Harmonized Landsat/Sentinel-2 Reflectance Dataset

October 2, 2020

Contents

1 Introduction

This User GuideATBD references the Provisional HLS Version 1.5 data, which have not been validated for their science quality and should not be used in science research or applications.

Many land monitoring applications require more frequent observations than can be obtained from a single "Landsat-class" sensor. Examples of these applications include crop type and condition monitoring, vegetation phenology, disaster response, and surface water quality. These and numerous other applications all require near-daily imagery at medium spatial resolution.

Since trying to achieve near-daily observation frequency from a single, well-calibrated satellite system may be prohibitively expensive, a reasonable alternative is to combine data from multiple international sources. The Harmonized LandsatSentinel-2 (HLS) Project supports this goal by generating a harmonized surface reflectance product from Landsat 8 and Sentinel-2 inputs. Specifically, by "harmonized" we mean that the products are:

- Gridded to a common pixel resolution, projection, and spatial extent (i.e., tile);
- Atmospherically corrected to surface reflectance using a common radiative transfer algorithm ;
- Normalized to a common nadir-view geometry via Bi-directional Reflectance Distribution Function (BRDF) estimation;
- Adjusted to represent the response from a common spectral bandpass.

HLS generates two tiled image products at 30-meter spatial resolution, "L30" and "S30", derived from individual Landsat-8 and Sentinel-2, acquisitions, respectively. These products constitute a "stackable" Analysis Ready Data (ARD) format, such that a user may examine any given pixel through time, and treat the near-daily reflectance time series as though it came from a single sensor.

2 Historical Perspective

The Landsat series of satellites have provided land imagery of the globe since 1972. Each satellite provides 16-day repeat coverage of global land areas. Landsat 8, launched in February 2013, introduced significant improvements in the number of spectral bands and the radiometric quality of the data. Landsat 8 includes two instruments: the Operational Land Imager (OLI) covering the visiblenearshortwave reflective bands, and the Thermal Infrared Sensor (TIRS) covering the longwave (thermal) infrared bands. The European Sentinel-2 satellites (part of the Copernicus environmental monitoring

service) provide a similar type of imagery to Landsat, but with additional spectral bands and higher resolution via its Multispectral Imager (MSI) instrument. Sentinel-2a was launched in 2015, and Sentinel-2b in 2017. Each Sentinel-2 satellite provides global land coverage every 10 days, or every 5 days for the two-satellite constellation. Combining the data from Landsat 8, Sentinel-2a, and Sentinel-2b provides some 100 observations per year for equatorial regions, and over 200 acquisitions per year in mid-latitudes and regions of orbit overlap (Li and Roy, 2017).

While similar, Landsat 8 and Sentinel-2 are not identical. The table below lists some of the differences between the systems. HLS attempts to adjust each product to create a harmonized "Landsat like" reflectance image.

Orbital altitude	705 km	786 km
Equatorial crossing time (MLT)	10:00 am	10:30 am
Image swath \ view angle field of view	180km \ 15 deg FOV	290 km \ 20.1 deg FOV
Spatial resolution (VISSWIRTIR)	3030100m	1020m (no TIR)
Spectral bands	9 VSWIR, 2 TIR	12 VSWIR
Ground track repeat	16 days	10 days (per platform)

3 Algorithm Description

3.1 Scientific Theory

HLS uses a processing chain involving several separate radiometric and geometric adjustments, with a goal of eliminating differences in retrieved surface reflectance arising solely from differences in instrumentation. The overall chain is shown in Fig. 1. Input data products from Landsat 8 (Collection 2 Level 1T top-of-atmosphere reflectance or top-of-atmosphere apparent temperature) and Sentinel-2 (L1C top-of-atmosphere reflectance) are ingested for HLS processing. A series of radiometric and geometric corrections are applied as described below to convert data to surface reflectance, adjust for BRDF differences, and adjust for spectral bandpass differences.

Three types of products are then generated: "S10" products – atmospherically corrected Sentinel-2 images in their native resolution and geometry; and the harmonized products "S30" and "L30". These products have been radiometrically harmonized to the maximum extent, and then gridded to a common 30-meter UTM basis using the Sentinel-2 tile system. Note that S10 products are not normally archived. The S30 and L30 products are resampled as needed to a common 30-meter resolution UTM projection, and tiled using the Sentinel-2 Military Grid Reference System (MGRS) UTM grid.

Below we describe the algorithms use for (1) atmospheric correction; (2) BRDF adjustment; (3) bandpass normalization; and (4) geometric processing. More detailed descriptions can be found in Claverie et al. (2018); we note below cases where algorithms have been changed or updated from that reference .

1. Atmospheric Correction

HLS relies on the operational LaSRC (Landsat Surface Reflectance Correction) algorithm for atmospherically correcting top-of-atmosphere reflectance to surface reflectance. As described in Vermote et al., (2016), the LaSRC approach relies on the inversion of the relatively simple equation in the Lambertian case, with no adjacency effects that account for a simplified coupling of the absorption by atmospheric gases and scattering by molecules and aerosols as it is implemented in the 6SV radiative transfer code (Vermote et al. 1997b, Kotchenova et al. 2006):

$$\rho_{TOA}\left(\theta_S, \theta_\nu, \psi, P, \overbrace{\tau_A, \omega_0, P_A}^{Aerosols}, U_{H_2O}, U_{O_3}\right) = Tg_{OG}\left(m, P\right) Tg_{O_3}\left(m, U_{U_{O_3}}\right) \Big[\rho_{atm}\left(\theta_S, \theta\nu, \psi, P, \tau_A, \omega_0, \right. \quad (1)$$

where ρ_{TOA} is the reflectance at the top of the atmosphere, ρ_{atm} is the atmosphere intrinsic reflectance, Tr_{atm} is the total atmosphere transmission (downward and upward), S_{atm} is the atmosphere spherical albedo, and ρ_{s} is the surface reflectance to be retrieved by the atmospheric correction procedure: the geometric conditions are described by the solar zenith angle qs, the view zenith angle qv, and the relative azimuth f (or the difference between the solar and view azimuth angles); P is the pressure that influences the number of molecules and the concentration of absorbing gases in the atmosphere, Tg designates the gaseous transmission by water vapor (Tg_{H2O}), ozone (Tg_{O3}), or other gases (Tg_{OG}), U_{H2O} is the integrated water vapor content, U_{O3} is the integrated ozone content, and m is the so-called "air-mass" computed as $1\cos(qs) + 1\cos(qv)$; t_{A}, w_0 and P_{A} describe the aerosol properties and are spectrally dependent: τ_{a} is the aerosol optical thickness, w_0 is the aerosol single scattering albedo, and P_{A} is the aerosol phase function.

The main limitation of 6SV is the plane parallel assumption for the atmosphere, which limits the quoted accuracy (0.4%) to Sun and view zenith angles lower than 75 degrees. In addition to this limitation, Equation [1] supposes that scattering and absorption could be decoupled, which is not true where both strong absorption and scattering regimes occur for example near strong water vapor absorption lines in the near-infrared (e.g. in MODIS band 18 and 19).

The atmospheric pressure P is obtained from a combination of data available from a coarse-resolution (one deg., six hours time step) weather prediction model available from NCEP GDAS (2015), sea level pressure P_{sl} and the altitude z [km] given by a Digital Elevation Model at 0.05 degree resolution

(ETOPO5, 1988) and computed as:

$$P = P_{sl} e^{-\frac{z}{8}} \tag{2}$$

The pressure should be representative of the average atmospheric pressure along the path from the Sun to the target and back to the satellite for primary scattering, and along an even more complicated path for multiple scattering.

Gaseous absorption terms include ozone and water vapor. The ozone amount U_{O3} is obtained via NCEP GDAS (2015) (at 1deg., 6 hours time step) via the ancillary information included in the MODIS surface reflectance Climate Modeling Grid (MOD09CMA). The surface reflectance Climate Modeling Grid (CMG) adopted a simple Geographic latitudelongitude projection at 0.05 degree ($\sim$ 5.5 km). The water vapor is also extracted from the ancillary information included in the MODIS surface reflectance Climate Modeling Grid (MOD09CMA) for Terra, which is itself computed from the MODIS near-infrared band 18 (931-941 nm) and 19 (915-965 nm) at 1 km spatial resolution (Gao and Kaufman, 2003).

Estimating aerosol optical thickness and phase function remains the most challenging aspect of atmospheric correction. In LaSRC a simplifying assumption is that a single aerosol model ("urban clean", Dubovik, 2002) can adequately represent global aerosol distributions. Then the aerosol optical thickness is inverted using the two blue bands available on Landsat8OLI (band 1 and 2) and red band for each non-water pixel that has not been flagged as cirrus. The approach relies on empirical correlation between ratios of the blue and red bands and aerosol optical thickness observed from

MODIS, which has been used also as the basis for MODIS Collection 6 implementation.

The method for inverting the AOT from OLI or MSI is relatively simple if the ratio between the red and blue bands can be known for every 10-30m pixel. First, this ratio is computed at coarse resolution (0.05 degree) from 10 years of MODIS (Terra) and MISR data. The MISR AOT product is used as input to the atmospheric correction of the MODIS TOA data for each valid observation, providing the MODIS surface reflectances that can be used to derive a ratio at 0.05 degree. The data are carefully filtered for clouds and high AOT values. This data processing enables one to account for the accuracy of that ratio globally and across several years and seasons. The ratio is computed for each valid observation and subsequently fitted as a linear function of $\mathrm{NDVI}_{\mathrm{MIR}}$, a vegetation index analogous to NDVI that uses the Mid-IR (2.1µm) channel instead of Red. The per-pixel ratio at the OLIMSI resolution is then calculated from the 30m $\mathrm{NDVI}_{\mathrm{MIR}}$ values.

The version of the LaSRC atmospheric correction code used for HLS v1.5 is derived from a C-language implementation of LaSRC version 3.5.5. used operationally for Landsat processing at USGS EROS. The only significant change from the version documented in Vermote et al. (2016) is that the aerosol optical thickness is calculated on a coarser (1km) spacing in order to speed processing time.

2. Bi-directional Reflectance Distribution Function (BRDF) Correction

The relative view angle between a Landsat 8 OLI and Sentinel-2 MSI observation of a single ground target can be as great as $7.5 + 10.0 = 17.5 degrees$. This view angle difference is sufficient to generate several percent absolute reflectance difference for normal vegetation materials (Gao et al., 2009). The

HLS BRDF correction attempts to normalize the surface reflectance to an optimal nadir-view value.

HLS has opted to use the c-factor technique and global coefficients provided by Roy et al. (2016) because the technique is very stable, reversible, easy to implement for operational processing and has been evaluated for Sentinel-2 data (Roy et al., 2017). The c-factor technique uses fixed BRDF coefficients for each spectral band, i.e., a constant BRDF shape, derived from a large number of pixels in the MODIS 500 m BRDF product (MCD43) that are globally and temporally distributed (>15 billion pixels). The technique has been evaluated using ETM+ data off-nadir (i.e. on the overlap areas of adjacent swaths, Roy et al., 2016) and MSI data (Roy et al., 2017). The technique is applied in HLS on OLI and MSI bands equivalent to MODIS ones; MSI red-edge spectral bands are therefore not normalized. Normalized reflectance is calculated for original reflectance and a c-factor (Eq. 3). The latter is deduced (Eq. 3) from BRDF coefficients for the three kernels (isotropic, volumetric and geometric). The kernel definitions are described in the ATBD of the MOD43 product (Strahler et al., 1999), and the specific c-factor coefficients are provided in Roy et al. (2016) and Claverie et al. (2018).

$$\rho\left(\lambda, \theta^{Norm}\right) = c\left(\lambda\right) \times \rho\left(\lambda, \theta^{sensor}\right) \tag{3}$$

$$c\left(\lambda\right) = \frac{f_{iso}\left(\lambda\right) + f_{geo}\left(\lambda\right) \times K_{geo}\left(\theta_{Norm}\right) + f_{vol}\left(\lambda\right) \times K_{vol}\left(\theta_{Norm}\right)}{f_{iso}\left(\lambda\right) + f_{geo}\left(\lambda\right) \times K_{geo}\left(\theta_{Sensor}\right) + f_{vol}\left(\lambda\right) \times K_{vol}\left(\theta_{Sensor}\right)} \tag{4}$$

where θ_{Sensor} refers to the sun-illumination geometry configuration (i.e., θ_{v}, θ_{s}, $\Delta\phi$) of the input data and θ_{Norm} refers to the sun-illumination geometry configuration of the normalized data ($\theta_{\mathrm{v}} = 0$, $\theta_{\mathrm{s}} = \theta_{\mathrm{s\_out}}$, $\Delta\phi = 0$).

It should be noted that HLS v1.4 applied Equation 4 to correct both view and solar elevation angles, the latter an attempt to normalize for BRDF

changes associated with solar elevation changes during the growing season. However, further investigation and discussions suggested that the solar elevation correction was not appropriate for the simplified C-factor formulation. As a result, HLS v1.5 only corrects for view angle differences.

3. Bandpass Adjustments

The harmonization also requires adjustment of the small differences between the equivalent spectral bands of MSI and OLI. The OLI spectral bandpasses are used as reference, to which the MSI spectral bands are adjusted. No bandpass adjustment is defined for the (i) MSI red-edge bands (B05, B06 and B07), (ii) broad NIR band (B08), and (iii) atmospheric bands (B09 and B10). MSI bandpasses are based on the revised Sentinel-2a relative spectral responses (RSR's) for bands 1 and 2 provided by ESA in 2017.

As described in Claverie et al. (2018), the bandpass adjustment algorithm was derived from a selection of EO-1 Hyperion hyperspectral imager spectra. 160 million per-pixel spectra were extracted from a set of 158 Hyperion scenes, distributed globally by latitude, Using the RSR for OLI and MSI, the raw (MSI) and target (OLI) spectral reflectance values were calculate for each pixel from the hyperspectral spectra. A global linear regression was then developed to transform MSI spectral reflectance to "pseudo-OLI" spectral reflectance.

4. Geometric Processing

The S30 and L30 products are coregistered in the Sentinel-2 Military Grid Reference System (MGRS), partially resulting from the use of an HLS inter-

nal set of geolocation reference images when necessary. Since the processing baseline 2.04 initiated in June 2016 and up to the current 2.09 the Sentinel-2 L1C geolocation is quite stable, other than a few incidental anomalies, with long-term absolute accuracy close to 11 meters at 95% for both satellites (7 Apr 2020 L1C data quality report). This accuracy is sufficient for the HLS 30-meter pixel size. When ESA applies its precisely geolocated Global Reference Image (GRI) in late 2020, the Sentinel-2 geolocation accuracy will be further improved. Therefore, HLS does not adjust the geolocation of Sentinel-2 data of this period. However, the Sentinel-2 L1C data before processing baseline 2.04 can show large geolocation error, especially a yaw angle bias apparent at the swath edges of adjacent orbits. To mitigate this problem, the HLS processing system has selected for each tile a 10-meter near-infrared summer image of minimal cloud contamination from processing baseline 2.04 to build its own internal geolocation reference images to which images of earlier processing baselines are to be registered.

The coregistration is aided by the Automated Registration and Orthorectification Package (AROP) (Gao, Wolfe and Masek, 2009), which automatically identifies tie points and fits a coordinate transformation function between a target image and the reference image for the tile. The spectral measurement in the target image is resampled with the cubic convolution technique during coregistration and saved in native 102060m spatial resolutions for S10. In the subsequent production of S30, the 102060m pixels of S10 are resampled to 30 meters with a simple area-weighted average. The HLS internal set of geolocation reference images will continue to be used until ESA reprocesses all images before processing baseline 2.04 to a better quality.

Collection-1 Landsat-8 data often do not align with Sentinel-2 data and in general show lesser geolocation accuracy (Storey et al 2016). HLS v1.4 has applied AROP to register Landsat-8 data to the HLS internal Sentinel-2 based geolocation reference images and uses cubic convolution to resample the spectral data. With the use of ESA-provided Sentinel-2 GRI to improve the density and accuracy of Landsat ground control points, USGS will release Collection-2 Landsat data with a better geolocation accuracy in mid-

2020. AROP will not be needed on Collection-2 Landsat-8 data by that time, but resampling is still necessary because the UTM coordinate origin in the Landsat-8 system corresponds to a pixel center but in the Sentinel-2 system corresponds to a pixel corner.

5. Implementation

The HLS processing flow has been implemented on the University of AlabamaMarshall Space Flight Center (UAHMSFC) IMPACT cloud computing environment, utilizing Amazon Web Services (AWS) services.

Input data sources include:

- Landsat 8 Collection 2 "Real Time" TOA reflectance products from USGS EROS. To reduce HLS latency, IMPACT uses the "real time" Landsat 8 products rather than the final Tier 1 products. Users should note that this can introduce positional uncertainty in the Landsat 8 TIR observations compared to the final Tier 1 products.
- Sentinel-2 L1C TOA products. Input data for forward processing (2020 onwards) from the ESA International Hub; Input data for archival products (2015-2020) from the USGS EROS Sentinel-2 L1C mirror site.
- LaSRC atmospheric correction inputs include ozone concentration, water vapor, and atmospheric temperature from MODIS CMA products, and surface topography (for pressure calculation) based on the Global Climate Model DEM.

Data production is kicked off daily based on new input granules from the ESA or USGS archives. HLS latency is scaled by the availability of both input TOA imagery, as well as the availability of the atmospheric correction

inputs. HLS products are typically available within 2-3 days of image acquisition.

3.1.1 Assumptions

The following assumptions apply to the HLS algorithms:

LaSRC atmospheric correction assumes a plane-parallel atmosphere and Lambertian surface

BRDF correction is valid for small ranges of view angle (<20 degrees) near nadir

HLS does not attempt BRDF or bandpass corrections for bands that have no MODIS counterpart (e.g. Sentinel-2 red edge bands).

HLS uses "Real Time" Landsat 8 Collection 2 products to reduce latency; positional accuracy of the Landsat 8 TIR bands may be lower than Tier 1 data available 2 weeks later.

3.2 Mathematical Theory

3.2.1 Assumptions

3.3 Algorithm Input Variables

3.4 Algorithm Output Variables

4 Algorithm Implementations

5 Algorithm Usage Constraints

6 Performance Assessment Validation Methods

The HLS Project distinguishes between "Quality Assurance" and "Validation". Quality Assurance (or QA) provides per-granule or per-pixel information on the relative quality of the observation, as a flag for users to either use or discard that observation. Validation presents a quantitative assessment of product accuracy and uncertainty against and absolute reference.

Validation

HLS Surface Reflectance products have been validated in several ways. The LaSRC atmospheric correction has been validated by comparing Landsat 8 surface reflectance products with imagery corrected by the 6S radiative transfer model using aerosol optical thickness (AOT) derived from AERONET in-situ observations (Vermote et al., 2016). Since aerosols remain the primary source of uncertainty for retrieving surface reflectance from Landsat data, this comparison primarily tests the ability of the LaSRC algorithm to accurately retrieve AOT. Results indicate an overall uncertainty of 0.001-0.011 absolute reflectance, depending on band, with minimal dependence on target brightness (Fig. 1, Fig. 2), and similar performance for Sentinel-2. Complete

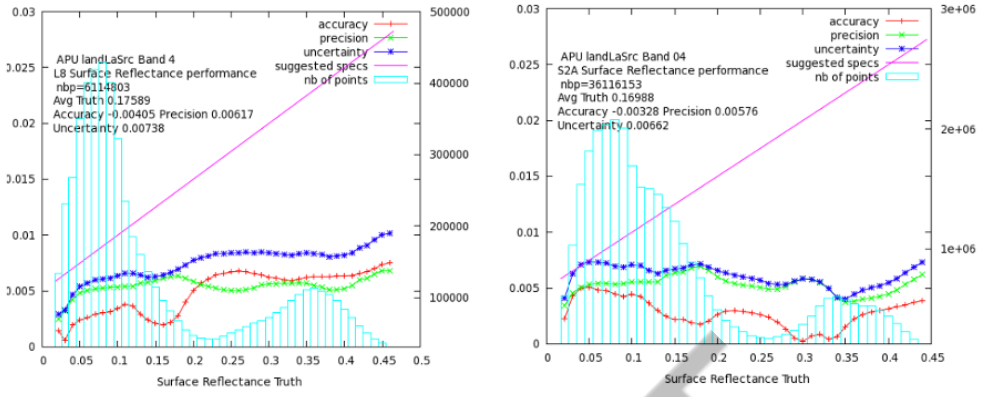

Figure 1: Accuracy (bias), precision, and uncertainty of the LaSRC atmospheric correction applied to Landsat 8 and Sentinel-2 data for the red band (Band 4).

results are presented in Vermote et al. (2016) and have also been included in the CEOS Atmospheric Correction Intercomparison Experiment (ACIX) (Doxani et al., 2018).

A second validation exercise focused on using ground-based albedometer networks (SURFRAD and OZFLUX) to compare HLS nadir-adjusted reflectance to observed albedo. There are several factors that must be considered when performing this comparison:

- Comparing directional reflectance with albedo requires a hemispheric integration using a known bi-directional reflectance distribution function (BRDF), as well as a spectral integration to compare the relatively narrow HLS band passes with the broader channels of the albedometer instruments.

- The view footprint of the SURFAD and OZFLUX instruments is much smaller than an HLS 30-meter pixel, so local heterogeneity beyond the albedome-

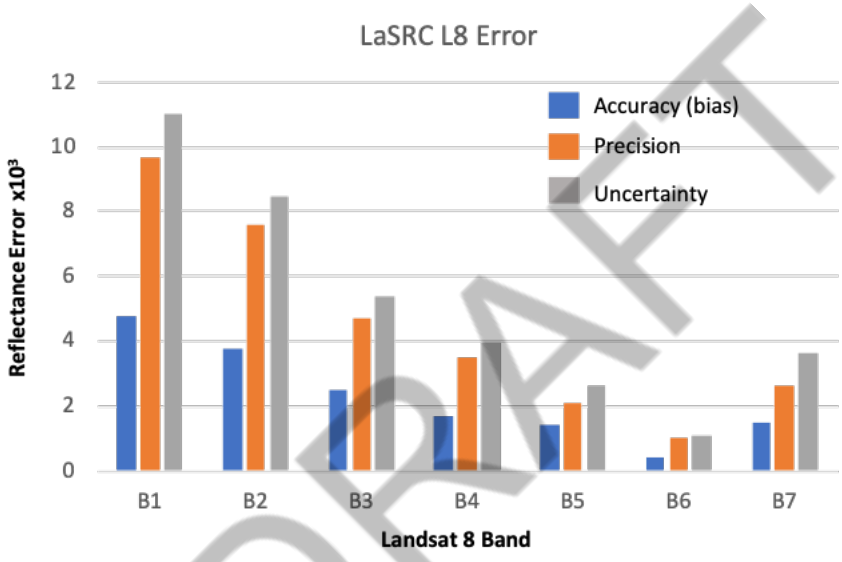

Figure 2: LaSRC accuracy, precision, uncertainty compared to Aeronet-derived surface reflectance for Landsat 8 OLI, data from Vermote et al. (2016). Units are absolute reflectance $\times 10^3$ (e.g. 10 = 1% reflectance or 0.01)

ter field of view will affect the comparison.

Results are presented in Franch et al. (2019) (Fig. 3). Using sites in both the US and Australia, albedo estimates using S30 and L30 products as input provide comparable RMSE (0.015 – 0.03 albedo).

Finally, Claverie et al. (2018) presented a theoretical error budget for the HLS product derived from published assessments of component errors from each of the algorithms (atmospheric correction, BRDF adjustment, spectral band pass correction). Rolling up the published component errors and assuming they are independent (e.g. total error is the root sum square of component errors) indicates per-band uncertainty of 0.01 to 0.02 absolute reflectance.

It should be noted that the above analyses compare HLS products to absolute reflectance or albedo estimates. Users may care more about the temporal stability (comparability) of reflectance from the S30 and L30 products, rather than the absolute error. Time series analysis over stable (e.g. invariant desert) sites are ongoing, but short-term variability from these sites is generally less than 0.5% absolute reflectance over a period of days.

Quality Assurance (QA)

Per-pixel QA information is included with each v1.5 HLS S30 or L30 image via the QA layer. Clouds and cloud shadow are identified using the

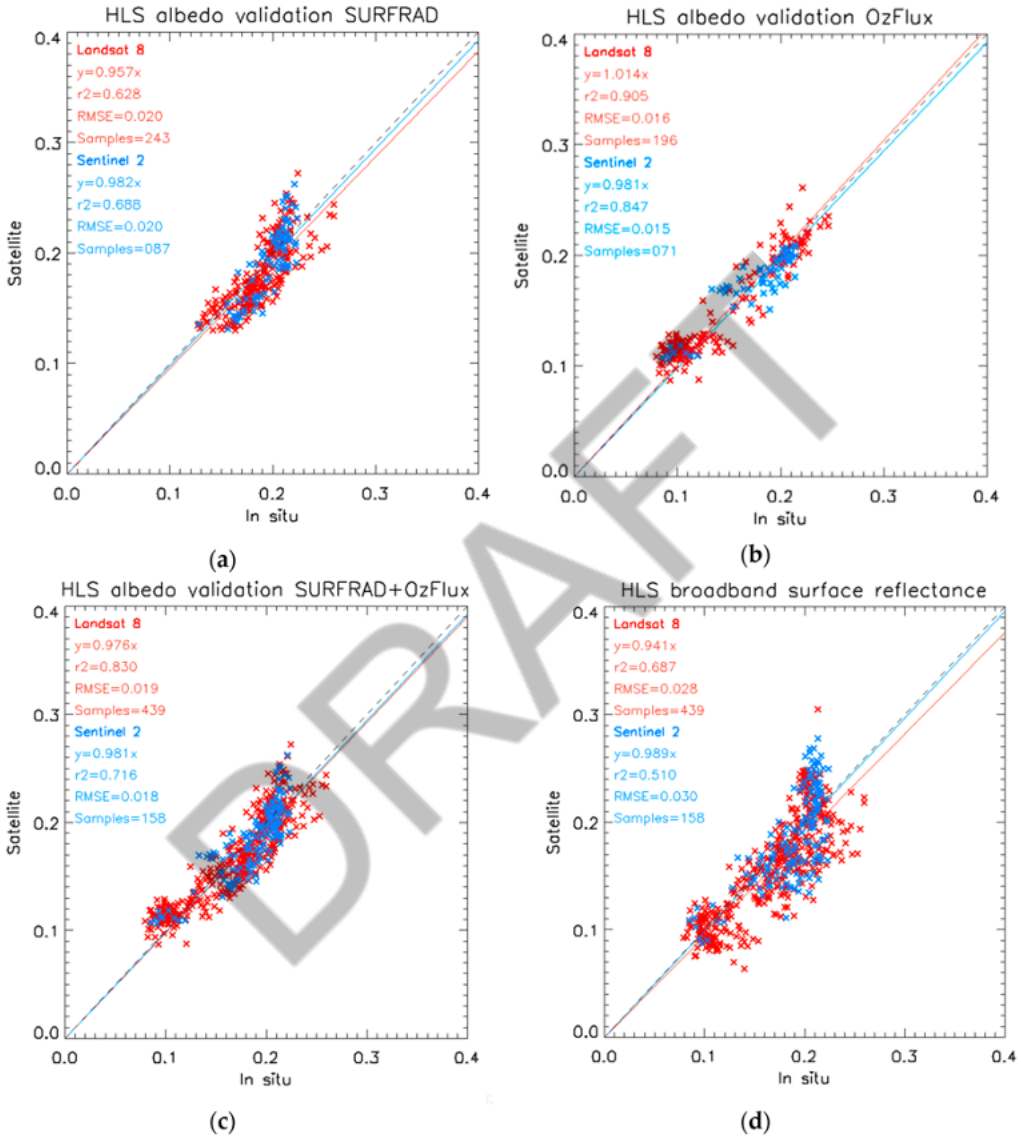

Figure 3: Comparison of SURFAD and OZFLUX albedometer measurements with HLS-derived albedo. From Franch et al. (2019)

Fmask 4.0 algorithm published by Zhu et al (2015).

In previous HLS versions, per-granule QA statistics were derived by comparing aggregated HLS reflectance values to near-simultaneous, cloud-free MODIS CMG reflectances (Claverie et al., 2018, Fig. 1). In general, this approach showed that HLS reflectance values were consistent with MODIS reflectance values, except for those cases where HLS cloud masking failed. Version 1.5 has discontinued the per-granule MODIS comparisons as the approach has not proved critical for flagging "poor" granules.

7 Performance Assessment Validation Uncertainties

8 Performance Assessment Validation Errors

9 Data Access Input Data

10 Data Access Output Data

10.1

HLSS30 dataset landing page

Access Url: https://lpdaac.usgs.gov/products/hlss30v015/

10.2

HLSL30 dataset landing page.

Access Url: https://lpdaac.usgs.gov/products/hlsl30v015/

11 Data Access Related URLs

11.1

HLS S30 imagery in GIBS. This link can be used to view HLS S30 imagery in the GIBS Worldview client (will be updated once data becomes available)

Url: https://GIBS_layer_S30.com

11.2

S30 Data Access URL. This URL can be used to directly access HLS S30 data

Url: https://search.earthdata.nasa.gov/search/granules?p=C1711924822-LPCLOUD&pg[0][gsk]=-start_date&q=HLS&m=-0.140625!0!2!1!0!0%2C2&tl=1582733901!4!!

11.3

L30 Data Access URL. This URL can be used to directly access HLS L30 data

Url: https://search.earthdata.nasa.gov/search/granules?p=C1711972753-LPCLOUD&pg[0][gsk]=-start_date&q=HLS&m=-0.140625!0!2!1!0!0%2C2&tl=1582733901!

4!!

11.4

S30 Collection metadata URL. This URL can be used to directly access the collection level metadata record for the S30 data product

Url: https://cmr.earthdata.nasa.gov/search/concepts/C1711924822-LPCLOUD.html?token=229aabe8b195cfe39ee6c58e45885832f98baeb7bf46463e93915c88c26d779a:OLpAZlE4HqIOMr0TYqg7UQ

11.5

L30 Collection metadata URL. This URL can be used to directly access the collection level metadata record for the L30 data product

Url: https://cmr.earthdata.nasa.gov/search/concepts/C1711972753-LPCLOUD.html?token=4d94c4278610e8bec249b1cfd623af4ff10867f54bdc6588979e4ccceb5faf51:OLpAZlE4HqIOMr0TYqg7UQ

11.6

HLS L30 imagery in GIBS. This link can be used to view HLS L30 imagery in the GIBS Worldview client (will be updated once data becomes available)

Url: https://GIBS_layer_L30.com

12 Contacts

12.1

Jeffrey Masek

12.1.1 Contact Mechanisms

jeffrey.g.masek@nasa.gov

12.2

Junchang Ju

12.2.1 Contact Mechanisms

junchang.ju@nasa.gov

12.3

Brian Freitag

12.3.1 Contact Mechanisms

brian.m.freitag@nasa.gov

References