The Matsu Project

June 18, 2013

Robert L. Grossman

University of Chicago

Open Cloud Consortium

What is the Matsu Project?

Matsu is an open source project for processing satellite imagery to support earth sciences researchers using a community science cloud.

Matsu is a joint project between the Open Cloud Consortium and NASA's EO-1 Mission (Dan Mandl, Lead)

OSDC Console Apply Public Data Status Projects Support News/PIRE Open Science Data Cloud

OCC Project Matsu

An open source project for cloud-based processing of satellite imagery to support the earth sciences.

Project Matsu

Project Matsu is a collaboration between NASA and the Open Cloud Consortium to develop open source technology for cloud-based processing of satellite imagery to support the earth sciences. Technology developed by the collaboration include:

- The Namibia Flood Dashboard.
- MapReduced based analytics for identifying floods and CO_2 concentrations.
- Using elastic infrastructure-as-a-service to create Level 1 images each day.
- A Hadoop-based OGC-compliant tiling service and Web MapService.

You can learn more from these five minute videos introducing some Matsu tools.

Matsu Resources

- Daily Namibia Flood Dashboard
- Hadoop-supported Web Map Service Areas of Interest:
  - Carbon Monoxide cluster centers from volcanic eruption
  - Irrigation Patterns in the Sahara
  - Elevation
  - Water Classifier Namibia
  - Water Classifier Italian coast
- Available Matsu images

Matsu Support

- The code can be found at the OCC Github site.
- Description of Matsu Tile and WMS Services.

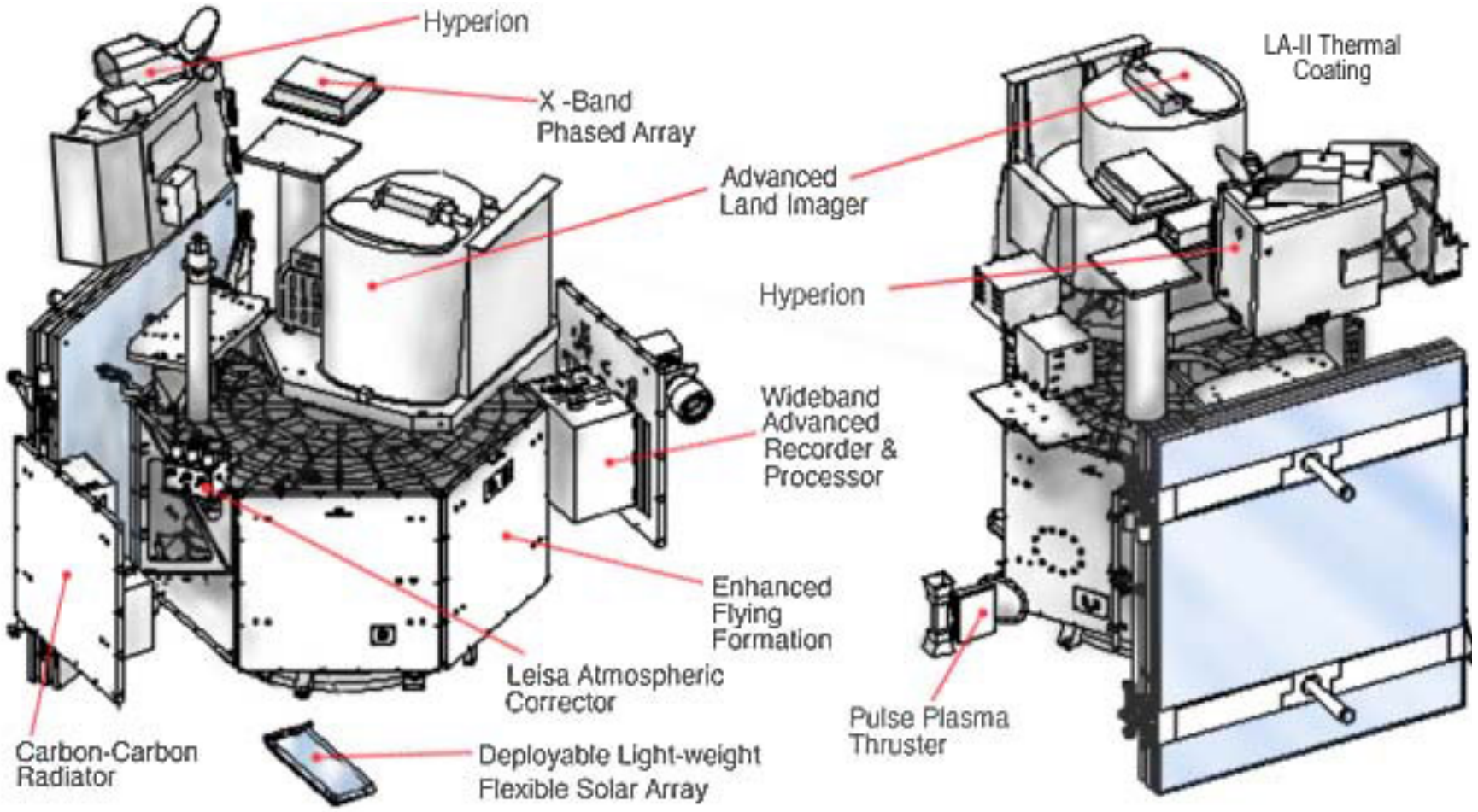
The OSDC is a resource of the Open Cloud Consortium and made possible by our sponsors.

matsu.opensciencedatacloud.org

EO-1 mission

- Approved in March 1996 and launched on November 21, 2000 from Vandenburg Air Force Base, California on a Delta 7320
- All technologies were flight-validated by December 2001
- EO-1 is now in an Extended Mission

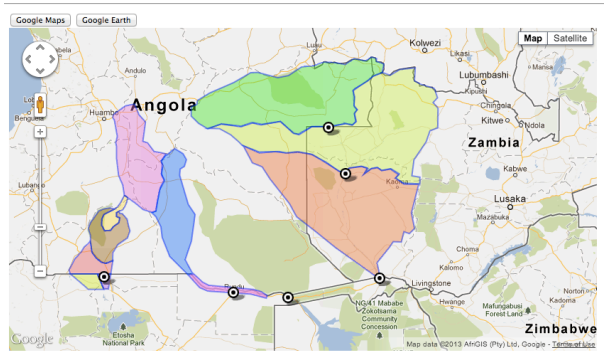
EO-1's ALI and Hyperion Instruments



Data - Instruments

- Hyperion Imaging Spectrometer
  - Designed to gather data from a given region on the Earth by viewing the surface in terms of 242 distinct 'bands' of light.
- Advanced Land Imager (ALI)
  - Used to validate and demonstrate technology for the Landsat Data Continuity Mission (LDCM)

All available L1G images (2010-now)

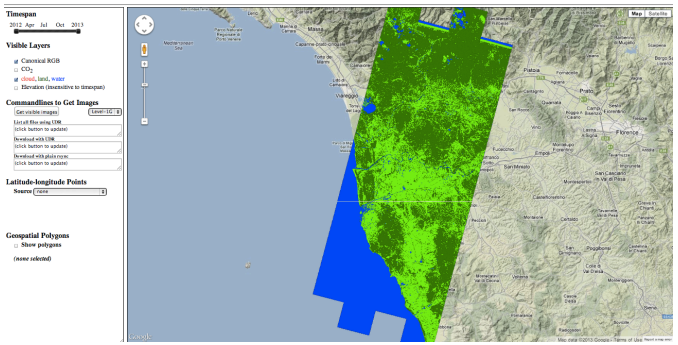


Legend:

ALI Flood Classification	Class 1 - Background:	Class 2 - Opaque Clouds:	Class 3 - Cloud Shadow:	Class 4 - Haze and Thin Clouds:	Class 5 - Clear Water:	Class 6 - Turbid Water:	Class 7 - Dry Land:

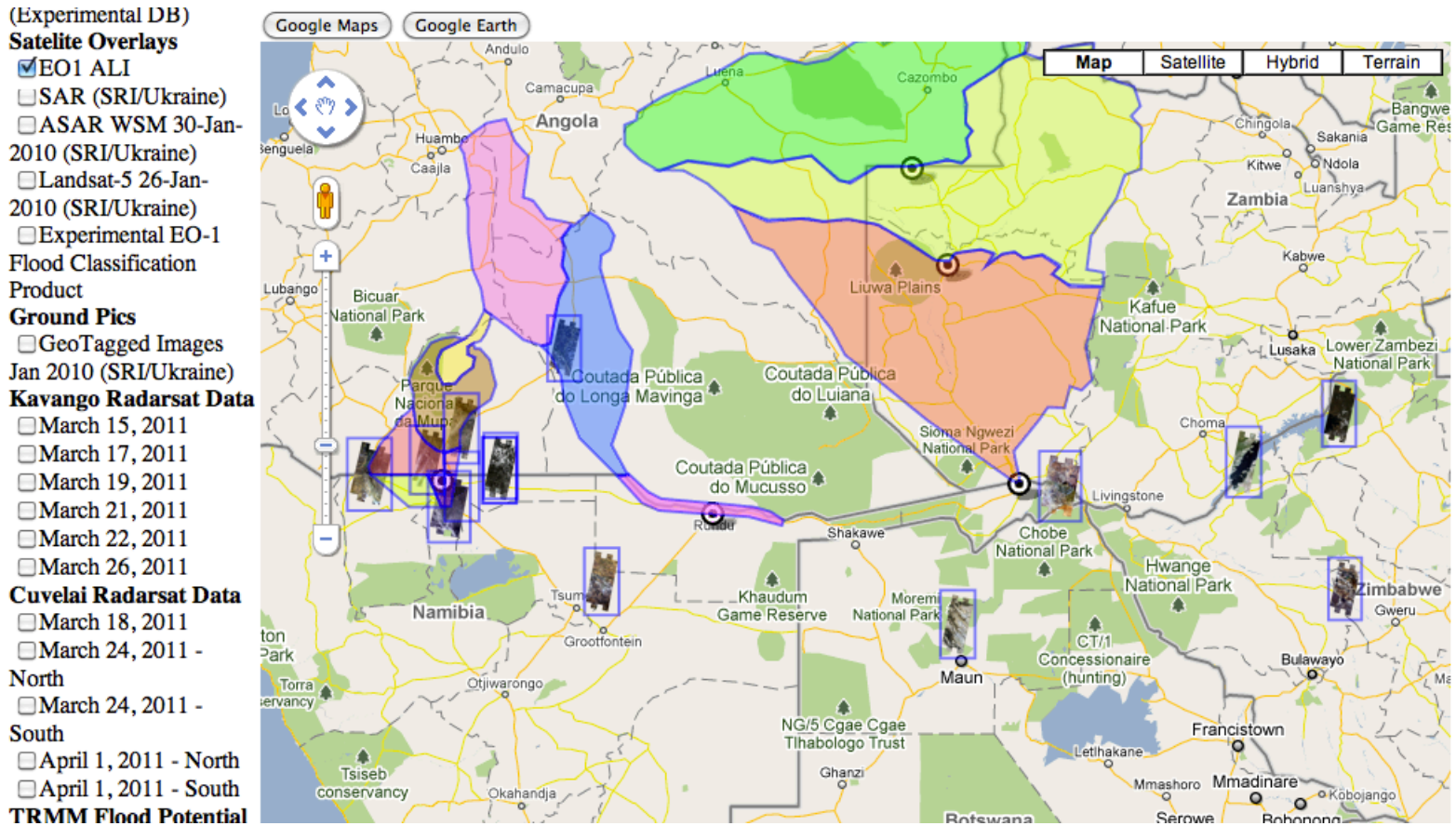
1. Open Science Data Cloud (OSDC) stores Level 0 data from EO-1 and uses an OpenStack-based cloud to create Level 1 data.

2. OSDC also provides OpenStack resources for the Nambia Flood Dashboard developed by Dan Mandl's team.



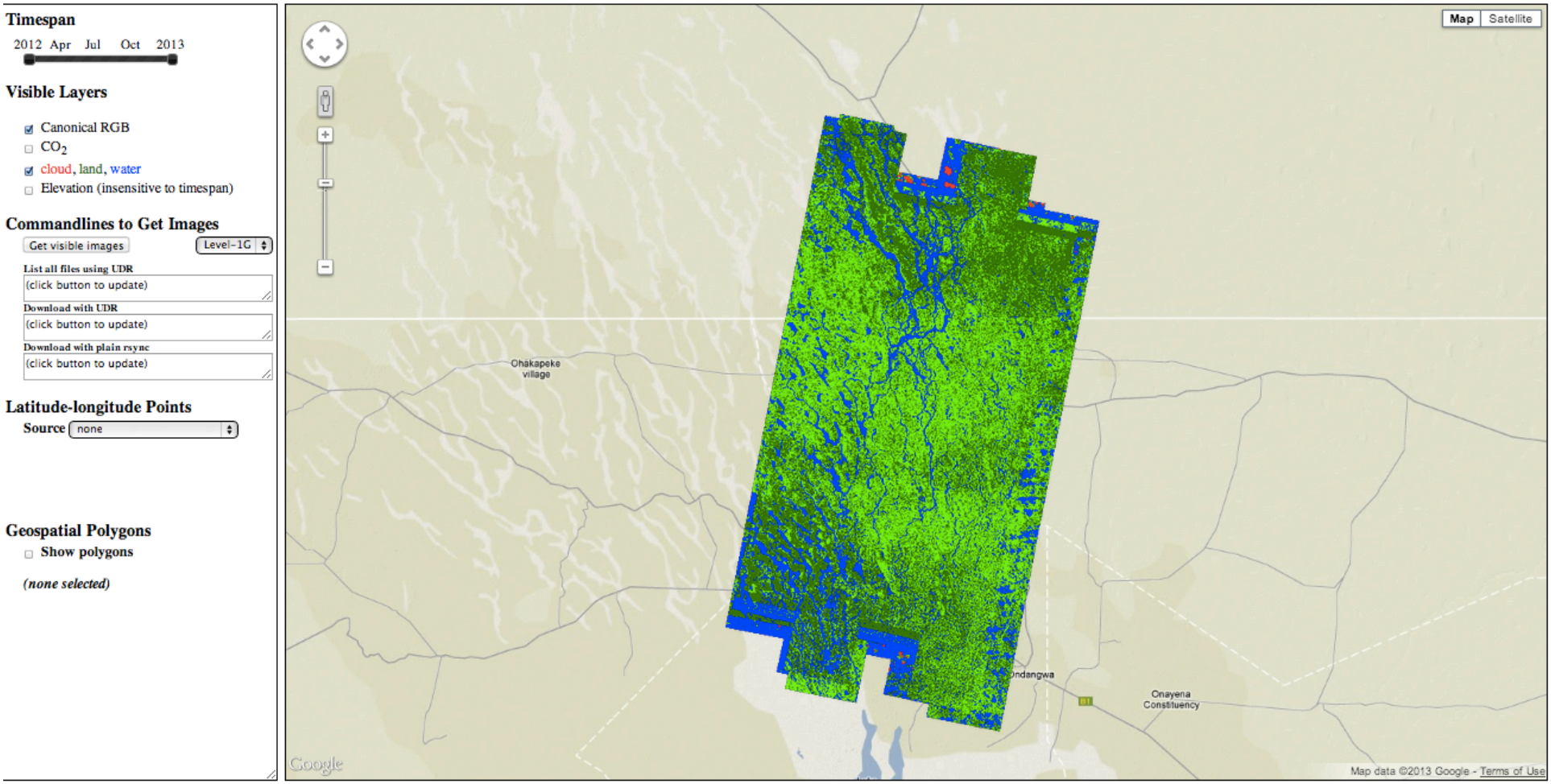
3. Project Matsu uses a Hadoop/Accumulo system to run analytics nightly and to create tiles with OGC-compliant WMTS.

NASA's Matsu Mashup

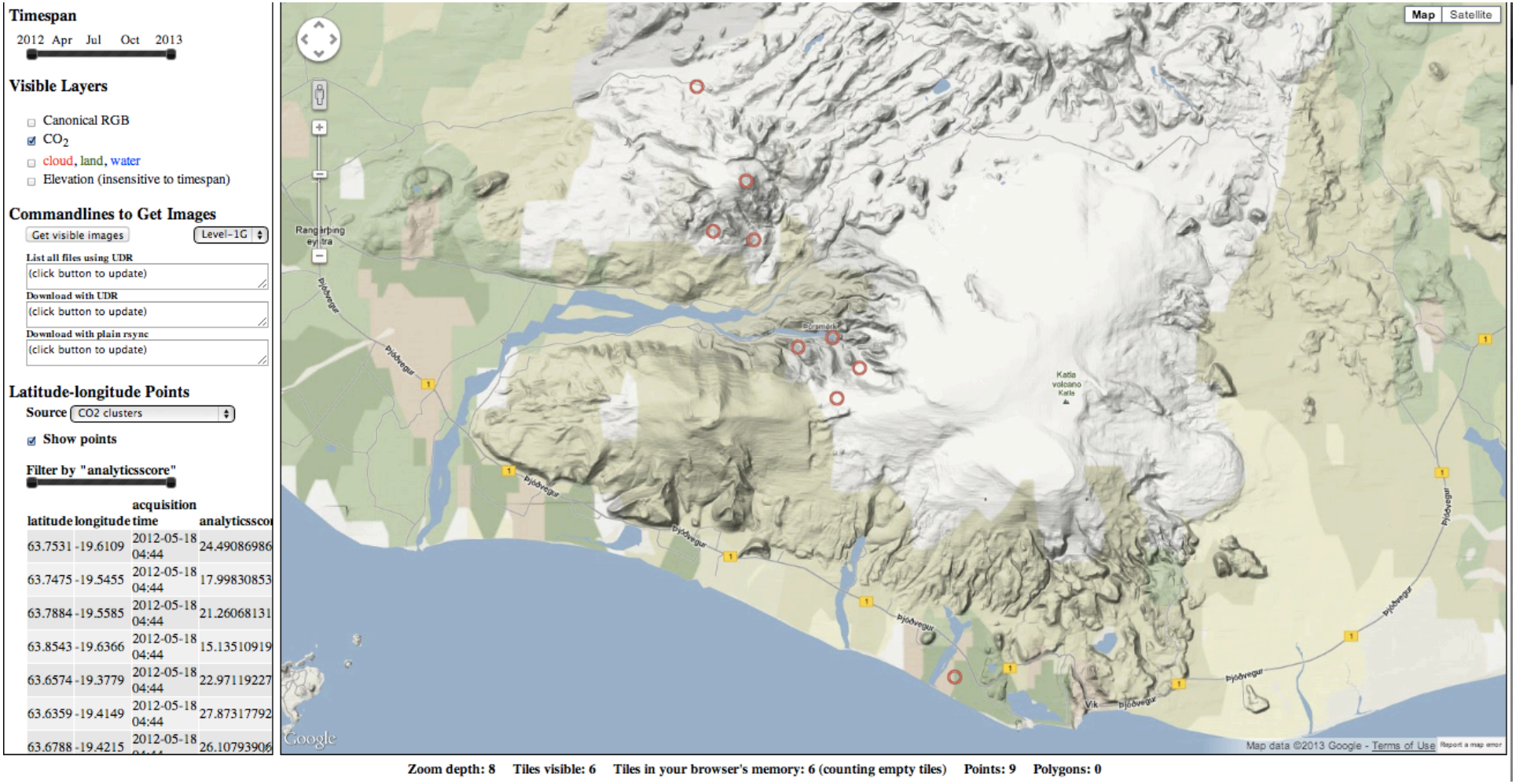


OSDC Satellites

- EO-1 (2012)
- Landsat7 – GLS 2000 (2013)
- MODIS (2013)
- TBD (2014)
- TBD (2015)

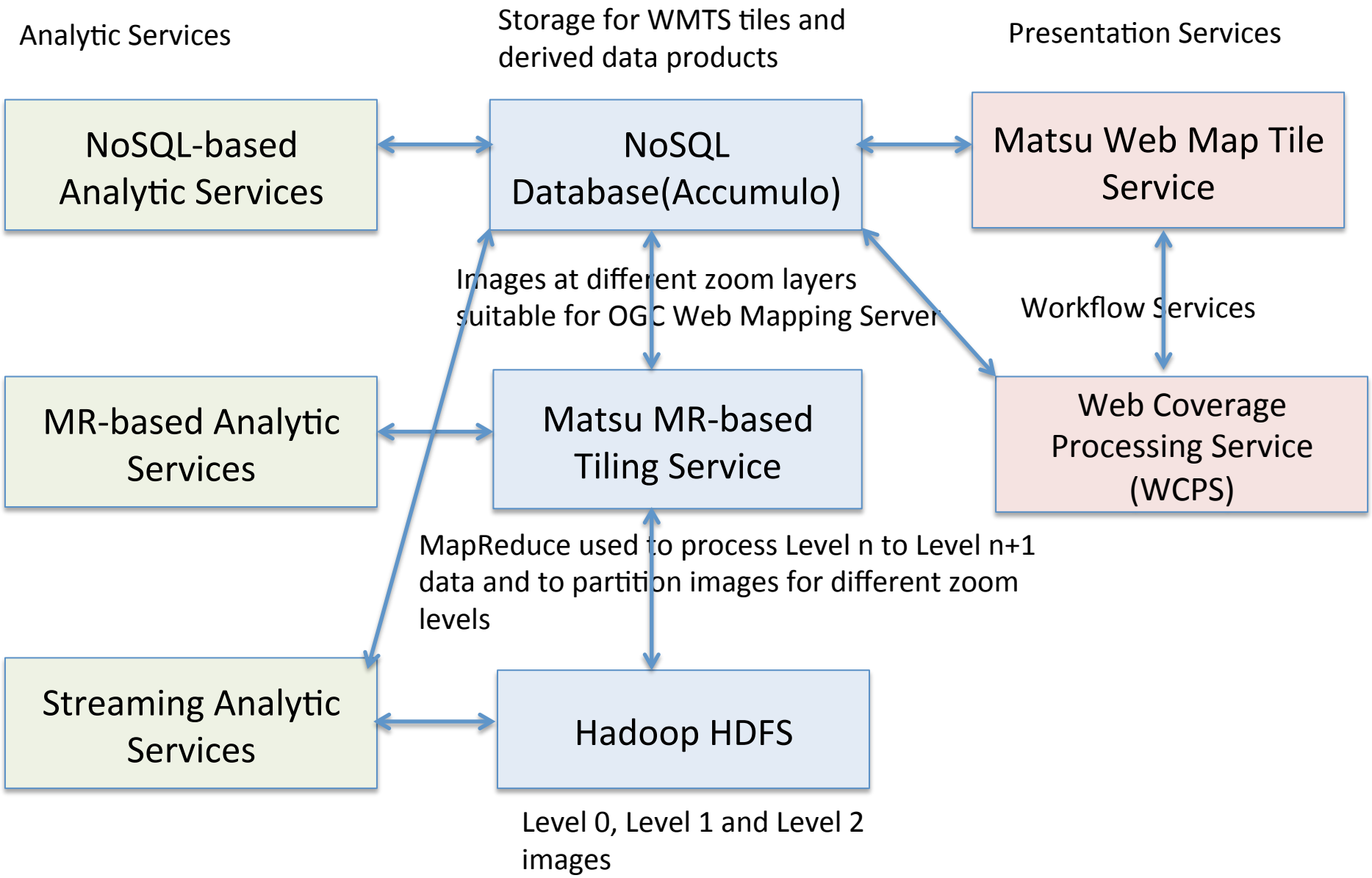


Matsu Web Map Tile Service



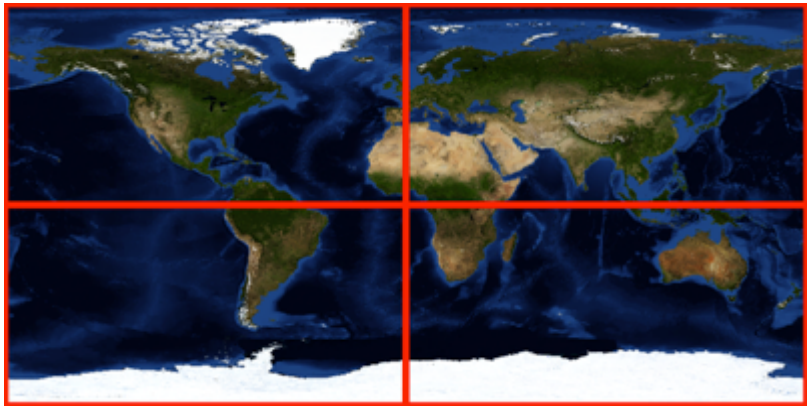
It is easy to layer analytics over the Web Map Tile Service (WMTS). Here is one identifying CO_2

Matsu Hadoop Architecture

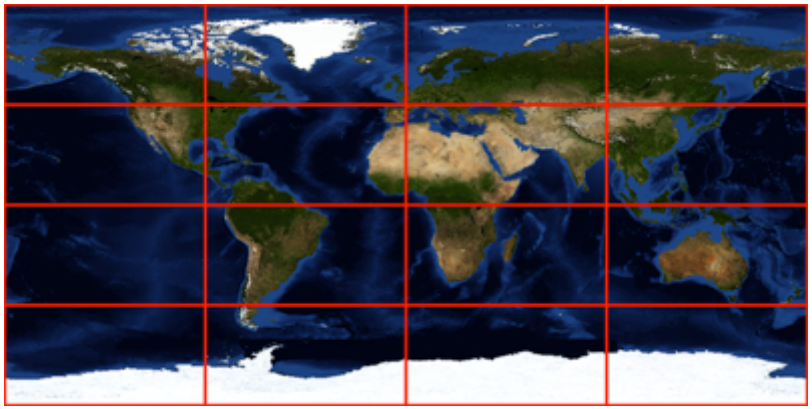


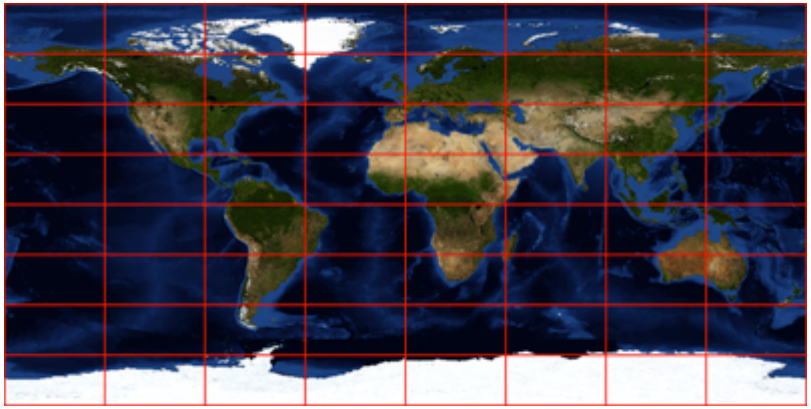
Zoom Levels

Zoom Level 1: 4 images

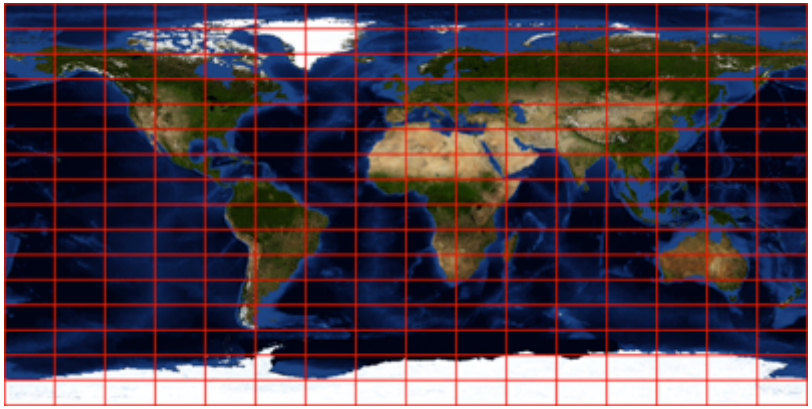


Zoom Level 2: 16 images





Zoom Level 3: 64 images



Zoom Level 4: 256 images

Build Tile Cache: Map

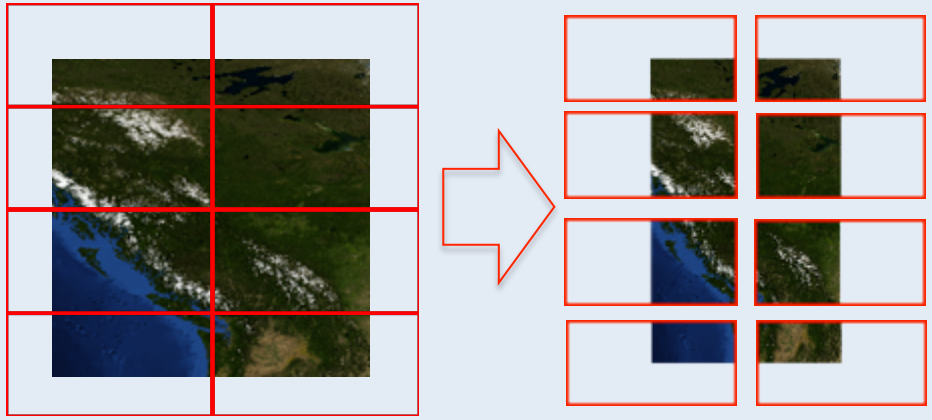
Mapper Input Key: Bounding Box

(minx = -135.0 miny = 45.0 maxx = -112.5 maxy = 67.5)

Mapper Input Value:

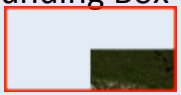
Step 1: Input to Mapper

Mapper resizes and/or cuts up the original image into pieces to output Bounding Boxes

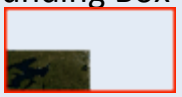


Step 2: Processing in Mapper

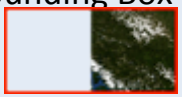
Mapper Output Key: Bounding Box
Mapper Output Value:



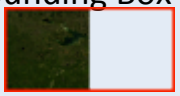
Mapper Output Key: Bounding Box
Mapper Output Value:



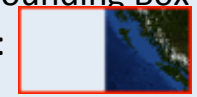
Mapper Output Key: Bounding Box
Mapper Output Value:



Mapper Output Key: Bounding Box
Mapper Output Value:

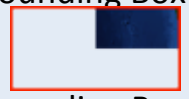


Mapper Output Key: Bounding Box
Mapper Output Value:

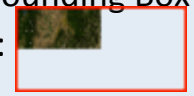


Mapper Output Key: Bounding Box
Mapper Output Value:

Mapper Output Key: Bounding Box
Mapper Output Value:



Mapper Output Key: Bounding Box
Mapper Output Value:



Step 3: Mapper Output

Build Tile Cache: Reduce

Reducer Key Input: Bounding Box
(minx = -45.0 miny = -2.8125 maxx = -43.59375 maxy = -2.109375)
Reducer Value Input:

...

Step 1: Input to Reducer

Assemble Images based on bounding box

Step 2: Reducer Output

- Reducer assembles tiles at each zoom level
- Tiles written to Accumulo (a NoSQL database)

Map Phase

- Map
  - Read in images by Bands, Date, and Region
  - Fix a zoom level for sending to reducers
    - Based on number of reducers and processing power, ***not on*** the zoom you want for display
  - Emit as <key>, <value>
    - Key = <Bounding Box at Fixed Zoom Level>
    - Value = <Bounding Bounding Box at Smallest Zoom Level, Bands, Projection, Timestamp, Image Bytes>

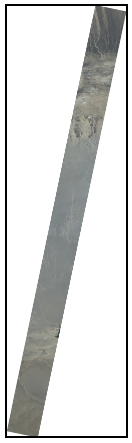
Reduce Phase

- All bytes for bands and satellite strips in this bounding box are mapped to the same reducer
- The key can be identified by the Lat/Long of the upper right corner of the box

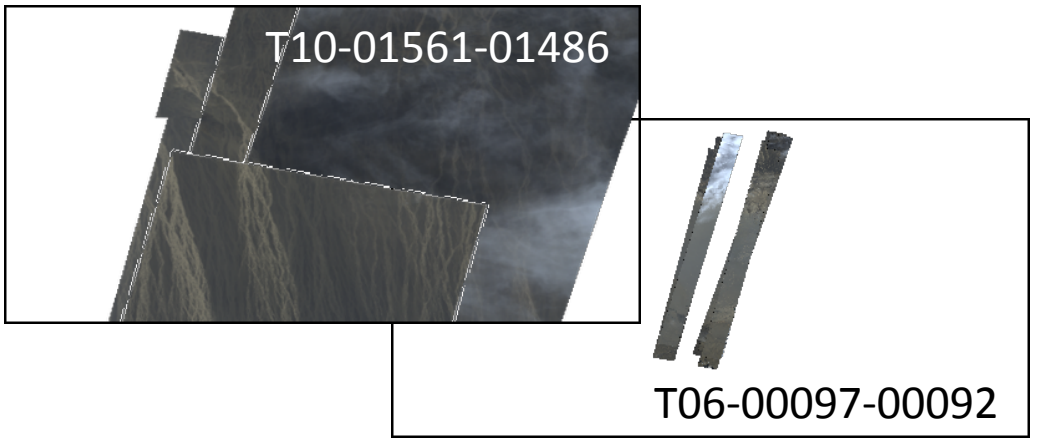
Level 1 Images - Details

location in Google Maps

L1R



L1G



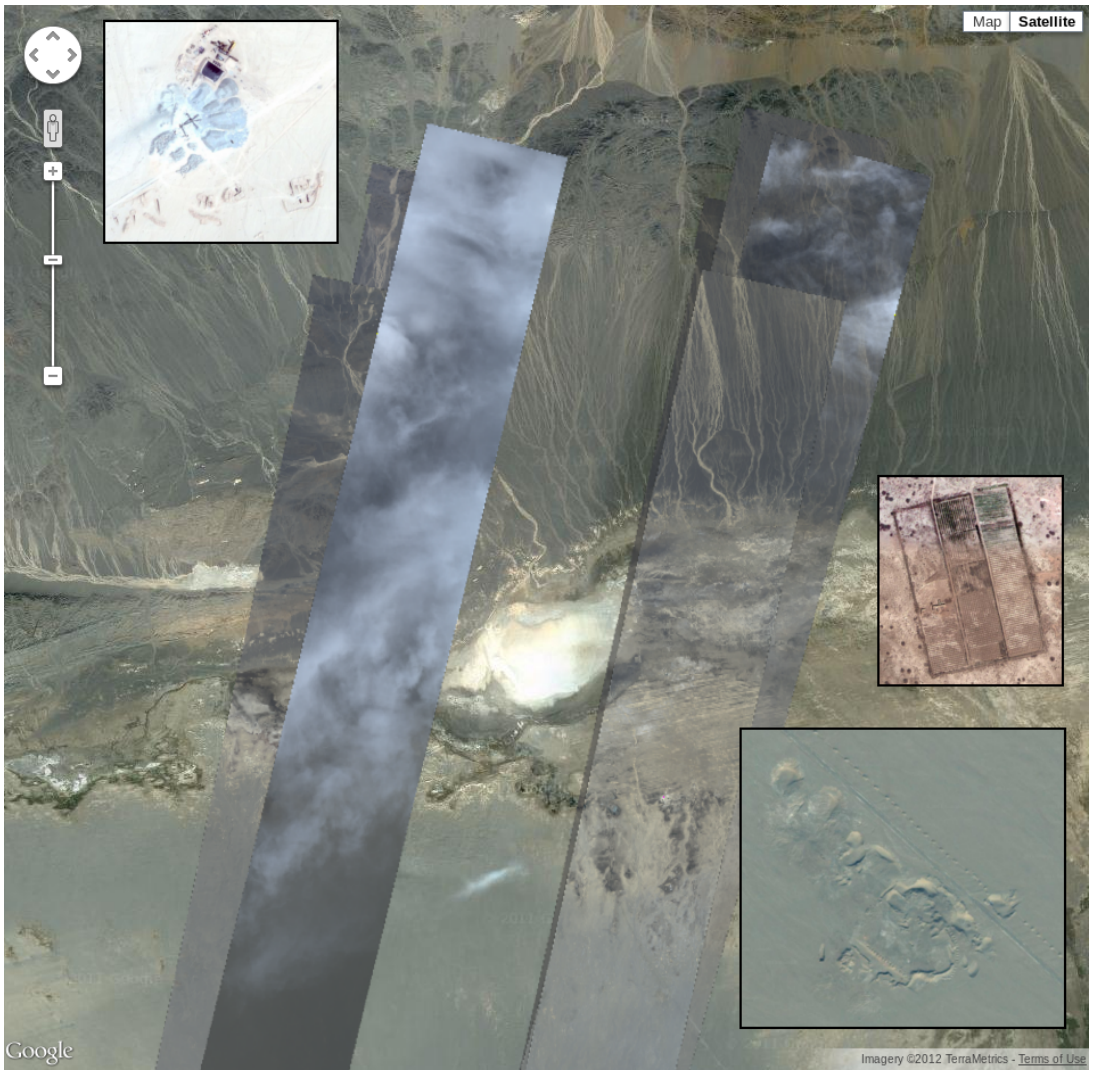
Level-2 tiles made in Map-Reduce, prepared for WMS

- Satellite track images (L1R) are rotated and geolocated (L1G) by NASA
- We overlay L1G images into Level-2 dyadic tiles in Map-Reduce

Some example images

Gobi Desert

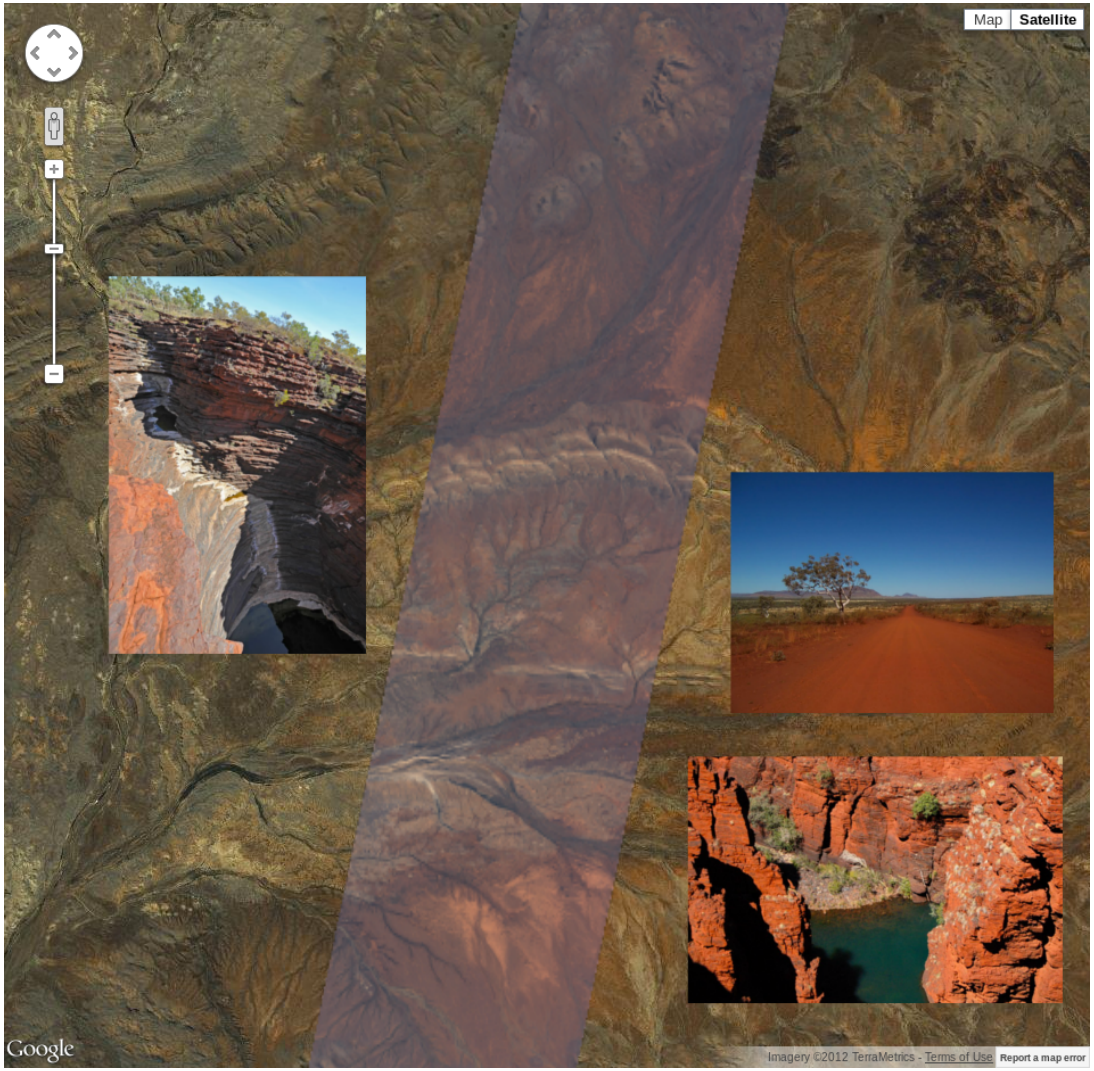
- same as previous page
- contains some strange structures that are too small to spatially resolve with Hyperion, but they might have interesting spectral features



Some example images

Karijini, Australia

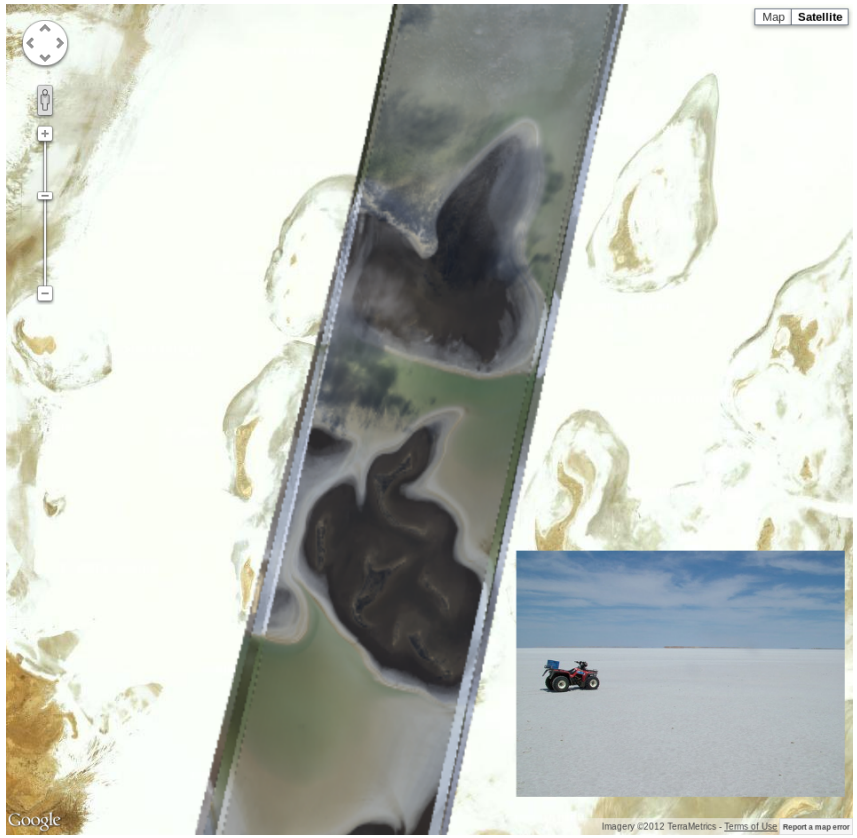
- lots of colorful minerals
- should have a very rich spectrum



Some example images

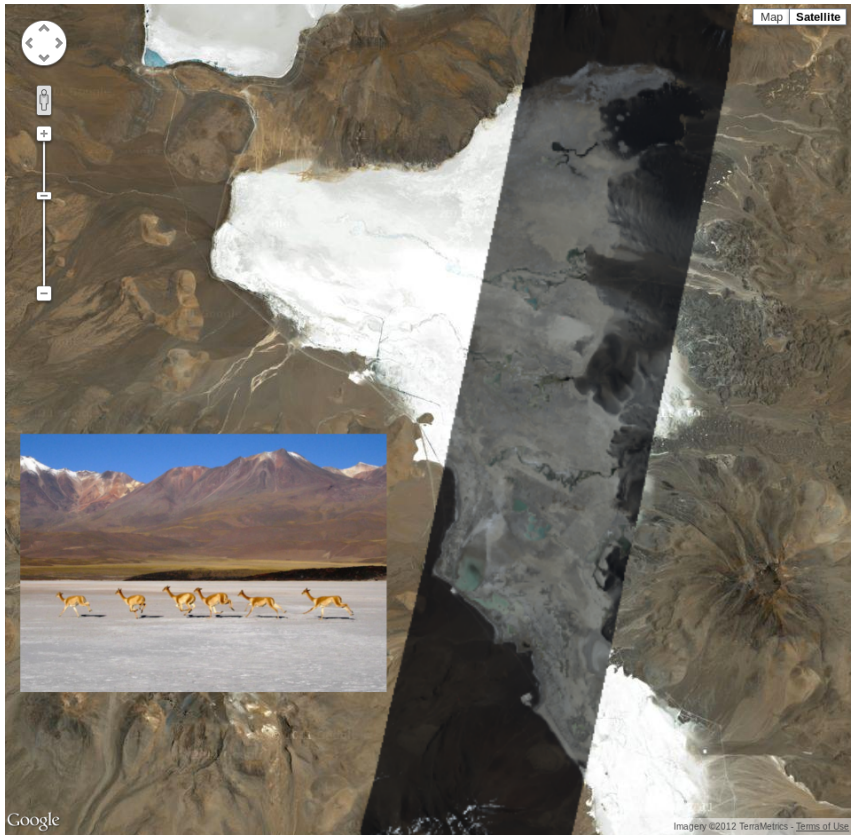
Lake Frome, Australia

- salt bed is a standard calibration target



Atacama Desert, Chile

- salt bed in the driest part of the world



Algebraic combination of spectral bands to make a more sensitive image

- CO_2 has three absorbtion lines within Hyperion's spectral range

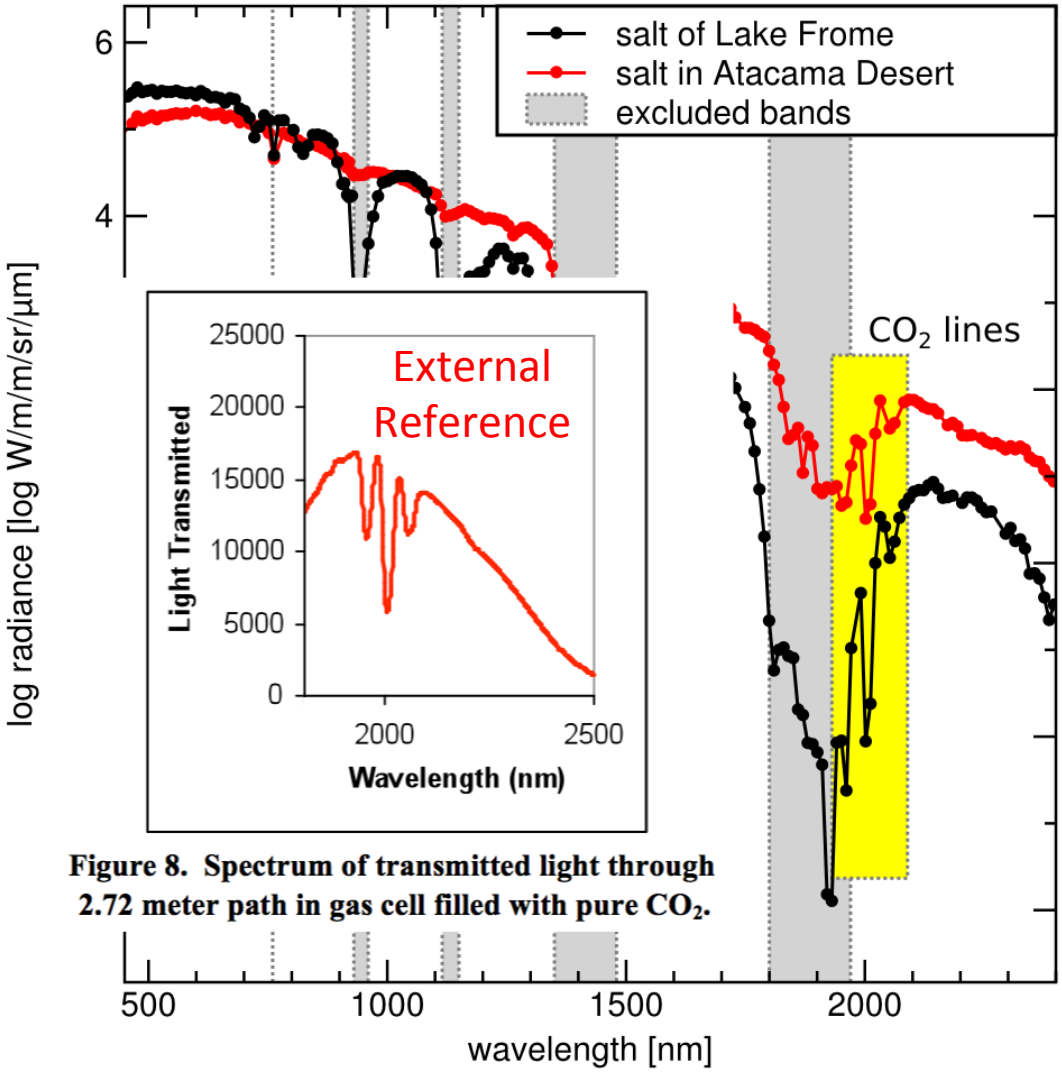
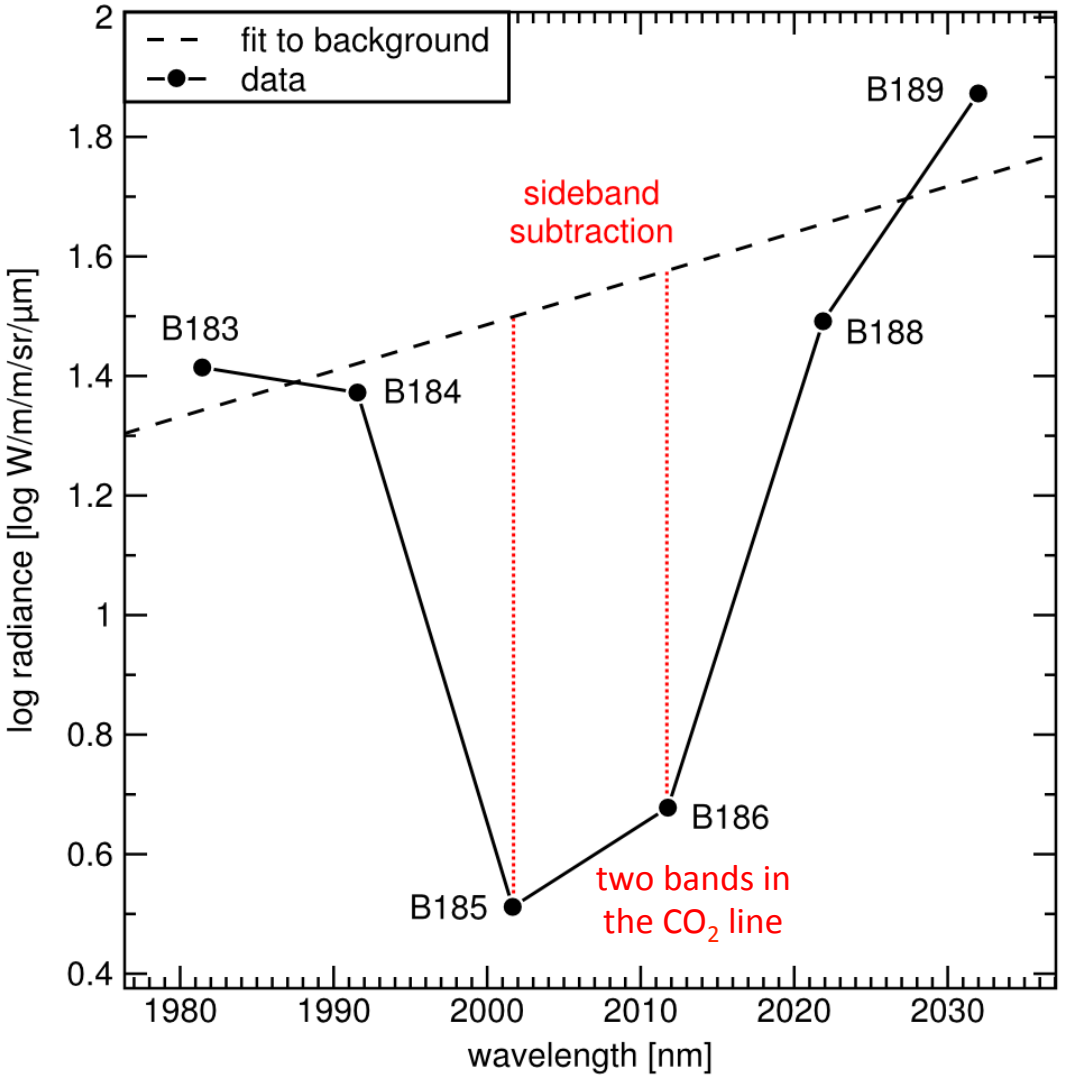


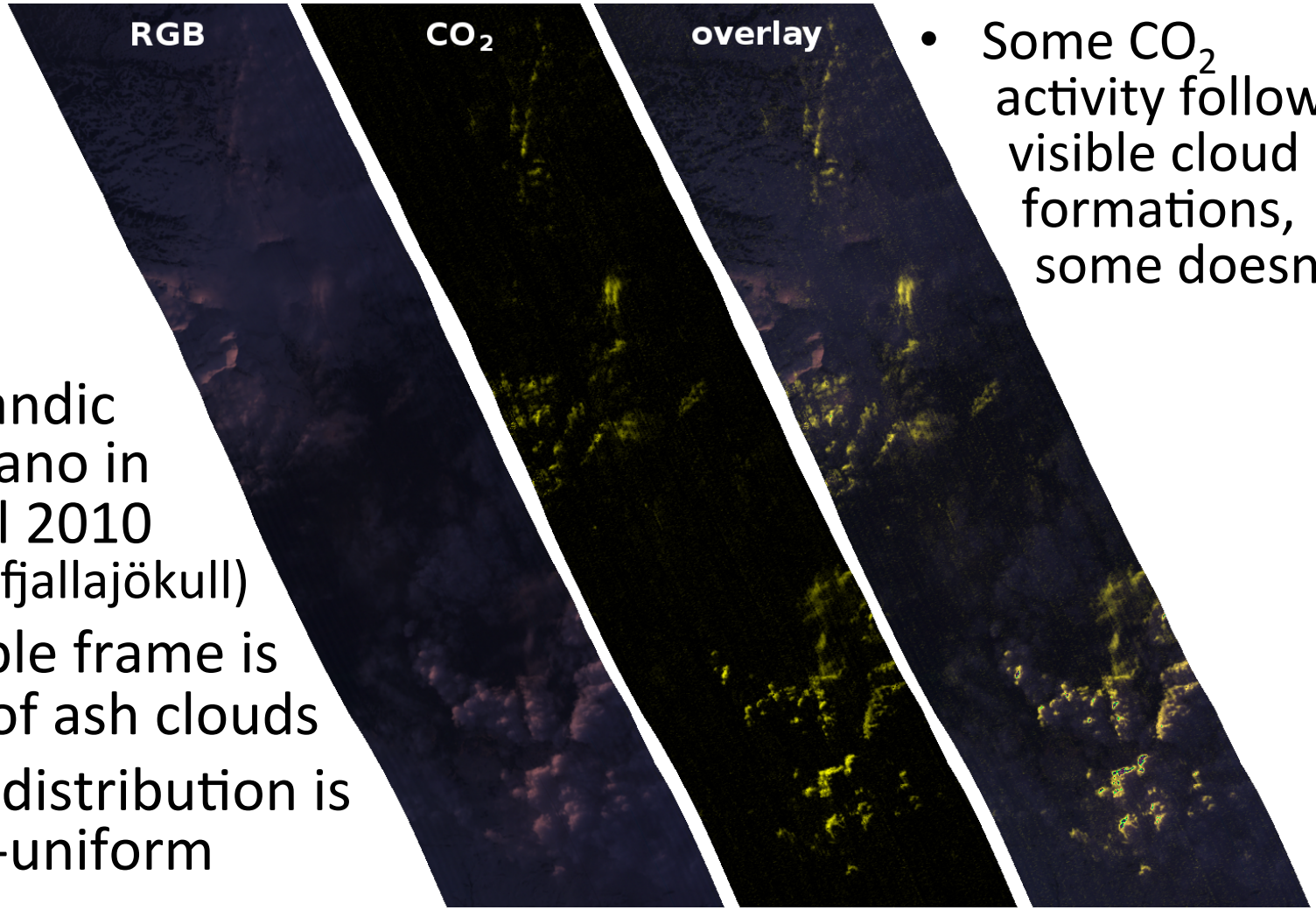
Figure 8. Spectrum of transmitted light through 2.72 meter path in gas cell filled with pure CO_2.

Algebraic combination of spectral bands to make a more sensitive image

- CO_2 has three absorbtion lines within Hyperion's spectral range
- Sideband subtraction technique extracts a pure sample of data in a peak by fitting nearby datapoints to a curve and subtracting peak values from the curve
- In this case, we invert the subtraction because it's an anti-peak



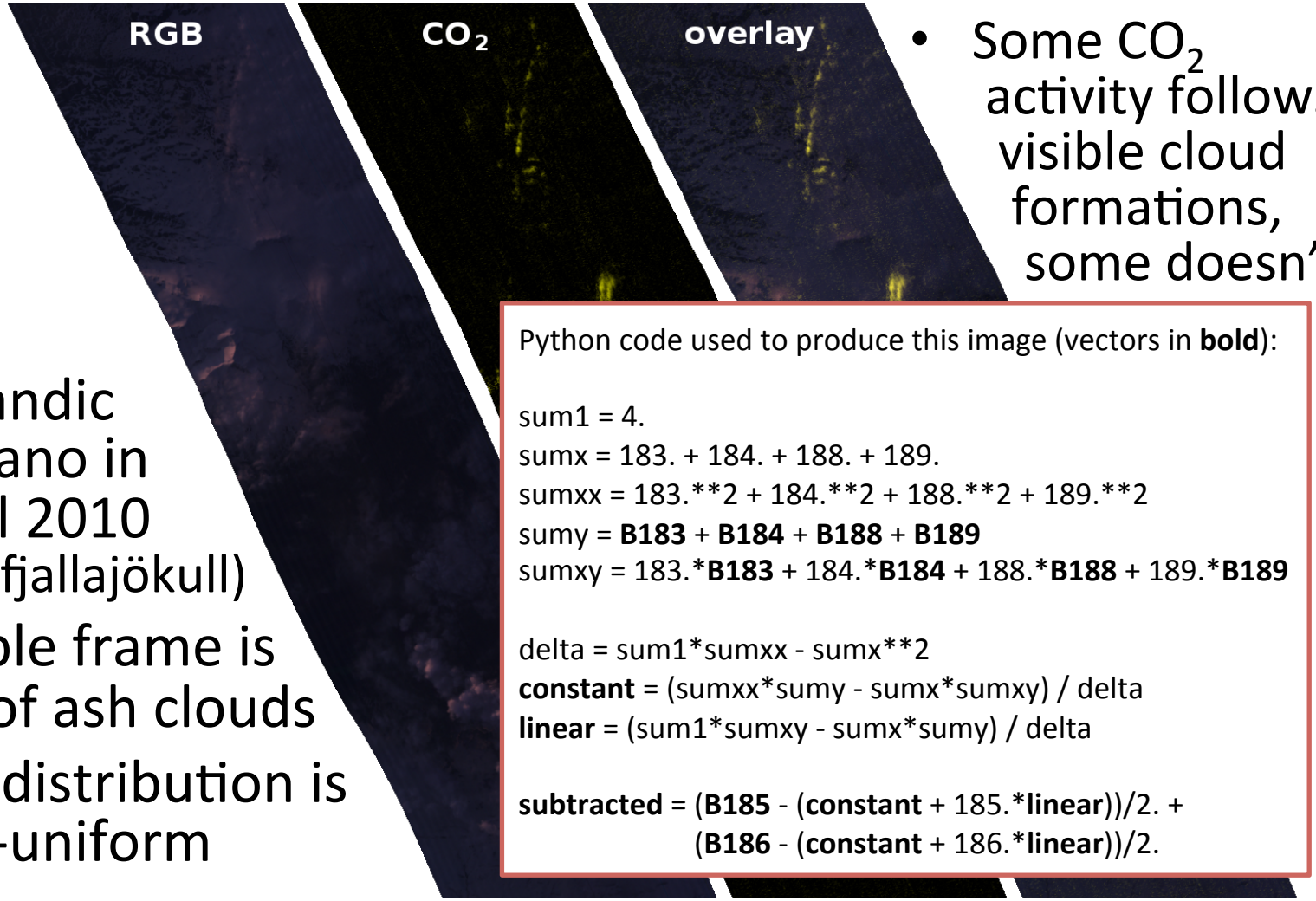
Algebraic combination of spectral bands to make a more sensitive image



- Some CO_2 activity follows visible cloud formations, some doesn't

- Icelandic volcano in April 2010 (Eyjafjallajökull)
- Visible frame is full of ash clouds
- CO_2 distribution is non-uniform

Algebraic combination of spectral bands to make a more sensitive image



- Some CO_2 activity follows visible cloud formations, some doesn't

- Icelandic volcano in April 2010 (Eyjafjallajökull)
- Visible frame is full of ash clouds
- CO_2 distribution is non-uniform

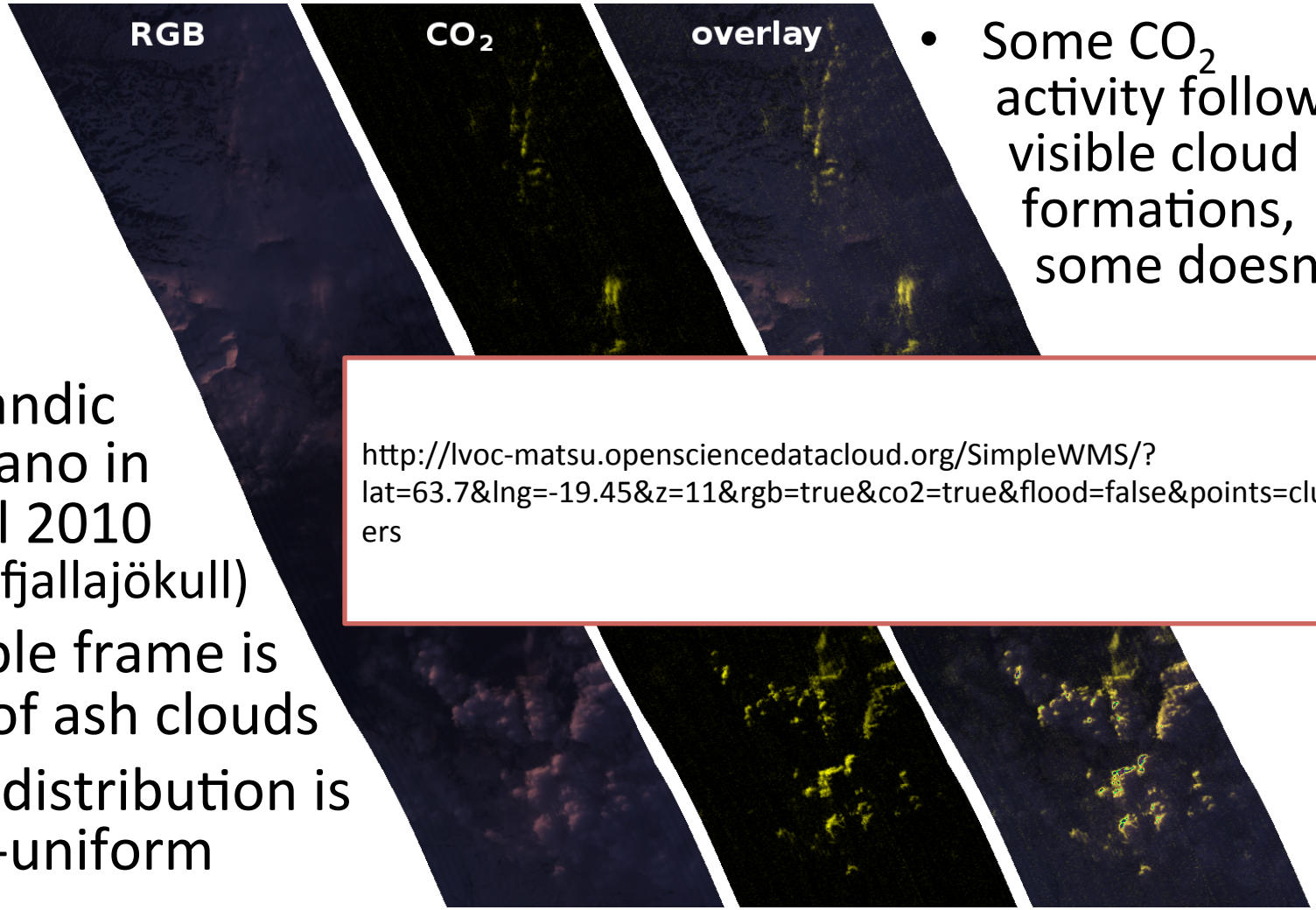
Python code used to produce this image (vectors in **bold**):

```
sum1 = 4.
sumx = 183. + 184. + 188. + 189.
sumxx = 183.**2 + 184.**2 + 188.**2 + 189.**2
sumy = B183 + B184 + B188 + B189
sumxy = 183.*B183 + 184.*B184 + 188.*B188 + 189.*B189

delta = sum1*sumxx - sumx**2
constant = (sumxx*sumy - sumx*sumxy) / delta
linear = (sum1*sumxy - sumx*sumy) / delta

subtracted = (B185 - (constant + 185.*linear))/2. +
             (B186 - (constant + 186.*linear))/2.
```

Algebraic combination of spectral bands to make a more sensitive image



- Some CO_2 activity follows visible cloud formations, some doesn't

- Icelandic volcano in April 2010 (Eyjafjallajökull)
- Visible frame is full of ash clouds
- CO_2 distribution is non-uniform

http://lvoc-matsu.opensciencedatacloud.org/SimpleWMS/?lat=63.7&lng=-19.45&z=11&rgb=true&co2=true&flood=false&points=clusters

Questions

For More Information

- Project Matsu is managed and operated by the Open Cloud Consortium (www.opencloudconsortium.org).
- Project Matsu is supported in part by grants from Gordon and Betty Moore Foundation and the National Science Foundation (Grants OISE - 1129076 and CISE 1127316).
- For more information about Project Matsu, please see the Project Matsu website: matsu.opensciencedatacloud.org
- The Project Director is Robert Grossman, who can be reached at

robert.grossman@uchicago.edu

Here is some detail of how we process EO-1 satellite imagery data using Hadoop in Project Matsu…

Step 1 – Storage & Archiving

From Space to Goddard to the OSDC

1. Transmit data from NASA's EO-1 Satellite to NASA ground stations and then to NASA Goddard
2. At Goddard, align data, perform radiometric corrections and generate Level 0 images (16-bit radiance values)
3. Transmit Level 0 data from NASA Goddard to the OCC's Open Science Data Cloud (OSDC)
4. Store images in a distributed, fault tolerate, file system

Step 2 – Creating Level 1 Images

Building Level 1 Images on the OSDC

1. Each day, the new Level 0 images stored on the OSDC are processed
2. Within the OSDC, NASA launches Virtual Machines (VMs) specifically built to render Level 1 images from Level 0 data.
   - Each Level 1 band is saved as a distinct image
3. Level 1 bands are written to storage facility in the OSCD for long-term public access

Step 3 – Tiling

Matsu Processing

1. Build Web Mapping Tile Service Tiles from Level 1 images using MapReduce
2. Store tiles in Accumulo
   - Index them so that they are accessible via Web Mapping Service
3. Run Analytics on Level 1 images
   - Move results of the analytics to Accumulo

Tiling - Detail

- Use MapReduce to build Web Tiles
  1. Each day, the Level 1 images created by NASA and stored on the OSDC are processed
  2. The Date and Bands (to create a visible image) are specified
  3. Run MapReduce Job
     1. Map – FILL-IN
     2. Partition – FILL-IN
     3. Reduce – FILL-IN

Tile Details, cont'd

- Images are handles as byte streams
- Divide (chunk) the Level 1 images into manageable sizes.
- Dyadic decomposition
  - Divide each image into 4 equal size pieces
  - For each additional zoom, subdivide each piece into 4 equal size pieces
- Tag each chunked images with the bounding box, date, time, dyadic level and bands.
- Convert the bytes into PNG files

Processing the Data

- Reduce
  - Once all images are received for a Bounding Box, sort by the most granular zoom level
  - Process that Zoom Level
  - Once a zoom level in is completed, combine images and scale the build the next zoom level

Accumulo Storage

- Images are stored by Bounding Box
  - -180.0_-90.0_180.0_90.0
- Column family
  - The tile style, zoom, and projection
- Column qualifier
  - Dimensions (width and height, 512 x 256)
- Value
  - the corresponding PNG image in raw bytes

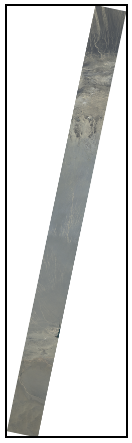
Serve to WMTS

- The WMTS query:
  - Bounding Box
  - Date
  - Layer name as a string
    - Haiti
  - Style name as a string
    - The bands used to build the Level 1 image or an alias: “B058:B023:B015” or “agricultural”
- Not supported
  - Map Project could be used, but for now, we only support a single projection

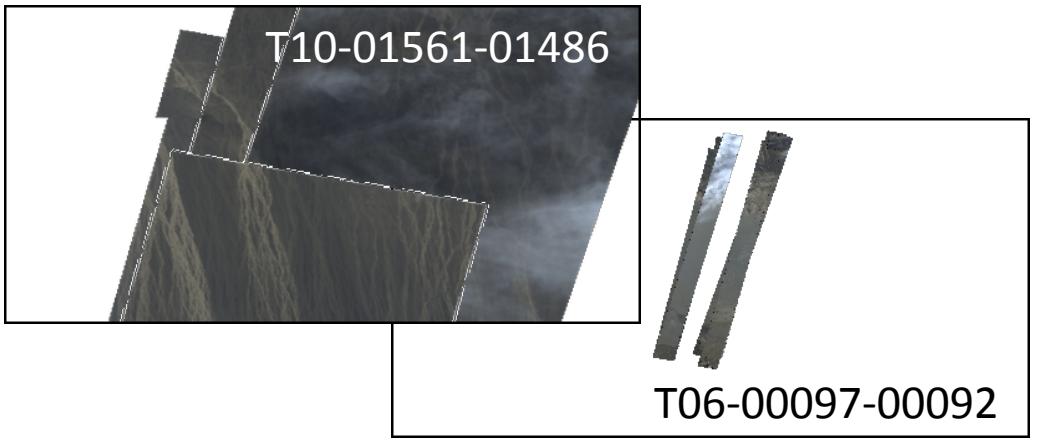
Images: stages of processing

image locations (viewed in Google Maps)

L1R



L1G



Level-2 tiles made in Map-Reduce, prepared for WMS

- Satellite track images (L1R) are rotated and geolocated (L1G) by NASA
- We overlay L1G images into Level-2 dyadic tiles using Map-Reduce