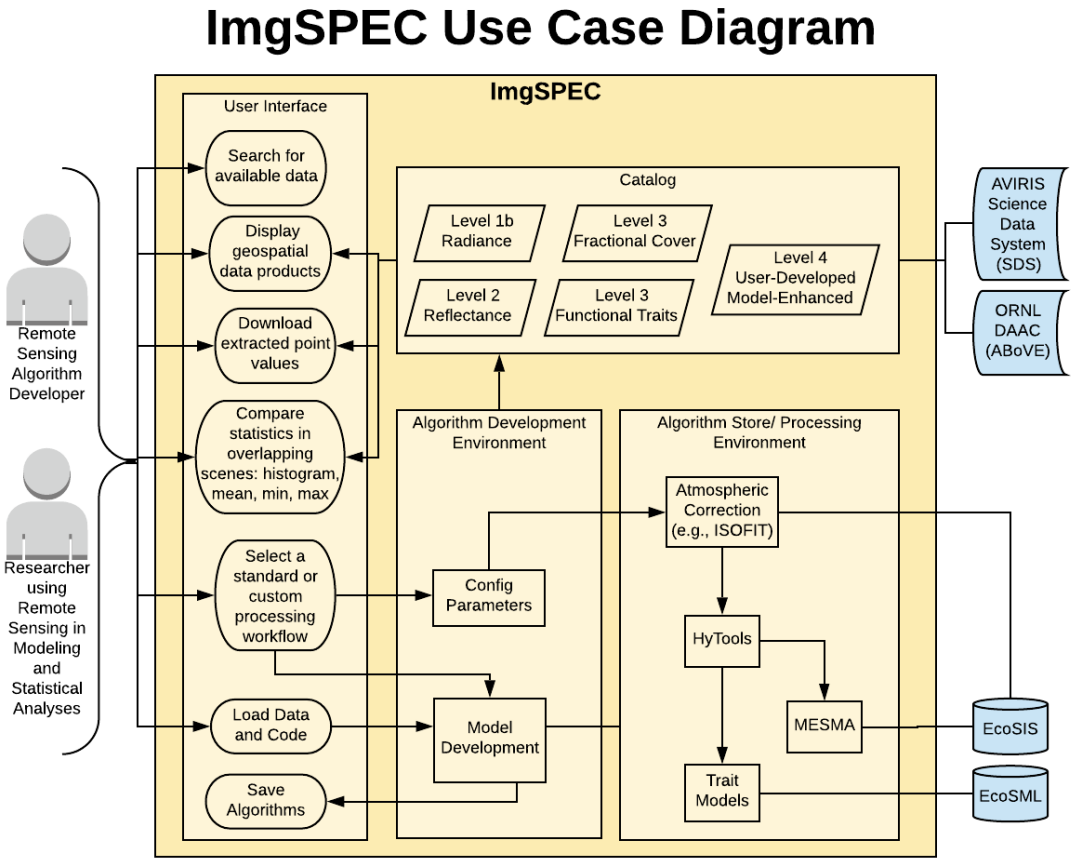
Test 7 (Dec 2021)

- Graphical User Interface
- Demonstrate Return on Investement

Public, open-source release (Feb 2022)

ABSTRACT

The geospatial Imaging Spectroscopy Processing Environment on the Cloud (ImgSPEC; formerly GeoSPEC) pioneers an on-demand science data processing system (SDPS) producing user-customized Level 1 calibrated radiance to Level 3+ data products in anticipation for the 2017-2027 Earth Decadal Survey prioritized spaceborne global imaging spectrometer to advance the study of Surface Biology and Geology (SBG). SBG data volumes (~20 TB/day) of high dimensionality (>224 bands) would be infeasible to download and the breadth of applications of the data across dozens of disciplines presents a need to evolve the traditional NASA SDPS. ImgSPEC streamlines processing data into key SBG observables that have demonstrated algorithms at local-to-regional scales and may vary locally. As such, a traditional, monolithic SDPS could not fully exploit the information in SBG measurements. To remove this barrier to use, ImgSPEC demonstrates an on-demand SDPS prototype that improves imaging spectroscopy data discovery, access, and utility enabling shared knowledge transfer from advanced imaging spectroscopy users to less experienced users such as decision makers and the general public. We test three use cases: 1) standard data processing workflows, 2) customized variants of standard workflows, and 3) algorithm development of new workflows. We create collaborative algorithm development environments that offer services typically restricted to NASA SDPSs such as data product provenance and bulk processing. We leverage existing NASA-funded information technologies such as the hybrid on-premise/ cloud science data system (HySDS), the Multi-mission Algorithm and Analysis Platform (MAAP), ECOSIS – a crowd-sourced spectral database, and ECOSML – a crowd-sourced model database. We demonstrate ImgSPEC on the Terrestrial Ecosystem use case processing through to foliar traits and fractional cover, thus aligning with driving thrusts for the SBG Science and Applications Communities.



(https://agu.confex.com/data/abstract/agu/fm20/7/2/Paper_674727_abstract_660235_0.png)