The global availability of Landsat 5 TM and Landsat 7 ETM+ land surface observations and implications for global 30 m Landsat data product generation

V. Kovalskyy *, D.P. Roy

Geographic Information Science Center of Excellence, South Dakota State University Brookings, SD 57007, USA

ABSTRACT

With the advent of the free U.S. Landsat data policy it is now feasible to consider the generation of global coverage 30 m Landsat data sets with temporal reporting frequency similar to that provided by the monthly Web Enabled Landsat (WELD) products. A statistical Landsat metadata analysis is reported considering more than 800,000 Landsat 5 TM and Landsat 7 ETM+ acquisitions obtained from the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center archive. The global monthly probabilities of acquiring a cloud-free land surface observation for December 1998 to November 2001 (2000 epoch) and from December 2008 to November 2011 (2010 epoch) are reported to assess the availability of the Landsat data in the USGS Landsat archive for global multi-temporal land remote sensing applications. The global probabilities of acquiring a cloud-free land surface observation in each of three different seasons with the highest seasonal probabilities of cloud-free land surface observation are reported, considering one, two and three years of Landsat data, to assess the availability of Landsat data for global land cover mapping. The probabilities are derived considering Landsat 5 TM only, Landsat 7 ETM+ only, and both sensors combined, to examine the relative benefits of using one or both Landsat sensors. The results demonstrate the utility of combing both Landsat 5 TM and Landsat 7 ETM+ data streams to take advantage of their different acquisition patterns and to mitigate the deleterious impact of the Landsat 7 ETM+ 2003 scan line failure. Sensor combination provided a greater global acquisition coverage with a 1.7% to 14.4% higher percentage of land locations acquired monthly compared to considering Landsat 7 ETM+ data alone. The mean global monthly probability of a cloud-free land surface observation for the combined sensors was up to nearly 1.4 and 6.7 times greater than for ETM+ and TM alone respectively. The probability of acquiring a cloud-free Landsat land surface observation in different seasons was greater when more years of data were considered and when both Landsat sensor data were combined. Considering combined sensors and 36 months of data, 86.4% and 84.2% of the global land locations had probabilities ≥0.95 for the 2000 and 2010 epochs respectively, with a global mean probability of 0.92 (σ 0.24) for the 2000 epoch and 0.90 (σ 0.28) for the 2010 epoch. These results indicate that 36 months of combined Landsat sensor data will provide sufficient land surface observations for 30 m global land cover mapping using a multi-temporal supervised classification scheme.

1. Introduction

Landsat data provide a balance between the requirements for localized high-spatial resolution studies and global monitoring (Goward et al., 2001) and the free Landsat data policy opens a new era for utilizing the more than three million scenes stored in the U.S. Landsat archive at the United States Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (henceforth termed the USGS Landsat archive) (Wulder et al., 2012). Prior to the free Landsat data policy, global coverage Landsat data sets were developed through NASA and USGS data buys but only a fraction of the USGS Landsat archive was exploited (Gutman et al., 2008; Masek, 2007; Tucker et al., 2004). These data sets, originally called

Geocover, were reprocessed and are now termed the Global Land Survey (GLS) data and provide global, orthorectified, low cloud cover Landsat images centered on the years 1975, 1990, 2000, 2005, selected with a preference for leaf-on conditions (Gutman et al., 2008). Collectively, the GLS data are designed to provide a consistent set of observations to assess land-cover changes at quasi-decadal scale. However, each GLS data set is composed of from only one to three Landsat images per land location which provides insufficient data to capture surface changes required for optimal land cover classification. The NASA funded Web Enabled Landsat (WELD) project has demonstrated the capability to generate near-continental scale Landsat composited mosaics with a weekly, monthly, seasonal and annual reporting frequency (Roy et al., 2010a; WWW1). With the advent of the free Landsat data policy it is now feasible to consider the generation of global coverage 30 m Landsat data sets with reporting frequency similar to that provided by the WELD products. This

---

* Corresponding author. Tel.: +1 9176596046.
  E-mail address: Valeriy.Kovalskyy@sdstate.edu (V. Kovalskyy).

paper investigates this potential, specifically for global 30 m Landsat data product generation.

Global 30 m observations have been provided by the Landsat 5 Thematic Mapper (TM) and Landsat 7 Enhanced Thematic Mapper Plus (ETM+) from 1984 to present, corresponding to about 75% of the duration of the 40 year Landsat temporal data record (Loveland & Dwyer, 2012; Williams et al., 2006). In this paper the global probabilities of acquiring a cloud-free Landsat land surface observation in each month and in three different seasons are reported to assess the availability of the Landsat data in the USGS Landsat archive for global multi-temporal remote sensing applications and land cover mapping respectively. The probabilities are derived for the Landsat 5 TM and Landsat 7 ETM+ independently and for both sensors together. Combination of the data from both sensors is investigated because it may provide improved global cloud-free land coverage and is likely needed because the free USGS Landsat archive does not contain data for every Landsat land overpass (Arvidson et al., 2006; Goward et al., 2006; Wulder et al., 2012), cloud cover reduces the number of Landsat surface observations (Ju & Roy, 2008), and in May 2003 the Landsat 7 ETM+ scan line corrector failed, reducing the usable data in each Landsat ETM+ scene by 22% (Markham et al., 2004). A total of 36 months of global Landsat data centered on climate years 2000 and 2010 are considered because these two periods occur before and after the 2003 Landsat ETM+ scan line corrector failure (Markham et al., 2004). The geographic location, extent and overlap of Landsat acquisitions imposed by the Landsat sensor and orbit geometry is considered explicitly in the reported analysis. Maps and global summary statistics are generated for 12, 24 and 36 month periods centered on the 2000 and 2010 epochs to quantify the utility of using one, two and three years of Landsat acquisitions for global land cover mapping.

2. Data

2.1. USGS Landsat archive holdings

The amount of Landsat data in the USGS Landsat archive is not constant from year to year, geographically, or among sensors, because of differing data reception capabilities and Landsat system health issues and because of selective Landsat scene acquisition due to factors including payload, ground station and mission cost constraints (Arvidson et al., 2006; Goward et al., 2006; Loveland & Dwyer, 2012; Markham et al., 2004).

The Landsat 5 TM was launched in March 1984 and the Landsat 7 ETM+ was launched in April 1999. They have 15° fields of view and are in approximately 710 km sun-synchronous polar orbits that each overpass every Earth location every 16 days. For several days in early June 1999 the Landsat 7 ETM+ was in nearly the same orbit as Landsat 5 for sensor cross-calibration (Teillet et al., 2001). The Landsat 7 ETM+ obtained its operational orbit by the end of June 1999 with the 16-day repeat cycle of Landsat 5 TM and Landsat 7 ETM+ offset from each other by 8 days (Arvidson et al., 2006; Teillet et al., 2001). Landsat 7 ETM+ acquisitions are scheduled systematically using a Long-Term Acquisition Plan (LTAP) that refreshes annually the USGS Landsat archive with sunlit, substantially cloud-free acquisitions that capture seasonal land surface dynamics (Arvidson et al., 2006). In contrast, the Landsat 5 TM has had no systematic acquisition plan, reflecting in part the different federal agency and commercial company operational mandates for Landsat 5 (Goward et al., 2006). Unlike the Landsat 7 ETM+, the Landsat 5 TM has no onboard data recorder and consequently used the Tracking and Data Relay Satellite System (TDRSS) to downlink to the USGS Landsat archive data that were collected outside the United States. The Landsat 5 TM TDRSS antenna failed in 1987 and data transmission became limited to direct, real-time X-band transmission and only Landsat 5 TM data sensed within a U.S. ground station line-of sight

were copied to the USGS archive (Chander et al., 2007). Non-U.S. international ground stations collect Landsat 5 TM data via the X-band transmission but only a proportion of these data have been copied to the USGS archive (Goward et al., 2006; Loveland & Dwyer, 2012). After the 1999 launch of the Landsat 7 ETM+ some international ground stations switched their reception from Landsat 5 TM to Landsat 7 ETM+. When the Landsat 7 ETM+ scan line corrector system failed in May 2003 several international ground stations discontinued Landsat 7 ETM+ acquisitions and changed their operations to Landsat 5 TM. Most recently, the aging Landsat 5 TM is experiencing a number of issues including power limitations that are restricting the amount of worldwide continuous imaging (Loveland & Dwyer, 2012).

2.2. Landsat metadata used in this study

Landsat data are acquired in approximately 180 km × 170 km scenes defined in a Worldwide Reference System (WRS) of path (groundtrack parallel) and row (latitude parallel) coordinates (Arvidson et al., 2001). The specific metadata names used in this study are summarized in Table 1, they describe the WRS scene path and row, the date of acquisition, the cloud cover fraction (0, 1,…, 9), the sensor type (Landsat 5 TM or Landsat 7 ETM+), and the latitudes and longitudes of the four scene corners. The cloud cover fraction is defined using an automatic cloud cover assessment (ACCA) algorithm deployed to ascertain the cloud fraction of each Landsat scene prior to archiving in the USGS Landsat archive. The Landsat 5 TM ACCA algorithm was developed under the computer processing limitations of 1980's computers (Su, 1984) whereas the Landsat 7 ETM+ ACCA algorithm was developed without these constraints and uses a more complex and reliable algorithm (Irish, 2000; Irish et al., 2006).

The metadata for all daytime non-Antarctic acquisitions archived on 11th April 2012 were obtained from the USGS Landsat archive metadata database (WWW2). The date of metadata access is noted because the U.S. Landsat project has started repatriating Landsat data from other non-U.S. international Landsat stations and consequently the number of available scenes in the USGS Landsat archive is expected to increase (Loveland & Dwyer, 2012). Antarctic scenes were excluded from this study because for most of the year there are no daytime Antarctic Landsat observations and because of the substantial issues in processing Landsat data under snow-bound and high solar zenith angle conditions (Bindschadler et al., 2008).

Globally, a total of 844,797 Landsat metadata records were obtained for three climate years in two epochs from 1st December 1998 to 30th November 2001 (2000 epoch) and from 1st December 2008 to 30th November 2011 (2010 epoch) (Table 2). Thirty six months of metadata per epoch were examined in an attempt to ensure a sufficiently representative global data set that captures the variable amount of Landsat 5 TM and Landsat 7 ETM+ data in the USGS

Table 1
Landsat metadata used in the study and example metadata values.

Metadata name	Metadata value example
sceneID	LE70880662010334EDC00
sensor	LANDSAT_ETM_SLC_OFF
acquisitionDate	2010-11-30
path	88
Row	66
cloudCover	1
upperLeftCornerLatitude	−7.7466
upperLeftCornerLongitude	158.45857
upperRightCornerLatitude	−7.99601
upperRightCornerLongitude	160.17395
lowerLeftCornerLatitude	−9.36151
lowerLeftCornerLongitude	158.10541
lowerRightCornerLatitude	−9.61218
lowerRightCornerLongitude	159.82896

Table 2
Total global number of Landsat metadata records considered in this study summarized for 12, 24 and 36 month periods centered on May/June of 2000 and 2010.

2000 epoch	TM	ETM+	TM and ETM+
12 months 12/01/1999–11/30/2000	36,048	92,222	92,222
24 months 06/01/1999–05/31/2001	70,690	140,103	210,793
36 months 12/01/1998–11/30/2001	104,182	226,505	330,687
2010 epoch	TM	ETM+	TM and ETM+
12 months 12/01/2009–11/30/2010	57,576	100,333	157,909
24 months 06/01/2009–05/31/2011	174,423	199,426	373,849
36 months 12/01/2008–11/30/2011	203,412	310,698	514,110

Landsat archive and because several years of Landsat data may be needed to provide optimal Landsat land cover classification accuracies in persistently cloudy regions (Broich et al., 2011; Lindquist et al., 2008). The 2000 and 2010 epochs were selected as they occur before and after the May 2003 Landsat ETM+ scan line corrector failure. As this is a global study, a climate year definition is used where Winter is defined in a northern hemisphere sense by the months: December, January and February, Spring by the months: March, April, May, Summer by the months: June, July, August, and Autumn by the months: September, October, November (Rossow & Dueñas, 2004). In both epochs the greater proportion of metadata is Landsat 7 ETM+ (68% and 60% in the 2000 and 2010 epochs) rather than Landsat 5 TM (Table 2). The greater proportion of Landsat 7 ETM+ data in the USGS Landsat archive has been documented before (Roy et al., 2010b; Wulder et al., 2012) and occurs for the reasons discussed in Section 2.1.

We note that the Landsat 5 TM cloud fraction metadata for the 2000 epoch and for the first two months of the 2010 epoch (December 2008 and January 2009) were defined using the original L5 ACCA algorithm (Su, 1984) but for the remaining 34 months of the 2010 epoch (February 2009 to November 2011) were defined using an algorithm based on the L7 ETM+ ACCA algorithm (Irish et al., 2006). This Landsat 5 TM ACCA algorithm inconsistency however did not affect the global monthly 2010 epoch Landsat 5 TM cloud statistics that are presented in the Results section.

3. Methods

The availability of cloud-free Landsat 5 TM and 7 ETM+ land surface observations were considered capturing the geographic location, extent and overlap of Landsat acquisitions. A global grid of regularly spaced land points was compared with the Landsat scene corner coordinates and cloud fraction metadata to derive the probability at each land grid point of there being at least one cloud-free observation in a month and in multiple seasons. The seasonal probabilities were derived considering 12, 24 and 36 months of Landsat data per epoch to establish if 12, 24 or 36 months of Landsat data are needed to provide reliable global Landsat land cover mapping. The probabilities were derived considering Landsat 5 TM only, Landsat 7 ETM+ only, and both sensors combined, to examine the relative benefits of using one or both Landsat sensors.

3.1. Global land grid definition

The global grid was defined in an equal area projection to ensure that the surface area sensed by each Landsat acquisition was sampled with the same spatial grid density. The sinusoidal equal area projection was used as it provides a global uninterrupted projection (Snyder, 1987). The grid spacing was set sufficiently small to capture the variable geographic location and extent of Landsat acquisitions and scene overlap imposed by the Landsat sensor and orbit geometry. Fig. 1 illustrates the geographic areas sensed by four adjacent Landsat 7 ETM+ scenes derived from their corner coordinate metadata values. The Landsat

sensor orbits drift due to several factors, primarily due to Earth gravitational effects (Wertz, 2001), and are maintained by semi-periodic station keeping propellant usage (NASA, 1996). Table 3 summarizes the mean and maximum deviation in the geographic location of the upperLeftCornerLatitude and upperLeftCornerLongitude metadata values for the epochal Landsat 5 TM and Landsat 7 ETM+ metadata. The deviations were computed relative to the mean longitude and mean latitude for each WRS path/row having seven or more acquisitions within an epoch. The maximum deviations are 15.6 km East–West and 5.2 km North–South and occur very infrequently. The mean global deviation North–South is less than 1 km and no more than 3.05 km East–West (Table 3). Thus, the average global impact of Landsat orbit drift and orbit station keeping maneuvers on the location of the Landsat scenes is less than 2% of the 180 km scene dimension. Of greater magnitude, and occurring systematically, is the across-track overlap between scenes in adjacent Landsat orbits that increases significantly from Equatorial latitudes (Fig. 1a) further polewards (Fig. 1b) due to the poleward convergence of the Landsat orbits.

Fig. 2 shows global summary statistics of the across-track scene overlap distance for adjacent Landsat 5 TM and 7 ETM+ paths. The mean across-track overlap distance (Fig. 2a) is 0.196° at the Equator and increases to 12.535° and 12.826° at latitudes 80° South and 80° North respectively. At high latitudes the overlap is significant and is equivalent to more than to 90% of a scene. Fig. 2b shows the minimum, mean and maximum overlap distance for the equatorial latitudes 10° S to 10° N, where the overlap is the smallest. At the Equator (WRS row 60) the overlap ranges from a minimum of 0.105° to a maximum of 0.283°.

The global grid was defined with a spacing of 5.559752 km equivalent to 0.05° at the Equator and less than half the minimum 0.105° across-track overlap observed between adjacent Landsat paths (Fig. 2b). The grid was composed of 7200 by 3600 points defined in the sinusoidal projection. Typically, a single Landsat scene encompasses one thousand 5.559752 km spaced sinusoidal grid points. Only grid locations falling within the non-Antarctic Landsat scenes defined in the U.S. Landsat project "Land Definition" data base (WWW3) were considered. The resulting global land grid is illustrated in Fig. 3 and is composed of a total of 6,138,864 land non-Antarctic grid points.

3.2. Grid point in Landsat scene computation

A sorting algorithm was implemented to obtain efficiently the metadata of Landsat acquisitions encompassing each land grid point for a given period of interest. First the Landsat acquisition date was compared with the period of interest and then for those scenes falling within the period of interest their corner coordinates were projected into the sinusoidal grid and compared with the land grid point location using a point in polygon routine (O'Rourke, 1998). In the along-track direction the geographic areas sensed by successive scenes in the same Landsat path overlap by approximately 300 scanlines due to the timing of the onboard Landsat scene recording and transmission. The along-track overlapping areas were discarded by identifying any southern overlapping area and removing it from the southern part of the scene.

3.3. Probability of there being at least one cloud-free observation of a land grid point

The probability of there being at least one cloud-free observation of a land grid point within a given period is derived from the probability of all of the overpasses over that period being cloudy as:

$$P_{one+} = 1 - \prod_{i=1}^{n} p_i \qquad (1)$$

where P_{one+} is the probability of there being at least one cloud-free observation of the land grid point, n is the number of satellite
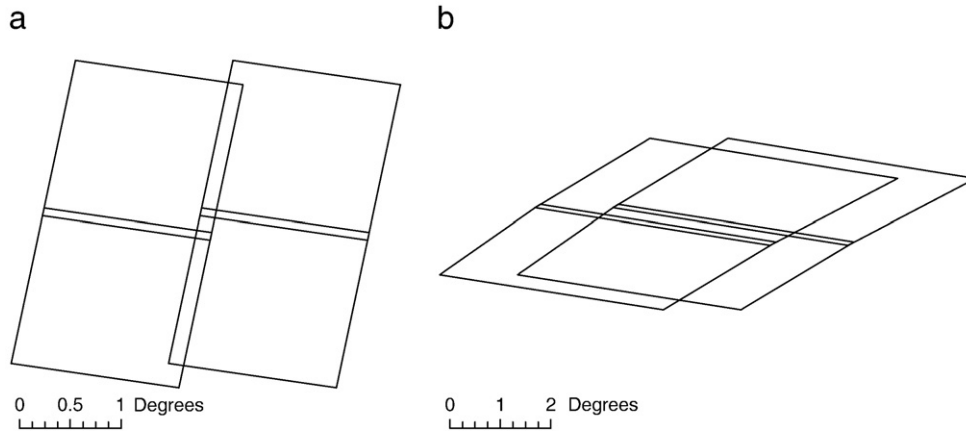
Fig. 1. Geographic locations of four Landsat 7 ETM+ acquisitions derived from their metadata latitude and longitude scene corner coordinates for two adjacent Landsat paths (left: 173, right: 172) acquired June 2010 over (a) the equator (rows 60 and 61) and (b) around latitude 70° (rows 9 and 10). The across-scan (West–East) and along-track (North–South) overlap of the adjacent scenes are clearly apparent.

overpasses in the period of interest over the land grid point, and p_i is defined as:

$$p_i = \frac{f_i}{10} + 0.05 \qquad (2)$$

where p_i is the probability of the i-th Landsat acquisition being cloudy, and f_i is the Landsat cloud fraction metadata (0,1 … 9), to give possible cloud probabilities of 0.05, 0.15, 0.25 … 0.85, 0.95 (Ju & Roy, 2008). To model the impact of the scan line corrector failure removal of 22% of the Landsat ETM+ scene pixels (Markham et al., 2004), the cloud probability for the Landsat ETM+ 2010 epoch data is defined as:

$$p_i = 0.22 + 0.78\left(\frac{f_i}{10} + 0.05\right) \qquad (3)$$

where p_i is the probability of the i-th Landsat 7 ETM+ 2010 epoch acquisition being cloudy and f_i is the cloud fraction metadata (0, 1 … 9). For mathematical tractability the scan line corrector failure missing pixels are assumed to occur in a random way across the scene and the reported cloud fraction metadata value is assumed to be representative of all the scenes. If there were no Landsat observations of the land grid point within the given period then P_{one+} is set as zero.

3.4. Probability of there being at least one cloud-free land observation of a land grid point in each of three seasons

To assess the suitability of Landsat data for global land cover mapping, an estimate that captures the probability of cloud-free land surface observation in three different seasons is developed. It is established that land cover classification algorithms have improved accuracy if multi-temporal rather than single-date satellite data are classified, particularly if the acquisition dates capture different seasonal stages of vegetation development (Defries & Townshend,

1994; Lo et al., 1986). Multi-temporal supervised classification schemes commonly exclude cloudy months or seasons as they preclude sufficiently representative training data collection (Hansen et al., 2008; Hansen et al., 2011). Accordingly, in this paper the probability of there being at least one cloud-free land observation occurring in each of the three seasons with the highest seasonal probabilities of cloud-free land surface observation is computed, defined as:

$$P_{3s+} = \prod_{k=1}^{3} P_{k,one+} \quad \text{where} \quad P_{1,one_+} \geq P_{2,one_+} \geq P_{3,one_+} \geq P_{4,one_+} \qquad (4)$$

where P_{3s+} is the probability of there being at least one cloud-free land observation of the land grid point in each of the three seasons with the highest seasonal probabilities of cloud-free land surface observation and $P_{k,\,one+}$ is probability of there being at least one cloud-free observation of the land grid point for season k, defined as:

$$P_{k,one+} = 1 - \prod_{i=1}^{n} p_i \qquad (5)$$

where $P_{k,\,one+}$ is the probability of there being at least one cloud-free observation of the land grid point in season k (Spring, Summer, Winter, or Autumn), n is the number of satellite overpasses over the land grid point for season k occurring in each epoch, and p_i is defined as [3] for the 2010 epoch Landsat 7 ETM+ (SLC-off) data and as [2] for the other Landsat data. If a location has less than three observations within four different seasons then P_{3s+} is set as zero.

3.5. Global probability calculation

Global summary statistics of P_{one+} and P_{3s+} were generated from the 6,138,864 land grid point probability values (Fig. 3). The statistics were computed independently considering the metadata describing the Landsat 5 TM acquisitions, the Landsat 7 ETM+ acquisitions, and both Landsat 5 TM and Landsat 7 ETM+ acquisitions combined.

Table 3

The mean and maximum absolute deviation in the geographic location of the upperLeftCornerLatitude and upperLeftCornerLongitude metadata values (Table 1) for Landsat 5 TM and Landsat 7 ETM+ and for the two 36 month epochs of metadata (Table 2). The tabulated deviations were computed relative to the mean upperLeftCornerLatitude and upperLeftCornerLongitude for those WRS path/row locations having 7 or more Landsat acquisitions within each epoch.

Sensor	Epoch	Mean E–W absolute deviation (km)	Maximum E–W absolute deviation (km)	Mean N–S absolute deviation (km)	Maximum N–S absolute deviation (km)	Number of Landsat acquisitions considered	Number of unique WRS path/row considered
TM 5	2000	3.05	15.56	0.77	3.81	98,956	4260
ETM+ 7	2000	1.72	7.55	0.19	3.02	220,691	9185
TM 5	2010	1.47	9.26	0.47	4.28	196,558	7145
ETM+ 7	2010	2.64	8.46	0.24	5.22	305,630	9247
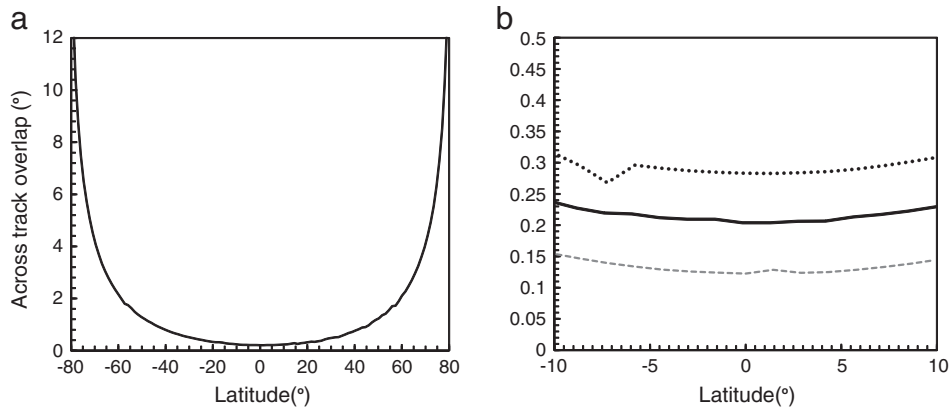
Fig. 2. Across track overlap in the geographic area sensed by Landsat 5 TM or Landsat 7 ETM+ adjacent paths, (a) mean overlap plotted against latitude (80° South to 80° North), (b) detail of minimum (dashed), mean (solid line), and maximum (dotted) overlap for equatorial latitudes (10° South to 10° North). These results derived from 96,301 globally distributed Landsat 5 TM and Landsat 7 ETM+ metadata records for twelve months 1st December 2009 to 30th November 2010. The scene corner coordinates were used to compute the longitudes of the mid-points of the scene Eastern and Western sides and then the across-track scene overlaps were derived from these longitudes for all scenes occurring in adjacent WRS paths for each WRS row.

Global summary statistics of P_{one+} were generated for each of the 36 months. Global summary statistics of P_{3s+} were generated with respect to 12, 24 and 36 month periods for each epoch (Table 2). We note that P_{one+} and P_{3s+} are computed from scene-level cloud fraction metadata that do not capture cloud variations within Landsat scenes. This may result in over- or under-estimation of P_{one+} and P_{3s+} for individual land grid points that are persistently cloudy or cloud-free compared to the majority of the grid points that fall with the same Landsat scene. This issue was also noted in the Landsat cloud study of Ju and Roy (2008).

4. Results

4.1. Global distribution of Landsat acquisitions

Fig. 4 illustrates the total number of Landsat acquisitions over each land grid point (Fig. 3) sensed during the 36 months of the 2000 epoch (left column) and the 2010 epoch (right column) for the Landsat 5 TM (top row), Landsat 7 ETM+ (middle row) and both Landsat sensors combined (bottom row). Where there were no acquisitions the land grid locations are colored white. The elliptical shapes in the Landsat 5 TM results (top row) show the X-band reception coverage of international ground stations whose TM data were copied to the USGS archive. The greater global availability of Landsat 7 ETM+ data compared to Landsat 5 TM (Table 2) in either epoch is evident and is due to differences in the Landsat acquisition strategies and

data reception capabilities described in Section 2.1. In the 2000 epoch the greatest number of acquisitions per land grid point over 36 months was 206 (Landsat 5 TM), 300 (Landsat 7 ETM+) and 421 (both sensors). In the 2010 epoch the greatest number of acquisitions per land grid point over 36 months was 248 (Landsat 5 TM), 520 (Landsat 7 ETM+) and 1040 (both sensors). All these maxima occurred in the conterminous United States. The percentage of land grid points with no acquisition over the 36 months with respect to Landsat 5 TM, Landsat 7 ETM+, and considering both sensors, was 43.7%, 1.3% and 1.0% (2000 epoch) and 18.6%, 1.1% and 1.1% (2010 epoch) respectively.

Fig. 5 shows the same results as Fig. 4 but for the single month of July selected because it was the month with the greatest number of acquisitions for either Landsat sensor in climate years 2000 and 2010. The Landsat sensors each overpass every Earth location every 16 days (i.e. no more than twice per month) but the observed number of acquisitions in any month can be greater than two because of the across-track scene overlap (Fig. 2). The maximum number of acquisitions of a land grid point in July was 9 for the Landsat 5 TM, over Western California, and 20 for the Landsat 7 ETM+, over Graham Island near Greenland.

Fig. 6 shows the monthly global mean (top row) and standard deviation (bottom row) of the number of Landsat acquisitions per land grid point location for the 2000 epoch (left column) and 2010 epoch (right column). The impact of any scan line corrector issues is not considered. There were a greater number of acquisitions in the 2010 epoch compared to the 2000 epoch, and more Landsat 7

Fig. 3. Global land grid defined by 9233 unique non-Antarctic land Landsat WRS path/rows defined the U.S. Landsat project "Land Definition" data base and from the geographic corner coordinates for these path/rows extracted from the 2000 and 2010 epoch Landsat TM and ETM+ metadata. The grid is defined in the sinusoidal equal area projection and is composed of 6,138,864 land grid points spaced every 5.559752 km in the X and Y axes of the sinusoidal coordinate system.
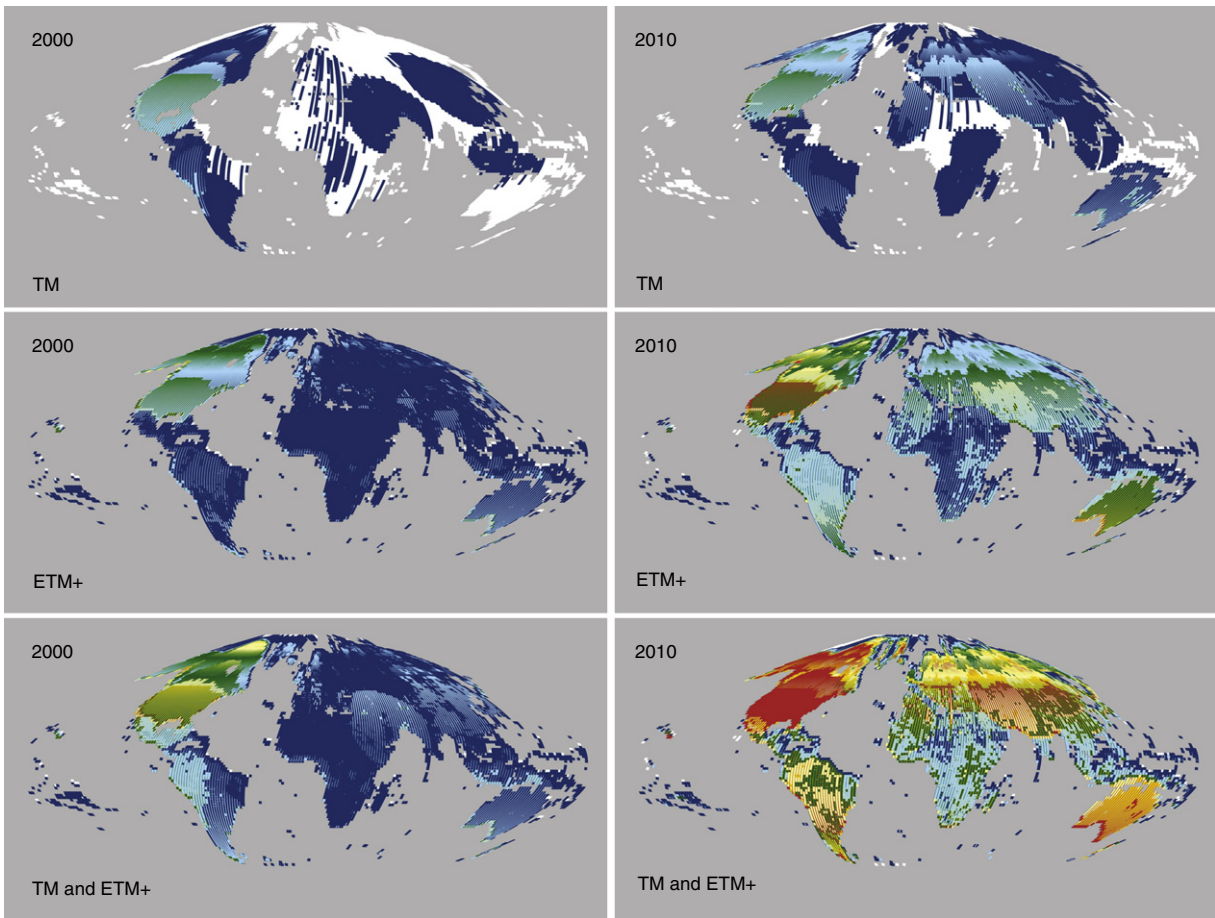
Fig. 4. Total number of Landsat acquisitions over each land grid (Fig. 3) sensed for the 36 months of the 2000 epoch (left column) and the 2010 epoch (right column) by Landsat 5 TM (top row), Landsat 7 ETM+ (middle row), and by both Landsat 5 TM and Landsat 7 ETM+ (bottom row). Number of acquisitions colored as: 0 = White, 1 ≤ Dark Blue ≤ 50, 51 ≤ Light Blue ≤ 100, 101 ≤ Green ≤ 150, 151 ≤ Yellow ≤ 200, 201 ≤ Orange ≤ 250, 251 ≤ Red < 1040. Grid points not defined as land (Fig. 3) are shown as gray. The impact of any scan line corrector issues is not considered.

ETM+ than Landsat 5 TM acquisitions in either epoch (Table 2, Fig. 4) and this pattern is apparent in the mean monthly values illustrated in the top row of Fig. 6. The impact of the Landsat 7 ETM+ Long-Term Acquisition Plan is evident in the seasonal variation in the number of Landsat acquisitions, with relatively fewer acquisitions in the northern hemisphere winter months than in the summer months. There were no Landsat 7 ETM+ acquisitions before 29th June 1999 because the sensor had only just begun operating. Disregarding Landsat ETM+ before July 1999, the average 36 monthly global mean number of Landsat 5 TM and Landsat 7 ETM+ acquisitions per land grid point are 0.42 and 1.23 (2000 epoch) and 0.87 and 1.39 (2010 epoch) respectively. When the sensors are combined the average 36 monthly global mean number of acquisitions per land grid point are 1.41 (2000 epoch) and 2.26 (2010 epoch). In certain months the mean number of Landsat acquisitions per land grid point for the combined sensors is nearly 2.2 and 8.2 times greater than for Landsat 7 ETM+ and Landsat 5 TM alone respectively. The monthly standard deviation number of acquisitions illustrated in Fig. 6 (bottom row) is correlated and typically greater than the monthly mean values.

Fig. 7 shows the monthly percentage of the global land grid point locations that had one or more Landsat acquisitions. In the 2000 epoch the monthly percentage of land grid points (disregarding Landsat ETM+ before July 1999) with one or more Landsat acquisitions varied from 9.0% to 31.1% (Landsat 5 TM), from 40.7% to 67.5% (Landsat 7 ETM+), and from 50.1% to 75.2% when both sensors are combined. In the 2010 epoch, the monthly percentage of land grid points with one or more Landsat acquisitions was higher than in

2000 and varied from 13.5% to 60.2% (Landsat 5 TM), from 47.0% to 73.6% (Landsat 7 ETM+), and from 53.6% to 79.6% when both sensors are combined. Combination of the sensors provides in each month a 1.7% to 14.4% higher percentage of acquired land grid point locations compared to Landsat 7 ETM+ data alone.

4.2. Probabilities of there being at least one cloud-free observation of each land grid point

The above results illustrate that combination of Landsat 5 TM and Landsat 7 ETM+ data will provide a greater number and more complete global Landsat land coverage. However, the degree of cloud cover is important as clouds generally preclude surface observation (Irish et al., 2006) and hence reduce the number of useful land surface observations (Ju & Roy, 2008). Fig. 8 shows the global monthly mean cloud cover percentage computed over all land grid points that had a Landsat acquisition. The global mean monthly cloud cover considering both epochs and both sensors combined is about 35% but with considerable among month variability. In the 2010 epoch the mean monthly Landsat 7 ETM+ cloud cover is lower than for Landsat 5 TM. This is expected because the Landsat 7 ETM+ Long-Term Acquisition Plan (LTAP) attempts to acquire substantially cloud-free acquisitions globally (Arvidson et al., 2006) and the Landsat 5 TM data are not acquired following an LTAP. In the 2000 epoch this difference is less apparent, this is not a failure of the LTAP but occurs because the Landsat 7 ETM+ data acquired much more of the globe than Landsat 5 TM in the 2000 epoch (Fig. 4, left column), including regions that are persistently cloudy at the time of Landsat overpass, such as Equatorial
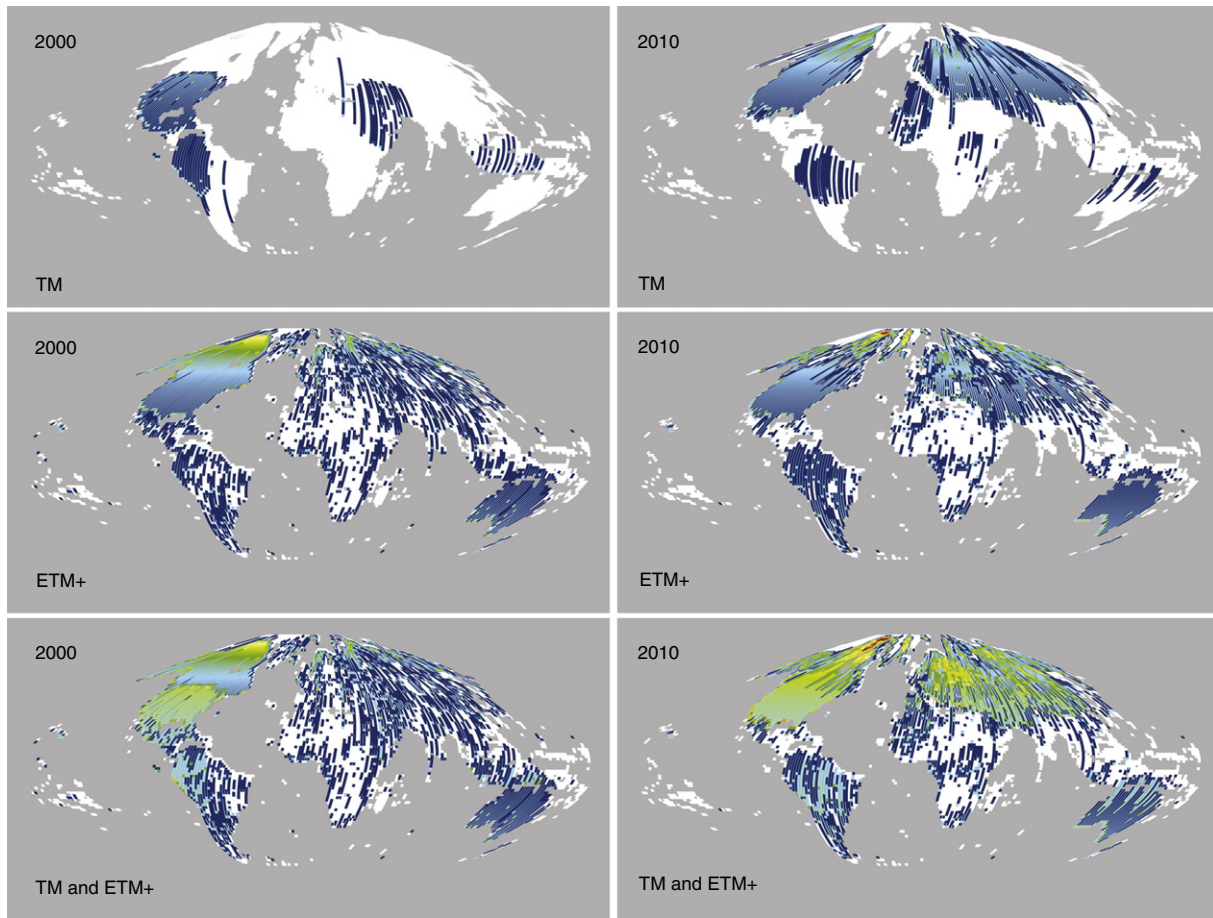
Fig. 5. Total number of Landsat acquisitions over each land grid (Fig. 3) sensed in July 2000 (left column) and July 2010 (right column) by Landsat 5 TM (top row), Landsat 7 ETM + (middle row), and by both Landsat 5 TM and Landsat 7 ETM + (bottom row). Number of acquisitions colored as: 0 = White, 1 ≤ Dark Blue ≤ 2, 3 ≤ Light Blue ≤ 4, 5 ≤ Green ≤ 6, 7 ≤ Light Green ≤ 8, 9 ≤ Yellow ≤ 10, 11 ≤ Orange ≤ 12, 13 ≤ Red < 20. Grid points not defined as land (Fig. 3) are shown as gray. The impact of any scan line corrector issues is not considered.

Africa, Amazonia, and northern boreal regions (Ju & Roy, 2008). We note that the Landsat 5 TM cloud fraction metadata for December 2008 and January 2009 were defined using the original L5 ACCA algorithm (Su, 1984) but were defined for the remaining 34 months of the 2010 epoch (February 2009 to November 2011) using an algorithm based on the L7 ETM + ACCA algorithm (Irish et al., 2006). This, however, does not cause any temporal inconsistency in the global monthly Landsat 5 TM cloud statistics at the beginning of the 2010 epoch (Fig. 8, right, filled circles).

Fig. 9 shows the probability of there being at least one cloud-free observation of each land grid point for July modeling cloud cover and the Landsat 7 ETM + scan line corrector failure. The probabilities are derived from Eqs. (1) to (3). In 2000 (Fig. 9, left column) the magnitude of the probabilities is broadly similar between Landsat 5 TM and Landsat 7 ETM + at the WRS path/rows acquired by both sensors. This is expected as there is no reason why the two Landsat sensors, which have the same overpass time but sense the same location on different days, would capture different cloud conditions when considered over large areas and time periods. In 2010 (Fig. 9, right column) the magnitude of the Landsat 7 ETM + probabilities is generally lower than Landsat 5 TM at WRS path/rows acquired by both sensors because of the impact of the Landsat ETM + scan line corrector failure (Eq. 3). For either year, combination of the July Landsat 5 TM and Landsat 7 ETM + data provides higher probabilities than considering one sensor alone.

Fig. 10 shows the monthly global mean (top row) and standard deviation (bottom row) of the probability of there being at least one cloud-free observation of each land grid point for the 2000 epoch (left column) and 2010 epoch (right column). The temporal variation in the mean global monthly probabilities is correlated with the monthly number of acquisitions and monthly mean cloud covers illustrated in Figs. 7 and 8. Disregarding Landsat ETM + before July 1999, the average 36 monthly mean Landsat 5 TM probability is 0.15 (2000 epoch) and 0.30 (2010 epoch) and the average 36 monthly mean Landsat 7 ETM + probability is 0.47 (2000 epoch) and 0.45 (2010 epoch). The Landsat 7 + ETM monthly mean probabilities are greater than for Landsat 5 TM in both epochs because of the much greater number of Landsat 7 ETM + acquisitions (Table 2). The average 36 monthly mean Landsat 7 ETM + probability in the 2000 epoch (0.47) is slightly lower than in the 2010 epoch (0.45) despite the nearly 50% more Landsat 7 ETM + observations in the 2010 epoch (Table 2). This is because of the global impact of the Landsat ETM + scan line corrector failure. Combination of sensors provides the highest average 36 monthly mean probabilities of 0.52 (2000 epoch) and 0.55 (2010 epoch). In certain months the mean monthly probability for the combined sensors is nearly 1.4 and 6.7 times greater than for Landsat 7 ETM + and Landsat 5 TM alone respectively. The monthly global standard deviation probabilities (Fig. 10, bottom row) are relatively temporally consistent, implying that the impact of between month differences in the locations of Landsat acquisitions and cloudiness are obscured at global scale. The standard deviation probabilities may be more temporally consistent for Landsat 7 ETM + than Landsat 5 TM because of the Long-Term Acquisition Plan that reduces Landsat 7 ETM + cloud seasonality (Fig. 8). The Landsat 5 TM standard deviation probabilities are consistently lower than for Landsat 7 ETM + in the 2000 epoch because the Landsat 5 TM acquired much less of the globe (Fig. 4, left column).
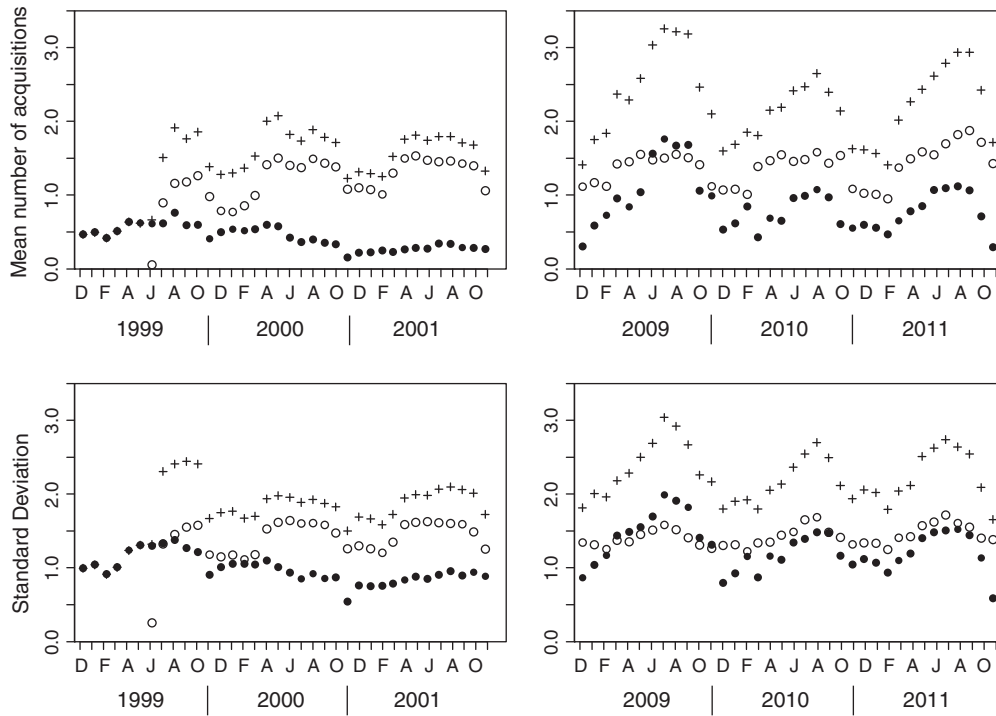
Fig. 6. Summary monthly global statistics (top row: mean, bottom row: standard deviation) of the number of land grid acquisitions sensed in the 36 months of the 2000 (left column) and 2010 (right column) epochs. Statistics computed with respect to all the 6,138,864 land grid points (Fig. 3) for Landsat 5 TM (filled circles), Landsat 7 ETM+ (open circles), both Landsat 5 TM and Landsat 7 ETM+ (crosses). The impact of any scan line corrector issues is not considered. There were no Landsat 7 ETM+ acquisitions before 29th June 1999.

4.3. Probabilities of there being at least one cloud-free land observation of a land grid point in each of three seasons

Figs. 11 and 12 show the probability of there being at least one cloud-free land observation occurring in each of the three seasons with the highest seasonal probabilities of cloud-free land surface observation (P_{3s+}). The figures were generated considering 12 months (Fig. 11) and 36 months (Fig. 12) of Landsat acquisitions. They reflect a conservative estimate of there being sufficient temporal Landsat data for multi-temporal land cover classification. The white areas show zero valued probabilities and occur where there was not at least one acquisition in any three seasons (December, January and February; March, April, May; June, July, August; or September, October, November). The impact of the limited global coverage of the Landsat 5 TM (Fig. 4) is evident in the high proportion of zero valued probabilities in the top rows of Figs. 11 and 12. Globally, lower probabilities occur at locations with persistent cloud cover at the time of Landsat overpass, including Equatorial Africa, Amazonia, northern boreal regions and much of South East Asia. These locations have been observed to be challenging to classify using Landsat data due to cloud contamination

(Broich et al., 2011; Hansen et al., 2008; Lindquist et al., 2008; Souza, 2006) particularly in high latitude Boreal regions with short growing seasons (Potapov et al., 2008). The probabilities computed using only 12 months of data (Fig. 11) are lower than those computed using 36 months of data (Fig. 12) because of the smaller availability of Landsat observations and so reduced likelihood of obtaining a cloud-free Landsat observation.

Table 4 summarizes the global mean and standard deviation of P_{3s+} to provide a conservative summary of there being sufficient temporal Landsat data for multi-temporal land cover classification. The probabilities are shown for 12, 24 and 36 month epochs encompassing climate years 2000 and 2010. The global mean probabilities increase when longer duration epochs are considered and when both Landsat sensor data are combined. The mean global probabilities considering Landsat 5 TM alone are less than 0.35 (12 months) and less than 0.6 (36 months). The global mean probabilities considering only Landsat ETM+ are 0.74, 0.89 and 0.90 in 2000 and 0.62, 0.81, 0.87 in 2010 for 12, 24 and 36 months respectively. The Landsat 7 ETM+ probabilities are higher than for Landsat 5 TM because of the greater global Landsat 7 ETM+ acquisition as
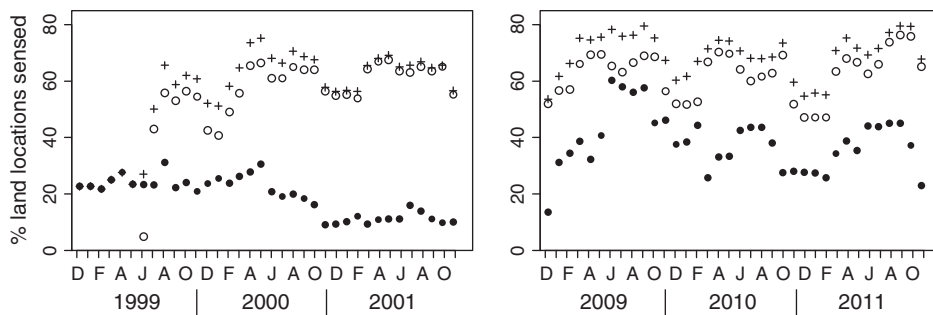
Fig. 7. Global monthly percentage of the 6,138,864 land grid points (Fig. 3) with a Landsat 5 TM (filled circles), Landsat 7 ETM+ (open circles), and both Landsat 5 TM and Landsat 7 ETM+ (crosses) acquisitions sensed in the 36 months of the 2000 (left column) and 2010 (right column) epochs. The impact of any scan line corrector issues is not considered. There were no Landsat 7 ETM+ acquisitions before 29th June 1999.
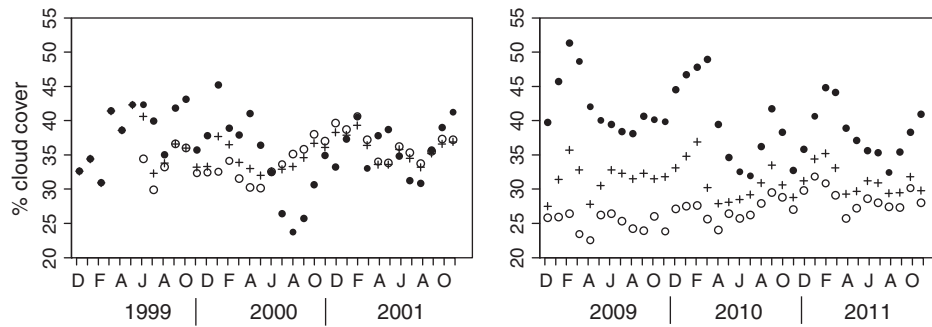
Fig. 8. Global monthly mean cloud cover percentage computed over the land grid points that had a Landsat acquisition in each of the 36 months of the 2000 (left column) and 2010 (right column) epochs. Results are shown for Landsat 5 TM (filled circles), Landsat 7 ETM+ (open circles), and both Landsat 5 TM and Landsat 7 ETM+ (crosses). The impact of any scan line corrector issues is not considered. There were no Landsat 7 ETM+ acquisitions before 29th June 1999.

discussed above. However, when the data from the Landsat sensors are combined the mean global probabilities are only about one decimal place greater than considering Landsat ETM+ data alone. Considering combined sensors and 36 months of data the global mean probability is 0.92 (σ 0.24) in 2000 and 0.90 (σ 0.28) in 2010. The Landsat 7 ETM+ and combined sensor mean probabilities are slightly smaller in the 2010 epoch than in the 2000 epoch because of the deleterious impact of the Landsat 7 ETM+ scan line corrector failure, which was observed in Fig. 10. These results indicate that 36 months of combined Landsat sensor data may provide sufficient data for global land cover mapping, although there are regional differences in the probabilities as noted above for Figs. 11 and 12.

Table 5 summarizes the percentage of the global land grid points with different P_{3s+} probability ranges. A greater percentage of the global land surface has higher probability of there being sufficient temporal Landsat data for multi-temporal land cover classification when longer duration epochs are considered and when both Landsat sensor data are combined. Some 56.4%, 81.0% and 86.4% of the global land grid points have probabilities ≥0.95 (red in Figs. 11 and 12) for combined sensors in 2000 considering 12, 24 and 36 months respectively. In 2010 these percentages are slightly smaller, 48.1%, 76.7% and 84.2% respectively, because of the scan line corrector issue discussed above. Combination of the Landsat sensor data provides the most benefit over using only one Landsat sensor for the shorter
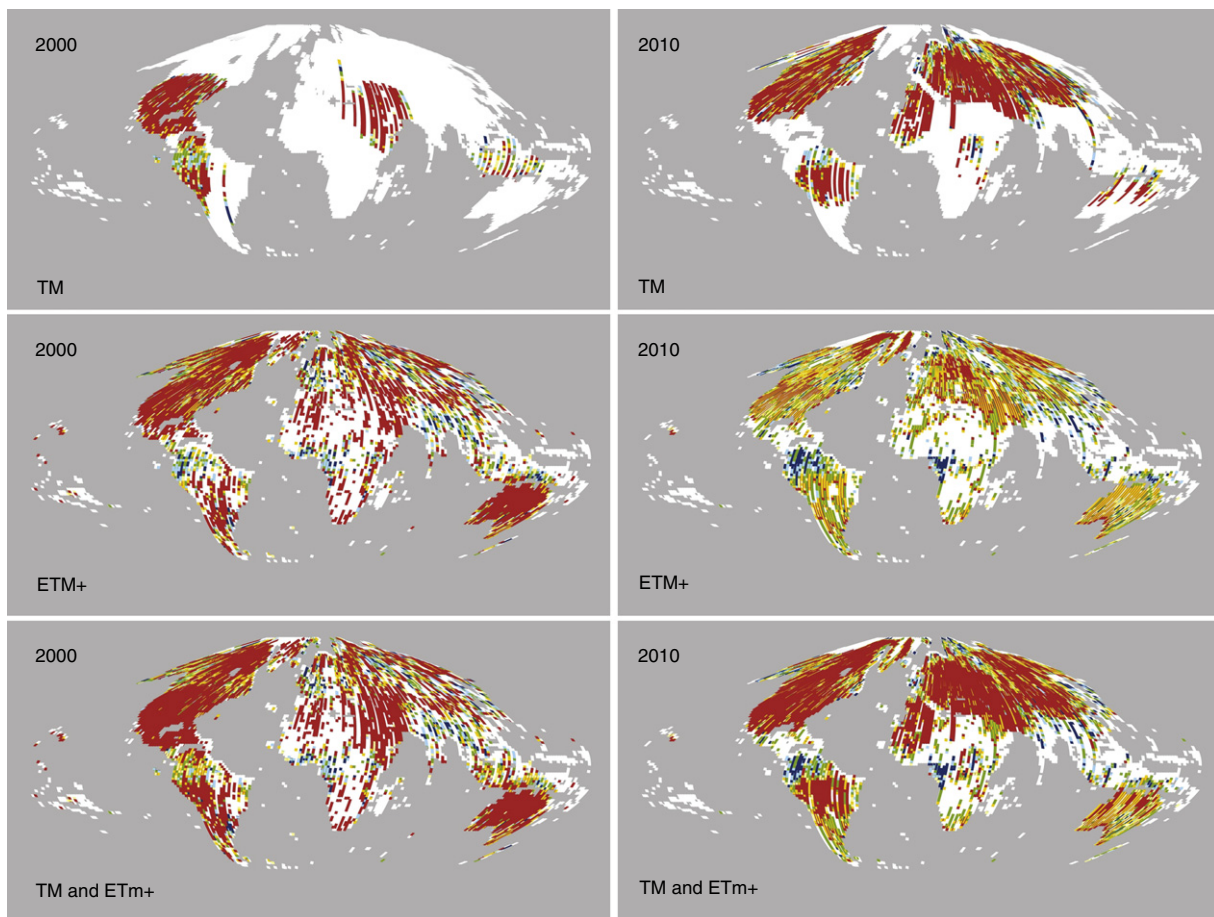
Fig. 9. Probability of there being at least one cloud-free observation of each land grid point (Eq. (1)) for Landsat acquisitions sensed in July 2000 (left column) and July 2010 (right column) by Landsat 5 TM (top row), Landsat 7 ETM+ (middle row), and by both Landsat 5 TM and Landsat 7 ETM+ (bottom row). Probabilities colored as 0=White, 0<Dark Blue<0.25, 0.25≤Light Blue<0.5, 0.5≤Green<0.75, 0.75≤Yellow <0.85, 0.85≤Orange<0.95, 0.95=<Red=<1. Grid points not defined as land (Fig. 3) are shown as gray.
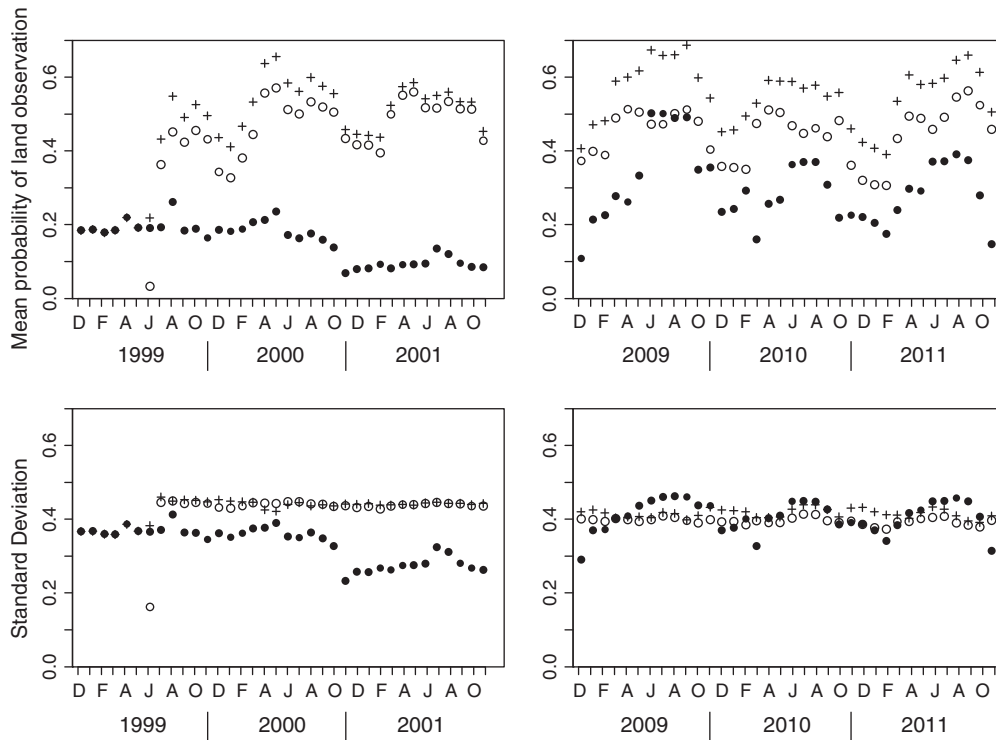
Fig. 10. Summary monthly global statistics of the probability of there being at least one cloud-free observation of each land grid point (Eq. (1)) (top row: mean, bottom row: standard deviation) sensed for the 36 months of the 2000 (left column) and 2010 (right column) epochs. Results are computed with respect to all the 6,138,864 land grid points (Fig. 3) for Landsat 5 TM (filled circles), Landsat 7 ETM + (open circles), and both Landsat 5 TM and Landsat 7 ETM + (crosses). The impact of any scan line corrector issues is considered. There were no Landsat 7 ETM + acquisitions before 29th June 1999.

duration epochs, even so, for the 36 month duration epochs there are 4.7% and 6.2% more global land grid points with probabilities ≥ 0.95 when the sensor data are combined compared to using Landsat 7 ETM + alone. The percentage of the global land grid points with $P_{3s+} = 0$ (white in Figs. 11 and 12) for the combined sensors are 12.9%, 6.0% and 5.3% for 12, 24, and 36 months in 2000, and 16.0%, 9.8% and 7.7% for 12, 24, and 36 months in 2010. That is using longer duration epochs and both Landsat sensor data ensures more of the globe can be classified. However, even using 36 months of data and combined Landsat sensors, 5.3% and 7.7% of the global land surface in 2000 and 2010 respectively had insufficient Landsat observations for multi-temporal land cover classification. The results illustrated in Figs. 11 and 12, and summarized in Tables 2 and 3, underscore the need for combined sensor multi-annual data sets for global land cover classification.

5. Conclusions

Global satellite land products have been generated systematically at coarse spatial resolution (100–1000 m) using MODIS and AVHRR data (Justice et al., 1998; Townshend et al., 1994; Tucker et al., 2005). The 2008 free Landsat data distribution policy opens a new era for utilizing Landsat data (Woodcock et al., 2008; Wulder et al., 2012). With the advent of free data it becomes feasible to process long term and/or large-area Landsat data sets to provide a medium spatial resolution (30 m) analog to the coarse spatial resolution land products generated from the MODIS and AVHRR data streams (Roy et al., 2010a). Prior to the free Landsat policy, global scale Landsat products have not been demonstrated widely due to high data cost, data availability and data processing constraints (Gutman et al., 2008). In addition, as with any optical wavelength satellite sensor, cloud contamination compromises global Landsat image usability for land surface studies (Ju & Roy, 2008). Further, the spatio-temporal availability of Landsat data in the USGS Landsat archive is dependent

on the Landsat sensor and orbit geometry and is reduced by selective Landsat scene acquisition due to payload, ground station and mission cost constraints, data reception capabilities and Landsat system health issues (Arvidson et al., 2006; Loveland & Dwyer, 2012; Markham et al., 2004).

In this paper the global probabilities of acquiring a cloud-free Landsat land surface observation in each month and in three different seasons were reported to assess the availability of the Landsat data in the USGS Landsat archive for global multi-temporal remote sensing applications and land cover mapping respectively. The global probabilities of acquiring a cloud-free Landsat land surface observation were reported for Landsat 5 TM and Landsat 7 ETM + independently and for both sensors together for 36 monthly epochs centered on 2000 and 2010. The geographic scene overlap imposed by the Landsat sensors and orbit geometries was quantified and considered explicitly in the analysis. The across-track overlap distance between adjacent Landsat paths varied from a minimum of 0.105° at the Equator to more than 12.5° at 80° latitude. At high latitude the overlap is significant and is equivalent to more than 90% of a Landsat scene, and raises the possibility for novel high-latitude Landsat studies to take advantage of the high temporal observation coverage. This increasing poleward overlap also suggests caution in the interpretations of large area multiple orbit Landsat cloud metadata that do not consider the scene overlap.

The results reported in this paper demonstrate the utility of combining both Landsat 5 TM and Landsat 7 ETM + data streams to take advantage of their different acquisition patterns and to mitigate the deleterious impact of the Landsat 7 ETM + 2003 scan line failure. Of the approximately 845,000 Landsat 5 TM and Landsat 7 ETM + metadata available in the USGS Landsat archive considered in this study, the greater fraction (more than twice and one third in the 2000 and 2010 36 month epochs) were Landsat 7 ETM + compared to Landsat 5. This was due to improved US data recording and ground system capacity and systematic acquisition policy for the Landsat 7 ETM + (Loveland & Dwyer, 2012). The average monthly global mean number
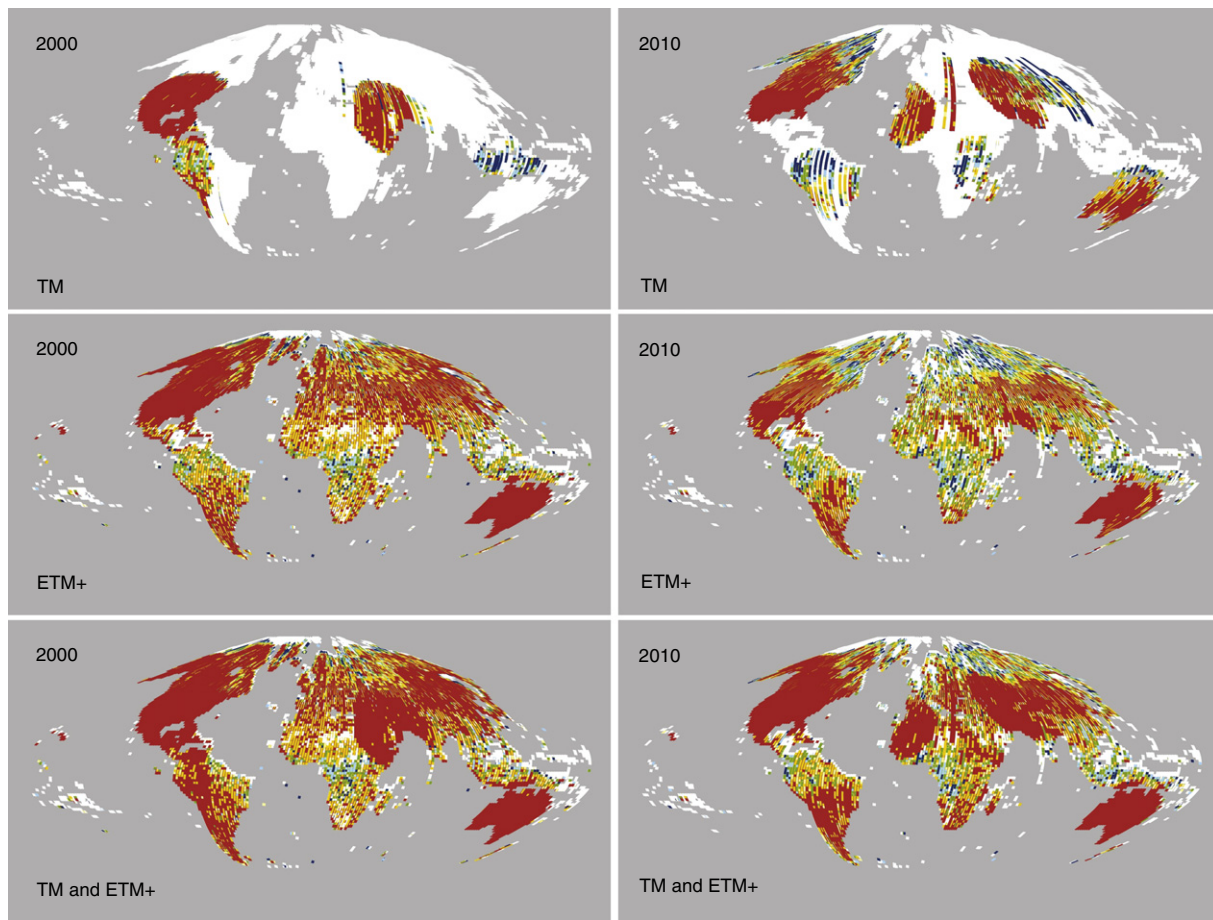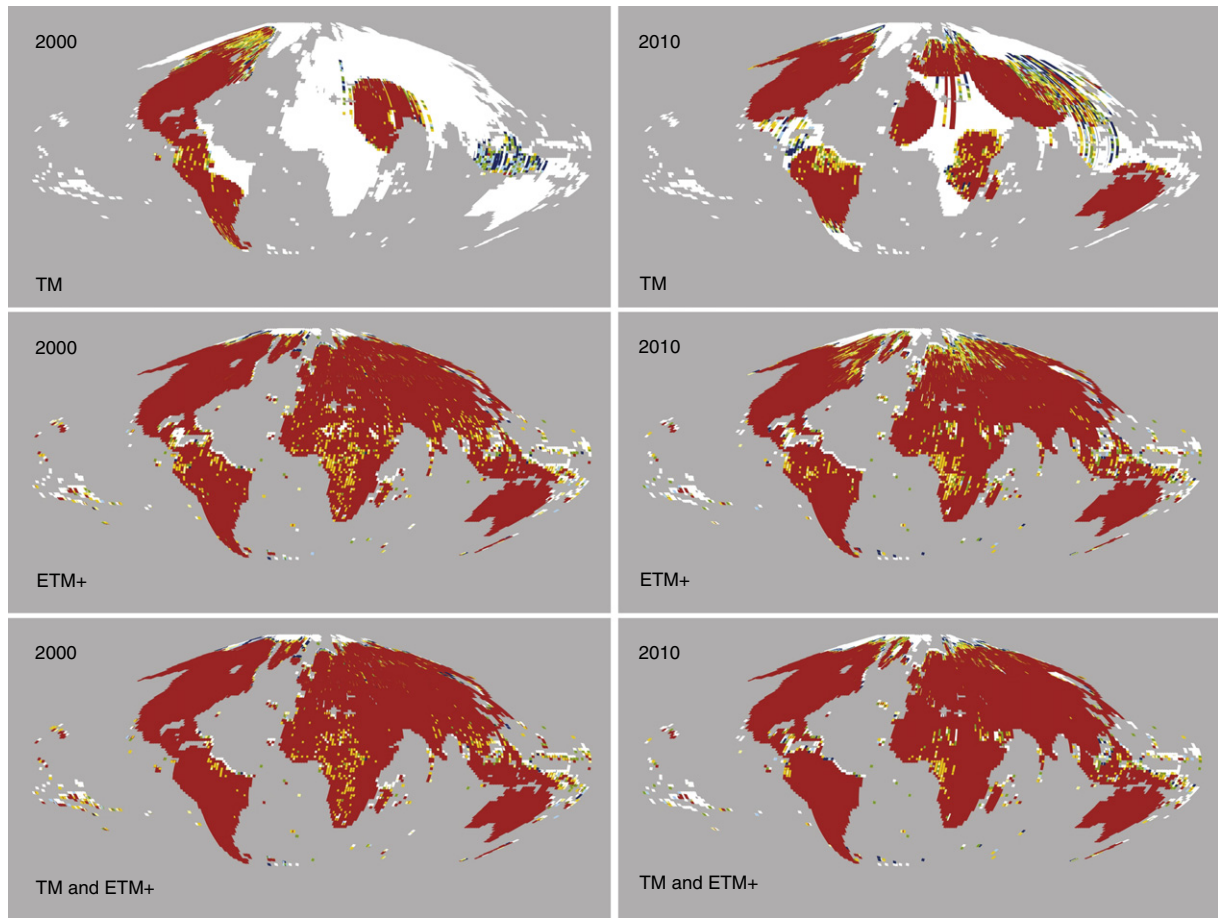
Fig. 11. Probability of there being at least one cloud-free land observation occurring in each of the three seasons with the highest seasonal probabilities of cloud-free land surface observation (Eq. (4)) for the 12 months of climate year 2000 (left column) and the 12 months of climate year 2010 (right column) considering Landsat 5 TM (top row), Landsat 7 ETM+ (middle row), and both Landsat 5 TM and Landsat 7 ETM+ (bottom row). Probabilities colored as 0=White, 0<Dark Blue<0.25, 0.25≤Light Blue<0.5, 0.5≤Green<0.75, 0.75≤Yellow<0.85, 0.85≤Orange<0.95, 0.95=<Red=<1. Grid points not defined as land (Fig. 3) are shown as gray.

of Landsat 5 TM and Landsat 7 ETM+ acquisitions per land location were 0.42 and 1.23 (2000 epoch) and 0.87 and 1.39 (2010 epoch) respectively. The average monthly global mean number of land acquisitions for combined Landsat 5 TM and Landsat 7 ETM+ sensing was 1.41 (2000 epoch) and 2.26 (2010 epoch). In addition to providing a higher global mean monthly number of acquisitions, the sensor combination provided a greater acquisition coverage with 1.7% to 14.4% higher percentage of global land locations acquired monthly compared to considering Landsat 7 ETM+ data alone.

The probability of there being at least one cloud-free observation of each global land location was computed modeling cloud effects and the Landsat ETM+ scan line corrector failure. The global mean monthly cloud cover considering both epochs and both Landsat sensors were approximately 35% with considerable among month variability reflecting the number and cloudiness of the acquired Landsat data. In the 2000 epoch the average 36 monthly mean probability of there being at least one cloud-free observation of each land location was 0.15, 0.47 and 0.52 for Landsat 5 TM, Landsat 7 ETM+ and both Landsat sensors combined respectively. In the 2010 epoch the average 36 monthly mean probabilities were 0.30, 0.45 and 0.55 for Landsat 5 TM, Landsat 7 ETM+ and both sensors combined. In certain months the mean monthly probability for the combined sensors was nearly 1.4 and 6.7 times greater than for Landsat 7 ETM+ and Landsat 5 TM alone respectively. These differences are not insignificant and illustrate the utility of combined Landsat sensor data products.

To date, no global coverage Landsat land cover product has been generated although Landsat sample-based studies illustrate the

potential for global mapping (Beuchle et al., 2011; Ridder, 2007) and a number of regional to near-continental scale Landsat land cover products have been developed to take advantage of the improved spatial resolution, and so reduced among vegetation class spectral confusion, of Landsat compared to data from coarser spatial resolution global polar orbiting systems (Hansen & Loveland, 2012; Herold et al., 2008). In this paper the probability of acquiring a cloud-free Landsat land surface observation in different seasons was derived to provide a conservative estimate of there being sufficient data in the USGS Landsat archive for global land cover mapping using a multi-temporal supervised classification scheme. Specifically, the probability of acquiring a cloud-free Landsat land surface observation in each of three different seasons with the highest seasonal probabilities of cloud-free land surface observation was computed. The probabilities were derived considering 12, 24 and 36 month epochs encompassing 2000 and 2010. Higher probabilities were found when longer duration epochs were considered and when both Landsat sensor data were combined. Combination of the Landsat sensor data provided the most benefit for the shorter duration epochs, even so, for the 36 month epochs there were 4.7% and 6.2% more global land locations with probabilities ≥0.95 when the sensor data were combined compared to using Landsat 7 ETM+ alone. Locations with persistent cloud cover at the time of Landsat overpass, including Equatorial Africa, Amazonia, northern boreal regions and South East Asia, had the lowest probabilities and have been observed to be challenging to classify with Landsat data by other researchers. In particular, northern boreal regions have both persistent cloud and a short

Fig. 12. Probability of there being at least one cloud-free land observation occurring in each of the three seasons with the highest seasonal probabilities of cloud-free land surface observation (Eq. (4)) for the 36 months of the 2000 (left column) and for the 36 months of the 2010 (right column) epochs considering Landsat 5 TM (top row), Landsat 7 ETM+ (middle row), and both Landsat 5 TM and Landsat 7 ETM+ (bottom row). Probabilities colored as 0=White, 0<Dark Blue<0.25, 0.25≤Light Blue<0.5, 0.5≤Green<0.75, 0.75≤Yellow<0.85, 0.85≤Orange<0.95, 0.95=<Red=<1. Grid points not defined as land (Fig. 3) are shown as gray.

Table 4

Summary global statistics computed with respect to the global land points (Fig. 3) of the probability of their being sufficient temporal Landsat data for multi-temporal land cover classification (P_{3s+}). Results shown for 12, 24 and 36 monthly periods encompassing 2000 and 2010 for Landsat 5 TM, Landsat 7 ETM+, and both sensors combined. P_{3s+} is defined from Eq. (4) as the probability of there being at least one cloud-free land observation of the land grid point in each of the three seasons with the highest seasonal probabilities of cloud-free land surface observation. The 12 and 36 month results are illustrated in a spatially explicit way in Figs. 11 and 12 respectively.

	TM	ETM+	TM & ETM+	TM	ETM+	TM & ETM+
	12 months 2000			12 months 2010		
Mean probability	0.194	0.742	0.790	0.332	0.621	0.727
Probability standard deviation	0.378	0.370	0.349	0.435	0.383	0.374
	24 months 2000			24 months 2010		
Mean probability	0.250	0.890	0.908	0.524	0.808	0.861
Probability standard deviation	0.415	0.273	0.255	0.471	0.338	0.313
	36 months 2000			36 months 2010		
Mean probability	0.315	0.895	0.923	0.577	0.869	0.895
Probability standard deviation	0.450	0.281	0.242	0.471	0.305	0.283

Table 5

The percentage of the global land points (Fig. 3) with different probability ranges that quantify the likelihood of their being sufficient temporal Landsat data for multi-temporal land cover classification (P_{3s+}). Results shown for 12, 24 and 36 monthly periods encompassing 2000 and 2010 for Landsat 5 TM, Landsat 7 ETM+, and both sensors combined. P_{3s+} is defined from Eq. (4) as the probability of there being at least one cloud-free land observation of the land grid point in each of the three seasons with the highest seasonal probabilities of cloud-free land surface observation. The 12 and 36 month results are illustrated in a spatially explicit way in Figs. 11 and 12 respectively.

Probability range	TM	ETM+	TM & ETM+	TM	ETM+	TM & ETM+
	12 months 2010			12 months 2010		
$P_{3s+}=0$	77.1	15.9	12.9	57.3	20.0	16.0
$0<P_{3s+}<0.5$	3.1	5.5	4.6	8.6	12.5	7.3
$0.5≤P_{3s+}<0.75$	1.9	7.7	5.8	4.2	15.2	9.3
$0.75≤P_{3s+}<0.95$	3.9	25.3	20.2	9.0	24.7	19.2
$0.95≤P_{3s+}≤1$	14.0	45.6	56.4	20.9	27.6	48.1
	24 months 2000			24 months 2010		
$P_{3s+}=0$	71.1	7.1	6.0	41.0	12.4	9.8
$0<P_{3s+}<0.5$	3.5	1.8	1.6	5.0	3.2	2.5
$0.5≤P_{3s+}<0.75$	1.9	2.7	2.2	4.0	5.8	3.0
$0.75≤P_{3s+}<0.95$	4.6	12.1	9.2	8.9	17.3	8.0
$0.95≤P_{3s+}≤1$	18.9	76.2	81.0	41.0	61.3	76.7
	36 months 2010			36 months 2010		
$P_{3s+}=0$	65.1	7.8	5.3	36.7	9.4	7.7
$0<P_{3s+}<0.5$	2.9	1.4	1.5	4.4	2.1	1.9
$0.5≤P_{3s+}<0.75$	1.8	1.5	1.4	2.9	2.7	2.1
$0.75≤P_{3s+}<0.95$	3.9	7.6	5.5	7.3	7.9	4.2
$0.95≤P_{3s+}≤1$	26.3	81.7	86.4	48.7	78.0	84.2

growing season that make land cover mapping difficult due to infrequent Landsat surface observation. Considering 36 months of Landsat ETM + data alone, 81.7% and 78.0% of the global land locations had probabilities ≥0.95 in 2000 and 2010 respectively, with a global mean probability of 0.89 (σ 0.2) in 2000 and 0.87 (σ 0.30) in 2010. Considering combined Landsat sensors and 36 months of data, 86.4% and 84.2% of the global land locations had probabilities ≥0.95 in 2000 and 2010 respectively, with a global mean probability of 0.92 (σ 0.24) in 2000 and 0.90 (σ 0.28) in 2010. These results indicate that 36 months of Landsat 5 TM and Landsat 7 ETM provide sufficient data for global land cover mapping.

Arguably, there are currently few global applications that require systematically derived global Landsat scale data products, other than land cover classification and land cover change applications (Gutman et al., 2008). However, future requirements for global mapping, monitoring, assimilation and prediction of land surface characteristics at Landsat resolution seem likely provided that computational efficiencies can be implemented (Lewis et al., 2012; Nemani et al., 2009; Running et al., 2009; Sleeter et al., 2012). The Landsat archive has largely been under-utilized but we suggest that the combination of the free Landsat data policy, demonstrated bulk Landsat processing capabilities (Roy et al., 2010a) and rapid improvements in data processing and storage capabilities, posit the future development of global Landsat products. If such products can be provided with systematic and consistent processing, and with formats and data distribution systems that make the products easy to obtain and apply, then, like global coarse spatial resolution products, they will be used for local to continental scale applications, in addition to global applications.

Acknowledgments

References

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., Gasch, J., & Williams, D. (2006). Landsat-7 long-term acquisition plan: Development and validation. *Photogrammetric Engineering and Remote Sensing, 72*, 1137–1146.

Beuchle, R., Eva, H. D., Stibig, H. -J., Bodart, C., Brink, A., Mayaux, P., Johansson, D., Achard, F., & Belward, A. (2011). A satellite data set for tropical forest area change assessment. *International Journal of Remote Sensing, 32*, 7009–7031.

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

Broich, M., Hansen, M., Stolle, F., Potapov, P., Margono, B. A., & Adusei, B. (2011). Remotely sensed forest cover loss shows high spatial and temporal variation across Sumatera and Kalimantan, Indonesia 2000–2008. *Environmental Research Letters, 6*, 014010.

Chander, G., Helder, D. L., Malla, R., Micijevic, E., & Mettler, C. J. (2007). Consistency of L4 TM absolute calibration with respect to the L5 TM sensor based on near-simultaneous image acquisition. In J. Butler, & J. Xiong (Eds.), *Proceedings of SPIE Conference on Earth Observing Systems XII in San Diego, CA. 6677* (pp. 66770F).

Defries, R. S., & Townshend, J. R. G. (1994). NDVI-derived land cover classifications at a global scale. *International Journal of Remote Sensing, 15*, 3567–3586.

Goward, S., Arvidson, T., Williams, D., Faundeen, J., Irons, J., & Franks, S. (2006). Historical record of Landsat global coverage: Mission operations, NSLRSDA, and international cooperator stations. *Photogrammetric Engineering and Remote Sensing, 72*, 1155.

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

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

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

Hansen, M. C., & Loveland, T. R. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment, 122*, 66–74.

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 in the Congo Basin. *Remote Sensing of Environment, 112*, 2495–2513.

Herold, M., Mayaux, P., Woodcock, C. E., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment, 112*, 2538–2556.

Irish, R. R. (2000). Landsat 7 automatic cloud cover assessment. *International Society for Optics and Photonics (SPIE), 4049*, 348–355, http://dx.doi.org/10.1117/12.410358.

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

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

Justice, C. O., Vermote, E., Townshend, J. R. G., Defries, R., Roy, D. P., Hall, D. K., et al. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing, 36*, 1228–1249.

Lewis, P., Gómez-Dans, J., Kaminski, T., Settle, J., Quaife, T., Gobron, N., Styles, J., & Berger, M. (2012). An Earth Observation Land Data Assimilation System (EO-LDAS). *Remote Sensing of Environment, 120*, 219–235.

Lindquist, E. J., 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 global land survey. *International Journal of Remote Sensing, 29*, 7269–7275.

Lo, T. H. C., Scarpace, F. L., & Lillesand, T. M. (1986). Use of multitemporal spectral profiles in agricultural land-cover classification. *Photogrammetric Engineering and Remote Sensing, 52*, 535–544.

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

Markham, B. L., Storey, J. C., Williams, D. L., & Irons, J. R. (2004). Landsat sensor performance: History and current status. *IEEE Transactions on Geoscience and Remote Sensing, 42*, 2691–2694.

Masek, J. (2007). White paper on use of gap-filled products for the mid-decadal Global Land Survey (GLS). Available at: http://gls.umd.edu/documents/MDGLS_gapfill.pdf (Date accessed 03/29/2012)

NASA (1996). Landsat 7 System: Data Format Control Book Volume I—Data Acquisition Plan. Available at: http://landsat.usgs.gov/documents/dfcb_vol_1.pdf (Date accessed 10/17/2012)

Nemani, R., Hashimoto, H., Votava, P., Melton, F., Wang, W., Michaelis, A., Mutch, L., Milesi, C., Hiatt, S., & White, M. (2009). Monitoring and forecasting ecosystem dynamics using the Terrestrial Observation and Prediction System (TOPS). *Remote Sensing of Environment, 113*, 1497–1509.

O'Rourke, J. (1998). Computational Geometry in C. Cambridge, MA: Cambridge University Press.

Potapov, P., Hansen, M. C., Stehman, S. V., Loveland, T. R., & Pittman, K. (2008). Combining MODIS and Landsat imagery to estimate and map boreal forest cover loss. *Remote Sensing of Environment, 112*, 3708–3719.

Ridder, R. M. (2007). Options and recommendations for a global remote sensing survey of forests. *Global Forest Resources Assessment 2010, Forest Resources Assessment Working Paper 141*. Rome: Food and Agriculture Organization of the UN.

Rossow, W. B., & Dueñas, E. N. (2004). The International Satellite Cloud Climatology Project (ISCCP) web site: An online resource for research. *Bulletin of the American Meteorological Society, 85*, 167–172.

Roy, D. P., Ju, J., Kline, K., Scaramuzza, P. L., Kovalskyy, V., Hansen, M., Loveland, T. R., Vermote, E., & Zhang, C. (2010). Web-enabled Landsat Data (WELD): Landsat ETM+ composited mosaics of the conterminous United States. *Remote Sensing of Environment, 114*, 35–49.

Roy, D. P., Ju, J., Mbow, C., Frost, P., & Loveland, T. (2010). Accessing free Landsat data via the Internet: Africa's challenge. *Remote Sensing Letters, 1*, 111–117.

Running, S. W., Nemani, R. R., Townshend, J. R. G., & Baldocchi, D. D. (2009). Next-generation terrestrial carbon monitoring. In B. J. McPherson, & E. T. Sundquist (Eds.), *Carbon sequestration and its role in the global carbon cycle* (pp. 359). Washington, D. C.: American Geophysical Union.

Sleeter, B. M., Sohl, T. L., Bouchard, M. A., Reker, R. R., Soulard, C. E., Acevedo, W., Griffith, G. E., Sleeter, R. R., Auch, R. F., Sayler, K. L., Prisley, S., & Zhu, Z. (2012). Scenarios of land use and land cover change in the conterminous United States: Utilizing the special report on emission scenarios at ecoregional scales. *Global Environmental Change, 22*, 896–914.

Snyder, J. P. (1987). Map projections—a working manual. Washington, DC: United States Government Printing Office.

Souza, C. M., Jr. (2006). Mapping land use of tropical regions from space. *Proceedings of the National Academy of Sciences, 103*, 14261–14262.

Su, J. J. (1984). Enhanced ACCA Algorithm. *Space Imaging Corporation Technical Memo IT81-LSD-SA&E Memo* (pp. 274).

Teillet, P. M., Barker, J. L., Markham, B. L., Irish, R. R., Fedosejevs, G., & Storey, J. C. (2001). Radiometric cross-calibration of the Landsat-7 ETM+ and Landsat-5 TM sensors based on tandem data sets. *Remote Sensing of Environment, 78*, 39–54.

Townshend, J. R. G., Justice, C. O., Skole, D., Malingreau, J. P., Cihlar, J., Teillet, P., Sadowski, F., & Ruttenberg, S. (1994). The 1 km resolution global data set: needs of the International Geosphere Biosphere Programme. *International Journal of Remote Sensing, 15*, 3417–3441.

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

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

Wertz, J. R. (2001). Mission geometry; orbit and constellation design and management. *The Space Technology Library* (pp. 619). : Microcosm Press & Kluwer Academic Publications.

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

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., et al. (2008). Free access to Landsat imagery. *Science*, *320*, 1011.

Wulder, M. A., Masek, J. G., Cohen, W. B., Loveland, T. R., & Woodcock, C. E. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, *122*, 2–10.

WWW1 (). Web Enabled Landsat Data Products. Available at: http://weld.cr.usgs.gov/ (Date accessed 03/29/2012)

WWW2 (). Landsat Metadata. Available at: http://edcsns17.cr.usgs.gov/EE/InventoryStream (Date accessed 04/11/2012)

WWW3 (). Land Definition Table. Available at: http://landsat.usgs.gov/documents/L7LandDB_R3_10-2002.xls (Date accessed 01/05/2012)