MAXMISING AUTOMATION IN LAND COVER MONITORING WITH CHANGE DETECTION

Konrad Wessels[1], Frans van den Bergh[1], Karen Steenkamp[1],

Derick Swanepoel[1], Bryan McAlister[1], Brian Salmon[1], David Roy[2] and Valeriy Kovalskyy[2]

1. Remote Sensing Research Unit, CSIR-Meraka, Pretoria, 0001, South Africa

2. Geospatial Sciences Center of Excellence, South Dakota State University, Brookings, SD 57007, USA

KEYWORDS: Land cover, change detection, Landsat, South Africa

ABSTRACT

With the availability of free moderate spatial resolution Landsat satellite data land cover mapping systems are moving away from classifying single date cloud-free images to classifying data time-series. This requires the ability to handle large volumes of data, which in turn requires high levels of automation in data pre-processing, image classification and change detection. This paper reports on the progress made towards the development of a more automated land cover monitoring system for South Africa. We firstly employed a local installation of the Web-enabled Landsat Data (WELD) system to serve as the data "backbone" for pre-processing and storing large amounts of Landsat data sensed over South Africa. A system was developed to rapidly update land cover maps for previous mapped areas using highly-automated, training data generation, scalable random forest classification, accuracy assessment, change detection and rapid, online operator validation. The technology is aimed at assisting government and industry to provide land cover data at a much higher update frequency to address ever-increasing demands for land cover products and services.

INTRODUCTION

Land cover is a fundamental description of the terrestrial surface. Globally the conversion of natural vegetation to croplands, pastures, monoculture plantations, and urban areas have contributed significantly to greenhouse gasses and caused considerable losses of biodiversity (Foley et al., 2005). Land cover change furthermore affects surface energy fluxes between the land and the atmosphere which drive climate and climatic change. Land cover has been identified as an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS) under the framework of United Nations Framework Convention on Climate Change (UNFCC).

In South Africa, there is a pressing demand for land cover data to support various legislative and operational needs of various national and provincial departments. A recent survey revealed that the majority of users require data every 2-4 years, completed within 2 years from the date of image acquisition (Wessels, 2014). Despite this demand, the last official, national land cover data set dates back to 2000 (Fairbanks et al., 2000). Since then individual provinces, have funded their own land cover maps for various reporting dates. As a consequence, CSIR-Meraka undertook to develop a highly automated and scalable land cover mapping systems to improve the efficiency of land cover mapping in South Africa.

Monitoring land cover over large areas ideally requires automated land cover classification methods with little interpretation by human operators. Rather than select individual cloud-free images and then classify them using photointerpretation techniques, application of automated classification approaches to all the available satellite data is increasingly the norm (Hansen and Loveland, 2012). As large volumes of historical and current Landsat data are now freely available (Wulder et al., 2012) and with the launch of the Landsat Data Continuity Mission (LDCM) in 2013, the user community requires systems that can pre-process Landsat data and provide robust data structures for time-series and large area analysis rather than analysis of individual scenes. To this end, we employed a local installation of the Web-enabled Landsat Data (WELD) system (Roy et al., 2010) to serve as the data "backbone" for pre-processing and storing large amounts of Landsat data sensed over South Africa. Furthermore, monitoring land cover over large areas will ideally require automated land cover classification methods with little interpretation by human operators (Hansen and Loveland, 2012).

The overall objective of the study was to develop a system which can rapidly update a land cover map for a desired year from a previously produced land cover map using machine learning methods, change detection and Landsat time series data. The system prototype was tested for the Kwa-Zulu Natal province (KZN) of South Africa.

METHODS

System overview

The system relies on generating training data from existing land cover maps for rapid update in subsequent years. In this case we used the land cover maps created by GeoTerraImage (GTI) for Ezemvelo KZN Wildlife using SPOT5 images for 2005 (GeoTerraImage, 2008), 2008 (GeoTerraImage, 2010) and 2011. This forms part of Ezemvelo KZN Wildlife long term monitoring of land use change program (Escott and Jewitt, 2014). The training and model development module (Fig. 1) firstly creates training samples from the original land cover map in a controlled manner which distributes training evenly across a study area, avoid edges of land cover polygons and provides more training for small polygons / classes so that it is not dominated by a few extensive classes e.g. grasslands. The input variables (or features) are the seven Landsat bands for each of the four seasonal WELD composites for a specific year, as well as static data layers: digital elevation model (DEM), mean annual precipitation (GMAP) and mean annual temperature (Tmean) (Fig. 2). The latter three static features enable the random forest classifier to distinguish the same land cover type in different environments. This is common practice in regional remote sensing based classifications.

Random forest (RF) supervised classifiers create an ensemble of decision tree classifications on different data subsets and decides on the mode of all the outputs of each tree to classify each input pixel. The classification and mapping module takes the random forest models developed using Waikato Environment for Knowledge Analysis (WEKA) for the four seasons and individual seasons and applies the models to classify each Landsat pixel. The benefit of using the four seasons is that they capture the phenological differences of various land cover types. In addition, if clouds occur within a pixel of a specific season, the system applies RF WEKA models that excludes that season, and classifies based on the remaining seasons. The classification system thus truly capitalises on a multi-temporal approach and compensates on a pixel level for missing data. Finally the land cover map was filtered according to heuristic, class-specific, size-specific rules to reduce speckle, but preserve edges and small linear classes (e.g. rivers and small water bodies).

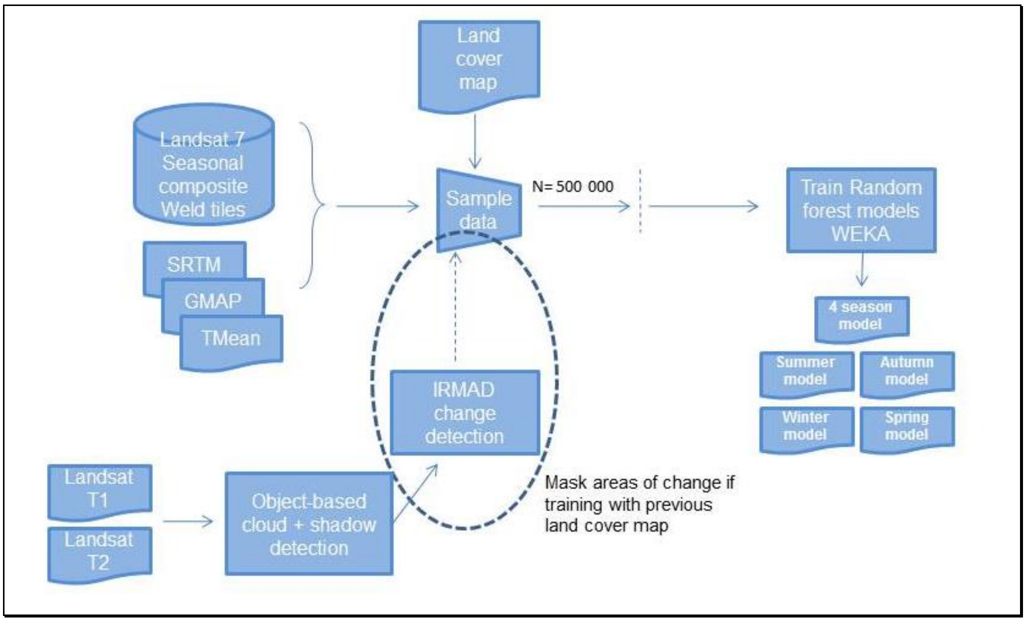

Figure 1. Training and model development module of land cover mapping system.

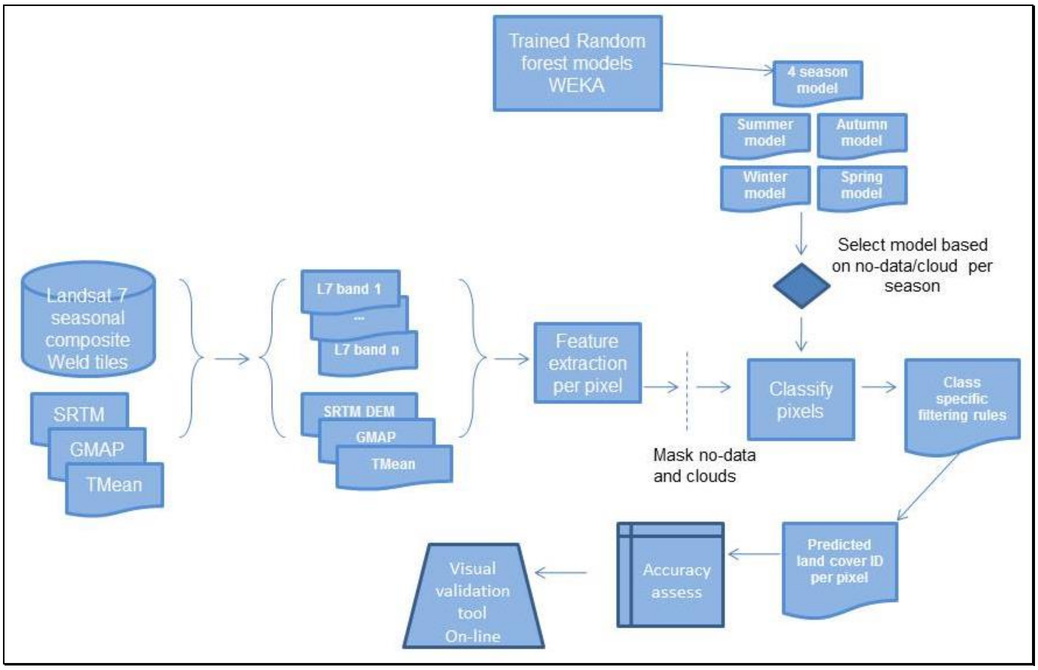

Figure 2. Classification and mapping module of land cover system.

Change detection is used to exclude areas from training which may have changed between the date of the original land cover training map and the subsequent date to which a new map must be updated (Fig. 1). Several Landsat change detection methods have been published - for review see (Hansen and Loveland, 2012). Following extensive literature review we selected the Iteratively reweighted multivariate alteration detection (IRMAD) approach for change detection since it is effective at normalising image pairs of different dates by compensating to linear transformations that are typically caused by seasonal changes (differences in greenness due to image date) and atmospheric change (Canty and Nielsen, 2008).

Study area

The KwaZulu-Natal (KZN) province of South Africa (92100 km^2) has diverse land cover types and land uses ranging from natural grasslands and savannas to exotic forestry plantations, sugarcane, dryland cultivation and urban areas. The natural vegetation range along a continuum of tree cover, from grasslands, to open savannah woodlands, bushland, thickets, and indigenous forests. KZN is the only province with repeat land cover data due to the lon term monitoring program (Escott and Jewitt, 2014). The land cover of KZN was officially mapped by GeoTerraImage (GTI) for 2005 (GeoTerraImage, 2008), 2008 (GeoTerraImage, 2010) and 2011 using SPOT5 imagery, with an overall accuracy of approximately 80%. The original 45 classes which included various land use classes were aggregated to 25 land cover classes.

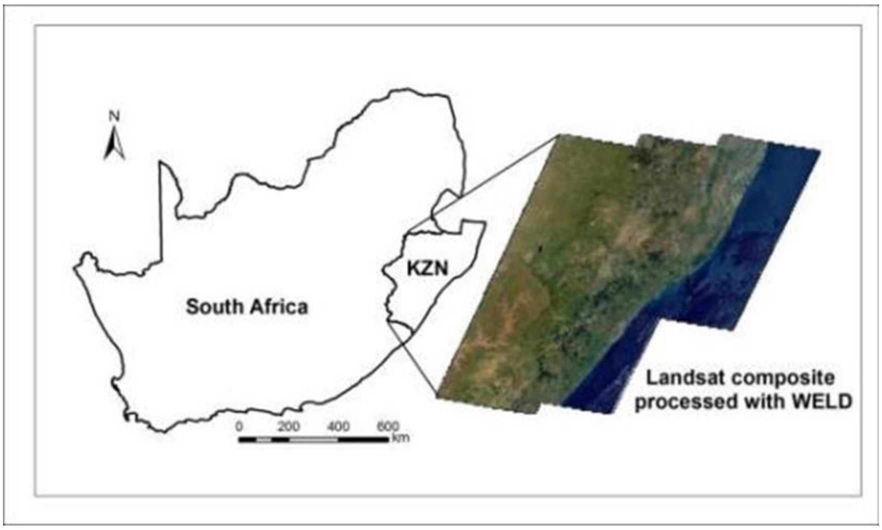

Figure 3. KwaZulu-Natal (KZN) province of South Africa and WELD seasonal composite.

Landsat ETM+ data pre-processing with WELD

A total of 84 Landsat 7 Enhanced Thematic Mapper plus (ETM+) images were obtained over KwaZulu-Natal (KZN) from 1 Dec. 2007 to 30 Nov. 2008. The WELD Version 1.5 algorithms (Roy et al., 2010) were used to generate 30m composited seasonal mosaics of KZN for each season; summer: December 2007 to February 2008 (20 images), autumn: March to May (17 images), winter: June to August (24 images), and spring: September to November (23 images). The Version 1.5 WELD products store the six reflective top of atmosphere Landsat 7 ETM+ bands, the two thermal bands, and other data for each 30m pixel (Roy et al. 2010). The WELD products are defined in fixed geolocated tiles defined in the equal area sinusoidal projection. Compositing procedures are applied independently on a per-pixel basis to the gridded WELD time series to reduce cloud and aerosol contamination, fill missing values due to the Landsat ETM+ scan line corrector failure that removed about 22% of each Landsat 7 image, and to reduce data volume.

RESULTS

Change detection validation

An on-line validation system (Fig. 4) was developed that presents an operator with 2008 and 2011 Landsat 30m and SPOT 2.5m true colour images (provided by South African National Space Agency – SANSA), overlaid with the change polygons generated by the IRMAD procedure. The system enables the operator rapidly, visually assess and record land cover changes with a bounding box around automatically selected change polygons. It also simulated no-change polygons (25% of total number of polygons) by randomly locating actual change polygons, whilst avoiding areas with a 50th percentile scaled change metric values. To test operator differences a common pool of 15% (N = 105) of all polygons validated by each operator (N = 700) were presented to each operator in random order.

Figure 4. Web-based land cover change detection validation interface using SPOT5 2.5m pan-sharpened merged product.

The change detection validation results can be summarised as follows:

- 11% of the change detected by IRMAD was due to remnant cloud and cloud shadow not detected
- 5% of change detected was due to burned area often in grassland, wetlands and agricultural fields, although the dates of the Landsat images used was before the burn season
- After removing the above instances of clouds or burns from the data, 16.5% of the change detected by IRMAD was not deemed to be land cover change by the operator (false positive)
- The change detection accuracy was 83.5%, i.e. the change detected by IRMAD that was considered to be true change by the operator (true positive)
- 92.7 % of the polygons containing no change (randomly located), were confirmed not to have changed by the operator – (true negative) – this is in agreement with the empirically determined threshold value, which implied a false negative value of 7%

Most of the change detected was dominated by two forms of change:

- Forestry Plantations clear felling and growth (35% of all changes)
- Changes within cultivated fields from active dryland, irrigated, sugarcane to bare soil or fallow fields (56% of all changes)
- Forestry and cultivation changes accounted for 91% of detected change after excluding erroneous changes due to clouds or burned area.

Land cover classification

The supervised land cover classification system was optimised in terms of the amount of training data generated, the random forest decision tree complexity and the number of trees included in the random forest.

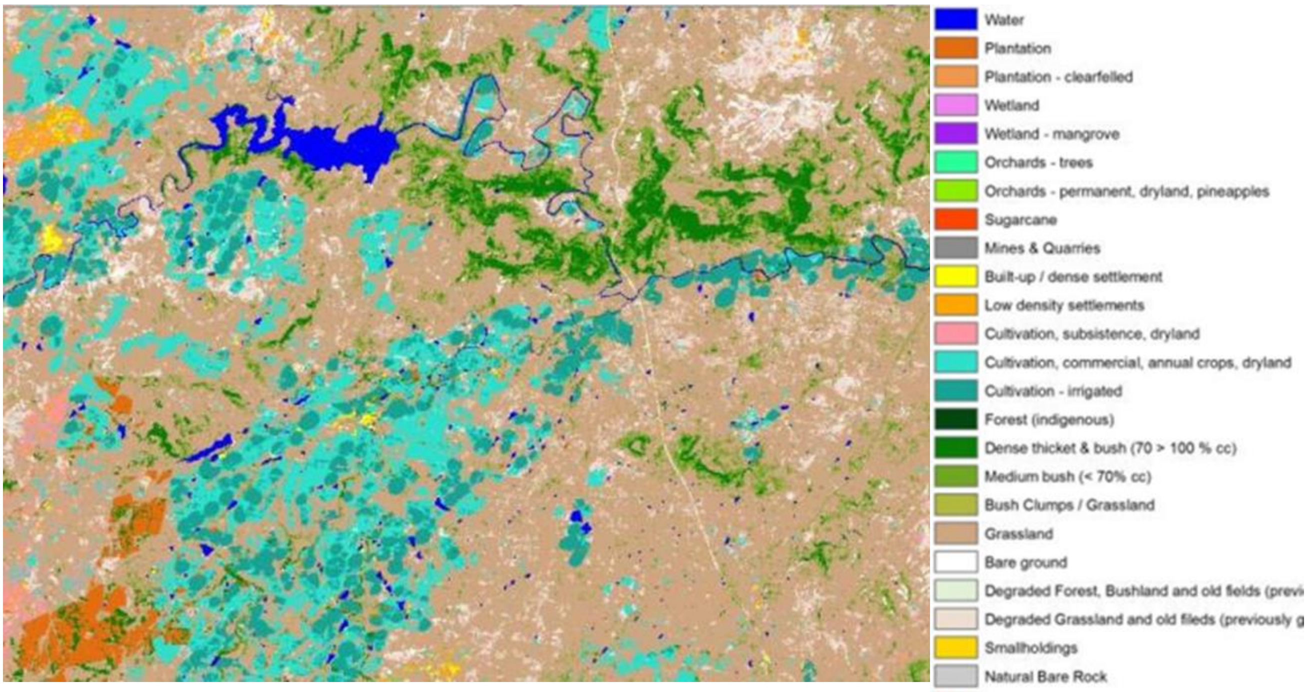

Figure 5. Zoom-in of Landsat 7 land cover data for KwaZulu-Natal.

The system produced coherent land cover maps (Fig. 5), however the scan line corrector error in the Landsat 7 data clearly led to misclassifications in some areas. This suggests that the random forest classifier was not always able to correctly classify pixels used to fill gaps using different dates when they were too different spectrally. Spectral differences within the seasonal composites were due to atmospheric effects and phenological changes within the three month seasonal composite period. The preliminary overall producer accuracies for 2005 and 2008 are summarized in Table 1.

Table 1. Overall producer accuracy of KwaZulu Natal land cover using the original Ezemvelo KZN Wildlife land cover classes and the same aggregated to the National Geospatial Information's Land Cover Classification System (LCCS) classes.

	2005		2008	
	original classes	NGI LCCS	original classes	NGI LCCS
Overall producer accuracy	65%	74%	61%	71%

The overall accuracy was calculated by comparing the classification label of every pixel in KZN to that in the original Ezemvelo KZN Wildlife map (resampled to 30m), with the exclusion of the pixels used to train the models. The accuracy assessment is therefore based on every pixel in the study area and not just a small sample. Note that the error of the original map is convolved in our error. The overall accuracy was encouraging considering the high level of automation, however, it was not yet at the level required by operational users (above 80%). The accuracy improved when the original 25 classes were aggregated to 15 National Geospatial Information's Land Cover Classification System

(LCCS) classes. Most of the misclassifications using the original classes were due to a number of classes that are spectrally similar and which cover large areas (thus having a disproportionate influence on overall accuracy), these include degraded grasslands, low density settlements, subsistence cultivation and grasslands (Figure 6). These classes represent areas of low fractional vegetation cover and high bare soil cover and were mainly distinguished based on ancillary data and visual interpretation based on the land use context. Other important land cover classes such as forestry plantations, indigenous forest, sugarcane, bare soil and irrigated cultivation were all classified with a high level of accuracy.

The 2011 land cover was created by first excluding areas of change between 2008 and 2011 from the training data generation using the 2008 class labels (which are still correct since no land cover change was detected), but the 2011 Landsat 7 WELD seasonal composites. The cross validation showed that the 2011 RF models trained with a high level of classification accuracy (comparable to 2008 Table 1), however a full validation could not yet be conducted as the independently generated 2011 Ezemvelo KZN Wildlife land cover has not yet been officially released at time of writing.

DISCUSSION

A highly automated and scalable system was developed and its operational potential demonstrated. The system included no manual training data generation or optimal image selection. It is however dependent on a historical land cover map, but it can rapidly be updated thereafter for at least six years. The ability to rapidly update land cover over multiple provinces to single baseline year in the manner described here for 2011, is of great value to national department for monitoring and reporting purposes. More tests need to be conducted to better understand the limitations of the system. The seasonal WELD composites contained significant spectral differences between dates for the same adjacent land cover pixels. These spectral differences inevitably led to classification error. It was initially expected that the machine learning approach would learn the differences for pixels with different observation dates, but these differences were too pronounced. Future research will experiment with reducing the composite period to months instead of seasons and compositing method within WELD. Results are also expected to improve with the new Landsat 8 data which does not have the scan line corrector error and has much higher global acquisition frequency, improved cloud detection and higher radiometric resolution (Roy et al., 2014). This research presents a significant advancement towards an automated and operational land cover monitoring system capable of rapidly updating maps to a desired date.

REFERENCES

CANTY, M. J. & NIELSEN, A. A. 2008. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation. *Remote Sensing of Environment,* 112**,** 1025-1036.

ESCOTT, B. J. & JEWITT, D. 2014. Ezemvelo KZN Wildlife Long Term Land Cover and Land Use Change Monitoring Project for KwaZulu-Natal, Unpublished Report. PO Box 13053, Cascades, 3202: Ezemvelo KZN Wildlife.

FAIRBANKS, D. H. K., THOMPSON, M. W., VINK, D. E., NEWBY, T. S., VAN DEN BERG, H. M. & EVERARD, D. A. 2000. The South African Land-cover characteristics database: a synopsis of the landscape. *South African Journal of Science,* 96**,** 69-82.

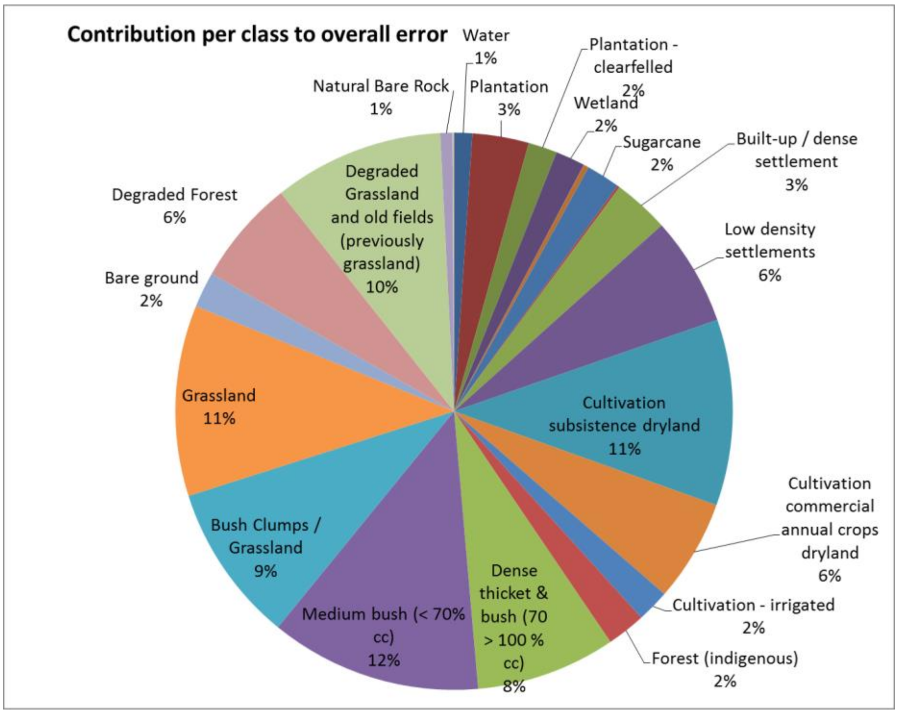

Figure 6. Percentage contribution to overall classification error by various land cover classes.

FOLEY, J. A., et al. 2005. Global consequences of land use. *Science,* 309**,** 570-574.

GEOTERRAIMAGE 2008. 2005 KZN Province Land-cover mapping (from SPOT 2/4 satellite imagery circa 2005-6): data users report and meta data (version 2.0). Published report by Geoterraimage (Pty) Ltd, South Africa for Biodiversity Conservation Planning Division. P. O. Box 13053, Cascades, Pietermaritzburg,3202.: Ezemvelo KZN Wildlife.

GEOTERRAIMAGE 2010. 2008 KZN Province Land-cover mapping (from SPOT 5 satellite imagery circa 2008): data users report and meta data (version 1.0). Published report by Geoterraimage (Pty) Ltd, South Africa for Biodiversity Conservation Planning Division, Ezemvelo KZN Wildlife. P. O. Box 13053, Cascades, Pietermaritzburg, 3202.: Ezemvelo KZN Wildlife.

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.

ROY, D. P., JU, J. C., KLINE, K., SCARAMUZZA, P. L., KOVALSKYY, V., HANSEN, M., LOVELAND, T. R., VERMOTE, E. & ZHANG, C. S. 2010. Web-enabled Landsat Data (WELD): Landsat ETM plus composited mosaics of the conterminous United States. *Remote Sensing of Environment,* 114**,** 35-49.

ROY, D. P., et al. 2014. Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment,* 145**,** 154-172.

WESSELS, K. J. 2014. REPORT ON USER REQUIREMENT SURVEY FOR LAND COVER DATA. Pretoria: CSIR.

WULDER, M. A., J.G. MASEK, W.B. COHEN, T.R. LOVELAND & 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.