Remote Sensing of Environment 114 (2010) 35–49

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States

David P. Roy [a,*], Junchang Ju [a], Kristi Kline [b], Pasquale L. Scaramuzza [b], Valeriy Kovalskyy [a], Matthew Hansen [a], Thomas R. Loveland [b], Eric Vermote [c], Chunsun Zhang [a]

[a] Geographic Information Science Center of Excellence, South Dakota State University, Brookings, SD 57007, USA
[b] U.S. Geological Survey Center for Earth Resources Observation and Science, 47914 252nd Street, Sioux Falls, SD 57198, USA
[c] Department of Geography, University of Maryland, 1113 LeFrak Hall, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 21 May 2009
Received in revised form 1 August 2009
Accepted 7 August 2009

Keywords:
Landsat ETM+
Time series
Free data
Continental
Long-term data record
Phenology
Composite
Mosaic

ABSTRACT

Since January 2008, the U.S. Department of Interior / U.S. Geological Survey have been providing free terrain-corrected (Level 1T) Landsat Enhanced Thematic Mapper Plus (ETM+) data via the Internet, currently for acquisitions with less than 40% cloud cover. With this rich dataset, temporally composited, mosaics of the conterminous United States (CONUS) were generated on a monthly, seasonal, and annual basis using 6521 ETM+ acquisitions from December 2007 to November 2008. The composited mosaics are designed to provide consistent Landsat data that can be used to derive land cover and geo-physical and bio-physical products for detailed regional assessments of land-cover dynamics and to study Earth system functioning. The data layers in the composited mosaics are defined at 30 m and include top of atmosphere (TOA) reflectance, TOA brightness temperature, TOA normalized difference vegetation index (NDVI), the date each composited pixel was acquired on, per-band radiometric saturation status, cloud mask values, and the number of acquisitions considered in the compositing period. Reduced spatial resolution browse imagery, and top of atmosphere 30 m reflectance time series extracted from the monthly composites, capture the expected land surface phenological change, and illustrate the potential of the composited mosaic data for terrestrial monitoring at high spatial resolution. The composited mosaics are available in 501 tiles of 5000 × 5000 30 m pixels in the Albers equal area projection and are downloadable at http://landsat.usgs.gov/WELD.php. The research described in this paper demonstrates the potential of Landsat data processing to provide a consistent, long-term, large-area, data record.

1. Introduction

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 the National Aeronautics and Space Administration (NASA), represents the longest temporal record of space-based land observations (Williams et al., 2006). Until recently the primary limitations to using Landsat data have been the cost and availability of data, which have precluded continental to global scale Landsat studies (Hansen et al., 2008). 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, over 2.2 million globally distributed Landsat acquisitions, at no cost via the Internet. Landsat acquisitions with a cloud cover of less than or equal to 40% are processed and made freely available as they are acquired, and users may request any other scene in the U.S. Landsat archive to be processed and made available via the Internet at no cost. 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 & Justice, 1988), during a time when the human population has doubled and the impacts of climate change have become noticeable (Woodcock et al., 2008).

The Landsat 7 Enhanced Thematic Mapper Plus (ETM+) is the most recent in a series of Landsat sensors that acquire high spatial resolution multi-spectral data over an approximately 183 km × 170 km extent, defined in a Worldwide Reference System of path (groundtrack parallel) and row (latitude parallel) coordinates, with a 16 day revisit capability (Williams et al., 2006). Every Landsat overpass of the conterminous United States (CONUS) is acquired by the U.S. Landsat Project, providing 22 or 23 acquisitions per year per path/row (Ju & Roy, 2008). The Landsat project does not acquire every acquisition globally due to ground system processing and archiving constraints (Arvidson et al., 2006). Cloud cover reduces the number of Landsat surface observations; for example, the annual mean Landsat ETM+ cloud cover for the CONUS and global scenes stored in the U.S. Landsat archive is about 40% and 35%

* Corresponding author.
E-mail address: david.roy@sdstate.edu (D.P. Roy).

0034-4257/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.rse.2009.08.011

respectively (Ju & Roy, 2008). In addition, in May 2003 the ETM+ scan line corrector (SLC) failed, reducing the usable data in each Landsat ETM+ SLC-off scene by 22% (Storey et al., 2005; Maxwell et al., 2007).

Arguably, the utility of Landsat data for long-term and/or large-area monitoring has not been fully assessed; to date, the majority of the data in the U.S. Landsat archive have not been used in applications science. A number of regional, continental and global Landsat data sets have been generated however. Regional mosaics of Landsat imagery are increasingly being developed to meet national monitoring and reporting needs across land-use and resource sectors, for example, in Canada (Wulder et al., 2002) and the Congo basin (Hansen et al., 2008). Large volume Landsat processing was developed by the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) that processed over 2100 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 1100 Landsat ETM+ austral summer acquisitions (Bindschadler et al., 2008). At global scale, the Global Land Survey decadal Landsat data set provides relatively cloud-free acquisitions selected for each path/row from the 1970s, 1990s and 2000s (Tucker et al., 2004). These data sets are composed of single date manually selected Landsat acquisitions. With the advent of free Landsat data it becomes feasible to apply temporal compositing approaches to multi-temporal Landsat acquisitions of the same path/row. Compositing procedures are applied independently on a per-pixel basis to gridded satellite time series data 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 et al., 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 a high likelihood of cloud, haze or shadow. Similar observations have been observed for compositing moderate and coarse spatial resolution satellite data (Holben, 1986, Cihlar et al., 1994, Roy, 2000).

In this paper we describe the generation of Landsat ETM+ composited mosaics of the CONUS for December 2007 to November 2008. The composited mosaics are designed to provide consistent Landsat data that can be used to derive land cover and geo-physical and bio-physical products needed for detailed regional assessments of land-cover dynamics and to study Earth system functioning (Gutman et al., 2008; Wulder et al., 2008). The composited mosaics are generated on a monthly, seasonal, and annual basis to provide data that capture temporal surface variations. The compositing approach is designed to preferentially select valid land surface observations with minimal cloud, snow, and atmospheric contamination; consequently the composited mosaics are not appropriate for Landsat studies of cloud, snow or the atmosphere. The processing steps and data products are informed by our MODIS Land processing, quality assessment and validation experience (Justice et al., 2002). The processing approach is intentionally designed to facilitate automated processing with minimal human intervention, including the requirement for composited mosaics to be updated regardless of the chronological order of the Landsat acquisition and processing dates, and to provide processing in near-real time i.e., updating composited mosaics shortly after the Landsat data are acquired. Information on how to obtain the composited mosaic products and further research is described.

2. Input web-enabled level 1T Landsat ETM+ data

The Landsat data made freely available by the U.S. Landsat project are sometimes called "web-enabled" data as they are made available via the Internet. These data are nominally processed as Level 1 terrain-corrected (L1T) data. The L1T data are available in GeoTIFF format in the Universal Transverse Mercator (UTM) map projection with World Geodetic System 84 (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 CONUS is less than 30 m even in areas with substantial terrain relief (Lee et al., 2004). While most web-enabled Landsat data are processed as L1T (i.e., precision and terrain-corrected), 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 is applied and, over the CONUS, the data are processed to Level 1G systematic (L1G) (WWW1). The product file metadata notes if the acquisition was processed to L1T or L1G.

Landsat acquisitions with cloud cover less than or equal to 40% are processed and made freely available as they are acquired, and users may request any other scene in the U.S. Landsat archive to be processed and made available at no cost via the Internet. Initially, the U.S. Landsat project implemented a 20% cloud cover threshold but this was increased to 40% in October 2008 after a processing load analysis and in response to requests from the user community including from this research project. In, addition, the U.S. Landsat project bulk processed all the CONUS ETM+ acquisitions with less than 40% cloud cover acquired from May 2007 to October 2008 in support of this project. The cloud cover of each ETM+ acquisition is estimated operationally by the automatic cloud cover assessment (ACCA) algorithm during the archiving process (Irish et al., 2006).

Fig. 1 shows the CONUS study area, defined by 459 Landsat path/row coordinates, covering about 11,000,000,000 30 m land pixels. All the web-enabled Landsat ETM+ data acquired over the study area for a 12 month period from December 2007 to November 2008 were obtained. A total of 6959 acquisitions were copied by dedicated file transfer protocol from the U.S Landsat project to computers at the Geographic Information Science Center of Excellence, South Dakota State University. Of the 6959 acquisitions, 438 (6.3%) were found to be processed as L1G and not as L1T. Fig. 2 shows histograms of the ACCA cloud cover percentage for the 6521 L1T (left) and the 438 L1G (right) acquisitions. The acquisitions that were processed as L1G have markedly higher cloud cover (mean 61%) compared to those processed as L1T (mean 18%). This is expected as higher cloud cover reduces the number of available ground control chips needed in the L1T processing, although a minority of acquisitions processed as L1T also have high cloud cover. The locations of acquisitions that had less than 20% ACCA cloud cover but that were processed to L1G are illustrated in Fig. 1 (black filled circles). For all 105 of these the availability of ground control chips and/or the reliability of the ground control matching algorithm was compromised for reasons other than cloud cover: 100 acquisitions were of 10 coastal water path/row locations where land ground control was less available, another (path 30, row 27) was extensively contaminated by snow reducing the availability of ground control, and four were at the same location in Kansas (path 30, row 34) where there is extensive center-pivot irrigation (Wardlow et al., 2007) that are thought to confuse the matching algorithm.

Table 1 summarizes the number, ACCA cloud cover, and seasonal distribution of acquisitions in the U.S. Landsat project archive. The number of acquisitions in the archive is about the same in each season (Table 1, column 2), but because of the higher cloud cover in winter and spring, fewer acquisitions were available with less than 40% ACCA