International Journal of Remote Sensing
Vol. 29, No. 24, 20 December 2008, 7269–7275

Taylor & Francis
Taylor & Francis Group

Letter

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

E. J. LINDQUIST*†, M. C. HANSEN†, D. P. ROY† and C. O. JUSTICE‡
†Geographic Information Science Center of Excellence, South Dakota State University, Brookings, SD, USA
‡Department of Geography, University of Maryland, College Park, MD, USA

(*Received 7 February 2007; in final form 2 June 2008*)

Landsat remote sensing of the central African humid tropics is confounded by persistent cloud cover and, since 2003, missing data due to the Landsat-7 Enhanced Thematic Mapper Plus (ETM+) scan line corrector (SLC) malfunction. To quantify these limitations and their effects on contemporary forest cover and change characterization, a comparison was made of multiple Landsat-7 image mosaics generated for a six Landsat path/row study site in central Africa for 2000 and 2005. Epoch 2000 mosaics were generated by compositing (i) two to three Landsat acquisitions per path/row, (ii) using the best single GeoCover 2000 acquisition for each path/row. Epoch 2005 composites were generated by compositing SLC-off data using (iii) five to seven acquisitions per path/row, (iv) three acquisitions per path/row. Eighty per cent of pixels were of suitable quality for change detection between (ii) and (iv), emulating that which is possible with current GeoCover and planned Global Land Survey (GLS) inputs. In a more data intensive change detection analysis using mosaics (i) and (iii), 96% of pixels had suitable quality. Compositing more acquisitions per path/row for the study area systematically reduced the percentage of SLC-off gaps and, when more than three acquisitions were composited, reduced the percentage of pixels with high likelihood of cloud, haze or shadow. The results indicate that additional input imagery to augment both the Geocover and GLS data may be required to enable forest cover and change analyses for regions of the humid tropics.

1. Introduction

Quantifying the rates and spatial pattern of the fine-scale, tropical forest cover change observed in central Africa (Wilkie and Laporte 2001) is important for making land management decisions that affect biodiversity, biogeochemical processes and human health (IPCC 2001, CBFP 2005). Remotely sensed regional land cover characterization at fine spatial resolution has typically utilized the Landsat sensor series (Townshend and Justice 1988, Goward *et al.* 2001, Williams *et al.* 2006). The GeoCover global decadal Landsat data set is composed of single-date acquisitions selected for each path/row from the 1970s, 1990s and 2000s but does not provide complete land surface observations due to persistent cloud (Tucker *et al.* 2004). The humid tropics are particularly cloudy at the time of Landsat overpass (Ju and Roy 2008) and cloud is limiting for many Landsat applications

*Corresponding author. Email: erik.lindquist@sdstate.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2008 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160802275890

(Asner 2001). For example, in the Congo Basin study area considered in this letter, 16% of the 2000 GeoCover data were cloud and cloud shadow contaminated. Contemporary studies using Landsat-7 Enhanced Thematic Mapper Plus (ETM+) are further complicated by the May 2003 failure of the scan line corrector (SLC) that decreased the usable Landsat data by 22% without respect to clouds or other atmospheric contamination (Markham *et al.* 2004, Trigg *et al.* 2006). Consequently, unobscured, remotely sensed observation of the humid tropics often requires multiple Landsat acquisitions. For these reasons, the planned Global Land Survey (GLS) 2005 data set being developed by the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS) will generate a circa 2005 GeoCover-like data set by compositing up to three low cloud cover Landsat acquisitions per path/row (Masek 2007, Gutman *et al.* 2008). The Geocover and planned GLS data sets are of unquestionable value for many monitoring applications (Laporte *et al.* 2007). However, for exhaustive characterization of forest cover and change within the Congo Basin, improved data sets may be required (Hansen *et al.* 2008). Recently, a decadal forest cover change mapping (DFCM) approach was developed to temporally composite best available pixels from multiple Landsat acquisitions and generate mosaics for forest classification and change detection analyses (Hansen *et al.* 2008). This letter examines the suitability of image inputs from the Geocover and planned GLS data sets compared to more intensive compositing methods using the DFCM approach.

2. Study area and data

The study area, defined by six adjacent Landsat path/rows (WRS PR 179059, 179060, 178059, 178060, 177059 and 177060, each about 185×185 km), is situated in the north-central Democratic Republic of the Congo (DRC) within the 'cuvette centrale' of the Congo River Basin (figure 1(*a*)). The area is characterized by low relief, meandering rivers and continuous dense Guineo-Congolian lowland tropical evergreen rain forest (White 1983) with settlements connected by unsurfaced roads.

The six GeoCover 2000 Landsat-7 acquisitions were obtained for the study area. In addition, Landsat-7 ETM+ data from two epochs (2000 and 2005) were selected based on minimal cloud cover as determined by metadata and browse imagery downloaded from GLOVIS (WWW1). A total of 14 Landsat-7 SLC-on acquisitions were selected from 2000 to 2003, providing two to three acquisitions per path/row for the 2000 epoch. A total of 36 SLC-off acquisitions were selected from 2004 to 2006, providing five to seven acquisitions per path/row for the 2005 epoch. These data were geometrically registered to the Geocover data using an automated ground control point matching algorithm and bilinear resampling (Kennedy and Cohen 2003). Visual examination of the coregistered data, focusing on regions containing distinct features, indicated that the data were misregistered by less than half a pixel, which did not significantly impact the subsequent compositing. Water bodies were removed from the analysis by applying an existing water mask (Hansen *et al.* 2008).

3. Methods

Epochal Landsat mosaics were created for both the 2000 and mid-decadal (2005) epochs using the DFCM compositing method described in Hansen *et al.* (2008). Every Landsat-7 pixel was assigned one of seven quality assessment (QA) values defined using classification trees and extensive training data applied to Landsat-7

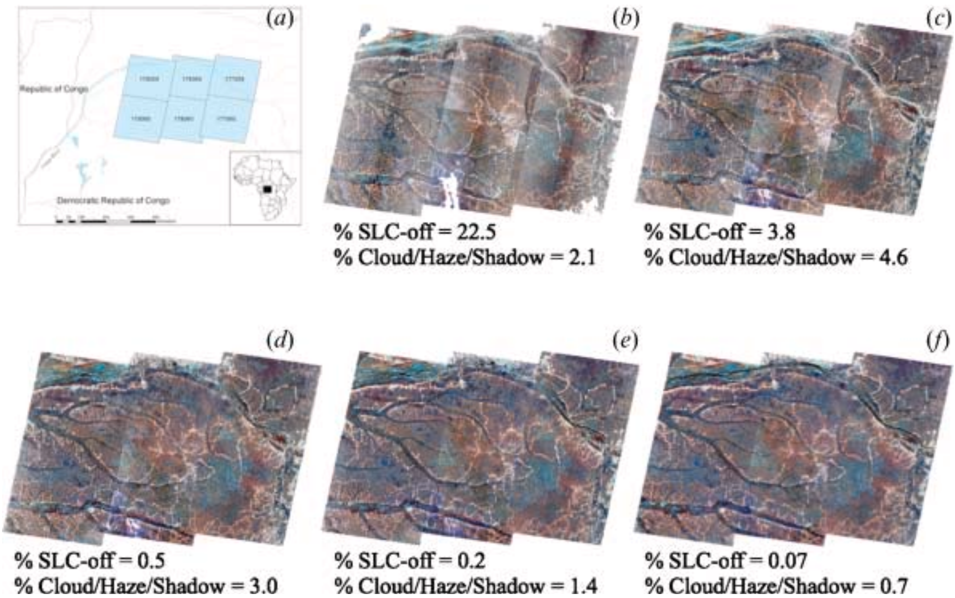

Figure 1. The study area (*a*), and 2005 epoch mosaics generated to illustrate DFCM compositing of (*b*) one, (*c*) two, (*d*) three, (*e*) four and (*f*) five Landsat-7 ETM+ acquisitions per path/row. Colour composites of Landsat-7 ETM+ bands 4, 5 and 7 are shown. Missing ETM+ SLC-off pixels are evident as white stripes and residual cloud contaminated pixels are evident in white, particularly in (*b*). Each mosaic is composed of more than 51 million $57\,\text{m} \times 57\,\text{m}$ pixels.

bands 4 (0.78–0.90 μm), 5 (1.55–1.75 μm), 6 (10.4–12.5 μm) and 7 (2.09–2.35 μm) and all combinations of possible two-band simple ratios (table 1). Composites were created by selecting pixels with the best quality as defined by the QA values. Pixels with no, or low, likelihood of cloud/haze or shadow (i.e. QA values <4) are suitable for land remote sensing applications.

Epoch 2000 mosaics were compiled (i) using single image GeoCover data, (Tucker *et al.* 2004) and (ii) by compositing two to three Landsat-7 pre-SLC-off acquisitions per path/row. Epoch 2005 composites were generated (i) simulating the planned GLS approach using three Landsat-7 SLC-off acquisitions per path/row and (ii) by compositing five to seven Landsat-7 SLC-off acquisitions per path/row. Selection of

Table 1. Per-pixel QA values, the relative quality, and the defining characteristics of those values obtained as part of a decadal forest cover change mapping algorithm developed by Hansen *et al.* (2008), and used to create best-pixel temporally composited Landsat mosaics in central Africa.

QA value	Quality	Characteristic
1	Best	No cloud/haze/shadow
2	Good	Low likelihood of cloud/haze
3	Good	Low likelihood of shadow
4	Buffer/Poor	Adjacent to QA 5 or 6
5	Poor	High likelihood of cloud
6	Poor	High likelihood of shadow
7	Missing	Missing due to SLC-off

the three acquisitions for the simulated GLS approach was undertaken by exhaustively computing which combination of three out of the available five to seven acquisitions provided the greatest number of composited pixels with QA values <4.

The quality of the individual epochal mosaics was quantified by computing the percentage of pixels with different QA values. To assess the utility of the epochal mosaics for mid-decadal change detection, the percentage of pixels with QA values <4 occurring at the same pixel locations was compared for the 2000 and 2005 mosaics.

4. Results

Figure 1(*b–f*) illustrates the results of the per-pixel DFCM compositing process, for the 2005 epoch mosaic generated by using one (*b*), two (*c*), three (*d*), four (*e*) and five (*f*) Landsat-7 SLC-off acquisitions per path/row. In each case, the Landsat acquisitions were selected by exhaustively computing which combination of the available five to seven acquisitions per path/row provided the greatest number of composited pixels with QA values <4. For example, the mosaic produced using one acquisition (figure 1(*b*)) was generated using the single Landsat acquisition for each path/row that had the least number of cloud, haze and shadow pixels. As the number of acquisitions composited is increased, the percentage of SLC-off gaps decreases monotonically (from 22.5, 3.8, 0.5, 0.2 to 0.07%, respectively). This is because the spatial phase of the SLC-off gaps is not constant between acquisitions of a path/row (USGS 2004). The percentage of cloud, haze or shadow pixels increases from 2.1% to 4.6% and then decreases to 3, 1.4 and 0.7% as the number of acquisitions composited is increased from one to five, respectively. This pattern occurs because clouds, haze and shadows may persist at certain locations and pixels from additional acquisitions that fill SLC-off gaps may be atmospherically contaminated.

Figure 2 shows histograms of the QA values in each of the four epochal mosaics. The epoch 2000 mosaics were generated with Landsat data acquired prior to the SLC-off failure and so have no SLC-off gaps (QA value 7). Best quality pixels comprised 77% and 93% of the epoch 2000 GeoCover and DFCM mosaics, respectively. The percentage of best quality pixels increased to 86% and 96% of the epoch 2005 simulated GLS and DFCM mosaics, respectively. The percentage of poor quality pixels totalled 16% and 3% in the 2000 Geocover and DFCM composites and 6% and 2% in the 2005 GLS and DFCM composites, respectively.

Figure 3 shows the percentage of good quality (QA value <4) and best quality (QA value 1) pixels occurring at the same pixel locations in the 2000 and the 2005 epochal mosaics. Specifically, the 2000 Geocover mosaic is compared with the 2005 simulated GLS mosaic (grey), and the 2000 and 2005 DFCM mosaics are compared (black). Increasing the number of image inputs in both epochs affects the amount of best and good quality data available for change detection. In a simulated GLS approach to change detection, using the single acquisition Geocover 2000 mosaic and the three acquisition simulated GLS 2005 mosaic, 80% of pixels were of suitable quality (69% of which were best quality) for change detection. The percentage of suitable quality pixels increased to 96% (89% of which were best quality) using the two to three and five to seven DFCM composited 2000 and 2005 mosaics.

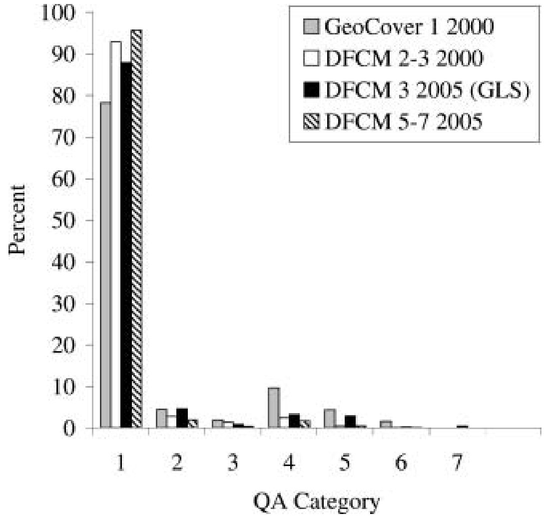

Figure 2. Histograms of the percentage of pixels in the four Landsat epochal mosaics for each QA value. QA values in descending order of quality are: 1, no cloud/haze or shadow (i.e. best quality); 2, low likelihood of cloud/haze; 3, low likelihood of shadow; 4, adjacent to high likelihood cloud/haze or shadow; 5, high likelihood of cloud/haze; 6, high likelihood of shadow; 7, missing due to SLC-off gaps. The four Landsat epochal mosaics are labelled in the key as method, number of Landsat acquisitions used per path/row, and epoch period. See text for further details.

5. Discussion and conclusion

The results reported in this letter demonstrate that multiple Landsat acquisitions per path/row are needed to generate high quality composites for the humid tropical forests of central Africa. It is generally difficult, however, to establish the number of acquisitions required because of spatio-temporal variation in cloud at the time of

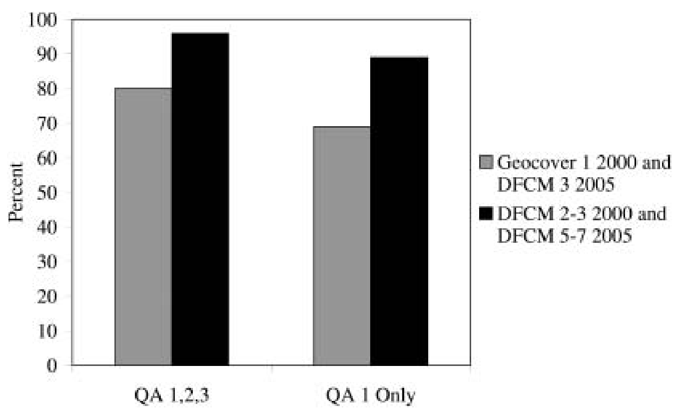

Figure 3. The percentage of good quality pixels (QA values 1, 2 and 3, i.e. low likelihood of cloud/haze or shadow) and best quality pixels (QA value 1, i.e. no cloud/haze or shadow) occurring at the same pixel locations in both epochal mosaics: grey shows a comparison of the 2000 Geocover (1 acquisition) and the 2005 simulated GLS (3 acquisitions) mosaics; black shows a comparison of the 2000 DFCM (2–3 acquisitions) and the 2005 DFCM (5–7 acquisitions) mosaics.

satellite overpass and the selective availability of Landsat acquisitions in many parts of the world, including the Congo (Ju and Roy 2008). Because of this lack of knowledge and the reality of resource constraints, generic approaches have been suggested and applied to the processing of tropical decadal datasets: best single-date imagery in the case of GeoCover (Tucker *et al.* 2004) and best two- to three-date imagery in the case of the planned GLS Landsat-7 SLC-off processing (Masek 2007).

We have demonstrated, for a limited study over the central African humid tropics, that the planned GLS approach of compositing up to three low cloud cover Landsat-7 ETM+ acquisitions for each path/row will produce composites that have minimal cloud, haze and shadow contamination and minimal SLC-off gaps. However, when comparing composites from different epochs for change detection purposes, SLC-off gaps and cloud-, haze- and shadow-contaminated pixels may not occur at the same locations, and the resulting number of pixels useful for change detection will be reduced. The current planned option for contemporary mid-decadal change detection by comparison of GeoCover and GLS data sets, for the study area, leaves 20% of the pixels unsuitable for change detection, which may preclude meaningful analysis of humid tropical forest cover change. This limited study shows that use of 2000 and 2005 epochal mosaics composited using more Landsat acquisitions per path/row (two to three in 2000 and five to seven in 2005) results in only 4% of the pixels remaining unsuitable for change detection. This reinforces the concept that, ideally, all of the data in the Landsat archive should be used to overcome the prevalence of cloud contamination, SLC-off gaps and other deleterious remote sensing variations (Hansen *et al.* 2008, Roy *et al.* 2008). Instead of using a generic number of image inputs per path/row, a minimum data quality threshold could be defined and the Landsat data archives mined until the threshold is met. This approach may become feasible when the current USGS plans to open up the Landsat-7 archive for free digital download are realized and when this data policy is more broadly adopted by other international satellite data providers.

Acknowledgements

Funding from NASA supported this research under grants NNX07AO57H and NNG06GC41G. The comments of three anonymous reviewers are also appreciated and greatly improved this manuscript.

References

ASNER, G.P., 2001, Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, **22**, pp. 3855–3862.

CONGO BASIN FOREST PARTNERSHIP, 2005, *The Forests of the Congo Basin: A Preliminary Assessment*. Available online at: carpe.umd.edu/resources/Documents/focb_aprelimassess_en.pdf/view (accessed 14 July 2007).

GOWARD, S.N., MASEK, J.G., WILLIAMS, D.L., IRONS, J.R. and THOMPSON, R.J., 2001, The Landsat 7 mission, terrestrial research and applications for the 21st century. *Remote Sensing of Environment*, **78**, pp. 3–12.

GUTMAN, G., BYRNES, R., MASEK, J., COVINGTON, S., JUSTICE, C., FRANKS, S. and KURTZ, 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**, pp. 6–10.

HANSEN, M., ROY, D.P., LINDQUIST, E., JUSTICE, C. and ALSTATT, A., 2008, A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in Central Africa. *Remote Sensing of Environment*, **112**, pp. 2495–2513.

IPCC, 2001, *Climate Change 2001: The Scientific Basis*. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Available online at: www.ipcc.ch/ipccreports/tar/wg1/index.htm (accessed 30 May 2008).

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.

KENNEDY, R.E. and COHEN, W.B., 2003, Automated designation of tie-points for image-to-image coregistration. *International Journal of Remote Sensing*, **926**, pp. 3467–3490.

LAPORTE, N., STABACH, J., GROSCH, R., LIN, T. and GOETZ, S., 2007, Expansion of industrial logging in Central Africa. *Science*, **316**, p. 1451.

MARKHAM, B., STOREY, J., WILLIAMS, D. and IRONS, J., 2004, Landsat sensor performance: history and current status. *IEEE Transactions on Geoscience and Remote Sensing*, **42**, pp. 2691–2694.

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/MDGLS-gapfill.pdf (accessed 29 January 2008).

ROY, D.P., JU, J., LEWIS, P., SCHAAF, C., GAO, F., HANSEN, M. and 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**, pp. 3112–3130.

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.

TRIGG, S.N., CURRAN, L.M. and MCDONALD, A., 2006, Utility of Landsat 7 satellite data for continued monitoring of forest cover change in protected areas in Southeast Asia. *Singapore Journal of Tropical Geography*, **27**, pp. 49–66.

TUCKER, C.J., GRANT, D.M. and DYKSTRA, J.D., 2004, NASA's global orthorectified Landsat data set. *Photogrammetric Engineering and Remote Sensing*, **70**, pp. 313–322.

USGS, 2004, *SLC-off Gap-Filled Products Gap-Fill Algorithm Methodology*. Available online at: http://landsat.usgs.gov/resources/files/L7SLCGapFilledMethod.pdf (accessed 29 January 2008).

WHITE, F., 1983, *The Vegetation of Africa: A Descriptive Memoir to Accompany the UNESCO/AETFAT/UNSO Vegetation Map of Africa* (Paris: UNESCO).

WILKIE, D. and LAPORTE, N., 2001, Forest area and deforestation in Central Africa: current knowledge and future directions. In *African Rain Forest Ecology and Conservation*, W. Weber, L. White, A. Vedder and L. Naughton-Treves (Eds), pp. 119–139 (New Haven, CT: Yale University Press).

WILLIAMS, D.L., GOWARD, S. and ARVIDSON, T., 2006, Landsat: yesterday, today, and tomorrow. *Photogrammetric Engineering and Remote Sensing*, **72**, pp. 1171–1178.

WWW1, USGS Global Visualization Viewer. Available online at: glovis.usgs.gov (accessed 20 December 2007).