Algorithm Theoretical Basis Document

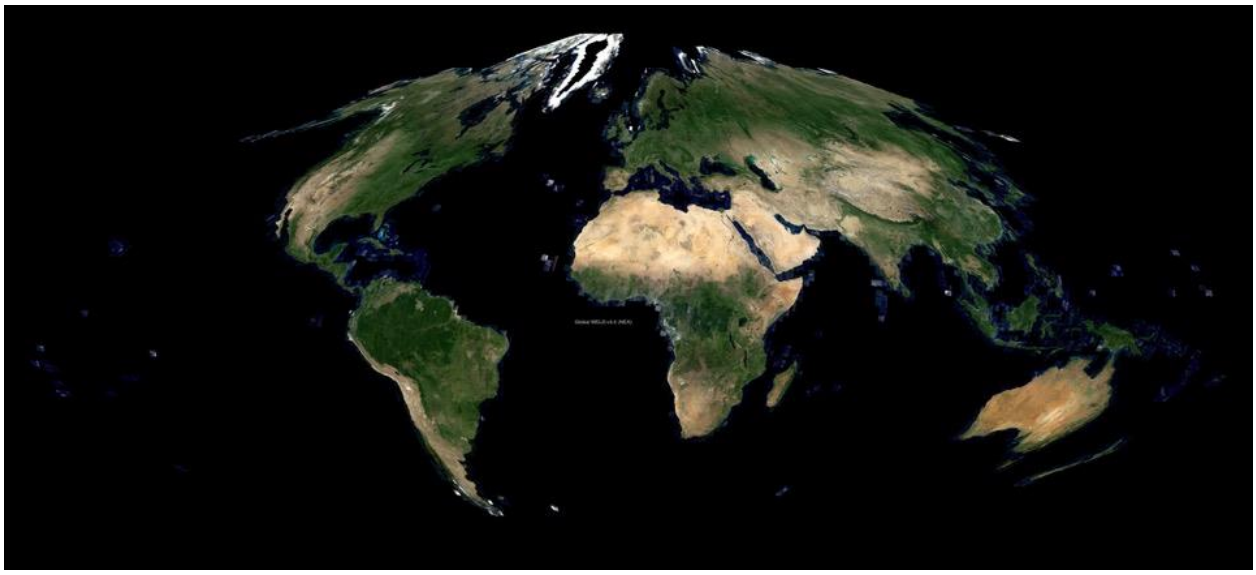
Global Web Enabled Landsat Data (GWELD) Products

Document Version 3.0, July 1 2019

David P. Roy [1] & Hankui K. Zhang [2]

[1] Department of Geography, Environment, & Spatial Sciences, and the Center for Global Change and Earth Observations, Michigan State University, East Lansing, MI 48824, USA

[2] Department of Geography and Geospatial Sciences, and the Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA



Example version 3.0 annual Global WELD (GWELD) 30 m product for climate year 2010

ABSRACT

Since January 2008, the U.S. Geological Survey has been providing free terrain-corrected and radiometrically calibrated Landsat data via the Internet. This revolutionary data policy provides the opportunity to use all the data in the U.S. Landsat archive and to consider the systematic utility of Landsat data for long-term large-area monitoring. With the advent of this free data policy, analysis ready data (ARD) are needed in user community to minimize Landsat pre-processing effort.

Built on the success of the conterminous United States (CONUS) Web Enabled Landsat Data (WELD) product, the NASA funded global WELD (GWELD) project seeks to provide global coverage Landsat ARD, specifically monthly and annual Landsat 30 m information for any terrestrial non-Antarctic location for six 3-year epochs spaced every 5 years from 1985 to 2010. They are generated from every available Landsat 4 and 5 Thematic Mapper (TM) and 7 Enhanced Thematic Mapper plus (ETM+) image held in the U.S. Landsat archive. The products define the "best" Landsat observation data available at the 30 m pixel tile location over monthly and annual product reporting periods. The GWELD products are developed specifically to provide consistent data that can be used to derive land cover as well as geophysical and biophysical products for regional assessment of surface dynamics and to study Earth system functioning.

The GWELD products are processed to nadir BRDF-adjusted surface reflectance (NBAR) for the reflective wavelength bands and to top of atmosphere (TOA) brightness temperature for the thermal bands. The products are defined in the same coordinate system and align precisely with the MODIS land products tiles. They aim to provide the first instance of global-scale science-quality Landsat data with a level of pre-processing comparable to the NASA MODIS land products.

Cite as
Roy, D.P., Zhang, H.K., Global Web Enabled Landsat Data (GWELD) Products - Algorithm Theoretical Basis Document, Version 3.0, July 1 2019.

Note this Algorithm Theoretical Basis Document may be changed in response to a formal NASA review process and as the GWELD product versioning is updated.

Table of Contents

1.0 OVERVIEW 4
2.0 VERSION 3.x GWELD PRODUCT THERORETICAL DESCRIPTION 5
2.1 Input Landsat data 5
2.2 Angular geometry computation 8
2.3 Reflective wavelength surface reflectance NBAR and TOA brightness temperature computation 11
2.3.1 Top of Atmosphere (TOA) reflectance computation 11
2.3.2 Atmospheric correction of the TOA reflectance bands 12
2.3.3 Reflective wavelength reflectance BRDF normalization 12
2.3.4 Normalized Difference Vegetation Index computation 14
2.3.5 TOA brightness temperature computation 14
2.4. Band saturation computation 15
2.5 Cloud masking 15
2.5.1 ACCA cloud detection 15
2.5.2 Classification tree cloud detection 16
2.6 Reprojection, resampling and tiling 18
2.7 Temporal compositing 21
2.8 GWELD Global Imagery Browse Services (GIBS) generation 26
3.0 VERSION 3.x PRODUCT DOCUMENTATION 29
3.1 Product storage format and contents 29
3.2 Product reporting periods 33
3.3 Product map projection and tiling scheme 35
3.4 Product file naming convention 37
3.5 GWELD product data volume 37
4.0 Known issues 38
5.0 LITERATURE CITED 40
Acknowledgements 48

1.0 OVERVIEW

The overall objective of NASA's Making Earth System Data Records for Use in Research Environments (MEaSUREs) program is to support projects providing Earth science data products and services driven by NASA's Earth science goals and contributing to advancing NASA's "missions to measurements" concept. Building on the Conterminous United States and Alaska WELD products the processing was expanded to global scale to provide monthly and annual Landsat 30 m information for any terrestrial non-Antarctic location for six 3-year epochs spaced every 5 years from 1985 to 2010. Herein the products will be referred to as global WELD (GWELD) products.

The GWELD products are developed specifically to provide consistent near-global coverage data that can be used to derive land cover, geophysical and biophysical products for assessment of surface dynamics and to study Earth system functioning. The GWELD products are processed so that users do not need to apply the equations and spectral calibration coefficients and solar information to convert the Landsat TM/ETM+ digital numbers to reflectance and brightness temperature, or atmospherically correct the reflective wavelength data, or correct for bi-directional anisotropy present in the reflective wavelength data, and successive products are defined in the same coordinate system and align precisely with the MODIS land products tiles, making them simple to use for multi-temporal applications. The GWELD products include spatially explicit quality assessment information, and appropriate metadata to enable further processing while retaining traceability of input data provenance. The GWELD products provide near global coverage Landsat Analysis Ready Data (ARD).

The GWELD processing, based on heritage techniques of MODIS data processing, is applied to all available Landsat TM/ETM+ acquisitions achieved in the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, (~125,000 Landsat TM/ETM+ images per year). Monthly and annual global WELD products are generated for six 3-year epochs spaced every 5 years from 1985 to 2010 and made freely available to the user community.

2.0 VERSION 3.x GWELD PRODUCT THERORETICAL DESCRIPTION

2.1 Input Landsat data

The GWELD products are made from Landsat 4 and 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper plus (ETM+) acquisitions. The Landsat 4 was launched on July 16 1982 and had been working until December 14, 1993 when the satellite lost its ability to transmit data. The Landsat 5 was launched on March 1, 1984 and had been observing the Earth surface until January 2013, when Landsat 5's instruments were powered off and the satellite was moved into a lower orbit. Landsat 7 was launched on April 15, 1999 and has been continuing to operate until present time. Landsat 4, 5 and 7 satellites maintain approximately 710 km sun-synchronous circular and 98.2° inclined near-polar orbits and have a 15° field-of-view that capture approximately 185 km × 170 km scenes defined in the Worldwide Reference System-2 (WRS2) path (groundtrack parallel) and row (latitude parallel) coordinates (Arvidson et al. 2001). Adjacent Landsat orbit paths are sensed 7 or 9 days apart and the same orbit path is sensed every 16 days, i.e., providing a 16 day revisit capability (Figure 1). Landsat 5 orbit was phased to the same Landsat 4 and Landsat 7 orbit to ensure a 8 day full earth coverage for Landsat 4 and 5 and for Landsat 5 and 7 (Hassett and Johnson 1984; Arvidson et al. 2006).

Every sunlit scene (solar zenith angle <75°) overpassed over the conterminous U.S. (CONUS) and main islands are acquired and archived at the USGS EROS (Ju and Roy 2008). For each sensor, scenes that are first overpassed between January 1 to 13 (January 14 for leap years) are overpassed a total of 23 times per year, while scenes first overpassed after January 14 (January 15 for leap years) are overpassed 22 times per year, i.e., each Landsat scene can be acquired a maximum of 22 or 23 times per year (Ju and Roy 2008). In many regions outside CONUS fewer images are acquired for a variety of reasons (Loveland and Dwyer 2012; Kovalskyy and Roy 2013). Notably, Landsat data acquired by international ground stations are being consolidated into the USGS EROS archive by the Landsat Global Archive Consolidation (LGAC) effort that began in 2010 (Wulder et al. 2016).

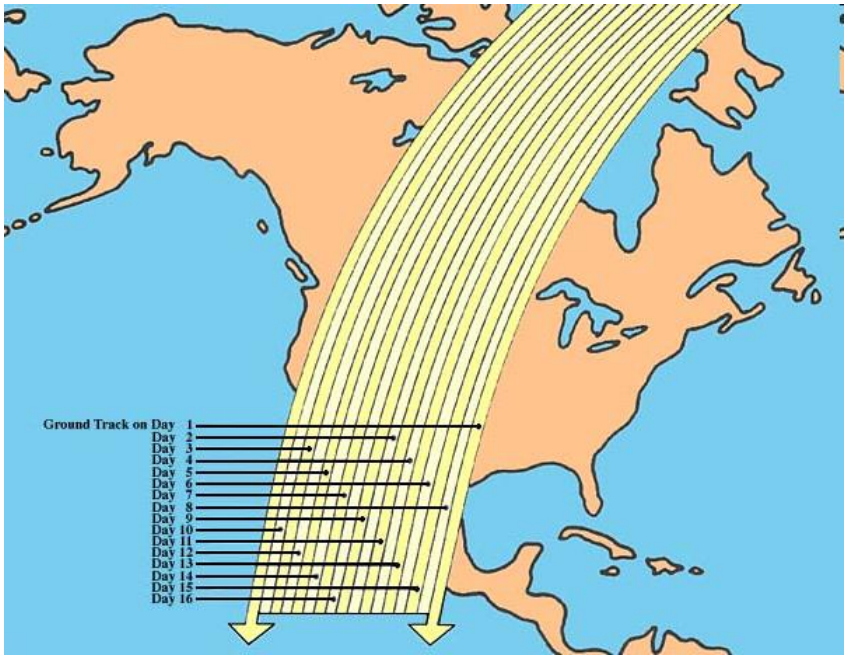


Figure 1 Landsat Orbit Geometry / Swath Pattern

Each Landsat image is approximately 183 km × 170 km but only Landsat imagery acquired over land and coastal areas and certain islands are acquired. Figure 2 shows every global Landsat WRS-2 path/center (nadir) scene location (Wulder et al. 2016).

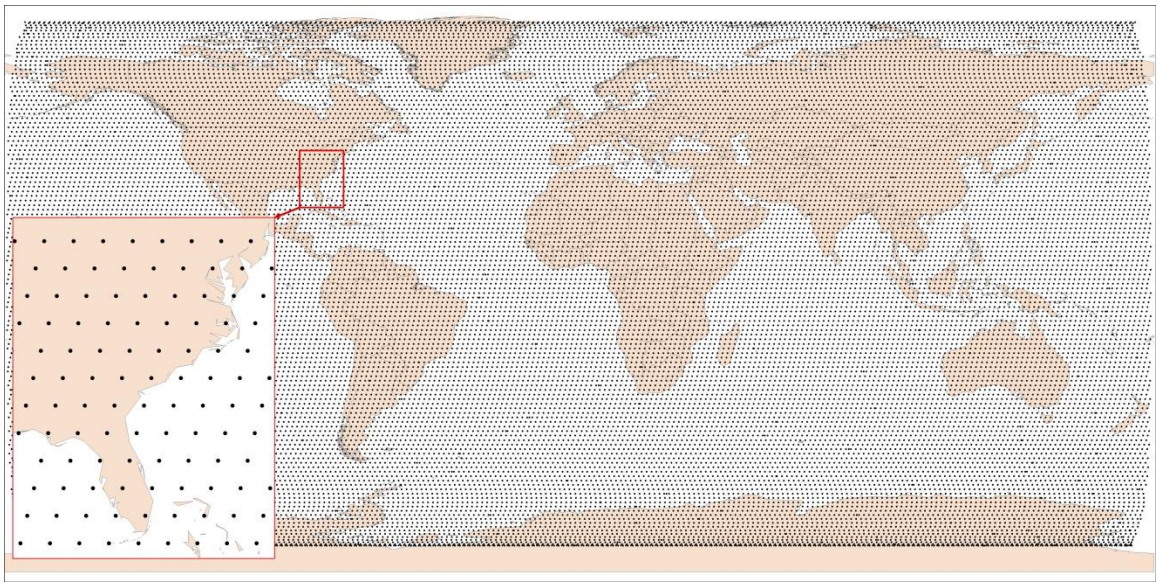


Figure 2 Landsat Path and Row locations. An area in the south east United States is highlighted.

Over the period of the GWELD project the Landsat image format and contents has changed. Notably, the Landsat Science Team helped advocate successfully for a processing/reprocessing collection scheme for Landsat. The Landsat Collection 1 products became available in 2017 and are now the standard products. They are designed to provide a consistent among-sensor archive to support time series analyses while controlling continuous improvement of the archive through periodic reprocessing, analogous to the MODIS processing/reprocessing approach (Justice et al. 2002). The Landsat Collection 1 products (https://www.usgs.gov/land-resources/nli/landsat/landsat-collections) have more rational filenames, improved per-pixel cloud mask, new quality data, improved calibration information, and improved product metadata including metadata that enable view and solar geometry calculation with code provided to us by the Landsat geolocation team.

Both the pre-Collection and the Collection 1 Landsat data are provided as 185 km × 170 km scenes defined in the WRS2 path (groundtrack parallel) and row (latitude parallel) coordinates in the Universal Transverse Mercator (UTM) map projection (WGS84 datum) in GeoTIFF format.

Some of the GWELD products were made using pre-Collection and some were made using Collection 1 input data. We do not anticipate significant differences due to the GWLED processing.

The GWELD products for the 2010 epoch (36 monthly products and annual products for 2009, 2010 and 2011) were made using pre-Collection Landsat input data. They are denoted as Version 3.0 products. Specifically, the pre-Collection Level-1 Precision and Terrain (L1TP) corrected data were used. The Level 1T processing includes radiometric correction, systematic geometric correction, precision correction using ground control chips, and the use of a digital elevation model to correct parallax error due to local topographic relief. While most pre-Collection Landsat data were processed as L1T, certain acquisitions did not have sufficient ground control or elevation data necessary for precision or terrain correction respectively. In these cases, the best level of correction was applied and the data were processed to Level 1G-systematic (L1G) with a geolocation error of less

than 250 meters (1σ) (Lee et al. 2004). The L1T file metadata records if the acquisition was processed to L1T or L1G. Only L1T data with geometry RMSD<30 m recorded in the metadata were used to make the GWELD 2010 epoch products. This was necessary to reduce the impact of L1G misregistration errors on the monthly and annual products (Roy 2000).

The GWELD products for the other epochs are made using Collection 1 Landsat input data. They are denoted as Version 3.1 products. As before, only L1T data with geometry RMSD<30 m recorded in the metadata are used. The processing also takes advantages of some of the new features of the Landsat Collection 1 data (e.g., new metadata for per-pixel solar and view geometry calculations).

Given the considerable global Landsat data volume (~125,000 Landsat images/year) the GWELD products are generated on the NASA Earth Exchange (NEX) high performance super computer (https://nex.nasa.gov/nex/) with an output volume of 24TB/year. The NEX pulled the Landsat acquisitions processed by the U.S. Landsat project from the USGS EROS and processed the data using the GWELD codes. The generated products are then sent back to the USGS EROS for dissemination.

2.2 Angular geometry computation

The Landsat viewing vector (Ω = view zenith angle, view azimuth angle) and the solar illumination vector (Ω' = solar zenith angle, solar azimuth angle) are defined for each Landsat 30 m pixel.

For the Collection 1 data (GWELD version 3.1), the Landsat per-pixel solar and viewing angles were calculated using the Landsat Angles Creation Tool provided by USGS (https://www.usgs.gov/land-resources/nli/landsat/solar-illumination-and-sensor-viewing-angle-coefficient-files) and using the angle coefficient file available with each Collection 1 L1TP file.

For the pre-Collection data (GWELD version 3.0), the solar illumination vector is computed using an astronomical model parameterized for geodetic latitude and longitude and time following the approach developed for MODIS geolocation (Wolfe et al. 2002). Computer code provided by Reda and Andreas (2005) was adapted to calculate the solar illumination vector for each Landsat pixel. This astronomical model is parameterized using the L1T UTM pixel coordinate data and the scene centre acquisition time available in the L1T metadata. The viewing vector can be computed precisely following the procedures described in the Landsat 7 Enhanced Thematic Mapper Plus (ETM+) Image Assessment System (IAS) if the satellite orientation is known. However, as this information is not provided in the L1T metadata, an alternative approach is adopted. The viewing vector is computed by first computing a vector normal to the surface of the WGS84 Earth model for the geodetic pixel coordinate, then the unit vector from the geodetic coordinate to the modeled satellite position, adjusting for the sensor-satellite attitude, and then the viewing vector zenith and azimuth components are derived using standard trigonometric formulae (Roy et al. 2008). Thus the viewing zenith (θ) and azimuth (ϕ) for the ground pixel *A* can be determined by Equations [1] and [2], given the locations of the satellite and the ground pixel (Figure 3).

The challenge is to estimate the satellite position. Nominally, the Landsat orbit follows the World Reference System-2 (WRS-2) with an orbital average altitude of 715.5 km and with each acquisition composed of 375 scans. Therefore, the satellite path can be estimated from the central location of each scan. The Landsat 15° field of view is swept over the focal planes by a scan mirror. The detectors are aligned in parallel rows on two separate focal planes: the primary focal plane, containing bands 1-4 and 8, and the cold focal plane containing bands 5, 6, and 7. The Landsat TM/ETM+ band 4 lies closest to the focal plane center with a displacement of around 10.4 IFOV to the sensor optical axis and is thus used to estimate the scan centers and the satellite positions for each scan.

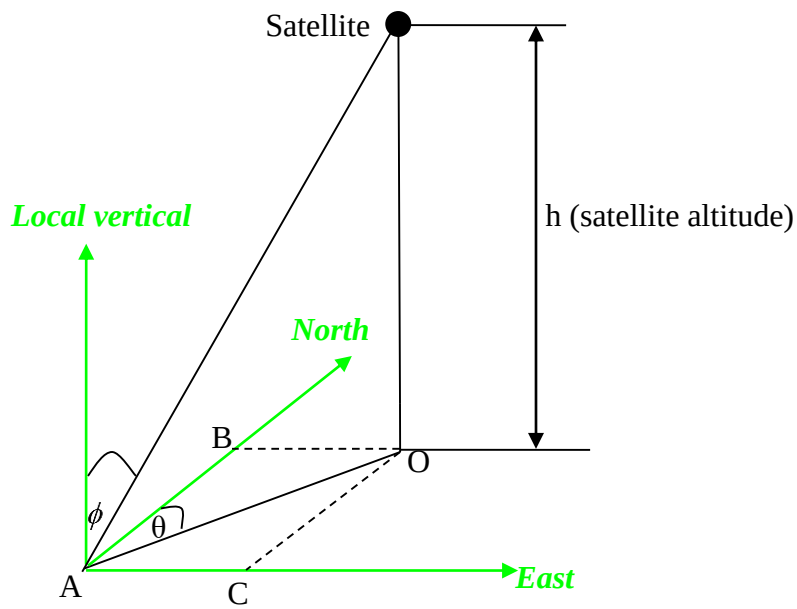


Figure 3 Landsat TM/ETM+ viewing geometry

$$\tan(\theta) = \frac{OB}{OC} \qquad [1]$$

$$\tan(\phi) = \frac{OA}{h} \qquad [2]$$

The edge pixels of the Band 4 image (that lies closest to the optical axis) are located and straight lines are fitted to determine the scene edges. The scene image is divided into 375 scans from north to south and the center for each scan is computed as:

$$P^{i}{}_{Scan,C} = (1 - \frac{i}{374})P_{NE,C} + \frac{i}{374}P_{SE,C} \qquad i = 0,1,\ldots 374 \qquad [3]$$

where $P^{i}{}_{Scan,C}$ is the scan center, and $P_{NE,C}$ and $P_{SE,C}$ are the centers of the north and south edges respectively. The satellite position $P^{i}{}_{Satellite}$ is estimated by displacing the scan centers by 10.4 IFOV as:

$$P^{i}{}_{Satellite} = P^{i}{}_{Scan,C} + 10.4 * 30.0 \qquad i = 0, 1, ... 374 \qquad [4]$$

This approach is computationally efficient although the accuracy of the viewing vector is a function of the errors of the L1T pixel geolocation and the spatial relations between the pixel and the sensor which may vary temporally.

2.3 Reflective wavelength surface reflectance NBAR and TOA brightness temperature computation

The GWELD processing is applied to all the Landsat TM/ETM+ bands, except the ETM+ 15 m panchromatic band, i.e., the 30 m blue (0.45-0.52μm), green (0.53-0.61μm), red (0.63-0.69μm), near-infrared (0.78-0.90μm), and the two mid-infrared (1.55-1.75μm and 2.09-2.35μm) bands, and the 60 m thermal (10.40-12.50μm) bands are processed as below.

2.3.1 Top of Atmosphere (TOA) reflectance computation

For the Collection 1 Landsat data, the calibration is derived using various on-board and vicarious calibration techniques (Markham and Helder 2012; Morfitt et al. 2015). A reflectance-based calibration is used as it has higher accuracy than the pre-Collection radiance-based approach (Markham et al. 2016). In the global WELD code the stored digital numbers are converted to reflectance as:

$$\rho_{\lambda} = \frac{g_{\lambda} \cdot DN_{\lambda} + b_{\lambda}}{\cos\theta_{s}} \qquad [5]$$

where ρ_{λ} is the top of atmosphere (TOA) reflectance (unitless), and DN_{λ} are the reflectance based sensor calibration gain and bias coefficients stored in the Collection 1 image metadata, DN_{λ} is the 8-bit DN values (0-255), and θ_s is solar zenith angle (radians) derived at 30 m resolution using the method described in Section 2.2 for the Collection 1 data.

For the pre-Collection Landsat data, a radiance-based calibration is used (Thorne et al. 1997). In the global WELD code the stored digital numbers are first converted to spectral radiance (units: W m^{-2} sr^{-1} μm^{-1}) using the sensor calibration gain and bias coefficients

stored in the Landsat L1T file metadata. Then the radiance is converted to top of atmosphere reflectance as:

$$\rho_\lambda = \frac{\pi \cdot L_\lambda \cdot d^2}{ESUN_\lambda \cdot \cos\theta_s} \qquad [6]$$

where ρ_λ is the top of atmosphere (TOA) reflectance (unitless), L_λ is the TOA spectral radiance (W m^{-2} sr^{-1} μm^{-1}), *d* is the Earth-Sun distance (astronomical units), $ESUN_\lambda$ is the mean TOA solar spectral irradiance (W m^{-2} μm^{-1}), and θ_s is solar zenith angle (radians) using the method described in Section 2.2 for the pre-Collection data. The quantities $ESUN_\lambda$ and *d* are tabulated by Chander et al. (2009).

2.3.2 Atmospheric correction of the TOA reflectance bands

The impact of the atmosphere is variable in space and time and is usually considered as requiring correction for quantitative remote sensing applications (Ju et al. 2012). Consistent Landsat surface reflectance data are needed in support of high to moderate spatial resolution geophysical and biophysical studies. The TOA reflectance (Section 2.3.1) are atmospherically corrected to surface reflectance.

The established Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Masek et al. 2006) method is used for atmospheric correction. The LEDAPS uses the 6SV radiative transfer code which has an accuracy better than 1% over a range of atmospheric stressing conditions (Kotchenova et al. 2006). The LEDAPS algorithm (Masek et al. 2006) derives the aerosol optical thickness independently from each Landsat acquisition using the Kaufman et al. (1997) dense dark vegetation approach and assuming a fixed aerosol type. The LEDAPS method also uses the NCEP/NCAR 6-hourly Reanalysis water vapor data and NASA's EP TOMS ozone data and surface atmospheric pressure from NCEP/NCAR 6-hourly Reanalysis data.

2.3.3 Reflective wavelength reflectance BRDF normalization

Most terrestrial surfaces are not Lambertian and so directional reflectance effects are present in satellite reflectance retrievals due to variable solar-surface-sensor geometry (Schaaf et al. 2002). Terrestrial reflectance anisotropy varies with the physical arrangement,

structural variability, and optical properties of the surface components (soil, grass, trees, etc.) and nominally may vary with the land cover type and condition. Directional reflectance effects, commonly described by the bidirectional reflectance distribution function (BRDF) (units of sr^{-1}), are relatively small in Landsat data, due to the narrow 15° sensor field of view and also because the acquisition view zenith angle is usually less than the solar zenith angle and so Landsat reflectance hot-spot effects do not occur (Zhang et al. 2016). However, across the Landsat swath the red and NIR reflectance can vary by up to 0.02 and 0.06 (reflectance units) due only to view variation effects (Roy et al. 2016a). These differences may constitute a significant source of noise for certain Landsat applications.

Landsat nadir BRDF-adjusted reflectance (NBAR) is derived from the surface reflectance (Section 2.3.2) using a semi-physical approach (Roy et al. 2016a) as:

$$\hat{\rho}_{Landsat}\left(\lambda,\Omega_{nbar},\Omega'_{nbar}\right) = c \times \rho_{Landsat}\left(\lambda,\Omega_{observed},\Omega'_{observed}\right) \qquad [7]$$

$$c = \frac{\hat{\rho}_{MODIS}\left(\lambda,\Omega_{nbar},\Omega'_{nbar}\right)}{\hat{\rho}_{MODIS}\left(\lambda,\Omega_{observed},\Omega'_{observed}\right)}$$

where $\hat{\rho}_{Landsat}\left(\lambda,\Omega_{nbar},\Omega'_{observed}\right)$ is the Landsat NBAR for wavelength λ for a 0° view zenith (Ω_{nbar}) and a defined optimal solar geometry (Ω'_{nbar}), $\rho_{Landsat}\left(\lambda,\Omega_{observed},\Omega'_{observed}\right)$ is the reflectance of a Landsat observation with viewing and solar illumination vectors $\Omega_{observed},\Omega'_{observed}$, and $\hat{\rho}_{MODIS}$ is the modeled reflectance for these angles computed at coarser spatial resolution using fixed global MODIS BRDF parameters described in (Roy et al. 2016a). The adjustment of surface reflectance to NBAR is conservative, i.e., it under-corrects BRDF effects, and has low sensitivity to the land cover type, condition, or surface disturbance. This is important as over the decadal GWELD product time series land cover, land use, and surface condition may have changed.

The optimal solar zenith in [7] is defined based on an established astronomical model (Blanco-Muriel et al. 2001) parameterized by the latitude, longitude, acquisition date and

a latitude-varying local time parameterized by a polynomial function of latitude (Zhang et al. 2016). The solar azimuth is not defined as the viewing zenith is set as 0˚.

The surface NBAR derived as [7] for each reflective Landsat band is stored as a signed 16-bit integer after being scaled by 10,000, in the same manner as the MODIS surface reflectance product (Vermote et al. 2002).

2.3.4 Normalized Difference Vegetation Index computation

The normalized difference vegetation index (NDVI) is the most commonly used vegetation index, derived as the near-infrared minus the red reflectance divided by their sum (Tucker 1979). The 30 m surface NBAR NDVI is computed from the red and near-infrared Landsat surface reflectance NBAR and is stored as signed 16-bit integers after being scaled by 10,000, in the same manner as the MODIS NDVI product (Huete et al. 2002).

2.3.5 TOA brightness temperature computation

For both the pre-Collection and Collection-1 data, the digital numbers stored in the Landsat image are first converted to spectral radiance (units: W m^{-2} sr^{-1} μm^{-1}) and the radiance sensed in the Landsat thermal bands are converted to TOA brightness temperature (i.e., assuming unit surface emissivity) using standard formula as:

$$T = \frac{K_2}{\log(K_1 / L_\lambda + 1)} \qquad [8]$$

where T is the 10.40-12.50μm TOA brightness temperature (Kelvin), K_1 and K_2 are thermal calibration constants set as values in Table 7 in (Chander et al 2009), and L_λ is the TOA spectral radiance. This equation is an inverted Planck function simplified for the TM/ETM+ sensor considering the thermal band spectral responses. Landsat thermal calibration constants are derived in the conventional manner considering the thermal band spectral responses (Chander et al. 2009) and used to derive calibrated TOA brightness temperature, i.e., the temperature of the observed surface if it was a perfect black body.

The 30 m TOA brightness temperature data are stored as signed 16-bit integers with units of degrees Celsius by subtracting 273.15 from the brightness temperature and then scaling by 100.

2.4. Band saturation computation

The Landsat TM/ETM+ calibration coefficients are configured in an attempt to globally maximize the range of land surface spectral radiance in each spectral band (Markham et al. 2006). However, highly reflective surfaces, such as snow and clouds, may over-saturate the reflective wavelength bands, with saturation varying spectrally and with the illumination geometry (solar zenith and surface slope) (Cahalan et al. 2001, Bindschadler et al. 2008). Similarly, hot surfaces may over-saturate the thermal bands (Flynn and Mouginis-Mark 1995), and cold surfaces may under-saturate the high-gain thermal band (Landsat Handbook, Chapter 6). Over and under-saturated pixels are designated by digital numbers of 255 and 1 respectively in the L1T data. As the radiance values of saturated pixels are unreliable, a 30 m 8-bit saturation mask is generated, storing bit packed band saturation (1) or unsaturated (0) values for the eight Landsat bands.

2.5 Cloud masking

It is well established that optically thick clouds preclude optical and thermal wavelength remote sensing of the land surface but that automated and reliable satellite data cloud detection is not trivial (Kaufman 1987, Platnick et al. 2003). Recognizing that cloud detection errors, both of omission and commission, will always occur in large data sets, both the Landsat automatic cloud cover assessment algorithm (ACCA) and a classification tree based cloud detection approach are implemented.

2.5.1 ACCA cloud detection

The U.S. Landsat project uses an automatic cloud cover assessment algorithm (ACCA) to estimate the cloud content of each acquisition (Irish 2000, Irish et al. 2006). The ACCA takes advantage of known spectral properties of clouds, snow, bright soil, vegetation, and water, and consists of twenty-six filters/rules applied to 5 of the 8 TM/ETM+ bands (Irish

et al. 2006). The primary goal of the algorithm is to quickly produce scene-average cloud cover metadata values, that can be used in future acquisition planning (Ardvidson et al. 2006), and that users may query as part of the Landsat browse and order process. The ACCA was not developed to produce a "per-pixel" cloud mask; despite this, the ACCA has an estimated 5% error for 98% of the global 2001 ETM+ acquisitions archived by the U.S. Landsat project.

The ACCA code is applied to every Landsat TM/ETM+ acquisition to produce a 30 m per-pixel cloud data layer, stored as an unsigned 8-bit integer.

2.5.2 Classification tree cloud detection

The state of the practice for automated satellite land cover classification is to adopt a supervised classification approach where a sample of locations of known land cover classes (training data) are collected. The optical and thermal wavelength values sensed at the locations of the training pixels are used to develop statistical classification rules, which are then used to map the land cover class of every pixel. Classification trees are hierarchical classifiers that predict categorical class membership by recursively partitioning data into more homogeneous subsets, referred to as nodes (Breiman et al. 1984). They accommodate abrupt, non-monotonic and non-linear relationships between the independent and dependent variables, and make no assumptions concerning the statistical distribution of the data (Prasad et al. 2006). Bagging tree approaches use a statistical bootstrapping methodology to improve the predictive ability of the tree model and reduce over-fitting whereby a large number of trees are grown, each time using a different random subset of the training data, and keeping a certain percentage of data aside (Breiman 1996). Conventionally multiple bagged trees are used to independently classify the satellite data and the multiple classifications are combined using some voting procedure. A single parsimonious tree from multiple bagged trees was developed so that only one tree was used to classify the Landsat data, reducing the GWELD computational overhead.

Supervised classification approaches require training data. A global database of Landsat Level 1T and corresponding spatially explicit cloud masks generated by photo-

interpretation of the reflective and thermal bands were used. This database was developed to prototype the cloud mask algorithm for Landsat 8 (Irons and Masek 2006). The Landsat interpreted cloud mask defines each pixel as thick cloud, thin cloud, cloud shadow or not-cloudy. These interpreted cloud states were reconciled into cloud (i.e., thick and thin cloud) and non-cloud (i.e. cloud shadow and not cloudy states) classes. In addition, to avoid mixed pixel cloud edge problems, the cloud labeled regions were morphologically eroded by one 30 m pixel and not used. A 0.5% sample of training pixels was extracted randomly from each Landsat scene, where data were present and not including the cloud boundary regions. A total of 88 northern hemisphere Landsat scenes acquired in polar (19 acquisitions), boreal (22 acquisitions), mid-latitude (24 acquisitions) and sub-tropical latitudinal zones (23 acquisitions) were sampled. The sampled Landsat data were processed to TOA reflectance, brightness temperature and the band saturation flag computed as described above. Only pixels with reflectance greater than 0.0 were used. A total of 12,979,302 unsaturated training pixels and 5,374,157 saturated training pixels were extracted.

Two classification trees; one for saturated training data and the other for the unsaturated training data were developed. The saturated TOA reflectance and brightness temperature values are unreliable but still provide information that can be classified. Consequently, better cloud non-cloud discrimination is afforded by classifying the saturated and unsaturated pixels independently.

For both the saturated and unsaturated classification trees, all the 30 m TOA reflective bands were used, except the shortest wavelength blue band which is highly sensitive to atmospheric scattering (Ouaidrari and Vermote 1999). The unsaturated classification tree also used reflective band simple ratios similar to those used by ACCA (Irish et al. 2006). The saturated classification tree did not use band ratios as they could not be computed when one or both bands in the ratio formulation were saturated.

Twenty five bagged classification trees were generated, running the Splus tree code on a 64 bit computer, each time, 20% of the training data were sampled at random with replacement and used to generate a tree. Each tree was used to classify the remaining ("out-of-bag") 80% of the training data, deriving a vector of predicted classes for each out-of-bag pixel. In this way, each training pixel was classified 25 or fewer times. The most frequent predicted class (cloud or non-cloud) for each training pixel was derived; and used with the corresponding training data to generate a single final tree, i.e. the final tree was generated using approximately $25 \times 0.8 \times n$ training pixels, where n was either the 12,979,302 unsaturated training pixels or the 5,374,157 saturated training pixels. To limit overfitting, all the trees were terminated using a deviance threshold, whereby additional splits in the tree had to exceed 0.02% of the root node deviance or tree growth was terminated. The final unsaturated and saturated classification trees were defined by 1595 nodes that explained 98% of the tree variance and 188 nodes that explained 99.9% of the tree variance respectively. These are used to classify every Landsat pixel according to its saturation status. The 30 m cloud classification results are stored as an unsigned 8-bit integer.

2.6 Reprojection, resampling and tiling

The processed data are reprojected from UTM image coordinates into global sinusoidal tiles nested within the 10° × 10° MODIS land product tiles (Figure 4) so it is straightforward to compare the 30 m GWELD products with any of the standard gridded MODIS land products (Wolfe et al. 1998; Justice et al. 2002). The sinusoidal projection is (i) uninterrupted, (ii) equal area, and (iii) has less pixel loss and duplication compared to other global equal-area projections (Seong et al. 2002; Mulcahy 2000).

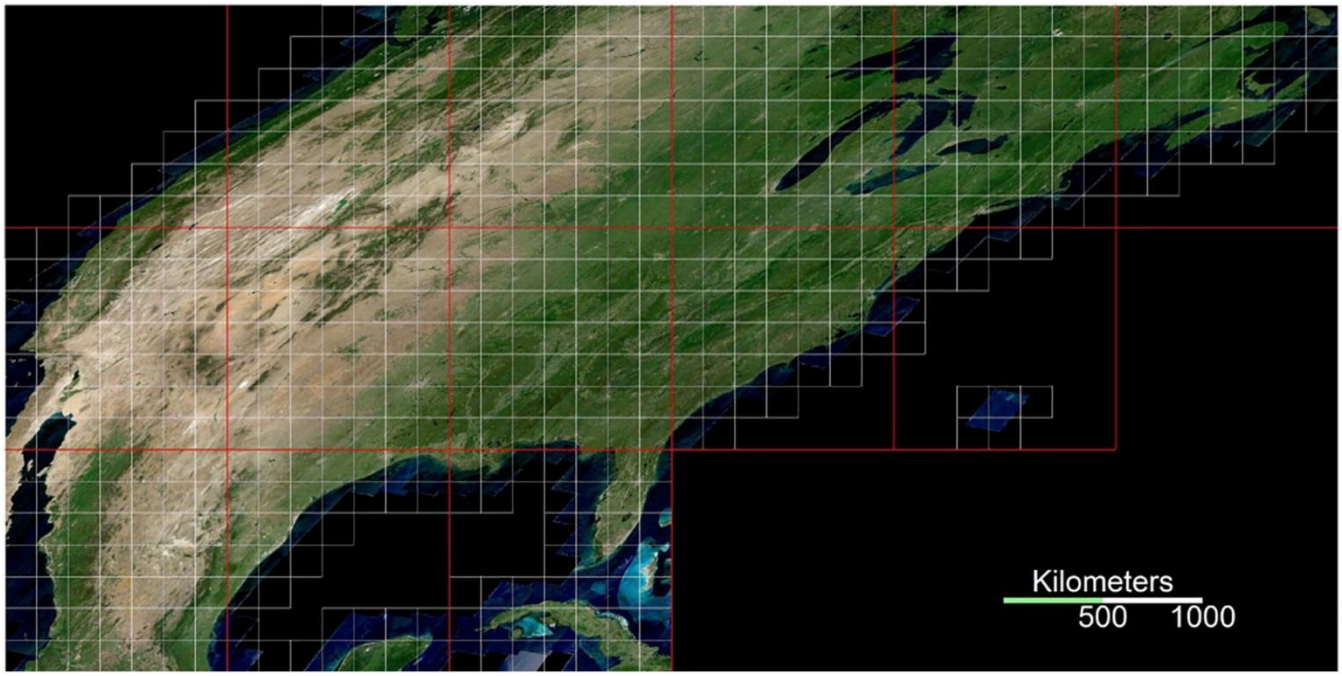


Figure 4 Illustration of nested Global WELD tiles (white) and the MODIS tiles (red) for North American 20°N to 50°N. This example is composed of 561 GWELD 5295 × 5295 30 m pixel tile boundaries (white) that are spatially nested within 14 standard MODIS land 10° × 10° tile boundaries (red) defined in the sinusoidal equal area projection. The background shows for geographic context the year 2010 version 3.0 GWELD true color surface NBAR product. Figure from Zhang and Roy (2017).

It is not physically possible to store global 30 m data in a single file. To ensure manageable file sizes, the 30 m Landsat data are reprojected into global sinusoidal tiles. Each GWELD tile is composed of 5295 × 5295 30 m Landsat pixels. There are 7 × 7 GWELD tiles within each 10° × 10° MODIS land tile (Figures 4 and 13). This tile pixel dimension (number of rows and columns) is smaller than the dimensions of an individual Landsat image.

The Landsat pixels are allocated to the sinusoidal coordinate system using the inverse gridding approach, sometimes known as the indirect method (Konecny 1979). In this approach the center coordinates of each sinusoidal 30 m pixel are mapped to the nearest pixel center in the Landsat data, and the TM/ETM+ processed data for that pixel are allocated to the sinusoidal output grid. This processing approach is computationally efficient and geometrically is the equivalent of nearest neighbor resampling (Wolfe et al. 1998). The General Cartographic Transformation Package (GCTP) developed by the

USGS and used to develop a number of applications including the MODIS global browse imagery (Roy et al. 2002) is used to transform coordinates between the UTM and sinusoidal map projections. The GCTP is computationally expensive. Consequently, a sparse triangulation methodology was used where the GCTP is invoked to project sinusoidal 30 m pixels to UTM coordinates only at the vertices of triangles, and sinusoidal 30 m pixel locations falling within the triangles are projected to UTM coordinates using a simplicial coordinate transformation (Saalfeld 1985). In this approach, any point (p_x, p_y) in a triangle with vertices (x_1, y_1), (x_2, y_2), (x_3, y_3) can be represented by three simplicial coordinates (s_1,s_2,s_3) defined:

$$s_1 = a_1 p_y + b_1 p_x + c_1$$
$$s_2 = a_2 p_y + b_2 p_x + c_2 \qquad [9]$$
$$s_3 = 1 - s_1 - s_2$$

where

$$a_1 = (x_3 - x_2)/t \qquad a_2 = (x_1 - x_3)/t$$
$$b_1 = (y_2 - y_3)/t \qquad b_2 = (y_3 - y_1)/t$$
$$c_1 = (x_2 y_3 - x_3 y_2)/t \qquad c_2 = (x_3 y_1 - x_1 y_3)/t$$
$$t = x_1 y_2 + x_2 y_3 + x_3 y_1 - x_3 y_2 - x_2 y_1 - x_1 y_3$$

Given a point (p_x, p_y) defined in sinusoidal coordinates the corresponding location in UTM coordinates is:

$$p_x^{'} = s_1 x_1^{'} + s_2 x_2^{'} + s_3 x_3^{'} \qquad [10]$$
$$p_y^{'} = s_1 y_1^{'} + s_2 y_2^{'} + s_3 y_3^{'}$$

where $(x_1^{'}, y_1^{'}),(x_2^{'}, y_2^{'}),(x_3^{'}, y_3^{'})$ are the coordinates of the triangle vertices in UTM calculated by projecting the corresponding sinusoidal triangle vertices (x_1, y_1), (x_2, y_2), (x_3, y_3) using the GCTP. A regular lattice of triangles is defined by bisecting squares with side lengths of 450 m (i.e., fifteen 30 m pixels) defined from the north-west origin of the sinusoidal coordinate system so that in each square there were two triangles with different topologies. This approach is computationally efficient as the GCTP is only called for each triangle vertex and the coefficients *a*, *b*, *c* and *t* are computed only once for each triangle.

2.7 Temporal compositing

Compositing procedures are applied independently on a per-pixel basis to gridded satellite time series and provide a practical way to reduce cloud and aerosol contamination, fill missing values, and reduce the data volume of moderate resolution global near-daily coverage satellite data (Cihlar 1994). Compositing was developed originally to reduce residual cloud and aerosol contamination in AVHRR time series to produce representative *n*-day data sets (Holben 1986). Compositing criteria have included the maximum NDVI, maximum brightness temperature, maximum apparent surface temperature, maximum difference in red and near-infrared reflectance, minimum scan angle, and combinations of these (Roy 2000). Ideally, the criteria should select from the time series only near-nadir observations that have reduced cloud and atmospheric contamination. Composites generated from wide field of view satellite data, such as AVHRR or MODIS, often contain significant BRDF effects (Cihlar et al. 1994, Gao et al. 2002, Roy et al. 2006). Compositing algorithms that model BRDF have been developed to compensate for this problem and combine all valid observations to estimate the NBAR (Schaaf et al. 2002). However, this approach does not provide a solution for compositing thermal wavelength satellite data, and is less appropriate for application to Landsat data as the comparatively infrequent 16 day Landsat repeat cycle and the narrow 15º Landsat sensor field of view do not provide a sufficient number or angular sampling of the surface to invert bidirectional reflectance models (Danaher et al. 2001, Roy et al. 2008). Consequently, the GWELD compositing is based on the selection of a “best” pixel over the compositing period.

The GWELD compositing is applied before atmospheric correction, i.e. to the TOA reflectance, because (i) atmospheric correction is imperfect (Ju et al. 2012) and (ii) because in this way only the composited gridded WELD pixel values need to be atmospherically corrected rather than every pixel in every input Landsat L1T acquisition (Roy et al. 2014b). Table 1 summarizes the GWELD compositing logic.

Table 1 GWELD compositing criteria. n_{valid} is the number of the valid (non-filled, non-saturated and non-cloudy) observations over the composition period; n_{water} is the number of the water (valid, but DT cloud state is 200 OR $\rho_{TOA}^{1} > \rho_{TOA}^{2} > \rho_{TOA}^{3} > \rho_{TOA}^{4}$) observations over the composition period; n_{soil} is the number of the soil (valid, non-water, $\rho_{TOA}^{2} \leq \rho_{TOA}^{3} \leq \rho_{TOA}^{4} \leq \rho_{TOA}^{5}$) observations over the composition period; n_{snow} is the number of the snow (valid, non-water, non-soil and *NDSI*>0.4) observations over the composition period. *SAM* is the spectral angle mapper (SAM) metric calculated over two Landsat TOA spectra using bands TM/ETM+ 2, 3, 4, 5 and 7. Minimum blue means selected the observation with the minimal blue band TOA reflectance. The maximum weighted NDVI and ND51 will be detailed in the below paragraph.

Priority	Compositing selection criteria
1	If n_{valid}=0: minimum blue over all observations
2	If n_{valid}=1 & (n_{water}==1 OR n_{snow}==1): minimum blue over all observations
3	If n_{valid}=1 & (n_{water}==0 & n_{snow}==0): valid observation
4	If n_{valid}=2 & n_{water}==2: minimum blue over the two valid observations
5	If n_{valid}=2 & n_{water}==1 & n_{soil}==0: (1) if *SAM* > 0.7: minimum blue over the two valid observations (2) else: maximum weighted NDVI and ND51
6	If n_{valid}=2 & n_{water}==1 & n_{soil}==1: (1) if *SAM* ≤ 0.7: minimum blue over the two valid observations (2) else: maximum weighted NDVI and ND51
7	If n_{valid}=2 & n_{water}==0: maximum weighted NDVI and ND51
8	If n_{valid}>2: (1) if n_{water}/n_{valid} ≥ 0.5: minimum blue over the valid observations (2) else: maximum weighted NDVI and ND51

Each row of Table 1 reflects a unique compositing path. If the criterion in a row is not met then the criterion in the row beneath is used and this process is repeated until the last row. This implementation enables the composites to be updated on a per pixel basis shortly after the input TM/ETM+ data are processed and regardless of the chronological processing order. For example, after 16 days the same pixel location may be sensed again, and the compositing criteria are used to decide if the more recent TM/ETM+ pixel data should be allocated to overwrite the previous data. For each composited pixel, the day of the year

that the selected pixel was acquired on, and the number of different valid acquisitions considered at that pixel over the compositing period, are stored in the GWELD product.

In Table 1, the maximum weighted NDVI and ND51 parameters are used to improve the original (version 2.2) CONUS WELD compositing approach that was based on a maximum NDVI compositing heritage (Roy et al. 2010). Figure 5 illustrates the improvement over the original version 2.2. The ND51 is derived in a similar way to NDVI but replacing the red and NIR bands with the Landsat TM/ETM+ band 1 (blue) and 5 (SWIR).

Figure 5 Illustration of Version 2.2 compositing results (left) and refined Version 3.x (right) compositing applied to Landsat 7 ETM and Landsat 5 TM data for a 400 by 400 30 m pixel subset over Columbia River Valley, Eastern Washington state, sensed July 2010. A total of five Landsat 7 and two Landsat 5 input scenes were composited. Issues with soil variations, shadows and cloud over water are notably improved.

The weights for the NDVI and ND51 are stored in a Look Up Table (LUT) of atmospheric effects derived by comparing Landsat 5 and 7 TOA and LEDAPS surface reflectance of more than 200 million 30 m pixels for three summer months and three winter months over CONUS (Roy et al. 2016b). Using fixed LUTs to characterize the atmospheric effects can avoid abnormal values in the surface reflectance NDVI and ND51 derived using image based atmospheric profile characterizations (Roy et al. 2014b). The NDVI and ND51

weights are inversely weighted so as to (1) select healthy vegetated pixels when the surface is vegetated (high NDVI weights for high NDVI values), (2) select bright surfaces when the NDVI starts to be meaningless for non-vegetated surface (low NDVI weights for low NDVI values).

The GWELD cloud masks are used to complement the maximum NDVI and ND51 criterion and to provide a more reliable differentiation between clouds and the land surface. The two cloud masks do not always agree, but it is not possible to quantitatively evaluate their relative omission and commission errors as a function of different clouds and background reflectance and brightness temperature. Consequently, a pixel is considered cloudy and non-cloudy if both the ACCA and the Classification Tree algorithms detected it as cloud and non-cloud respectively, and a pixel is considered as uncertain cloud if only one cloud algorithm detected it as cloudy.

The above compositing algorithm is applied to the TOA reflectance data. However, after compositing, the NBAR surface reflectance for the selected pixels is recorded in the output GWELD product, rather than the TOA reflectance. This swapping procedure (Roy et al. 2014b) is illustrated in Figure 6, showing the composited TOA reflectance (top left) and the output composited NBAR surface reflectance (bottom left). The swapping uses 30 m pixel provenance information that includes an index value (0, 1, … 255) that is stored at each GWELD product tile location (Figure 6, top right) that refers to the Landsat input image filename. In addition, the column and row locations in the input Landsat image (0 to the maximum dimension of the input Landsat image) that the composited pixel was selected from are stored as GWELD product bands and are used in the swapping process. Importantly, these information also provide traceability of the input data provenance.

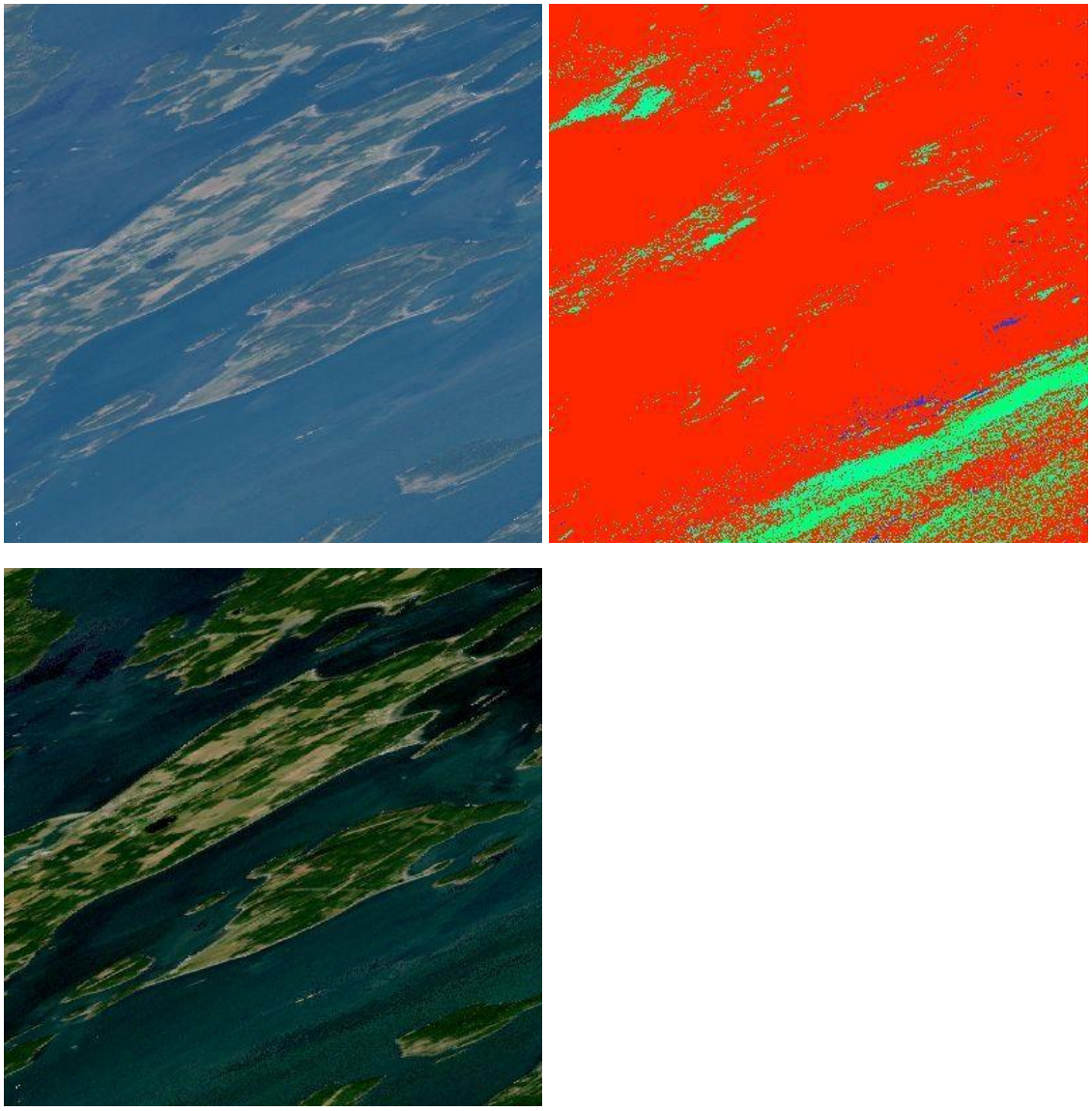


Figure 6 Illustration of composting swapping algorithm for a 400 × 400 30 m pixel subset of a July 2000 monthly GWELD product over the Strait of Juan de Fuca (located on the west coast international boundary between Canada and the United States and GWELD tile hh09v04.h6v1). The composited was generated from five Landsat 5 and 7 acquisitions. *Top Left* shows the composited top of atmosphere (TOA) true color red (0.63~0.69 μm), green (0.53~0.61 μm) and blue (0.45~0.52 μm), reflectance; *Top right* shows the L1T filename index values that the top left composted TOA reflectance is selected from (blue acquired on July 6 and 7, spring green acquired on July 14 and 15; and red acquired on July 30 2000); *Bottom Left* shows the swapped atmospherically corrected NBAR equivalent of the TOA data using the L1T index (*top right*) and the L1T pixel column and row coordinates (not shown here).

2.8 GWELD Global Imagery Browse Services (GIBS) generation

The Global Imagery Browse Services (GIBS) is a core NASA EOSDIS component that provides a scalable, responsive, highly available, and community standards based set of imagery services. Monthly and annual Global WELD 30 m Version 3.0 GIBS browse image products defined in the Geographic (i.e. latitude/longitude) projection were generated and are available in NASA Worldview (**Figure 7**).

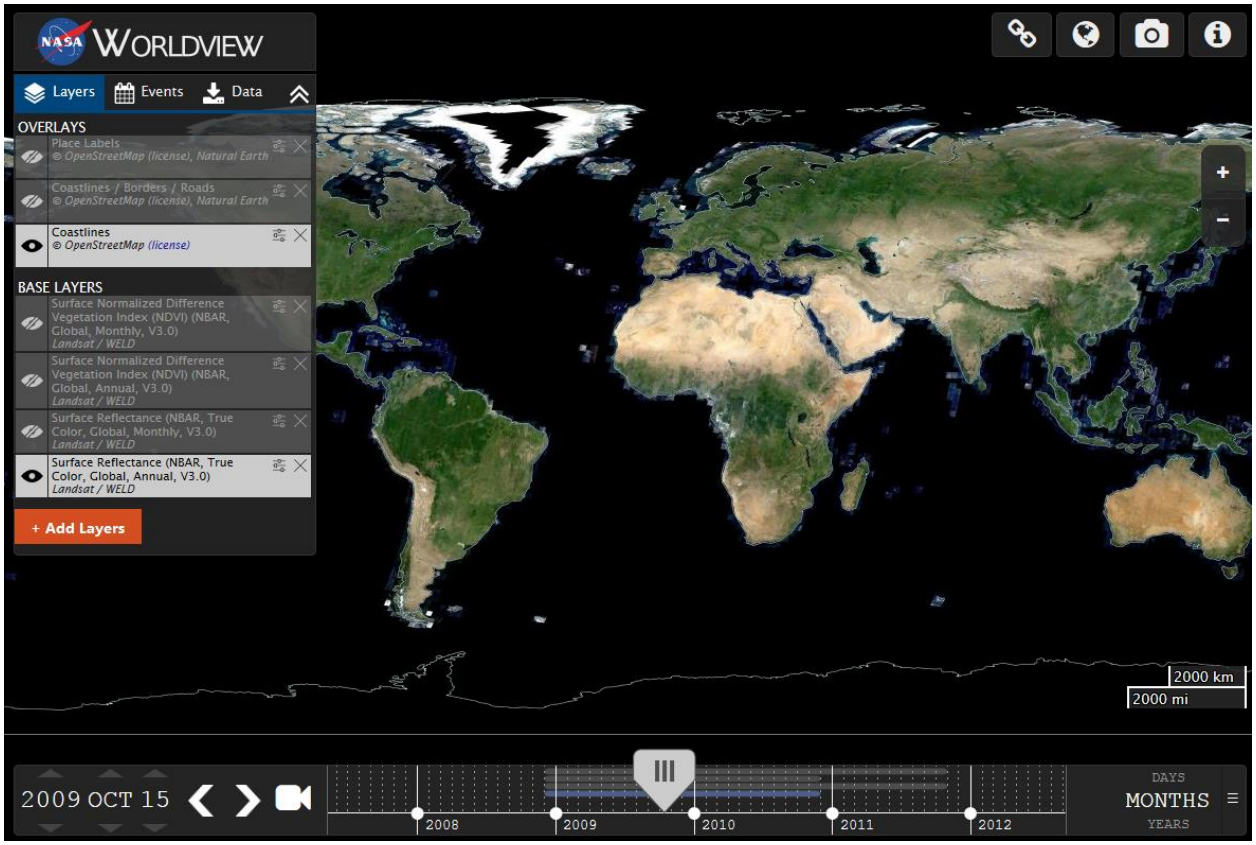


Figure 7 GIBS browse of the global WELD annual Version 3.0 30 m product for climate year 2009 in the NASA GIBS (see https://worldview.earthdata.nasa.gov/ or more directly http://go.nasa.gov/2kLcKto).

False color GIBS browses that highlight burned-areas (2.22, 0.83, 0.66 micron false color Red, Green, Blue bands) and snow/ice surfaces (0.48, 1.65, 2.2 micron false color Red, Green, Blue bands) were also generated. Figures 8 and 9 show examples.

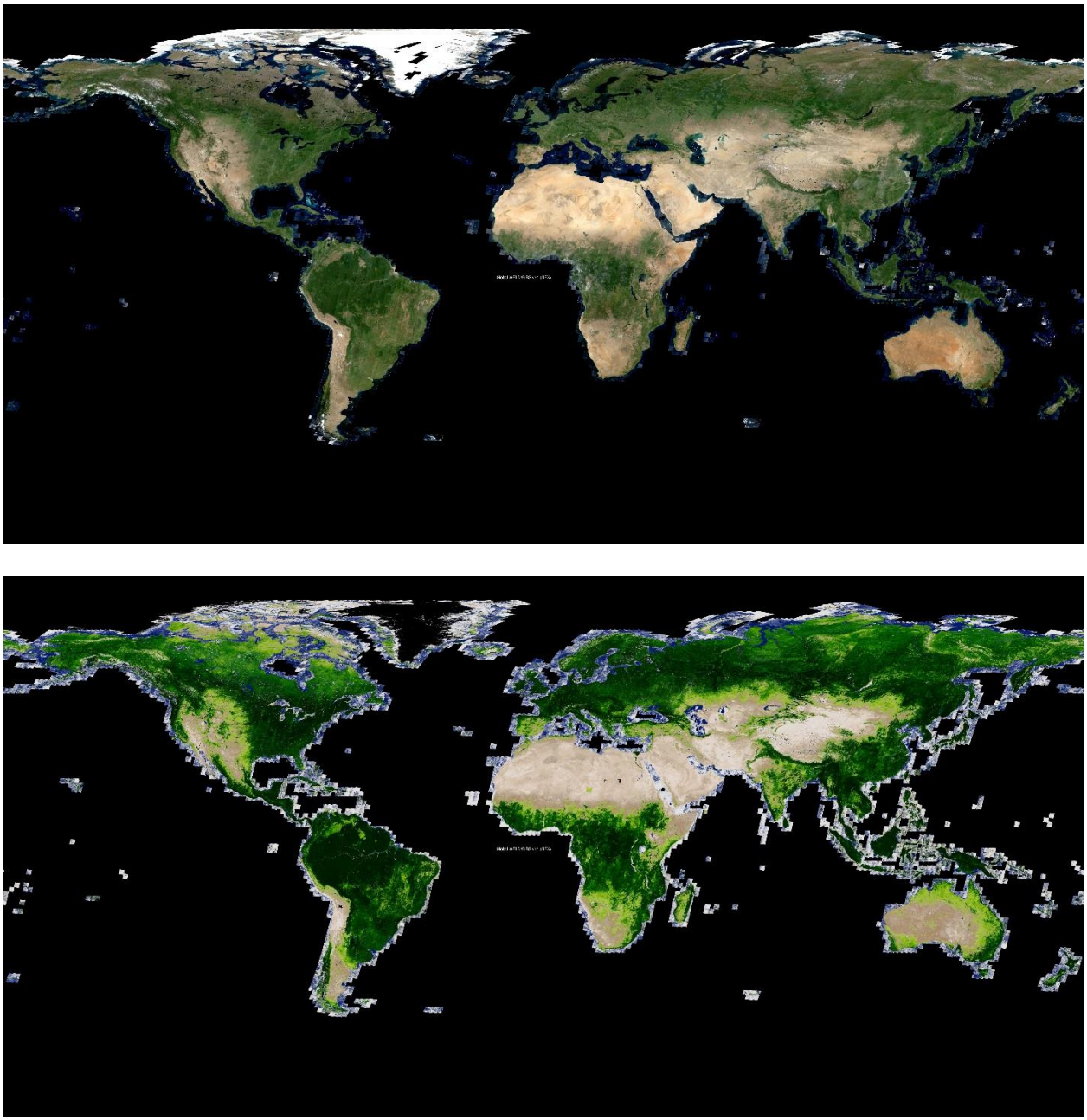


Figure 8 GIBS browses of the global WELD annual Version 3.1 30 m product for climate year 2000; true color (red, green, blue) bands (top) and surface NBAR NDVI (bottom).

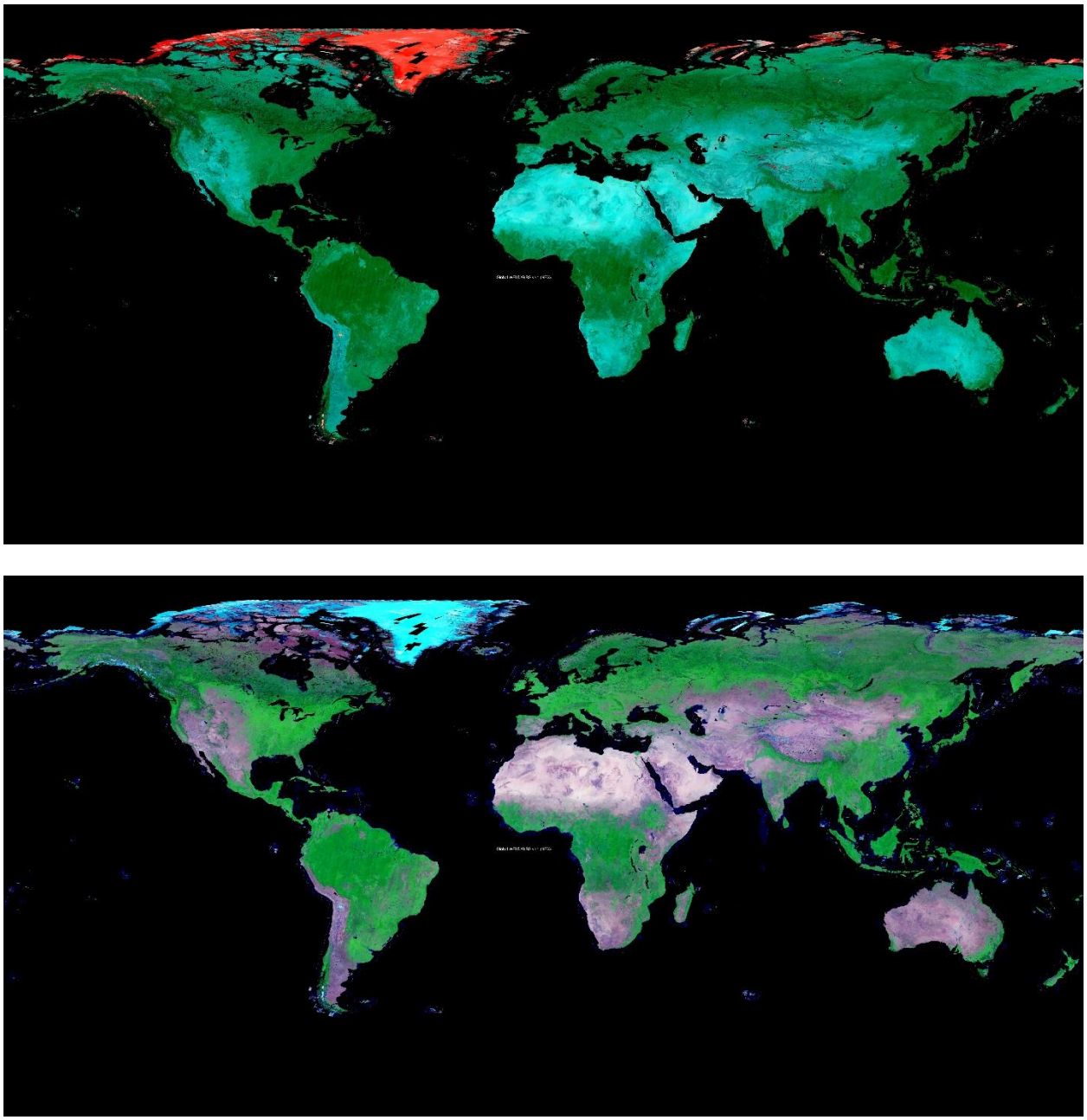


Figure 9 GIBS browses of the global WELD annual Version 3.1 30 m product for climate year 2000; false color bands that highlight snow/ice surfaces (0.48, 1.65, 2.2 micron bands) (top) and highlight burned-areas (2.22, 0.83, 0.66 micron false color bands) (bottom).

3.0 VERSION 3.x PRODUCT DOCUMENTATION

3.1 Product storage format and contents

The GWELD products are stored in Hierarchical Data Format (HDF), a self-descriptive data file format designed by the National Center for Supercomputing Applications to assist users in the storage and manipulation of scientific data across diverse operating systems and machines.

The version 3.0 GWELD 2010 epoch products (36 monthly products and annual products for 2009, 2010 and 2011) are defined in HDF4, the other version 3.1 GWELD epochs are defined in HDF4EOS.

The products are generated in separate 5295 × 5295 30 m pixel tiles. Each GWELD product tile is composed of 24 bands (24 HDF science data sets) stored with appropriate data types to minimize the file size and with band-specific attributes (fill value, scale factor, units, valid range) (**Table 2**). Each GWELD product tile carries the default HDF metadata and a number of product specific metadata that summarize the pixels in each tile (**Table 3**).

Table 2 GWELD product 30 m **bands**

Band Name	Data Type	Valid Range	Scale factor	Units	Fill Value	Notes
Band1_SRF_REF	int16	-2000 to 16000	0.0001	Unitless	-32768	The conventional Landsat 4/5/7 band numbering scheme is used. Surface reflectance (SRF) derived using the LEDAPS atmospheric correction code and then subsequently BRDF adjusted to nadir view (0 degree view zenith) with a modelled solar zeith defined in NBAR_Solar_Zenith. Top of atmosphere (TOA) brightness temperature (BT) is computed using standard formulae and calibration coefficients associated with each acquisition. Band 6 brightness temperature data are defined at 30 m. The Band62 pixel value is set to FILL if the pixel was from Landsat 4 or 5 as this band does not exist on the Landsat 4 and 5 TM sensors.
Band2_SRF_REF	int16	-2000 to 16000	0.0001	Unitless	-32768	
Band3_SRF_REF	int16	-2000 to 16000	0.0001	Unitless	-32768	
Band4_SRF_REF	int16	-2000 to 16000	0.0001	Unitless	-32768	
Band5_SRF_REF	int16	-2000 to 16000	0.0001	Unitless	-32768	
Band61_TOA_BT	int16	-32767 -- 32767	0.01	Degrees Celsius	-32768	
Band62_TOA_BT	int16	-32767 -- 32767	0.01	Degrees Celsius	-32768	
Band7_SRF_REF	int16	-2000 to 16000	0.0001	Unitless	-32768	
NDVI_SRF	int16	-10000 -- 10000	0.0001	Unitless	-32768	Normalized Difference Vegetation Index (NDVI) value generated from Band3_SRF_REF and Band4_SRF_REF.
Day_Of_Year	int16	1 -- 366	1	Day	0	Day of year the selected Landsat pixel was sensed on. Note (a) days 1-334 (or 1-335) were sensed in January-November of the nonleap (or leap) current year; (b) days 335-365 (or 336-366) were sensed in December of the nonleap (or leap) previous year; (c) in the annual composite of a leap year, day 335 always means November 30.
Saturation_Flag	uint8	0 -- 255	1	Unitless	None	The least significant bit to the most significant bit corresponds to bands 1, 2, 3, 4, 5, 61, 62, 7; with a bit set to 1 signifying saturation in that band and 0 not saturated.
DT_Cloud_State	uint8	0, 1, 2, 200	1	Unitless	255	Decision Tree Cloud Classification, 0 = not cloudy, 1 = cloudy, 2 = not cloudy but adjacent to a cloudy pixel, 200 = could not be classified reliably.
ACCA_State	uint8	0, 1	1	Unitless	255	ACCA Cloud Classification, 0 = not cloudy, 1 = cloudy.

Num_Of_Obs	uint16	0 -- 65535	1	Unitless	None	Number of observations considered over the compositing period.
Composite_Path	uint8	0 to 15	1	Unitless	255	Internal compositing algorithm pathway code.
Sensor	uint8	4, 5, or 7	1	Unitless	255	Landsat satellite that the pixel was selected from (4 = Landsat 4 TM, 5 = Landsat 5 TM, 7 = Landsat 7 ETM+)
Sensor_Zenith	int16	0 to 9000	0.01	Degrees	-32768	Sensor_Zenith (nadir = 0 degrees).
Sensor_Azimuth	int16	-18000 to 18000	0.01	Degrees	-32768	Sensor_Azimuth.
Solar_Zenith	int16	0 to 9000	0.01	Degrees	-32768	Solar_Zenith of pixel observation (directly overhead = 0 degrees).
NBAR_Solar_Zenith	int16	0 to 9000	0.01	Degrees	-32768	Solar_Zenith used for NBAR generation of pixel surface refelectance value.
Solar_Azimuth	int16	-18000 to 18000	0.01	Degrees	-32768	Solar_Azimuth.
L1T_Index	uint16	0 to 65534	1	Unitless	65535	Index to the L1T image that the pixel was selected from (reference L1T_Index_Metadata to find the Level 1T filename).
L1T_Column	uint16	0 to 10000	1	Unitless	65535	Pixel column number in the input Level 1T image.
L1T_Row	uint16	0 to 10000	1	Unitless	65535	Pixel row number in the input Level 1T image.

Table 3 GWELD product **metadata**

Name	Data Type	Valid Range	Units	Description
PRODUCT_VERSION	String	N/A	Unitless	Product version number (3.0., …) also reflected in the filename.
PGE_VERSION	String	N/A	Unitless	Internal version number of Product Generation Executive used to produce particular tile, not necessarily reflected in the filename.
INPUT_POINTER	String	N/A	Unitless	Internal list of input filenames used for production.
L1T_Index_Metadata	String	N/A	Unitless	List of Level 1T filenames and indices (reference with L1T_Index band pixel value) and the corresponding Level 1T image MTL file specified center solar zenith and solar azimuth values.
Mean_B1	Float	-0.1 to 1.0	Unitless	Mean band 1 reflectance computed from all non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile.
Mean_B2	Float	-0.1 to 1.0	Unitless	Mean band 2 reflectance computed from all non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile.
Mean_B3	Float	-0.1 to 1.0	Unitless	Mean band 3 reflectance computed from all non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile.
Mean_B4	Float	-0.1 to 1.0	Unitless	Mean band 4 reflectance computed from all non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile.
Mean_B5	Float	-0.1 to 1.0	Unitless	Mean band 5 reflectance computed from all non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile.
Mean_B6	Float	-200 to 300	Degrees Celsius	Mean band 6 brightness temperature computed from all non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile
Mean_B7	Float	-0.1 to 1.0	Unitless	Mean band 7 reflectance computed from all non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile.
Mean_NDVI	Float	-0.1 to 1.0	Unitless	Mean NDVI computed from all non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile.
Mean_Solar_Zenith	Float	0 to 90	Degrees	Mean solar zenith angle computed from all non-fill pixels in the tile.
Mean_NBAR_Solar_Zenith	Float	0 to 90	Degrees	Mean NBAR solar zenith angle computed from all non-fill pixels in the tile.

Percent_Saturated	Float	0.0 to 100.0	Percentage	Percentage of non-fill pixels in the tile that were flagged as saturated in any band (Saturation_Flag != 0).
Percent_ACCA_Cloudy	Float	0.0 to 100.0	Percentage	Percentage of non-fill pixels in the tile that were flagged as ACCA Cloudy (ACCA == 1).
Percent_DT_Cloudy	Float	0.0 to 100.0	Percentage	Percentage of non-fill pixels in the tile that were flagged as DT cloudy (ACCA == 1).
Mean_JDOY	Int	1 to 366	Day	Mean Julian Day of Year of non-fill pixels in the tile.
Min_JDOY	Int	1 to 366	Day	Minimum Julian Day of Year of non-fill pixels in the tile.
Max_JDOY	Int	1 to 366	Day	Maximum Julian Day of Year of non-fill pixels in the tile.
Number_Valid_Obs	Int	0 to 28037025	Count	The number of non-fill pixels in the tile.
Number_Valid_Noncloudy_Obs	Int	0 to 28037025	Count	The number of non-fill and non-cloudy (ACCA == 0 && DT != 1) pixels in the tile.
Count_L1T	Int	0 to 10000	Count	Count of the number of unique L1T images present in the tile.
Sensor_List	String	4 | 5 or 5 | 7	Unitless	List of Landsat satellites present in the tile.
Number_Valid_Sensor_Obs	String	0 to 28037025 | 0 to 28037025	Count	The number of non-fill pixels in the tile for each satellite indexed by Sensor_List.

3.2 Product reporting periods

The GWELD products are available as monthly and annual composited products. Both the monthly and annual products contain the same 24 bands (Table 2) and metadata (Table 3). They are defined in a temporally nested manner following climate modeling conventions where a year is defined by the December of the previous calendar year, and then January to November of the current year (Table 4).

Table 4 GWELD product types

Product Type	**Temporal Definition**
Annual	The preceding year's December through the current year's November.
Monthly	The days in each calendar month

Figures 10 and 11 show example global true color, red (0.63-0.69μm), green (0.53-0.61μm) and blue (0.45-0.52μm), browse images for the monthly and annual composites respectively. All the L1T TM/ETM+ data acquired in each temporal period are composited; for the longer periods more L1T data are available and so there are less gaps and less obvious cloudy data.

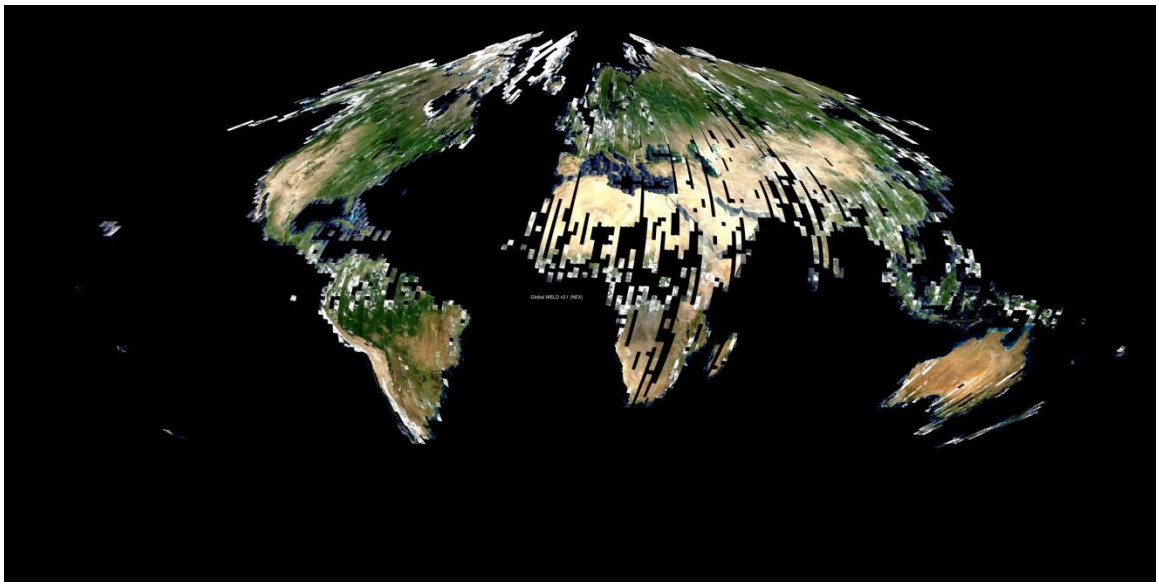


Figure 10 Example monthly GWELD composite (July 2000).

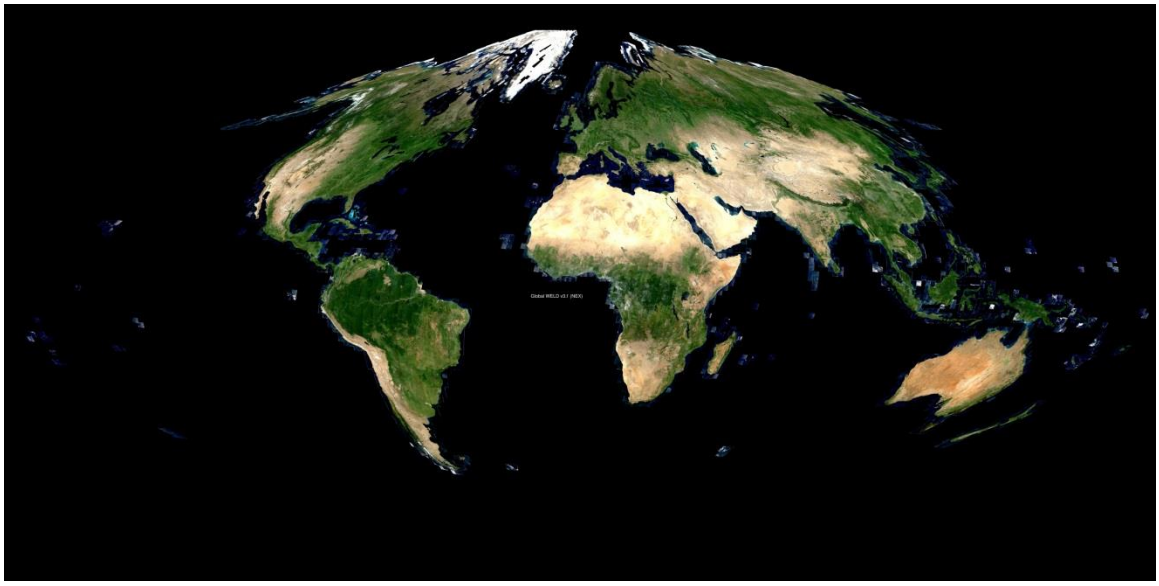


Figure 11 Example annual GWELD composite (Annual 2000).

3.3 Product map projection and tiling scheme

The GWELD products are defined in the same coordinate system and align precisely with the MODIS land products tiles. The GWELD products are defined in the equal area sinusoidal projection. The projection parameters for the USGS General Cartographic Tansformation Package (GCTP) are summarized in Table 5. The datum is World Geodetic System 84 (WGS84). The most upper left pixel coordinate is defined as:
ULX = -20015109.3557974174618721, ULY = 10007554.6778987087309361.

Table 5. Sinusoidal projection parameters (GCTP format)

Num	Parameter	Value	Description
0	Sphere	6371007.181	Radius of reference (meters) sphere
1-3		0.0	not used
4	CentMer	0.0	Longitude of the central meridian
5		0.0	not used
6	FE	0.0	False Easting in the same units as the sphere radius
7	FN	0.0	False Northing in the same units as the sphere radius
8-14		0.0	not used

The GWELD products are defined in tiles nested to the standard MODIS land product tiles (Figure 4). Figure 12 illustrates the MODIS land product tiles that are defined in a global non-overlapping grid of 36 × 18 tiles, each covering approximately 10º × 10º at the Equator (Wolfe et al. 1998). A table of the bounding MODIS tile coordinates is provided at https://landweb.modaps.eosdis.nasa.gov/cgi-bin/QA_WWW/newPage.cgi?fileName=sn_bound_10deg.

Within each MODIS 10º × 10º tile there are 7 × 7 GWELD product tiles (Figure 13). Each GWELD product tile is composed of 5295 × 5295 30 m pixels.

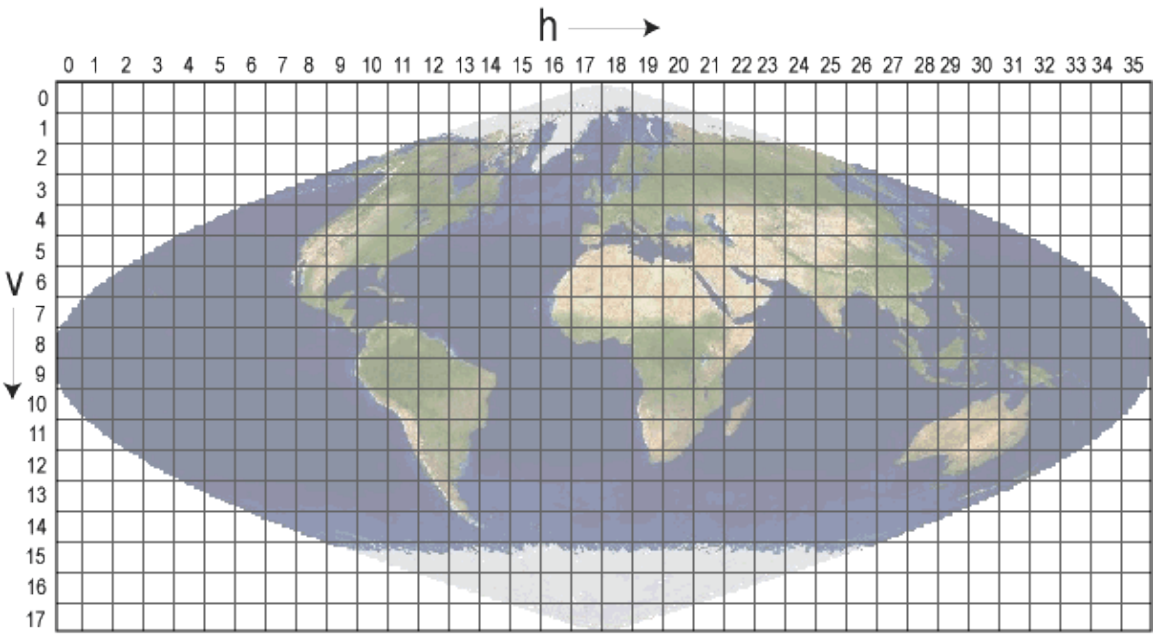


Figure 12 MODIS land product tiling scheme. Each 10º × 10º tile is referenced by a horizontal ($0 \leq h \leq 35$) and vertical ($0 \leq v \leq 17$) coordinate.

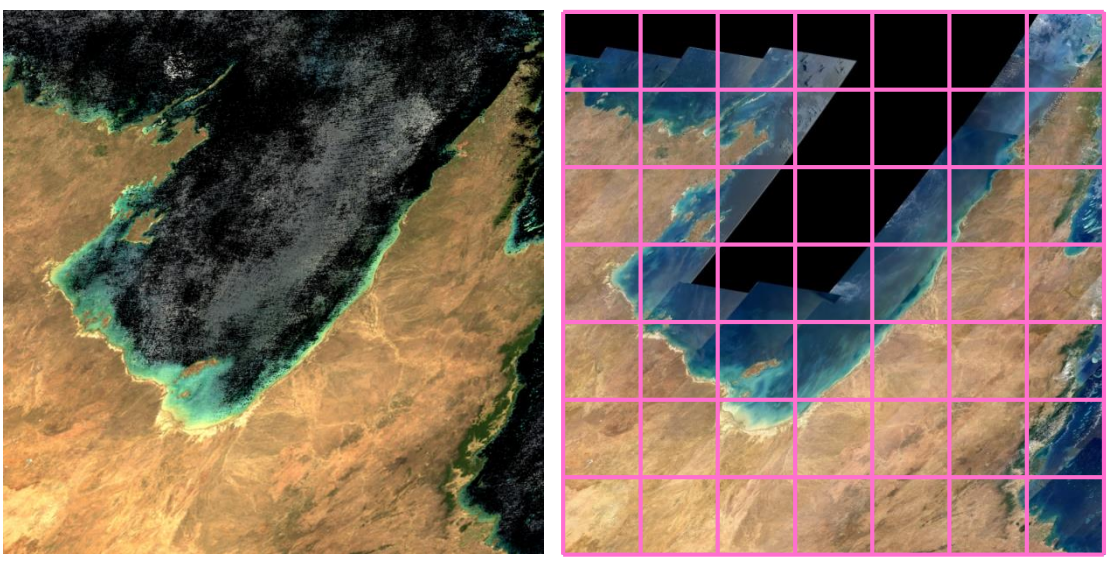


Figure 13 MODIS tile (h31v10), Gulf of Carpentaria, Australia, October 2009 data. *Left:* MODIS NBAR 500 m true color surface reflectance; *Right:* Landsat 5 and 7 GWELD NBAR 30 m true color surface reflectance composite, with the 7 × 7 global WELD tile boundaries superimposed (magenta). Each GWELD tile is referenced within the MODIS tile by a horizontal ($0 \leq x \leq 6$) and vertical ($0 \leq y \leq 6$) coordinate.

3.4 Product file naming convention

Table 6 Version 3.X GWELD Product filename convention - descriptive, simple, and amenable to scripting		
L<*ss*>.Globe.<*period*>.<*year*>.hh<*xx*>vv<*yy*>>.h<*x*>v<*y*>.doy<*min DOY*>to<*max DOY*>.NBAR.v<*version number*>.hdf		
	Valid Range	**Notes**
<ss>	04 / 05 / 07 / 45 / 57	Combination of contemporaneous Landsat sensors (4, 5, 7) used in the product.
<Period>	annual month01/month02/, …,/month12	Annual products are generated from a year of Landsat data sensed from December 1st of the previous year to November 30th of the current year. Monthly products generated from the Landsat data sensed in that month.
<Year>	1983, 1984, ... , 2011	Year the data sensed (monthly products). Year the January to November data sensed (annual products).
<xx>	00, 01, ..., 35	Horizontal MODIS land tile coordinate.
<yy>	00, 01, ..., 17	Vertical MODIS land tile coordinate.
<x>	0, 1, …, 6	Horizontal WELD tile coordinate within the MODIS land tile.
<y>	0, 1, …, 6	Vertical WELD tile coordinate within the MODIS land tile.
<min DOY>	001, 002, …, 366	Minimum non-fill Day_Of_Year pixel value present in the tile.
<max DOY>	001, 002, …, 366	Maximum non-fill Day_Of_Year pixel value present in the tile.
<Version Number>	3.0, …	Major and minor algorithm version changes reflected in the first and second digits respectively.

3.5 GWELD product data volume

The HDF format tiles are stored with HDF internal compression on and are typically 280 MB and 370 MB for each monthly and annual tile product, respectively. The annual products have greater file sizes since there are fewer fill value pixels. The total number of monthly composited files is around ~80,000/year and the total volume about >20TB/year. The total number of annual composited files is around ~8,000/year and the total volume about >3TB/year.

4.0 Known issues

The version 3.0 GWELD 2010 epoch products (36 monthly products and annual products for 2009, 2010 and 2011) are defined in HDF4, the other GWELD epochs are defined in HDF4EOS. Consequently, the version 3.0 GWELD 2010 epoch HDF product geographic coordinates cannot be read correctly by GDAL/ArcGIS. This is not an issue for the version 3.1 GWELD products.

The version 3.0 GWELD 2010 epoch product geographic coordinates (*ax*, *ay*) for the center of a pixel (*icol*, *irow*) can be determined in the following ways:

(1) using the tile corner pixel coordinates stored in the metadata
The tile corner coordinates can be extracted using command line "ncdump –h $file | grep sin". For example:

>ncdump –h L57.Globe.month01.2009.hh25vv04.h6v5.doy002to027.NBAR.v3.0.hdf | grep sin
will output the following:

```
:sinus_UL_X = 8736753.638365664 ;
:sinus_UL_Y = 4765502.598832616 ;
:sinus_UR_X = 8895603.638365664 ;
:sinus_UR_Y = 4765502.598832616 ;
:sinus_LL_X = 8736753.638365664 ;
:sinus_LL_Y = 4606652.598832616 ;
:sinus_LR_X = 8895603.638365664 ;
:sinus_LR_Y = 4606652.598832616 ;
```

Note that these coordinates are defined for the pixel corners, i.e., the UL pixel coordinate is the upper left coordinate of the most upper left pixel. Thus:

```
ax = sinus_UL_X + 30*(icol+0.5);
ay = sinus_UL_Y - 30*(irow+0.5);

// +0.5 means these are pixel center coordinates.
```

(2) using the original coordinates defined for the Sinusoid projection and tile ID (*modis_tile_h*, *modis_tile_v*, *mini_tile_h* and *mini_tile_v*):

```
sinus_global.sinus_ULX = -20015109.3557974174618721;
sinus_global.sinus_ULY =  10007554.6778987087309361;
sinus_global.modis_tile_size = 1111950.5197665231923262;
tiledim_nrow=5295;
tiledim_ncol=5295;
ax = sinus_global.sinus_ULX + (double)(modis_tile_h * sinus_global.modis_tile_size) +
```

```
                    ((double)(mini_tile_h * tiledim_ncol + icol) + 0.5) * 30;  /* pixel
center */
ay = sinus_global.sinus_ULY - (double)(modis_tile_v * sinus_global.modis_tile_size) -
                    ((double)(mini_tile_v * tiledim_nrow + irow) + 0.5) * 30;

 // +0.5 means these are pixel center coordinates.
```

(3) using a conversion tool

The conversion tool between the geographic location (latitude and longitude) and pixel location (tile ID and pixel column and row) is available at

https://globalmonitoring.sdstate.edu/projects/weldglobal/gweld.html#software_tools

For example:

> *./global.weld.tile.calculator.v2 forward tp.glw030 116.381 39.9769*
Output: *longitude 116.38100000 latitude 39.97690000 -> WELD.GLOBAL HH 26 VV 5 H 6 V 0 column 2266.91326 row 85.85698*
> *./global.weld.tile.calculator.v2 inverse 9 4 0 4 2467 5043*
Output: *WELD.GLOBAL HH 9 VV 4 H 0 H 4 column 2467.00000 row 5043.00000 -> longitude -122.00117713 latitude 42.92520343*

5.0 LITERATURE CITED

Arvidson, T., Gasch, J., & Goward, S. N. (2001) Landsat 7's long-term acquisition plan---an innovative approach to building a global imagery archive. *Remote Sensing of Environment, 78*, 13-26.

Arvidson, T., Goward, S. N., Gasch, J., & Williams, D. (2006). Landsat-7 long-term acquisition plan: development and validation. *Photogrammetric Engineering & Remote Sensing, 72* (10), 1137-1146.

Bindschadler, R., Vornberger, P., Fleming, A., Fox, A., Mullins, J., Binnie, D., Paulsen S. J., Granneman B., & Gorodetzky, D. (2008). The Landsat image mosaic of Antarctica. *Remote Sensing of Environment, 112*(12), 4214-4226.

Blanco-Muriel, M., Alarcón-Padilla, D. C., López-Moratalla, T., & Lara-Coira, M. (2001). Computing the solar vector. *Solar Energy, 70*(5), 431-441.

Breiman, L., Friedman, J., Olshen, R., Stone, C. (1984). *Classification and Regression Trees*. Wadsworth and Brooks/Cole, Monterey, CA.

Breiman, L. (1996). Bagging predictors. *Machine learning, 24*(2), 123-140.

Cahalan, R. F., Oreopoulos, L., Wen, G. Y., Marshak, A., Tsay, S. C., & DeFelice, T. (2001). Cloud characterization and clear-sky correction from Landsat-7. *Remote Sensing of Environment, 78*, 83−98.

Chander, G., Huang, C., Yang, L., Homer, C., & Larson, C. (2009a). Developing consistent Landsat data sets for large area applications: The MRLC 2001 protocol. *IEEE Transactions on Geoscience and Remote Sensing, 6*, 777−781.

Chander, G., Markham, B. L., & Helder, D. L. (2009b). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment, 113*, 893−903.

Cihlar, J. (1994). Detection and removal of cloud contamination from AVHRR images. *IEEE Transactions on Geoscience and Remote Sensing, 32*(3), 583–589.

Cihlar, J., Manak, D., D'Iorio, M. (1994). Evaluation of compositing algorithms for AVHRR data over land. *IEEE Transactions on Geoscience and Remote Sensing, 32*, 427-437.

Danaher, T., Wu, X. and Campbell, N. (2001). Bi-directional Reflectance Distribution Function Approaches to Radiometric Calibration of Landsat TM imagery. *Proc. IEEE Geoscience and Remote Sensing Symposium (IGARSS 2001)*, 6, 2654–2657.

Gao, F., Jin, Y., Li, X., Schaaf, C. B., & Strahler, A. H. (2002). Bidirectional NDVI and atmospherically resistant BRDF inversion for vegetation canopy. *IEEE Transactions on Geoscience and Remote Sensing, 40*, 1269−1278.

Global Marketing Insights, Inc. Remote Sensing Data, *2008 USGS Africa Remote Sensing Study, Aerial and Spaceborne Ten-Year Trends*, 2009, available online at http://www.globalinsights.com/USGS2008AfricaRSS.pdf

Goward, S. N., Masek, J. G., Williams, D. L., Irons, J. R., & Thompson, R. J. (2001). The Landsat 7 mission, Terrestrial research and applications for the 21st century. *Remote Sensing of Environment, 78*, 3-12.

Hansen, M. C., Egorov, A., Roy, D. P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., & Loveland, T. R. (2011). Continuous fields of land cover for the conterminous United States using Landsat data: First results from the Web-Enabled Landsat Data (WELD) project. *Remote Sensing Letters, 2*(4), 279-288.

Hansen, M. C., Roy, D. P., Lindquist, E., Adusei, B., Justice, C. O., & Altstatt, A. (2008). A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change and preliminary results for Central Africa. *Remote Sensing of Environment, 112*, 2495-2513.

Hansen, M. C., DeFries, R. S., Townshend, J. R., Carroll, M., DiMiceli, C., & Sohlberg, R. A. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions, 7*(10), 1-15.

Hansen, M. C., DeFries, R. S., Townshend, J. R. G., Marufu, L., & Sohlberg, R. (2002). Development of a MODIS tree cover validation data set for Western Province, Zambia. *Remote Sensing of Environment, 83*(1-2), 320-335.

Hansen, M. C., Townshend, J. R., DeFries, R. S., & Carroll, M. (2005). Estimation of tree cover using MODIS data at global, continental and regional/local scales. *International Journal of Remote Sensing, 26*(19), 4359-4380.

Hassett, P.J., Johnson, R.L. (1984). LANDSAT-5 orbit adjust maneuver report, NASA contract NAS 5-27888, Task Assignment 14300, Computer Sciences Corp (63.p).

Huete, A., (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment, 25*, 295–309.

Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote sensing of environment, 83*(1-2), 195-213.

Holben, B. (1986). Characteristics of maximum-value composite images from temporal AVHRR data. *International Journal of Remote Sensing, 7*, 1417-1434.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote E., Reagan J. A., Kaufman Y. J., Nakajima T., Lavenu F., Jankowiak I., & Lavenu, F. (1998). AERONET—A federated instrument network and data archive for aerosol characterization. *Remote sensing of environment, 66*(1), 1-16.

Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M. (2004). Development of a 2001 national land-cover database for the United States, *Photogrammetric Engineering and Remote Sensing, 70,* 829-840.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., Mckerrow, A., Vandriel, J.N., & Wickham, J. (2007). Completion of the 2001 national land cover database for the counterminous United States. *Photogrammetric Engineering and Remote Sensing, 73*(4), 337.

Irish, R. I., (2000). Landsat7 automatic cloud cover assessment. *In Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VI, Proc. SPIE*, vol. 4049, pp. 348-355.

Irish, R. I. , Barker, J. L., Goward, S. N., and Arvidson, T. (2006). Characterization of the Landsat-7 ETM+ automated cloud-cover assessment (ACCA) algorithm. *Photogrammetric Engineering & Remote Sensing, 72*, 1179 – 1188.

Irons, J. R., & Masek, J. G. (2006). Requirements for a Landsat Data Continuity Mission. *Photogrammetric Engineering and Remote Sensing, 72* (10), 1102-1108.

Ju, J., & Roy, D. P. (2008). The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sensing of Environment, 112*(3), 1196-1211.

Ju, J., Roy, D.P., Vermote, E., Masek, J., Kovalskyy, V. (2012). Continental-scale validation of MODIS-based and LEDAPS Landsat ETM+ atmospheric correction methods, *Remote Sensing of Environment*, 122, 175-184.

Justice, C., Townshend, J., Vermote, E., Masuoka, E., Wolfe, R., Saleous, N., Roy, D., Morisette, J. (2002). An overview of MODIS Land data processing and product status. *Remote Sensing of Environment, 83*, 3-15.

Justice, C., Belward, A., Morisette, J., Lewis, P., Privette, J., & Baret, F. (2000). Developments in the'validation'of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing, 21*(17), 3383-3390.

Kaufman, Y. J., Tanré, D., Remer, L. A., Vermote, E. F., Chu, A., & Holben, B. N. (1997). Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer. *Journal of Geophysical Research: Atmospheres, 102*(D14), 17051-17067.

Konecny, G. (1979). Methods and possibilities for digital differential rectification. *Photogrammetric Engineering and Remote Sensing, 6*, 727-734.

Kotchenova, S., Vermote, E., R. Matarrese, R., & and Klemm, F., Jr. (2006). Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance. *Applied Optics, 45*, 6762-6774.

Kovalskyy, V. and Roy, D.P. (2013). The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30m Landsat data product generation, *Remote Sensing of Environment*, 130, 280-293.

Landsat 7 ETM+ Geometric ATBD is on line at: http://landsathandbook.gsfc.nasa.gov/handbook/pdfs/L7_geometry_ATBD.pdf

Lee, D. S., Storey, J. C., Choate, M. J., & Hayes, R. (2004). Four years of Landsat-7 on-orbit geometric calibration and performance. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 2786−2795.

Liu, H. Q., & Huete, A. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE transactions on geoscience and remote sensing, 33*(2), 457-465.

Loveland, T.R. and Dwyer, J.L. (2012). Landsat: Building a strong future. *Remote Sensing of Environment, 122*, 22-29.

Markham, B., Goward, S., Arvidson, T., Barsi, J., & Scaramuzza, P. (2006). Landsat-7 long-term acquisition plan radiometry-evolution over time. *Photogrammetric Engineering & Remote Sensing, 72*(10), 1129-1135.

Markham, B. L., & Helder, D. L. (2012). Forty-year calibrated record of earth-reflected radiance from Landsat: A review. *Remote Sensing of Environment, 122*, 30-40.

Markham, B.; Barsi, J.; Kvaran, G.; Ong, L.; Kaita, E.; Biggar, S.; Czapla-Myers, J.; Mishra, N.; Helder, D. (2014). Landsat-8 operational land imager radiometric calibration and stability. *Remote Sensing, 6*(12), 12275-12308.

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

Morfitt, R., Barsi, J., Levy, R., Markham, B., Micijevic, E., Ong, L., Scaramuzza, P., & Vanderwerff, K. (2015). Landsat-8 Operational Land Imager (OLI) radiometric performance on-orbit. *Remote Sensing, 7*(2), 2208-2237.

Mulcahy, K. (2000). Two new metrics for evaluating pixel-based change in data sets of global extent due to projection transformation. *Cartographica: The International Journal for Geographic Information and Geovisualization, 37*(2), 1-12.

Ouaidrari, H., and Vermote, E.F. (1999). Operational atmospheric correction of Landsat TM data. *Remote Sensing of Enviroment, 70*, 4–15.

Platnick, S., King, M.D., Ackerman, S.A., Menzel, W. P. , Baum, B.A., Riédi, J.C., Frey, R.A. (2003). The MODIS cloud products: Algorithms and examples from Terra, *IEEE Transactions on Geoscience and Remote Sensing, 41*, 459–473.

Reda, I., Andreas, A., 2005. Solar position algorithm for solar radiation applications. Internal report, NREL/TP-560-343-2, National Renewable Energy Laboratory. 56p.

Roy, D.P. (1997). Investigation of the maximum normalised difference vegetation index (NDVI) and the maximum surface temperature (Ts) AVHRR compositing procedures for the extraction of NDVI and Ts over forest, *International Journal of Remote Sensing, 18*, 2383-2401.

Roy, D. P. (2000). The impact of misregistration upon composited wide field of view satellite data and implications for change detection. *IEEE Transactions on geoscience and remote sensing, 38*(4), 2017-2032.

Roy, D. P., Borak, J. S., Devadiga, S., Wolfe, R. E., Zheng, M., & Descloitres, J. (2002). The MODIS land product quality assessment approach. *Remote Sensing of Environment, 83*(1-2), 62-76.

Roy, D.P., Lewis, P., Schaaf, C., Devadiga, S., Boschetti, L. (2006). The Global impact of cloud on the production of MODIS bi-directional reflectance model based composites for terrestrial monitoring. *IEEE Geoscience and Remote Sensing Letters, 3*,452-456.

Roy, D. P., Ju, J., Lewis, P., Schaaf, C., Gao, F., Hansen, M., & Lindquist, E. (2008). Multi-temporal MODIS–Landsat data fusion for relative radiometric normalization, gap filling, and prediction of Landsat data. *Remote Sensing of Environment, 112*(6), 3112-3130.

Roy, D.P., Ju, J., Kline, K., Scaramuzza, P.L., Kovalskyy, V., Hansen, M.C., Loveland, T.R., Vermote, E.F., and Zhang, C. (2010). Web-enabled Landsat Data (WELD) Preliminary Results: Landsat ETM+ Composited Mosaics of the Conterminous United States. *Remote Sensing of Environment, 114*, 35-49.

Roy, D.P., Wulder, M.A., Loveland, T.R., Woodcock, C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C. B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z. P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne,

R.H., Zhu, Z. (2014a). Landsat-8: science and product vision for terrestrial global change research. *Remote Sensing of Environment, 145*, 154–172.

Roy, D. P., Qin, Y., Kovalskyy, V., Vermote, E. F., Ju, J., Egorov, A., Hansen M. C., Kommareddy I., & Yan, L. (2014b). Conterminous United States demonstration and characterization of MODIS-based Landsat ETM+ atmospheric correction. *Remote Sensing of Environment, 140*, 433-449.

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

Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., & Egorov, A. (2016b). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment, 185*, 57-70.

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.

Saalfeld, A. (1985). A fast rubber-sheeting transformation using simplicial coordinates. *The American Cartographer, 12*(2), 169-173.

Schaaf, C., Gao, F., Strahler, A., Lucht, W., Li, X., Tsang, T., Strugnell, N., Zhang, X., Jin, Y., Muller, J-P., Lewis, P., Barnsley, M., Hobson, P., Disney, M., Roberts, G., Dunderdale, M., d'Entremont, R., Hu, B., Liang, S., Privette, J., Roy, D. (2002). First operational BRDF, albedo and nadir reflectance products from MODIS. *Remote Sensing of Environment, 83*, 135-148.

Schaepman-Strub, G., Schaepman, M. E., Painter, T. H., Dangel, S., and Martonchik, J. V. (2006). Reflectance quantities in optical remote sensing–definitions and case studies. *Remote Sensing of Environment, 103*, 27-42.

Seong, J. C., Mulcahy, K. A., & Usery, E. L. (2002). The sinusoidal projection: A new importance in relation to global image data. *The Professional Geographer, 54*(2), 218-225.

Thorne, K., Markharn, B., Barker, P. S., & Biggar, S. J. P. E. (1997). Radiometric calibration of Landsat. *Photogrammetric Engineering & Remote Sensing, 63*(7), 853-858.

Townshend, J. R., & Justice, C. O. (1988). Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations. *International Journal of Remote Sensing, 9*(2), 187-236.

Townshend, J. R. (1994). Global data sets for land applications from the Advanced Very High Resolution Radiometer: an introduction. *International Journal of Remote Sensing, 15*(17), 3319-3332.

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote sensing of Environment, 8*(2), 127-150.

Tucker, C. J., Grant, D. M., & Dykstra, J. D. (2004). NASA's global orthorectified Landsat data set. *Photogrammetric Engineering & Remote Sensing, 70*(3), 313-322.

Tucker, C. J., Pinzon, J. E., Brown, M. E., Slayback, D. A., Pak, E. W., Mahoney, R., ... & El Saleous, N. (2005). An extended AVHRR 8‐km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. *International Journal of Remote Sensing*, 26(20), 4485-4498.

Vermote, E. F., El Saleous, N., & Justice, C. (2002). Atmospheric correction of the MODIS data in the visible to middle infrared: First results. *Remote Sensing of Environment*, *83*(1–2), 97−111.

Vermote, EF, Kotchenova, S (2008). Atmospheric correction for the monitoring of land surfaces. *Journal of Geophysical Research-Atmospheres, 113*(D23), D23S90.

Williams, D. L., Goward, S., & Arvidson, T. (2006). Landsat: yesterday, today, and tomorrow. *Photogrammetric Engineering & Remote Sensing, 72*(10), 1171-1178.

Wittich, K. P., & Kraft, M. (2008). The normalised difference vegetation index obtained from agrometeorological standard radiation sensors: a comparison with ground-based multiband spectroradiometer measurements during the phenological development of an oat canopy. *International Journal of Biometeorology, 52*(3), 167-177.

Wolfe, R. E., Roy, D. P., & Vermote, E. (1998). MODIS land data storage, gridding, and compositing methodology: Level 2 grid. *IEEE Transactions on Geoscience and Remote Sensing, 36*(4), 1324-1338.

Woodcock, C.E., Allen, A.A., Anderson, M., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Goward, S.N., Helder, D., Helmer, E., Nemani, R., Oreapoulos, L., Schott, J., Thenkabail, P.S., Vermote, E.F., Vogelmann, J., Wulder, M.A., Wynne, R. (2008). Free access to Landsat imagery. *Science, 320*(5879), 1011-1011.

Wulder, M.A., E. Loubier, and D. Richardson (2002). A Landsat-7 ETM+ Orthoimage Coverage of Canada. *Canadian Journal of Remote Sensing, 28*, 667-671.

Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., Cohen W.B., Loveland T.R., & Woodcock, C. E. (2008). Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment, 112*(3), 955-969.

Wulder, M. A., White, J. C., Loveland, T. R., Woodcock, C. E., Belward, A. S., Cohen, W. B., Fosnight E. A., Shaw J., Masek, J. G., & Roy, D. P. (2016). The global Landsat archive: Status, consolidation, and direction. *Remote Sensing of Environment, 185*, 271-283.

Wulder, M.A., Coops, N.C., Roy, D.P., White, J.C., & Hermosilla, T. (2018). Land Cover 2.0. *International Journal of Remote Sensing, 39*(12), 4254-4284.

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*(3), 1410-1418.

Zhang, H.K. and Roy, D.P. (2017). Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification, *Remote Sensing of Environment*. 197, 15-34

Zhang, H. K., Roy, D. P., Yan, L., Li, Z., Huang, H., Vermote E., Skakun S., & Roger J. (2018). Characterization of Sentinel-2A and Landsat-8 top of atmosphere, surface, and nadir BRDF adjusted reflectance and NDVI differences. *Remote Sensing of Environment, 215*, 482-494.

Zobrist, A. L., Bryant, N. A., & McLeod, R. G. (1983). Technology for large digital mosaics of Landsat data. *Photogrammetric engineering and remote sensing, 49*(9), 1325-1335.

Acknowledgements

The global WELD project is funded by the National Aeronautics and Space Administration (NASA) Making Earth System Data Records for Use in Research Environments (MEaSUREs) program under Cooperative Agreement NNX13AJ24A. The NASA MEaSUREs Program Officers were Martha Maiden and Lucia Tsaoussi.

The GWELD products were generated on the NASA Earth Exchange (NEX) high performance super computer (https://nex.nasa.gov/nex/).

The free Landsat data access and the GWELD product distribution hardware and maintenance is funded by the United States Geological Survey (USGS).

We acknowledge and thank the participation and support (listed in alphabetical order) of: Hirofumi Hashimoto, Junchang Ju, Valeriy Kovalskyy, Zhongbin Li, Tom Maiersperger, Andrew Michaelis, Rama Nemani, and Petr Votava. This ATBD could not have been written without regular visits to http://www.cottonwoodcoffee.com/.