Algorithm Theoretical Basis Document

Web Enabled Landsat Data (WELD) Products

(Document Version 1.0, February 2011)

David Roy, Junchang Ju, Indrani Kommareddy, Matthew Hansen, Eric Vermote,
Chunsun Zhang, Anil Kommareddy

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.

The NASA funded Web-enabled Landsat Data (WELD) project is systematically generating 30m composited Landsat Enhanced Thematic Mapper Plus (ETM+) mosaics of the conterminous United States and Alaska from 2005 to 2012. The WELD 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 WELD products are processed so that users do not need to apply the equations and spectral calibration coefficients and solar information to convert the ETM+ digital numbers to reflectance and brightness temperature, and successive products are defined in the same coordinate system and align precisely, making them simple to use for multi-temporal applications. They aim to provide the first instance of continental-scale science-quality Landsat data with a level of pre-processing comparable to the NASA MODIS land products.

Cite as
Roy, D.P., Ju, J, Kommareddy, I., Hansen, M, Vermote, E. Zhang, C., Kommareddy, A.,
Web Enabled Landsat Data (WELD) Products - Algorithm Theoretical Basis Document,
February 2011, http://globalmonitoring.sdstate.edu/projects/weld/WELD_ATBD.pdf

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

Table of Contents

1.0 OVERVIEW
- 1.1 Project Goals
- 1.2 Rationale and Research Applications

2.0 VERSION 1.5 WELD PRODUCT THERORETICAL DESCRIPTION
- 2.1 Input Data
- 2.2 Top of Atmosphere Reflectance and Brightness Temperature Computation
- 2.3 Normalized Difference Vegetation Index Computation
- 2.4. Band Saturation Computation
- 2.5 Cloud Masking
  - 2.5.1 ACCA Cloud Detection
  - 2.5.2 Classification Tree Cloud Detection
- 2.6 Angular Geometry Computation
- 2.7 Reprojection, Resampling and Tiling
- 2.8 Compositing
- 2.9 Browse Generation

3.0 VERSION 1.5 PRODUCT DOCUMENTATION
- 3.1 WELD Product Types
- 3.2 WELD Product Contents
- 3.3 WELD Product Map Projections
- 3.4 WELD Product Data Formats
- 3.5 WELD Product Data Volume

4.0 PRODUCT MANAGEMENT AND DISTRIBUTION

4.1 Data Management Plan

4.2 Production Hardware

4.3 WELD Product Distribution

4.3.1 WELD Product FTP Distribution

4.3.2 What You See Is What You Get (WYSIWYG) WELD Internet Distribution

4.3.3 WELD Open Geospatial Consortium Browse Imagery Distribution

4.4 WELD Product Distribution Metrics

4.5 WELD Product Long Term Archive Strategy

5.0 PLANNED VERSION 2.0 WELD PRODUCTS

5.1 Atmospheric Correction of the Top of Atmosphere Reflectance Bands

5.2 Reflective Wavelength Gap Filling

5.3. Thermal Wavelength Gap Filling

5.4 Reflective Wavelength Radiometric Normalization

5.5 Annual Percent Tree, Bare Ground, Vegetation and Water Classification

6.0 QUALITY ASESSMENT PLAN

7.0 VALIDATION PLAN

8.0 USER SUPPORT

9.0 GLOBAL WELD PATHFINDING

10.0 LITERATURE CITED

Acknowledgements

1.0 OVERVIEW

1.1 Project goals

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. The Web-enabled Landsat Data (WELD) project seeks to contribute to the Land measurement theme of the MEaSUREs program, working at high spatial resolution and using state of the art and validated MODIS land products to systematically generate "seamless" consistent mosaiced Landsat Enhanced Thematic Mapper plus (ETM+) data sets with per-pixel quality assessment information at weekly, monthly, seasonal (3 month), and annual time scales. In addition, annual percent tree, bare ground, vegetation and water will be generated.

The WELD products are developed specifically to provide consistent 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 WELD products are processed so that users do not need to apply the equations and spectral calibration coefficients and solar information to convert the ETM+ digital numbers to reflectance and brightness temperature, and successive products are defined in the same coordinate system and align precisely, making them simple to use for multi-temporal applications.

The WELD processing, based on heritage techniques and contemporaneous fusion of MODIS data, is applied to all Landsat ETM+ L1T acquisitions with cloud cover < 80% sensed over the conterminous United States (CONUS) and Alaska, approximately 8,000 and 1,800 acquisitions per year respectively. Weekly, monthly, seasonal and annual CONUS and Alaska WELD products will be generated for at least an 8-year period (2005-2012) and made freely available to the user community. The geographic scope of

the WELD products will be increased in the last year of the grant period to pathfind the challenges to expanding the WELD production to global scale.

1.2 Rationale and Research Applications

The Landsat satellite series, operated by the U.S. Department of Interior / U.S. Geological Survey (USGS) Landsat project, with satellite development and launches supported by National Aeronautics and Space Administration (NASA), represents the longest temporal record of space-based land observations (Williams et al. 2006). In January 2008, NASA and the USGS implemented a free Landsat Data Distribution Policy that provides Level 1 terrain corrected data for the entire U.S. Landsat archive at the United States Geological Survey (USGS) Center for Earth Resources Observation and Science (EROS). Free Landsat data will enable reconstruction of the history of the Earth's land surface back to 1972, with appropriate spatial resolution to enable chronicling of both anthropogenic and natural changes (Townshend and Justice, 1988), during a time when the human population has doubled and the impacts of climate change have become manifest (Woodcock et al. 2008).

Regional mosaics of Landsat imagery have been developed previously to meet national monitoring and reporting needs across land-use and resource sectors (Wulder et al. 2002, Hansen et al. 2008). Large volume Landsat processing was developed by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) that processed over 2,100 Landsat Thematic Mapper and ETM+ acquisitions to provide wall-to-wall surface reflectance coverage for North America for the 1990s and 2000s (Masek et al. 2006). More recently, a Landsat mosaic of Antarctica was generated from nearly 1,100 Landsat ETM+ austral summer acquisitions (Bindschadler et al. 2008). Global Landsat data sets have been developed through NASA and USGS data buys but only a fraction of the U.S. Landsat archive has been exploited (Tucker et al. 2004, Masek 2007, Gutman et al., 2008). These data sets, originally called Geocover, have been reprocessed and are now termed the Global Land Survey (GLS) datasets. The GLS datasets provide global, orthorectified, low cloud cover Landsat imagery centered on the years 1975, 1990, 2000,

and 2005 with a preference for leaf-on conditions (Gutman et al. 2008). Collectively, the GLS data sets are designed to provide a consistent set of observations to assess land-cover changes at a quasi-decadal scale. However, the GLS data sets do not provide spatially coherent data as neighboring scenes may have been acquired on different dates with different surface, atmospheric and solar illumination conditions, nor do they provide sufficient temporal resolution to capture vegetation phenology and surface changes required for optimal land cover classification and other terrestrial monitoring applications. With the advent of free Landsat data it becomes feasible instead to apply pixel based temporal compositing approaches – this is the basis of the WELD approach.

The capability to monitor the surface at Landsat resolution using the WELD data products is unprecedented. National science initiatives such as NASA's Land Cover and Land Use Change, Terrestrial Ecology, and North America Carbon programs, and the U.S. Strategic Plan for the Climate Change Science Program (CCSP), as well as international programs including the Global Observation of Forest Cover-Land Cover Dynamics (GOFC-GOLD) and the Group on Earth Observations (GEO), are calling for timely, accurate, land cover assessments at high spatial resolution (GCOS 2003, CCSP 2003, GEO 2005). Recently, the United Nations Collaborative Programme on Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD) stated the need "To establish robust and transparent national forest monitoring systems … using a combination of remote sensing and ground-based forest carbon inventory approaches" (U.N. 2010); Landsat data are recognized as the most suitable satellite data to assess historical deforestation rates and patterns (REDD Sourcebook, 2009). The case for the development of a high spatial resolution land cover land use change Earth System Data Record (ESDR) has been articulated in a NASA White Paper and it is needed to address several of the fundamental science questions posed in the NASA Research Plan and US Climate Change Science Program (Masek et al. 2006b). In addition, there is an established recognition of the opportunities that Landsat scale data provide for environmental and public sector applications (National Academy of Sciences 2002; Rowland et al. 2007, Global Marketing Insights 2009; Williams et al. 2006, Wulder et al. 2008).

3.0 VERSION 1.5 WELD PRODUCT THERORETICAL DESCRIPTION

2.1 Input Data

The WELD products are made from Landsat 7 Enhanced Thematic Mapper plus (ETM+) acquisitions. The Landsat 7 ETM+ was launched in 1999 and has a 15° field of view that captures approximately 183km x 170km scenes defined in a Worldwide Reference System of path (groundtrack parallel) and row (latitude parallel) coordinates (Arvidson *et al*. 2001). Adjacent Landsat orbit paths are sensed 7 days apart and the same orbit path is sensed every 16 days, i.e., providing a 16 day revisit capability (Figure 1).

Every sunlit scene (solar zenith angle <75° in the Northern hemisphere) overpassed over the conterminous U.S. and main islands are acquired and archived at the USGS EROS (Ju and Roy 2008). 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.

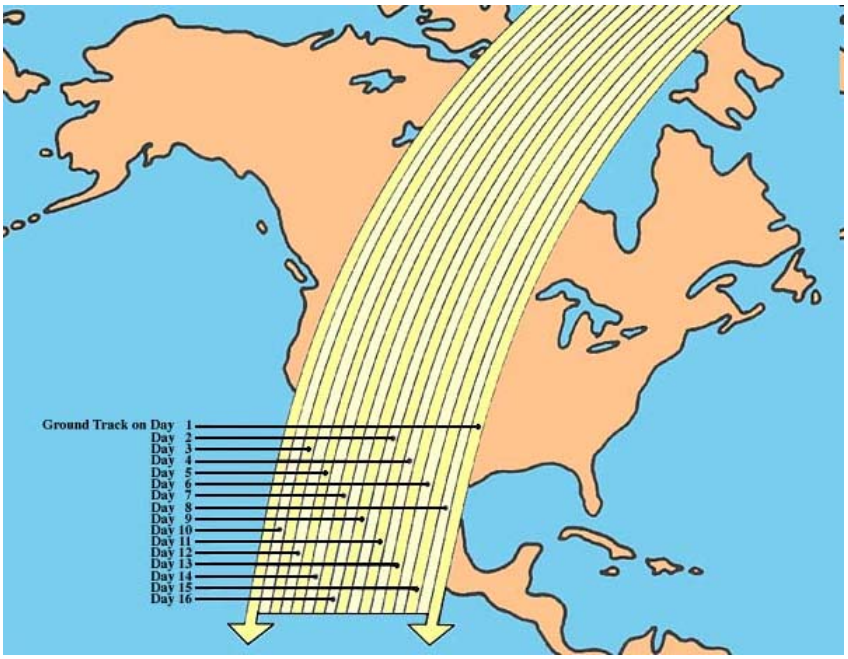


Figure 1 Landsat Orbit Geometry / Swath Pattern (from http://landsathandbook.gsfc.nasa.gov/handbook/handbook_htmls/chapter5l)

Figure 2 shows the conterminous United States (CONUS) and Alaska WELD study area, defined by 455 and 213 Landsat path/row coordinates, covering about 11,000,000,000 and 3,100,000,000 30m land pixels respectively.

Figure 2 WELD study area: Landsat Path and Row map for the conterminous United States (CONUS) and Alaska

The Landsat data are processed at the USGS EROS to Level 1 terrain corrected (L1T) level. The L1T data are available in GeoTIFF format in the Universal Transverse Mercator (UTM) map projection with WGS84 datum which is compatible with heritage GLS and Landsat MSS data sets (Tucker et al. 2004). 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. The L1T geolocation error in the conterminious United States

(CONUS) is less than 30m even in areas with substantial terrain relief (Lee et al. 2004). While most Landsat data are processed as L1T, certain acquisitions do not have sufficient ground control or elevation data necessary for precision or terrain correction respectively. In these cases, the best level of correction are applied and the data are 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 the L1T data are used to make the WELD products in order to reduce the impact of L1G misregistration errors on the WELD monthly, seasonal and annual composites (Roy 2000).

Landsat acquisitions with cloud cover less than 40% are processed by the U.S. Landsat project as they are acquired. The cloud cover of each acquisition is estimated operationally by the automatic cloud cover assessment algorithm (ACCA) (Irish et al. 2006). Users may request any other scene in the archive to be processed and made available at no cost via the Internet.

The WELD project staff manually order all ETM+ acquisitions with cloud cover between 40% and 80%. Once these, and the ETM+ acquisitions with cloud cover less than 40% are L1T processed, they are copied automatically via dedicated file transfer protocol from the USGS EROS to WELD project computers. Approximately 8,000 and 1,800 ETM+ L1T processed acquisitions are obtained per year for the CONUS and Alaska respectively.

2.2 Top of Atmosphere Reflectance and Brightness Temperature Computation

All the Landsat ETM+ bands, except the 15m panchromatic band are processed, i.e., the 30m 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 60m thermal (10.40-12.50μm) low and high gain bands are processed.

The spectral radiance sensed by each ETM+ detector is stored as an 8-bit digital number (Markham et al. 2006). The digital numbers are converted to spectral radiance (units: W

m^{-2} sr^{-1} μm^{-1}) using the sensor calibration gain and bias coefficients derived from the ETM+ L1T file metadata. This conversion minimizes remote sensing variations introduced by changes in the instrument radiometric calibration, sun-earth distance, the solar geometry, and exoatmospheric solar irradiance arising from spectral band differences (Chander et al. 2009).

The radiance sensed at the Landsat reflective and thermal wavelengths is then converted to reflectance (unitless) and brightness temperature (units: kelvins) respectively to provide data that has physical meaning and, for example, can be compared with laboratory and ground based measurements, model outputs, and data from other satellite sensors (Masek et al. 2006), and importantly provides data that can be used to derive higher level geo-physical and bio-physical products (Justice et al. 2002).

The radiance sensed in the Landsat reflective wavelength bands, i.e., the 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, are converted to top of atmosphere reflectance using standard formula as:

$$\rho_\lambda = \frac{\pi \cdot L_\lambda \cdot d^2}{ESUN_\lambda \cdot \cos\theta_s} \qquad [1]$$

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). The quantities $ESUN_\lambda$ and d are tabulated by Chander et al. (2009).

The TOA reflectance computed as [1] is the TOA bi-directional reflectance factor and can be greater than 1, for example, due to specular reflectance over snow or water under certain solar and viewing geometries (Schaepman-Strub et al. 2006). In addition, due to instrument artifacts not accommodated for by the calibration, the retrieved TOA reflectance can be negative, for example, over water bodies. The 30m TOA reflectance for each reflective band are stored as signed 16-bit integers after being scaled by 10,000, in the same manner as the MODIS surface reflectance product (Vermote et al. 2002).

The radiance sensed in the Landsat low and high gain thermal bands are converted to TOA brightness temperature (i.e., assuming unit surface emissivity) using standard formula as:

$$T = \frac{K_2}{\log(K_1 / L_\lambda + 1)} \quad [2]$$

where T is the 10.40-12.50μm TOA brightness temperature (Kelvin), K_1 and K_2 are thermal calibration constants set as 666.09 (W m^{-2} sr^{-1} μm^{-1}) and 1282.71 (Kelvin) respectively (Chander et al 2009), and L_λ is the TOA spectral radiance. This equation is an inverted Planck function simplified for the ETM+ sensor considering the thermal band spectral responses.

Since February 26th 2010 the Landsat L1T data have been produced at USGS EROS with the 60m thermal bands resampled to 30m (resampling applied in the L0 to L1T USGS processing). Prior to February 26th 2010 the two 60m thermal bands were nearest neighbor resampled in the WELD processing to 30m (see Roy et al. 2010). The WELD staff compared contemporaneous USGS EROS and WELD 30m resampled thermal band data granules and found small differences but judged them to not affect the subsequent WELD compositing procedures. This is posted as a *Known Issue* on the WELD Project website. The 30m low and high gain 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.

2.3 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), and is used in Maximum NDVI compositing to preferentially select pixels with reduced cloud and atmospheric contamination (Holben 1986). The 30m TOA NDVI is computed from the TOA red and near-infrared Landsat reflectance and 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).

2.4. Band Saturation Computation

The Landsat 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 30m 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 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 ETM+ acquisition to produce a 30m 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 WELD computational overhead.

Supervised classification approaches require training data. A global database of Landsat Level 1G 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 the future Landsat Data Continuity Mission (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 30m 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 30m TOA reflective bands were used, except the shortest wavelength blue band which is highly sensitive to atmospheric scattering (Ouaidrari and Vermote 1999). For both trees, the low gain thermal band was also used. The high gain thermal band was not used because it under-saturated or over-saturated frequently over the wide surface brightness temperature range of the CONUS e.g., from hot Summer desert to Winter snow covered surfaces. 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 * 0.8 * 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 30m cloud classification results are stored as an unsigned 8-bit integer.

2.6 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 ETM+ L1T pixel. 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. As shown in Figure 3, the viewing zenith (θ) and azimuth (ϕ) for the ground pixel *A* can be determined by Equations [3] and [4], given the locations of the satellite and the ground pixel.

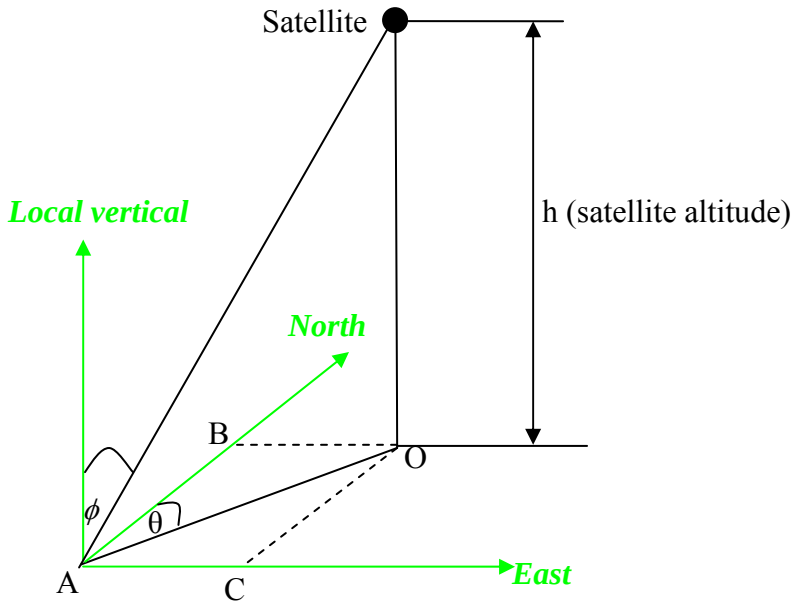


Figure 3 Landsat ETM+ viewing geometry

$$\tan(\theta) = \frac{OB}{OC} \quad [3]$$

$$\tan(\phi) = \frac{OA}{h} \quad [4]$$

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 ETM+ 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.

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.

The edge pixels of the Band 4 image 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:

$$P^{i}{}_{Scan,C} = (1 - \frac{i}{374})P_{NE,C} + \frac{i}{374}P_{SE,C} \qquad i = 0, 1, \ldots 374 \qquad [5]$$

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:

$$P^{i}{}_{Satellite} = P^{i}{}_{Scan,C} + 10.4 * 30.0 \qquad i = 0, 1, \ldots 374 \qquad [6]$$

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.

2.7 Reprojection, Resampling and Tiling

After each Landsat ETM+ L1T acquisition is processed as above, the 30m TOA reflective bands, TOA NDVI, TOA brightness temperature bands, band saturation mask, the solar and viewing geometry, and the two cloud masks are reprojected from the L1T UTM coordinates to a continental map projection. The high Landsat L1T data volume restricts provision of multiple product instances in different map projections, even though users will inevitably prefer different projections (Teillet et al. 2000). The Albers Equal Area projection was selected as it suitable for large areas that are mainly east-west

oriented such as the CONUS (Snyder 1993), and was defined with standard parallels and central meridians to provide heritage with the USGS EROS National Land Cover Database (Homer et al. 2004, Chander et al. 2009b).

It is not physically possible to store the reprojected Landsat 30m data in a single file. The largest file size achievable is limited by the amount of addressable memory on a user's personal computer, usually conservatively considered to be 32 bit computer with a maximum file size of 2GB. To ensure manageable file sizes, the 30m Landsat data are reprojected into 501 and 162 fixed Albers tiles where each tile is composed of 5000 x 5000 30m Landsat pixels. This tile pixel dimension (number of rows and columns) is smaller than the dimensions of individual L1T ETM+ acquisitions.

The Landsat ETM+ pixels are allocated to the Albers coordinate system using the inverse gridding approach, sometimes known as the indirect method (Konecny 1979). In this approach the center coordinates of each Albers 30m pixel are mapped to the nearest pixel center in the Landsat data, and the ETM+ processed data for that pixel are allocated to the Albers 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) and the MODIS Reprojection Tool (https://lpdaac.usgs.gov/lpdaac/tools/modis_reprojection_tool) is used to transform coordinates between the UTM and Albers map projections. The GCTP is computationally expensive. Consequently, a sparse triangulation methodology was used where the GCTP is invoked to project Albers 30m pixels to UTM coordinates only at the vertices of triangles, and Albers 30m 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:

$$s_1 = a_1 p_y + b_1 p_x + c_1$$
$$s_2 = a_2 p_y + b_2 p_x + c_2 \quad [7]$$

$$s_3 = 1 - s_1 - s_2$$

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 Albers coordinates the corresponding location in UTM coordinates is:

$$p_x^{'} = s_1 x_1^{'} + s_2 x_2^{'} + s_3 x_3^{'} \qquad [8]$$

$$p_y^{'} = s_1 y_1^{'} + s_2 y_2^{'} + s_3 y_3^{'}$$

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 Albers 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 450m (i.e., fifteen 30m pixels) defined from the north-west origin of the Albers 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. The maximum coordinate difference between the simplical interpolation and GCTP projected coordinates, occurs at the triangle centers, and for the CONUS is less than 1cm east-west and north-south.

2.8 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 (Holben 1986, Cihlar 1994). Thus, instead of spatially mosaicing select relatively cloud-free Landsat acquisitions together (Zobrist et

al. 1983), all the available multi-temporal acquisitions may be considered and at each gridded pixel the acquisition that satisfies some compositing criteria selected. In this way, the Global Land Survey (GLS) 2005 Landsat ETM+ data set is generated by compositing up to three circa 2005 low cloud cover acquisitions per path/row (Gutman et al. 2008). Recently, Lindquist et al. (2008) examined the suitability of the GLS data sets compared to more data intensive Landsat compositing methods (Hansen et al. 2008) and showed that over the Congo Basin compositing an increasing number of acquisitions reduced the percentage of SLC-off gaps and pixels with high likelihood of cloud, haze or shadow. Similar observations have been observed for compositing coarser resolution satellite data (Holben 1986, Cihlar 1994, Roy 2000).

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 procedures either select from colocated pixels in different orbits of geometrically registered data the pixel that best satisfies some compositing criteria or combine the different pixel values together. 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 bi-directional reflectance effects caused by angular sensing and illumination variations combined with the anisotropy of reflectance of most natural surfaces and the atmosphere (Cihlar et al. 1994, Gao et al. 2002, Roy et al. 2006). Compositing algorithms that model the bidirectional reflectance have been developed to compensate for this problem and combine all valid observations to estimate the reflectance at nadir view zenith for some consistent solar zenith angle (Roujean et al. 1992, Schaaf et al. 2002). This approach does not provide a solution for compositing thermal wavelength satellite data, and is not 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, WELD compositing is based on the selection of a "best" pixel over the compositing period.

Table 1 summaries the WELD compositing logic; each row reflects a comparison of two acquisitions of the same pixel. 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 ETM+ data are processed and regardless of the chronological processing order. For example, after 16 days the same Albers pixel location may be sensed again, and the compositing criteria are used to decide if the more recent ETM+ pixel data should be allocated to overwrite the previous data. For each composited Albers 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.

Table 1 WELD compositing criteria used to compare two acquisitions of a pixel

Priority	Compositing Criteria
1	If either fill: Select non-fill
2	If either saturated: Select unsaturated
3	If both saturated: Select the one with maximal Brightness Temperature
4	If one cloudy and one non-cloudy: Select non-cloudy
5	If one cloudy and one uncertain cloud: Select uncertain cloud if it has maximal Brightness Temperature or maximal NDVI, else select cloudy
6	If one non-cloudy and one uncertain cloud: Select non-cloud if it has maximal Brightness Temperature or maximal NDVI, else select uncertain cloud
7	If either or both "unvegetated" and both have NDVI < 0.5: Select the one with maximal Brightness Temperature
8	Select the one with maximal NDVI

The WELD compositing approach incorporates the heritage maximum NDVI and maximum brightness temperature compositing criteria as clouds and aerosols typically

depress NDVI and brightness temperature over land surfaces (Holben 1986, Cihlar et al. 1994, Roy 1997). The maximum NDVI compositing criterion is the primary compositing criterion, rather than the maximum brightness temperature criterion, because among cloud-free observations it preferentially selects vegetated observations and arguably the majority of terrestrial Landsat applications are concerned with vegetation. For certain low vegetation covers, including certain dark and bright soils, water and snow, the top of atmosphere (TOA) NDVI of a cloud can be higher than the TOA NDVI of the cloud free surface. The sensitivity of NDVI to the brightness of soil beneath vegetation canopies (Huete 1988) and to atmospheric effects (Liu and Huete, 1995) is well established. Consequently, a pixel is considered as "unvegetated" if NDVI < 0.09 AND 2.09-2.35μm TOA reflectance < 0.048. These two thresholds were derived empirically. When there are "unvegetated" pixels with NDVI < 0.5 the maximum brightness temperature compositing criterion is used, as cloud brightness temperatures tend to be lower than the cloud-free brightness temperature. The NDVI < 0.5 constraint is used to preclude the maximum brightness temperature selection of warm smoke. In these tests, the low gain 10.40-12.50μm TOA brightness temperature is used as it has a wider range than the high gain TOA brightness temperature and does not saturate (Chander et al. 2009). For compositing purposes only, a pixel is considered saturated if either the red or near-infrared TOA reflectance is saturated (as the NDVI is unreliable when one or both of these bands are saturated).

The cloud masks are used to complement the maximum NDVI and maximum brightness temperature 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.

2.9 Browse Generation

Browse images with reduced spatial resolution are generated from the weekly, monthly, seasonal and annual composited mosaics to enable synoptic product quality assessment with reduced data volume (Roy et al. 2002), and to provide browse imagery for the What You See Is What You Get (WYSIWYG) WELD Internet distribution system (Section 4.3.2).

CONUS and Alaska browse images are generated in the JPEG format with fixed contrast stretching and color look-up tables to enable consistent temporal comparison. The browse images are defined at different levels of generalization (Boschetti et al. 2008) using the median pixel values falling in a given window size defined with dimensions set as an integer multiple of the 30m Albers pixels. True color multispectral browses are generated from the TOA red (0.63-0.69μm), green (0.53-0.61μm) and blue (0.45-0.52μm) reflectance. Within a window, a pixel with the median red reflectance is located and then the blue and green reflectance values for that pixel selected. In this way, the reflectance for the same pixel is obtained which produces more coherent browse imagery than selecting the median reflectance values for each wavelength independently. The red reflectance is used as the "master" since it is less sensitive to atmospheric contamination than shorter wavelength blue and green reflectance (Ouaidrari and Vermote, 1999).

3.0 VERSION 1.5 PRODUCT DOCUMENTATION

3.1 WELD Product Types

The WELD products are available for the CONUS and Alaska as weekly, monthly, seasonal and annual composited products. The monthly, seasonal and annual products are defined in a temporally nested manner following climate modeling conventions where Winter is defined by the months December, January and February. The weekly products are defined more simply with respect to each calendar year (Table 2).

Table 2 WELD product types

Product Type	**Temporal Definition**
Annual	The preceding year's December through the current year's November.
Seasonal:	
Winter	December, January, February
Spring	March, April, May
Summer	June, July, August
Autumn	September, October, November
Monthly	The days in each calendar month
Weekly	Consecutive 7-day products with *Week01*: January 1 to January 7, *Week02*: January 8 to January 14, … , *Week 52*: December 24 to December 30 (non-leap years) or December 23 to December 29 (leap years), *Week 53*: December 30th to December 31st (leap years) or December 31st (non-leap years).

Figures 4 to 7 show example CONUS true color, red (0.63-0.69µm), green (0.53-0.61µm) and blue (0.45-0.52µm), TOA reflectance browse images for the weekly, monthly, seasonal and annual composites respectively. All the L1T 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.

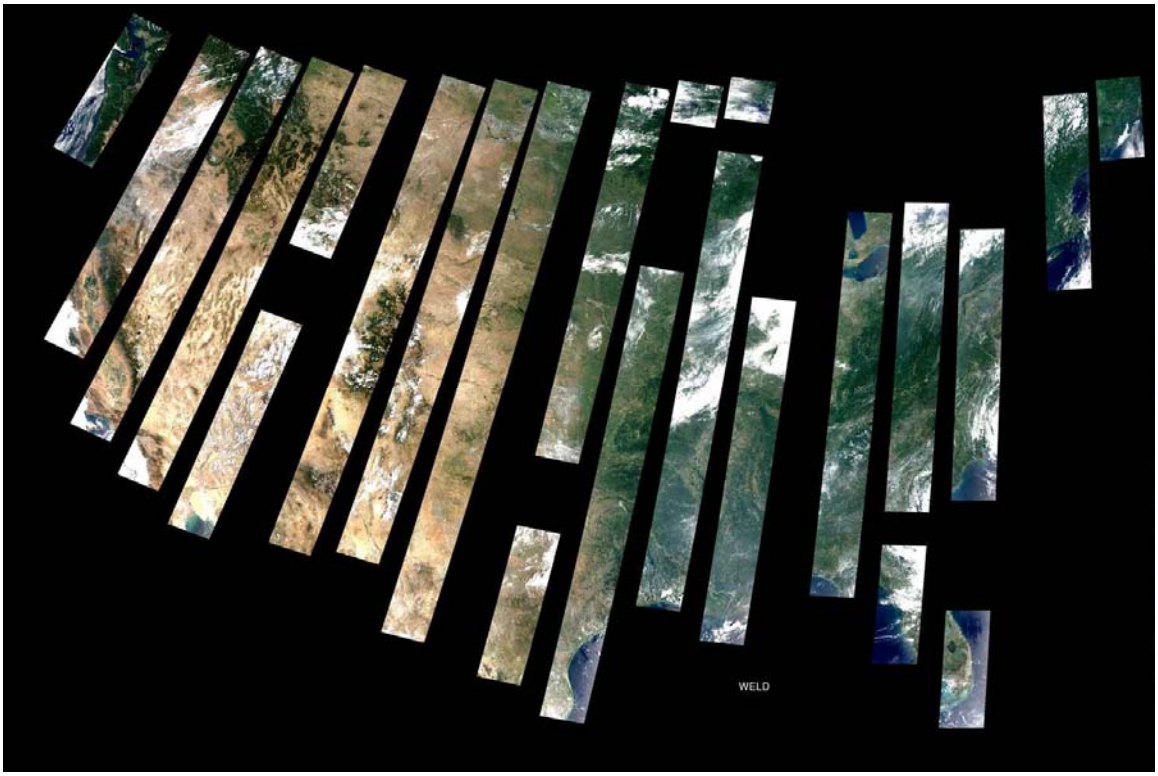


Figure 4 Example weekly WELD CONUS composite (July 15 - 21, 2008)

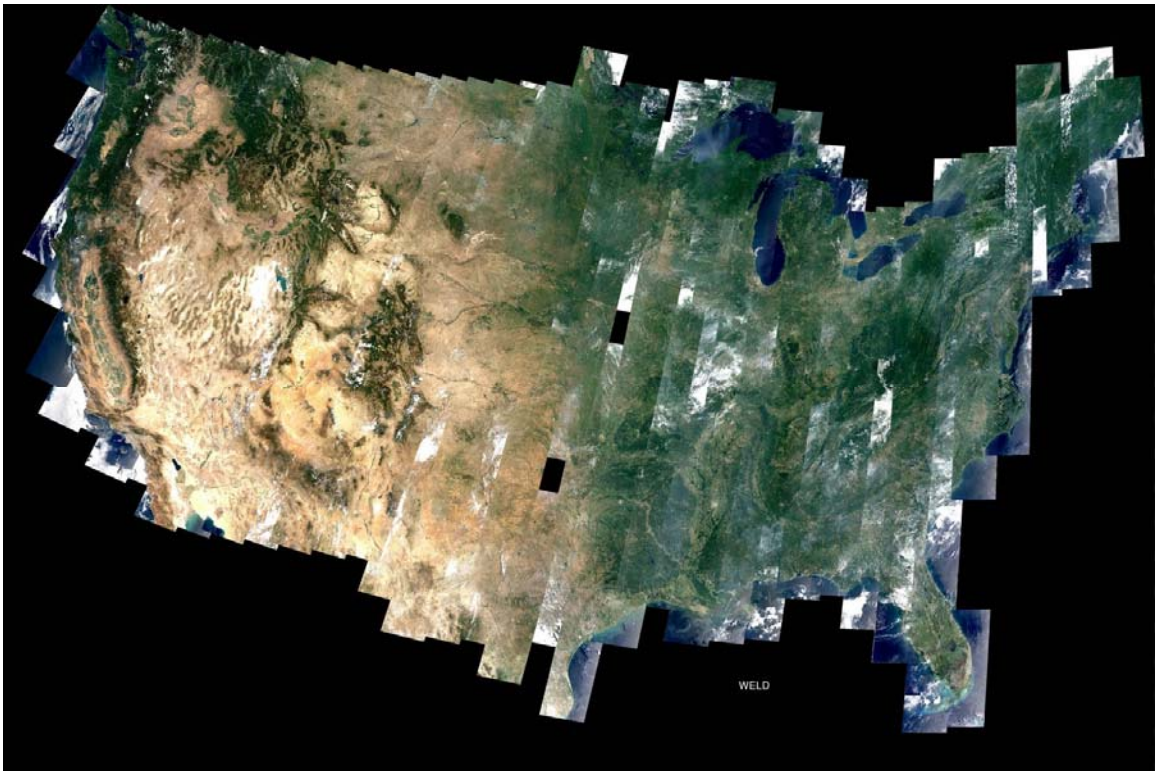


Figure 5 Example monthly WELD CONUS composite (July 2008)

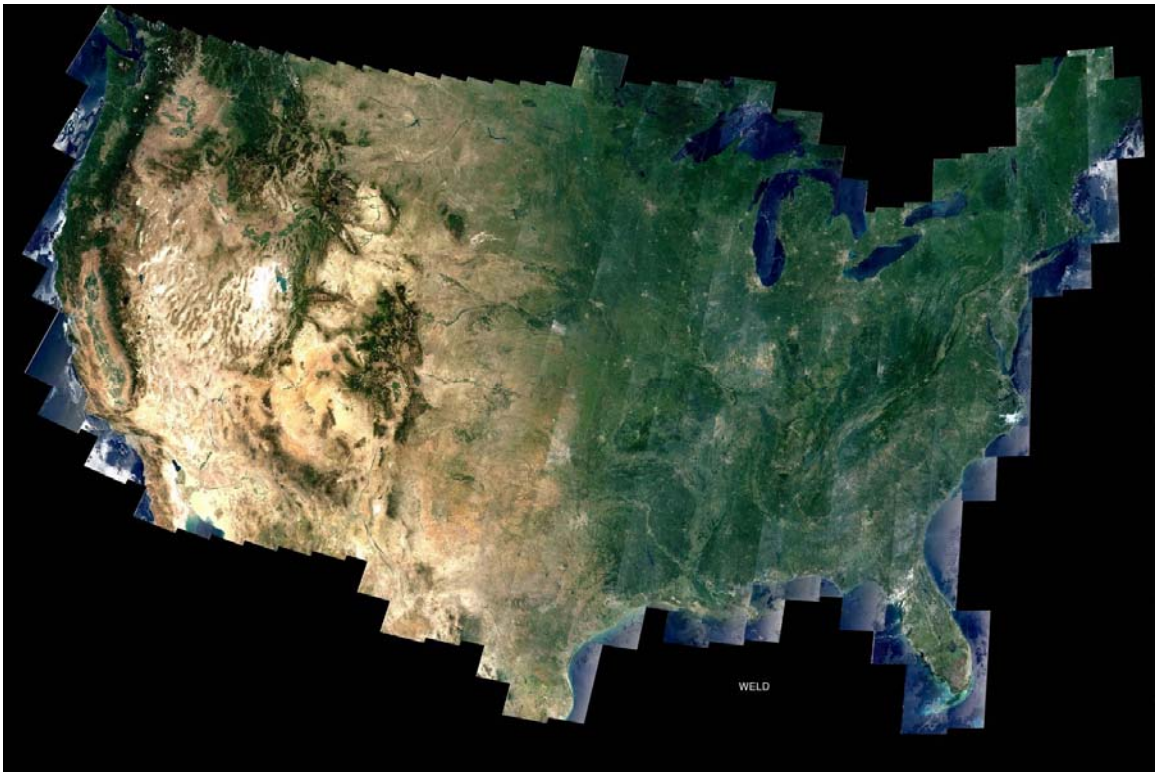


Figure 6 Example seasonal WELD CONUS composite (Summer 2008)

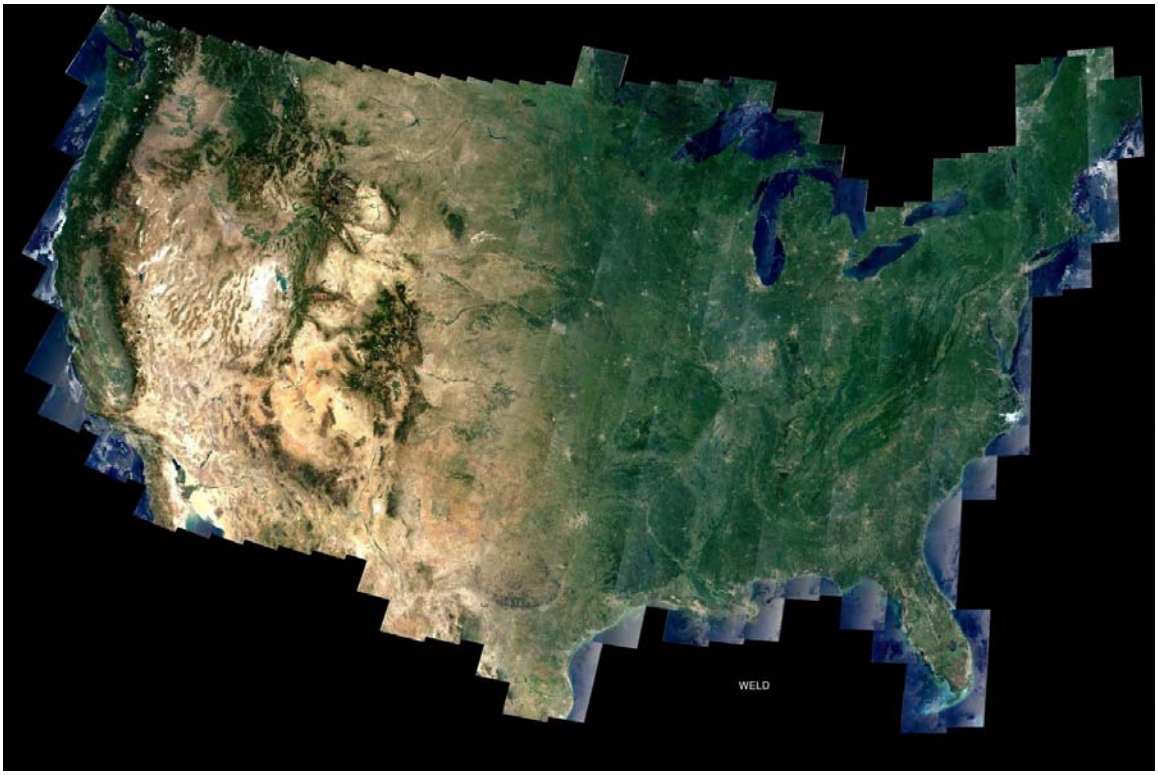


Figure 7 Example annual WELD CONUS composite (2008)

3.2 WELD Product Contents

Each WELD product 30m pixel has 14 bands stored with appropriate data types to minimize the file size.

Table 3 WELD Product Contents and Storage Attributes

Band Name	Data Type	Valid Range	Scale factor	Units	Fill Value	Notes
Band1_TOA_REF	int16	-32767 -- 32767	0.0001	unitless	-32768	Top of atmosphere (TOA) reflectance and brightness temperature (BT) are computed using standard formulae and calibration coefficients associated with each ETM+ acquisition. Band 6 brightness temperature data are resampled to 30 m. The conventional ETM+ band numbering scheme is used.
Band2_TOA_REF	int16	-32767 -- 32767	0.0001	unitless	-32768	
Band3_TOA_REF	int16	-32767 -- 32767	0.0001	unitless	-32768	
Band4_TOA_REF	int16	-32767 -- 32767	0.0001	unitless	-32768	
Band5_TOA_REF	int16	-32767 -- 32767	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_TOA_REF	int16	-32767 -- 32767	0.0001	unitless	-32768	
NDVI_TOA	int16	-10000 -- 10000	0.0001	unitless	-32768	Normalized Difference Vegetation Index (NDVI) value generated from Band3_TOA_REF and Band4_TOA_REF.
Day_Of_Year	int16	1 -- 366	1	Day	0	Day of year the selected ETM+ 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	uint8	0 -- 255	1	unitless	None	Number of ETM+ observations considered over the compositing period.

3.3 WELD Product Map Projections

The Albers Equal Area projection was selected as it suitable for large areas that are mainly east-west oriented such as the CONUS (Snyder 1993), and was defined with standard parallels and central meridians to provide heritage with the USGS EROS National Land Cover Database (Homer et al. 2004, Chander et al. 2009b). The map projection parameters are summarized in Table 4. The latitude of the CONUS and Alaskan projection origins fall outside the WELD product regions so that the Albers Northing value is always positive.

Table 4 WELD Product Projection Parameters

Projection: Albers Equal Area		
Datum: World Geodetic System 84 (WGS84)		
	CONUS	**Alaska**
First standard parallel	29.5˚	55.0˚
Second standard parallel	45.5˚	65.0˚
Longitude of central meridian	-96.0˚	-154.0˚
Latitude of projection origin	23.0˚	50.0˚
False Easting	0.0	0.0
False Northing	0.0	0.0

3.4 WELD Product Data Formats

The WELD products are processed in Hierarchical Data Format (HDF). HDF is 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. For example, it is used to store the standard MODIS Land products (Justice et al. 2002).

The WELD products are generated in HDF4 in separate 5000 x 5000 30m pixel tiles defined in the Albers Equal Area projection. Each tile has 14 bands storing the information described in Table 3 with the band (HDF science data set) specific attributes (fill value, scale factor, units, valid range). The WELD Version 1.5 HDF products contain only default HDF metadata. The planned Version 2.0 HDF products will also store product specific metadata mandated for long term WELD product archiving.

The HDF product filename convention is described in Table 5 and is designed to be descriptive, simple, and amenable to scripting.

Table 5 WELD HDF Product Filename Convention

Convention: **<Region>. <Period> . <Year> .h<xx>v<yy>.doy<min DOY>to<max DOY>.v<Version Number>.hdf**		
	Valid Range	**Notes**
<Region>	CONUS / Alaska	
<Period>	Annual, spring/summer/autumn/winter, month01/month02/, …,/month12, week01/week02/, …, /week52/week53	See Table 2
<Year>	2005, 2006, 2007, ..., 2012	
<xx>	00, 01, ..., 32 (CONUS) or 00, 01, ..., 16 (Alaska)	horizontal WELD tile coordinate.
<yy>	00, 01, ..., 21 (CONUS) or 00, 01, ..., 13 (Alaska)	vertical WELD tile coordinate.
<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>	1.1, 1.3, …, 2.0, 2.1, 2.2, …	Major and minor algorithm version changes reflected in the first and second digits respectively.

There are a total of 501 CONUS and 162 Alaskan tiles referenced using a two digit horizontal and vertical tile coordinate system (Figures 8 and 9) that is reflected in the HDF product filename.

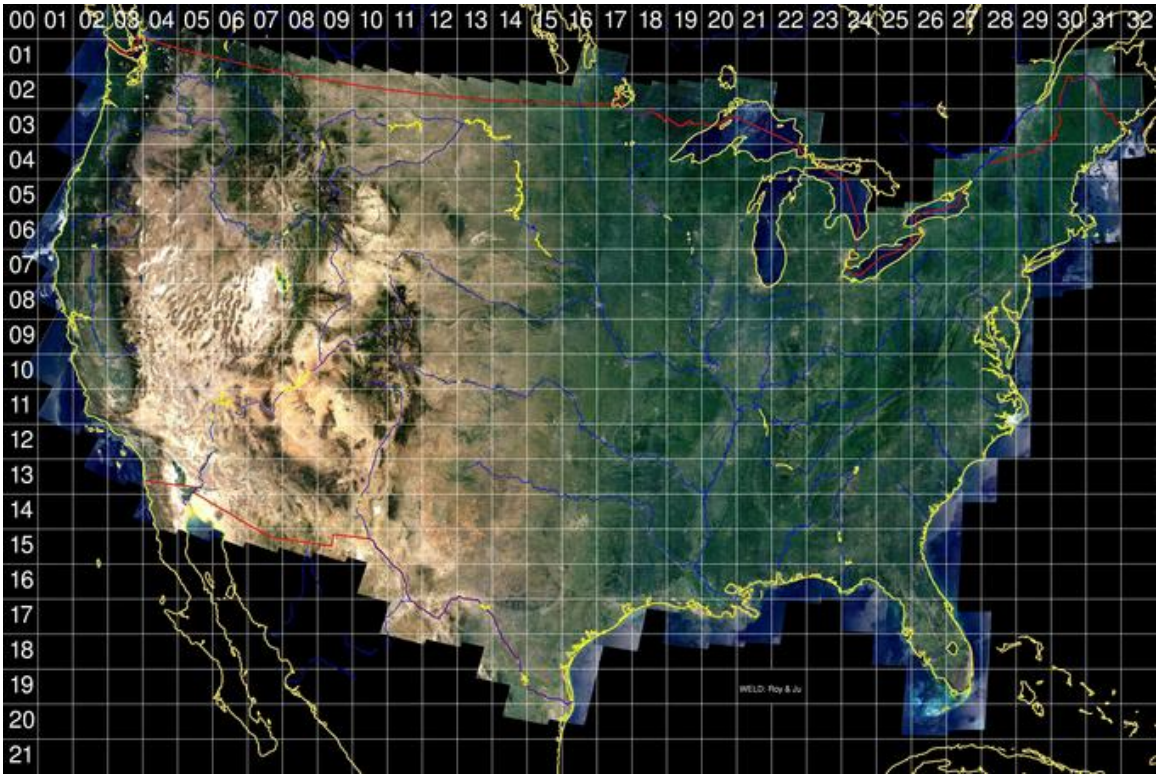


Figure 8 CONUS HDF horizontal and vertical Albers tile coordinate scheme

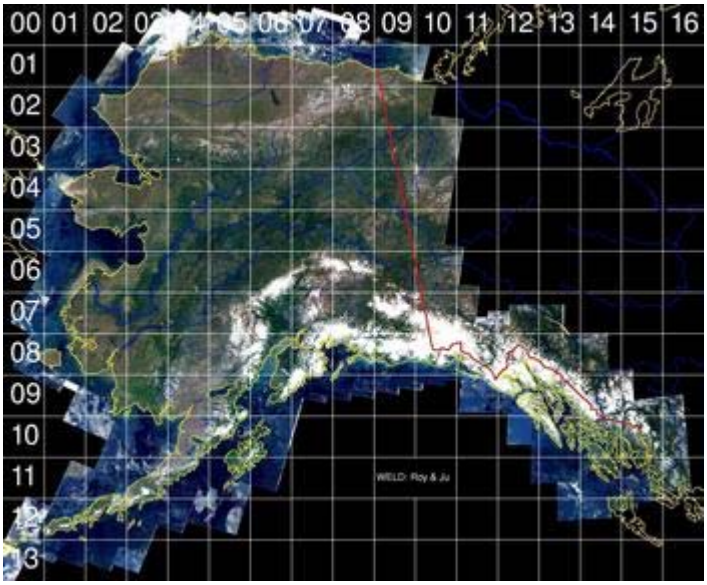


Figure 9 Alaska HDF horizontal and vertical Albers tile coordinate scheme

3.5 WELD Product Data Volume

The HDF format tiles are stored with HDF internal compression on and are typically 200MB. Table 6 summarizes the average tile data volume for the different WELD product types computed from the 2008 V1.5 products. The tabulated volumes vary because the amount of unobserved land data varies in space and time. For example, there are gaps in the weekly composited products imposed by the Landsat orbital geometry (Figure 4) and there are gaps in the Alaskan Winter products because there are fewer day time observations at high latitude.

Table 6 Typical WELD Product HDF Tile File Sizes

Product Type	**CONUS**	**Alaska**
Annual	226 MB	180 MB
Winter	225 MB	112 MB
Spring	220 MB	181 MB
Summer	219 MB	174 MB
Autumn	213 MB	162 MB
Weekly	95 MB	94 MB

The total annual WELD product volume is not the product of the volumes tabulated in Table 6 and the number of CONUS and Alaska tiles because for some periods tiles are not generated if they are all Fill values.

The total WELD product volume, for all the product types for both CONUS and Alaksa, is on average 4TB per year.

4.0 PRODUCT MANAGEMENT AND DISTRIBUTION

4.1 Data Management Plan

The WELD products are generated in Hierarchical Data Format (HDF) on the WELD project computers at the South Dakota State University (SDSU) Geographic Information Science Center of Excellence (GIScE). Periodically, after the WELD products have been generated, and after consultation with United States Geological Survey (USGS) Center for Earth Resources Observation and Science (EROS) engineers, the products are copied via secure file transfer protocol (rsync) to the USGS EROS for distribution.

Table 7 summarizes the Version 1.5 WELD processing steps. These steps are broadly split into processing performed on individual L1T acquisitions that are each defined in the UTM coordinate system (referred as UTM processing), processing on individual WELD tiles (referred to as TILE processing), and processing on multiple WELD tiles to produce continental reduced spatial resolution true color browse HDF and JPG imagery.

Table 7 WELD Version 1.5 processing steps overview

Steps in UTM processing o View and Solar Geometry o Digital Number to Calibrated Radiance o TOA reflectance & brightness temperature & band saturation & NDVI o Cloud masking saturated and unsaturated (ACCA and Decision Tree)
Steps in TILE processing o Albers to UTM projection o Temporal compositing (weekly, monthly, seasonal, annual)
CONUS and Alaska Browse Generation

The WELD processing is sequenced by the availability of L1T ETM+ data. A WELD project file transfer protocol (ftp) script is run weekly at SDSU using a cron job to retrieve all the CONUS and Alaskan Landsat ETM+ data as they are generated by the USGS EROS Level-1 Product Generation System. A code that ranks the sequence of

new Landsat L1T data available at the EROS site is used so that L1T acquisitions are ftp'd in order of the WELD tile they fall within to facilitate computationally efficient processing.

The WELD production system is being developed following a spiral development approach, largely implemented in a modular fashion in the C computer programming language, and running on the Linux Operating System. The processing modules are integrated via scripts in a manner designed to maximize CPU and memory resources and to reduce the number of disk read/write operations. Currently the scripts are run manually but the possibility of using an open source resource manager providing control over batch jobs and distributed compute nodes is being investigated to automate the production.

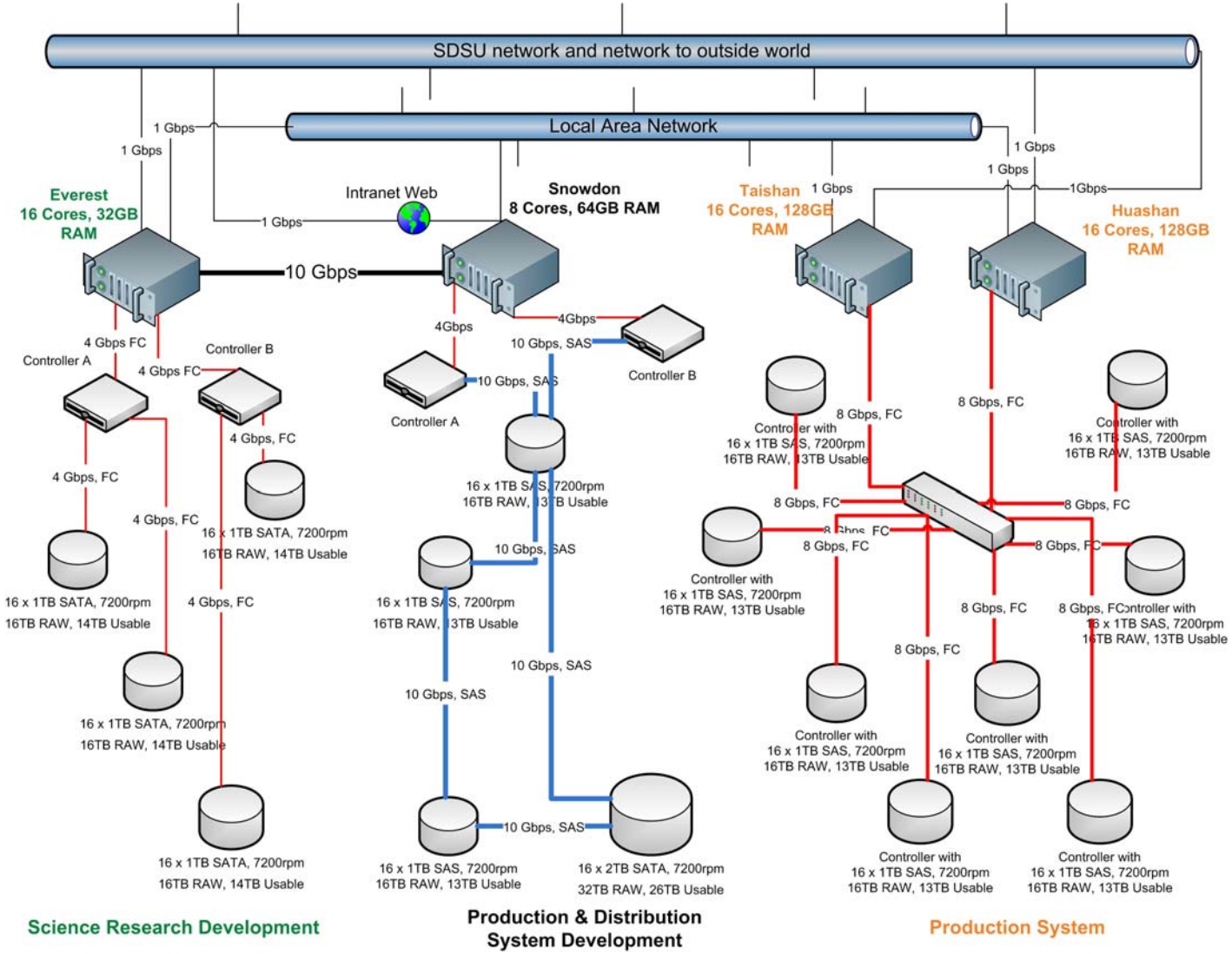
Care is taken to use strict algorithm, product filename and documentation versioning control. All the code modules produce exit codes and log files that are amenable to scripting and provide diagnostic resources for graceful code failures. The code is documented following standard function, input and output description conventions.

A WELD product versioning scheme is used to reflect product reprocessing using improved algorithms, ancillary data, sensor knowledge and input data. Major and minor algorithm version changes are reflected in the first and second digits respectively of the version number. The currently available WELD products are Version 1.5. There are insufficient resources to distribute more than one WELD product version. Users are encouraged to use the latest WELD product version.

4.2 Production Hardware

Figure 10 illustrates the WELD development and production resources. The production system is based on two Intel Xeon-based compute servers attached through a high performance storage area network switch connected via 8Gbps fiber channel to online RAID storage devices consisting of Serial Attached SCSI disks. This architecture offers high performance at low cost of ownership.

Figure 10 WELD Production Hardware



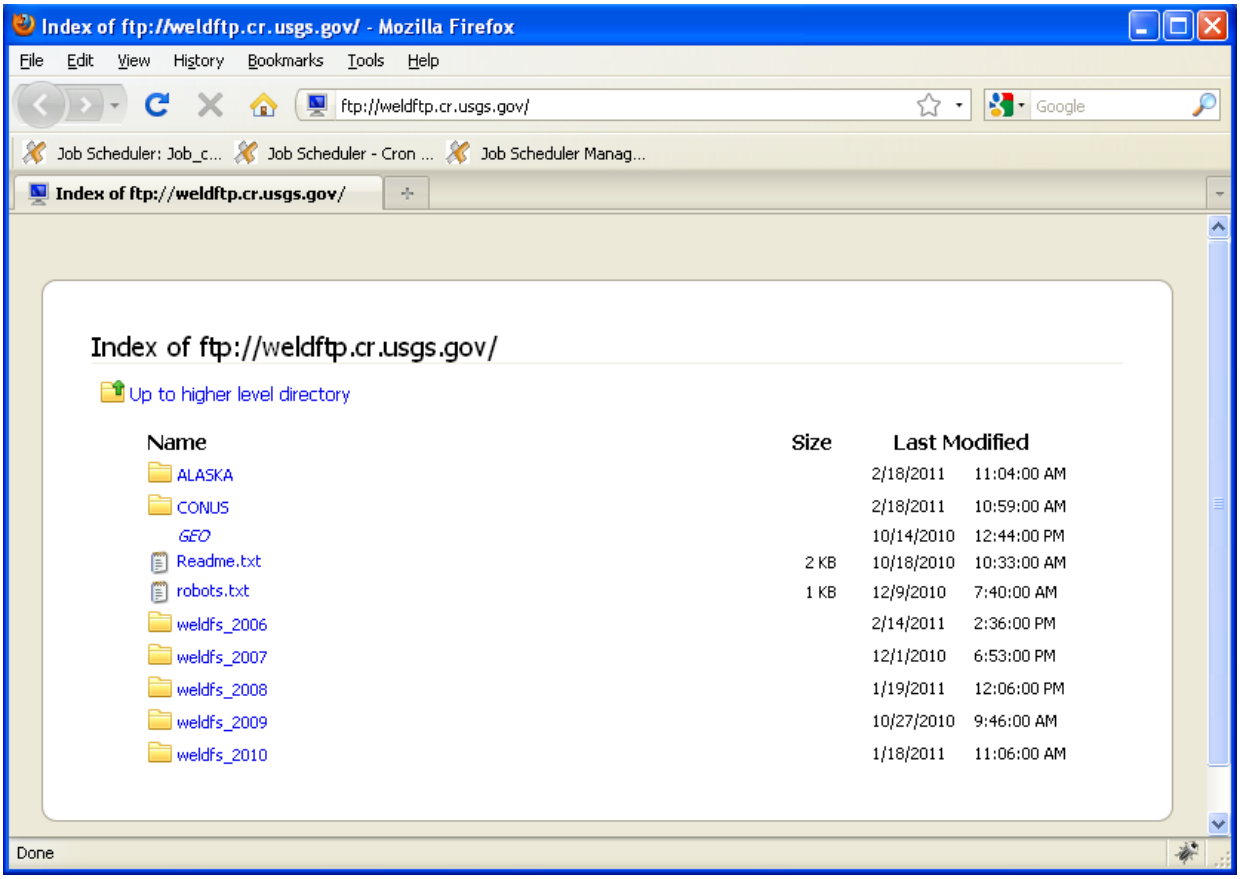
4.3 WELD Product Distribution

The WELD products are made freely available over the Internet from the USGS EROS in both HDF and GeoTIFF formats. In addition, select WELD true color browse images have been made Open Geospatial Consortium (OGC) compliant and are served from the Oak Ridge National Laboratory Distributed Active Archive Center.

4.3.1 WELD Product FTP Distribution

The HDF tiled products are available via anonymous FTP at ftp://weldftp.cr.usgs.gov/ . Currently the Version 1.5 weekly, monthly, seasonal and annual WELD products for the CONUS and Alaska are available for 2006, 2007, 2008, 2009 and 2010. This provides a total of about 20TB of data with HDF internal compression on.

Figure 11 Screen shot of the top page of the WELD FTP distribution



4.3.2 What You See Is What You Get (WYSIWYG) WELD Internet Distribution

In response to user requests concerning improvements to the original Version 1.0 WELD HDF product distribution, an intuitive what you see is what you get (WYSIWYG) WELD product Internet distribution interface was developed at SDSU using open source OpenLayers, MapServer and MySQL software.

Users of the WYSIWYG system require a web browser with JavaScript enabled. The system allows users to interactively order any rectangular spatial subset of any WELD product, up to 2GB, in a way that the WELD tile structure is transparent to the user. The interface design follows an easy-to-use and intuitive design philosophy and provides a user experience similar to commercial software such as Google Maps and the iPhone interface.

Users are able to interactively select and view any of the CONUS or Alsaka products, and pan and zoom (to a spatial resolution of 210m) within user selected product browse imagery. Users may order an arbitrary rectangular geographic area of interest, either interactively by moving a rubber band box over the displayed browse image, or by specifying geographic coordinates in a text field. Once a region of interest has been selected, the appropriate WELD tiles are assembled, subset and mosaiced, and placed on an HTTP site in GeoTIFF format. The user is sent an email with the relevant HTTP access information. The WYSIWYG distribution interface was also developed to harvest user information and product distribution metrics.

An early version of the WYSIWYG was demonstrated at a one day Landsat User meeting (Landsat User Workshop, Monday, September 21, 2009, USGS EROS, Sioux Falls) attended by Landsat users who were asked to present and discuss their experiences with Landsat data and make recommendations for improved Landsat distribution and processing. The WYSIWYG system has been subsequently improved, whilst avoiding the temptation to overbuild its functionality.

Figure 12 Screen shot of the top web page of the What You See Is What You Get (WYSIWYG) WELD Internet distribution interface at http://weld.cr.usgs.gov showing the 2006-2010 CONUS and Alaska annual products.

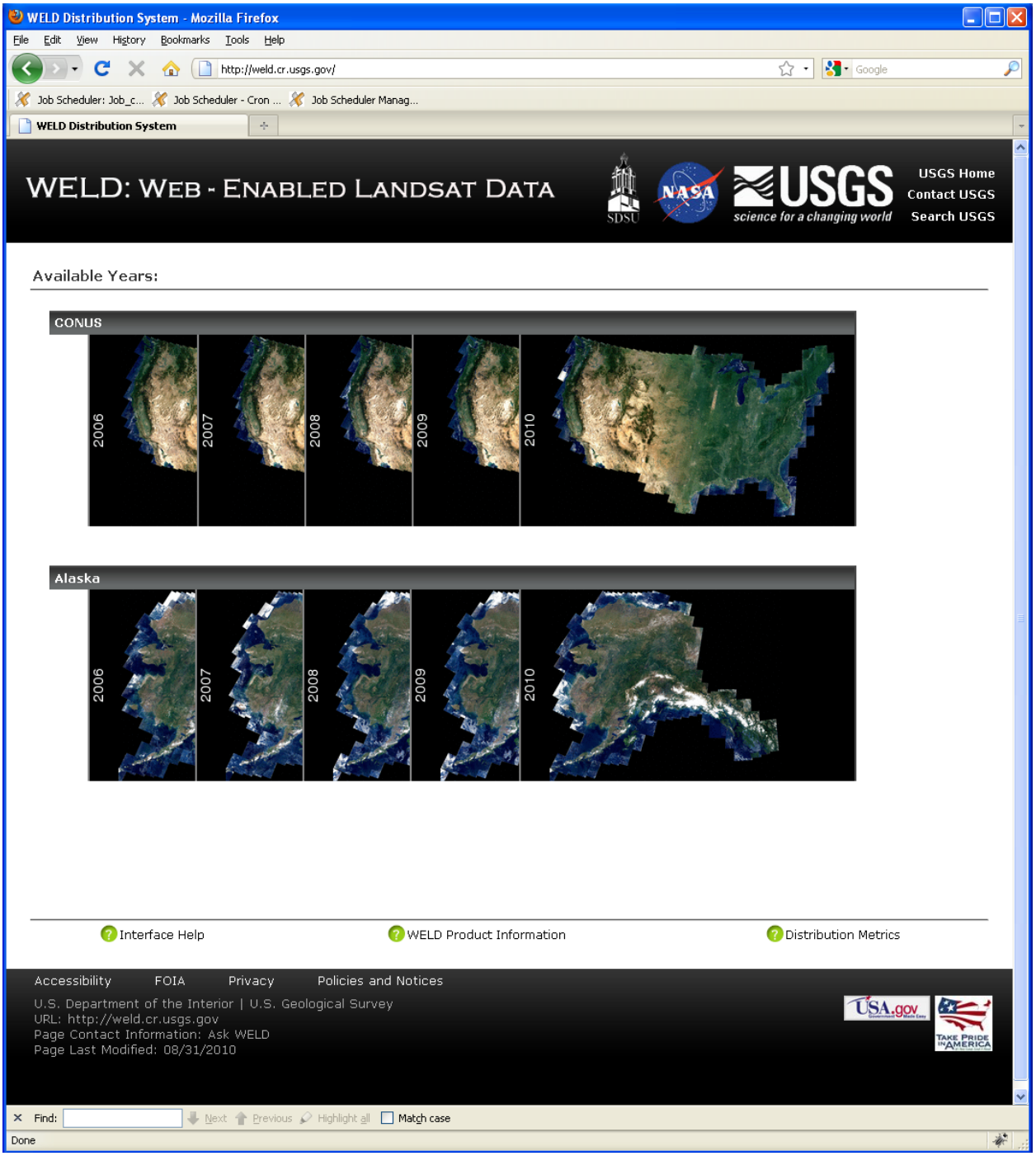


Figure 13 Screen shot of the WELD What You See Is What You Get (WYSIWYG) WELD Internet distribution interface, showing the 2008, annual and 4 seasonal (top row), 12 monthly (middle row) and 53 weekly (bottom row) CONUS products.

User ordered WYSIWYG WELD product subsets are defined in the Albers Equal Area projection and follow the filename convention described in Table 8. Due to the filename length constraint of the Microsoft Windows operating system, each of the 14 WELD bands are saved as separate GeoTIFF files in a sub-directory. The sub-directory name conforms to the HDF filename convention (Table 5) and also includes the bounding longitude and latitude of the ordered data area so that users can quickly locate and identify different WELD product orders.

Table 8 WELD GeoTIFF Product Filename Convention generated by the WELD What You See Is What You Get (WYSIWYG) WELD Internet distribution interface.

Convention: ***sub-directory name:* <Region>. <Period> . <Year>.lon<*min lon*>to<*max lon*> .lat<*min lat*>to<*max lat*>. doy<*min DOY*>to<*max DOY*>.*v*<*Version Number*>** ***band name:* <band name>.TIF**		
	Valid Range	**Notes**
<Region>	CONUS / Alaska	
<Period>	annual, spring/summer/autumn/winter, month01/month02/, …,/month12, week01/week02/, …, /week52/week53	See Table 2
<Year>	2005, 2006, 2007,..., 2012	
<min lon>	-127.000000 to -65.000000 (CONUS) -175.000000 to -125.000000 (Alaska)	minimum pixel center longitude of the ordered data area; specified to 6 decimal places.
<max lon>	-127.000000 to -65.000000 (CONUS) -175.000000 to -125.000000 (Alaska)	maximum pixel center longitude of the ordered data area; specified to 6 decimal places.
<min lat>	23.000000 to 52.000000 (CONUS) 50.000000 to 72.000000 (Alaska)	minimum pixel center latitude of the ordered data area; specified to 6 decimal places.
<max lat>	23.000000 to 52.000000 (CONUS) 50.000000 to 72.000000 (Alaska)	maximum pixel center latitude of the ordered data area; specified to 6 decimal places.
<min DOY>	001, 002, …, 366	minimum non-fill Day_Of_Year pixel value in the data.
<max DOY>	001, 002, …, 366	maximum non-fill Day_Of_Year pixel value in the data.
<Version Number>	1.1, 1.3, …, 2.0, 2.1, 2.2, …	Major and minor algorithm version changes reflected in the first and second digits respectively.
<band name>	Band1_TOA_REF, Band2_TOA_REF,…, Num_Of_Obs	See Table 3

4.3.3 WELD Open Geospatial Consortium Browse Imagery Distribution

Select WELD true color browse images have been made Open Geospatial Consortium (OGC) compliant and are served from the Oak Ridge National Laboratory Distributed Active Archive Center (http://webmap.ornl.gov/wcsdown/dataset.jsp?ds_id=111112).

The OGC compliance enables services to be placed against the WELD browse imagery and the data rendered into different applications over the internet from the Oak Ridge National Laboratory Distributed Active Archive Center.

Figure 14 Screen shot of a Google Earth rendering of the OGC compliant 2009 annual WELD true color browse image.

4.4 WELD Product Distribution Metrics

The WYSIWYG distribution interface was developed to harvest user information and product distribution metrics. First-time users attempting to order WELD products via the WYSIWYG interface are asked to register (by entering their email and user generated password) and to provide the information summarized in Table 9.

Table 9 WELD What You See Is What You Get (WYSIWYG) WELD Internet Distribution Interface User Information

User Country	246 countries [copied from the USGS EROS GLOVIS distribution system country list]
User Affiliation	• Educational/Academic Research institution • Non Governmental institution (NGO) • Commercial • General Public • Government institution not US • US Federal Government - Executive Branch • US Federal Government - Legislative Branch • US Federal Government - Judicial Branch • USGS Business Partner [adapted from the USGS EROS GLOVIS distribution system]
Intended WELD Product Primary and Secondary Uses	Agriculture, Climate Change, Cryosphere, Ecosystem Studies, Education, Emergency Response, Energy, Fire, International Land Issues, Geology, Human Ecology, Human Health, Insurance, Forestry, Land Change, National Security, Natural Resources, Planning, Socioeconomics, Water, Telecommunications, Terrestrial Monitoring, Visualization, Other [copied from the USGS EROS GLOVIS distribution system list]]

Distribution metrics for the WYSIWYG WELD Internet distribution interface are available online at http://weld.cr.usgs.gov/WYSIWYG/request_metrics.php .

The WYSIWYG was ported to USGS EROS in October 2010 and is actively distributing WELD products (http://weld.cr.usgs.gov). At the time of writing more than 155 users, from 11 countries, predominantly from Educational/Academic Research institutions, with a diversity of primary uses (the majority of uses are Land Change and Education), have placed more than 10,000 orders.

At the time of writing no tracking of the WELD FTP site distributions statistics has been undertaken by USGS EROS engineers.

4.5 WELD Product Long Term Archive Strategy

At the end of the 5 year grant funding period, a long term archive strategy for the most recent version of the WELD products will be negotiated with the USGS/EROS DAAC and any other agency suggested by the NASA program management.

5.0 PLANNED VERSION 2.0 WELD PRODUCTS

All years from 2005 to 2012 will be reprocessed as improved versions of the WELD algorithms are developed. In general it is preferable to use the latest WELD product version which reflects improvements to the WELD processing algorithms and input data. The next major WELD reprocessing will be Version 2.0 and will have the following elements:

- atmospheric correction of the top of atmosphere reflectance bands
- radiometric normalization of the reflectance to nadir view and fixed solar zenith angle
- gap filling of the Landsat ETM+ SLC-off and cloud gaps in the reflectance and thermal bands
- 30m annual percent tree, bare ground, vegetation and water classification

The algorithms for these improvements are described only briefly below, as they are currently submitted or in preparation for publication in peer reviewed journals.

5.1 Atmospheric Correction of the Top of Atmosphere 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. Consistent Landsat surface reflectance data are needed in support of high to moderate spatial resolution geophysical and biophysical studies. Two candidate atmospheric correction methods are being considered: a new MODIS-based method and the established Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) method. Both atmospheric correction methods use the 6SV radiative transfer code which has an accuracy better than 1% over a range of atmospheric stressing conditions (Kotchenova et al. 2006).

The MODIS-based method uses the atmospheric characterization data used in the generation of the standard MODIS Terra land surface reflectance product suite (Vermote et al. 2002) to correct the Landsat ETM+ data sensed in the same MODIS Terra orbit. The MODIS Terra aerosol optical thickness and aerosol type (dust, polluted urban, clear

urban, high absorption smoke, low absorption smoke), derived using an approach based on the Kaufman et al. (1997) dense dark vegetation (DDV) methodology, and MODIS derived water vapor, in conjunction with daily ozone derived from NASA's Earth Probe Total Ozone Mapping Spectrometer (EP TOMS) daily data and surface atmospheric pressure from the National Centers for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR) Reanalysis 6-hourly data are used (Vermote and Kotchenova, 2008).

The LEDAPS method (Masek et al. 2006) derives the aerosol optical thickness independently from each Landsat acquisition using the Kaufman et al. (1997) DDV approach and assuming a fixed continental aerosol type. The LEDAPS method also uses the NCEP/NCAR 6-hourly Reanalysis water vapor data, and like the MODIS-based method, uses the NASA's EP TOMS ozone data and surface atmospheric pressure from NCEP/NCAR 6-hourly Reanalysis data.

The MODIS instrument has superior spectral and radiometric characteristics and senses a much larger swath compared to the Landsat ETM+ and so should provide more reliable atmospheric characterization than the LEDAPS approach. However, the MODIS atmospheric characterization describes the atmosphere approximately 27 minutes after the Landsat ETM+ overpass and so dynamic aerosols may be better defined from the ETM+ acquisition itself under the LEDAPS approach provided that DDV targets are available in the ETM+ acquisition.

5.2 Reflective Wavelength Gap Filling

A semi-physical approach for Landsat cloud and Scan Line Corrector (SLC)-off gap filling, and also absolute radiometric normalization, that uses the MODIS 500m BRDF/Albedo product to describe the surface BRDF modulated by sub-pixel variations at the 30m ETM+ pixel scale has been developed (Roy et al. 2008) and is implemented as:

$$\hat{\rho}_{ETM+,t2}\left(\lambda_{ETM+},\Omega_{new},\Omega'_{new}\right) = c \times \rho_{ETM+,t1}\left(\lambda_{ETM+},\Omega_{observed},\Omega'_{observed}\right) \quad [9]$$

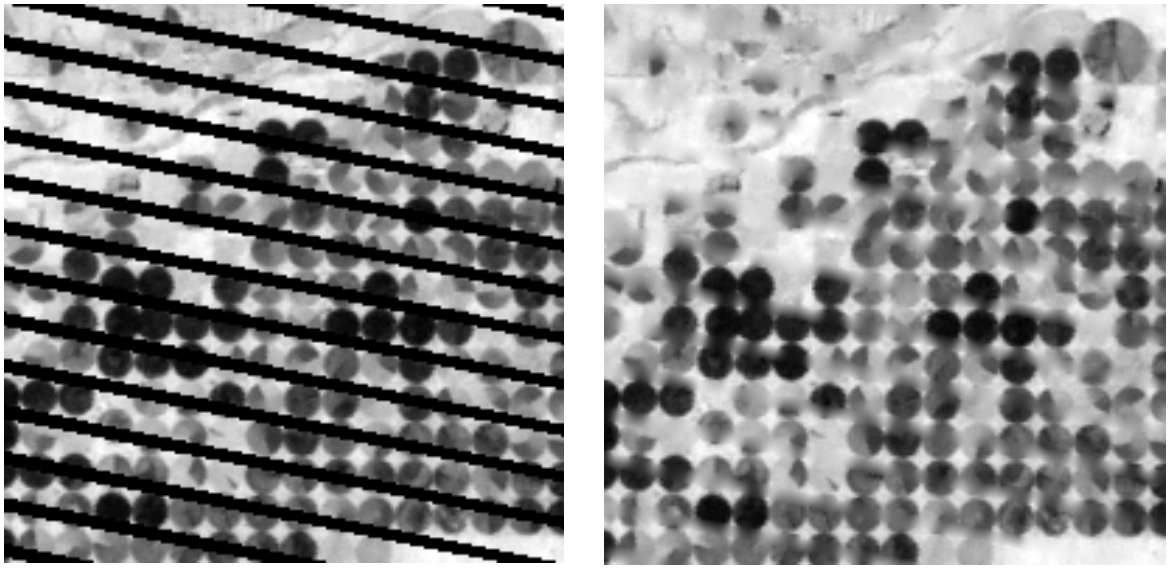
$$c = \frac{\hat{\rho}_{MODIS,t2}\left(\lambda_{MODIS},\Omega_{new},\Omega'_{new}\right)}{\hat{\rho}_{MODIS,t1}\left(\lambda_{MODIS},\Omega_{observed},\Omega'_{observed}\right)}$$

where $\hat{\rho}_{ETM+,t2}\left(\lambda_{ETM+},\Omega_{new},\Omega'_{new}\right)$ is the modeled Landsat reflectance for ETM+ wavelength λ_{ETM+} for any desired viewing and solar illumination vectors $\Omega_{new},\Omega'_{new}$ at time t_2, $\rho_{ETM+,t1}\left(\lambda_{ETM+},\Omega_{observed},\Omega'_{observed}\right)$ is the reflectance of a Landsat observation of the pixel sensed at time t_1 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 the MODIS 500m BRDF/Albedo product (Schaaf et al. 2002). Particular advantages of the method are: it does not require any tuning parameters and so may be automated; it is applied on a per-pixel basis and is unaffected by the presence of missing or contaminated neighboring Landsat pixels; it uses a band ratio and so is largely insensitive to spectral band pass differences between the Landsat and MODIS bands; it allows for future improvements through BRDF model refinement and error assessment.

5.3. Thermal Wavelength Gap Filling

The above approach cannot be applied to the Landsat thermal bands (Band61 and Band62) because the physics of emitted wavelength radiation is different to the physics at reflective wavelengths. Gap filling missing thermal band pixel data is further complicated because emitted radiation changes very rapidly in space and time. Consequently local spatial gap filling methods are being investigated: Geostatistical interpolants (kriging etc.) are computationally expensive; spline based interpolants fit to a large surrounding sample data area and the interpolated values may be outside the range of the sample data; inverse distance weighting interpolants are computationally inexpensive but perform poorly for irregular sample data distributions; natural neighbor interpolation has elegant properties (no tuning parameters, interpolated values are guaranteed to be within the range of the samples used, pass through the input samples and are smooth everywhere except at locations of the input samples). A natural neighbor interpolation code was developed to provide a computationally efficient interpolation (Figure 15).

Figure 15 Top of Atmosphere Band 61 Brightness Temperature, Kansas Pivot Irrigation 200 x 200 60m pixel detail, June 11th 2008. This example shows arguably the worst case gap filling scenario. **Left** SLC-off gaps (dark stripes). **Right:** Natural Neighbor Interpolation Gap Filled version.



5.4 Reflective Wavelength Radiometric Normalization

The radiometric consistency of reflective wavelength Landsat data may change spatially and temporally, due to atmospheric variations, sensor calibration changes, cloud and shadow contamination, and differences in illumination and observation angles. The Version 1.5 WELD processing (conversion to top of atmosphere reflectance, cloud screening, and compositing) will largely remove all of these variations in the monthly, seasonal and annual composites, except for reflectance differences due to illumination and observation angles.

This reflective wavelength gap-filling approach (Equation 9) allows for radiometric normalization of the Landsat reflective wavelength observations, by setting $\hat{\rho}_{ETM+,t2}(\lambda_{ETM+},\Omega_{new},\Omega'_{new})$ as $\rho_{ETM+,t1}(\lambda_{ETM+},\Omega_{new},\Omega'_{new})$ with $\Omega_{new},\Omega'_{new}$ set to nadir viewing and local solar noon for the day that the Landsat pixel was sensed.

Research efforts on reflective wavelength normalization and gap-filling have been focused on refining the Roy et al. (2008) method by using (a) the MODIS BRDF parameters from the one, two, or three closest dates (i.e. t_{1-1}, t_1, t_{1+1}) rather than the closest date t_1, (b) smoothing the c scaling factor by taking the mean of the MODIS BRDF parameters over a 35 x 35 30m pixel local area, (c) weighting the MODIS BRDF parameters using the MODIS BRDF product QA state information, (d) temporally weighting the MODIS BRDF parameters based on their temporal distance from the Landsat date. Generally as the window size increases and the number of closest BRDF inversion periods increase, the efficacy of the method increases. Also the efficacy tends to improve with the use of BRDF QA weighting and temporal weighting.

5.5 Annual Percent Tree, Bare Ground, Vegetation and Water Classification

WELD 30m land cover products are being developed based on the MODIS Vegetation Continuous Fields (VCF) bagged decision tree classification approach (Hansen et al. 2003). The 30m WELD land cover product suite will consist of annual maximum percent tree cover, minimum percent bare ground extent (including snow/ice), annual maximum percent vegetation cover (excluding tree cover), annual minimum percent surface water, and annual minimum percent snow/ice extent (nested within bare ground). Initial 2008 CONUS WELD results have proven promising (Hansen et al. 2011). In addition, the possibility of generating weekly percent bare ground, weekly percent surface water, and weekly percent snow/ice, will be investigated.

Training data are being collated to generate 30m WELD prototype maps of annual maximum percent tree cover, minimum percent bare ground extent, maximum percent vegetation cover (excluding tree cover), and minimum surface water probability, for 2006 to 2010. A preliminary water probability map is currently being converted to annual minimum percent surface water using a new sub-pixel percent water training data set. For the preliminary maximum percent tree cover layer, over 7 million 30m pixels of percent tree cover training data have been derived from crown/no crown classifications of

4m Ikonos and 2.8m QuickBird multi-spectral images distributed across the CONUS based on the method described by Hansen et al. (2002). Visual interpretation of very high spatial resolution (< 10m) GoogleEarth imagery was performed to fill regional gaps where Ikonos and QuickBird imagery was lacking. An additional 3.3 million 30m pixels of either 100% or 0% canopy cover have been interpreted from several thousand locations across the CONUS. Percent bare ground training data were defined from the maximum value of the coregistered 30m National Land Cover Database (NLCD) percent impervious surface (Homer et al. 2007) and the 500m MODIS VCF percent bare ground (Hansen et al. 2003) products. The NLCD layer captures impervious surface and not naturally bare areas, such as deserts and semi-arid lands that are mapped by the MODIS VCF layer. Water probability training data are being extracted by sampling the water pixels defined in the 2000 30m SRTM water body dataset (Rabus et al. 2003).

All the WELD land cover products will employ sub-pixel percent cover training data and will be generated from 25 bagged regression trees grown using a random 5% sampling of the training data sets with replacement for each tree. All per pixel results are ranked over the 25 trees, and the median percent cover estimate taken as the final classification output. The annual products will be generated using temporal metrics, based on ranked growing season quantiles similar to those used in the MODIS VCF product suite, derived from growing season WELD inputs from March to December for band 3, 4, 5, 7 TOA reflectance and TOA NDVI (Hansen et al. 2005). The shorter wavelength visible blue and green ETM+ bands 1 and 2 are not used due to their greater sensitivity to atmospheric effects (Ouaidrari and Vermote 1999). In addition, based on previous Landsat land cover mapping research (Hansen et al. 2008) simple band ratios of WELD TOA reflectance in band3/band5, band3/band7, band4/band5, band4/band7 and band5/band7 will be used. To help generate the surface water product, a slope layer derived from the 30m US National Elevation Dataset (http://ned.usgs.gov/) will be used as an input variable.

6.0 QUALITY ASESSMENT PLAN

Since a major WELD project objective is to produce a consistent Landsat data set, it is of paramount importance to identify deleterious product artifacts so that improvements may be implemented for future reprocessing and users informed.

Quality assessment (QA) is integral to the testing and development of the WELD products following a reduced-scope version of the operational QA approach developed and implemented by the MODIS Land Science Team (Roy et al. 2002).

In the first years of code and product development, the quality of the input products have been tracked through our affiliations and memberships with the MODIS Land and Landsat Science Teams and by communication with associated quality assessment personnel. The quality and consistency of the processed WELD products are tracked by systematic examination of the true color browse images (and browse images that are generated for QA of other WELD product bands) which enable rapid identification of problematic areas/periods for detailed inspection. In addition, informed feedback from the user community who are encouraged to participate in the QA process by sending email feedback to the WELD project email (sdsu.weld@sdstate.edu) has proven useful.

In the later years of the WELD funding period, as many years of WELD products become available, time series evaluations will be undertaken. Product time series analyses are important because they capture algorithm sensitivity to surface (e.g., vegetation phenology), atmospheric (e.g., aerosol loading) and remote sensing (e.g., sun-surface-sensor geometry) conditions that change temporally, and because they allow changes in the instrument characteristics and calibration to be examined (Roy et al. 2002).

7.0 VALIDATION PLAN

Inter-comparison of products made with different satellite data and/or algorithms provides an indication of their gross differences but comparison with independent reference data is required to assess satellite product accuracy (Justice et al. 2000).

Validation of the Version 1.5 products is quite restricted because of difficulties in finding reliable and contemporaneous independent reference data to compare with the WELD top of atmosphere (TOA) reflectance, brightness temperature, NDVI and cloud product data.

Flux tower measurements have been shown to provide suitable independent reference data for validation of the MODIS BRDF-albedo product (Román et al. 2009) and research to compare WELD NDVI with NDVI derived from the instantaneous radiance measured by flux tower pyranometers and photosynthetically active radiation sensors (Wittich and Kraft 2008) is being undertaken. Many of the AMERIFLUX flux towers have pyranometers and photosynthetically active radiation sensors, typically located several meters above the surface, that provide derived NDVI sensed over an area comparable to a Landsat pixel. Statistical correlation of contemporaneous non-cloudy WELD NDVI and flux tower NDVI over several years across the CONUS will provide an indication of the accuracy of the WELD NDVI product.

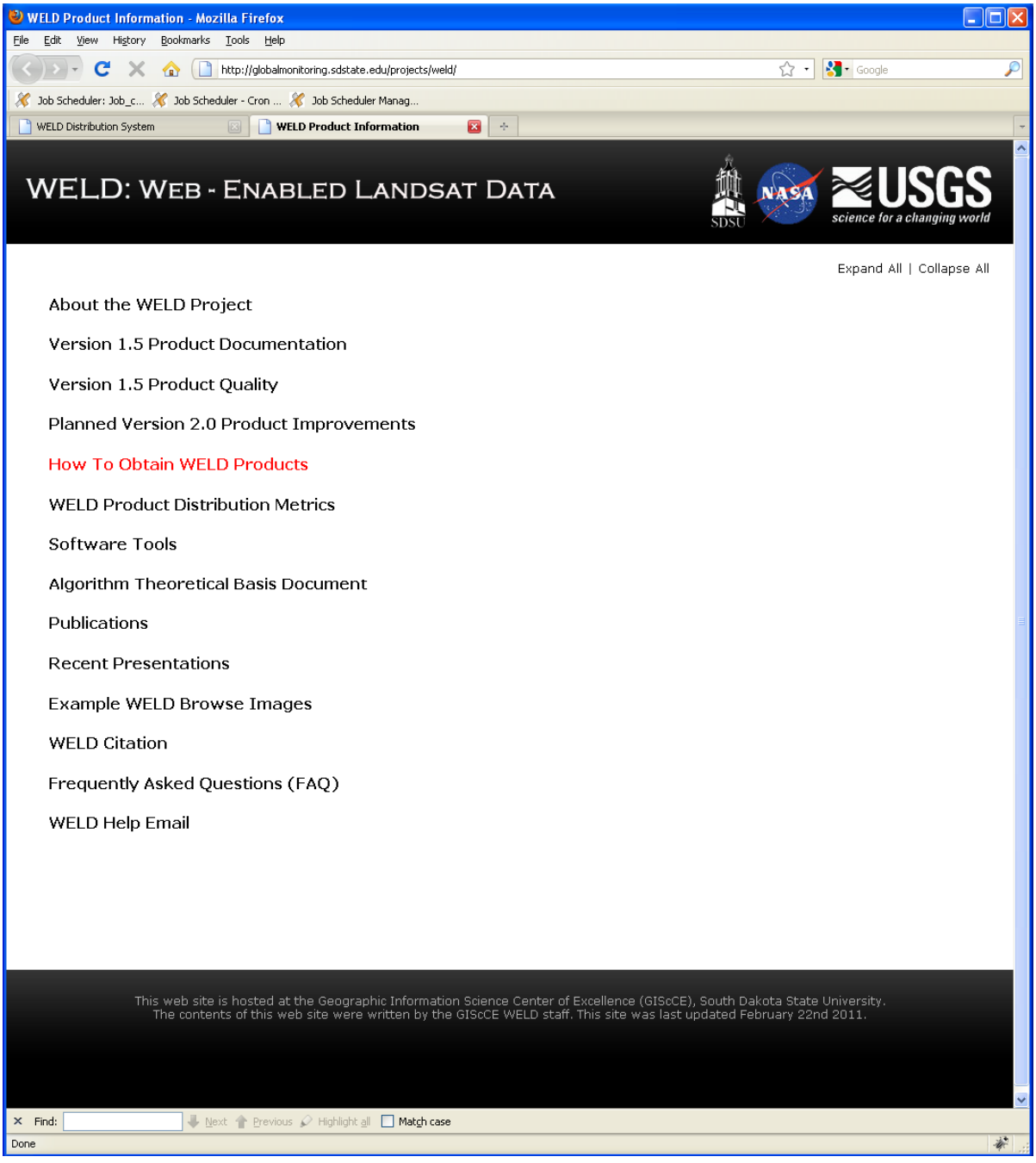
Since it is very difficult to validate cloud detection algorithms (Irish et al. 2006, Platnick et al. 2003), the global database of Landsat Level 1G and corresponding spatially explicit cloud masks used to derive the WELD classification tree cloud mask (Section 2.5.2) will be revisited. The WELD cloud classification tree used 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). The classification tree and ACCA WELD cloud mask algorithms will be validated by their application to other L1G scenes not included in the 88 scenes and then compared with the corresponding spatially explicit cloud masks.

Validation of the Version 2.0 WELD products will focus on validation of the atmospherically corrected reflectance by comparison with AEORNET data (Holben et al. 1998), internal consistency checking of the BRDF radiometrically normalized reflectance data sensed in successive overlapping orbits in the Landsat back and forward scatter directions, and comparison of the land cover products with the training data used to generate the recent 2006 National Land Cover Data (NLCD 2006) product.

8.0 USER SUPPORT

The WELD Product Documentation website http://globalmonitoring.sdstate.edu/projects/weld/ provides a comprehensive repository of WELD product information and documentation and is updated on a frequent basis.

Figure 15 Screen shot of the WELD Product Documentation Web Site



The WELD Product Documentation web site is maintained and hosted at SDSU GIScCE with clear links to the distribution web sites hosted at the USGS EROS and the Oak Ridge National Laboratory Distributed Active Archive Center. The WELD Product Documentation web site is compliant with public access via standard web browsers, and carries the NASA, USGS EROS and SDSU GIScCE logos. In addition to describing the WELD product formats and contents, information on the WELD product quality including known product issues, product improvements on previous WELD product versions, software tools, publications, recent presentations, a Frequently Asked Questions (FAQ) section, and a WELD help email, are available.

A workshop to facilitate outreach and to evolve the WELD products and aspects of the processing and distribution in step with user community requirements will be undertaken in the penultimate year of the project funding period. Prior to the workshop the same year of Version 1.5 (or later) and Version 2.0 WELD products will be made available for user comparison and assessment for at least a three month period. At the workshop, presentations on the WELD products and documentation will be made and participant feedback on user requirements and suggestions for improvements will be solicited and used for clarifying and refocusing these project elements. An emphasis will be on collating user reactions to the different processing paths used in the Version 1.5 (or later) and Version 2.0 WELD products. These reactions will help guide the selection of the most appropriate reprocessing for the entire 2005-2012 product data set. The workshop will be convened at a MODIS Science Team, Landsat Data Continuity Mission, or DAAC Land User Workshop, meeting.

9.0 GLOBAL WELD PATHFINDING

The WELD project has demonstrated the potential of large volume Landsat data processing to provide, on a systematic repeat basis, near continental scale data products that provide a high spatial resolution analogue to the moderate and coarse spatial resolution land products generated from the MODIS and AVHRR data streams (Justice et al. 2002, Tucker et al. 2005). The geographic extent of the WELD study area will be

increased in the last year of the grant period to pathfind the challenges to expanding the WELD production to global scale. The challenges to processing global scale Landsat data are considerable. Global Landsat processing is a "Big Data" issue - there are approximately 150 million MODIS 1km land pixels globally and three orders of magnitude more Landsat 30m pixels; the conterminous United States alone is composed of 11,000,000,000 30m pixels (Roy et al. 2010). The processing challenge is not only data volume; global satellite product generation experience has shown a need for algorithm refinement and product reprocessing (Townshend, 1994, Masuoka et al. 2010).

10.0 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., Patricia Vornberger, Andrew Fleming, Adrian Fox, Jerry Mullins, Douglas Binnie, Sara Jean Paulsen, Brian Granneman and David Gorodetzky (2008). The Landsat Image Mosaic of Antarctica, *Remote Sensing of Environment* 112, 4214–4226

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*, **26**, pp. 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.*, *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.

CCSP 2003. *Strategic Plan for the US Climate Change Science Program*. A report by the Climate Change Science Program and the Subcommittee on Global Change Research. 202p.

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 Climate Observing System, 2003, The Second Report on the Adequacy of the Global Observing Systems for Climate in Support of the UNFCCC, WMO-IOC-UNEP-ICS, GCOS-82, Technical Document No. 1143, World Meteorological Organization, Geneva, Switzerland, 85 pp.

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 Env., 78*, 3-12.

Group on Earth Observations, 2005, *Global Earth Observation System of Systems, 10-year implementation plan reference document,* ESA publications division: Noordwijk, Netherlands.

Gutman, G., Byrnes, R., Masek, J., Covington, S., Justice, C., Franks, S., and Headley, R. 2008. Towards monitoring Land-cover and land-use changes at a global scale: the global land survey 2005. *Photogrammetric Engineering and Remote Sensing.* 74(1):6-10.

Hansen, M.C., Egorov, A, Roy, D.P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., Loveland, T., 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. G., Carroll, M., Dimiceli, C., And 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**, paper no. 10, 15 pp. [online journal]

Hansen, M.C., Defries, R.S., Townshend, J.R.G., Marufu, L., Sohlberg, R., 2002, Development of a MODIS percent tree cover validation data set for Western Province, Zambia. *Remote Sensing of Environment*, **83**, pp. 320-335.

Hansen, M. C., Townshend, J. R. G., 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**, pp. 4359-4380.

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

Huete A, Didan K, Miura T, Rodriguez EP, Gao X, Ferriera LG (2002) Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens Environ* 82:195–213.

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

Holben, B. N. ,F . Eck, I. Slutsker, D. Tanre, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov, (1998), AERONET—A federated instrument network and data archive for aerosol characterization, *Remote Sensing of Environment. 66,* 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 conterminous United States. *Photogrammetric Engineering and Remote Sensing*, **73**, pp. 337-341.

Irish, R. I., (2000). Landsat7 automatic cloud cover assessment, in algorithms for multispectral, hyperspectral, and ultraspectral imagery, *Proc. SPIE*, vol. 4049, 2000.

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. and Roy, D.P., 2008, The Availability of Cloud-free Landsat ETM+ data over the Conterminous United States and Globally, *Remote Sensing of Environment*, **112**, pp. 1196-1211.

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. and Baret, F., 2000, Developments in the 'validation' of satellite sensor products for the study of the land surface. *International Journal of Remote Sensing*, **21**, pp. 3383 - 3390.

Kaufman, Y. J., Tanre, D., Remer, L. A., Vermote, E. F., Chu, A., and Holben, B. N., 1997, Operational remote sensing of tropospheric aerosol over the land from EOS-MODIS. Journal of Geophysical Research-Atmosphere , 102(14), 17051-17068

Konecny, G., "Methods and possibilities for digital differential rectification," *Photogramm. Eng. Remote Sensing*, vol. 45, no. 6, pp. 727–734, 1979.

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.

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. I*EEE Transactions on Geoscience and Remote Sensing*, 42, 2786−2795.

Lindquist, E., Hansen, M.C**.,** Roy, D.P., Justice, C.O., 2008, The suitability of decadal image data sets for mapping tropical forest cover change in the Democratic Republic of Congo: implications for the mid-decadal global land survey, *International Journal of Remote Sensing*, 29: 7269–7275

Liu, H. Q., and Huete, A. R. (1995). A feedback based modification of the NDVI to minimize canopy background and atmospheric noise," *IEEE Trans. Geosci.Remote Sens.*, vol. 33, pp. 457-465.

Markham, D., Goward, G., Arvidson, T., Barsi, J., Scaramuzza, P., 2006, Landsat-7 Long-Term Acquisition Plan Radiometry - Evolution Over Time, *Photogrammetric Engineering & Remote Sensing*, Vol. 72, No. 10, October 2006, pp. 1129–1135.

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. Geoscience and Remote Sensing Letters, 3, 68-72.

Masek, J.G., Friedl, M., Loveland, T., Brown de Colstoun, E., Townshend, J., Hansen, M., Ranson, K.J., 2006b, *A Community Land Earth System Data Record (ESDR) White Paper on Land Cover Land Cover Change,* requested by NASA, ftp://ftp.iluci.org/Land_ESDR/Landcover-change_Masek_whitepaper.pdf

Masek, J., 2007, White Paper on Use of Gap-Filled Products for the Mid-Decadal Global Land Survey (GLS). Available online at: http://mdgls.umd.edu/documents/MDGLSgapfill.pdf.

Masuoka, E., Roy, D.P., Wolfe, R., Morisette, J., Sinno, S., Teague, M., Saleous, N., Devadiga, S., Justice, C., and Nickeson, J., 2010, MODIS land data products—generation, quality assurance and validation, chapter 22 in Ramachandran, B., Justice, C.O., and Abrams, M.J., eds., *Land remote sensing and global environmental change—NASA's Earth Observing System and the science of ASTER and MODIS*, New York, Springer, p. 511-534.

National Academy of Sciences. *Down to Earth Geographical Information for Sustainable Development in Africa*. National Academies Press, 2002. available online at http://www.nap.edu/catalog/10455.html.

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.

Rabus, B., Eineder, M., Roth, A., Bamler, R., 2003, The shuttle radar topography mission- a new class of digital elevation models acquired by spaceborne radar. *Photogrammetric Engineering and Remote Sensing* , **57**, pp. 241-262.

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

REDD Sourcebook, 2009, A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals caused by deforestation, gains and losses of carbon stocks in remaining forests, and forestation. GOFC-GOLD Project Office, Natural Resources Canada, Alberta, Canada, available at http://www.gofc-gold.uni-jena.de/redd/.

Román, M. O., Schaaf, C. B., Woodcock, C. E., Strahler, A. H., Yang, X., Braswell, R. H., Curtis, P. S., Davis, K. J., Dragoni, D., Goulden, M. L., Gu, L., Hollinger, D. Y., Kolb, T. E., Meyers, T. P., Munger, J. W., Privette, J. L., Richardson, A. D., Wilson, T. B. and Wofsy, S. C., 2009, The MODIS (Collection V005) BRDF/albedo product: Assessment of spatial representativeness over forested landscapes. *Remote Sensing of Environment*, **113**, pp. 2476-2498.

Rowland, J., Wood, E., and Tieszen, Larry L., 2007, Review of remote sensing needs and applications in Africa: U.S. Geological Survey Earth Resources Observation and Science Center, 124 p., available only online at http://lca.usgs.gov/lca/rsca/dataproducts.php.

Roy, D. (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,* 38:2017-2032.

Roy, D., Borak, J, Devadiga, S., Wolfe, R., Zheng, M., Descloitres, J., 2002, The MODIS land product quality assessment approach, *Remote Sensing of Environment*, 83:62-76.

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: 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**, pp. 35-49.

Saalfeld, A. (1985). A fast rubber-sheeting transformation using simplicial coordinates. *The American Cartographer*, vol 12, No. 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.

Snyder, J.P. (1993). *Flattening the Earth: Two thousand years of map projections*, The University of Chicago Press, Chicago and London.

Teillet, P. M., Saleous, N. El., Hansen, M. C., Eidenshink, J. C., Justice, C. O., Townshend, J. R. G. (2000). An evaluation of the global 1-km AVHRR land data set. International Journal of Remote Sensing, 21, 1987– 2021.

Townshend, J.R.G. and Justice, C.O., 1988, Selecting the spatial resolution of satellite sensors required for global monitoring of land transformations, *International Journal of Remote Sensing*, **9**, pp. 187–236.

Townshend, J. R. G. (1994). Global data sets for land applications from the Advanced Very High-Resolution Radiometer—an introduction. Int. Journal of Remote Sensing, 15, 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 Dataset. *Photogrammetric Engineering & Remote Sensing.* 70(3), 313-322.

Tucker, C., Pinzon, P., Brown, M., Slayback, D., Pak, E., Mahoney, R., "An extended AVHRR 8-km NDVI data set compatible with MODIS and SPOT vegetation NDVI data", *International Journal of Remote Sensing*, Vol. 26(20), p.p. 4485−4498, 2005.

U.N.2010, *Report of the Subsidiary Body for Scientific and Technological Advice on its thirty-first session*, held in Copenhagen from 8 to 12 December 2009, FCCC/SBSTA/2009/8, United Nations Office at Geneva.

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. and 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**, pp. 167-177.

Wolfe, R., Roy, D., Vermote, E. (1998). The MODIS land data storage, gridding and compositing methodology: L2 Grid. *IEEE Transactions on Geoscience and Remote Sensing, 36*,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, pp. 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., J.C. White, S.N. Goward, J.G. Masek, J.R. Irons, M. Herold, W.B. Cohen, T.R. Loveland, and C.E. Woodcock, 2008, Landsat continuity: Issues and opportunities for land cover monitoring, *Remote Sensing of Environment,* **112,** pp. 955-969.

Zobrist, A.L., Bryant, N.A. and R.G. McLeod. (1983). Technology for Large Digital Mosaics of Landsat Data, *Photogrammetric Engineering and Remote Sensing*, *49,* 1325-1335.

Acknowledgements

The 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 NNX08AL93A. The NASA MEaSUREs Program Officer is Martha Maiden.

The free Landsat L1T data access and the WELD HDF and GeoTIFF product distribution hardware and maintenance is funded by the United States Geological Survey (USGS).

The WELD true color browse images have been made Open Geospatial Consortium (OGC) compliant and are served from the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center funded by the ORNL.

This ATBD could not have been written without regular visits to http://www.cottonwoodcoffee.com/