Introduction to Daily SDC on AWS

In the fields of agriculture, wildlife conservation, human settlement studies, resources, energy, and environment, more and more observation time series provided by Landsat and other sensors are used for ground monitoring.

However, the widespread cloud contamination in the remotely sensed data, the abnormal state of the satellite sensors, or random noises have exacerbated observation incompleteness and reduced the effectiveness and availability of Earth observations (EO). For reliable monitoring, a complete observation record in space and time is required.

Besides, due to technical and budget limitations, remote sensing instruments trade spatial resolution, and swath width. As a result, not one sensor provides both high spatial resolution, and high temporal frequency. Therefore, high spatial resolution sensors are usually accompanied by a low temporal frequency. However, EO tasks involving frequent dynamic monitoring of land processes require urgently both high spatial resolution and preferably daily temporal frequency for regional and global applications.

To these ends, we developed a spatial-temporal data reconstruction and fusion framework based on an automated, serverless production chain on the AWS to build our seamless data cube (SDC) – the daily surface reflectance record with spatial completeness and temporally equal interval. Based on the data reconstruction algorithm, raw Landsat data are optimally predicted, so all missing values and outliers (caused by cloud effects, data damage or loss, saturation, etc.) are effectively recovered, greatly overcoming the shortcoming of data incompleteness that conventional EO data has.

To further deal with the conflict between high spatial resolution and frequent time coverage, by constructing a virtual constellation of complementary types of satellites, we integrated the merit of high spatial resolution Landsat data with that of high temporal frequency MODIS data through data fusion, to achieve better fused daily EO. The produced SDC has a spatial resolution of 30 m and a daily interval, which improves the Landsat data availability level by 16 times. Such kind of data with high spatial resolution and temporal frequency is a long-term dream in the remote sensing application community, that has never been realized before.

Following Analysis Ready Data (ARD) standards, all data used first underwent a rigorous data pre-processing process (including geometric correction, topographic correction, radiometric correction, and atmospheric correction) and configured with a detailed quality assessment layer and general metadata to allow users to assess the suitability of data. In this way, the built SDC can significantly reduce the pre-processing burden of users, lower the threshold for using remote sensing data, broaden the use of remotely sensed data to a wider range of communities.

Data in SDC is organized and aligned into a fixed and regular tiling grid, to improve parallel computing performance, optimize and simplify the data query and loading process. Besides, the SDC is stored in the Cloud Optimized GeoTIFF (COG) format along with built Spatiotemporal Asset Catalog (STAC) metadata, serving in a cloud native geospatial architecture to support adaptive, accurate, and smart online geospatial query and analysis.

The spatial-temporal consistent SDC can provide high-quality seamless observations, which is equivalent to realizing a daily scanning of the Earth at a 30 m resolution and obtaining a daily cloud-free and seamless mosaic photograph. It significantly improves the spatial-temporal consistency of observations, to serve more higher-level applications of Landsat data, such as to derive various essential variables such as albedo, leaf area index, burnt area mapping, land cover, and land use mapping, soil moisture, etc. that are required in terrestrial, oceanic, and atmospheric environmental monitoring to support tasks such as water resource management, ensuring food security, climate, and ecosystem modeling, biodiversity protection, and environmental and health disaster prevention. For these different remote sensing applications, the SDC can be more readily used and thus are more promising. The knowledge gained from this uniquely comprehensive data set can form a new foundation for achieving the United Nations (UN) Sustainable Develop Goals (SDGs).

Data Access from AWS

The SDC is stored in the cloud using AWS object storage service, S3. Users can use local computers or flexible cloud computing resources on AWS in the form of EC2 instances to access data.

Data Directory and Location

The SDC is organized using a directory structure based on the data acquisition date of each scene and the tile's ID in the SDC tile system (as shown in Figure 1). The SDC is stored under the World Geodetic System (WGS84) coordinate reference system (EPSG:4326). The size of each tile is $0.3° \times 0.3°$, and the tile's ID (XXX, YYY) can be calculated as: XXX = fix(longitude/$0.3°$), YYY = fix(latitude$0.3°$). For each scene, the data is provided in:

s3://sdc-daily-thu/XXX/YYY/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd/

where,

- XXX = e.g. 370 - tile ID on the longitude direction.
- YYY = e.g. 080 - tile ID on the latitude direction.
- VX = e.g. V2 – Product version.
- YYYY = e.g. 2019 - acquisition year.
- MM = e.g. 03 - acquisition month.
- DD = e.g. 20 - acquisition day.
- yyyy = e.g. 2020 - processing year.
- mm = e.g. 01 - processing month.
- dd = e.g. 15 - processing day.

For instance, the files associated with the scene for the tile (370,119) taken on 2020/06/21 is available at s3://sdc-daily-thu/370/119/SDC_V2_370119_20200621_20201004/.

If you use the AWS Command Line Interface, you can access the bucket with this simple command:

aws s3 ls s3://sdc-daily-thu/

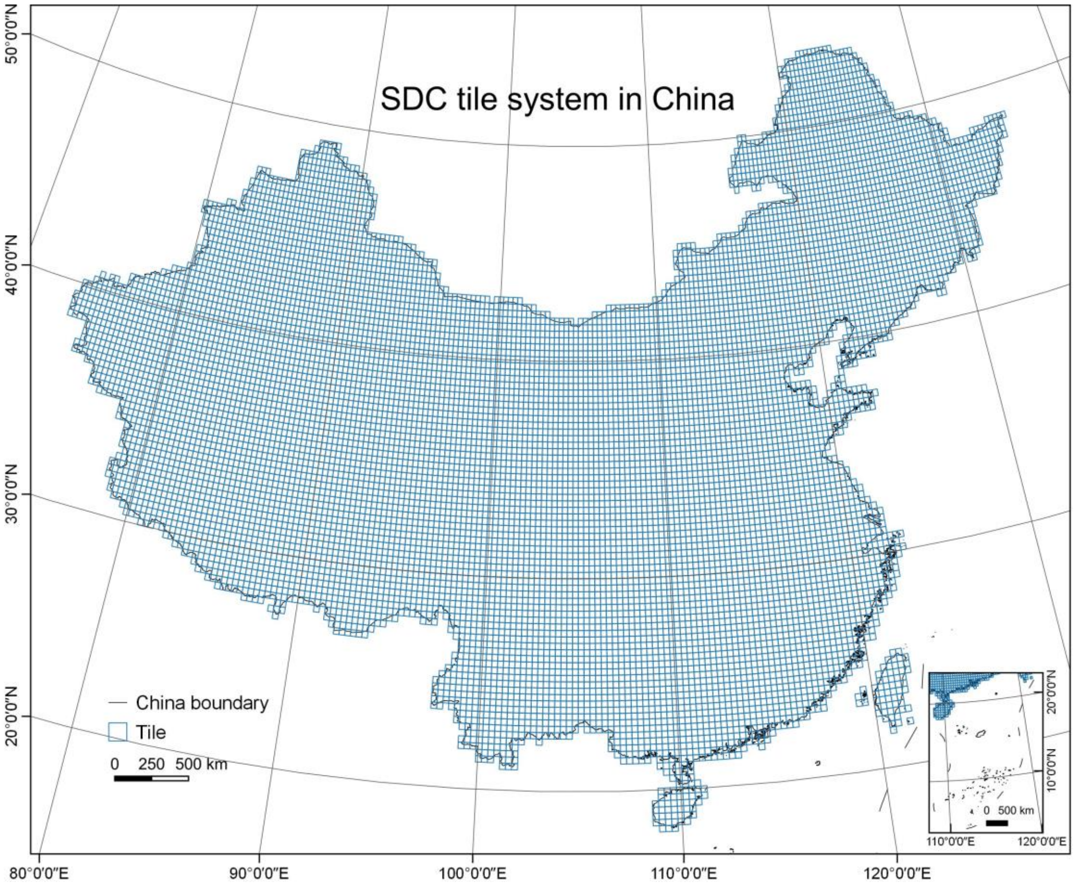

Figure 1 SDC tile system in China, where a tile size is $0.3° \times 0.3°$.

Data Types and Structure

Each scene's directory includes:

- a .tif Cloud Optimized GeoTIFF for each of the scene's up to 6 spectral bands' surface reflectance
- a .tif Cloud Optimized GeoTIFF recording Quality Assessment (QA) information
- a thumbnail jpeg
- a .json STAC metadata file

Specifically, the directory (s3://sdc-daily-thu/XXX/YYY/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd/) and contain the following files.

- s3://sdc-daily-thu/XXX/YYY/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd_SR.tif
- s3://sdc-daily-thu/XXX/YYY/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd_QA.tif

- s3://sdc-daily-thu/XXX/YYY/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd_thumbnail.jpg
- s3://sdc-daily-thu/XXX/YYY/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd/SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd.json

Raster Data

Due to the differences in band configuration between Landsat and MODIS, 6 matching spectral bands (as shown in Table 1) are processed. The raw scenes in Landsat Collection 1 Level 1 Tier 1 terrain and precision corrected products (L1TP) with cloud cover lower than 80% were first converted to surface reflectance first with USGS EROS Center Science Processing Architecture (ESPA). For each scene, missing values or outliers are predicted by data reconstruction. For left dates without observation, daily surface reflectance is predicted by data fusion of Landsat and MODIS (from daily MCD43A4 product) data. In addition to daily surface reflectance, a sperate band is used to store QA information.

The SDC data is provided in COG raster format, which can be opened with a variety of GIS software tools, such as QGIS software, GDAL, and Rasterio library in Python. The spatial resolution of the SDC data is $0.00025°$ per pixel (27.83 m on the Equator).

Table 1 Raster layer description

Band	**Description**	**Data Type**	**Fill Value**	**Valid Range**	**Scale Factor**
Blue	Surface reflectance for blue band	16-bit signed integer	-32767	0 to 10000	0.0001
Green	Surface reflectance for green band	16-bit signed integer	-32767	0 to 10000	0.0001
Red	Surface reflectance for red band	16-bit signed integer	-32767	0 to 10000	0.0001
NIR	Surface reflectance for nir band	16-bit signed integer	-32767	0 to 10000	0.0001
SWIR1	Surface reflectance for swir1 band	16-bit signed integer	-32767	0 to 10000	0.0001
SWIR2	Surface reflectance for swir2 band	16-bit signed integer	-32767	0 to 10000	0.0001
QA	Quality assessment information	16-bit signed integer	-32767	0 to 32767	N/A

Metadata file

The JSON metadata file provides basic descriptions of the scene (Table 2).

Table 2 Content in the metadata file

Item	Description
collection	sdc
id	SDC_VX_XXXYYY_YYYYMMDD_yyyymmdd
XXX	Tile ID on the longitude direction
YYY	Tile ID on the latitude direction
type	Feature
datetime	Data acquisition time
geometry	Full footprint
bbox	Bounding box
spatialresolution	0.00025°
platform	landsat, modis
instrument	oli, tm, etm
epsg	4346
rawgoodper	Raw good-quality observation percentage
qalevel	Prediction quality flag
contact	liuhan18@mails.tsinghua.edu.cn
managedby	Department of Earth System Science, Tsinghua University

Tutorials

Tutorials on how to use the data on AWS are available at:

https://github.com/thu-hanliu/sdc/blob/main/SDC_tutorial.pdf

Contact

If you have questions about the data, please contact

liuhan18@mails.tsinghua.edu.cn