

Harmonized Landsat Sentinel-2 (HLS) Product User Guide

Product Version 1.5

Principal Investigator: Dr. Jeffrey G. Masek, NASA/GSFC

Web site: <http://hls.gsfc.nasa.gov>

Correspondence email address: Jeffrey.G.Masek@nasa.gov



Acronyms

AROP	Automated Registration and Orthorectification Package
BRDF	Bidirectional Reflectance Distribution Function
BT	Brightness temperature
CMG	Climate Modelling Grid
ETM+	Enhanced Thematic Mapper Plus
GDAL	Geospatial Data Abstraction Library
GLS	Global Land Survey
HDF	Hierarchical Data Format
HLS	Harmonized Landsat and Sentinel-2
KML	Keyhole Markup Language
MGRS	Military Grid Reference System
MSI	Multi-Spectral Instrument
NBAR	Nadir BRDF-normalized Reflectance
OLI	Operational Land Imager
QA	Quality assessment
RSR	Relative spectral response
SDS	Scientific Data Sets
SR	Surface reflectance
SZA	Sun zenith angle
TM	Thematic Mapper
TOA	Top of atmosphere
UTM	Universal Transverse Mercator
WRS	Worldwide Reference System

DRAFT

1 Introduction

The Harmonized Landsat and Sentinel-2 (HLS) project is a NASA initiative and collaboration with USGS to produce compatible surface reflectance (SR) data from a virtual constellation of satellite sensors, the Operational Land Imager (OLI) and Multi-Spectral Instrument (MSI) onboard the Landsat-8 and Sentinel-2 remote sensing satellites respectively. The combined measurement enables global land observation every 2-3 days at moderate (30 m) spatial resolution. The HLS project uses a set of algorithms to derive seamless products from OLI and MSI: atmospheric correction, cloud and cloud-shadow masking, spatial co-registration and common gridding, illumination and view angle normalization and spectral bandpass adjustment. The HLS data products can be regarded as the building blocks for a “data cube” so that a user may examine any pixel through time, and treat the near-daily reflectance time series as though it came from a single sensor.

The HLS suite contains two products, S30 and L30, derived from Sentinel-2 L1C and Landsat L1T (Collection 2) products, respectively. They are derived using the same atmospheric correction code, the same view/illumination angle correction algorithm, and gridded into the same MGRS tiles using a 30m pixel size. Thus, the S30 and L30 products are physically co-registered to common tiles. The S30 products have also been adjusted to mimic the Landsat-8 spectral response for common spectral bands.

2 New in v1.5

HLS v1.5 builds on v1.4 by updating and improving processing algorithms, expanding spatial coverage, and providing validation. Particular updates are as follows:

- *Global coverage.* All global land, including major islands but excluding Antarctica, is covered.
- *Input data.* Landsat 8 Collection-2 data from USGS are used as input; better geolocation is expected as C2 data use the Sentinel-2 Global Reference Image (GRI) as an absolute reference.
- *Atmospheric correction.* A C version of LaSRCv3.5.5 has been applied for both Landsat 8 and Sentinel-2 data for computational speedup. LaSRCv3.5.5 has been validated for both Landsat 8 and Sentinel-2 within the CEOS ACIX-I (Atmospheric Correction Inter-Comparison eXercise, <http://calvalportal.ceos.org/projects/acix>).
- *QA band.* The QA band is generated exclusively by and named after Fmask, consistently for the two HLS products (S30 and L30).
- *BRDF adjustment.* BRDF adjustment mainly normalizes the view angle effect, with the sun zenith angle largely intact. This adjustment is applied to the Sentinel-2 red-edge bands as well.
- *Sun and view angle bands are provided.*
- *Product format.* The product is delivered in individual Cloud Optimized GeoTIFF (COG) files to allow for spectral and spatial subsetting in applications.
- *Temporal Coverage and Latency.* Version 1.5 moves toward “keep up” processing. The intent is to continually update products with <5 day latency. Users are cautioned however that HLS is still a research product.

3 Products overview

3.1 Input data

The Operational Land Imager (OLI) sensor is a moderate spatial resolution multi-spectral imager onboard the Landsat-8 satellite, in a sun-synchronous orbit (705 km altitude) with a 16-day repeat cycle. The sensor has a field of view of 15 degrees (approximately a 185km swath). The OLI sensor has 9 solar reflective bands and the data are co-registered with the data from the 2-band instrument TIRS (Thermal Infrared Sensor) onboard the same Landsat-8 satellite (Irons et al., 2012). The native resolution for OLI is 30 m and for TIRS is 100 m, but TIRS data are resampled to 30 m for distribution. HLS v1.5 uses Landsat 8 Collection-2¹ Level-1 top-of-atmosphere (TOA) product as input: for “keep-up” processing, the Real-Time data are used and, for back processing, the tier-based data are used. The Real-Time TOA OLI data have the same quality as the tier-based data do, but the Real-Time TIRS data may have lesser geolocation and radiometric quality.

The Sentinel-2 Multi-Spectral Instrument (MSI) is onboard the Sentinel-2A and -2B satellites with a 786 km orbit altitude. The ground sampling distance varies with the spectral bands: 10 m for the visible bands and the broad NIR band, 20 m for the red edge, narrow NIR and SWIR bands, and 60 m for the atmospheric bands (Drusch et al., 2012). The sensor has a 20.6° field of view corresponding to an image swath width of approximately 290 km. HLS v1.5 uses Level-1C (L1C) Top of Atmosphere product. Table 1 provides an overview of Landsat 8 and Sentinel-2 characteristics.

Table 1: Input data characteristics

	Landsat 8/OLI-TIRS	Sentinel-2A/MSI	Sentinel-2B/MSI
Launch date	February 11, 2013	June 23, 2015	March 7, 2017
Equatorial crossing time	10:00 a.m.	10:30 a.m.	10:30 a.m.
Spatial resolution	30 m (OLI) / 100 m (TIRS)	10 m / 20 m / 60 m (see spectral bands)	
Swath / Field of view	180 km / 15°	290 km / 20.6°	
Spectral bands (central wavelength)	Ultra blue	443 nm	443 nm (60 m)
	Visible	482 nm, 561 nm, 655 nm	490 nm (10 m), 560 nm (10 m), 665 nm (10m)
	Red edge	-	705 nm (20 m), 740 nm (20 m), 783 nm (20 m)
	NIR	865 nm	842 nm (10 m), 865 nm (20 m)
	SWIR	1609 nm, 2201 nm	1610 nm (20 m), 2190 nm (20 m)
	Cirrus	1373 nm	1375 nm (60 m)
	Water Vapor	-	945 nm (60 m)
	Thermal	10.9 μm, 12 μm	-

¹ Landsat Collections — <https://landsat.usgs.gov/landsat-collections>

3.2 Overall HLS processing flowchart

The same processing methods are applied to generate S30 and L30 (Fig. 1). LaSRC is used for atmospheric correction, Fmask for cloud masking (QA). Earlier MSI data need coregistration improvement, and Landsat data are gridded into the tiles that MSI use; all pixels are resampled to 30m. Surface reflectance is corrected for view and illumination angle effect. MSI bandpasses are adjusted to the Landsat ones. A detailed description of processing methods can be found in Section 4.

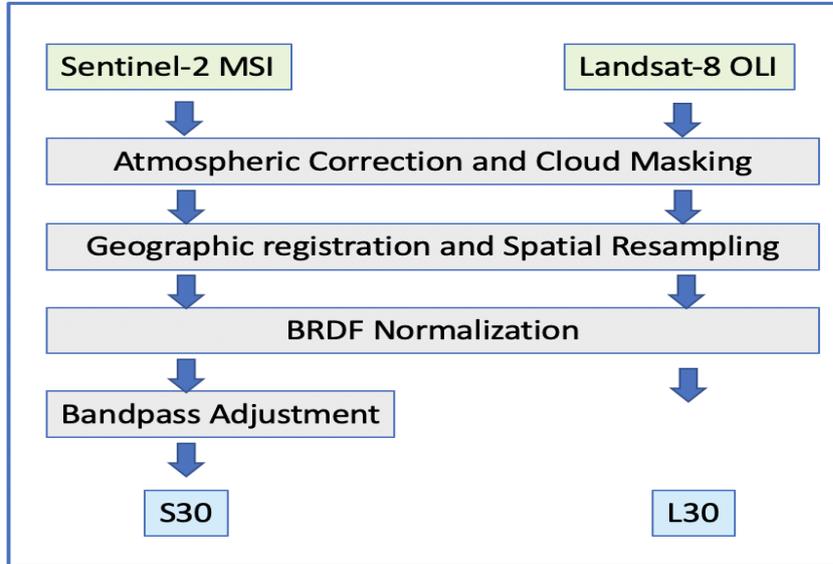


Figure 1: HLS science algorithm processing flow

3.3 Products specifications

As noted above, the HLS suite contains two products, S30 and L30. They are derived by using the same atmospheric correction code, same view/illumination angle correction algorithm, and gridded into the same MGRS tiles with a 30m pixel size. S30 has been adjusted to the Landsat-8 spectral response. The product specifications are given in Table 2.

Table 2: HLS products specifications

Product Name	S30	L30
Input sensor	Sentinel-2A/B MSI	Landsat-8 OLI/TIRS
Spatial resolution	30 m	30 m
BRDF-adjusted	Yes (except bands 09, 10)	Yes (except band 09)
Bandpass-adjusted	Adjusted to OLI-like (except red edge, water vapor and cirrus bands)	No (HLS uses OLI bandpasses)
Projection	UTM	UTM
Tiling system	MGRS (110*110)	MGRS (110*110)

3.4 Spectral bands

All Landsat-8 OLI and Sentinel-2 MSI reflective spectral bands nomenclatures are retained in the HLS products (Table 3).

Table 3: HLS spectral bands nomenclature

Band name	OLI band number	MSI band number	HLS band code name L8	HLS band code name S2	Wavelength (micrometers)
Coastal Aerosol	1	1	B01	B01	0.43 – 0.45*
Blue	2	2	B02	B02	0.45 – 0.51*
Green	3	3	B03	B03	0.53 – 0.59*
Red	4	4	B04	B04	0.64 – 0.67*
Red-Edge 1	–	5	–	B05	0.69 – 0.71**
Red-Edge 2	–	6	–	B06	0.73 – 0.75**
Red-Edge 3	–	7	–	B07	0.77 – 0.79**
NIR Broad	–	8	–	B08	0.78 – 0.88**
NIR Narrow	5	8A	B05	B8A	0.85 – 0.88*
SWIR 1	6	11	B06	B11	1.57 – 1.65*
SWIR 2	7	12	B07	B12	2.11 – 2.29*
Water vapor	–	9	–	B09	0.93 – 0.95**
Cirrus	9	10	B09	B10	1.36 – 1.38*
Thermal Infrared 1	10	–	B10	–	10.60 – 11.19*
Thermal Infrared 2	11	–	B11	–	11.50 – 12.51*

* from OLI specifications

** from MSI specifications

3.5 Output projection and gridding

HLS has adopted the tiling system used by Sentinel-2. The tiles are in the Universal Transverse Mercator (UTM) projection, and are 109,800 m (110km nominally) on a side. The tiling system is aligned with the UTM-based Military Grid Reference System (MGRS). The UTM system divides the Earth's surface into 60 longitude zones, each 6° of longitude in width, numbered 1 to 60 from 180° West to 180° East. Each UTM zone is divided into latitude bands of 8°, labeled with letters C to X from South to North. A useful mnemonic is that latitude bands N and later are in the Northern Hemisphere. Each 6°×8° polygon (grid zone) is further divided into the 110km ×110km Sentinel-2 tiles labeled with letters. For example, tile 11SPC is in UTM zone 11, latitude band S (in Northern Hemisphere), and labeled P in the east-west direction and C in the south-north direction within grid zone 11S. Users should note that there is horizontal and vertical overlap of around 8-10 km between two adjacent tiles in the same UTM zone. The overlap between two adjacent tiles both straddling a UTM zone boundary may be much greater. A KML file produced by ESA showing the location of all Sentinel-2 tiles is available at https://sentinel.esa.int/documents/247904/1955685/S2A_OPER_GIP_TILPAR_MPC_20151209T095117_V20150622T000000_21000101T000000_B00.kml

One trivial difference with the ESA gridding is that HLS inherits the USGS UTM convention of keeping the Y coordinate for the Southern Hemisphere negative, therefore with no need for hemisphere specification. In contrast, many spatial data handling tools use a convention of

adding 10,000,000 meters to make the southern coordinate positive (i.e. use of a false northing 10,000,000) and indicating which hemisphere to avoid confusion. These tools may textually report a Southern Hemisphere dataset with a false-northing 0 and no indication of hemisphere as being in Northern Hemisphere, but correctly handle the geolocation of the data in processing.

4 Algorithms description

4.1 Atmospheric correction

The same atmospheric correction algorithm, Land Surface Reflectance Code (LaSRC) developed by Eric Vermote (NASA/GSFC) (Vermote et al., 2016), is applied to both sensors data. LaSRC is based on the 6S radiative transfer model and a heritage from the MODIS MCD09 products (Vermote and Kotchenova 2008) as well as the earlier LEDAPS algorithm implemented for Landsat-5 and Landsat-7 (Masek et al. 2006). A detailed description of the method is given in Vermote et al. (2016), and results of surface reflectance validation for Landsat 8 and Sentinel-2 within CEOS ACIX-I are provided in Doxani et al. (2018).

LaSRC uses atmospheric inputs (ozone, water vapor) from MODIS to correct for gaseous absorption, and surface pressure based on topographic elevation to correct for molecular (Rayleigh) scattering. Aerosol optical thickness (fixed continental type) is derived via an image-based algorithm using the ratio of the red and blue spectral bands (Vermote et al., 2016). The output is directional surface reflectance. HLS also includes the two thermal infrared bands from the Landsat 8 TIRS sensor in the L30 product – these values are not atmospherically corrected, but are rescaled apparent brightness temperature (no atmosphere, unity emissivity).

HLS 1.5 uses a C version LaSRC v3.5.5 implemented by USGS, mainly for computational speedup.

4.2 Spatial co-registration of input data

Our objective in HLS is to maintain the geodetic accuracy requirement of the Sentinel-2 images (<20 m error, 2σ) and improve the multi-temporal co-registration among Sentinel-2 images and between Sentinel-2 and Landsat 8 images (<15 m 2σ) for the 30 m products. This specification supports time series monitoring of small fields, man-made features, and other spatially heterogeneous cover types.

Coregistration is less of a concern in HLS v1.5, but we describe the methodology as it still has relevance for earlier MSI L1C data. Before HLS v1.5, two issues impeded a direct registration of Landsat 8 and Sentinel-2 imagery without additional processing. First, while the relative co-registration of Landsat 8/OLI Collection-1 imagery was quite accurate (<6.6 m, Storey et al. 2014), the absolute geodetic accuracy varied with the quality of the Global Land Survey 2000 (GLS2000) ground control around the world. In some locations, the GLS geodetic accuracy was in error by up to 38 m (2σ , Storey et al. 2016). As a result, Sentinel-2/MSI and Landsat 8/OLI Level-1 products did not align to sub-pixel precision for those locations (Storey et al. 2016). Second, an error in the yaw characterization for the MSI L1C images processed before v02.04 (May 2016) caused misregistration between the edges of MSI images acquired from adjacent orbits (ESA 2018). The misregistration of up to 2.8 pixels at 10 m resolution between Sentinel-2A images from adjacent orbits has been observed by Skakun et al. (2017) and Yan et al. (2018). Although the issue was fixed with L1C version 02.04 (yielding to a measured absolute

geolocation of less than 11m at 95.5% confidence, ESA 2018), archived Sentinel-2 data from 2015-2016 will continue to have this error until the entire archive is reprocessed by ESA.

Earlier HLS versions used the Automated Registration and Orthorectification package (AROP, Gao et al. 2009) to resample Landsat imagery to a “master” Sentinel-2 image for each tile (see Claverie et al., 2018 for details). However, HLS v1.5 is based on USGS Collection 2 Landsat data, which now uses the Sentinel-2 Global Reference Image (GRI) as an absolute control. In HLS v1.5, the improved Landsat ground control in Collection 2 eliminates the need for AROP for L30 production. Cubic convolution resampling is still needed because USGS aligns the UTM coordinate origin to a pixel center while ESA aligns it to a pixel corner. In addition, AROP is still required for 2015-2016 Sentinel-2 input data due to the yaw steering issue described above. When the entire Sentinel-2 archive is consistently processed to be on a collection basis the use AROP for early Sentinel-2 data will be retired too.

4.3 Cloud masks

HLS provides per-pixel cloud, cloud shadow, snow, and water masks. In earlier versions of HLS, the cloud mask was a union of cloud masks accompanying the Level-1 input, the internal cloud mask of atmospheric correction code LaSRC, and the cloud mask by Fmask (Zhu et al. 2015). In HLS v1.5, the cloud mask was generated exclusively by Fmask 4.2, an update of Fmask 4.0 reported in Qiu et al. (2019).

The internal cloud mask of LaSRC is not available in HLS v1.5, but may be added as a separate data layer in future versions.

4.4 View and illumination angles normalization

The S30 and L30 Nadir BRDF-Adjusted Reflectance (NBAR) products are surface reflectance normalized for the view angle and the illumination angle effect, using the *c*-factor technique by Roy et al. (2016). The view angle is set to nadir for all pixels in normalization. The illumination angle for a tile is set to the mean of the solar zenith angles at the tile center at the respective times when Landsat-8 and Sentinel-2 overpass the tile center’s latitude on the day; this angle is derived using the code described in Li et al (2018).

The BRDF normalization uses a constant set of BRDF coefficients, derived from 12-month MODIS 500m global BRDF product (MCD43) (more than 15 billion pixels). The derived BRDF coefficients are applied to OLI and MSI bands equivalent to MODIS ones. The technique has been evaluated using off-nadir (i.e. in the overlap areas of adjacent swaths) ETM+ data (Roy et al. 2016) and MSI data (Roy et al. 2017). For the normalization of MSI red-edge spectral bands that have no MODIS equivalents, the linearly interpolated BRDF coefficients from the enclosing MODIS red and NIR wavelength bands are used (Roy et al 2017). BRDF coefficients for the three kernels (isotropic, geometric, and volumetric) are shown in the Table 4. The kernel definitions are described in the ATBD of MOD43 product (Strahler et al. 1999).

Table 4: BRDF coefficients used for the c-factor approach (Roy et al. 2016 and 2017)

Band name	HLS band code name L8	HLS band code name S2	Equivalent MODIS band	f _{iso}	f _{geo}	f _{vol}
Coastal/Aerosol	B01	B01	3	0.0774	0.0079	0.0372
Blue	B02	B02	3	0.0774	0.0079	0.0372
Green	B03	B03	4	0.1306	0.0178	0.0580
Red	B04	B04	1	0.1690	0.0227	0.0574
Red-Edge 1	–	B05	-	0.2085	0.0256	0.0845
Red-Edge 2	–	B06	-	0.2316	0.0273	0.1003
Red-Edge 3	–	B07	-	0.2599	0.0294	0.1197
NIR Broad	–	B08	2	0.3093	0.0330	0.1535
NIR Narrow	B05	B8A	2	0.3093	0.0330	0.1535
SWIR 1	B06	B11	6	0.3430	0.0453	0.1154
SWIR 2	B07	B12	7	0.2658	0.0387	0.0639

$$\rho(\lambda, \theta^{Norm}) = c(\lambda) \times \rho(\lambda, \theta^{sensor}) \quad (2)$$

$$c(\lambda) = \frac{f_{iso}(\lambda) + f_{geo}(\lambda) \times K_{geo}(\theta^{Norm}) + f_{vol}(\lambda) \times K_{vol}(\theta^{Norm})}{f_{iso}(\lambda) + f_{geo}(\lambda) \times K_{geo}(\theta^{sensor}) + f_{vol}(\lambda) \times K_{vol}(\theta^{sensor})} \quad (3)$$

where: $\theta^{Norm} \Leftrightarrow (\theta_v = 0, \theta_s = \theta_s^{out}, \Delta\varphi = 0)$

$$\theta_{sensor} \Leftrightarrow (\theta^{sensor} = \theta_v^{sensor}, \theta_s = \theta_s^{sensor}, \Delta\varphi = \Delta\varphi^{sensor})$$

The BRDF effect is caused predominantly by the view angle variation and secondarily by the solar angle variation. The normalization of the solar zenith angle is out of two considerations. First, Landsat-8 and Sentinel-2 overpass the same latitude 30 minutes apart; on these rare days when the Landsat-8 and a Sentinel-2 overpass the same ground location, the solar zenith will be different due to the 30-minute time difference. Second, since the solar zenith angle increases from east to west within a swath, the solar zenith angle for the overlapping area of two swaths can be different due to the tile's relative location change within the swaths. These points are illustrated for a tile near the Equator where the solar zenith angle changes most dramatically in these cases.

Tile 19NGA with its center at 0.41N and 66.71W is right on the equator. The mean solar zenith angle for the Landsat-8 image on the tile is greater than that for a temporally close Sentinel-2 image because Landsat-8 overpasses 30 minutes earlier (Fig 2). There is also temporal oscillation in solar zenith angle in a sensor's image time series. The mean solar zenith angle of each Sentinel-2 granule in 2019 follows two curves, which in fact originated from two adjacent orbits, not from the coexistence of S2A and S2B; when the tile is located to the east of the nadir view in the original image swath, the solar zenith angle was smaller (the lower curve of Sentinel-2 in Fig. 2) than in a temporally close image when the tile is located to the west of the nadir view (the upper curve of Sentinel-2). The observed solar zenith angle oscillates day to day between the two orbits. Similar pattern was present in the Landsat images over this tile, which was also observed from two adjacent Landsat orbits (Fig 2). The solar zenith angle temporal oscillation in Landsat time series is smaller because Landsat image swath is narrower than Sentinel-2's (185km vs 290 km).

The solar zenith angle used in normalization is the mean of the solar zenith angles at the respective times that Landsat-8 and Sentinel-2 overpass a tile center's latitude. This prescribed solar zenith angle is calculated using the software provided by Li et al (2018). The idea is based on the fact that a sensor overpasses the same latitude at the same local solar time and therefore the solar zenith angle will be the same at nadir for the same latitude on the same day. The angle normalization takes into account all types of variations presented in Fig 2, but at the same time allows for the smooth change of daily solar zenith angle due to daily solar declination change.

For high-latitude tiles with their centers above the highest latitude that Landsat-8 and Sentinel-2's nadir view can reach (81.8 degrees and 81.38 degrees respectively), the NBAR solar zenith angle prescribed by Li et al (2018) cannot be applied, and the mean observed solar zenith angle in the tile is used instead. Only about 20 land tiles fall into this category.

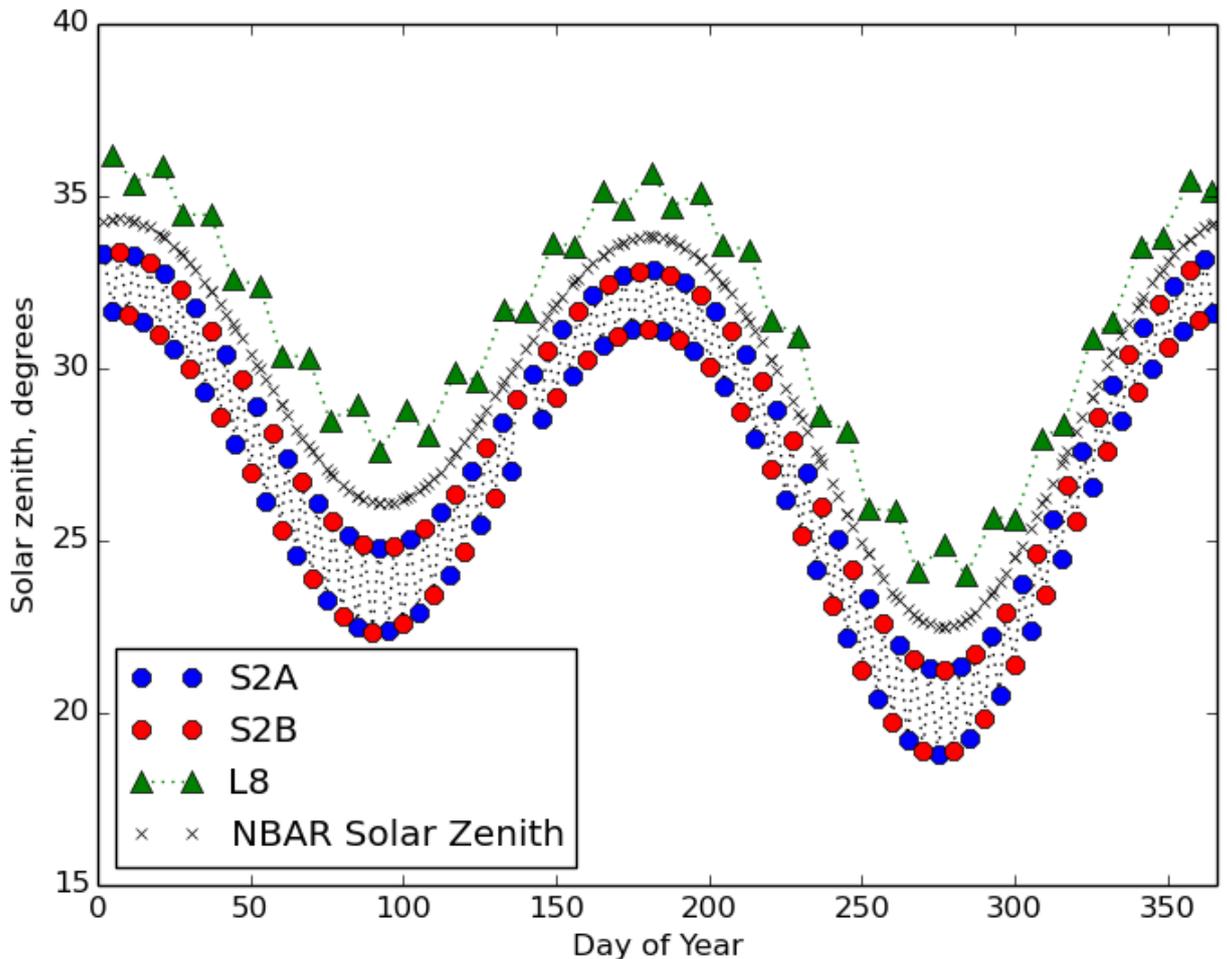


Fig. 2. The observed mean solar zenith angle in each tiled Sentinel-2 and Landsat-8 image and the solar zenith angle used in each image's BRDF normalization, for an equatorial tile 19NGA in 2019. The observed mean solar zenith angle in a Landsat image is higher than that in a temporally close Sentinel-2 image because Landsat overpasses 30 minutes earlier. There is also day-to-day oscillation in

mean observed solar zenith angle for each sensor due to the alternating observation from two adjacent orbits.

4.5 Bandpass adjustment

MSI and OLI have slightly different bandpasses for equivalent spectral bands, and these differences need to be removed in HLS products. The OLI spectral bandpasses are used as reference, to which the MSI spectral bands are adjusted. The bandpass adjustment is a linear fit between equivalent spectral bands. The slope and offset coefficients were computed based on 500 hyperspectral spectra selected on 160 globally distributed Hyperion scenes processed to surface reflectance and used to synthesis MSI and OLI bands. MSI's RSRs correspond to the version v2.0 (Claverie et al., 2018). The spectral differences between MSI onboard Sentinel-2A (S2A) and Sentinel-2B (S2B) are accounted. Note that the S2A CA and Blue bands RSRs correspond to S2B RSRs. The coefficients are given in Table 5, and scatterplots are given in Figure 2 and Figure 3.

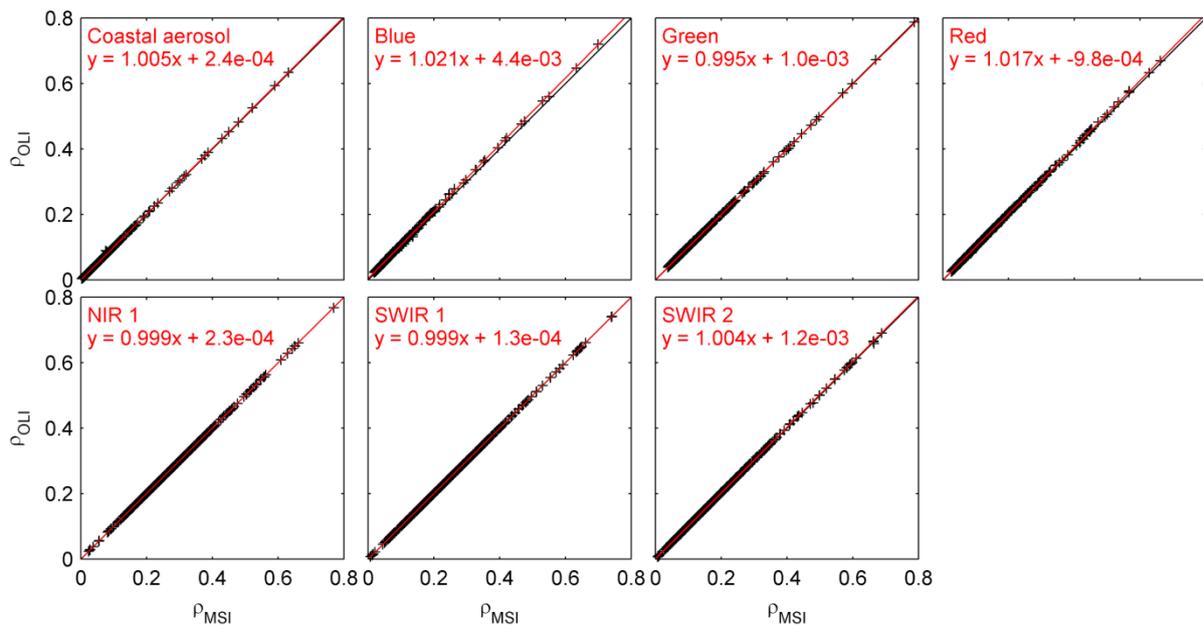


Figure 2: Sentinel-2A MSI vs OLI surface reflectance for the seven equivalent bands, using a synthetic dataset built with 500 surface reflectance spectra from Hyperion.

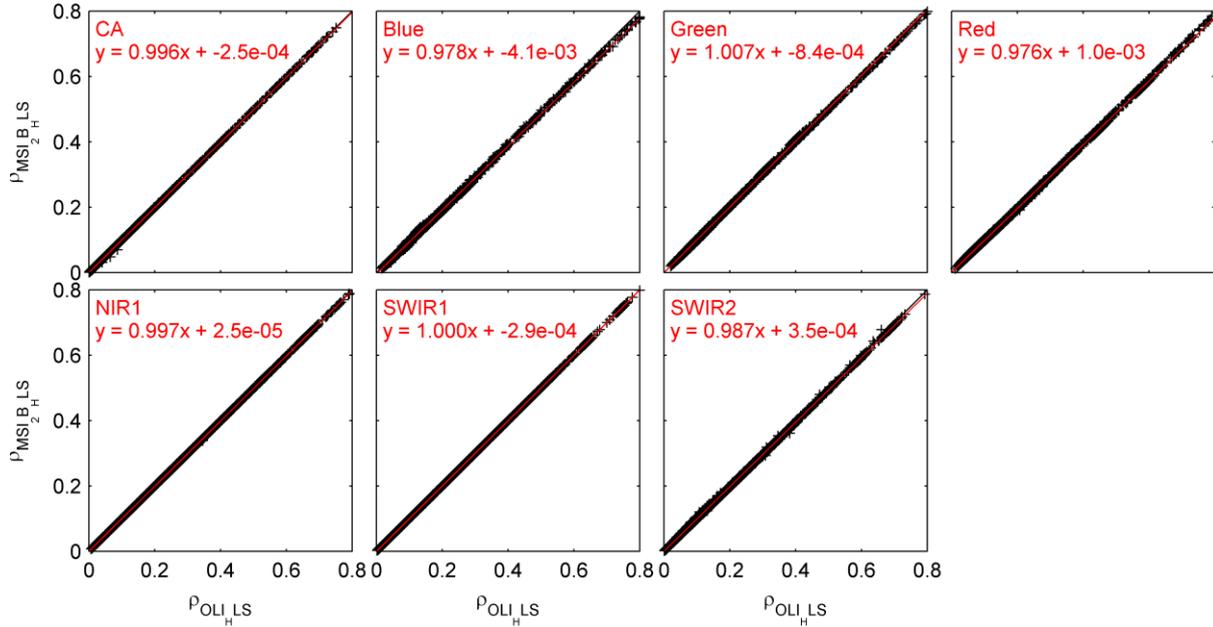


Figure 3: Same as Figure 2, but for Sentinel-2B.

$$\rho_{OLI} = a \times \rho_{MSI} + b \quad (5)$$

Table 5: Bandpass adjustment coefficients

HLS Band name	OLI band number	MSI band number	Sentinel-2A		Sentinel-2B	
			Slope (a)	Intercept (b)	Slope (a)	Intercept (b)
CA	1	1	0.9959	-0.0002	0.9959	-0.0002
BLUE	2	2	0.9778	-0.004	0.9778	-0.004
GREEN	3	3	1.0053	-0.0009	1.0075	-0.0008
RED	4	4	0.9765	0.0009	0.9761	0.001
NIR1	5	8A	0.9983	-0.0001	0.9966	0.000
SWIR1	6	11	0.9987	-0.0011	1.000	-0.0003
SWIR2	7	12	1.003	-0.0012	0.9867	0.0004

4.6 Spatial resampling

Resampling in S30 generation is the simple area-weighted average because of the nesting relationship between the 10m, 20m, 60m and 30m pixels. To produce an S30 pixel, a set of 3x3 10m pixels are averaged with equal weights, four 20m pixels are averaged with weights 4/9, 2/9, 2/9, and 1/9, and a 60m pixel is duplicated to produce 2x2 S30 pixels.

Additional resampling is applied to surface reflectance generated from LIC from processing baseline prior to 2.04 (approximately mid-2016). Cubic convolution is applied to resample the 10m/20m/60m spectral bands, as part of the AROP-based temporal coregistration, preceding the above area-weighted average. Only a small proportion of images are involved.

With the improved geolocation in Landsat-8 Collection-2, AROP-based temporal coregistration has phased out. Spatial resampling of Landsat-8 is still needed in gridding the data into the MGRS used by Sentinel-2, because even if the Landsat image and the intended MGRS tile are in the same UTM zone, the UTM coordinate origin corresponds to a corner of a Sentinel-2 pixel, but the center of a Landsat pixel. Cubic convolution is used.

Resampling of the QA layer is implemented differently. QA layers are created during cloud masking, performed at 10 m over the MGRS grid for Sentinel-2 and at 30 m over the WRS path-row grid for Landsat. The resample of the QA is performed for each bit in turn.

- S30: 10 m QA are resampled to 30m using a “presence” rule. A QA bit 1 in any of the 9 nesting 10m pixels causes the output bit set to 1 for S30. That is, a QA bit for S30 is set to 0 if only all the nine 10m pixels bit values are 0.
- L30: while a 4x4 window is used in cubic convolution resampling of the spectral data, only the innermost 2x2 pixels are examined for QA resampling because almost all the cubic convolution weights are in the 2x2 window. The “presence” rule is used; a QA bit for an L30 pixel is set to 0 only if all the 2x2 pixels have 0 at the corresponding bit.

Because a resampling window may contain a mixture of QA bits, the “presence” rule may make the output QA bits not mutually exclusive. For example, the QA bits may indicate a pixel is cloud, cloud shadow and water at the same time. This is not a mistake; it allows the user to discard as much suspicious data as possible.

5 Spatial coverage

HLS v1.5 covers all the global land except Antarctica, as depicted in a land mask (Fig. 5) derived from a 10m resolution coastline dataset (<https://www.naturalearthdata.com/downloads/10m-physical-vectors/10m-land/>). Note that the data acquisition over some small oceanic islands by the Landsat and Sentinel-2 sensor may not be made regularly. Antarctica is excluded because of low solar elevations which compromise the plane-parallel atmospheric correction.



Figure 4: HLS v1.5 covers the global land, including major islands but excluding Antarctica.

6 Product formats

6.1 File format

HLS products are distributed in Cloud Optimized GeoTIFF (COG), one file per data layer to offer the flexibility of only downloading the needed data layers and, for cloud-based applications, the needed spatial subsets within a tile. The COG files are internally compressed.

For S30, the files are saved in a directory such as

`HLS.S30.T17SLU.2020117T160901.v1.5/`,

which indicates S30 over tile 17SLU from data acquired on day 117 of 2020, and acquisition time UTC 160901; the data version is 1.5. The individual COG files can be found in the directory:

```
HLS.S30.T17SLU.2020117T160901.v1.5.B01.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B02.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B03.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B04.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B05.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B06.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B07.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B08.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B8A.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B09.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B10.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B11.tif
HLS.S30.T17SLU.2020117T160901.v1.5.B12.tif
HLS.S30.T17SLU.2020117T160901.v1.5.Fmask.tif
HLS.S30.T17SLU.2020117T160901.v1.5.SZA.tif
HLS.S30.T17SLU.2020117T160901.v1.5.SAA.tif
HLS.S30.T17SLU.2020117T160901.v1.5.VZA.tif
HLS.S30.T17SLU.2020117T160901.v1.5.VAA.tif
HLS.S30.T17SLU.2020117T160901.v1.5.cmr.xml
HLS.S30.T17SLU.2020117T160901.v1.5.json
HLS.S30.T17SLU.2020117T160901.v1.5.jpg
```

The filenames for individual spectral bands and Fmask cloud mask are self-explaining. Sun zenith angle (SZA), sun azimuth angle (SAA), view zenith angle (VZA) and view azimuth angle (VAA) file are also provided; see Section 6.4 for details. The `HLS.S30.T17SLU.2020117T160901.v1.5.cmr.xml` is the metadata file, `HLS.S30.T17SLU.2020117T160901.v1.5.json` contains the size and checksum value of each file, and `HLS.S30.T17SLU.2020117T160901.v1.5.jpg` is a natural-color browse image.

Note that during the early years of Sentinel-2 operation the acquisition time information reported in the Sentinel-2 L1C product may not be the sensing time exactly over the tile, but the time when the sensor began to sense the sun-lit side of the earth for each orbit. When several observations are available on the same day at the high latitude, this time can still separate the observations in a sequence for the day.

L30 data are stored in the same format. An example product consists of the following files:

```
HLS.L30.T17SLU.2020209T155956.v1.5.B01.tif
HLS.L30.T17SLU.2020209T155956.v1.5.B02.tif
```

HLS.L30.T17SLU.2020209T155956.v1.5.B03.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.B04.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.B05.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.B06.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.B07.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.B09.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.B10.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.B11.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.Fmask.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.SZA.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.SAA.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.VZA.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.VAA.tif
 HLS.L30.T17SLU.2020209T155956.v1.5.cmr.xml
 HLS.L30.T17SLU.2020209T155956.v1.5.json
 HLS.L30.T17SLU.2020209T155956.v1.5.jpg

The UTC time in the filenames are the sensing time at the scene center of the input Landsat-8 scene. If two scenes overlap a MGRS tile, the sensing time of one the scenes is chosen by chance. So, this timing information is not accurate for the tile center; it is intended mainly as an identifier.

6.2 S30

The product S30 contains MSI surface reflectance at 30 m spatial resolution. Table 6 lists all the data layers of the S30 product.

Table 6: list of the SDS of the S30 product (SR = Surface Reflectance, NBAR = Nadir BRDF-Adjusted Reflectance, TOA Refl. = Top of Atmosphere Reflectance).

Data layer	MSI band number	Units	Data type	Scale	Fill value	Spatial Resolution	Description
B01	1	reflectance	int16	0.0001	-9999	30	NBAR
B02	2	reflectance	int16	0.0001	-9999	30	
B03	3	reflectance	int16	0.0001	-9999	30	
B04	4	reflectance	int16	0.0001	-9999	30	
B05	5	reflectance	int16	0.0001	-9999	30	
B06	6	reflectance	int16	0.0001	-9999	30	
B07	7	reflectance	int16	0.0001	-9999	30	
B08	8	reflectance	int16	0.0001	-9999	30	
B8A	8A	reflectance	int16	0.0001	-9999	30	
B09	9	reflectance	int16	0.0001	-9999	30	
B10	10	reflectance	int16	0.0001	-9999	30	
B11	11	reflectance	int16	0.0001	-9999	30	NBAR
B12	12	reflectance	int16	0.0001	-9999	30	
FMASK (Table)	-	none	uint8	-	255	30	Quality bits

6.3 L30

The product L30 contains Landsat-8 OLI surface reflectance and TOA TIRS brightness temperature gridded at 30 m spatial resolution in MGRS tiles. Table 7 lists all the data layers of the L30 product.

Table 7: All the data layers of the L30 product (SR = Surface Reflectance, NBAR = Nadir BRDF-normalized Reflectance, TOA Refl. = Top of Atmosphere Reflectance, TOA BT = Top of Atmosphere Brightness temperature).

Data layer	OLI band number	Units	Data type	Scale	Fill value	Spatial Resolution	Description
B01	1	reflectance	int16	0.0001	-9999	30	NBAR
B02	2	reflectance	int16	0.0001	-9999	30	
B03	3	reflectance	int16	0.0001	-9999	30	
B04	4	reflectance	int16	0.0001	-9999	30	
B05	5	reflectance	int16	0.0001	-9999	30	
B06	6	reflectance	int16	0.0001	-9999	30	
B07	7	reflectance	int16	0.0001	-9999	30	
B09	9	reflectance	int16	0.0001	-9999	30	TOA Refl.
B10	10	degree °C	int16	0.01	-9999	30	TOA BT
B11	11	degree °C	int16	0.01	-9999	30	
FMASK (Table 11)	-	none	uint8	-	255	30	Quality bits

6.4 The sun and view angles

HLS v1.5 also provides the sun zenith/azimuth and view zenith/azimuth angles used in BRDF correction. The S30 angle data is interpolated from the ESA-provided 5km angles in a text form; HLS selects the view angle of the 2nd red-edge band and uses it on all bands. The L30 angle data is provided in the Collection-2 data; it is originally derived for the red band and is representative of all bands.

Table 8: Description of the sun and view angles.

Angle band	Units	Data type	Scaling factor	Fill value	Spatial resolution
Sun zenith	degrees	uint16	0.01	40,000	30m
Sun azimuth	degrees	uint16	0.01	40,000	30m
View zenith	degrees	uint16	0.01	40,000	30m
View azimuth	degrees	uint16	0.01	40,000	30m

6.5 Quality Assessment layer

HLS 1.5 products have one Quality Assessment (QA) layer, generated from Fmask 4.2 and named after Fmask. The Fmask integer output is converted to the bit representation (Table 10) as in HLS v1.4. The HLS processing dilates the Fmask cloud and cloud shadow by 5 pixels for S30 and L30 and labels the dilation as “Adjacent to cloud/shadow.”

Table 11: Description of the bits in the one-byte Quality Assessment layer. Bits are listed from the MSB (bit 7) to the LSB (bit 0)

Bit number	QA description	Bit value	Description
7-6	Not used	Not used	Not used
5	Water	1	Yes
		0	No
4	Snow/ice	1	Yes
		0	No
3	Cloud shadow	1	Yes
		0	No
2	Adjacent to cloud/shadow	1	Yes
		0	No
1	Cloud	1	Yes
		0	No
0	Cirrus	Reserved, but not used	NA

See Appendix A on how to decode the QA bits with simple integer arithmetic.

6.6 Metadata

Metadata about the S30 and L30 products is presented in the xmr.xml file.

6.5.1 Key metadata elements for S30 from the complete XML list include:

- **PRODUCT_URI**
The input L1C granule URI, for processing backtracing
- **SENSING_TIME**
Sensing time at the center of the granule, or at the start of the datatake for earlier L1C data
- **SPATIAL_COVERAGE**
The area percentage of the tile with data
- **CLOUD_COVERAGE**
The percentage of cloud and cloud shadow in observation based on Fmask
- **ULX and ULY**
The UTM X/Y coordinate at the upperleft corner of the tile
- **SPATIAL_RESAMPLING_ALG**
Algorithm used in resampling 10m/20m/60m data to 30m: area weighted average. For L1C data prior to baseline 2.04, cubic convolution is used in coregistration.
- **ADD_OFFSET**
Value added to the scaled int16 reflectance data. Should be 0.
- **REF_SCALE_FACTOR**
Multiplier applied to the scaled int16 reflectance data
- **FILLVALUE**
Pixel value in the spectral bands where no observation was made
- **QA_FILLVALUE**
The pixel value in Fmask where no observation was made
- **MEAN_SUN_AZIMUTH_ANGLE**
The mean solar azimuth in the tile
- **MEAN_SUN_ZENITH_ANGLE**
The mean solar zenith in the tile
- **MEAN_VIEW_AZIMUTH_ANGLE(B01)**
The mean view azimuth in B01. Should be representative of all spectral bands.
- **MEAN_VIEW_ZENITH_ANGLE(B01)**
The mean view azimuth in B01. Should be representative of all spectral bands.
- **NBAR_SOLAR_ZENITH**
The solar zenith angle used in NBAR derivation. It can be the same as MEAN_SUN_ZENITH_ANGLE for very high latitude.
- **MSI_BAND_01_BANDPASS_ADJUSTMENT_SLOPE_AND_OFFSET**
The slope and offset applied to the Sentinel-2 B01 reflectance in the linear bandpass adjustment
- **MSI_BAND_02_BANDPASS_ADJUSTMENT_SLOPE_AND_OFFSET**
- **MSI_BAND_03_BANDPASS_ADJUSTMENT_SLOPE_AND_OFFSET**
- **MSI_BAND_04_BANDPASS_ADJUSTMENT_SLOPE_AND_OFFSET**
- **MSI_BAND_11_BANDPASS_ADJUSTMENT_SLOPE_AND_OFFSET**
- **MSI_BAND_12_BANDPASS_ADJUSTMENT_SLOPE_AND_OFFSET**

- MSI_BAND_8A_BANDPASS_ADJUSTMENT_SLOPE_AND_OFFSET
- AROP_AVE_XSHIFT(METERS)
 - The average shift in the X direction derived by AROP to register earlier L1C to a reference image. After L1C processing baseline 2.04, AROP is not applied and all the AROP related metadata elements are set to 0.
- AROP_AVE_YSHIFT(METERS)
- AROP_NCP
- AROP_RMSE(METERS)
- AROP_S2_REFIMG
- ANGLEBAND
 - For some earlier L1C data, the view angle information can be missing for some spectral bands. Therefore a substitute band was used in NBAR derivation. This 0-based 13-element array indicates the ID of the substitute band for each band. No substitution is needed for later versions of L1C.
- ACCODE
 - The version of LaSRC used by HLS
- IDENTIFIER_PRODUCT_DOI
 - This S30 product's DOI.

6.5.2 Key metadata elements for L30 from the complete XML list include:

- LANDSAT_PRODUCT_ID
 - The Landsat-8 input L1 scene product ID for processing backtracing. For example, if the user only wants Tier-1 based HLS products, he/she can make the selection based on this metadata element. If two scenes of the same WRS path overlap the MGRS tile, two product IDs are reported.
- SENSING_TIME
 - The carried over scene center sensing time; not precisely represented the data gridded into the tile. When two scenes overlap the tile, the sensing time for one of them was chosen by chance.
- SPATIAL_COVERAGE
 - The percentage of the tile with data
- CLOUD_COVERAGE
 - The percentage of cloud and cloud shadow in the data
- SPATIAL_RESAMPLING_ALG
 - Resampling algorithm in gridding Landsat data into the tile. Cubic convolution. Even if there is no UTM zone change, resampling is still needed because the UTM coordinate origin correspond to a corner of a Sentinel-2 L1C pixel but the center of a Landsat L1 pixel. UTM zone.
- ULX and ULY
 - The UTM X/Y coordinate at the upperleft corner of the tile
- ADD_OFFSET
 - See above for S30.
- REF_SCALE_FACTOR
 - See above for S30.
- THERM_SCALE_FACTOR
 - Multiplier to be applied to the thermal bands to get temperature in Celsius.
- FILLVALUE
 - See above for S30.

- QA_FILLVALUE
See above for S30.
- MEAN_SUN_AZIMUTH_ANGLE
The mean solar azimuth in the tile
- MEAN_SUN_ZENITH_ANGLE
The mean solar zenith in the tile
- NBAR_SOLAR_ZENITH
The solar zenith angle used in NBAR derivation. It can be the same as MEAN_SUN_ZENITH_ANGLE for very high latitude.
- ACCODE
The version of LaSRC used by HLS
- TIRS_SSM_MODEL
Metadata carried over from Landsat L1 data, indicating the quality of the thermal data
- TIRS_SSM_POSITION_STATUS
Metadata carried over from Landsat L1 data, indicating the quality of the thermal data
- IDENTIFIER_PRODUCT_DOI
This L30 product's DOI.

7 Known issues (TBD)

References

- Claverie, M., Vermote, E., Franch, B., & Masek, J. (2015). Evaluation of the Landsat-5 TM and Landsat-7 ETM + surface reflectance products. *Remote Sensing of Environment*, 169, 390-403.
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., Skakun, S.V., & Justice, C.O. (2018). The Harmonized Landsat and Sentinel-2 surface reflectance data set, in press, *Remote Sensing of Environment*.
- Doxani, G., Vermote, E., Roger, J. C., Gascon, F., Adriaensen, S., Frantz, D., ... & Louis, J. (2018). Atmospheric correction inter-comparison exercise. *Remote Sensing*, 10(2), 352.
- Drusch, M. et al. (2012) Sentinel-2: ESA's optical high-resolution mission for GMES operational services, *Remote Sensing of Environment*, 120, 25-36.
- ESA (2018). Sentinel-2 Data Quality Report S2-PDGS-MPC-DQR.
- Franch, B., Vermote, E.F., Claverie, M., (2014a). Intercomparison of Landsat albedo retrieval techniques and evaluation against in situ measurements across the US SURFRAD network. *Remote Sensing of Environment*, 152, 627-637.
- Franch, B., Vermote, E. F., Sobrino, J. A., & Julien, Y. (2014b). Retrieval of surface albedo on a daily basis: Application to MODIS data. *IEEE Transactions on Geoscience and Remote Sensing*, 52(12), 7549-7558.
- Gao, F., Masek, J.G., & Wolfe, R.E. (2009). Automated registration and orthorectification package for Landsat and Landsat-like data processing. *Journal of Applied Remote Sensing*, 3(1), 033515.

Irons, J.R., Dwyer, J.L, and J. Barsi (2012) The next Landsat satellite: The Landsat Data Continuity Mission, *Remote Sensing of Environment*, 122,11-21, 10.1016/j.rse.2011.08.026

Li, Z., Zhang, H.K., Roy, D.P., 2018, Investigation of Sentinel-2 bidirectional reflectance hot-spot sensing conditions, *IEEE Transactions on Geoscience and Remote Sensing*, 10.1109/TGRS.2018.2885967. (<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8594675>)

Masek, J. G., Vermote, E. F., Saleous, N. E., Wolfe, R., Hall, F. G., Huemmrich, K. F., ... & Lim, T. K. (2006). A Landsat surface reflectance dataset for North America, 1990-2000. *IEEE Geoscience and Remote Sensing Letters*, 3(1), 68-72.

Qiu S., Zhu Z., and He B., Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4-8 and Sentinel-2 imagery, *Remote Sensing of Environment*, (2019), doi.org/10.1016/j.rse.2019.05.024

Roy, D. P., Li, J., Zhang, H. K., Yan, L., Huang, H., & Li, Z. (2017). Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 199, 25-38.

Roy, D.P., Zhang, H.K., Ju, J., Gomez-Dans, J.L., Lewis, P.E., Schaaf, C.B., Sun, Q., Li, J., Huang, H., & Kovalskyy, V. (2016). A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, 176, 255-271.

Roy, D.P., Li, Z., Zhang, H.K., 2017, Adjustment of Sentinel-2 multi-spectral instrument (MSI) red-edge band reflectance to nadir BRDF adjusted reflectance (NBAR) and quantification of red-edge band BRDF effects, *Remote Sensing*, 9(12), 1325. (<http://www.mdpi.com/2072-4292/9/12/1325>)

Schaaf, C. B., Gao, F., Strahler, A. H., Lucht, W., Li, X., Tsang, T., ... & Lewis, P. (2002). First operational BRDF, albedo nadir reflectance products from MODIS. *Remote Sensing of Environment*, 83(1-2), 135-148.

Shuai, Y., Masek, J. G., Gao, F., & Schaaf, C. B. (2011). An algorithm for the retrieval of 30-m snow-free albedo from Landsat surface reflectance and MODIS BRDF. *Remote Sensing of Environment*, 115(9), 2204-2216.

Skakun, S., Roger, J. C., Vermote, E. F., Masek, J. G., & Justice, C. O. (2017). Automatic sub-pixel co-registration of Landsat-8 Operational Land Imager and Sentinel-2A Multi-Spectral Instrument images using phase correlation and machine learning based mapping. *International Journal of Digital Earth*, 10(12), 1253-1269.

Storey, J., Choate, M., & Lee, K. (2014). Landsat 8 Operational Land Imager On-Orbit Geometric Calibration and Performance. *Remote Sensing*, 6, 11127-11152

Storey, J., Roy, D. P., Masek, J., Gascon, F., Dwyer, J., & Choate, M. (2016). A note on the temporary misregistration of Landsat-8 Operational Land Imager (OLI) and Sentinel-2 Multi Spectral Instrument (MSI) imagery. *Remote Sensing of Environment*, 186, 121-122.

Strahler, A.H., Lucht, W., Schaaf, C.B., Tsang, T., Gao, F., Li, X., Lewis, P., & Barnsley, M. (1999). MODIS BRDF/Albedo Product: Algorithm Theoretical Basis Document Version 5.0. In M. documentation (Ed.). Boston.

Vermote, E., Justice, C. O., & Bréon, F. M. (2009). Towards a generalized approach for correction of the BRDF effect in MODIS directional reflectances. *IEEE Transactions on Geoscience and Remote Sensing*, 47(3), 898-908.

Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46-56.

Vermote, E. F., & Kotchenova, S. (2008). Atmospheric correction for the monitoring of land surfaces. *Journal of Geophysical Research: Atmospheres*, 113(D23).

Yan, L., Roy, D. P., Li, Z., Zhang, H. K., & Huang, H. (2018). Sentinel-2A multi-temporal misregistration characterization and an orbit-based sub-pixel registration methodology. *Remote Sensing of Environment*, 215, 495-506.

Zhang, H.K., Roy, D.P., & Kovalskyy, V. (2016). Optimal Solar Geometry Definition for Global Long-Term Landsat Time-Series Bidirectional Reflectance Normalization. *IEEE Transactions on Geoscience and Remote Sensing*, 54, 1410-1418.

Zhu, Z., Wang, S., & Woodcock, C.E. (2015). Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. *Remote Sensing of Environment*, 159, 269-277.

DRAFT

Acknowledgment

We thank Feng Gao for providing and spending many hours adapting the AROP code for HLS with quick turnaround. We also thank Jan Dempewolf for offering his Python script which works around the GDAL-incompatible issue in HLS v1.3 and it has proved very useful for many people. We also thank Shuang Li, Min Feng, and Mark Broich for GDAL-test HLS v1.4.

Appendix A. How to decode the bit-packed QA

Quality Assessment (QA) encoded at the bit level provides concise presentation but is less convenient for users new to this format. This appendix shows how to decode the QA bits with simple integer arithmetic and no explicit bit operation at all. An analogy in the decimal system illustrates the idea. Suppose we want to get the digit of the hundreds place of an integer 3215. First divide the integer by 10^2 (i.e. 100) to get an integer quotient 32, then the digit of the ones place (the least significant digit) of the quotient is what we want. By computing $32 - ((32 / 10) * 10)$, we get 2, the digit in the hundreds place of 3215. (Note that in integer arithmetic $32/10$ evaluates to 3.) The same idea applies to binary integers. Suppose we get a decimal QA value 100, which translates into binary 01100100, indicating that the aerosol level is low (bits 6-7), it is water (bit 5), and adjacent to cloud (bit 2). Suppose we want to find whether it is water, by examining the value of bit 5. It can be achieved in two steps:

- Divide 100 by 2^5 to get the quotient, 3 in this case for integer arithmetic
- Find the value of the least significant bit of the quotient by computing $3 - ((3/2) * 2)$, which is 1

The pixel is water based on the QA byte. Note that Step 2 above is essentially an odd/even number test. All the bits can be decoded with a loop.