Ingestion of KOMPSAT Satellite Data in Open Data Cube

Kiwon LEE

Seoul, Korea
Korea Multi-Purpose Satellite (KOMPSAT) series

KOMPSAT 3
Specifications:

<table>
<thead>
<tr>
<th>Product Resolution</th>
<th>Panchromatic: 0.5m, (0.7m GSD)</th>
<th>Multispectral: 2.0m @ altitude 685.13km nadir, (2.8m GSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral bands:</td>
<td>PAN: 450 – 900 nm</td>
<td>MS1 (B): 450 – 520 nm</td>
</tr>
<tr>
<td></td>
<td>MS2 (G): 520 – 600 nm</td>
<td>MS3 (R): 630 – 690 nm</td>
</tr>
<tr>
<td></td>
<td>MS4 (N): 760 – 900 nm</td>
<td></td>
</tr>
<tr>
<td>Swath width:</td>
<td>16km</td>
<td></td>
</tr>
<tr>
<td>Imaging Modes:</td>
<td>Strip, Multipoint, Stereo, Wide area</td>
<td></td>
</tr>
<tr>
<td>Orbit:</td>
<td>Sun Synchronous orbit</td>
<td></td>
</tr>
<tr>
<td>Altitude:</td>
<td>685.13km</td>
<td></td>
</tr>
<tr>
<td>Inclination:</td>
<td>98.14 deg</td>
<td></td>
</tr>
<tr>
<td>MLTAN:</td>
<td>13:30 (local time)</td>
<td></td>
</tr>
<tr>
<td>Map projection:</td>
<td>UTM/WGS84</td>
<td></td>
</tr>
</tbody>
</table>

KOMPSAT 3A
Specifications:

<table>
<thead>
<tr>
<th>Product Resolution</th>
<th>Panchromatic: 0.4m, (0.54m GSD)</th>
<th>Multispectral: 16m @ altitude 528km nadir, (2.16m GSD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spectral bands:</td>
<td>PAN: 450 – 900 nm</td>
<td>MS1 (B): 450 – 520 nm</td>
</tr>
<tr>
<td></td>
<td>MS2 (G): 520 – 600 nm</td>
<td>MS3 (R): 630 – 690 nm</td>
</tr>
<tr>
<td></td>
<td>MS4 (N): 760 – 900 nm</td>
<td></td>
</tr>
<tr>
<td>Swath width:</td>
<td>13km</td>
<td></td>
</tr>
<tr>
<td>Imaging Modes:</td>
<td>Strip, Multipoint, Stereo, Wide area</td>
<td></td>
</tr>
<tr>
<td>Orbit:</td>
<td>Sun Synchronous orbit</td>
<td></td>
</tr>
<tr>
<td>Altitude:</td>
<td>528km</td>
<td></td>
</tr>
<tr>
<td>Inclination:</td>
<td>97.5 deg</td>
<td></td>
</tr>
<tr>
<td>MLTAN:</td>
<td>13:30 (local time)</td>
<td></td>
</tr>
<tr>
<td>Map projection/Datum</td>
<td>UTM/WGS84</td>
<td></td>
</tr>
</tbody>
</table>

KOMPSAT 5
Specifications:

<table>
<thead>
<tr>
<th>Sensor Type:</th>
<th>SAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAP Projection/Datum</td>
<td>UTM/WGS84</td>
</tr>
<tr>
<td>Orbit:</td>
<td>Dawn-Dusk orbit</td>
</tr>
<tr>
<td>Imaging Modes:</td>
<td>UH/HR/EH, ES/ST, EW/WS</td>
</tr>
<tr>
<td>Altitude:</td>
<td>550km</td>
</tr>
<tr>
<td>Inclination:</td>
<td>97.06 deg</td>
</tr>
<tr>
<td>MLTAN:</td>
<td>06:00, 18:00 (local time)</td>
</tr>
<tr>
<td>File Format:</td>
<td>HDF5/GeoTIFF</td>
</tr>
<tr>
<td>Orbit repeat cycle:</td>
<td>28 days</td>
</tr>
<tr>
<td>Location accuracy:</td>
<td>4.82m RMSE, 7.32m CE90</td>
</tr>
</tbody>
</table>
AOGEOSS Task 11: AO-DataCube

GEO Work Programme 2017-2019 Application
Asia-Oceania GEOSS Initiative (AOGEOSS)

AOGEOSS Activities

Applications and services
- Task 1. AWCI
- Task 2. AP-BON
- Task 3. Carbon and GHG Initiative
- Task 4. Ocean and Society
- Task 5. Asia-RICE
- Task 6. Monitoring and evaluation of drought in Asia-Oceania region
- Task 7. Environmental Monitoring and Protection
- Task 8. Ocean and Islands
- Task 9. Himalayan GEOSS

Foundational tasks
- Task 10. Data Sharing
- Task 11. AO-DataCube
- Task 12. Users Engagement and Communication

Fig. 1 The AOGEOSS Activities
(Boxes with blue background represent existing AP activities; other boxes represent newly proposed ones)
In computer programming contexts, a data cube (or datacube) is a multi-dimensional ("n-D") array of values.

Typically, the term data cube is applied in contexts where these arrays are massively larger than the hosting computer's main memory; examples include multi-Terabyte/Petabyte data warehouses and time series of image data.

Adam Lewis et al. (2016) International Journal of Digital Earth. Vol 9(1) Rapid, high-resolution detection of environmental change over continental scales from satellite data – the Earth Observation Data Cube
What is Analysis Ready Data (ARD)?

Analysis Ready Data are time-series stacks of overhead imagery that are prepared for a user to analyze without having to pre-process the imagery themselves. Those who don’t work with satellite imagery every day likely underestimate the amount of labor involved in preparing imagery for analysis.

Once these steps are completed, you are ready to start analysis!

ARD Types of Satellite Data in Open Data Cube

ARD: Analysis Ready Data

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Type</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landsat-5/7/8</td>
<td>ARD (surface reflectance, USGS Collection-1, UTM projection, 30m)</td>
<td>CEOS</td>
</tr>
<tr>
<td>Landsat-5/7/8</td>
<td>ARD (surface reflectance, from LEDAPS and NBAR, Albers projection, 25-m)</td>
<td>GA</td>
</tr>
<tr>
<td>Sentinel-1</td>
<td>ARD (gamma nought, 10m)</td>
<td>CEOS</td>
</tr>
<tr>
<td>Sentinel-1</td>
<td>ARD (gamma nought, Albers projection, 12.5m)</td>
<td>GA</td>
</tr>
<tr>
<td>Sentinel-2</td>
<td>Level-1C (MSI TOA reflectance, Albers projection, 10/20/60m)</td>
<td>GA</td>
</tr>
<tr>
<td>ALOS-1/2 PALSAR Annual</td>
<td>ARD (gamma nought, WGS84, 25m)</td>
<td>CEOS</td>
</tr>
<tr>
<td>Mosaics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASTER Digital Elevation</td>
<td>ARD (elevation)</td>
<td>CEOS</td>
</tr>
<tr>
<td>Model (DEM)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MODIS - MCD43</td>
<td>ARD (BRDF Albedo, 16-Day L3 Global 500m)</td>
<td>GA</td>
</tr>
</tbody>
</table>
ARD Tasks and Missions

The Move Toward Analysis Ready Data and the Opportunities / Challenges Ahead

April 14, 2016
Presented By:
Thomas Cecere
Land Remote Sensing Program International Liaison, USGS
tcecerel@usgs.gov, +1 (703) 648-5551

U.S. Landsat Analysis Ready Data (ARD) Mission

Moving Towards Analysis Ready SAR Data
January 30, 2017
Big Geospatial Data Analytics (BDGA) Workshop
Canada Centre for Mapping and Earth Observation

Functionality
The ARD Tools package provides users with the following capabilities:

- Support for reading and writing metadata
- Radiometric normalization techniques
- Spectral pre-classification
- Topographic normalization

remote sensing

Article
Analysis Ready Data: Enabling Analysis of the Landsat Archive
John L. Dwyer 1,*, David P. Roy 2,*, Brian Sauer 1, Calli B. Jenkerson 3, Hankui K. Zhang 2, and Leo Lymburner 4

Special Issue "Science of Landsat Analysis Ready Data"
## Analysis Ready Data (ARD): Comparison of Sentinel-1 SAR

<table>
<thead>
<tr>
<th>step</th>
<th>Description</th>
<th>Google Earth Engine</th>
<th>CSIRO</th>
<th>Geoscience Australia</th>
<th>Satellite Application Catapult</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Orbit Updates</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>2</td>
<td>GRD Border Noise</td>
<td></td>
<td></td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>3</td>
<td>Thermal Noise</td>
<td>O</td>
<td>O</td>
<td></td>
<td>O</td>
</tr>
<tr>
<td>4</td>
<td>Radiometric Calibration</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>5</td>
<td>Multilooking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Terrain Flattening</td>
<td>?</td>
<td>O</td>
<td>O</td>
<td>?</td>
</tr>
<tr>
<td>7</td>
<td>Speckle Filtering</td>
<td></td>
<td>O</td>
<td></td>
<td>?</td>
</tr>
<tr>
<td>8</td>
<td>Terrain Correction</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>9</td>
<td>db Conversion</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>

Daniel Wicks, Thomas Jones, Cristian Rossi, 2018. Testing the interoperability of Sentinel 1 analysis ready data over the united kingdom, IGARSS 2018
Indexing and Ingestion - Configuration Files for KOMPSAT 3 Data

Three configuration files to ingest a dataset into the Data Cube

- A dataset type configuration file
- Dataset metadata, describing a single acquisition
- An ingestion configuration file describing the output dataset characteristics for a given dataset type

One configuration file for KOMPSAT 3 Products

One configuration file for each KOMPSAT 3 Scene

One configuration file for KOMPSAT 3 Ingested data
Data ingestion is the process of obtaining and importing data for immediate use or storage in a database. To ingest something is to "take something in or absorb something."

Data can be streamed in real time or ingested in batches. When data is ingested in real time, each data item is imported as it is emitted by the source. When data is ingested in batches, data items are imported in discrete chunks at periodic intervals of time. An effective data ingestion process begins by prioritizing data sources, validating individual files and routing data items to the correct destination.

https://whatis.techtarget.com/definition/data-ingestion
Ingestion Scripts for KOMPSAT 3 Optical Data

Installation and Setting of Open Data Cube

1. Install Data Cube Package and Dependencies
2. Create a Database to hold the Index
3. Initialize Database

Define Product Type
- Data Definition in YAML

Index Datasets
- Metadata for Data Indexing in YAML
- Write Prepare Script
- Run Prepare Script
- Create Dataset Metadata
- Run Index Dataset

Data Ingestion
- Write Ingestion Config
- Run Ingestion

Complete Ready to Analysis Data

KOMPSAT Scripts
Ingestion Scripts for KOMPSAT 3 Data in Open Data Cube

Test System
OS: Windows / Ubuntu / Mac OS
Setup: Python pip or Miniconda

Now Experimental Stage

Hardware Components

CPU
Intel E5-2630 v4 *1 (10 cores/20Threads)

Memory
ECC 16GB * 4 = 64GB

SSD 512GB
(Operating System/Software/Open Data Cube)

HDD(RAID) 8TB
(Remote Sensing Storage)

Configuration for Open Data Cube Deployment

Miniconda 3 (datacube_kompsat)
OPEN DATA CUBE 1.6.1
jupyterhub
Python 3.6
PostgreSQL

Location “/” Installed Operating System & Open Data Cube
Location “/datacube/**” stored KOMPSAT & etc.
Test Case of KOMPSAT 3 Dataset Ingestion

- Using Data Cube API and Jupyter Notebook by ODC
- Indexed Image set in GeoTiff (or netCDF)
  - ODC Accessing
  - Image Searching (Time, Lat/Lon, Resolution, Geo-coordinates, Band number etc.)
  - RGB Color Composting

```
In [1]:
import datacube
from datacube.storage.masking import mask_invalid_data
query = {
    'time': ('2013-01-01', '2016-01-01'),
}
dc = datacube.Datacube(app='plot-rbg-recipe')
data = dc.load(product='kompasat3_11g_ms', measurements=['red', 'green', 'blue'],
               output_crs='EPSG:4326', resolution=(0.000028, -0.000028), **query)
data = mask_invalid_data(data)
fake_saturation = 8000
rgb = data.to_array(dim='color')
rgb = rgb.transpose((rgb.dims[1]*rgb.dims[1]))
rgb = rgb.where((rgb <= fake_saturation).all(dim='color'))
rgb /= fake_saturation
rgb.plot.imshow(x=data.crs.dimensions[1], y=data.crs.dimensions[0],
                col='time', col_wrap=4, add_colorbar=False)

Out [1]:
<xarray.plot.facetgrid.FacetGrid at 0x7f4ab8606a90>
```

KOMPSAT Samples around the southern Mekong River
On-going Works and Plan

- Developing ingestion scripts for KOMPSAT data series
- Scripts and modules releasing to AO KOMPSAT users via GitHub
- Preparation studies for KOMPSAT ARD for ODC implementation in an open source cloud computing environment and its testing, e.g., on openstack.
- Linking GIS layers and functionalities via QGIS plugin with