LAND-COVER MAPPING USING LANDSAT FOR SUSTAINABLE GREEN HOUSE GAS (GHG) INVENTORY DEVELOPMENT

Conference Paper · November 2013

4 authors, including:

Benson Kenduiywo
University of California, Davis

Eunice Wanjiku
Jomo Kenyatta University of Agriculture and Technology

Charles N Mundia
Dedan Kimathi University of Technology

LAND-COVER MAPPING USING LANDSAT FOR SUSTAINABLE GREEN HOUSE GAS (GHG) INVENTORY DEVELOPMENT

Benson Kipkemboi Kenduiywo[1], Eunice W. Nduati[2], Charles N. Mundia[3] and Tesfaye Korme[4]

[1,2,3]Jomo Kenyatta University of Agriculture and Technology Department of Geomatic Engineering and Geospatial Information Systems, E-Mail: bkenduiywo@jkuat.ac.ke, enduati@jkuat.ac.ke and cmndegwa@jkuat.ac.ke
[4]Regional Center for Mapping of Resources for Development, E-Mail: korme@rcmrd.org

Abstract

Information on land-cover is important for verification of land use, land-use change and forestry (LULUCF). LULUCF is significant in assessing anthropogenic Green House Gas (GHG) emissions. This study aimed at developing a simple and computationally efficient yet accurate methodology for national land-cover mapping. The longstanding Landsat data freely available with a renewed and sustainable future archive after the launch of Landsat 8 was used. Data of two epochs namely 2000 and 2010 were selected. A chain classification approach using maximum likelihood classification (MLC) coupled with decision tree was used. Chain classification approach was significant in classifying images of different seasons given that in national land-cover projects it is rare to obtain images of the same date. Six classes recommended by IPCC were adopted. The developed approach attained an accuracy of 86% with a kappa coefficient of 0.8. The study concluded the freely available Landsat data, computational efficiency of MLC and decision tree can be tapped for sustainable land-cover mapping for GHG. This method is replicable and therefore can be used to produce complete and comparable national land-cover products.
.

Keywords: Green House Gas (GHG), chain classification, maximum likelihood classification (MLC), decisition tree

1 INTRODUCTION

Land-use/land-cover (LULC) mapping is very significant source of anthropogenic Green House Gas (GHG) emissions. Land use refers to man's activities on earth, which are directly related to land, whereas land cover refers to objects on land surface natural or manmade (Bhatta, 2010) . LULC change, mainly due to deforestation, has been found to contribute to about 20% of the GHG emissions from anthropogenic sources (Metz, 2007). Land use, land-use change and forestry (LULUCF) sector in general has an aggregate share of over 30% of the gross global emissions. This makes LULUCF a critical component in accounting for GHG emissions. To account for emissions, baseline spatial information of changes in LULC is of paramount importance. Moreover, international requirements for reporting on the environment status dictate the need to monitor land cover and land cover change through time (Intergovernmental Panel On Climate Change (IPCC), 2003; Metz, 2007). According to (Intergovernmental Panel On Climate Change (IPCC), 2003) remote sensing methods are suitable for independent verification of national LULUCF. However, most developing countries are faced with limited technological, financial and personnel resources. Consequently, this has so far challenged development of land-cover maps often resulting into incomplete and/or incomparable products. This study seeks to demonstrate the use of simplistic, replicable and sustainable remote sensing land-cover mapping approach.

Sustainable land-cover mapping requires that remote sensing processes that are used to generate maps can be implemented using limited resources. In order to be able to report LULUCF through time it is mandatory to monitor land-cover. Remotely sensed data are an integral component of large area monitoring. Therefore, longevity and continuity of remote sensing programs are indispensable to the success and feasibility of large area monitoring programs (Wulder et al., 2008). (Intergovernmental Panel On Climate Change (IPCC)) requires that: the smallest spatial unit for assessing land-use changes to be 0.05 ha which is approximately 500 m^2, images of appropriate temporal resolution and be consistently available over time. Landsat is the longstanding sensor whose data are available for free and the only cost involved is technical capacity. The Landsat Data Continuity Mission objective to collect, archive and distribute multispectral imagery for global repetitive coverage (Irons & Dwyer, 2010) saw the launch of Landsat 8 which has extended opportunities for LULUCF. Furthermore, Landsat satellite data is the most widely used data type for land cover mapping because of its 35-year data record and its global coverage (Sexton et al., 2013; Wulder, et al., 2008). Therefore, this study found it appropriate, economical and sustainable to use Landsat images for land-cover mapping.

A supervised approach using maximum likelihood classification (MLC) method coupled with decision tree was used for mapping. MLC is a statistical method that uses Bayesian formula to allocate a pixel to a class with the highest probability see (Tso & Mather, 2009). The performance of MLC is limited by frequency distribution assumption and thus decision tree was used to enhance the classification. Expert knowledge and ancillary information was used in selecting training samples for each representative class. Class separability analysis informed choice of final classes for classification. Therefore, the objective of this paper is to demonstrate a remote sensing procedure that can be utilized for land-cover mapping for GHG inventory.

2 MATERIALS

2.1 Landsat Data

Landsat freely available data of two epochs (2000 and 2010), see (USGS, 2013), was used for land-cover classification. The data scene numbers are Path 169 Row 067 to 070. General land-cover of the area include: shrub-land, cropland, forest, grassland, other-land, wetland and settlements. A total of four scenes of Landsat images were used. In 2000 epoch; scenes of 10th September 2001 and 1st October 2000 corresponding to Path 169 Row 067 to 68 and Path 169 Row 069 to 070 respectively were used. As for 2010 epoch, all the four scenes (Path 169 Row 067 to 070) corresponding to one date, 4th June 2009 were available. The images of the two epochs were selected within the dry season of the area, normally between May to October months of a year.

Figure 1: Landsat Images used for land-cover mapping

3 METHODOLOGY

3.1 Data pre-processing

Images with high spectral quality were given priority during selection. This was done in consideration of the dry season and minimum cloud cover. The images were downloaded from USGS site unzipped, Landsat bands 1, 2, 3, 4 and 7 stacked and re-projected to WGS 84 datum on Universal Transverse Mercator

(UTM) zone 36 south projection automatically using a designed python program. Band 6 (thermal band) and 8 (panchromatic band) were exempted. Clouds present on 2000 epoch images were removed by masking. Mosaicking was done only for scenes acquired in one date so as to maintain spectral coherency between scenes. Images corresponding to 2010 epoch were all acquired in one date and thus one mosaic was produced. However, in 2000 epoch two mosaics were produced corresponding to the dates of acquisition. The images of the two epochs were co-registered pixel to pixel so as to facilitate pixel to pixel change comparison.

3.2 Training site selection

Image interpretation was done to identify classes of interest. Six major classes namely: Forestland, grassland, cropland, wetland, other-land and settlement as defined by (Intergovernmental Panel On Climate Change (IPCC), 2003) were used. Training sites of the chosen classes were selected using expert knowledge with aid of Google Earth GeoEye images and ancillary data. The training sites were distributed throughout the study area taking sub-classes of the main categories. The sub-classes include: dense, moderate and sparse forest, open and closed grassland, open and closed shrub-land, vegetated wetland, annual and perennial cropland and water body. Training sites of the different categories of forest were selected with the guide of Normalized Difference Vegetation Index (NDVI). Forested areas with NDVI value greater than 70%, 50% to 70% and below 50% were considered dense, moderate and sparse forest respectively. The sub-classes were later merged to IPCC recommended classes. Transformed divergence was used to compute separability between classes as recommended by (Swain & King, 1973) for feature selection in multispectral remote sensing. Classes with separability values less than 1.5 were omitted from the classification (Mthembu & Marwala, 2008). Only one pair of classes fell below the established threshold; cropland and settlements. Therefore, we opted to omit settlements from MLC algorithm to avoid pixel class confusion. A strategy of chain classification was adopted in selection of training sites within scene overlap areas. There were spectral differences at overlap areas in 2000 epoch due to different imagery dates. This is anticipated because when classifying large areas it is sometimes difficult to obtain images of the same date especially if they are in archive. However, (Knorn et al., 2009) recommend representing classes well in the overlap area using chain classification. Following this, we ensured similar class interpretation on overlapping areas so as to avoid mismatch on scene boundaries. This approach was better than averaging pixels in overlap areas which resulted in class misinterpretation.

3.3 Image classification

Maximum likelihood classifier is prone to produce noisy classification (Kenduiywo, 2012) especially if class selection is not properly done. In contrast, advanced methods like Support Vector Machines (SVM), Markov Random Fields and Conditional Random Fields incorporate spatial context during classification hence minimize noise, see (Kenduiywo et al., 2012). However, their SVM algorithm in Envi 4.7 took more than a day to classify one scene of Landsat. MLC took less than 10 minutes to produce results of one Landsat scene using a core i7 computer. Thus, the method was deemed cost effective and sustainable for national and regional land-cover mapping for GHG projects where multiple scenes are used. We used maximum likelihood classifier in Envi 4.7 software to classify the selected classes except settlements. Settlements were digitized from GeoEye images and rasterized to 30 m pixel resolution equivalent to the classified images. The decision tree was used to integrate the settlements into the outcome of MLC to produce a final thematic map.

A Decision tree performs multistage classifications by using a series of binary decisions to place pixels into proper classes (Matinfar & Roodposhti, 2012). It partitions a data set into homogeneous subsets using nodes where trees branch or split the data set (Punia et al., 2011). The decision tree hierarchical structure for labelling objects provides flexibility in understanding relationships between objects/classes (Tso & Mather, 2009). We exploited this attribute and used decision trees for post-classification. Particularly we used it to integrate settlements, combine classes, code and add colours to the map produced by MLC. As illustrated in **Error! Reference source not found.** classes falling in settlements and that were initially classified as cropland (annual or perennial) or other-land were reclassified as settlements using the decision tree. The final thematic map from decision tree was passed through a majority analysis using a 3 by 3 majority filter in order to eliminate any noisy pixels. A 3 by 3 majority filter was sufficient given that the smallest spatial unit for assessing land-use changes as per (Intergovernmental Panel On Climate Change (IPCC), 2003) is 0.05 ha.

Figure 2: Sample of decision tree in Envi 4.7 used to integrate settlements, combine classes, coding and adding colours to the thematic maps

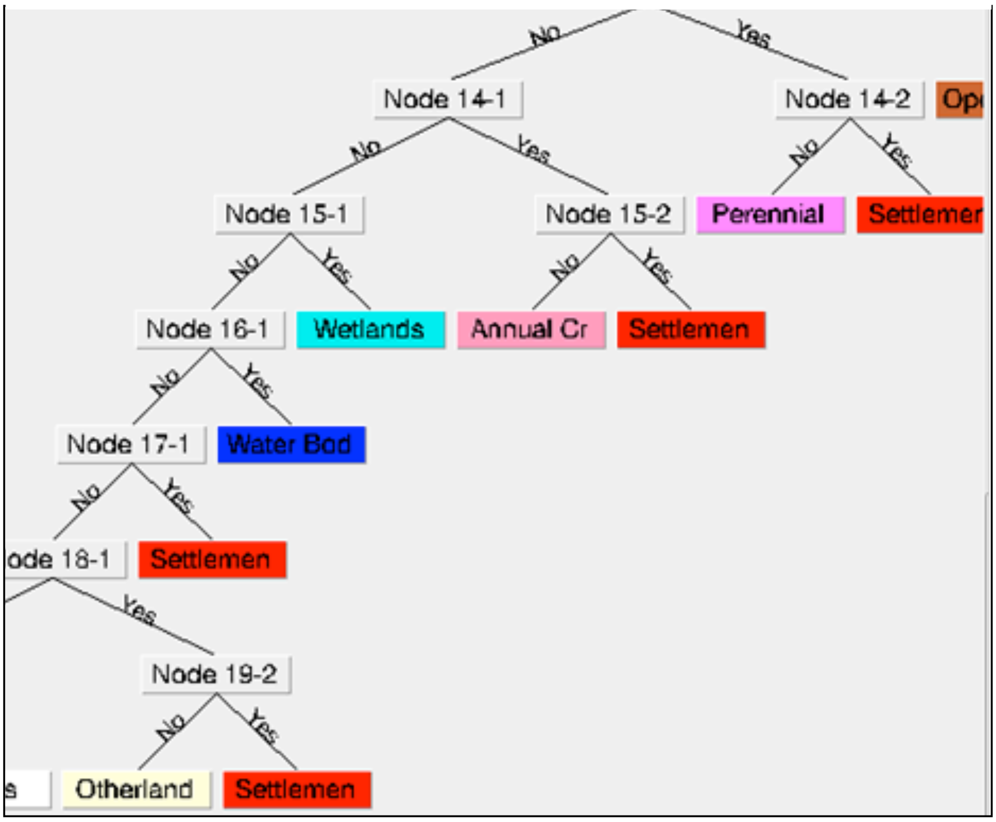
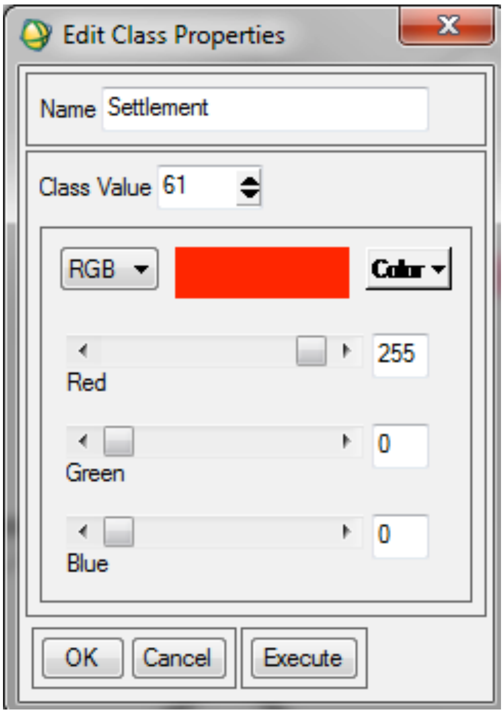


Figure 3: Properties used to code and assign colours to the classified map



3.4 Accuracy assessment

In any land-cover mapping procedures it is essential to evaluate the performance of the designed classification method. This gives a chance to experts to have a degree of confidence to the results. This study adopted the commonly used accuracy assessment method in remote sensing; confusion matrix/error matrix. It shows the proportions of correctly classified (overall accuracy) and misclassified pixels in a table matrix. In this way, several accuracy measures can be derived from it (ITC, 2010). Some of the measures; overall accuracy, false positives, false negatives and kappa statistics, are used as quality measures. False positives and false negatives are synonymous to type I and type II errors which indicate the proportions of pixels omitted and incorrectly classified respectively. They are a consequence of producer and user accuracies. Kappa statistics is useful in evaluating different remote sensing methods because it accounts for the degree of accuracy that can be attained when labels are assigned at random.

Stratified random sampling was used to generate samples of representative classes as recommended for land-cover data (ITC, 2010). The sampled points were compared with field and ancillary data to generate ground truth points. The ground truth samples were used to generate an error matrix for accuracy assessment.

4 RESULTS

Results of the study include land-cover maps for two epochs and accuracy assessment results. **Error! Reference source not found.** illustrate the final land-cover maps. Sub-classes were merged to produce the six IPCC classes. It is evident that grassland (shrub-land and forestland) converted to forestland.

	Forestland	Grassland	Cropland	Wetland	Settlement	Other-land	Total	User Accuracy
Forestland	39	5	3	1	0	0	48	81.3%
Grassland	3	57	5	1	0	0	66	86.4%
Cropland	1	0	42	0	1	0	44	95.5%
Wetland	7	0	0	13	0	0	20	65.0%
Settlement	0	0	0	0	12	0	12	100.0%
Other-land	0	0	1	0	0	9	10	90.0%
Total	50	62	51	15	13	9	200	
Producer Accuracy	78.0%	91.9%	82.4%	86.7%	92.3%	100.0%	86.0%	

show accuracy assessment results of the six classes. The overall accuracy achieved is 86.0% ((172 pixels out of a total of 200 were correctly classified) with a kappa coefficient value of 0.8. Wetland class has the lowest user accuracy of 65.0% (corollary of commission error of 35%). Consequently, forestland has the lowest producer accuracy of 78.0% (corollary of omission error of 22%).

Figure 4: Final classified images

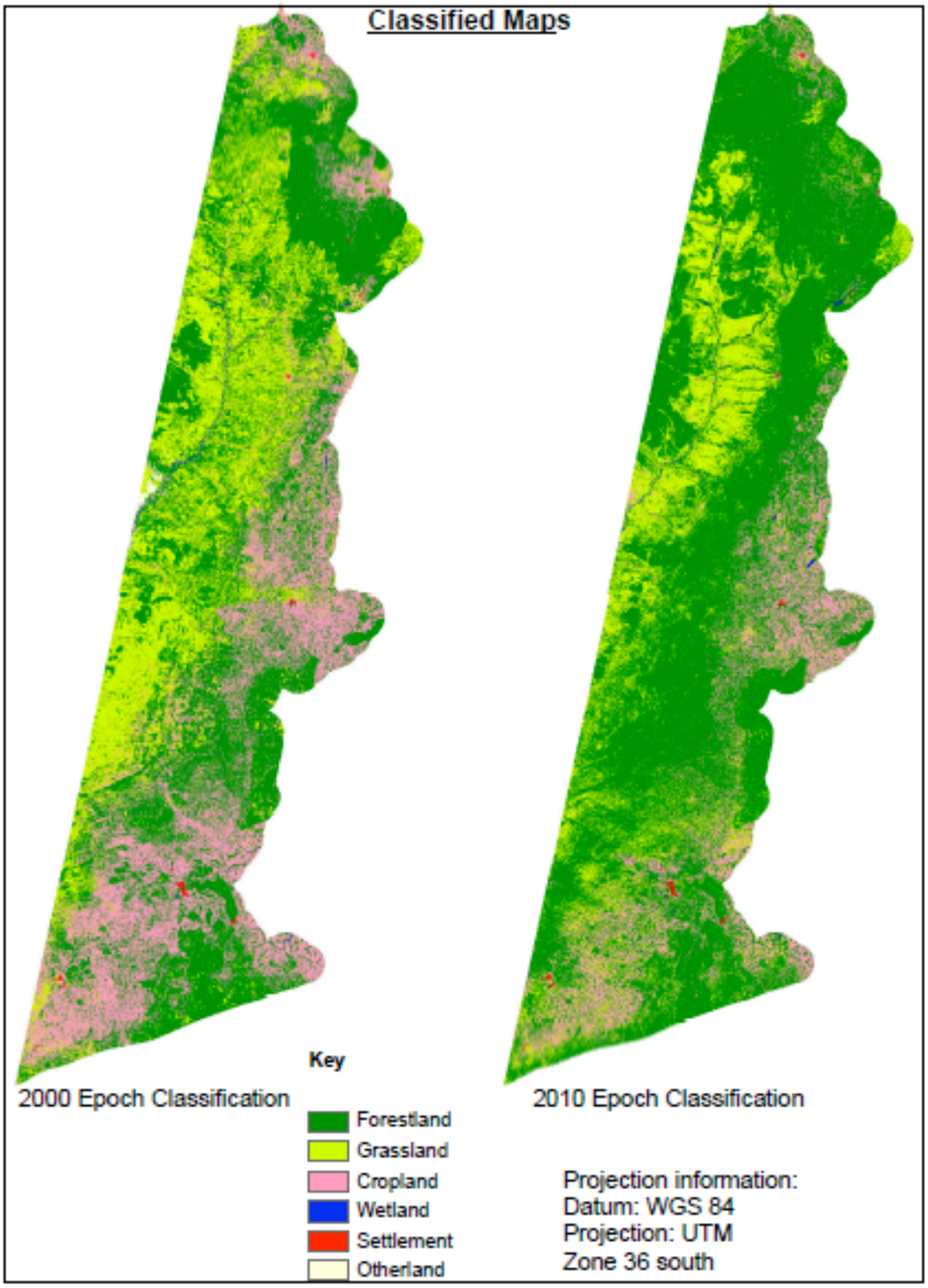


Table 1: Accuracy assessment by error matrix

	Forestland	Grassland	Cropland	Wetland	Settlement	Other-land	Total	User Accuracy
Forestland	39	5	3	1	0	0	48	81.3%
Grassland	3	57	5	1	0	0	66	86.4%
Cropland	1	0	42	0	1	0	44	95.5%
Wetland	7	0	0	13	0	0	20	65.0%
Settlement	0	0	0	0	12	0	12	100.0%
Other-land	0	0	1	0	0	9	10	90.0%
Total	50	62	51	15	13	9	200	
Producer Accuracy	78.0%	91.9%	82.4%	86.7%	92.3%	100.0%	86.0%	

5 DISCUSSION AND CONCLUSION

This study sought to demonstrate that freely available Landsat data can be used to support land-cover mapping initiatives so as to support GHG inventory in developing countries. We produced a mosaicked land-cover map of the study area with considerably acceptable overall accuracy of 86%. No specific established accuracy thresholds for land-cover mapping exist. Though, (Anderson, 1976; Thomlinson et al., 1999) agree on a target of an overall accuracy of 85% with (Thomlinson, et al., 1999) setting a threshold of not less than 75% accuracy per class. Moreover, acceptability of accuracy values lies solely on user requirement. In this study, only one class, wetland, had a user accuracy value below the 75% threshold. In the study area, most trees do exist in wetland areas i.e. along the river basin and within wet areas. Therefore, this explains why there was significant misclassification of wetland and forestland.

It is evident from the results that MLC coupled with decision tree classification produce acceptable land-cover maps. Computational efficiency and reliability are some of the advantages of the method. The approach is also replicable and thus can be used to build capacity of national GHG land-cover mapping agencies. Such replicable mapping approaches are important in obtaining complete and/or comparable land-cover both national wide and regionally. Developed land-cover maps will provide baseline information for determination of carbon credits to developing countries. Our future research will investigate the challenge of mosaicking national land-cover map across different UTM zones.

6 ACKNOWLEDGEMENT

We are indebted to Regional Center for Mapping of Resources for Development (RCMRD) where land-cover mapping for GHG inventory project is on-going and the GHG team at RCMRD who we shared ideas together.

7 REFERENCES

Anderson, J. R. (1976). A land use and land cover classification system for use with remote sensor data. *2001*(964).

Bhatta, B. (2010). *Analysis of urban growth and sprawl from remote sensing data*: Springer.

Intergovernmental Panel On Climate Change (IPCC). (2003). Good Practice Guidance for Land Use, Land-Use Change and Forestry. *IPCC National Greenhouse Gas Inventories Programme 2007*, 632.

Irons, J. R., & Dwyer, J. L. (2010). An overview of the Landsat Data Continuity Mission. 769508-769508.

ITC. (2010). *GI science and earth observation : a process - based approach* Enschede: University of Twente Faculty of Geo-Information and Earth Observation (ITC).

Kenduiywo, B. K. (2012). *Detection of built - up areas in SAR and Aster images using conditional random fields.* University of Twente Faculty of Geo-Information and Earth Observation (ITC), Enschede.

Kenduiywo, B. K., Tolpekin, V. A., & Stein, A. (2012). Detection of built-up area expansion in ASTER and SAR images using conditional random fields. 85370N-85370N.

Knorn, J., Rabe, A., Radeloff, V. C., Kuemmerle, T., Kozak, J., & Hostert, P. (2009). Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sensing of Environment, 113*(5), 957-964.

Matinfar, H. R., & Roodposhti, M. S. (2012). Decision Tree Land Use/Land Cover Change Detection of Khoram Abad City Using Landsat Imagery and Ancillary SRTM Data. *Annals of Biological Research, 3*(8), 4045-4053.

Metz, B. (2007). *Climate Change 2007-Mitigation of Climate Change: Working Group III Contribution to the Fourth Assessment Report of the IPCC* (Vol. 4): Cambridge University Press.

Mthembu, L., & Marwala, T. (2008). *A note on the separability index*. Johannesburg, South Africa: University of the Witwatersrand

Punia, M., Joshi, P. K., & Porwal, M. C. (2011). Decision tree classification of land use land cover for Delhi, India using IRS-P6 AWiFS data. *Expert Systems with Applications, 38*(5), 5577-5583.

Sexton, J. O., Urban, D. L., Donohue, M. J., & Song, C. (2013). Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. *Remote Sensing of Environment, 128*(0), 246-258.

Swain, P., & King, R. (1973). Two effective feature selection criteria for multispectral remote sensing.

Thomlinson, J. R., Bolstad, P. V., & Cohen, W. B. (1999). Coordinating Methodologies for Scaling Landcover Classifications from Site-Specific to Global: Steps toward Validating Global Map Products. *Remote Sensing of Environment, 70*(1), 16-28.

Tso, B., & Mather, P. M. (2009). *classification methods for remotely sensed data* (Second ed.). London: Taylor and Francis Group.

USGS. (2013). Earth Resources Observation and Science Center (EROS). Retrieved 15th May, 2013, from http://glovis.usgs.gov/

Wulder, M. A., White, J. C., Goward, S. N., Masek, J. G., Irons, J. R., Herold, M., et al. (2008). Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment, 112*(3), 955-969.